



**ISO 9001:2015
Certified**

**NBA
Accredited**

**NAAC
Accredited**

www.facebook.com/mcw.tcet

www.twitter.com/multiconw



CRTCE 2018

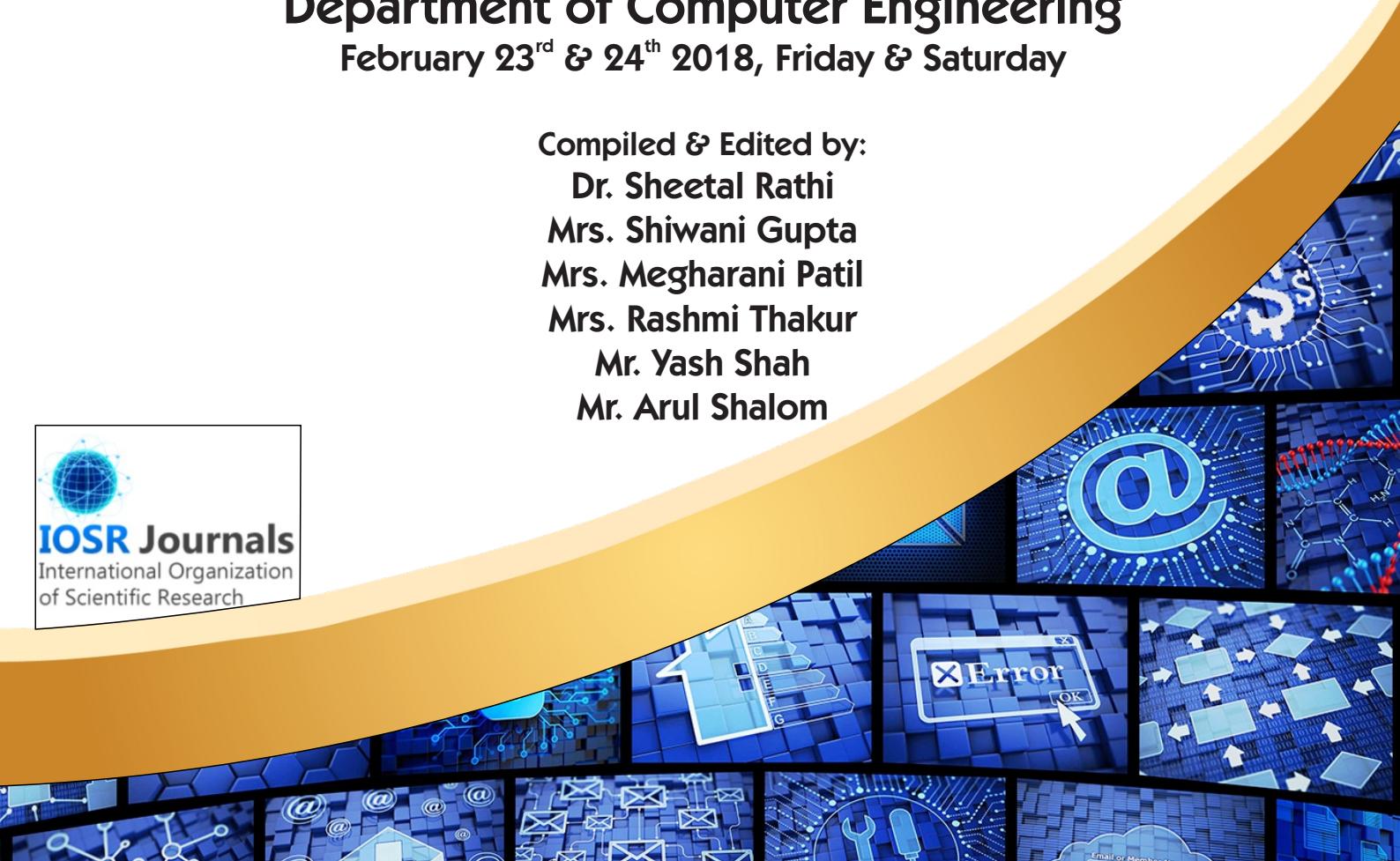
Proceedings of **Conference on** **Recent Trends in Computer Engineering**

Organized by,

Department of Computer Engineering
February 23rd & 24th 2018, Friday & Saturday

Compiled & Edited by:

**Dr. Sheetal Rathi
Mrs. Shiwani Gupta
Mrs. Megharani Patil
Mrs. Rashmi Thakur
Mr. Yash Shah
Mr. Arul Shalom**



Zagdu Singh Charitable Trust's (Regd.)

THAKUR COLLEGE OF ENGINEERING & TECHNOLOGY

(Approved by AICTE, Govt. of Maharashtra & Affiliated to University of Mumbai*)

Institute Accredited by National Assessment and Accreditation Council (NAAC), Bangalore#

• ISO 9001 : 2015 Certified • Accredited Programmes by National Board of Accreditation, New Delhi**

*Permanent Affiliated UG Programmes : • Computer Engineering • Electronics & Telecommunication Engineering • Information Technology (w.e.f. A.Y. 2015-16)
• Electronics Engineering (w.e.f. A.Y. 2017-18)

**1st time Accredited UG Programmes : • Computer Engineering • Electronics & Telecommunication Engineering • Information Technology (3 years w.e.f. 16-09-2011)

**2nd time Accredited UG Programmes : • Computer Engineering • Electronics & Telecommunication Engineering • Information Technology • Electronics Engineering (3 years w.e.f. 01-07-2016)

1st cycle of Accreditation : • "A" Grade for 5 years (w.e.f. 30-10-2017)



ISO 9001:2015
Certified

NBA
Accredited

NAAC
Accredited

CRTCE - 2018 Conference Proceedings



MULTICON-W 2018

In Association with





In Association with



"Conference on Recent Trends in Computer Engineering(CRTCE 18)"

Chief Patron

Mr. V. K. Singh, Chairman, TEG

Patron

Mr. Jitendra R. Singh, Trustee, TEG

Mr. Karan V. Singh, CEO, TCET, TIMSR, TIMSCDR

Program Chair

Dr. B. K. Mishra, Principal, TCET

Advisory Committee

Dr. Bernard Menezes, Professor, KReSIT, IIT Bombay

Mr. Sanjay Sinha, (COO, UNOTECH Software Ltd)

Dr. Vivek Bharatiya (TCS)

Mr. Saurabh Srivastava, Assistant Vice President, Reliance JioInfocomm Ltd

Technical Program Committee

Mr. Gautam Mishra, Sr. S/W Developer, IBM, Boston, USA

Mr. Sunil Pawar, Business Analyst, Tech Mahindra, Pune

Ms. Shruti Gawade, S/W Developer, Infosys Ltd. U.K.

Mr. Gajanand Pund, Sr. S/W Developer, Cape Gemini, U.S.A

Prof. Giovanni Saggio, Member and Principal Investigator

of Italian Space Biomedical Society, Promoter and coordinator of HITEG Lab

Prof. Luigi Bianchi, DICII

Dr. Vito Errico, Researcher

Dr. Ramjee Prasad, Founding Director, CTIF and Prof. Arhus Univ. Denmark

Program Co-Chair (CRTCE2018)

Dr. R. R. Sedamkar, Professor, CMPN Dept

Convenor

Dr. Sheetal Rathi, Associate Professor, CMPN Dept.

Co-Convenor

Mrs. Shiwani Gupta, Assistant Professor, CMPN Dept.

Mrs. Megharani Patil, Assistant Professor, CMPN Dept.

Mrs. Rashmi Thakur, Assistant Professor, CMPN Dept.

Internal Review Committee

Dr. Rekha Sharma, Associate Professor, CMPN Dept

Mrs. Harshali Patil, Assistant Professor, CMPN Dept

Mr. Anand Khandare, Assistant Professor, CMPN Dept

Treasurer

Mr. Vikas Singh, Assistant Professor, CMPN Dept.

Publication Committee

Mr. Yash Shah, Lecturer, CMPN Dept.

Mr. Arul Shalom, Lecturer, CMPN Dept.

External Review Committee

Dr. Subhash Shinde, Dean Academic, LTCOE

Dr. Narendra Shekokar, Professor and HOD CMPN, DJSCOE

Dr. Nupur Giri, Professor and HOD CMPN, VESIT

Dr. Dhananjay Kalbande, Professor and HOD CMPN, SPIT

Dr. Amiya Tripathi, Dean R&D, DBIT

Dr. Madhuri Rao, Professor and HOD IT, TSEC

Dr. Tanuja Sarode, Professor and HOD CMPN, TSEC

Dr. Sunil Surve, Professor and HOD CMPN, Fr. Conceicao Rodrigues College of Engineering

Thakur College of Engineering & Technology, Mumbai, India

Copyright©2018 by the International TCET. Permission to make digital or hard copies of portions of this work for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page, Copyrights for components of this work owned by others than above mentioned bodies must be honored, Abstracting with credit is permitted.

To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and or a fee. Request permission to republish from : Publications Depat. TCET. For other copying of articles that carry a code at the bottom of the first or last page, copying is permitted provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center.

Notice to Past Authors of Multicon-W Published Articles

TCET intends to create a complete electronic archive of all articles and or other material previously published by TCET, if you have written a work that was previously published by TCET in any journal or conference proceedings prior to 2015, and you do NOT want this work to appear in the TCET digital Library, please inform multiconw@thakureducation.org, stating the Title of the work, the author(S), and where and when published.

ISBN Numbers

ISBN : 978-0-9994483-0-4

Printed in India

Thakur College of Engineering & Technology, Mumbai India

PREFACE

It is a great honor for us to write a message for the CONFERENCE ON RECENT TRENDS IN COMPUTER ENGINEERING (CRTCE 18) scheduled during Feb.23-24, 2018.

This is the ninth event in the series of international conferences organized by TCET since the first event ICWET 2010. This event is organized in view of strengthening and achievement of, Vision and Mission statements (VMS) of the Institute; Technological need of the time Objective behind VMS is to provide the quality education of inter-national standards so that the students graduating from TCET must get the recognition as successful scientists, professionals, and business leaders. Moreover, need of the time is technological development which can simplify our life in the eco-friendly environment with better connectivity and security. The 9th Annual International Conference, CRTCE 2018 serves as a premier platform that gathers all academicians, researchers and professionals, in the relevant engineering disciplines and domains, to present their research contributions and share their thoughts. CRTCE 18 has gained wide publicity through website, media coverage, and visits to various colleges by the team of faculty members and our well-wishers. As a result this event has got over-whelming response.

TCET has strong belief in quality and relation building. Therefore, a lot of care has been taken for branding the event, identifying the resources, logistic support required for the event, compilation and printing of conference proceeding, souvenir etc. Effort has been taken to make the delegates feel at home though away from home with the confidence of doing something for the development of nation.

TCET has successfully held eight international events in a row. ICWET 2010, 11, 12 and Multicon-W 2013, 14, 15,16 and 17 in association with international journals of repute and multinationals in the IT industry with an objective to reach a new horizons. We feel that the present endeavor could not have been possible without the moral support, strong belief and time to time motivation from management of Thakur Educational Group

During the two days there will be about 100 presentations from national as well as international academicians and industrial personnel and deliberation by the delegates and the resource persons. Our best wishes and good luck to all.

We thank to all the members of the organizing and editorial committee for supporting the event and extending their cooperation to make it a grand successful event.

Dr. Sheetal Rathi
Convenor - CRTCE 2018

Contents

Sr.No.	Paper ID	Title & Author	Page No.
<i>Advanced Computing and System Design</i>			
1	9	Cloud Based Face Detection and Analysis <i>Saurabh Indoria, Kalindi Awasthi, Shubham Chhaparia, Vidyadhari Singh</i>	1-4
2	38	Internet of Things (IoT) In The Smart Automotive Sector: A Review <i>Ruhi Kiran Bajaj, Madhuri Rao, Himanshu Agrawal</i>	5-10
3	67	Privacy Challenges Internet Of Things Applications <i>Shiwani Gupta, Priyankar Tiwari, Loukik Salvi</i>	11-16
4	73	Hacking In Blind: Use Of User Interface As Attack <i>Aishwarya R Verma, Snehal T Patil, Sheetal Rathi</i>	17-22
5	98	QR Code Based Authentication System for Banking <i>Ashwini Patil, Nihir Shah, Ashlesh Patel, Pragnesh Patil, Harsh Shah</i>	23-26
6	112	Smart City Traveler On Android <i>Suyash Singh, Abhijeet Prasad, Tanmay Patil, Pratik Singh, Manish Rana</i>	27-28
7	118	Android Controlled Fire Fighting Robot <i>Chinmay Verlekar, Pratik Patel, Manoj Dixit, Aakash Kanojiya, Vaishali Nirgude</i>	29-32
8	133	Blockchain Technology and its application in Retail <i>Kanishk Verma, Vikas Singh</i>	33-36

9	134	Design and Development of Pulse Oximetry for Continuous Monitoring of Pregnant Ladies using Arduino <i>Naziya Pathan, Mukti E. Jadhav</i>	37-40
10	10	Cloud Based Vehicle Monitoring System <i>Amit Dubey, Jainil Jain, Vamsi Akhumukhi, Rahul Gupta, Vidyadhari Singh</i>	41-46
11	25	Smart Ambulance System with Traffic Overriding <i>Chaitanya Zarekar, Sandeep Pardeshi, Mayannk Kumaar, Lydia Suganya</i>	47-52
12	50	Android Application for Conference Alert <i>Deshmukh Ravish Zafar, Khan Darakhshan Zain, Shaikh Faizan, Ansari Bilal ZakirAli</i>	53-54
13	53	Technical Approach for Agriculture <i>Musa Asmat Pawle, Nilofer Peerzade, Atif Thakur, Meenaz Gigani</i>	55-58
14	66	Cloud-based Core Banking Software for Small Finance Institutes <i>Manjiri Bhat, Nishant Shishodia, Vikram Rajpurohit, Narayan Mundhara, Anjali Yeole</i>	59-60

Intelligent Systems, Multimedia and Network Security

15	74	A Scalable Hybrid Network Control Plane for Cloud Computing <i>Shreyash Sharma, Neha Kamble, Anand khandare</i>	61-64
16	79	Optimizing AODV Protocol for Preventing Black Hole Attack On MANET <i>Saurabh Bagde</i>	65-68
17	82	Design of Secure Robustness Speaker Authentication System <i>Amit Tiwari</i>	69-72
18	97	Data Mining for Internet of Things <i>Deepali Joshi</i>	73-74

19	99	Scalable Data Reduction Technique in Cloud Storage <i>Ankur Sinha, Anand Khandare</i>	75-80
20	110	Quantum Key Distribution: An Application of Quantum Cryptography <i>Yash Shah, Vidyadhari R. Singh</i>	81-84
21	111	An Image Encryption and Decryption Using Chaotic Technique <i>Anita A. Lahane, Pranjali Sankhe, Surabhi Singh, Shruti Pimple</i>	85-88
22	120	Multi Touch Technology Interface and Computer Interactions <i>Swastik Chavan, Salonee America, Meher Dodhia, Mahima Jain</i>	89-92
23	125	Polymer Memory - Use of Plastic As Secondary Storage <i>Vipul Agarwal, Adwait Arayakandy, Ankit Gupta, Kalpana Gangwar</i>	93-96
24	128	Intelligent Traffic Light System(ITLS) <i>Akib Maredia, Rishabh Mishra, Ashutosh Mishra, Nitin Harane</i>	97-100
25	131	Study on types of Trojans in Security <i>Veena Kulkarni</i>	101-104
26	6	Billing System Using Raspberry-Pi <i>Shivam Mishra, Ritesh Mishra, Bharat Joshi, Rashmi Thakur</i>	105-110
27	7	Enhanced Classification Framework <i>Rishi Shah, Chandrashekhar Bora, Amit Mehra, Yash Kaul, Harshali Patil</i>	111-116
28	11	Enhanced System for Delivery and Efficient Path Finding <i>Pratiksha Chourasia, Shivam Pandey, Simran Malawat, Shradha Mhatre, Harshali Patil</i>	117-120

29	13	Canteen Management System Using Payment Gateway <i>Ashish Kothari, Garima Choudhary, Dhruvi Bhagat, Rashmi Thakur</i>	121-124
30	19	Smart Automated Computer Education Management System <i>Ravishankar Singh, Pratik Singh, Swaraj Kasliwal, Veena Kulkarni</i>	125-128
31	39	Graphical Approach for Frequent Pattern Mining <i>Anupriya Babbar, Vipul Babbar, Tanushsree Bhattacharjee</i>	129-132
32	101	School Management ERP <i>AnandKhandare, Raj Shah, Vishwa Solanki, Aakash Patel, Abhishek Anand</i>	133-136
33	12	Roadways Safety Suggestions <i>Deep Suchak, Nishant Shishodia, Vikram Rajpurohit, Kunal Singhal, Prachi Janrao</i>	137-140
34	20	Content Management System Using Web Technology <i>Shubham Vishwakarma, Saurabh Singh, Rahul Singh, Ajaykumar Verma, Sheetal Rathi</i>	141-144
35	29	Real Time Monitoring and Control in Irrigation System Using Iot <i>Vishal Rajak, Ninad Virkar, Anurag Shetty, Prakash Shukla, Veena Kulkarni</i>	145-146
36	45	Agile in Automotive Manufacturing <i>Tasneem Mirza, Archana Patankar</i>	147-150
37	49	Accident Detection and Alert System <i>Prashant Kapri, Shubham Patane, Suraj Maurya, Abhimanyu Singh, Arul Shalom</i>	151-154
38	52	News Validation System <i>Sufiyan Pawaskar, Nooralam Shaikh, Pradnyesh Rane, Tabrez Khan</i>	155-158

Information Management and Data Analytics

39	81	Exam Cell Automation System and Timetable Generator <i>Dnyaneshwar Dhangar, Aishwarya Sontakke, Jhanvi Shah, Namrata Bodas, Yash Shah</i>	159-162
40	84	Decision Making in Agriculture Using Data Analytics <i>Sudha Shankar, Madhuri Rao</i>	163-166
41	86	Shared Nearest Neighbor Approach for Clustering Documents <i>Kalpana Gangwar</i>	167-170
42	105	Automated Village - A Way to Progress <i>Pratiksha Singh, Tanya Talwar, Shaheen Sayyed, Tanvi Sanzgiri, Anand Khandare</i>	171-174
43	126	Augmented Situation Awareness and Intelligence Using Data Mining <i>Rohit Shashi Dubey, Akash Mishra, Sunil Yadav, Devyashwar Raj</i>	175-178
44	62	Artificial Limbic Brain Models Using Neural Networks for Emotional Analysis <i>Akanksha Joshi, Anushri Arora</i>	179-184
45	71	How to explore the Art of Computer Security <i>Vinitta Sunish, Drashti Shrimai, Veena Kulkarni</i>	185-190
46	75	Text Preprocessing and Polarity check on Social Media Data <i>Prajakta Samant, Ashuta Thakur, Harshali Patil</i>	191-196
47	77	Database Access for Non-Technical Users Using NLP and Voice Recognition <i>Anuradha Maurya, Aishwarya Manjalkar, Rashmi Thakur</i>	197-200
48	87	Multimedia Recoloring Technique for Propanopic CVD <i>Nazneen A. Pendhari</i>	201-204

49	90	Machine Learning Algorithm Analysis <i>Sruthi Harish, Ishita Sharma, Sakshi Singh, Manisha Sharma, Megharani Patil</i>	205-210
50	96	A Review Paper On Automatic Text Summarization Methods <i>Vaishali Nirgude, Vikas Nirgude</i>	211-216
51	104	Handwritten Pattern Recognition Using Artificial Neural Network <i>Ruchit Tripathi, Jenil Shah, Vishal Singh, Sudarshan Dasarathy, Rekha Sharma</i>	217-220
52	4	Commodity Search Engine for Online Shopping Using Web Datasets <i>Saurabh Rai, Varun Shah, Drishti Sisodiya, Riddhi Vaidya, Shiwani Gupta</i>	221-224
53	17	Monitoring and Assessment of Daily Activities Using IOT and Machine Learning <i>Aditya Sarvankar, Omar Dalvi, Jonathan Kumar, Shailesh Sangle</i>	225-228
54	26	Home Automation System Using Raspberry Pi <i>Uday Menon, Rahul Yadav, Shubham Singal, Kaustubh Shah, Vikas Singh</i>	229-232
55	32	Real Time Drowsiness Detection System <i>Aakashkumar Yadav, Rishabh Rathod, Bharat Rathore, Abhishek Yadav, Shailesh Sangle</i>	233-236
56	33	Secured ATM Management using Image Steganography <i>Piyush Akoliya, Rushikesh Shinde, Sakshi Singh, Sujay Salvi, Ruta Pathak</i>	237-240
57	46	Natural Language Processing Based SQL Engine <i>Abhijeet Saha, Jaineel Parmar, Pranit Patil, Tanushree Bhattacharjee</i>	241-244
58	55	Business Intelligence and Business Ratting System <i>Ali Akbar Nathani, Shaikh Mehboob Siraj, Aavez Kazi, Shaikh Sharukh, Prof. Sameer Panwala</i>	245-246

59	78	Security of Multimedia Data Using Hybrid Model Gautam Kumar Jha	247-250
60	95	Proposed Improved Text Summarization Method Using Adaptive Neuro Fuzzy Inference System Pratiksha Warule, Sudhir Sawarkar, Archana Gulati	251-254
61	103	Use of Long Short-Term Memory for Enhancing Bi-Lingual Machine Translation Approach Manish Rana, Mohammad Atique	255-258
62	106	Activity Recognition Based Machine Learning Models Keval Chavda, Viral Haria, Bijal Patel, Anupriya Babbar	259-262
63	121	Incorporating Copying Mechanism in Sequence-to-Sequence Learning Gautam Mishra, Sansriti Mishra, Anand Khandare	263-266
64	122	Auto Text Summarization Dnyaneshwar Dhangar, Richa Tanna, Akanksha More, Ankita Thakur	267-270

Appendix

Author Index

Cloud Based Face Detection And Analysis

Saurabh Indoria¹, Kalindi Awasthi², Shubham Chhaparia³, Vidyadhari Singh⁴

Computer Engineering, Thakur College of Engineering and Technology, Mumbai, India

Abstract—Surveillance devices capture a lot of information but still it requires a human to interpret data. This wastes a lot of potential usage of the valuable data collected. If instead, there is a solution which reduces the human effort required in the process, the efficiency and the utilization of the surveillance system will increase. The main idea behind this is to deploy an application server on AWS cloud which will be connected to a Mobile Application which will be working as a surveillance device. Due to the use of AWS Cloud, the backend costs would be reduced by a significant level. To reduce the development time, face recognition libraries will be used. This system can be used for various applications such as office premises, educational institutions, or other premises.

Keywords—AWS, Cloud, Face recognition, Surveillance, Serverless, Server-based.

I. INTRODUCTION

Security devices have been increasingly installed in various forms of businesses and institutional premises. The most common of them all are the surveillance devices such as CCTVs. All kinds of premises usually have them to keep a check on the activities of the areas remotely monitored by a human agent. This itself introduces a point of failure as human error. Humans cannot be always efficient in what they do, and also, they have a limited scope of work. The man hours required for increasing the utilization of such complex CCTV systems incur an additional burden on the patron. Hence, in many scenarios, surveillance devices are often just used to record events in case they are used in future reference. In general, a surveillance system would consist of the following components:

- A number of VGA Night Vision Camera devices.
- Network Video Recorder (NVR).
- Ethernet backbone network.
- Internet routers for remote monitoring and storage (Optional).
- Monitoring Stations (Optional).

In many places, it is found that IP cameras are being used which communicate through the internet to the remote servers and allow remote monitoring from any part of the world. Also taking into consideration the increasing internet expansion in India, we can infer that all surveillance systems would likely be connected to the internet in near future. As of the current scenario, the internet connectivity is only used for allowing the security systems to store the data on remote servers and also enabling the remote monitoring of the systems. However, since the data is being collected in such a huge quantity, security is not the only aspect which could be targeted. Processing data locally would be a challenging task, but since it is being sent over the internet to remote servers, we can think of other ways which can affect the way the data is utilized. Considering the fact that the surveillance

systems are connected to the internet, we can come to an inference that having specialized data processing servers in the cloud can help utilize the data in a more efficient way. Till date, surveillance systems have been used for pure monitoring purposes, but we think that we can expand the scope of these systems serving various other purposes as the data collected is huge and can be utilized in various ways. Reducing the human work in the process and increasing utilization is the main goal behind this project. Using the powers of AWS Cloud and its services combined with the surveillance systems, we aim at solving real life problems put up by various renowned institutions.

II. PROPOSED METHODOLOGY

The main objective of this project is to provide a working proof of concept for the advanced monitoring system. Mobile devices would be used as monitoring devices. Application servers would be deployed on cloud.

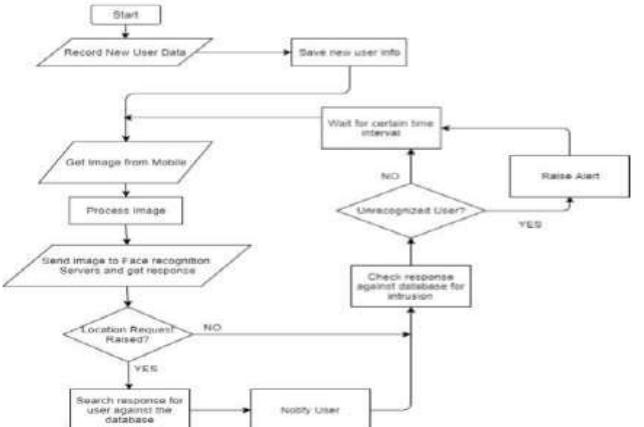


Figure I. Flowchart

The basic flow of the project would be as such:

- Step 1: Mobile device connects to remote servers.
- Step 2: At fixed intervals, the device sends an image up to the cloud.
- Step 3: These images are run through the face recognition servers.
- Step 4: The face ids are retrieved and are checked against the database for intrusion detection.
- Step 5: If some room gives a paging request for an individual, the application servers get in touch with all the mobile devices (monitoring devices).
- Step 6: A signal is sent to mobile device to send an image of the area its monitoring.
- Step 7: All the images are run through face recognition servers and the corresponding face ID is determined by checking the individual against the database.

Step 8: Appropriate alarm is raise and notification is sent.
 Step 9: All the activity is logged into the database for analysis purposes.

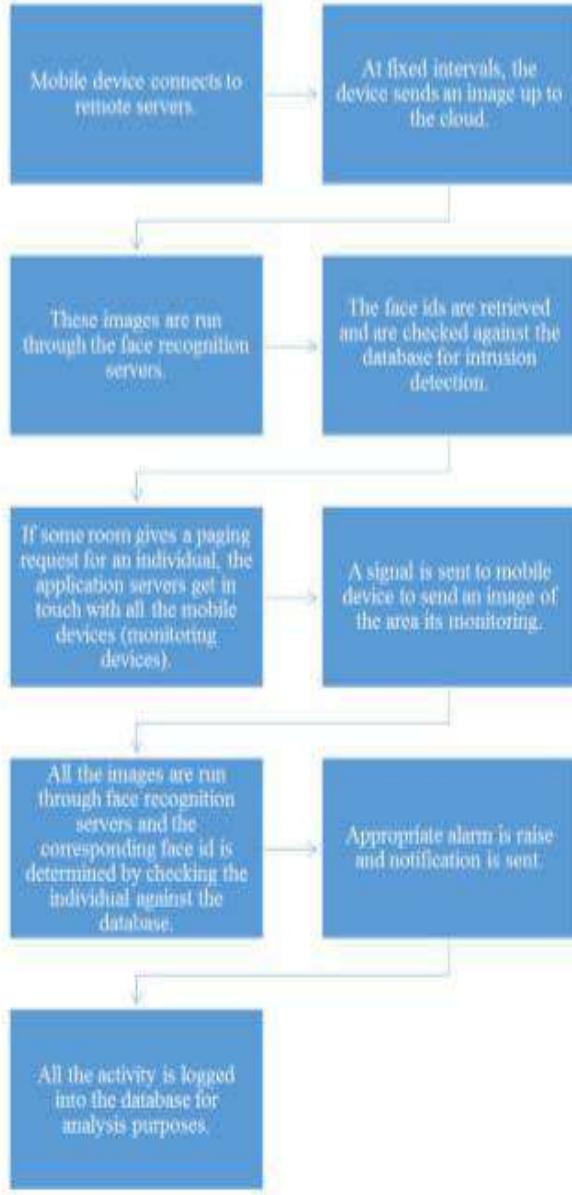


Figure 2. Methodology

III. SYSTEM ARCHITECTURE

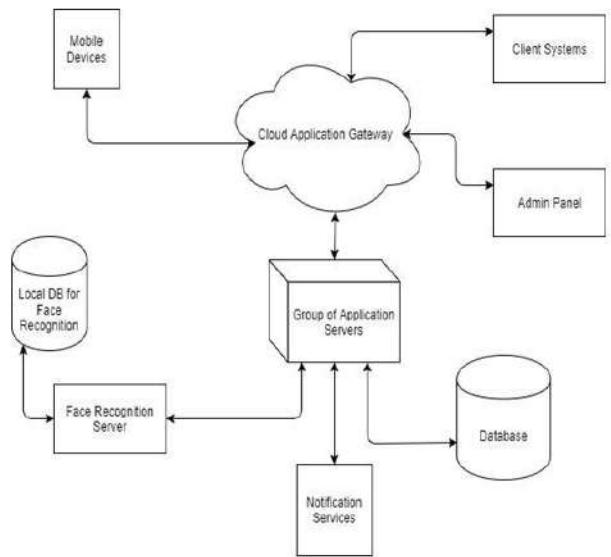


Fig 3. Architecture

The structural model is described as a general software architecture model. Various modules and their functions are listed below:

1. Mobile Device: It will act as the data collection device, or simply, a CCTV.
2. Client System: This is basically the interface for the alert to be raised from client side to locate a person.
3. Admin Panel: Overall system control and monitoring panel.
4. Cloud application gateway: All the external connections would be through the gateway device.
5. Group of application servers: These will be clusters of instances running to serve the requests and would be under an auto scaling group.
6. Face Recognition Servers: These will be processing the image received from the application server and will return back the response containing the detected and recognized faces.
7. Notification Services: For sending various forms of notifications such as SMS, email, etc.
8. Database: For storing user information, logs, etc.

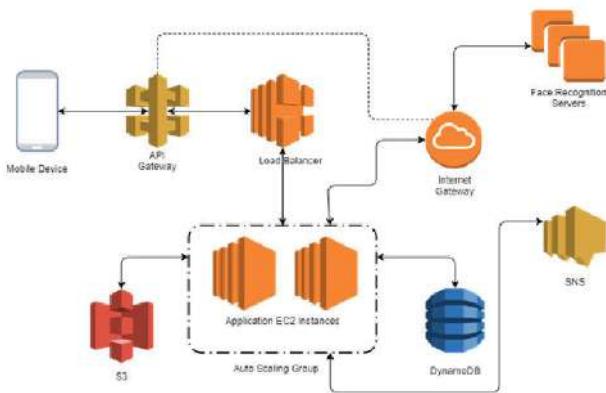


Figure 4. System Architecture

The above is a proposed system architecture using public cloud infrastructure from AWS. Various services are described below:

1. API Gateway: It handles the API requests and routes them to proper application servers. It also manages usage quota, API keys, etc.
2. Load Balancer: It distributes the incoming traffic among the underlying instances using various algorithms.
3. EC2 Instances: Elastic compute cloud instances are basically individual instance of a server, which are used in clusters for load balancing.
4. Auto Scaling Group: This functionality basically monitors the load on the cluster of instances and can be configured to scale up or scale down horizontally or vertically based on various performance metrics.
5. S3: Simple storage service is a key-value store service used to store flat files. This can be used to store images, logs, metadata, etc.
6. DynamoDB: It a fully managed, NoSQL database which concurrently replicates the database in 2 regions, each having 3 availability zones, thus creating 6 redundant copies of the data. It also can be configured to auto scale based on the amount of database traffic being generated. Failover is managed by AWS only.
7. SNS: Simple notification system is used to send SMS, email notifications to users.
8. Internet Gateway: It is used to connect the VPC (Virtual private cloud) to the outside world via internet. ACLs are defined to restrict the type of traffic flow for security purposes.

IV. SCOPE

1. Location Tracking: Determining the location of individuals in a high security premise.
2. Head Count: Determining or approximating the number of individuals present at a given place at a given time.

3. Intrusion Detection: Recognizing unauthorized personnel trying to enter or forcefully entering a secured zone with restricted access and raising appropriate alarm.
4. Authorization: Uninterrupted access using facial recognition for authorization.
5. Management Reports: Generate reports based on certain analysis.

V. Applications

1. High Security Premises:
It can be deployed in high security premises for automated intrusion detection.
2. Business Institutions:
It can be used to authorize and authenticate access to secured locations without any hindrance.
3. Research Institutions:
It can be used to track location of important personnel at the time of emergencies.
4. Educational Institution:
It can be used for attendance management.

VI. DISCUSSION

Since the naive systems are not capable of using the collected data as efficiently as our system, the following will be the advantages and disadvantages of the proposed system.

Advantages:

- Less human intervention.
- Reduction in operating costs.
- Lower chances of human error.
- Wide scope to expand functionalities.
- Unlimited capacity.
- Multiple use cases where the system can be implemented.

Disadvantages:

- Depends on IP based surveillance infrastructure.
- Chances of seldom false alarms.
- Requirement of some level of human assistance, not fully automated.

We also tried to discover the server-based and serverless architectures for this purpose. The findings and observations are explained below:

Serverless: The main advantage here is that you do not need to worry about load balancing, auto scaling policies, etc. Serverless architecture uses AWS Lambda, which basically runs a piece of code on demand, and you are charged only for the time duration the code is run. So, for a less frequent application, this benefits as there is no upfront cost or extra servers which costs you money. In case of large scale applications, Lambda is again cost effective as you can scale up instantly to any level, without incurring much cost, as Lambda is relatively cheap. However, the downside is that the entire application has to be managed on stateless requests as sessions, etc. cannot be used in Lambda. There are other services which we can use in combination to tackle this problem, but, it is not so straight forward to transfer a server-based application to a serverless architecture, it requires complete redesign of the entire system architecture.

Server-Based: Here, you need to manage the auto scaling policies, failovers, load balancing, etc. There are few services which can take the load off your shoulders, but still, it requires a lot of administrative help. The advantage here is, compute intensive applications can be seamlessly run on server-based architecture, and large tasks can be completed collaboratively by smaller instances. However, for general application development, we have noticed that serverless architecture suits well for mobile applications, and is beneficial in terms of both cost and quality.

V. CONCLUSION

Security is a rising concern for many institutions and simply installing monitoring devices won't help solve the issue completely. Also, it is practically infeasible to manually monitor each and every location all the time. Looking at the increasing pace of internet connectivity expansion, and considering the fact that more areas would be needing surveillance in the coming years, we proposed the idea of a cloud based application that will use face recognition and will be able to assist humans in scenarios like burglary, etc.

REFERENCES

- [1] Jinesh Varia, "Architecting for the cloud: Best practices", amazon web services white paper Jan 2011.
- [2] Khan R.Z., Ahmad M.O. (2016) Load Balancing Challenges in Cloud Computing: A Survey. In: Lobiyal D., Mohapatra D., Nagar A., Sahoo M. (eds) Proceedings of the International Conference on Signal, Networks, Computing, and Systems. Lecture Notes in Electrical Engineering, vol 396. Springer, New Delhi
- [3] Kevin Hamlen, Murat Kantarcioglu, Latifur Khan, Bhavani Thuraisingham," Security Issues for Cloud Computing", International Journal of Information Security and Privacy, 4(2), 39-51, April-June 2010.
- [4] Shang-Hung Lin," An Introduction to Face Recognition Technology", Informing science special issue on multimedia informing technologies-part 2 volume 3 no 1-2000.
- [5] Monjur Ahmed and Mohammad Ashraf Hossain," CLOUD COMPUTING AND SECURITY ISSUES IN THE CLOUD", International Journal of Network Security & Its Applications (IJNSA), Vol.6, No.1, January 2014.

INTERNET OF THINGS (IoT) IN THE SMART AUTOMOTIVE SECTOR.

A REVIEW

Ruhi Kiran Bajaj, Madhuri Rao, Himanshu Agrawal

*Thadomal Shahani Engineering College, Bandra, Mumbai-India,
ruhibajaj2007@gmail.com, my_rao@yahoo.com, himanshu.agrawal@sitpune.edu.in*

Abstract - The Internet of Things (IoT) is a new wave of Internet that is expected to transform our lives. Internet has connected people and now it is connecting ‘Things’ to make seamless communication and intelligence pooling. IoT is a disruptive technology that has tremendous potential to change the world and is changing the way we live. It uses low cost Internet connected devices and sensors that create new possibilities. Not long ago, the idea of IoT in the automotive sector was being seen as a futuristic theoretical concept and today we are already seeing possibilities of connected cars, driverless cars and application of IoT in the car ecosystem covering parking, environment, supply chain and transport governing bodies. This paper discusses the evolution and developments on Internet of Things with regard to the automotive sector to provide a perspective on the various areas such as- Connected Car services/applications, Vehicle communications, IoT in Intelligent Transportation, IoT based Supply Chain Management in Automotive Industry and New Generation Cars, where tangible progress is being made.

Keywords - Internet of Things(IoT), Automotive Industry, Connected Cars.

I. INTRODUCTION

IoT is a disruptive technology where cyber world meets the physical world. It is autonomous communication between inanimate objects [28], to benefit human beings. IoT encompasses all technologies in SMAC (Social, Mobile, Analytics and Cloud). Automotive industry is on course to a disruptive transformation using developments around smarter vehicles and related infrastructure. IoT is at the heart of this digital transformation in Auto sector. It connects people, machines, vehicles, auto parts, and services to streamline the flow of data, enable real-time decisions, and improve automotive experiences [25]. Leading automotive manufacturers, suppliers, and dealers have started investing heavily in Internet of Things and are gaining returns in the form of ultra-efficient inventory management, real-time promotions that grow sales, reduced operational expenses and increase in revenue. They are beginning to change their business processes and recognize that, in time, IoT will touch every area of automotive operations and customer engagement [27]. Tesla motors is a big example in the Internet of Things domain. In addition to being a luxury car, performance vehicle, the car is one big IoT device with a lot of IoT properties inbuilt. Applications are developed that use data collected by connected cars [30] in many ways. For example, traffic control systems can provide real-time data collected from connected cars to avoid traffic jams and accidents, Automotive

components manufacturers can benefit from data about wear and tear to pre-order the components to be replaced and notify customers before there are equipment failures. Car sharing mobile apps can use real-time location data to encourage car pools. Also, in insurance systems, premium rates would be based on geolocation of vehicles and driving behavior of drivers. Figure I.1 shows these and many more developments on IoT in Automotive sector

A. Connected Car Ecosystem

IoT has a significant impact on Automotive industry. Automobile manufacturing companies, telecommunication service providers [33] and software companies are coming together to build the Connected Car.

A connected car is a car which using its onboard sensors and internet connectivity enhances the in-car experience of its users [3]. Connected car, just doesn't mean the capacity to surf the internet on the move, but the communication between cars, communication of cars with other devices. As of date, there are only a small number of cars which are internet enabled but it is expected that the number will rise considerably in less than a decade's time. And seeing the lifestyle experiences required by today's generation and the exposure to smartphones will definitely help the connected cars market to grow exponentially. The Connected car services and applications along with other developments on IoT in Automotive sector are discussed below.

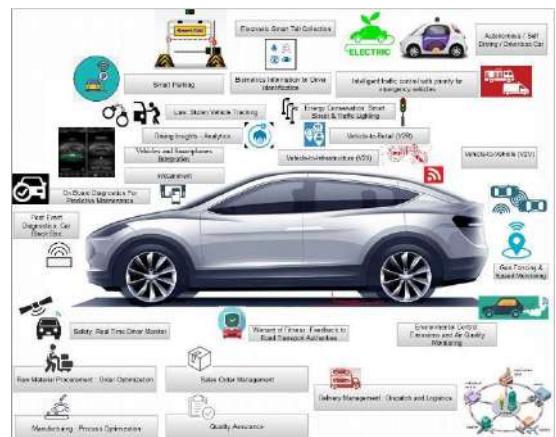


Fig 1: Developments on IoT in Automotive sector

B. Connected Car services/applications

1. INFOTAINMENT

Infotainment refers to a system in vehicles that delivers a combination of information and entertainment [9] content/services. Typical features of an In-Vehicle-Infotainment (IVI) system are – providing navigation features while driving, managing audio/visual entertainment content, delivering rear-seat entertainment, connectivity with smart phones for hands free experience with the help of voice commands [2]. Infotainment Options should create a Safer In-Car Experience [26] helping drivers keep their eyes on the road and their hands on the wheel. Infotainment options like apps and menus need to be well organized and accessing features should be very intuitive to avoid drivers' distraction. Also, voice and audio commands need to anchor the in-car app platform and drivers should be able to navigate menus or create messages using voice commands. Audio should be the primary way to accept commands from the users. With rapid growth in smart phone and Cloud technologies, consumers are demanding for live streaming of music and Internet radio. Ideas for advanced infotainment features integrate both user behavior and the next level of cloud-based infotainment systems [29]. Some of the examples of smartphone apps integration software units are Car Play, Google Projected Mode, Mirror Link etc.

2. VEHICLES AND SMARTPHONES INTEGRATION

In today's fast-paced world, people need to be both mobile and online for most of the time with no exception when it comes to being onboard of a car or even when being the driver [31]. Using the On-Board Diagnostics OBD/OBD-II port, (which is like computer which monitors emissions, mileage, speed, and other useful data) information regarding engine and other crucial vehicle parameters can be displayed on the driver's smartphones and same can be sent to service provider for analysis [2]. Alerts related to the car like Open doors, Lights ON and Hand brake ON and performing actions on certain vehicle parts such as Lock/Un-lock vehicle doors, roll windows up/down and AC temperature +/- are becoming seamless.

3. DRIVING INSIGHTS - ANALYTICS

Smartphones sensors such as GPS, Gyroscope or Orientation sensor and Accelerometer can be used to model the driving behavior. By mounting the smartphone in the vehicle, data from these sensors can be used to detect driving patterns such as sharp turns, sudden acceleration, hard braking, drifting and speeding [2]. This can be used to profile the driver as safe or aggressive, to rate and compare different drivers and share such data with insurance companies for customized premiums. Pay as You Drive (PAYD) and Pay How You Drive (PHYD) are the upcoming use based insurances packages [12] provided by auto insurance companies that reward safe drivers and penalize rash ones with different premiums.

4. ON-BOARD DIAGNOSTICS FOR PREDICTIVE MAINTENANCE

The On-Board Diagnostics (OBD/OBD-II) port is commonly used in automobile service and maintenance for self-diagnosis and reporting of any issues that may occur or have occurred within the system [12]. Using this, information such as emissions, mileage, faults, vehicle and engine speed, engine temperature, fluid levels, gear shifts, battery status, etc. can be monitored-II is connected to the Check Engine light or MIL-Malfunction Indicator Light, which

illuminates when the system detects a problem. Till now it was largely used for post-facto analysis; i.e. only when some problem arises. However, by pairing smartphones with vehicles, this information can be readily made available to the vehicle owners and service stations [2], giving them a better picture of the car performance. Monitoring these parameters actively and with some level of on-device analytics, drivers can get proactive service alerts on their smartphones and potential faults can be identified for early diagnosis and care.

5. SAFETY: REAL TIME DRIVER MONITOR

To encourage efficient and safe driving, drivers are screened and evaluated on driving habits. Advanced sensor-based technologies to detect and monitor behavior and fatigue levels of driver [43] are emerging which makes the cars more intelligent for avoiding accidents on roads. Systems are being developed for real time monitoring of vehicles [30] which controls the speed of the vehicle and fatigue level of the driver to prevent accidents [4]. The primary components of such a system, as shown in Figure I.2, will be microcontrollers along with some sensors [44] like eye blink, gas, impact sensors, alcohol detecting sensor and fuel sensors. GPS and Google Maps API's is used to track the location of the vehicle which can sent to a predefined number in the system.

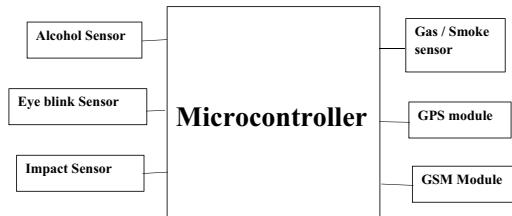


Fig 2: Real time driver monitoring system

6. GEOFENCING AND SPEED MONITORING

The geofencing and speed monitoring applications [34] can be used to inform the car owner if the vehicle has gone out of the predefined geographical area or is being driven faster than a preset threshold speed [6]. Speed of the vehicle can be measured speed sensors [45] and geofencing can be achieved with the help of GPS. This can be used for parental control or for remotely monitoring the use of company cars.

7. LAW: STOLEN VEHICLE TRACKING

GSM and GPS based Vehicle Tracking System as shown in Figure I.3 is used with a tracking device [46] which is hidden in the vehicle to monitor and track the location of vehicles [24]. Satellite signals will be received by a remotely located application server and then position co-ordinates with latitude and longitude are determined. Exact position of the vehicle can be determined from these coordinates and using the GSM system [35], owners can be notified.

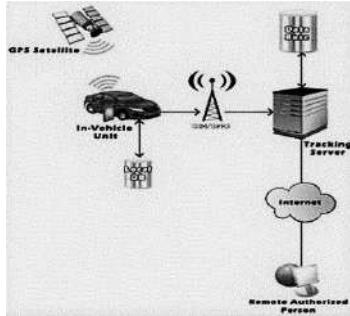


Fig 3: Vehicle Tracking System

8. BIOMETRICS INFORMATION FOR DRIVER IDENTIFICATION

Biometrics refers to the physical, biological or behavioral characteristics of a person. It can be used to identify and authenticate a driver. Biometric identifiers include face recognition [6], fingerprints or voice recognition. Voice samples [12] of a driver can also be used to provide a hands-free experience for navigating through the apps in a connected car environment. Biometric data can be used as a powerful anti-theft protection tool for increased safety. Cameras and sensors within the car can be useful for such biometric based driver authentication and also for comfort features inside the car by quickly changing car settings like seat position, mirror settings etc. to accommodate different driving patterns of the people authorized to use the car. Real time health parameters of a driver like pulse rate and breathing patterns can be monitored with the help of sensors attached on the steering wheel and seatbelt which in turn can be used to monitor their stress levels, other health conditions and prevent accidents.

C. Vehicle Communications

As more and more connected cars emerge, and in-vehicle embedded connectivity becomes common, a whole new paradigm of vehicle communications is set to unfold.

1. VEHICLE TO VEHICLE (V2V)

V2V communication comprises of a wireless network where the vehicles send speed and location data to the vehicles near-by to prevent accidents and offer the opportunity to improve the safety of commuters significantly [2]. Each vehicle continuously broadcasts a message with the speed and position data to the nearby vehicles over an ad hoc mesh network. Dedicated Short Range Communications (DSRC) [47] technology designed specifically for automotive use allows vehicles to communicate with other vehicles [12]. V2V relies on Vehicular Adhoc Network (VANET) [48] which is a wireless adhoc network of the vehicles for exchanging data.

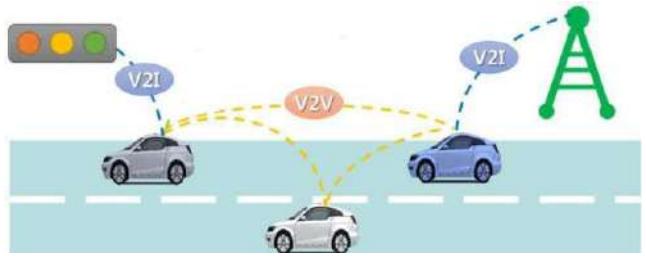


Fig 4: V2V and V2X

2. VEHICLE TO INFRASTRUCTURE(V2X)

V2X communication is the wireless transmission of information between the vehicles and roadside infrastructure [32] to avoid or mitigate accidents and also to provide a wide range of other safety, mobility, and environmental benefits. The vehicles can communicate with the roads, digital signage, traffic lights, safety and control systems [2] to avoid crashes and traffic congestion [47] through intelligent safety applications. Drivers can also be made aware of various other conditions such as roadwork, diversion and adverse weather conditions. This will not only help in reducing the number of accidents and casualties but also in tracking, tracing and monitoring vehicles on the move. Even in the event of emergency, vehicles would be able to quickly intimate details to roadside assistance, emergency services like ambulance, insurance providers as well as family members.

3. VEHICLE-TO-RETAIL (V2R)

The retail industry has been exploring the innovative use cases and presenting drivers with location-based advertising or even discount coupons while driving in the vicinity of a shopping mall. Similarly, applications to book restaurants and to place their saved ‘Easy Order’ from their vehicles through a voice interface are quite popular. OEMs are exploring other vehicle-to-retail (V2R) applications [7] such as Volvo has set up its Roam Delivery trial, which allows delivery companies to locate and unlock a connected vehicle to deposit an item purchased online.

D. IoT in Intelligent Transportation

1. ELECTRONIC SMART TOLL COLLECTION

Electronic Smart Toll Collection system on Toll plazas is used to collect the road toll charges in a convenient manner without any traffic jams and delay. This system relies on roadside sensors such as RFID readers or Automatic number plate Recognition cameras. Another telematics-based solution uses GPS [9] and cellular network to provide Electronic toll collection services. Also, smart card/tags can be used to identify vehicle details and for payment [36].

2. SMART PARKING

Traffic jams caused by vehicles is an alarming problem and is increasing day by day with the increase in number of vehicles. Lack of sufficient parking space is the main reason in this problem along with the increase in size of luxury cars. Finding a free parking space is often a tedious and frustrating task. Smart parking will optimize the use of parking space, resulting in efficient parking and smoother

traffic flow. Smart parking system can be deployed with the help of microcontrollers [37], sensors, real time data about available parking space [8] and automated payment of parking charges. Such system will allow people to reserve a parking space before they actually reach that location thereby saving a lot of time and energy. This also controls the air pollution and reduces the traffic jam problems.

3. ENERGY CONSERVATION: SMART STREET & TRAFFIC LIGHTING

Smart or Intelligent street lighting, also referred to as adaptive street lighting, dims or switched off when no activity is detected, but brightens when movement of pedestrians, cyclists and vehicles is detected. A new, innovative and optimized traffic light system as well as street light management system will achieve significant energy saving. It will use many sensors and cameras to control and improve the efficiency of the system, presence of vehicle or person can be detected by using the presence detector (IR sensor), street light glows with full intensity for some time and then becomes dim. Also, energy-efficient LED-based streetlights [38] which consumes lower energy can be coupled with providing IP connectivity and IoT based Smart Street lighting systems. Another use of this system can be to control the traffic jams. The RFID tags [10] are used for detection of emergency vehicle like ambulance. Priority can be assigned to each lane depending upon number and type of vehicles. Green traffic light will be given for the lane having highest priority and more time than another lane. This system saves more power and provide safety on the road.

4. POST EVENT DIAGNOSTICS: CAR BLACK BOX

An event data recorder (EDR) is a device for recording [6] vehicle data related to events such as an accident or a crash. Such "Black Box" technology [11] can be used as an accident investigations tool. It records vehicle data like speed, fast acceleration, heavy braking, use of seatbelt and airbag deployment [12] several seconds before, during and/or a fraction of a second after the accident. To detect and collect the information from the vehicle, many types of sensors like proximity sensor, ultrasonic sensor, pressure sensor and temperature sensor are used. To download the data recorded using an EDR some scanning tool can be connected to the vehicle and this data can be presented in an easy to read graphical form.

5. INTELLIGENT TRAFFIC CONTROL WITH PRIORITY FOR EMERGENCY VEHICLES

Intelligent traffic control systems in city traffic strives to give priority to selected type of users, such as public transport, VIP users, and emergency services [13]. Traffic light control system uses RFID tags attached to the vehicles. Priorities can be assigned to different types of vehicles. RF readers are installed on the road intersections considering the traffic density on the roads. On the two intersecting roads, two linked lights [14] are installed, RF reader will store the details of all the vehicles that pass from that road. For usual traffic, the traffic light controller follows the round robin sequence of the traffic lights. As an emergency vehicle such as an ambulance is detected, the controller generates the green signal for the lane with emergency vehicle, leaving the round robin sequence.

6. WARRANT OF FITNESS: FEEDBACK TO ROAD TRANSPORT AUTHORITIES

Warrant of Fitness (WoF) is a mandatory certification mark required on all LMV's (Light Motor Vehicle). It certifies that the vehicle has passed all the necessary inspections of safety. A WoF test checks the following parameters of the vehicle – [15] tyre and brake condition, lights, windscreen wipers, seat belts, doors, airbags, suspension, speedometer, exhaust, fuel system etc. To pass WoF test, a vehicle must meet some minimum criteria in each category. WoF is in the form of a sticker displayed on the windscreen indicating the date on which the vehicle passed its last WoF test and when it is due for the next inspection. Local car repair garages can perform this WoF test and issue WoF certificate. Traffic police can check this WoF anytime and if the WoF sticker is not found or it is expired, the driver has to pay the fine. This WoF checking can be automated using RFID based WoF stickers which can be read by RFID readers installed on roadside.

7. ENVIRONMENTAL CONTROL: EMISSIONS AND AIR QUALITY MONITORING

The problem of emission of gases from vehicles is increasing the pollution levels. Incomplete combustion of fuel is one of the main reasons of this high level of emissions. Vehicle emissions can be monitored and controlled [39] by using an IoT based system consisting of smoke and temperature sensors, along with GSM and GPS based modules [5]. Whenever the vehicle crosses the predefined threshold limit of pollution level, the system generates an alert message using the GSM module and the engine of the vehicles gets switched off automatically. To locate the position of the vehicle, GPS module can be used. Such a system can save the environment by controlling the air pollution caused by vehicles.

8. IOT BASED SUPPLY CHAIN MANAGEMENT IN AUTOMOTIVE INDUSTRY

IoT has a huge impact on Automotive Supply Chain Management. It is the interconnected network [17] which binds all the tiers of suppliers, manufacturers (OEM's), service providers, dealers, distributors and customers located at different geographical regions. Smarter Supply Chain [40] will address almost all the challenges faced by supply chain managers such as cost, changing customer demands, risk and globalization.

9. RAW MATERIAL PROCUREMENT: ORDER OPTIMIZATION

To gain more visibility in raw material procurement, Sensors and RFID tags can be used. Movement of material from suppliers to shop floor and throughout the assembly line can be tracked with the help of these 'smart' devices [16]. Specific automobile components can be monitored and tracked with the help of these smart tags for better customization and order optimization.

10. MANUFACTURING: PROCESS OPTIMIZATION

IoT has immense potential to improve visibility in manufacturing to the extent where each unit of production can be "monitored" at each step in the production process. With IoT interconnectivity, shop floor and top floor can be connected which reduces the human intervention to monitor deviations and changes. Operating parameters [17] such as alignment, temperature, pressure

etc. are periodically measured by the sensors. Any deviations beyond the preset values are controlled automatically by making adjustments in the process. Such real time visibility in the manufacturing process results in increased productivity, efficiency and reduced cost.

11. QUALITY ASSURANCE

Quality assurance refers to a systematic process of auditing and inspecting the process and product for meeting the specified requirements [18]. Quality control in all the stages of automotive supply chain is very critical to maintain the competitive edge in the market. With the increasing demand for connected cars, automotive industry is currently facing a transition in car production [19]. Connected cars are more software centric. Expertise in Software Quality Assurance will make a big difference and ensure that great software-driven connected cars are manufactured and delivered.

12. SALES ORDER MANAGEMENT

IoT and connected products are transforming traditional sales and marketing models in automotive industries.

Products such as ‘Cars’ are now connected and OEM’s can stay in touch with the customer throughout the lifecycle of the product. This creates tremendous potential for cross-selling and up-selling. The Internet of Things is the right platform to provide such connected environment. Companies can gain detailed insights into the customer relationship by collecting and analyzing product usage information to understand how the product is performing [20], how much is it being used, which features are being used, and which features are not and accordingly providing post sale services.

13. DELIVERY MANAGEMENT: DISPATCH AND LOGISTICS

Applying IoT to dispatch and logistics operations promises a huge impact. Data Analytics plays an important role in optimizing the distribution networks. By analyzing the parameters like production time, order quantities, production cost, customer locations and delivery cost, supply chain managers can decide on the correct number of distribution centers along with their ideal locations [16]. Such use cases ensure that high service standards are maintained and thereby reducing the transportation and warehousing costs.

E. New Generation Cars

1. AUTONOMOUS/ SELF DRIVING/DRIVERLESS CAR

An autonomous car/ self-driving car/ driverless car [41] is a vehicle that is designed to sense its environment and navigate without any human input. With the continuous developments in self-driving cars, network connectivity between the car and its environment will enhance thereby giving a whole new driving experience. The IoT powered self-driving car is in its own control once the destination is predetermined and communicates with its environment and the on-going traffic [21]. Drivers of a self-driving cars can enjoy the freedom to make use of their phones, laptops and other devices without any fear in mind.

2. ELECTRIC CARS

An electric car is a car that is powered by an electric motor [42], using electrical energy stored in array of rechargeable batteries [9].

Electric cars are growing in popularity as an alternative to vehicles powered by petrol/diesel as they are more sustainable, cheaper, easier to maintain and more efficient [22]. Electric cars are becoming more mainstream due to the increasing availability of plug-in technology, which has allowed owners to charge their cars easily at home, and more powerful lithium-ion batteries that allow them to travel over longer distances. Recently, many countries have started offering incentives for public charging stations making long distance travel a much more economical option for electric car owners. Also, there are companies beginning to develop wireless charging for electric cars, making the charging process much easier.

F. Electric Cars- IoT Applications

- Home charging solutions
- Scheduling a charging slot.
- Nearest charging station with tariff rates.
- Battery charge status with estimated driving time.

II. KEY FINDINGS

Table III.1 summarizes the developments discussed in this paper on IoT in Automotive sector.

Table.1: Developments on IOT in Automotive Sector

Connected Car services / applications	Vehicle Communication	IoT in Intelligent Transportation	IoT based SCM in Automotive Industry	New Generation Cars
Infotainment	Vehicle-to-Vehicle (V2V)	Electronic Collection Smart Toll	Raw Material Procurement: Order Optimization	Autonomous / Self driving/Driverless Car
Vehicles and Smartphones Integration	Vehicle-to-Infrastructure (V2X)	Smart Parking	Manufacturing: Process Optimization	Electric cars
Driving Insights - Analytics	Vehicle-to-Retail (V2R)	Energy Conservation: Smart Street & Traffic Lighting	Quality Assurance:	
On-Board Diagnostics for Predictive maintenance		Post Event Diagnostics: Car Black Box	Sales Order Management	
Safety: Real Time Driver Monitor		Intelligent traffic control with priority for emergency vehicles	Delivery Management: Dispatch and Logistics	
Geo-fencing and Speed Monitoring		Warrant of Fitness: Feedback to Road Transport Authorities		
Law: Stolen Vehicle Tracking		Environmental Control: Emissions and Air Quality Monitoring		
Biometrics Information for Driver identification				

III. SCOPE

Internet of Things is reshaping almost all industry sectors. Applications like Smart Home, Smart Manufacturing, Smart Healthcare, Smart City, Smart Farming, Connected Cars and Wearables amongst many others are transforming the Businesses and enhancing the Customer experience. This paper identifies the innovative applications of IoT in Automotive industry in the areas of Connected Car services/applications, Vehicle Communications,

IoT in Intelligent Transportation, IoT based Supply Chain Management in Automotive Industry and New Generation Cars.

IV. CONCLUSION

This paper outlines developments on IoT in Automotive sector such as Connected Car services/applications, Vehicle communications, IoT in Intelligent Transportation, IoT based Automotive Supply Chain Management and New Generation Cars. As cars get smarter and connected with the other cars, smartphones and things, suitable analytical processing can be applied to the operational parameters, allowing OEM's, automobile drivers as well as road safety authorities to get better visibility about the automobile's performance as well as the overall traffic situation, allowing for timely actions.

REFERENCES

- [1] White paper – “Digital disruption and the future of the automotive industry”- IBM Center for Applied Insights, Dirk Wollschlaeger, Mike Foden, Richard Cave and Matthew Stent,September 2015.
- [2] Whitepaper- “Transforming the Automotive industry with Connected Cars – An Internet of Things Perspective”,IGATE Global Soloution, Madhusudhan Reddy Nukala, Shreyas Bhargave, Bipin Patwardhan, April 2014.
- [3] Connected Cars – Future of Automobiles, by Manisha Salecha. Article in IoT India Magazine, July 20, 2016.
- [4] S.P. Bhumkar, V.V. Deotare, R.V.Babar, “Accident Avoidance and Detection on Highways”, International Journal of Engineering Trends and Technology- Volume3Issue2- 2012.
- [5] Anita Kulkarni, T. Ravi Teja, “Automated System for Air Pollution Detection and Control in Vehicles”, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 3, Issue 9, September 2014.
- [6] Whitepaper – “everis Connected Car Report”, everis - Consulting, IT & Outsourcing Professional Services.
- [7] Whitepaper- “Smart Cars and the IoT-Innovation at the Automotive Ecosystem Edge”, ABI Research, August 2014.
- [8] Whitepaper- “Smart Parking”, Aditya Basu, Happiest Minds Technologies Pvt. Ltd 2014.
- [9] Whitepaper –“2025 Every Car Connected: Forecasting the Growth and Opportunity”, GSMA- SBD Feb 2012.
- [10] Mr. Solanke Digamber, Mr. Jogdand Maruti, Mr.Jadhav Amol, Mr. Mandlik Sachin, “Intelligent Traffic Light And Automatic Street Lighting System According To Traffic Density”, IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 3, Issue 1, Feb-Mar, 2015.
- [11] P. Ajay Kumar Reddy, P.Dileep Kumar, K. Bhaskar Reddy, E.Venkataramana, M.Chandra Sekhar Reddy, “BLACK BOX FOR VEHICLES”, International Journal of Engineering Inventions, ISSN: 2278-7461, Volume 1, Issue 7(October2012).
- [12] Whitepaper –“The Connected Car and Privacy Navigating New Data Issues”, Future of Privacy Forum, November 13, 2014.
- [13] Mario Mlinaric , “Intelligent traffic control with priority for emergency vehicles”, April 2016.
- [14] Suresh Sharma, “Traffic Light Priority Control For Emergency Vehicle Using RFID”, International Journal of Innovations in Engineering and Technology (IJIET), Vol 2 Issue 2 April 2013.
- [15] Web link/: https://en.wikipedia.org/wiki/Warrant_of_Fitness
- [16] Whitepaper – “The Smarter Supply Chain of the Future”, IBM Institute for Business Value, 2009.
- [17] Whitepaper-Designing for Manufacturing’s ‘Internet of Things’ June 2014, Cognizant.
- [18] Whitepaper- “Integrated Quality Assurance”, TÜV SÜD AG, Munich, Germany.
- [19] Web link/: <http://www.smartindustryblog.com/worldwide/why-quality-assurance-is-so-important-to-the-automotive-industry/>
- [20] Web link: <http://www.ptc.com/product-lifecycle-report/how-the-iot-is-changing-sales-and-marketing>
- [21] Web link: <https://informationcounts.com/autonomous-vehicles-and-the-internet-of-things/>
- [22] “The connected car of the future”, <https://agsolar.com.au/blog/the-connected-car-of-the-future/>
- [23] “The New Auto Insurance Ecosystem: Telematics, Mobility and the Connected Car”, Cognizant Report, August 2012.
- [24] Mrs. K.P.Kamble, “Smart Vehicle Tracking System”, International Journal of Distributed and Parallel Systems (IJDPS) Vol.3, No.4, July 2012.
- [25] White paper – “CEO Perspective: The Internet of Things for the Automotive Industry”, 2014 SAP SE.
- [26] White paper- “The In-Car AppExperience: Convergenceand Integration”, UIEvolution.
- [27] Liu Tenghong,Yuan Rong,Chang Huating, “Research on the Internet of Things in the AutomotiveIndustry”, 2012 IEEE International Conference on Management of e-Commerce and e-Government.
- [28] Jayavaradhana Gubbi, Rajkumar Buyya, Slaven Marusic, Marimuthu Palaniswami, Internet of Things (IoT): A vision, architectural elements, and future directions, Future Generation Computer Systems 29 (2013) 1645–1660
- [29] White paper – “IntelligentIn-VehicleInfotainment(IVI)”, Mindtree
- [30] White paper – “Exploring the Connected Car”, Cognizant November 2012
- [31] White paper – “What’s Driving the Connected Car?”, Spirent, 2016
- [32] Ning Lu, Nan Cheng, Ning Zhang, Xuemin Shen and Jon W. Mark, “Connected Vehicles: Solutions and Challenges”,IEEE INTERNET OF THINGS JOURNAL, VOL. 1, NO. 4, AUGUST 2014
- [33] Shuhui Ma, “The Solution of an IoT Application: Smart Vehicle”, Proceedings of ICCTA2011
- [34] White paper – “Connected Vehicles - From Building Cars to Selling Personal Travel Time Well-Spent”, Cisco April 2011
- [35] Pritpal Singh, Tanjot Sethi, Bibhuti Bhusan Biswal, and Sujit Kumar Pattanayak, “A Smart Anti-theft System for Vehicle Security”, International Journal of Materials, Mechanics and Manufacturing, Vol. 3, No. 4, November 2015
- [36] M.S.Manivannan, “Smart Card and IOT Based Electronic Toll Collection System for Vehicles”, International Journal of Modern Embedded System (IJMES) ISSN: 2320-9003(Online)Volume No.-4, Issue No.-1, February, 2016
- [37] Mr. Basavaraju S R, “Automatic Smart Parking System using Internet of Things (IOT)”, International Journal of Scientific and Research Publications, Volume 5, Issue 12, December 2015 ISSN 2250-3153
- [38] White paper – “Smart cities and the Internet of Things”, Hewlett Packard Enterprise 2016
- [39] C. Balasubramaniyan and D. Manivannan, “IoT Enabled Air Quality Monitoring System (AQMS)using Raspberry Pi”, Indian Journal of Science and Technology,Vol 9(39), DOI: 10.17485/ijst/2016/v9i39/90414, October 2016ISSN (Print) : 0974-6846ISSN (Online) : 0974-5645
- [40] Suresh Bhattacharya, Dr. D. Mukhopadhyay and Dr. Sunil Giri, “Supply chain management in Indian automotive industry: complexities, challenges and way ahead”, International Journal of Managing Value and Supply Chains (IJMVSC) Vol.5, No. 2, June 2014
- [41] Technical Report - “Self-Driving Cars”, http://www.funginstitute.berkeley.edu/sites/default/files/Self_Driving_Cars.pdf
- [42] Service Training Report- “Basics of Electric Vehicles”, Volkswagen Group of America, July 2013

- [43]Boon-Giin Lee and Wan-Young Chung, "A Smartphone-Based Driver Safety Monitoring System Using Data Fusion", Sensors 2012, 12, 17536-17552; doi:10.3390/s121217536
- [44]D.Sowmya, I.Suneetha, N.Pushpalatha, "Driver Behavior Monitoring through Sensors and Tracking the Accident using Wireless Technology", International Journal of Computer Applications (0975 – 8887) Volume 102– No.2, September 2014
- [45]Thomas O'Kane and John V. Ringwood, "Vehicle Speed Estimation Using GPS/RISS (Reduced Inertial Sensor System)", ISSC 2013, LYIT Letterkenny, June 20–21
- [46]Mohammed Baqer M. Kamel, "Real-Time GPS/GPRS Based Vehicle Tracking System", International Journal of Engineering and Computer Science ISSN: 2319-7242 Volume 4 Issue 8 Aug 2015, Page No. 13648-13652
- [48]Buddhika R. Maitipe, U. Ibrahim, M. I. Hayee, and Eil Kwon, "Vehicle-to-Infrastructure and Vehicle-to-Vehicle Information System in Work Zones Dedicated Short-Range Communications, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 125–132. DOI: 10.3141/2324-15
- [49]"Vehicle-to-Vehicle Communications: Applications and Perspectives", DOI: 10.1002/9780470610893.ch12 In book: Wireless Ad Hoc and Sensor Networks, pp.285 – 308

Privacy Challenges Internet of Things Applications

Shiwani Gupta¹, Priyankar R. Tiwari², Loukik S. Salvi³

Department of Computer Engineering, Thakur College of Engineering and Technology, Mumbai, India

Abstract— The Internet of things paradigm believes the enormous interconnection and cooperation of smart matters over the modern-day and destiny gadgets. The internet of things is, consequently, the evolution of the Internet to cover the actual world, enabling many new offerings with the intention to enhance human beings' everyday lives, spawn new companies and make homes, cities and delivery smarter. Smart things permit certainly for ubiquitous statistics series or tracking; however, those beneficial capabilities are also examples of privacy threats which are already now proscribing the achievement of the internet of things imaginative and prescient while no longer applied efficaciously. Those threats involve new challenges which includes the pervasive privacy-conscious management of private facts or methods to control or avoid ubiquitous monitoring and profiling. This paper analyses the privacy troubles within the internet of factors in detail. To this cease, we first speak the evolving capabilities and traits inside the Internet of things with the intention of scrutinizing their privacy implications. 2nd, we classify and take a look at privacy threats on this new placing, declaring the demanding situations that want to be conquer to make certain that the Internet of factors becomes a reality.

Index Terms— Issues, Challenges, Vulnerabilities, Internet of Things, IoT, Application, Drawback, Privacy, Security

I. INTRODUCTION

The Internet of Things (IoT) anticipates the interconnection of billions to trillions[1,2], of savvy things around us – particularly identifiable and addressable regular things with the capacity to gather, store, process and impart data about themselves and their physical condition. IoT frameworks will convey propelled administrations of a radically new kind in light of progressively fine-grained information procurement in an environment [3] thickly populated with shrewd things. Cases of such IoT frameworks are inescapable social insurance, propelled constructing administration frameworks, savvy city administrations, open reconnaissance and information obtained, or participatory detecting applications [4].

The undeniably undetectable, thick and unavoidable accumulation, handling and spread of information amidst individuals' private lives offer to ascend to genuine security concerns. The obliviousness of those issues can have undesired results, e.g. Disapproval and disappointment of new administrations, harm to notoriety, or expensive claims. The general population blacklist of the Italian retailer Benetton in 2003 [6], the repudiation of the Dutch shrewd metering bill in 2009 [7], or the current clamor against the EU FP7 inquire about venture INDECT [9, 10] are just three cases of IoT related tasks that accomplished enormous issues because of uncertain security

issues. Security has been a hot research theme in various innovation and application regions that are vital empowering influences of the IoT vision, e.g. RFID, remote sensor systems (WSN), web personalization, and versatile applications and stages. Regardless of significant commitments from these groups, an all-encompassing perspective of emerging protection issues in the IoT is absent since the IoT is a developing idea that contains a developing number of advances and displays a scope of evolving highlights. Among these, we witness a blast in the quantity of brilliant things and better approaches for communicating with frameworks and introducing input to clients. As we will appear, these new highlights of the IoT will disturb security issues and present unanticipated dangers that stance testing specialized issues. These security dangers, regardless of whether known or new, should be considered (I) in a reference model of the IoT bookkeeping that records for its particular substances and information streams, (ii) from the point of view of existing protection enactment, and (iii) with respect to the interesting and developing highlights in the IoT.

II. PRIVACY DEFINITION

Privacy is a very vast and various notion for which literature offers various definitions and views. The perception of privacy shifted between media, territorial, communication, and physical privacy. With the increasing use and efficiency of digital facts processing data privacy has grown to be the essential trouble these days. Information privacy turned into defined by Westin in 1968 as "the right to pick out what private records about me is known to what human beings" [3]. Whilst Westin's definition, it noted non-electronic environments remains valid, it is also too general to permit focused discussion approximately privacy in the IoT. We as a result adapt and concretize definition:

Privacy in the Internet of Things is the threefold guarantee to the subject for

- Awareness of privacy risks imposed by smart things and services surrounding the data subject
- Individual control over the collection and processing of personal information by the surrounding smart things
- Awareness and control of subsequent use and dissemination of personal information by those entities to any entity outside the subject's personal control sphere.

Our definition of privacy captures in essence the idea of informational self-willpower by using enabling the concern (i) to assess his non-public privacy risks, (ii) to take appropriate action to shield his privacy, and (iii) to be assured that it's miles enforced past his immediate manipulate sphere. The running systems analogy described by way of Radomirovic in is a similar concept to represent what we talk over with because the personal sphere of

the statistics subject. In smart home situations it could be pictured as that person's household or instantaneous area, as Radomirovic fittingly observes. However, the exact scope of the situation's non-public sphere can vary from situation to state of affairs and its miles still doubtful what constitutes the man or woman's private sphere, or operating gadget limitations within the analogous phrases, in e.g. A place of business surroundings or public space. In addition, the notion of personal information is necessarily fuzzy, due to the fact privacy is a deeply social concept and situation to substantially varying person notion and requirements. Subsequently, care have to be taken when designing new systems and offerings to cautiously verify the sensitivity of the involved information and bearing on user requirements, e.g. As businesses are starting to enforce in privacy impact analysis's (pias). In the long run, our definition ought to be understood such that the user may additionally define what he considers private facts.

III. REFERENCE MODEL

The IoT reference model we endorse is based on the ITU and IERC visions of the IoT and may be summarized as: each person and anything is interconnected anywhere at any time through any community taking part in any service. Our reference model describes the entities and facts flows of IoT packages. In the model we take into account 4 major styles of entities as depicted in figure 1. Smart matters are regular matters augmented with records and verbal exchange era (ICT). They're capable of collect system and communicate facts approximately themselves and their environment as well as to interact with other things and humans. Backend host offerings that acquire, combine and examine information from many clever matters to provide a fee-introduced carrier to the cease-consumer. People have two exceptional roles in our reference model. They may be problem to records collection by way of the smart things surrounding them or recipients of facts or offerings. Be aware, that someone may be each situation and recipient at the equal time, e.g. In a personal healthcare software. Sooner or later, smart things are related to services thru an infrastructure with specific traits ranging from lowpower lossy networks to powerful Internet backbones probable traversing distinctive intermediate gateways and servers, e.g. Firewalls and protocol bridges.

Privacy legislation

Privacy legislation tries to attract barriers to the evermore information-hungry business models of many Internet businesses (e.g. Facts market locations, marketing networks, ecommerce web sites) and to define mandatory practices and processes for privacy safety. We rapidly evaluation the development and realistic impact of privacy legislation then discover applicable factors and issues in the context of the internet of things. Privacy is diagnosed as an essential human proper within the 1948 widespread statement of Human Rights and is anchored within the constitutional regulation of maximum countries today. The first main piece of law on statistics privacy changed into handed with the 1974 US privacy Act, which hooked up the fair facts practices (fips).

The fips contain the concepts of word, consent, man or woman get right of entry to and control, facts minimization, functional use, ok safety and accountability. They had been taken up in by the business enterprise for financial Co-operation and improvement (OECD), which anticipated trade barriers from the more and more numerous country wide privacy law.

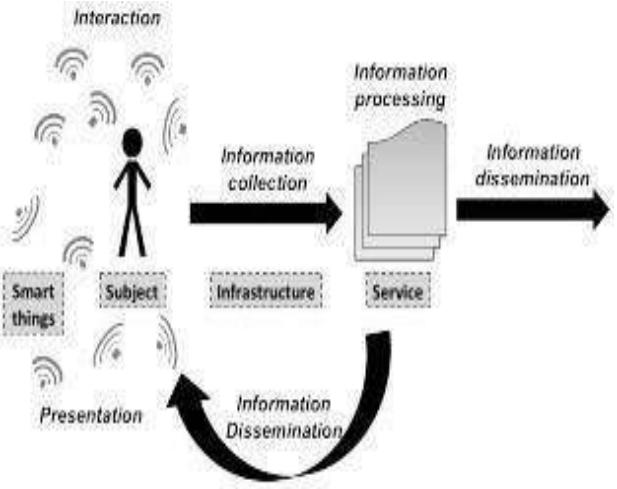


Figure 1: IoT Reference Model

awareness of incidents inside the first area. In the end, the economics of privacy are nevertheless in desire of these in push aside of privacy rules. On the one facet, improvement of pets, enforcement and audits of privacy protection guidelines is highpriced and may restrict enterprise fashions. On the other facet, violations of privacy legislation both move unpunished or result most effective in comparably small fines, at the same time as public awareness is still too low to result in intolerable harm of public popularity. Accordingly, dismiss of privacy legislation, as e.g. Google deliberately circumventing Safari's person monitoring protection, appears profitable. Over this incident, Google paid a report quality of \$22.5 Million in an agreement with the Federal exchange commission (FTC), but it's miles attainable that the income extra than compensated. It is going to be a chief undertaking to design a unified enduring legislative framework for privacy safety inside the internet of things, in place of passing speedy previous portions of rules on singular technologies. The important thing, however, can be a deep understanding of existing and lingering new threats to privacy in the IoT – these threats are what rules desires to protect against, ultimately.

IV. EVOLUTION IN IOT

Beginning with the imaginative and prescient of ubiquitous computing, the Internet of factors relies upon on technologic progress to deliver increasing miniaturization and availability of facts and communique generation at lowering fee and energy consumption. The IoT is hence not a disruptive new era, but a singular paradigm whose complete attention may be a slow procedure. This phase reviews key technology in the beyond evolution of the IoT and associated paintings in privacy and then analyses the current features of the IoT and how they evolve. This discussion of evolving technology and capabilities permits us to apprehend and examine privacy threats and challenges early on.

Evolving Technologies

RFID generation stands at the beginning of the IoT imaginative and prescient. It allows passive automated identity of factors at the price of a couple of cents. Certainly, the conclusion of the IoT imaginative and prescient continues to be frequently visible in the pervasive deployment of RFID tags. RFID privacy problems were thoroughly researched. The dominant threats are automated

identification and tracking of people through hidden tags, e.g. In clothes. Different countermeasures had been proposed, inclusive of reader authentication, tag encryption, randomizing tag identifiers, and blocking off or killing of tags. Dealing with the lifecycle of an RFID tag is a task in order to also be thrilling for smart matters inside the IoT context. Wireless sensor network (WSN) generation bureaucracy the subsequent evolutionary step of the IoT: matters turn out to be active, as they're augmented with sensing, processing and conversation talents to construct first interconnected networks of things. Today, sensor nodes range from tiny millimetre-sized sensor nodes (e.g. Clever dirt) to meter-scale GSM-ready climate stations. Sensor networks consist of each small-scale domestic deployment and also bigscale industrial tracking structures enabled by means of standards inclusive of zigbee, zwave, ANT or Bluetooth. Privacy research in wsns has focussed on privacy threats with regard to the collected sensor facts, queries to the community and the vicinity of records sources and base stations. Precise demanding situations which have been usually identified are uncontrollable environments and aid constraints, which additionally traits of the IoT. The arrival of clever telephones has further progressed the conclusion of the IoT imaginative and prescient, being the first mobile mass gadgets with ubiquitous internet connection. Smart phones accumulate critical amounts of private statistics about their proprietor, e.g. Identifiers, bodily area, and activity that bear massive privacy risks [1]. Huge privacy research has been performed focussing mostly on region-based offerings, detection of privacy breaches [5, 6] and privacy aware architectures for participatory sensing [7, 8]. The cloud computing paradigm has thrived over the last decade and offers means to deal with the expected records explosion within the IoT. Privacy research in cloud computing has focussed on adequate records protection and prevention of statistics leaks [5,6], auditing and provenance, and personal facts processing.

A. Evolving Features

Technology: The internet of things started out with the autoid Labs envisioning automated identification of factors through RFID technology. These days, we see that increasingly technologies are incorporated underneath the IoT umbrella term, e.g. Sensor networks, smart phones, and cloud offerings. Predictions for the destiny frequently foresee that information and verbal exchange technology more and more actions into things thereby making them smarter and self-conscious [5]. At the identical time, new technologies may be invented and made appropriate for the loads. In opposition to this background, it's miles but doubtful if, while, and to which volume the estimated kingdom of ubiquity and invisibility can be reached [4]. Size estimates from academia and foremost industry players, amongst them IBM, HP and Cisco, predict fantastic numbers of 50 billion to trillions [1] of connected matters through 2020 [2]. Even though predicted numbers and time horizons vary extensively, even defensive situations suggest that the IoT will increase the size of networks by orders of value. Such big amounts of devices will severely assignment the scalability of existing and imminent privacy era. Interconnection with the technologic advances and lowering charges of wireless communications, the IoT is predicted to adapt in the direction of a country where smart things are ubiquitously interconnected [2]. All-IP give up-to-end connectivity is the accepted imaginative and prescient for recognition of such

pervasive interconnectedness and is closely driven with the aid of the IETF 6lowpan and ETSI M2M working corporations. As smart matters emerge as question-able from any distance on this process, this will not best bring on terrific opportunities for new services but additionally difficult privacy troubles.

Thing interaction:

With growing numbers of internet linked things, it turns into critical what interfaces they offer to humans for configuration, debugging and interaction. We've witnessed a development from RFID tags with little interplay competencies to sensors and gadgets featuring constrained interfaces, which includes leds and small presentations responding to buttons and touches. Predictions that foresee mainly more haptic interfaces and the smart phone as principal mediator among people and things based on internet interfaces seem very sure, as we are already beginning to see those tendencies these days. It is a thrilling open question, whether or not web-based interfaces will be successful as number one interfaces to things. Different prediction including Intel's for concept-pushed control via 2020 remains very dubious. The lack of interfaces and interplay mechanisms can pose privacy threats as determined in RFID. Then again, too complex interfaces (as e.g. Found in social networks) do now not seem to help either, while increasingly personalized interplay which includes speech or, at the extreme, thoughts may even rise up new privacy issues.

System interaction:

Aside from the interaction with things, it is exciting how better degree interaction of human beings with the IoT, i.e. With groups of factors, structures and offerings, will expand. Comparable arguments and predictions as for aspect interaction observe also to system interplay. The principle difference is that the complete environment is anticipated to serve for interplay with people primarily based on elaborate interfaces realized through the collaboration and coordination of many smart matters and their specific abilities. Much less coherent predictions foresee thought-driven interaction and mechanisms that appeal to all human senses. An environment wherein human being is pervasively and passively exposed to interaction and comments from IoT structures must be carefully designed now not to violate their privacy.

V. PRIVACY THREATS AND CHALLENGES IN THE IOT

The evolving nature of the IoT concerning technologies and functions in addition to the rising new approaches of interplay with the IoT cause particular privacy threats and demanding situations. This section presents our category of those threats. Figure 2 arranges them into the 5 special stages from our reference version in step with where they're maximum vulnerable to seem. Every of the seven threat classes is analysed in four steps: First, a definition and characterization of the threat category with concrete times of privacy violations is given. Second, we examine how the IoT evolution influences, changes and aggravates the unique chance. Right here, table II serves as a precis of or three decided on features with most impact on a specific risk category. Third, we pick out methods and countermeasures from related work and ask whether or not they be carried out also inside the IoT or are insufficient. Fourth, we present the principle technical demanding situations and capacity methods to triumph over the ones threats in the IoT. Of path, the hazard categories aren't totally disjunct and we make sure to factor out dependencies and overlaps between exclusive

threats.

A. Identification

Identity denotes the threat of associating a (chronic) identifier, e.g. A name and cope with or a pseudonym of any kind, with an individual and information about him. The hazard as a result lies in associating an identity to a specific privacy violating context and it additionally permits and aggravates other threats, e.g. Profiling and tracking of individuals or combination of various statistics resources. The risk of identity is presently maximum dominant within the facts processing section on the backend offerings of our reference model, in which huge quantities of information is concentrated in a central region outside of the issue's control. In the IoT additionally the interaction and collection segment will become relevant, because the effect of the evolving technologies and interconnection and interaction functions aggravates the risk of identity: First, surveillance camera era is increasingly integrated and used in non-safety contexts, e.g. For analytics and advertising. As facial databases (e.g. From fb) end up to be had also to non-governmental events like advertising and marketing platforms, automated identification of individuals from digicam snap shots is already a truth. 2nd, the growing (wi-fi) interconnection and vertical verbal exchange of regular things, opens up opportunities for identity of gadgets thru fingerprinting [11]. It became diagnosed already for RFID technology that individuals may be identified through the aura in their things. 1/3, speech recognition is extensively utilized in mobile applications and massive records-bases of speech samples are already being constructed. The ones ought to probably be used to understand and perceive people, e.g. By governments asking for get right of entry to that facts. With speech reputation evolving as an effective way of interaction with IoT systems and the proliferation of cloud computing for processing tasks, this may in addition extend the assault vector and privacy risks. Identity protection and, complementary, protection against identity is a predominant topic in RFID privacy, however has also received tons attention in the regions of records anonymization, and privacy-enhancing identification management. Those procedures are difficult to suit to the IoT: most records anonymization techniques can be damaged using auxiliary facts, that is in all likelihood to grow to be available in some unspecified time in the future all through the IoT evolution. Identity control solutions, except depending closely on highly-priced crypto operations, are by and large designed for very confined environments, like business enterprise or home networks and thus tough to suit to the dispensed, numerous and heterogeneous environment of the IoT. Procedures from RFID privacy due to similarities in resource constraints and numbers of factors are the maximum promising.

B. Localization and Tracking

Localization and monitoring is the chance of figuring out and recording someone's location through time and area. Tracking requires identification of a few type to bind continues localizations to 1 man or woman. Already today, monitoring is viable through distinct manner, e.g. GPS, internet traffic, or cellular phone place. Many concrete privacy violations have been recognized related to

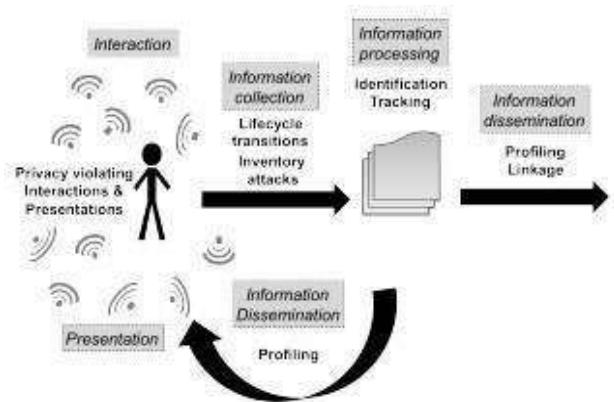


Figure 2: Threats in Reference Model

this chance, e.g. GPS stalking, disclosure of personal statistics consisting of an infection, or commonly the uneasy feeling of being watched. But, localization and tracking of people is likewise a vital functionality in many IoT structures. These examples show that users understand it as a violation once they don't have control over their place statistics, are ignorant of its disclosure, or if data is used and combined in a beside the point context. This coincides with our definition of privacy. Proximity, localization and tracking typically do not lead to privacy violations, as e.g. Everybody within the on the spot surrounding can directly examine the difficulty's region. Research on place privacy has proposed many processes that may be categorised by using their architectural angle into (i) client-server, (ii) trusted third birthday party, and (iii) distributed/peer to-peer. But, the ones tactics have been primarily tailor-made to outdoor situations wherein the user actively makes use of a LBS through his clever phone. As a consequence, these procedures do no longer healthy without sizable changes to the adjustments added via IoT.

The main challenges we perceive are (i) cognizance of tracking inside the face of passive data series, (ii) manipulate of shared location records in indoor environments, and (iii) privacy preserving protocols for interplay with IoT systems.

C. Profiling

Profiling denotes the chance of compiling statistics dossiers about individuals so as to infer interests by using correlation with other profiles and statistics. Profiling techniques are more often than not used for personalization in e-trade however additionally for inner optimization based totally on purchaser demographics and pursuits. Examples in which profiling leads to a violation of privacy violation are price discrimination, unsolicited commercials, social engineering, or inaccurate automatic decisions, e.g. By means of fakebooks' automated detection of sexual offenders. Also, collecting and selling profiles approximately human beings as practiced by way of numerous statistics marketplaces nowadays is usually perceived as a privacy violation. The examples display that the profiling threat appears specifically within the dissemination section, toward 1/3 parties, but also toward the difficulty itself in shape of inaccurate or discriminating choices. The impact of the evolving functions inside the IoT is especially twofold:

First, the IoT evolution leads to an explosion of records sources as an increasing number of ordinary things get related. Second, even as information series, for this reason, increases quantitatively through orders of significance, it additionally adjustments qualitatively as data is gathered from formerly inaccessible elements of people's private lives.

Additionally, the aggravation of identification and monitoring threats, in addition, fuels the possibilities for profiling and risk from dubious records-promoting organizations. Current methods to preserve privacy encompass client side personalization, data perturbation, obfuscation and anonymization, distribution and running on encrypted records. These tactics can possibly be applied to IoT scenarios, however, must be adapted from the same old version that assumes a primary database and accounts for the many distributed facts sources that are expected inside the IoT. This will require big efforts for recalibration of metrics and redesign of algorithms, as e.g. Latest work in differential privacy for distributed facts sources indicates [12].

A. Privacy-violating interaction and presentation

This hazard refers to conveying private facts through a public medium and in the system disclosing it to an unwanted audience. It can be loosely sketched as shoulder surfing but in real-international environments. Many IoT applications, e.g. Clever retail, transportation, and healthcare, envision and require heavy interplay with the consumer. In such structures, it's miles imaginable that statistics might be provided to users using smart things in their surroundings, e.g. Thru superior lighting fixtures installations, speakers or video displays. Vice versa, customers will manage structures in new intuitive ways the usage of the matters surrounding them, e.g. Moving, touching and speaking to clever things. However, lots of the ones interplay and presentation mechanisms are inherently public, i.e. People in the location can observe them. This turns into a threat to privacy whilst non-public data is exchanged between the device and its user. In smart towns e.g. A person would possibly ask for the way to a particular fitness clinic. The sort of question ought to now not be replied e.g. By way of showing the way on a public show nearby, visible to any passers-with the aid of. Every other instance are recommendations in shops that reflect private pursuits, which includes specific weight loss plan food and medicine, films or books on precarious topics. Due to its close connection to interaction and presentation mechanisms, the hazard of privacy-violating interactions and presentation seems more often than not inside the homonymous phases of our reference model. For the reason that such superior IoT offerings are nonetheless in the future, privacy violating interactions have now not obtained a good deal interest from research. Interaction mechanisms are but critical to usable IoT structures and privacy threats must therefore be addressed. We identify two specific challenges with the intention to should be solved: First, we want way for automated detection of privacy-touchy content. It's far without difficulty imaginable that the provisioning of content material and rendering it for the user are treated in two steps by means of extraordinary systems: E.g. Company A generates guidelines for customers of a shop, that are then

delivered to the consumer via agency B's device: either by way of special lighting fixtures and the use of audio system or via a push to his smart cell phone. How to choose between the ones interaction mechanisms, one public one personal? Ought to corporation A mark privacy touchy content material or ought to organization B hit upon it? How can employer B (committed to privacy) guard itself from A's lax privacy mindset? Automated detection of privacy-sensitive content can help to determine these questions. 2d, with the previous factor in mind, scoping could be essential, i.e. How can we scope public presentation medium to a selected subgroup of recipients or a selected physical vicinity? This method would evidence useful to help users, which haven't any clever smartphone (or some other tool providing a non-public channel for interactions and displays). But, it is going to be hard to accurately decide the captive target market of a selected presentation medium, separate the supposed goal institution and adjust the scope as a consequence. Packages for privacy-keeping pervasive interaction mechanisms are, e.g. Smart shops and department shops, clever cities and healthcare applications.

Inventory attack

Inventory assaults talk over with the unauthorized series of statistics about the life and traits of personal things. One evolving feature of the IoT is interconnection. With the realization of the all-ip and end-to-give up imaginative and prescient, smart matters end up question-capable over the Internet. At the same time as things can then be queried from everywhere through legitimate entities (e.g. The proprietor and licensed users of the machine), non-legitimate parties can question and take advantage of this to assemble an stock list of factors at a specific region, e.g. Of a household, office building, or manufacturing facility. Although smart matters could distinguish valid from illegitimate queries, a fingerprint in their verbal exchange speeds, reaction times and other unique characteristics may want to probably be used to determine their type and version. When you consider that stock assaults are particularly enabled by means of the increasing verbal exchange abilities of things, the danger arises inside the information collection segment of our reference version. The impact of latest technologies in this chance isn't always yet clean. Manifold concrete privacy violations based on inventory attacks are conceivable or have truly come about. First, burglars can use stock facts for focused run-ins at private homes, offices, and factories, much like how they already use social media today to stake out capacity victims [9]. Third, personal statistics is disclosed by way of the ownership of particular matters, consisting of non- public interests (e.g. Books, films, song) or health (e.g. Medicinal drug, medical gadgets). Fourth,

efforts for business espionage may be complemented by a stock assault, as noted by Mattern [13]. Radomirovic [13] and Van Deursen [12] have recognized the threat of profiling through fingerprinting inside the context of RFID. However, with RFID the trouble is at a much more local scope as RFID tags may be read handiest from a close distance and queries are mostly confined to reading the tag's identifier. As analyzed above, the problem will worsen inside the IoT evolution because the attack vector is greatly elevated by means of increasing proliferation of wireless communications, give up- to end connectivity, and greater sophisticated queries. With the intention to thwart stock assaults within the IoT, we identify the following two technical demanding situations: first, smart matters ought to be able to authenticate queries and handiest solution to the ones by using legitimate events to thwart active stock assaults via querying. Inventory attacks will certainly be hard to counter. The reality that the usage of pets, although meant to defend privacy, can clearly make fingerprinting even simpler, leaves hiding inside the (privacy-ignorant) hundreds presently because of the maximum feasible but genuinely suboptimal answer [13]

VI. CONCLUSION

This paper motivates the want for a detailed evaluation of privacy threats and demanding situations within the internet of things. We dissect this complicated topic right into a 4-step technique:

- First, we offer a formal foundation for discussing privacy in the IoT via concisely framing our belief of privacy and the applied reference version. A brief assessment of applicable privacy legislation identifies clear insufficiencies and similarly motivates the want for a detailed assessment of privacy threats.
- Inside the second step, we acknowledge that the internet of factors is constantly evolving and cannot be decreased to the sum of the technologies it builds upon. Right here, our discussions of evolving technologies and capabilities provide both a general and privacy-targeted view on the past, present and destiny evolution of the IoT.
- Thirdly, we summarize existing privacy threats into seven categories and review them in the light of the evolving IoT. Identity, tracking, and profiling are long known threats that, as we display, might be greatly aggravated in the IoT.
- The four threats of privacy-violating interactions and displays, lifecycle transitions, inventory attacks and records linkage rise up later inside the IoT evolution.

We don't forget that profiling stays one of the most excessive threats: Our analysis suggests that it is substantially irritated and that different threats like identification or tracking, although every scary unique very specific privacy violations, add to its dangers with the aid of offering even greater linkable information. We bear in mind privacy-violations within the interaction and presentation section a critical destiny threat, due to the corresponding interplay mechanisms with clever things and structures that are just evolving and are instead particular to the IoT. The involved technical challenges have subsequently obtained little attention within the associated work to date and require new methods of the use of generation as well as a honest amount of foresight and sensitivity for privacy implications. Ultimately, we stress two middle minds, that our paintings indicate for a privacy-aware Internet of things: First, the IoT is evolving – privacy is a consistent project and should be faced with the essential foresight. 2d, a fruitful outcome calls for the coordinated motion to offer technical solutions supported via the corresponding prison framework.

VII. REFERENCES

- [1] Evans D. The Internet of Things - How the Next Evolution of the Internet Is Changing Everything. CISCO white paper 2011
- [2] David K, Jefferies N. Wireless visions: A look to the future by the fellows of the wwrif. Vehicular Technology Magazine, IEEE dec 2012; 7(4):26 –36, doi:10.1109/MVT.2012.2218433.
- [3] Mattern F, Floerkemeier C. From active data management to event-based systems and more. Springer-Verlag, 2010.
- [4] Presser M, Krco Sa. IOT-I: Internet of Things Initiative: Public Deliverables – D2.1: Initial report on IoT applications of strategic interest 2010.
- [5] Atzori L, Iera A, Morabito G. The Internet of Things: A survey. Computer Networks 2010; 54(15):2787–2805, doi:10.1016/j.comnet.2010.05.010
- [6] Benetton to Tag 15 Million Items. RFID Journal. [Http://bit.ly/xxe4wi](http://bit.ly/xxe4wi) [Online. Last accessed: 2012-09-25], 2003.<https://www.buzzle.com/articles/pros-and-cons-ofinternetof-things-IoT.htm>
- [7] Cuijpers C. No to mandatory smart metering does not equal privacy! Tilburg Institute for Law, Technology, and Society Webblog 2009;
- [8] Renaud K, Ga andlvez Cruz D. Privacy: Aspects, definitions and a multi-faceted privacy preservation approach. Information Security for South Africa (ISSA), 2010, 2010; 1 –8, doi:10.1109/ISSA.2010.5588297.
- [9] Westin AF. Privacy and freedom. Washington and Lee Law Review 1968; 25(1):166.
- [10] Radomirovic S. Towards a Model for Security and Privacy in the Internet of Things. 1st International Workshop on the Security of the Internet of Things, Tokyo, Japan, 2010.
- [11] Moore B. Privacy: Studies in social and cultural history. M.E. Sharpe, 1984.Solove D. A taxonomy of privacy. University of Pennsylvania Law Review 2006; 154(3):477.
- [12] International Telecommunication Union (ITU). The Internet of Things. ITU Internet Reports 2005;
- [13] Vermesan O, Friess P, Guillemin P, Gusmeroli S,Sundmaeker , Bassi A, Jubert I, Mazura M,Harrison M, Eisenhauer10 M, et al.. Internet of things strategic research roadmap. Internet of Things: Global Technological and Societal Trends 2009;

Hacking in Blind: Use of User Interface as Attack

Aishwarya R Verma¹, Sneha T Patil¹, Sheetal Rathi²

M.E. Scholar¹, Associate. Professor

*Dept. of Computer Science Engineering, Thakur College of Engineering and Technology
aishwaryaverma.2015@gmail.com, snehal321.patil@gmail.com, sheetal.rathi@thakureducation.org*

Abstract - Many terminals are used in safety critical operations in which humans, through terminal user interfaces, become part of the system control loop. These terminals are generally embedded, single purpose devices with restricted functionality, sometimes air gapped and increasingly hardened. This paper gives information about the various ways of the hackers that can attack the systems using adaptive user interface attacks in which it attacks a small device of the interface that connects user input peripherals to the target system. Such device executes an attack making the authorized user feel safe and make them performing safety-, security-critical operations, by modifying or blocking user input channel, so that the hacker can able to overcome the number of challenges that makes inability of direct observation of state user interface along with getting avoided to be detectable by the legitimate user. Such attacks can be involved with clickjacking, likejacking, cursorjacking, password manager attacks, etc. which the user cannot able to recognize it easily. Here, we will determine the art of hacking human brain through various user interface designs through which the hackers attack the users' systems quickly.

Keywords - User interface attacks, input peripherals, authorized user, safety-, security-critical operations.

I. INTRODUCTION

GUI plays an instrumental role in the success of any browser. GUI enables user control and improves interaction with the browser. GUI is considered a part of the user trust model for all types of software including browsers. GUI in browsers includes notification bars, status bars, address bars, download dialog boxes, HTTP authentication dialog, and browser objects such as frames, buttons, etc. Users interact with the GUI components in their routine life jobs. GUI flaws are considered design errors in which an attacker can circumvent the normal functioning of the browser by running malicious JavaScript. Primarily, GUI bugs in browsers are mostly exploited by spoofing [1], clickjacking, likejacking, password manager attacks. Spoofing attacks are those kinds of attacks that tamper the UI component of software in order to fool users into performing false operations by exploiting their ignorance. Clickjacking attacks are under the categorization of UI redressing attacks in which an attacker embeds a hidden UI object such as buttons, frames, etc. For example, an attacker can easily place a hidden button over the real button in a browser window that executes a malicious function when a user clicks it. Likejacking attacks fall into the UI redressing attacks in which an attacker embeds a hidden UI object in the form of likes, dislikes, add to lists options etc., which on-click to this can executes some malicious functions. Password manager attacks can consist of SQL injections on the look-alike GUI website used with different server connectivity. This can lead to the hacking of sensitive password of

the sensitive sites like e-banking, industrial and educational login system, mail login password, important pins and other information.

Such attacks can take place through different servers' connectives', which are hard to recognize in a look-alike GUI. Therefore, this gives information about various attacks which are not easy to recognize in GUI.

II. PROBLEM STATEMENT

As these attacks are invisible to traditional malware detection, so it operates under uncertainty without any feedback from systems and it directly gives the visual indication of attack to the user. This is called as "hacking in blind". The attackers come up with the innovative approaches in which they propose a novel way to attack the system through UI. Such attacks are quick to deploy, very hard to notice, invisible to existing malware detection software. The attackers try to implement various new techniques in which UI are hard to track or hardly be detected fingerprinting techniques. They implement the attack on a small embedded device.

The goal of the attackers is to attack a security-critical UI through various physical attack techniques. Before approaching the attack through UI, many other techniques were studied and tried to implement. The following are the physical attacks techniques and their limitations.

A. Hardware Modification

In this case, the attacker has brief information of the physical access to a device. Therefore, it can lead to a severe compromising of the system by the user to the attacker. But, as attacker cannot shut down the devices without being noticed, therefore it prevents the attacker from opening the device and injecting advanced hardware backdoors or any hardware modifications. The attackers may not have that sufficient time to performs such attacks

B. Software Injection

It is based on local injection of malicious code, which can be performed through configured terminals in the system, i.e. for e.g. connected USB devices.

C. Operate user Interface

This attack is based on user input towards the local user interface. In this the attacker needs to manipulate the device through user interface. But, if the screen is locked and password protected, then the attacks can be avoided easily. In security-critical UI's the damage of such attacks are limited. Therefore, some systems keep second-factor authentication systems, so that the attackers cannot access the system easily.

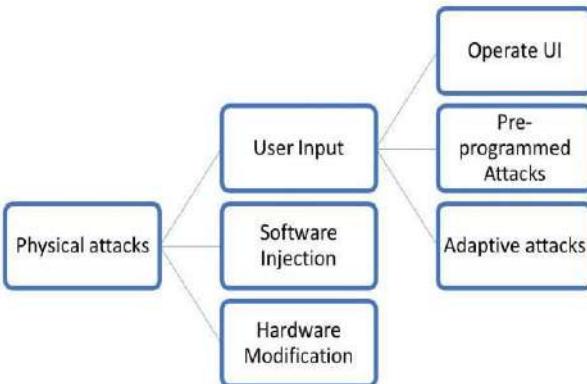


Fig.1: Classification of physical attack techniques

D. Pre-Programmed Attacks

This User Interface attacks rely on external devices connected to the targeted system. The purpose of this attack is to passively intercept the user input or actively inject preprogrammed sets of commands. Keylogger in local device is an example of passive attacks as it collects all the user input from the local system. BadUSB is an example of pre-programmed attack where a malicious input device injects keystrokes into the target system. Such attacks modify the system settings by opening a console or an administrative window. As attackers are not aware of the current system state, such attacks cannot be used to compromise (i.e. hijacking a system) an e-banking user session, without resorting to installing malware to the device.

III. UI POSSIBLE ATTACKS

As UI is one of the medium for the attackers to attack the target system, their main aim is to get the critical and most sensitive information through the input events when the authorized user is performing a security-critical operation.

A. Clickjacking

Clickjacking is also known as UI readdressing or I frame Overlay. In this, the attacker uses multiple transparent or opaque layers so that he can trick a user, by making him/her click on a button, document or a link of another page when they were intending to click on the top-level page. Therefore, through one click, the attacker tries to "hijack" the user system through one click. Sometimes, the attacker tries to send the malicious virus/worms/Trojan through documents. If the mails are not recognized properly then the system can be hacked by the user and all the data inside it can get compromised.

Clickjacking is a careful crafted combination of stylesheets, iframes and text boxes. This combination makes the UI so real that it makes the user feel to believe that the page is genuine and safe to provide sensitive information such as username, email id, passwords, pins etc. But they are not aware that they are typing into an invisible frame controller by the attacker.

Sometimes, in the busy schedule of everyday life, the attackers take the advantages by sending genuine mails of cloud storage,

security updates, message inbox full notifications, do and don't document. In such mails, they embed the malicious ransomwares, virus, Trojans etc. which forces the system's information to get compromised in industries.



Cloud Storage Requirement

The latest Cloud Storage solution is available. Locally saved files will be deleted from your machine. You must move files to the new Cloud Storage solution by end of month. We design the upgrade to be easy and compatible with the hardware and software you already use. It will provide a improved functionality, security, and reliability for sensitive work files. Follow the link below to download the installer to get started.

[Get Started!](#)

Fig.2: Example of Clickjacking through emails

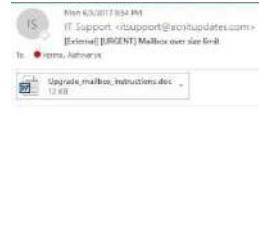


Fig.3: Clickjacking through documents on emails

B. Likejacking

Likejacking is a Facebook-specific version of clickjacking attack. The purpose of the attack is to get attacked through many clicks on a webpage items without any knowledge.

Facebook page has two layers when attackers represent the page. The back layer consists of the "Like" button which requires a mouse click cursor. The front layer consists of the UIs' which can be in the form of pictures, thoughts, videos, advertisements etc. which tricks the user making them click on it. The clicking of the Facebook "Like" button leads further spreading of spam to other users.

Likejacking have been evolved for money-making schemes. Such techniques are called as affiliate marketing. In affiliate marketing, it pays for every person who clicks and views the ads, click on the images for more information related to donations, sign-up for services or register for same.



Fig.4: Likejacking via donation advertisements

C. Cursorjacking

Cursorjacking is UI readdressing technique in which it deceived the users by using a custom cursor image, in which the pointer is displayed with an offset. The displayed mouse cursor is usually shifted to the right from the actual mouse position. Therefore, using the clever positioning of page elements and mouse cursor, the attacker can direct the user to click to the desired elements, making the user confused about the navigations in the page.



Fig.5: Cursorjacking Example

D. Password Manager Attacks

In our day-to-day life, users need to login in many web services, therefore they set up the authentication credentials with the enormous number of sites. As a result, the user who needs to keep different passwords for different sites is driven to use password manager. These password managers are available through browser built-in password managers, mobile password managers or 3rd party manager. This helps us in autofill the required password to the required sites.

The manual entering of the password on every login page get reduced via password manager. Once the password and username are entered, the password manager store the information for the future login in less time. Every password manager has its own autofill policies on different platforms with reference to browser, operating system, application software etc.

There are many vulnerabilities found in the password manager, due to which the attacker took the advantages for same. These attacks, not only stolen a million of passwords, but also it has compromised the complete website

1. THE EVIL COFFEE SHOP ATTACK

This attack is a kind of man-in-middle attack. In this, the attack take place through wi-fi or network router connectivity. They take the temporary control of the network router and modifies the arbitrary network traffic that are originating from or to a user's

machine. The user is unaware as he/she is not using the victim website and passwords are also getting extracted.

2. SWEEP ATTACKS

This attack works against any password manager that supports autofocus of password field. The target user needs to be connected to the Wi-fi hotspot which is controlled by the attacker.

There are 3 basic approaches that the attacker can use the landing page to sweep the passwords. In first approach, the landing page contains invisible I Frames, which is pointing to the arbitrary pages at multiple target sites. When the browser is loading the frame, then attacker injects a login form and JavaScript into every field. As user is unaware and feels it's a genuine one, it fills the details accordingly and the autofocus password manager auto-populates the corresponding password accordingly. In second approach, the attacker makes use of multiple windows, so that the user can gain access over the window, making them encourage to disable any popup blocker. Multiple windows are more noticeable than I Frames. They can be closed as soon as the passwords have been stolen. In third approach, it uses the chain of redirects. In this, the user requests some pages, the attacker responds by redirecting it to the site, where the attacker can able to learn the password. The injected JavaScript add a login page and hides the other page details. As soon as the password manager autofocus the password, the password gets exfiltrated and the browser is redirected to the next target page. This makes the user feels that the wi-fi internet is slow to load.

3. INJECTION

This is the one of the simplest attack where the login page is vulnerable page. Such site has login page of "http", after entering the username and password, it redirects to the "https" page for submission. Most browser block the "http" pages as it is vulnerable, but if only the login page is of "http" and rest is "https", the browser fails to block the same. Broken "https" connections are also vulnerable and leads to vulnerabilities as attacker can modify the login page using their own self-signed certificates.

4. EXFILTRATION

In this technique, the JavaScript in the attacker page has the username and password. It loads the invisible I Frame and pass the credentials as parameters. It also modifies the action of login form to submit to an attacker-controlled site. If the password manager does not have autofocus, then the attacker tries to trick the user to interact with the login form without making them realized about it, the same exfiltration techniques can be used to steal the password as soon as the password form is filled.

5. DEFENSES

The main motto of defense is secure filling of information, which requires modified passwords to work with modified browser. But, in some cases, while filling the login form, we are unable to recognize the security level of the page as login page is of "http" but after entering your credentials it redirects to "https". In such case, neither a password manager nor the browser provides security once the login form is filled with required credentials as javascript can read the password directly from the form. It can change the

forms' action so that it can submit the passwords to the stealing page, which is hosted by the attacker.

IV. CACHE POISONING

The concept of “www1, www2, www3...” etc. came into picture when “www” server was unable to handle all the load of the webpages alone. Therefore, to handle all the web servers and distribute the load to balance the complete website, this solution came into picture, which is also known as round-robin Domain Name Service.

For example, there are two web servers, www1.yourdomain.com(192.168.1.30) to www2.yourdomain.com(192.168.1.40). To balance the load for www.yourdomain.com on these two servers using round-robin DNS, the following lines are needing to be added in yourdomain.com zone file:

```
www1      IN A    192.168.1.30
www2      IN A    192.168.1.40
www      IN CNAME www1
www      IN CNAME www2
```

We need to restart the system and then ping www.yourdomain.com host. The address 192.168.1.30 will be coming as output. After restarting and pinging the same website, 192.168.1.40 will display. This is because the preceding configuration tells the name server to cycle through CNAME records for www.yourdomain.com host is both www1.yourdomain.com and www2.yourdomain.com.

As attackers need the medium to attack through look alike UI, therefore they make use of clone UI of the same website with different server connectivity, which are hard to be noticed by the authorized user.

Cache poisoning is the malicious techniques in which it manipulates DNS queries to insert data into an unprotected DNS server's cache. This poisoned data is later given out in response to client queries. Such data can direct clients to host that are running trojan web servers or mail servers, where the hacker may retrieve valuable information from users.

The best example for this is to login into a page which has different server of “<http://www1.example.com/loginpage>” which may be vulnerable and after entering the username and password, it redirects the page to “<https://www2.example.com/login successful>”

V. SOLUTIONS

To prevent these attacks, here are some measures that can result in reducing the adverse attacks to some extent, but it is hard to guarantee foolproof solutions.

Table V.1: Solution Table

Attacks	Solution
---------	----------

Clickjacking	X-Frame should be included as options HTTP header in all your Web Pages User should be asked to solve CAPTCHA. One-time code in URLs should be included to crucial pages. Frame buster JavaScript can be included. Evaluation should take place for Email protection. There should Use web application firewall in every browser.
Cursor jacking	This could be solved by “picking up” users cursor before it enters frame area, so technically user will hover from attackers' page
Password Manager Attack	Check the use HTTPS on both the login page and page where it submits to Use CSP (Content Security Policy) to prevent the execution of inline scripts, making injection of JavaScript directly into the login page ineffective Host the login page in a different sub domain than the rest of the site. This limits the number of pages considered same –origin with the login page reducing the attack surface

- An appropriate browser filter should be used while surfing the Internet. For example: NoScript is an appropriate choice. It works on Mozilla Firefox and it has some built-in capabilities to take control of certain UI redressing attacks such as clickjacking.
- Browsers should be upgraded regularly, and security measures must be applied in a timely manner. Most browser software vendors like Microsoft, Apple, and Mozilla release security advisories about potential vulnerabilities. These security advisories contain an updated fix and patch that should be installed to upgrade the requisite browsers. However, if automated updates are enabled, the system is updated regularly against potential threats. A user can also download individual security updates manually from vendor websites.
- Browser design requires a significant amount of change in the way User Interface components are handled. However, it becomes hard for the vendors to change UI on a regular basis. This attack projects how the design issues in XSS filters result in exploitation of vulnerability. Download Dialog Box Spoofing Browsers use a download dialog box in order to download a file from a server. This process acts as a notification to the user about the characteristics of the file. The download dialog box is displayed when a user clicks a hyperlink to download a specific file. It is a type of GUI displayed to the user for raising an alert. Attackers are spoofing download dialog boxes to trick users into downloading malicious files instead of authorized files. This attack is triggered on a wide scale to infect user machines with malware. Recently performed tests on Internet Explorer have CrossTalk—May/June 2011 33 PEOPLE SOLUTIONS TO SOFTWARE PROBLEMS dox in

the field of browsers, but vendors should take appropriate steps to secure the design interface.

- Users should not visit those pages that they are not sure of. Sometimes, being paranoid is an effective way to be secure. Always think twice about what you click. There are certain client-side browser filters available that help users substantially to make smart decisions if a potential threat is detected. For example, the NoScript plug-in works as an inline component with Mozilla Firefox to increase security. It enables the user to surf in a secure manner and raises notification against insecure objects and attacks such as XSS. Browsers as Internet Explorer come with built-in client-side protection against XSS attacks. Thus, potential combinations of client-side filters and user awareness can reduce the exploitation ratio of vulnerabilities.
- Users should be aware of the basic attacks on the Internet that can help them to understand exploitation attempts. There are several websites such as Threat post, Security Focus, and Register, etc. that provides substantial information about new research and attacks.

VI. DISCUSSION

In deterministic user interfaces where the same element clicks always transit to the same predefined UI state. However embedded terminals can have more complex UIs where this is not the case. For examples, clicking a button can lead to different states depending on the values in the current state. Our attack can be extended to encompass such cases as well. Clickjacking, likejacking, cursorjacking, password manager attacks are the powerful attacks against modern web applications.

We observe that UI security attacks are fundamentally attack on human perception. We discussed possible defense against our perceptual attacks and find that most defenses either have an unacceptable usability cost or do not provide comprehensive defense.

Table.2: UI Attacks table

Attacks	Strategy/area used for attack
Clickjacking	Buttons, embedded links in files
Likejacking	Like buttons in Facebook and other websites, social media advertisements
Cursorjacking	Introduction of new working cursor in the website used
Password Manager Attacks	3rd party software, JavaScript, I Frame, internet connectivity

As shown in the table, Clickjacking / Likejacking / Cursorjacking takes place through website click and explore. At times, this attacks also takes place while filling up the form, in which the users are not aware about the security of the website and to make the password memorable for them, at times they enter their common passwords. As a result, the attacker tries to attack the email account entered by the user entered while filling up the form and then it explores the other accounts linked with the email. In such

cases in which the email id, username (rare), password is same, can face a major compromise with the data. The attacks can also misuse the hacked account by sending the malicious links with the disguised UI, in which the linked user will feel safe while browsing without recognizing the planned attacks behind the UI. Therefore, the other users account and their sensitive information also gets affected and compromised.

In today's life, internet is the most necessary thing in their smartphones or desktop/laptops. At times, the user while exploring for the need of internet, gets connected through free Wifi hotspots in cafeteria, schools, colleges, hospitals, car services, railways etc. Attackers takes the advantages of such free hotspots and tries to attack the other devices by sending the vulnerable software, documents, messages for feedback along with the link which contains malicious software that get installed through one click. At times, the attackers try to put the server down such that they can able to attack the password manager of their device and get all the passwords. In such case, the user feels that the server is busy, or it is not available, but unknowingly their usernames and passwords get compromised.

The overall scenario is clear that if the user has well known knowledge of the safe/unsafe websites, mails, files etc. then the user can able to recognize and be careful while accessing the website.

VII. SCOPE

Thus, we understood that how the user can be fooled and their security can be threatened with the help of User Interface. The User Interface is one of the medium through which the user interacts with the system throughout the world. Such attacks can be compared with the story in which the donkey was disguised in lion costume to have food in farms and to fool other farmers. In this, the attackers are disguised themselves in the form embedded links, file contents, form filling, while login session to 404 page not found, pop-ups etc.

If the people are not aware about what is going on during the important activity like login, uploading etc.; all the sensitive information along with their account can get compromised. In such case, the attacker can take the advantage of the files and ask for money to safe their files from compromising in their system. Ransomware is one of the most dangerous software, which encrypts all the data and for their decryption, a major amount of ransom is asked. At times, when the attackers get their ransom they do not decrypt the information also.

Therefore, they need to be alertness in the user itself, which can be done through education and other mass media awareness. Also, care must be taken to follow up the solutions such that the people and their information can be safe in this cyber world. For that, the user must make sure about all the solutions as mentioned in paper must be maintained.

There should be timely scan of the computer so that if there is a malicious software or trojan, if present, can be removed. There should be a check of the user interface by scanning them from top to bottom; including their grammar, spell check, document extension, link tooltip detail (if possible) etc.

Also, care must be taken to access and use of the sensitive information only on secured website (i.e. https). It should also have copyright security too. Use of username/email along with the password must be used in secured web login page, must be avoided in unsecured pages (i.e. http). The use of one common password and email id must be avoided. Multiple password which are easy to memorize can be used. Before opening the attachments or links, it should be properly checked. If it is not proper, report the same. The unknown hotspot connectivity to your devices must be avoided.

This small check of the user interface helps the user to stay secure to minimum extent. This information should be present among all the people while browsing the network

REFERENCES

- [1] Present Hacking in the Blind: (Almost) Invisible Runtime User Interface Attacks – Luka Malisa, Kari Kostiainen, Thomas Knell, David Sommer, and SrdjanCapkun – Department of Computer Science, ETH Zurich.
- [2] Ethical Hacking: A Technique to Enhance Information Security, Gurpreet K. Juneja – International Journal of Innovative Research in Science, Engineering and Technology, Vol.2, Issue 12, December 2013.
- [3] Hacking in the Blind: (Almost) Invisible Runtime User Interface Attacks – Luka Malisa, Kari Kostiainen, Thomas Knell, David Sommer, and SrdjanCapkun – Department of Computer Science, ETH Zurich.
- [4] Ethical Hacking: A Technique to Enhance Information Security, Gurpreet K. Juneja – International Journal of Innovative Research in Science, Engineering and Technology, Vol.2, Issue 12, December 2013.
- [5] Password Managers: Attacks and Defenses - David Silver, Suman Jana, Eric Chen, Collin Jackson, and Dan Boneh, Stanford University Carnegie Mellon University
- [6] Web Hacking: Attacks and Defence, by McClure Stuart
- [7] <https://www.linkedin.com/pulse/future-user-interface-hacking-brain-anshuman-singh>
- [8] <https://en.wikipedia.org/wiki/Clickjacking>
- [9] <http://whatstechtarget.com/definition/clickjacking-user-interface-or-UI-redressing-and-IFRAME-overlay>
- [10] <https://www.sophos.com/en-us/security-news-trends/security-trends/what-is-likejacking.aspx>
- [11] <https://www.sophos.com/en-us/security-news-trends/security-trends/what-is-likejacking.aspx>
- [12] <https://koto.github.io/blog-kotowicz-net-examples/cursorjacking/>
- [13] <http://blog.kotowicz.net/2012/01/cursorjacking-again.html>
- [14] <https://blog.acolyer.org/2017/02/06/password-managers-attacks-and-defenses/>
- [15] https://www.owasp.org/index.php/Clickjacking_Defense_Cheat_Sheet
- [16] <https://www.troyhunt.com/clickjack-attack-hidden-threat-right-in/>
- [17] http://www.nds.rub.de/media/attachments/files/2011/09/UIRedressing_Marcus-Niemietz.pdf

QR Code based Authentication System for Banking

Ashlesh Patel, Pragnesh Patil, Harsh Shah, Nihir Shah, Ashwini Patil
Computer Science, Thakur College of Engineering and Technology, Mumbai, India

Abstract—The main idea of our project is to build up a new system of banking and to overthrow the old system of using OTP's and instead use an inventive secure authentication method which utilizes a QR code; System which uses two way authentication by using a random number and registered IMEI number, acting as a token of authentication. As the information stored in the QR code is in encrypted format it is secured. QR code is scanned with the help of scanner in smartphone. The result generated by scanning a QR code is a combination of a random number which is generated by random number function and IMEI number registered by the user. If there is internet connectivity in the smartphone, the generated string is automatically entered in the login page and is redirected to the home screen of the banking page. If there is no internet connectivity in the smartphone, the encrypted string generated by the scanning of QR code gets decrypted and a new six digit pin code is generated which is to be entered manually by the user which then redirects to the home screen of the banking website. The objective is to develop an security system using a two factor authentication: a trusted device which will scan a QR code and act as a token and a password known by the user. Our aim is to enhance the security of the banking transactions and provide users with a convenient way of performing the transactions.

Keywords—QR code, IMEI number, authentication, AES algorithm, registration, encryption

I . INTRODUCTION

A . BACKGROUND

As most of the transactions in today's era is getting digitalized that fears the users of losing their credentials as we are still using the primitive measures of providing security. So there arises a need of providing more enhanced security measures to all these online transactions which will assure the user's information not getting tampered. Providing security by means of QR code is more efficient than password, fingerprints and face detection system. The QR code is a matrix which is an array of square modules arranged in a square pattern[1]. The three corners of the QR code forms a unique pattern that assists easy location of its position, size and inclination.

B . PROBLEM STATEMENT

The current system consists of OTP which is sent to user via SMS or email but email spoofing or man in the middle attack can occur. The password system provided security against unauthorized

access but the evolution of different attacks like brute force dictionary attack made this system ineffective.

An alternative of this system is given in form of a two factor authentication which uses password as the 1st factor and a randomly generated code as the 2nd factor. With advantages of this system also came the disadvantages. For example spoofing of network, delay in delivery of the OTP. This system was replaced by more efficient system which used a QR code instead of OTP but didn't solve the problem related to delivery of the code.

The new system we offer generates QR code which consist of the IMEI number and a 4 digit code. The 2nd factor of authentication is replaced by an android application installed on the registered phone. This system removes the problem of network spoofing, man in the middle and the delay in receiving the unique code.

II. LITERATURE SURVEY

In the literature survey we did the survey of certain systems which are common used.

The traditional password system fails to confirm the user identity as there is a threat of phishing of users credential. Using QR code will thus eliminate the problem of phishing as it uses user's registered IMEI number encrypted with a random number and then hashed to generate a string[2]. As a string is hashed, the attacker cannot retrieve the data in its original format. A variety of size of symbols is provided together with four levels of error correction which learns something new from each system.

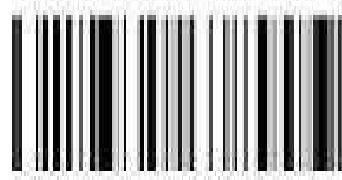
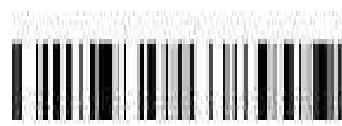


Fig 1. Bar Code

TABLE I

Paper Name	Disadvantages
OTP Encryption Techniques in Mobiles for Authentication and Transaction Security	<ul style="list-style-type: none"> 1. Most OTP systems are susceptible to real-time replay and social engineering attacks. 2. OTPs are also indirectly susceptible to man in the middle (MITM) and man in the browser (MITB) attacks.
SURVEY ON INFORMATION HIDING TECHNIQUES USING BARCODE	<ul style="list-style-type: none"> 1. They Can Breakdown 2. Label damage. 3. Scratched or crumpled barcodes may cause problems
A Secure Credit Card Protocol over NFC	<ul style="list-style-type: none"> 1. Security problems. 2. Sensitive data can be accessed if card is lost.

III. PROPOSED METHODOLOGY

The ideology behind developing the system using the iterative model is that it will allow the software developers to take the advantage of previously developed versions of the system. The key steps in developing the system are to start with the implementation of the smaller essential modules and then as per the software requirements iteratively enhance the system. By using the current system, the developer can identify the flaws in the system and make improvements by providing new functionality to the system.

At each iteration, design modifications are made and new functional capabilities of the system are implemented. This process continues until the main objective of providing the finest security is achieved. The development consists of initialization step, Iteration step and the control list of the module to design a system which replaces the current OTP based system. The QR based authentication system lets the user input the password, if the user is authenticated then an encrypted string in the form of QR code is displayed on the screen.

The user gets authenticated if the encrypted string matches the IMEI number present in the database[3]. Designing a system for visually challenged people in which the person will be able to hear the code once the QR code is scanned. The visually impaired can enter the code via text-to-speech feature of the web application.

Prior to QR code there were some primitive authentication techniques available namely Username and password, Barcode, Fingerprints, Face recognition. The security was compromised as the Username and password faced the problem of phishing and eavesdropping. The limitation with barcode is it can only store up to 20 digits so complex passwords cannot be generated using barcode. The crumpled and scratched barcode does not provide effective security.

The devices and the technology used for the Fingerprints and Face identity is not cost effective and it also suffers from accuracy problem. Thus to overcome from all the disadvantages of the existing system, QR code is introduced. QR

code is Quick Response code. It was introduced in 1994 by Denso-Wave, a Japanese company subsidiary of Toyota. QR code can generate more complex passwords as it can store up to 4296 alphanumeric characters which comes over the disadvantage of the barcode. As it is a two-dimensional barcode, it can be read from any direction.



Fig 2. QR- code

There are two sections in this system. In the encoding section conversion of input data to a QR Code symbol takes place. In this the data analysis and encoding is done then error correction, coding and the final message is structured. The second section decodes the QR Code image and displays the data contained in that QR code[4]. The decoding procedure starts with differentiating the black and white modules and then reorganizes the modules to obtain the decoded format information.

IV. PROPOSED SYSTEM

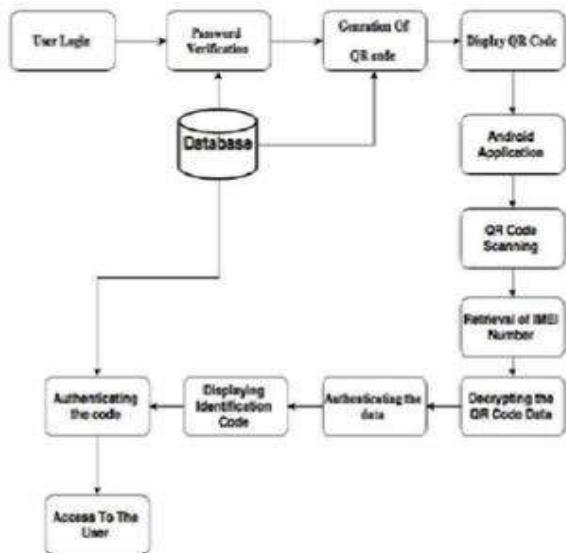


Fig 3. Registration system diagram

The following steps give the information on how to complete the registration process:- Firstly user would go

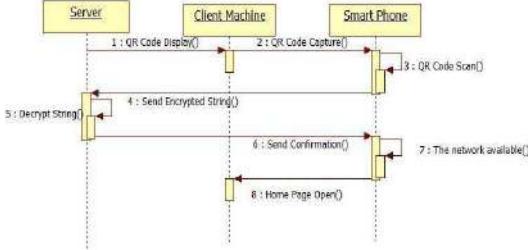


Fig 4. System overview diagram

into the registration section and submit the details like his/her username, Password, IMEI number of the phone. Once the data is validated it will be stored in the database. The data of the database server will produce the public key and private key and store it in the server.

After this, the user will proceed to download the application and install it on his/her phone. When user runs the application for the first time, the class files of public key and private are created and stored into the internal storage of mobile phone. In a registration if the user does not enter all the values like username, password, IMEI number, mobile number and email address then registration process will not get completed. Validation is most important part in the registration process; if validation is not successful then user will not be able to login

V. IMPLEMENTATION

A. Online Authentication System

This method is implied when the user's mobile has access to internet and is online, in which using the public key, random number and the IMEI number are encrypted forming a string. With help of this encrypted string, QR code is generated using QR code generation function. Once it is generated, it will be displayed on the client's machine and the client will scan this QR code with his mobile phone. As it is online mode, after scanning the generated string (combination of a random number and IMEI number) automatically enters the login page with help of the internet. If login is successful, the home page of banking website is displayed. So in our system there is no real need of remembering the password. The user's public key is used to decrypt the string and also makes sure it exists in our transaction table with the random number and then modifies the row of the table. Then the server checks whether the IMEI is correct or not. If found correct, it assigns that IMEI to the legitimate user[5]. Once the login is successful, the transaction row is deleted and a new QR code image will be generated when the user wish to login again. Now the PHP session is created and when user finishes his transaction and logs off, the session is destroyed.

B. Offline Authentication System

This method is implied when the user's mobile doesn't have access to internet and is offline, in which authentication system, a unique

six-digit number is generated using pin code generation algorithm which is formed by the encrypted string(IMEI number and random number). This unique code has to be entered manually on the login page with his username.

After the pin code is entered, the IMEI number is verified with the stored database by the server. If the number matches the database records then the user is a legitimate user and homepage of bank gets displayed and the timestamp is also checked. If the user does not login in the available timestamp of 5 minutes then the session is destroyed and the user won't be able to login.

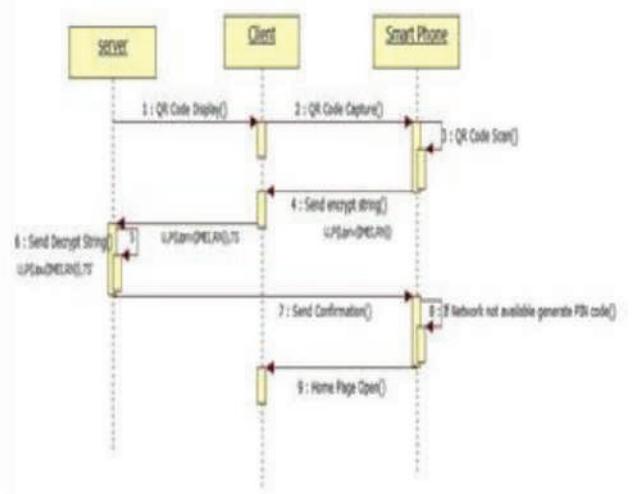


Fig 5. Offline authentication

VI. SECURITY

The QR code and the encryption algorithm provide a more powerful security to our system. It does not get vulnerable to the man-in-the-middle attack because the communication between the user and the server is always in the encrypted form. Username also cannot be reused again as it gets deleted after the user logs out of the system. Also for mobile application, person needs to have the password so that it can't be attacked by any other means. If the untrusted person knows to handle the internal storage then only the security problem is created. A phishing attack is possible on the mobile phone if we replace the application by some other application and the pass code also gets covered but it is still not possible if he does not own a certificate[6].

One of the key security feature in offline authentication mode is the timestamp, if user fails to login within the given time period then the login fails.

VII. CONCLUSION & FUTURE SCOPE

This work provides additional security with the traditional way of online authentication of banking; which includes username and password. However, by adding QR code authentication the security measures for banking are enhanced. Two factor authentications are considered in this system. With the help of this QR code security is increased during the login of the particular bank. Depending on the authentication only the client will be able to perform the transaction. In future we would like to add voice input command feature to our website and android application. It will help the user to do his work comfortably. We would like to

use some advanced encryption and decryption algorithm, better than AES.

REFERENCES

- [1].Snehal.Kalbhor,Ashwini.Mangulkar,Mrs.SnehalKulkarni“Android App for Local Railway Ticketing Using GPS Validation”.International Journal of Emerging Trends in Science and Tech.,IJETST-Vol-01,Issue-01,Mar-2014,Pages71-74.
- [2].Fu-HauHsu,Min-HaoWu,Shiu-JengWANG,“Dual-watermarking by QR-code Applications in Image Processin”.9th International Conference on Ubiquitous Intelligence and Autonomic and TrustedComputing,DOI10.1109,2012,Pages 638-643.
- [3]. Mrs.ShantaSondur, Ms.TanushreeBhattacharjee “QR-Decoder and Mobile Payment System for Feature Phone”, VESIT,International Technological Conference(I-TechCON)-Jan.03(2014),Pages13-15.
- [4]. SomdipDey, B. JoyshreeNath and C. AsokeNath “OTP Encryption Techniques in Mobiles for Authentication and Transaction Security” Institute of Information Systems Argentinierstrasse -2009.
- [5].Dr.A.P.Adslu,GayatriKumbhar VrundaChincholkar, YogeshKamble, AnujaBankar “Automated Exam Process using QR Code Technology” International Journal of Application or Innovation in Engineering &Management,(IJAIEM)-ISSN 319-4847,Vol.3,Issue 4,April-2014,Pages-296-298.
- [6]. Ben Dodson, DebangsuSengupta, Dan Boneh, and Monica S. Lam “Secure, Consumer-Friendly Web Authentication and Payments with a Phone”, International Journal of Applied Engineering Research, ISSN 0973-4562, Vol. 8, No. 17 (2013).

Smart City Traveler on Android

Tanmay Patil, Abhijeet Prasad, Pratik Singh, Suyash Singh, Manish Rana

Department of Computer Engineering Thakur College of Engineering & Technology, Mumbai, India.

tanmaypatil.tct@gmail.com, abhijeetprasad333@gmail.com, pratik.singh287@gmail.com, Suyasssssssh09@gmail.com,
manish.rana@thakureducation.org

Abstract: We are pleased to present “Smart City Traveler” project and take this opportunity to express our profound gratitude to all those people who helped us in completion of this project. We thank our college for providing us with excellent facilities that helped us to complete and present this project. We would also like to thank the staff members and lab assistants for permitting us to use computers in the lab as and when required. We express our deepest gratitude towards our project guide for him/her valuable and timely advice during the various phases in our project. We would also like to thank him/her for providing us with all proper facilities and support as the project co-coordinator.

I. INTRODUCTION

There has been a huge development in information technology recently. In addition to this, GIS has been commonly used in a variety of fields such as tourism industry enabling people from around the globe to reach out to one another other.

A network is a set of linear features that are interconnected in GIS. Tourists visiting a new locality are provided with city map pamphlets or maps that need to be viewed by the tourists to find the fine printed locations to visit.

So, here we propose a Smart City Traveler system which helps a traveler to explore a city visiting for first time and wanted to explore the city.

Our proposed system is a GIS based system that automatically creates schedule for the traveler for a whole day or a user specific timing.

The development of this system is done in android using Foursquare API which provides real world pre-defined places like Restaurants, Tourist Spot, etc.

The system also provides info about these places that includes, how to reach them, description of the place along with images of the place and address contact numbers if available.

Thus, the proposed system is a complete guide to help new comers to know about a new city and navigate the city with an internet connection.

II. COMPARATIVE ANALYSIS OF EXISTING TECHNOLOGY

a) Human Intervention

If the information about eateries in a particular area are not present in the provided database or the websites then the human intervention would be required to do the ground work and find the information regarding the same.

b) Regional Impact

It may so happen that in the same area or say different areas for example borivali west and Kandivali east there may be presence of a food stall with the exact same and features. Also it may so

happen that a particular food type is dominated in an area specified so that may have a regional impact.

c) Financial Impact

The system at its initial stages is not having any considerable financial cost added to it development but at later stages if all the progress goes according to the planning then it may prove as a financial income for the developers.

III. PROPOSED METHOD

- a) The purpose of developing this android application is to create a schedule for the traveler travelling to city and wanted to explore the city by specifying the time in hours. System then promptly analyzes the questionnaire and creates a schedule for traveler based on provided information and time.
- b) The development is done in two technical languages as Java for Android Application for User/Traveler and Asp .net for Web portal which is used by Admin. Initially, the traveler needs to register himself by filling up the details using android application.
- c) After successful registration, user can login now using login credentials which then proceeds with questionnaire where application ask user about their likings and habits.
- d) Based on questionnaire, application promptly analyzes for the place based on user specified information and time.
- e) The application is capable enough to search the place automatically based on Foursquare API.
- f) This application also helps you to find the places nearby you or around the world.
- g) After searching a locality, the map will show the specific details such as name, area, location, phone no. & kilometers from the current location of the given user.

IV. DATA EMBEDDING PHASE

- a) This project has a login page which prevents unauthorized access.
- b) This system can be used to view the location view in map that the user wishes to reach.

- c) The user can also find the paths to follow to reach the final destination in map which gives a better view to the users.
- d) Since the location can be viewed in map, the user can even zoom in and zoom out to get a better view.
- e) The usage of this application greatly reduces the time required to search for a place.
- f) The application also leads to quicker decision making with respect to places to visit.

V. EXPECTED OUTCOME

A. Input:-

- a) Query for places

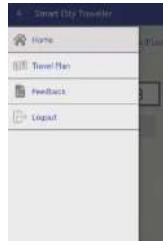


Figure 2: Input image

- b) List of places one can travel to



VI. FUTURE SCOPE

- a) The scope of this project is to develop a complete application regarding this issue that the traveler's might encounter. For implementing this system we will be using
- b) Active Server Pages.NET, SQL Server, Active Data Objects.Net Overview
- c) Android Studio, Microsoft .NET Framework
- d) To achieve the objectives of the project.
- e) Thus, the central aim of this project will be to:-
- f) Our proposed system is a GIS based system that automatically creates schedule for the traveler for a whole day or a user specific timing
- g) Traveler can perform task such as navigating through city maps: zoom into areas and also displays various amenities

such as tourist places: hospitals: Institutes as well as bus stops and train station.

- h) The development of this system is done in android using Four Square API, which provides real world pre-defined places like restaurants: Tourist Spot:
- i) The system also provides info about these places that includes: how to reach their, description of the place along with images of the place arid address contact numbers if available.
- j) Thus: the proposed system is a complete guide to help new comers to know about a new city arid navigate the city with all internet connection.

VII. CONCLUSIONS

This project has a login page which prevents unauthorized access. This system can be used to view the location view in map that the user wishes to reach. The user can also find the paths to follow to reach the final destination in map which gives a better view to the users. Since the location can be viewed in map, the user can even zoom in and zoom out to get a better view. The usage of this application greatly reduces the time required to search for a place. The application also leads to quicker decision making with respect to places to visit.

REFERENCES

- [1] Jian Meng, Neng Xu, "A Mobile Tourist Guide System Based on Mashup Technology "ISBN 978-1-4244- 7618-3 /10 @2010 IEEE.
- [2] Xiaoyunshi," Tour-Guide: Providing Location-Based Tourist
- [3] Information on Mobile Phones "ISBN 978-1-4244-7547-6/10 @2010 IEEE.
- [4] "DROID M-Learning" paper published in International Conference on "Global Economic Turmoil and Strategic Advantage (2012)" [ISBN: 978-81-910118-1-4].
- [5] V. Kostakos, T. Ojala and T. Juntunen, "Traffic in the Smart City: Exploring City-Wide Sensing for Traffic Control Center Augmentation," in IEEE Internet Computing, vol. 17, no. 6, pp. 22-29, Nov.-Dec. 2013.
- [6] Sassi and F. Zambonelli, "Coordination Infrastructures for Future Smart Social Mobility Services," in IEEE Intelligent Systems, vol. 29, no. 5, pp. 78-82, Sept.-Oct. 2014.
- [7] A. Jahangiri and H. A. Rakha, "Applying Machine Learning Techniques to Transportation Mode Recognition Using Mobile Phone Sensor Data," in IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 5, pp. 2406-2417, Oct. 2015.

AUTHORED BOOK

- [1] Recommender Systems: The Textbook, Springer, April 2016, Charu C. Aggarwal.

Android Controlled Fire Fighting Robot

Chinmay Verlekar, Pratik Patel, Manoj Dixit, Aakash Kanojiya, Vaishali Nirgude

Computer Engineering, Thakur College of Engineering, Mumbai, India

Abstract—The proposed work is to develop a fire fighting robot using Arduino Uno R3 for remote operation. A robot detects and extinguishes fire that can be controlled from a specified distance. It uses flame sensor for fire detection and Arduino board for processing. Fire extinguisher is used to extinguish the detected fire. The robotic vehicle is loaded with water tanker and a pump which is controlled over wireless communication to throw water. An Atmega 8 microcontroller is used for controlling the desired operation of Robot at the transmitter side, User sends command to the microcontroller which is on the robot and then translated into robotic movement either to move forward, backward and left or right etc. At the receiving end five motors are interfaced to the microcontroller where four of them are used for the movement of the vehicle and the remaining one to position the arm of the robot. Remote operation is achieved by any smart-phone with Android OS, upon a GUI based touch screen operation. The android application device transmitter acts as a remote control that has the advantage of adequate range, while the receiver have Bluetooth device fed to the microcontroller to drive DC motors via motor driver IC for necessary work. A water tank along with water pump is mounted on the robot body and its operation is carried out from the microcontroller output through appropriate signal from the transmitting end. A motor driver IC is interfaced to the microcontroller through which the controller drives the motors. The robot is also equipped with cameras and ultrasonic sensors. The camera plays role in giving feedback to user and in finding the source of fire. Ultrasonic sensors are used to avoid collisions during movement. Feedback provided by camera

on the robot will display on a screen of Smartphone. Further the robot works on IOT mechanism which will help to identify the exact location of the fire and providing an alert message to user.

Keywords—Arduino, Ultrasonic Sensor, Temperature sensor, Bluetooth module, Auto sprinkler, Camera Movable pipes

I. INTRODUCTION

Our proposed project aims to develop an android controlled fire fighter robot that can be used to extinguish fires through remote handling. The vehicle consists of a water tank along with a pump which can throw water when needed. The system uses an 8051 microcontroller for this purpose.

The android device is used as a transmitter to send over controlling commands to the vehicle. The android device provides a good touch based GUI for controlling the robotic vehicle.

The Bluetooth receiver on the vehicle is used to receive those commands sent by the android device. These are then fed to the motors responsible for controlling the vehicle movements in front, back, left and right directions. The Bluetooth receiver is interfaced with an 8051 microcontroller for this purpose. The microcontroller after receiving input commands, operates the motors through a driver IC for vehicle movements.

The use of android has one more advantage in addition to improved GUI. It allows use of Bluetooth technology for communication allowing the vehicle to operate in a good range from the device. The system can also be later enhanced through the use of a wireless camera to be used for monitoring purposes.

II. LITERATURE SURVEY

1. International Journal of Innovative Research in Science, Engineering and Technology (An ISO 3297: 2007 Certified Organization)	FEATURES:	FUTURE WORK:
	<ul style="list-style-type: none">1. Intelligent fire detection and suppressions.2. Locate the position of fire origin.3. Protection of property from loss.4. Minimization of ecological consequences	<p>The project can be implemented with advancements like: A robot mounted with camera so that the camera will record the situation and the firemen can monitor it from control room. A headset with full color display. A mission control centre.</p>

	<p>5. Reliable.</p> <p>6. Path tracking.</p>	
<p>2. International Journal of Engineering Research and Development e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com Volume 10, Issue 4 (April 2014), PP.43-47</p>	<p>1) The Robot is able to move autonomously for one meter on the competition field</p> <p>2) Find the points of possible fire focuses using a sound signal to indicates when they are found</p> <p>3) Robot needs to climb up and down in two slopes located randomly on the competition field</p>	<p>1) The fire sensors are capable of detecting fire from 2 m away.</p> <p>2) The sensor detection range is 360 degrees.</p> <p>3) The fire extinguisher works for a continuous time till fire extinguishers.</p> <p>4) The robot moves at 2.5 ft/sec on a flat surface, accelerate from 0 to 1 ft/sec in two seconds.</p> <p>5) The robot turns within a 6" radius and utilizes circular or octagon design in order to minimize possible impact area.</p>
<p>3. H. P. Singh, AkanshuMahajan-d, N. Sukavanam-b, VeenaBudhraja-c, Swarn Singh-a, Amit Kumar-a and AnadiVashisht-c harendramaths@gmail.com (a- Department of Mathematics, Sri Venkateswara College, University of Delhi, b-Department of Mathematics, IIT Roorkee, c-Department of Statistics, Sri Venkateswara College, University of Delhi and d- Department of Electronics, Sri Venkateswara College, University of Delhi)</p>	<p>1. Can be used as a mobile surveillance system.</p> <p>2. Can be used as a fire extinguisher at places out of human reach.</p> <p>3. Can be used in security system.</p> <p>4. Can be used in chemical and oil industry, nuclear plants, mine fields and dangerous substance transport.</p>	<p>1. For detecting fire with 100% accuracy so that the robot can differentiate between industrial fire and an ordinary flame, we will be adding three more type of sensors i.e. temperature sensor, smoke sensor and thermal sensor.</p> <p>2. To save people who get trapped in the fire, we will again use transmission of wireless signals to the fire fighting person so that they can easily locate the people and hence save a lot of precious time</p> <p>3. We can replace water in pumping system with pressurized carbon dioxide to fight with fires caused due to electric short circuits.</p> <p>4. For domestic use, we will try to implement motion planning using neural networks so that the errors can be minimized in mapping of the house.</p>

III. DESIGN PHASE

A. Block Diagram

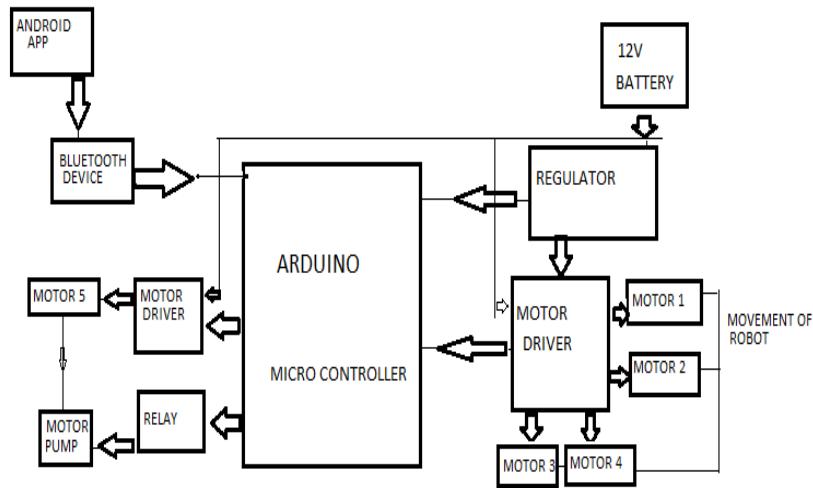


Figure 1: Block Diagram

IV. REQUIREMENT SPECIFICATION -

A. Software

1) Eclipse (Keil Compiler)

Keil an ARM Company makes C compilers, macro assemblers, real-time kernels, debuggers, simulators, integrated environments, evaluation boards, and emulators for

ARM7/ARM9/Cortex-M3, XC16x/C16x/ST10, 251, and 8051 MCU families.

Compilers are programs used to convert a High Level Language to object code. Desktop compilers produce an output object code for the underlying microprocessor, but not for other microprocessors. the programs written in one of the HLL like 'C' will compile the code to run on the system for a particular processor like x86 (underlying microprocessor in the computer). For example compilers for Dos platform is different from the Compilers for Unix platform So if one wants to define a compiler then compiler is a program that translates source code into object code.

2) Android Code:

Android Code: The first program makes use of the continuous data stream to detect, and stop the car when it has lost connection to the phone. Because of this continuous stream of characters, it also needs to filter repeated characters. The program is called "RC_Motor_Shield_Continuous.ino". It uses the "millis()" function which returns the number of milliseconds since the Arduino's execution started. This function lets the Arduino know when 500 milliseconds have elapsed since the last command was received. Once 500 milliseconds have gone by without receiving a command

B. Hardware

1) Microcontroller:

a. Features:

8K Bytes of In-System Programmable (ISP) Flash Memory

4.0V to 5.5V Operating Range

Fully Static Operation: 0 Hz to 33 MHz

256 x 8-bit Internal RAM

32 Programmable I/O Lines

Three 16-bit Timer/Counters

Eight Interrupt Sources

Full Duplex UART Serial Channel

2) Motors and Drivers

The L293 and L293D are quadruple high-current. Half-H drivers. The L293 is designed to provide bidirectional drive currents of up to 1 A at voltages from 4.5 V to 36 V. The L293D is designed to provide bidirectional drive currents of up to 600-mA at voltages from 4.5 V to 36 V. Both devices are designed to drive inductive loads such as relays, solenoids, dc and bipolar stepping motors, as well as other high-current/high-voltage loads in positive-supply applications. All inputs are TTL compatible. Each output is a complete totem-pole drive circuit, with a Darlington transistor sink and a pseudo-Darlington source. Drivers are enabled in pairs, with drivers 1 and 2 enabled by 1,2EN and drivers 3 and 4 enabled by 3,4EN. When an enable input is high, the associated drivers are enabled, and their outputs are active and in phase with their inputs. When the enable input is low, those drivers are disabled, and their outputs are off and in the high-impedance state. With the proper data inputs, each pair of drivers forms a full-H (or bridge) reversible drive suitable for solenoid or motor applications.

3) Bluetooth Module

Bluetooth is a wireless technology standard for exchanging data over short distances (using short-wavelength UHF radio waves in the ISM band from 2.4 to 2.485 GHz[3]) from fixed and mobile devices, and building personal area networks (PANs). Invented by telecom vendor Ericsson in 1994,[4] it was originally conceived as a wireless alternative to RS-232 data cables. It can

connect several devices, overcoming problems of synchronization.

Bluetooth is managed by the Bluetooth Special Interest Group (SIG), which has more than 20,000 member companies in the areas of telecommunication, computing, networking, and consumer electronics.[5] Bluetooth was standardized as IEEE 802.15.1, but the standard is no longer maintained. The SIG oversees the development of the specification, manages the qualification program, and protects the trademarks.[6] To be marketed as a Bluetooth device, it must be qualified to standards defined by the SIG.[7] A network of patents is required to implement the technology, which is licensed only for that qualifying device.

V. CONCLUSION

This project entitled as “Android Bluetooth Base Control Fire Fighting Robot” was created and developed effectively. For that demonstration purpose a prototype module is built for live demonstration, answers are discovered to be acceptable. As it is a demo module a toy type machine is built with less cost. While creating and developing this proto type module, we've consulted couple of expert's individuals who're understand in a variety of fields, these professionals working at different organizations goes to Hyderabad assisted us while fabricating the robot. As it is a prototype module, much amount isn't invested, the entire machine is built with in your area available components, particularly the mechanical components utilized in this project work are acquired from mechanical fabricators, and they're less than the necessity, large amount of modifications should be transported in design and it is essential to really make it just as real working system. Hence, the module will be enhanced further for acquiring better results. This project says creating a relatively inexpensive, high precision fire fighting robot that is targeted to manage through android mobile (remote). The thought of controlling through remote would be to boost the operator safety. The end result from the thesis is a straightforward robot that is controlled with a wise android phone & also has got the voice instructions. This thesis aims to supply simple recommendations for individuals thinking about building robots. As pointed out earlier, the work continues to be transported out several occasions and also the goal of the thesis would be to familiarize the scholars with basic principles of Adriano and Android to construct anything possible.

REFERENCES

- [1] W. Budiharto, Membuat Robot Cerdas, Jakarta: Gramedia, 2006.
- [2] Ratnesh Malik, “Fire Fighting Robot : An Approach”, Indian Streams Research Journal Vol.2,Issue.II/March; 12pp.1-4
- [3] Kristi Kosasih, E. Merry Sartika, M. Jimmy Hasugian, danMuliady, “The Intelligent Fire Fighting Tank Robot”, Electrical Engineering Journal, Vol. 1, No. 1, October 2010
- [4] H. P. Singh, AkanshuMahajan, N. Sukavanam, VeenaBudhraja ,”Control Of An Autonomous Industrial Fire Fighting Mobile Robot”, DU Journal of Undergraduate Research and Innovation
- [5] Swati A. Deshmukh, Karishma A. Matte and Rashmi A. Pandhare, “Wireless Fire Fighting Robot”, International Journal For Research In Emerging Science and Technology
- [6] LakshayArora, Prof.AmoJoglekar, “Cell Phone Controlled Robot with Fire Detection Sensors”, (IJCSIT) International Journal of Computer

Blockchain Technology and its Application in Retail

Kanishk Verma¹, Vikas Singh²

Application Development Analyst, Accenture Mumbai¹, Assistant Professor, TCET,Mumbai²

Abstract—In the late 1990s e-commerce revolution had struck the retail market and the retailers who had then migrated from their traditional systems to e-commerce know the potential of disruptive revolutionizing technologies. Given its wide range of potential applications and required changes at infrastructure level, blockchain being a shared ledger where all financial transactions are recorded, eliminating the errors that can occur when each member participating in a transaction maintains its own data set for the transaction acts as the perfect lubricant and catalyst between trusted as well as untrusted entities in a business

Keyword – Blockchain, Bitcoin, Cryptocurrency, Retail, Hash, Loyalty, Hash, Transparency

I. INTRODUCTION

Satoshi Nakamoto back in 2008 proposed a system for electronic transactions without relying on trust. It is nothing but peer-to-peer networking coupled with proof of public history of transactions that in turn becomes computationally impractical for an attacker to change and this system was implemented only a year later through Bitcoin. The underlying technology which decentralizes transactional data for sharing across a large network of untrusted participants is coined as Blockchain. [3] Although this technology is mainly adopted in digital currency but it is also a promising technology for other wide areas. From healthcare industry to finance industry or throughout retail, blockchain technology has emerged to be the next big thing. This research paper serves to provide a simplified detail on blockchain and also serves to introduce its various applications and emphasizes on its application for Retail industry.

Blockchain is a way of digitally recording data and transactions. It acts as a distributed database of records or a public ledger of digital events or transactions that have been shared among participating parties across a large network of untrusted participants. Requirement of a third party verification is eliminated, thus disrupting any sector that leverages it traditionally. Each transaction in the public ledger is supposed to be verified by consensus of the majority of participants in the system. Once entered information will never be erased as it is immutable.

Even in its nascence, blockchain being a shared ledger where all financial transactions are recorded, eliminating the errors that can occur when each member participating in a transaction maintains its own data set for the transaction. Within the blockchain network aggregated data on transactions that have occurred are listed in blocks including timestamped blocks, act as data structure of blockchain. Blockchain technology itself is non-controversial and has worked flawlessly over the years and is being successfully applied to both financial and non-financial applications.

As a part of this paper to understand how retail organizations view blockchain, we conducted a survey of many retail professionals and the findings have been shared and explained in subsequent sections. [9]

II. WHAT IS BLOCKCHAIN ?

Before getting ahead of ourselves, we must understand what does the term blockchain exactly mean? Standard chartered defines blockchain as the technology behind Bitcoin and other cryptocurrencies as a distributed ledger database for recording transactions, more commonly defined as blocks. [1]

A tight coupling of three principals' i.e. Cryptography keys, distributed network with a shared ledger and record keeping & security together define a blockchain technology.

A. Cryptography keys: To create a secure a digital identity reference. Combination of these public and private keys is a dexterous form of consent, forming unique digital signature

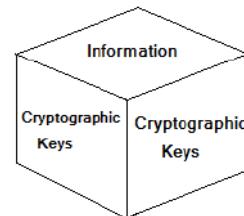


Fig 1. Diagrammatic representation of a block composed of cryptographic keys and other relevant information

B. P2P network: The understanding of a distributed network can be done by “if a tree falls in the forest” experiment. [6] If cameras are setup in the forest to record the falling of a tree, we can obtain an analogy of the entire process of tree falling. Similarly, the distributed network acts as set of cameras, validating the entire movement of blocks in a chain.

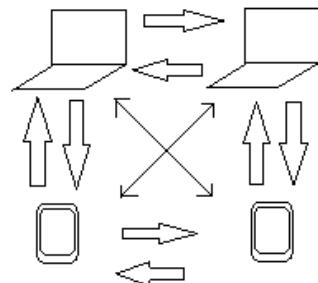


Fig 2. A P2P Distributed Network

C. Protocol: A protocol enacts as a platform for the communication of a cryptographic block consisting of timestamp, digital signature and relevant information

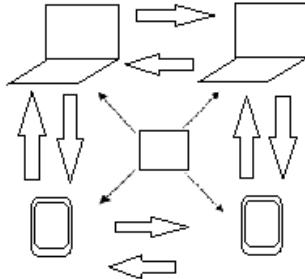


Fig 3. A distributed network of a block

The entire blockchain technology is a mere replica of these tightly coupled three principals. [2]

To understand blockchain better, we developed a basic blockchain wherein a block will store information like “NAME” and “Timestamp”, which is encrypted using SHA256 encryption and this information is passed onto another block where the previous hashed value is hashed with current details, thus forming a blockchain. We have made use of Python programming language for this purpose. Below is a snippet, which consists of the code for storing information in the distributed ledger i.e. a block and a snippet illustrating the output for the same

```
class Block:
    def __init__(self, index, timestamp, data, previous_hash):
        self.index = index
        self.timestamp = timestamp
        self.data = "Kamish"
        self.previous_hash = previous_hash
        self.hash = self.hash_block()

    def hash_block(self):
        sha = hasher.sha256()
        sha.update("%s-%s-%s" % (self.index, self.timestamp, self.data, self.previous_hash))
        return sha.hexdigest()

    def __str__(self):
        return "Block <%s>: %s" % (self.hash, self.data)
```

Fig 4. Programming code for a sample blockchain

```
Block #9 has been added to the blockchain!
Hash: 81a18fb08a42d65cf72bc09d5446632a73056638b5c990f69a9b5a3b4148a5ee

Block #10 has been added to the blockchain!
Hash: 873bbe429f0810d7a58a6602fc227e1058c218e03333d08d17f37f71958ccb40

Block #11 has been added to the blockchain!
Hash: 6f91a0fdde285d2e81ed68f22d815dfc116a57ac21031899bd93ce650b20948b

Block #12 has been added to the blockchain!
Hash: d4c888cd73df36dc105871cdb6cdad493c3afe9e09b60211a93d892973b9832

Block #13 has been added to the blockchain!
Hash: f9e4db4909565cf37a83991c379128348ccb64c5e761fc704299b8e71230e1f6
```

Fig 5. Output of the program

III. SURVEY

To understand how retail organizations view blockchain, we conducted an online survey of 108 retail professionals between November 2017 and early January 2018. 53% of the respondents represented a division of the retail firm and the rest represented the entire firm. Respondents employ following divisions as functional areas: 38% in IT, 41% in Supply Chain Management, 29% in Data management, 26% in Finance and Logistics and HR,

and 18% in Customer Service, 15% in marketing, compliance and procurement. [10]



Fig 6. Survey results indicating respondents employ a variety of divisions in their organizations

The main purpose for the survey was to understand the flexibility of Indian retail professionals to adapt blockchain technology as a platform for their business as usual activities.

Currently, all of the retail professionals in scope maintain a customer repository in form of a database and 85% use a delivery tracking system to enable smooth logistics tracking. For securing digital payments, only 15% have their own firewall for payment protection, rest all rely on firewall provided by banking partners. On having frequent purchasing behavioral patterns of the customers presented in form of reports, majority of the respondents rated the idea 8 on 10

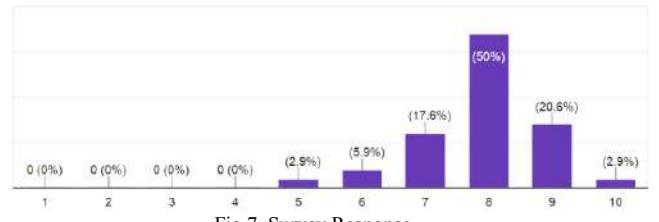


Fig 7. Survey Response

Each respondent was from a different retail industry segment. Majority of the retail professionals in scope were from Convenience stores or discount retail [10]

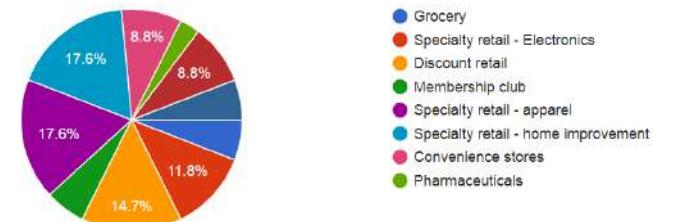


Fig 8. Survey Response

IV. BLOCKCHAIN'S RETAIL POTENTIAL

In the late 1990s e-commerce revolution had struck the retail market and the retailers who had then migrated from their traditional systems to e-commerce know the potential of disruptive revolutionizing technologies. Given its wide range of potential applications and required changes at infrastructure level, blockchain will change the way of retail market.

Customer serves the key role player in any retail domain and its trust is hard earned and easily lost. In an era where more and more transactions are happening on the internet, blockchain is a new headway. While blockchain is not going to revolutionize retail

tomorrow, there are enough changes in the coming years that through meticulous innovation and appropriate investment can change the way retail is done now. [7]

Further in the paper, it can be so viewed that adoption of blockchain technology by retailers can contribute a lot to assist the retailers in improving their existing business processes which in turn will lead to growth of their businesses. Following are a few key points under which this headway can be achieved.

4.1 Revolutionizing Supply Chain Management

During our survey, it so emerged that the fashion apparel industry retailers with their growing complex supply chain and knowledge unit management systems and shorter product life cycles, sales forecasting has become increasingly difficult. A blockchain solution, one where single source of truth systems coupled with smart contracts enable automatic execution of orders and payments, serves best to resolve the problems by accurate sales forecasting among many others.

For other wings of retail industry, shipment tracking and product/order recall are more of an integral part of current existing SCM systems. The SCM systems employed by many of the retailers does not facilitate immediate exact tracking for each stakeholder in the supply chain as they employ their own SCM for such a purpose. To revolutionize SCM in retail a blockchain can be used to store data about the shipment at every stage. It can include information like location, timestamp, check-in time, shipment handling personnel details, temperature, condition of package etc. Such a block of information can be accessed and added by each stakeholder in the supply chain. Thus providing accurate and concrete transparent information of the package. So in-case of product recall, such an accurate record will help retrieve and identify the point of failure and help Supply Chain Managing Retailers to run their businesses more efficiently. [5]

4.2 Reduction in counterfeit goods

For designers of luxury consumable goods, deterioration of brand value by counterfeit goods can cause decline in value of the luxury commodity and hence cause the manufacturers and retailers to incur huge losses. Blockchain acts a verifiable and the only viable solution for such a cause.

The proliferation of forgeries can be avoided by enabling a blockchain digital ledger record to store all details of each product including the ones that are counterfeited. Hence, in such cases when the end-customer receives the package of goods or commodities scan a code that is permanently etched into the product in order to access all of the information stored in the blockchain's digital ledger record. Retailers prone to counterfeiting of goods and services can use such blockchain to renew their lost trust with their customers.

4.3 Revolutionizing tracking systems of provenance

In an agrarian nation such as India, farmers and retailers involved in supplying organic farmed goods to customers, the existing tracking systems are somewhat non-existent. The current supply chain is so horrendous that neither the end customer nor the source i.e. farmer are aware of the source of a product nor the justification for the high price incurred on a product.

Existing supermarket chains such as Organic India, Godrej's Nature Basket, Hyper City, Food Bazaar, Spencers and many others are currently facing a great hurdle in building strong

confident relations with their consumers. As the consumers are growing suspicious that organic labels such as these are just a marketing tool and strategy to charge higher prices. Supermarket chains all across the nation can deploy a blockchain solution to raise this confidence by letting their customers to track the journey of a product from the farm to the store. This in turn can create an atmosphere of trust amongst all consumers. [8]

4.4 Ever-changing Customer Profile

Significant changes in the psychographic profile of the consumers are an outcome of the changes in a country's economic, social, legal and political and most importantly the technological changes. The customers purchasing habits are never static, it may seem to the consumers that tracking their purchasing behavior may be an easy job but as published in a few research studies by a think tank Project Guru it doesn't seem so. A major hurdle encountered by marketers today is managing and maintaining such complex and vast data.

Marketers serving many retailers can employ a blockchain solution. Advanced data warehousing systems can be developed for the retailers using blockchain technology as the records stored and accessed across the distributed ledger are immutable. This data then can be used to forecast the specific demands from a particular locality and suggest and store adequate stocks beforehand enhancing their just-in-time inventory facility. In today's ecommerce world, where same hour/day delivery are being employed by all retailers, data generated is enormous and to contain and being able to retrieve the same data along-with having developed algorithms which predict the next purchase of a customer, blockchain serves the best platform. The distributed ledger system across its computing network is already the next big thing. [4]

V. IMPLEMENTATION

The internet's commercial coming of age is instructive. The technological spectrum has advanced to new heights since ecommerce in 1990s. Early bird retailers who switched to ecommerce are the only ones with-standing the ripple effect of technology. Blockchain technology being the next big thing, retailers need to move quickly to gain experience so that they can understand the usage and various techniques to employ blockchain in multiple divisions of their organization.

During the survey, all of the respondents seemed ecstatic to take the blockchain leap but only a few of them were employing blockchain technology in the next year or so.

Out of all the respondents, a whopping majority were in the current phase of employing blockchain technology. [9]

But only 55% of the respondents were keen to employ blockchain technology throughout the organization or in a particular division. [9]

To implement the blockchain technology in organizations, retailers need to develop strategies and make certain considerations such as

1. Set clear goals: Retailers need to assess whether blockchain technology is well suited for business issues

2. Flexibility: In its nascent stage, retailers need to stay flexible to adapt the changes and developments in blockchain technology.
3. Understanding real world potential: Retailers need to adapt and understand blockchain technology's real word potential and its large scale implementations.

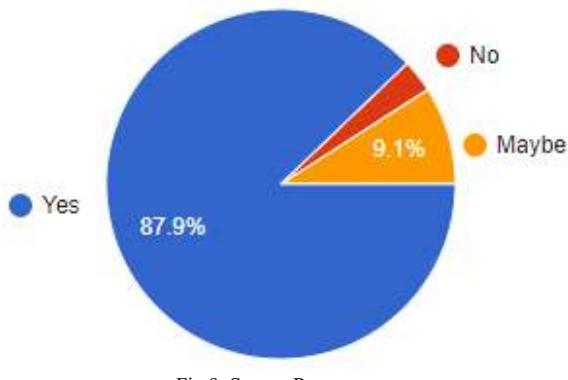


Fig 9. Survey Response

● No, but we are seriously investigating the possibility of implementing block...

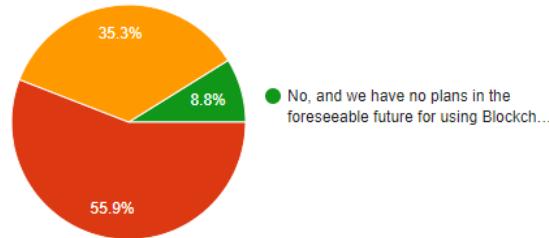


Fig 10. Survey Response

4. Setting up cross-functional teams: The initiative should have all stake holders throughout the organisation set up cross-functional teams to address business specific problems and opportunities.
5. Selecting both permissioned and permission-less platforms: In order to have a custom tailored blockchain as per the business requirements, retailers need to select platforms as per each use case. [4]

VI. RESULTS AND DISCUSSIONS

The respondents for the survey belonged to a huge variety of retail domains. 18% of the respondents were apparel retailers and electronic retailers, along with them there were retailers from discount retail, membership clubs and pharmaceuticals. 88% of the respondents were ready to deploy blockchain in their businesses and only a few were skeptical of this disruptive technology. It is important that retailers across all domains employing blockchain need to consider a few prerequisites which include

1. Understanding the extent of employing blockchain

2. Setting up cross-functional teams backed by all stakeholders
3. Maintaining only relevant customer profiled data
4. Transparency at all levels.

Furthermore, blockchain in its nascence is a disruptive technology and with the proper funding for blockchain technology projects, one can reinvent the retail domain.

TABLE I. FIGURES REFERENCE TABLE

Sr No	Figure Number	Figure Description
1.	Fig.1	Representation of a block composed of cryptographic keys and other information
2.	Fig 2	P2P distributed network
3.	Fig 3	Distributed network of a block
4.	Fig 4	Programming code for sample blockchain
5.	Fig 5	Output of the program
6.	Fig 6	Survey Response
7.	Fig 7	Survey Response
8.	Fig 8	Survey Response
9.	Fig 9	Survey Response
10.	Fig 10	Survey Response

VII. CONCLUSION

Features of blockchain technology and its benefits have been limited to financial applications such as bitcoin and other cryptocurrencies. Still in its nascent stages, this technology is still evolving with lot of scope for different businesses. Just like in 1990s ecommerce which disrupted the traditional way of doing businesses for retailers, adapting to blockchain technology will change the way of doing business in the world. Improved transparency, better loyalty tracking system, better delivery tracking systems coupled with increased and efficient supply chain management are just a few of the many benefits this key technology aims to serve. Retailers in order to stand out from all the rest of their competitors need to adapt this disruptive technology, else their technique of doing business will be a thing of the past.

REFERENCES

- [1] https://www.sc.com/fightingfinancialcrime/av/SCB_Fighting_Financial_Crime_Deep_dive_Blockchain_August_2017.pdf
- [2] Blockchain and its Scope in Retail (IRJET) - Arijit Chakrabarti1, Ashesh Kumar Chaudhury
- [3] Bitcoin: A Peer-to-Peer Electronic Cash System - Satoshi Nakamoto
- [4] Retail: Opening the Doors to Blockchain – Cognizant
- [5] <https://cointelegraph.com/news/camelgram-to-develop-blockchain-based-point-of-sale-network>
- [6] https://en.wikipedia.org/wiki/Single_source_of_truth
- [7] <https://www.accenture.com/in-en/insight-highlights-cgs-blockchain-cpg-and-retail-industry>
- [8] <https://www.coindesk.com/chinese-retail-giant-jd-com-joins-blockchain-in-transport-alliance>
- [9] <https://docs.google.com/forms/d/1G8-Wlm1pznAPv0mXc5kHtAKIiVGBFojkz0YMUswY>

Design and Development of Pulse Oximetry for Continuous Monitoring of Pregnant Ladies using Arduino

Naziya Pathan¹, Mukti E. Jadhav²

Department of management science and computer studies Maulana Azad College of Arts, science &commerce
Aurangabad, India¹,

Department of computer science and IT,MIT College Aurangabad, India²
nazi.pathan@yahoo.com¹, muktijadhav@gmail.com²

Abstract--The continuous monitoring and caring patient's condition is an important area of research in the field of Biomedical. These required multi-parameters system for continuous sensing processing and displaying the condition of the patients. Pregnancy brings many changes in a woman's body. Including pulse rate, which increases during pregnancy. A pulse rate is different for every woman, and is based on level of fitness before pregnancy and age. Doctor will monitor pulse rate throughout the pregnancy, and recommendations to care accordingly. The objective of this paper is to design a non-invasive Arduino based optical pulse oximetry .The pulse oximetry consists of a fingertip sensor and an Arduino microcontroller. In it Photo transistor sensor is used for the pulse wave's detection and the Arduino microcontroller is used to analyze the pulse wave to calculate the oxygen saturation (sao2).The pulse oximetry uses the Arduino as a microcontroller.

Keywords: Pulse rate, Pulse oximetry, Arduino microcontroller, Phototransistor.

I. INTRODUCTION

The pulse oximeter is one of the medical device used to measure spo2 and pulse rate of a person. It became a standard procedure for the measurement of blood oxygen saturation in the hospital operating room and recovery room [1].It is also one of the important parameters for old people, pregnant women in several critical situation. It is non-invasive and allows immediate and real time monitoring, its use has expanded to include other purpose such as screening, diagnosis, patients follow up and self-monitoring. Good nutrition and regular determination of finger pulse is the guarantee of improved health conditions of a pregnant women and future children. Physicians recommend pregnant women pulse oximeter for accurately measure the oxygen level. Pregnant women have been using pulseoximeter technology throughout the decade. Digital technology has produced new medical devices such as networked glucose reader, digital thermometers and stethoscopes as well as innovative application such as motion sensors and video conferencing tools. A pulse oximeter is one of them it is intended for the non -invasive measurement of arterial blood oxygen saturation and pulse rate [2].It provides an important function in the intensive care unit, as an early warning system for patient emergencies [3].

II. BIOLOGICAL PRINCIPLES OF OXYGEN MOVEMENT

The oxygen is important for functioning of each cell in the human body. Without oxygen cells will die. Oxygen saturation is an indication of oxygen transport in the body. Several methods have been developed to analyze oxygen delivery.

Blood red cells contains a protein called hemoglobin. Hemoglobin is a stable only when bound to 1 to 3 molecules of oxygen. Red cells with oxygenated hemoglobin circulate in the blood through the whole body .When blood gets in contact with a cell the red cells hemoglobin releases oxygen and becomes Deoxyhemoglobin (Hb) (deoxygenated hemoglobin)[4].The volume of the arteries become larger before the blood enter the capillaries. This changes makes possible for the oximetry system to differentiate the arterial blood from all other absorbing substances [5] [6].

III. METHODOLOGY

During the last few years there has been a significant increase in the number of various pulse oximetry on the market ranging from pulse monitors to portable wireless digital oximeters[7][8]The study of working principle and the components of pulse oximeter module is derived from[9][10].It is Characterized into three different stages.

- 1) Sensing stage: - Design and development of pulse oximetry probe sensor.
- 2) Processing stage: - Development of pulse oximeter module.
- 3) Displaying stage: Display of results.

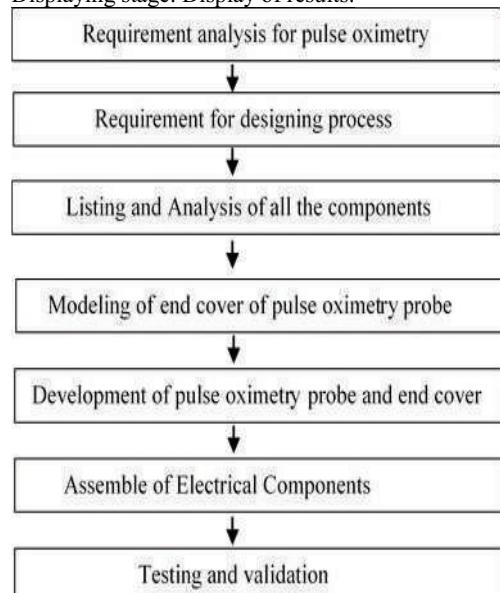


Fig1: Block diagram of pulse oximetry

IV. THE WORKING PROCESS

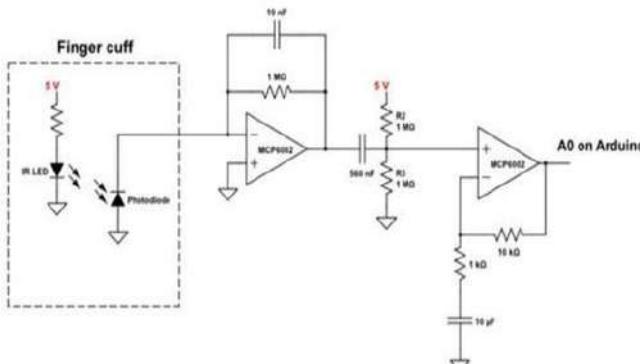


Fig 2: Electrical circuit diagram for the detection of pulse rate

As blood is pumped through the body, the volume of blood in extremities such as fingers increases and decreases with the pumping of the heart. The change in blood volume in the finger tips can be detected by shining a light through the finger and detecting the amount of light that passes through the finger using a photodiode. A photodiode is a semiconductor that produces a current proportional to the amount of light that hits it. So when the blood volume in the fingers increases, less light is getting through the finger and hitting the photodiode. The current produced by the photodiode is converted to a voltage by an amplifier and read by the Arduino. The finger cuff has a photodiode and IR LED integrated into it. Both components have wires that connect them to the Arduino.

The next portion of circuit is known as a transimpedance amplifier or current-to-voltage converter. A transimpedance amplifier is an amplifier configuration that allows to convert a current to a voltage. In this case, the current produced by the photodiode flows through the resistor in the op amp's feedback loop. A voltage at the output of the op amp with accordance to Ohm's Law, $V = I \cdot R$. The current produced by a photodiode can be small (micro-amps), this is why we are using such a large resistor (1 M Ohm) as the gain-setting resistor. A high gain-resistor such as this is typical for transimpedance amplifiers. The capacitor in the feedback loop helps reduce high-frequency noise.

Next stage is a high-pass filter. A high-pass filter allow us to remove low-frequency signals. In this case, we are removing signals that are slower than a normal pulse signal. Specifically, we are removing the DC bias from our signal. When the IR LED shines light through the finger, most of the light is absorbed by the tissue. Our circuit will in turn produce a voltage that corresponds to the absorbance of light by the tissue, not the pulsatile of the artery (which is what we really want). For this reason, we use a high pass filter to get only the absorbance due to the pulsatile of the blood in the arteries (our pulse).

The final portion of our circuit is an AC gain stage using a non-inverting op amp. A non-inverting op amp amplifies and input voltage in accordance to the equation $1 + R_2/R_1$ where R_2 is the resistor in the feedback loop and R_1 is the resistor that is connected to ground. Notice the capacitor in the non-inverting op amp. A capacitor blocks DC signals and only allows AC signals to pass. This means that only the AC signal which corresponds to our pulse gets amplified, not the voltage that we applied in the previous stage.

IV. Factors for an Irregular Pulse Rate and Oxygen Level during Pregnancy

- 1) Increased Load during Pregnancy:-During pregnancy hearts accommodate to the changes in body. Because heart is working harder. Due to that “skipped beat” or a “fluttering” in neck or chest occur. Additional symptoms like fatigue, lightheadedness, dizziness, chest pain, shortness of breath, and a rapid or pounding heart point to an irregularity in the rhythm of heart that's impacting its functioning[12].
- 2) Lifestyle or Health Factors:-Some factors are related to lifestyle, health, and state of mind that may also cause irregular pulse Rate. As pregnancy hormones tussle with system, it feel the impact on the body and mind quite intensely. Deal these through some simple lifestyle changes or quick remedies, without medical intervention

These factors include:

- Anxiety or stress: Ease pregnancy pangs with relaxation or breathing exercises.
 - Dehydration: Take in plenty of fluids or have an oral rehydration solution.
 - Low blood sugar: Deal with it by having juice, biscuits, or even a tablespoon of sugar or honey [13].
 - Too much chocolate or caffeine. Cut down on these or avoid them as much as possible.
- 3) Medical Conditions:-Sometimes, an irregular heart rate can be indicative of a problem like a heart disorder or thyroid imbalance.
 - Heart Disorder: Additional symptoms like chest pain, dizziness, shortness of breath, or fainting along with palpitations could point to heart disease[14].
 - Hyperthyroidism: Symptoms like fatigue, hair loss, difficulty concentrating, hand tremors, frequent bowel movements, increased sweating and appetite, problems with your weight or sleep, and heat intolerance may be indicative of hyperthyroidism.

V. FACTORS FOR AN IRREGULAR PULSE RATE AND OXYGEN LEVEL DURING PREGNANCY.

- 1) Adopt Healthy Habits:-Eat regular meals, and get plenty of fluids as well as sufficient sleep. These habits can good for general health and wellbeing during pregnancy.
- 2) Cut out the Stress:-Anxiety and stress can make your heart rate irregular. Practicing relaxing techniques like meditation, tai chi, or yoga can help you deal with stress.
- 3) Take A Breath:-Try deep breathing when you get palpitations. Breathe in deeply and slowly through your nose so that you feel your abdomen move.
- 4) Splash Water:-Splashing some cold water on face can be helpful for palpitations.

VI. HARDWARE AND SOFTWARE REQUIREMENT

The pulse oximeter uses the ARDUINO as microcontroller, which has ultra-low power capability so the system power consumption is low [11]. One Arduino, One general Purpose Op Amp (MCP6002 or similar). Three Capacitors (1 x 10 nF 1 x 470 nF and 100 nF). One Photodiode. One IRLED. (Figure no 3). Matlab for automation of experiments, data visualization Arduino UNO

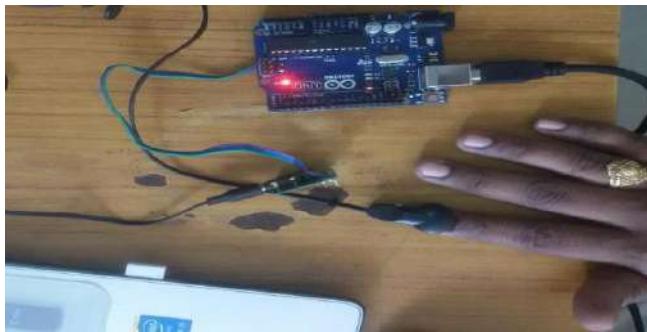


Figure 3 Device interface with Arduino

VII. RESULTS AND DISCUSSION

A Program is developed in Matlab which allow viewing and stored the signals for further analysis with algorithms in order to minimize the noise and to make the peak detection to determine the pulse rate and oxygen level in the blood and store the patient detail in the database for further used like patient name, age, weight, pulse rate and oxygen level and due date of patient(Figure 4)

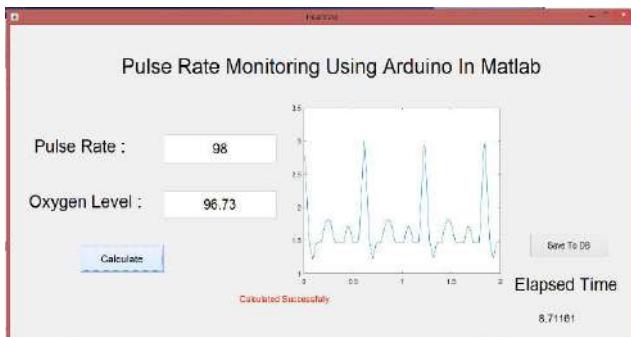


Fig 4: Representing data after Matlab processing

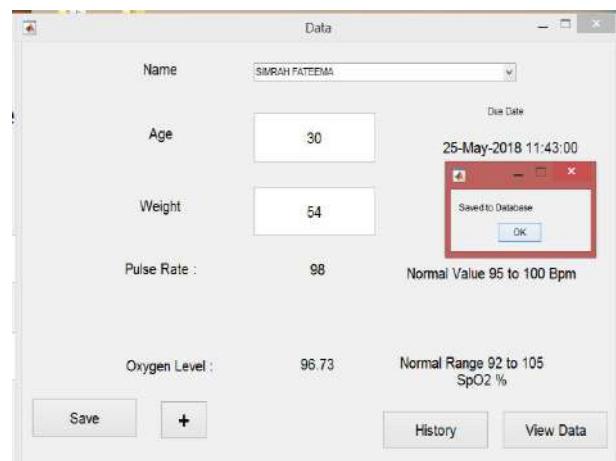


Fig 5: Storing of the data after Matlab processing

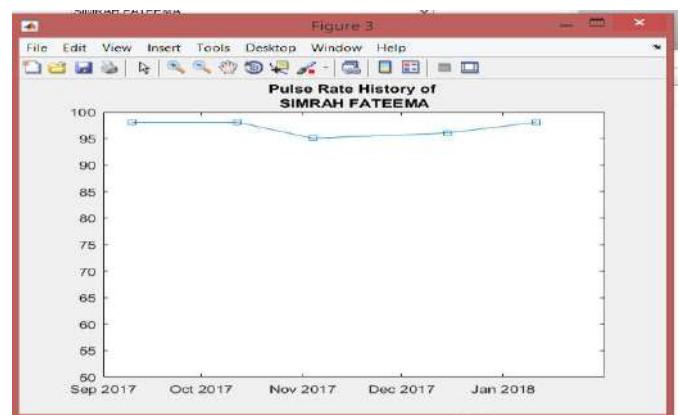


Fig 6: Graphical Representation for History of Pulse Rate

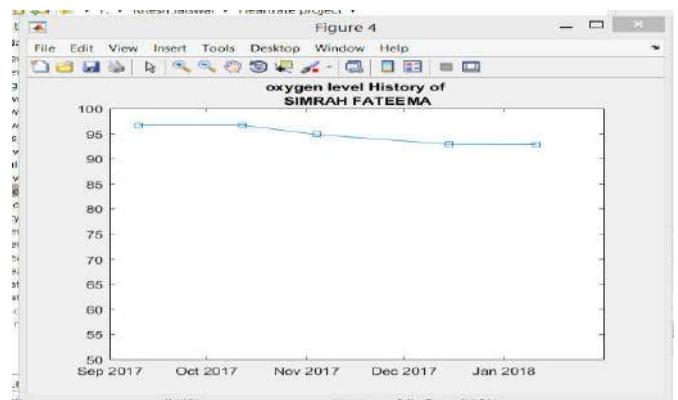


Fig 7: Graphical Representation for History of Oxygen Level

The above Figure no 6and 7 describes the Pulse Rate and Oxygen Level of a Patient. It helps to view the history of the patient throughout the nine month for the patient and doctors.

VIII. CONCLUSION

The main aim of this paper to developed a portable pulse oximeter unit that receives data and process in order to return the values of pulse rate and oxygen level in the blood. The need for the development of pulse oximeter was emerged to enhance the technical competency of India and to provide the cost effective products for the hospital requirements both in urban and rural areas. The data is acquired by Arduino module, which communicate with a platform that would have implemented in order to perform the signal processing.

REFERENCES

- [1] Bronzino S and Joseph D.IEEE Biomedical engineering hand book,CRC press, IEEE Press 1995, 0-8493-8346-3.J.ClerkMaxwell,ATreatiseonElectricityandMagnetism,3rded „vol.2.Oxford:Clarendon,1892,pp.68–73.
- [2] John w. Severinghaus, MD, “Takuo Aoyagi: Discovery of pulse oximetry”. Anesthesia and Analgesia, vol.105 no.65 supp/s1-s4 2007K.Elissa, “Titleofpaperifknown,” unpublished.
- [3] Brinkman,R and W.G.Zijlstra, “Determination and continuous Registration of the percentage oxygen saturation in clinical conditions”, Archivumchirurgicum Neerlandicum,Vol.1,pp,177-183,sep 2003.
- [4] Santiago L. (2012) pulse oximeter fundamentals and design, free scale semiconductor Document Number: AN4327.
- [5] Hoff.D.Zhang.R.Stalter,T and carlson,M (2003).Pulse oximetry .Undergraduate Thesis Electrical and Computer Engineering, North Carolina State University, USA..
- [6] Oximetry.org: Principles of pulse oximetry Technology retrieved on: 16th of June 2013. URL :(<http://www.oximeter.org/pulseox/principles.html>)
- [7] Aoyagi T, Kishi M, Yamaguchi K, Watanabe S. (1974), annual meeting of the Japanese society for the medical electronics and biological engineering.p.99-91.
- [8] Castillo,J.M,Olivares,J,Palomares,J.M. “Design of a wireless pulse oximeter using a Mesh ZigBee sensor Network”, Department of Computer Architecture, University of Cordoba, Spain, Bio Devices,2011,pp410-404.
- [9] Ashoka Reddy K (2008), “Novel methods for performance enhancement of pulse oximeters, Department of Electrical Engineering”. Indian Institute of Technology, Madras. . <http://www.ni.com> (website).
- [10] P.C.Branche, W.S.Johnston, C.J. Pujary, and Y.Mendelson, “Measurement Reproducibility and sensor placement considerations in designing a wearable pulse oximeter”.30th Annual Northeast Bioengineering conference, 2004.
- [11] Symptoms, Diagnosis & Monitoring of Arrhythmia. American Heart Association.
- [12] Low blood sugar. National Institutes of Health
- [13] Skipping a beat — the surprise of heart palpitations. Harvard Health Publications
- [14]

Cloud Based Vehicle Monitoring System

Amit Dubey¹, Jainil Jain², Vamsi Akhumukhi³, Rahul Gupta⁴, Vidyadhari Singh⁵

Thakur College of engineering and technology, Mumbai, India

Abstract - The expectation of Increasing Number of Vehicles is rising globally due to economy growth in countries like India and hence the ownership is becoming much more affordable. However, the adoption of vehicle tracking is still very much lagging. Currently there are very few systems wherein the user can track the vehicle location of its peers and family members and monitor it. We are planning to build an application wherein the user can get the GPS co-ordinates of the vehicle he is monitoring. This application uses the GPS function, available in most of Smart phones today, to pin point current location accurately. The Co-ordinates of location will be sent to the cloud, and server will send current location to the user and will provide additional services like notifications. The System just involves the use of any GPS enabled smart mobile Phones by the users. With an installed App on any smartphone we can enable user to track the location of vehicle and get other details. Such a system can be used for many applications including Educational Institutes, public transportation systems, fleet management etc.

Keywords- GPS, AWS, CLOUD, SERVERLESS

I. INTRODUCTION

Vehicle monitoring system's main aim is to provide security to all vehicles by keeping a track on it using minimal hardware supports and being highly scalable. Tracking of any vehicle involves two major aspects, getting the exact location of the vehicle and transmitting it to the user without a time lag [1]. Tracking application involves sending and receiving requests to and from the servers which makes them inefficient in large scale uses due to their manual server management. Vehicle tracking systems that are available in the market are somewhat application specific, region specific and are costly [2-4]. The proposed system of monitoring vehicle using cloud-based architecture serves users with minimum cost and fast computation speed. With advancement in technology where existing systems are shifting to new platforms, developing a system for large group of people with minimum cost and utilization of best possible platform is one of the major concern that needs to be handled. As adoption of cloud has been swift and global in recent years due to its fast implementation, no upfront costs and infinite scalability in least price, the system is planned to be deployed using AWS serverless cloud computing. The project tries to propose an accurate map-matching algorithm for location tracking of a mobile user that can be potentially used by privacy adversaries to accurately reconstruct a user's actual trace [5].

The main idea behind this is to deploy a GPS tracking application on a Serverless architecture using Amazon Web Services (AWS) which will enable us to cut down costs to a significant level and improve scalability.

The plan is to deploy this system for school buses and other business plans to enable product's easy acceptance. The goal of the project is:

1. To develop a cost effective and highly efficient vehicle monitoring system and deploy it for School bus tracking.
2. To replace existing system's manually configured implementation with fully managed and infinite scalable AWS services.

3. To provide a top notch pervasive platform that is to extend the reach from a beginner to a big business.

The project can be extended as an effective security measure for Public Transportation facilities, Asset Tracking and Ambulance services [6].

A. Scope of the Project:

1. Transportation Services: This can be used to monitor vehicles which carry goods and public transport facilities like buses.
2. Hospital: This can be integrated with hospitals by providing ambulance tracking services
3. Real Time monitoring: This can provide premium features to live track cab services for employees in a company by integrating it with IoT devices.
4. Asset tracking: Companies needing to track valuable assets for insurance or other monitoring purposes can plot the real-time asset location on a map and closely monitor movement and operating status.
5. Field service management: Vehicle tracking allows companies to quickly locate a field engineer and dispatch the closest one to meet a new customer request or provide site arrival information.

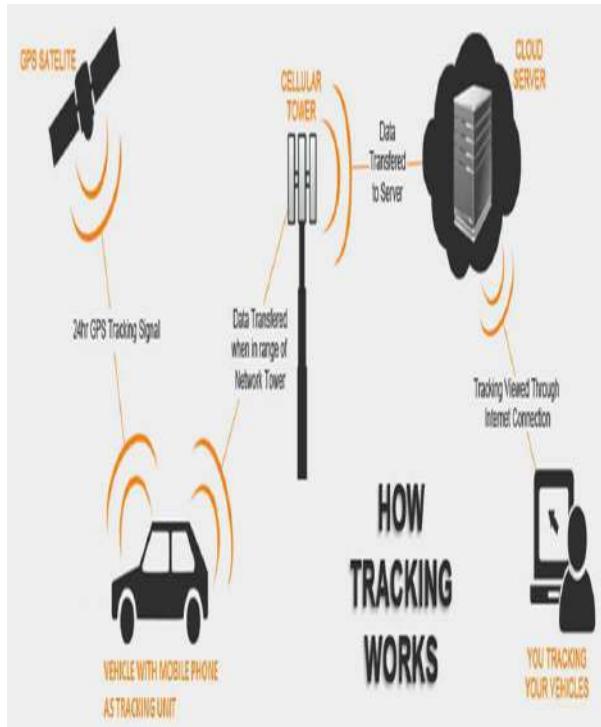


Figure 1: Basic Tracking Workflow

II.LITERATURE SURVEY

A simple bus tracking system proposed by Lau [7] in UCSI University, Kuala Lumpur, Malaysia provides students with real time location information and status of a bus after specific time intervals using LED panel and a Smartphone Application within a fixed route. Real-time bus tracking systems are beneficial to college students who attend colleges with large campuses. Spending less time waiting for a bus improves the comfortable and effective time management of the students.

Here Comes the Bus [8] is easy-to-use school bus tracking software that uses RFID Proximity readers on the bus and barcode on student ID card which gives parents and students the ability to track the location of their school buses and accurately gauge when their bus will arrive.

Track school bus[9] provides real-time updates on school bus locations.Tracker sends alerts in the form of SMS, Android push notifications or iOS notifications and parents get to know about bus delays, over speeding, unscheduled bus stops or other emergencies.Helps you save energy and time by automatic routing and planning and scheduling bus stops.

CHAKRAVIEW [10] offers an innovative communication tool between the schools / bus operators and parents. They use slightly expensive technology to track the school buses on mobile phones and send various types of alerts about the buses. Instead of using GPS devices, they use the GPS and GPRS capabilities of smartphones to offer a much more versatile solution.

Vehicle tracking system using Social Network Service [11] Social network services such as Twitter and Facebook have also been used to build a base for Vehicle tracking systems and attracted interest in a number of users. Each in-vehicle device has an account of the twitter social network and can identify the vehicle location in social network on a regular basis. A web interface is used to display a vehicle location placed on Google maps, and a status of a vehicle like door open/close, and ignitions on/off. The proposed system can be accessed from a Smartphone. So, the system would become more efficient to users of social network as they allow quick monitoring of the location and status of the vehicle.

III. EXISTING SYSTEMS VS PROPOSED SYSTEM

The overall goals of this section were firstly to establish the significance of the general field of study, and identify a place where a new contribution could be made. The idea of the application emerged by identifying following facts from other similar applications that is described below.

A basic Vehicle Tracking and monitoring System can be developed by three configurations namely, Hardware configured system, Software configured system or combination of both [12].

The Barriers found when using a hardware configured systems are:

1. Requires a manual configuration of each hardware on separate vehicle.
2. Expensive in terms of cost as well as maintenance.
3. The System is not resilient.

While Software configured system too will require a tracking component to get the current location of the vehicle.

Moreover, managing Server instances for large group of people is difficult.

Auto-scaling needs to be configured and Availability is a major concern.

Our Proposed System of Serverless architecture eliminates all these barriers and thus provides a cost effective and a highly automated mechanism for tracking the vehicles.

Serverless architecture is a cloud computing code execution model in which cloud provides fully managed starting and stopping virtual machines necessary to serve requests and requests are billed by an abstract measure of resources required to satisfy the request [13]. So, the total cost of the system gradually decreases as the bill is only initiated when a request is made rather than the amount of time the computing continues in the background. The overhead of managing servers is removed as everything is managed by AWS services. Thus, the system is adoptable due to its secured storage of data, Cost-effectiveness, Efficiency, limited use of hardware component and resilient Architecture.

IV. PROPOSED METHODOLOGY

A. Project Description:

The basic idea is to develop a cross-platform GPS tracking application with the help of AWS Lambda and other AWS services. The application uses GPS function which is available in most of the smart phones. The mobile phone is used as a GPS tracking device with the application installed in it. The plan is to deploy it for school bus tracking wherein the parents and the school both can monitor the activities of the bus and track their children thereby eliminating security risks. The application uses Google Maps API and location services for determining the location of the bus.

B. System Analysis:

The Figure shown below depicts the structural overview in which the overall flow can be seen at a broader level.

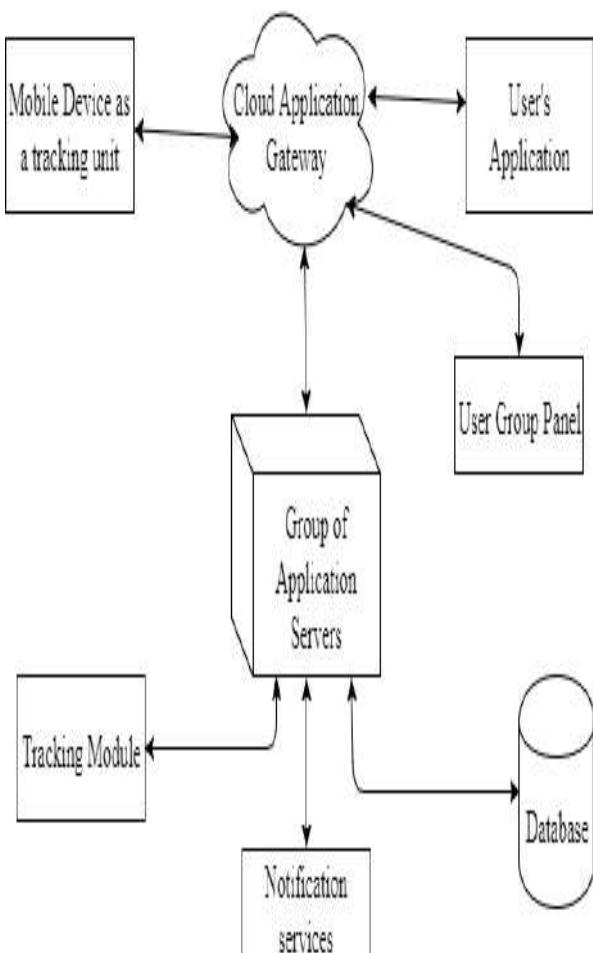


Figure II: Structural overview

C. Proposed Work:

The main objective of the system is to provide a GPS enabled vehicle monitoring system with a resilient architecture.

Mobile devices would be used as tracking unit. Application servers would be deployed on cloud. The basic flow of the project would be as such:

1. As the plan is to deploy for school bus tracking, the mobile device of the driver of bus will send latitude and longitudinal co-ordinates and on the other side, users (School and Parents) can monitor the live status of the buses.
2. Each bus will be assigned with a unique bus-id.
3. The Driver will login with his credentials and the bus-id he is associated with.
4. The GPS enabled tracking application which is installed on the phone will check for the Google location services and automatically detect the current location through Google Maps API and location services.
5. At fixed intervals, the device detects the location and sends it to the application server and is stored in the database.
6. On boarding and departure to and from the bus, a notification would be sent to parents.
7. At the backend, the distance between two points is evaluated and interpolated in n different sub-points for smooth animation and request to appropriate Google location API is made.

smooth animation and request to appropriate Google location API is made.

8. The Parents need to login and the registered bus-id with parent would be tracked and monitored.
9. The system would use AWS lambda functions to provide fully managed and infinite scalable product.
10. AWS DynamoDB is used for databases and Serverless Architecture is used to bind everything in a workflow.

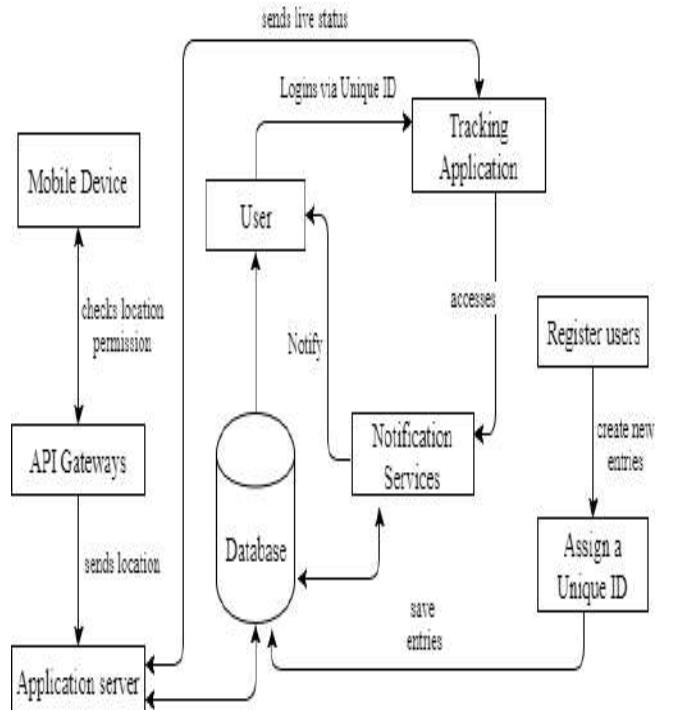


Figure III: Basic Implementation flow

The figure above depicts the overall flow of the system illustrating how mobile device used as tracking unit is connected with the dedicated users and responds via notification services.

D. System Architecture:

The overall system architecture using AWS services [14] is depicted below:

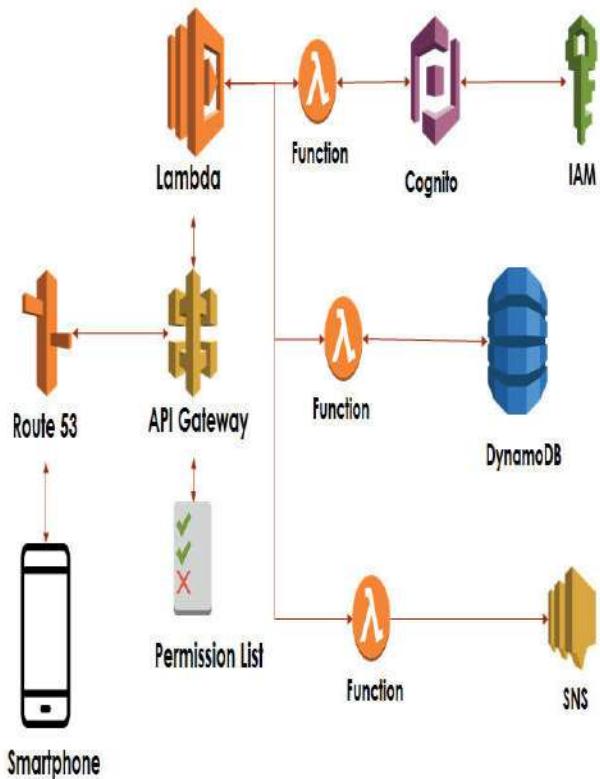


Figure IV: Overall System Architecture

Smartphone: Any GPS enabled mobile device used as a tracking unit.

Route 53: Amazon Route 53 is a highly available and scalable cloud Domain Name System (DNS) web service. Amazon Route 53 effectively connects user requests to infrastructure running in AWS – such as Amazon EC2 instances, Elastic Load Balancing load balancers, or Amazon S3 buckets [13]. It is also used for

1. Registering domain names.
2. Routing internet traffic to the resources for our domain.
3. Checking the health of the resources.

API Gateway: Amazon API Gateway handles all the tasks involved in accepting and processing up to hundreds of thousands of concurrent API calls. In our system we will be working with Google maps API.

AWS Lambda: AWS Lambda is a compute service that lets you run code without provisioning or managing servers. AWS Lambda executes your code only when needed and scales automatically, from a few requests per day to thousands per second [13]. You pay only for the compute time you consume - there is no charge when your code is not running.

AWS Cognito: Amazon Cognito is a service with which you can create unique identities for your users and authenticate these identities with identity providers and save user data on cloud.

IAM: Identity and access management is a web service that helps you securely control access to AWS resources. It is free to use and manages Identity federation and provides Multi-factor authentication.

Dynamo DB: It is NoSQL database service provided by AWS that can store and retrieve any amount of data and serve any level request traffic.

SNS: Simple Notification service is a web service that enables end-users, application and devices to send and receive notification from cloud.

The overall system flow would be:

1. The application installed on the smartphone will send the Geo co-ordinates via GPS
2. Route 53 that acts as DNS will route the information to the desired API Gateways and triggers the AWS Lambda function
3. Permission list contains the policies that we want to attach to the IAM and corresponding actions for which we can grant permissions to use particular AWS resources and perform the action.
4. AWS lambda when triggered for the first time uses Cognito services for creating new user identity which is further authenticated with the help of IAM services.
5. Once Authenticated, another Lambda function is Triggered and the information is stored in Dynamo Database for future use.
6. The Triggering of AWS Lambda functions makes the system serverless and finally AWS SNS sends the

appropriate monitoring information in the form of Notification when required.

7. Thus in a nutshell, the overall system is highly secured and resilient.

V.RESULTS AND DISCUSSIONS

The results are somewhat tentative and based upon the comparative study with other systems and response of existing AWS projects. Amazon web services Free Tier is capable of managing this kind of applications and therefore any advancement in this system will not be expensive and management will be at ease. According to Amazon Web services Documentation and pricing models, use of its services is based upon the number of requests made and by our estimation regarding the use of services, we found out that the monthly expenses of the proposed system were affordable. Google Maps API provides 2,500 free requests per day, calculated as the sum of client-side and server-side queries and 50 requests per second, calculated as the sum of client-side and server-side queries. Beyond these limited quotas pricing will be by ‘Pay-as-you-go’ billing model [15]. The Tracking would vary from smartphone to smartphone depending upon the accuracy of the GPS and quality of Device used as tracking unit. The implementation of the proposed system till date is depicted below:

Figure 5 shown below is a basic Android Application that sends GPS co-ordinates of the mobile at a regular fixed interval of time via AJAX.

Figure V: User Interface of Android Application

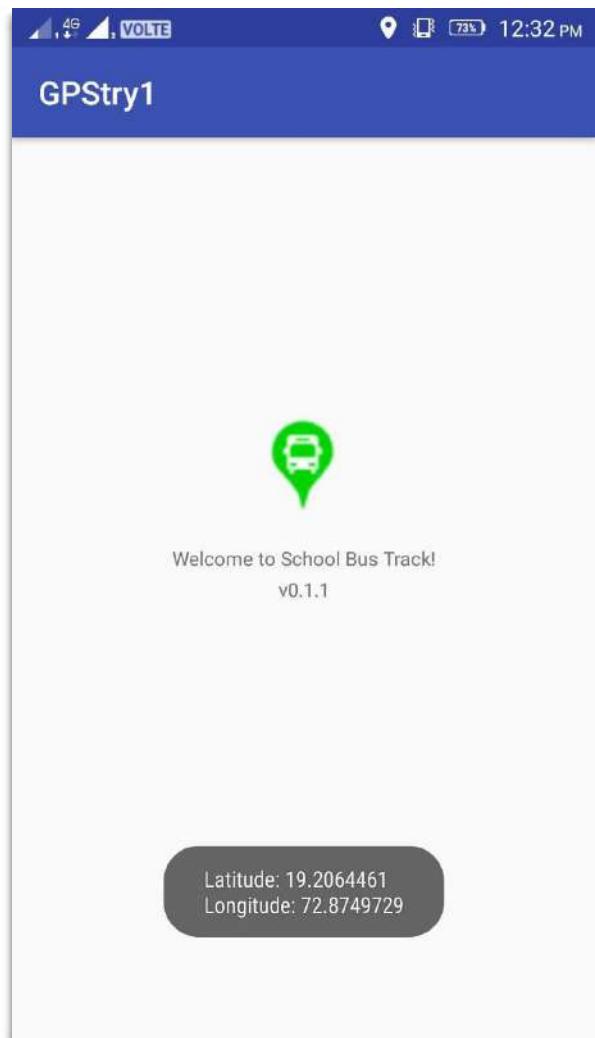


Figure 6: Web View to live track vehicle

The above figure depicts the current location of the user in form of Marker on the website view of the map.

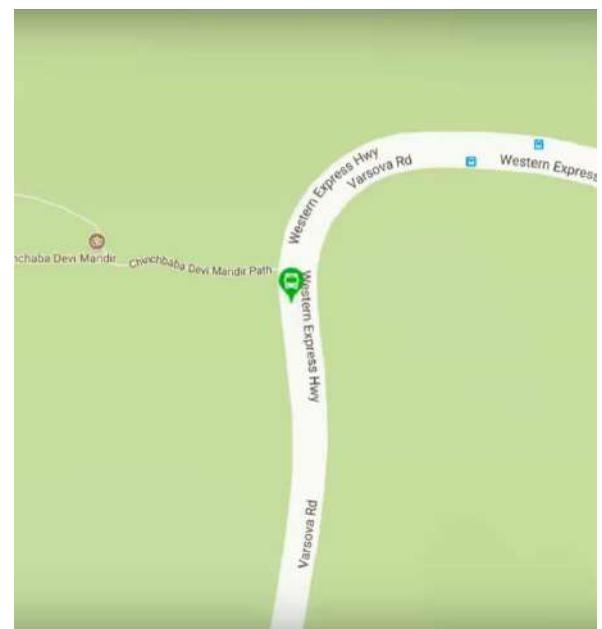


Figure VII: Snapshot of live tracking

The above image is a screenshot taken from a live tracking of a vehicle. The application was tested on a route of about 20kms to check the smoothness of the animation on the website.

VI. CONCLUSION

In this paper we have presented the development of a cloud-based vehicle monitoring system which is able to obtain a vehicle's GPS coordinate and transmit it using the appropriate API services to the user's phone. Development of country towards digitization and increased internet connectivity in recent years has made people adaptive to the use mobile phones, hence the acceptability of the system by people in the near future will be definite. The development of vehicle monitoring system demonstrates the feasibility of near real-time tracking of vehicles, which can be used for security of personal vehicle, public transportation systems, fleet management and many other applications. The system can provide improved customizability, global operability and cost when compared to existing solutions. Once the system is complete, the vehicle tracking system has the potential to be commercialized as a standalone product since its utility is quite popular.

REFERENCES

- [1] Dhanshri K. Kulkarni, Rakhi R. Patel, Bhavna S. Pal and Prof. Lavina D. Panjwani, "Android Application for College Bus Tracking", International Journal of Innovative and Emerging Research in Engineering Volume 3, Issue 5, 2016.
- [2] Youjing Cui; Shuzhi Sam Ge, "Autonomous vehicle positioning with GPS in urban canyon environments," IEEE Transactions on Robotics and Automation, vol.19, no.1, pp.15-25, Feb 2003.
- [3] Pradip V Mistary, R H Chile, "Real Time Vehicle Tracking System Based on ARM7 GPS and GSM Technology", IEEE INDICON 2015 1570187057
- [4] Official Website ofLoJack(2013). <https://www.lojack.com/>
- [5] G. Jemilda, R. Bala Krishnan, B. Johnson, G. LingaSangeeth, "Mobile Application for College Bus Tracking", International Journal of Computer Science and Mobile Computing, Vol. 4, Issue. 3, March 2015, pg.500 – 507
- [6] Pham Hoang Oat, MichealDrieberg and Nguyen Chi Cuong, "Development of Vehicle Tracking System using GPS and GSM Modem", IEEE Conference on Open Systems (ICOS), December 2 - 4, 2013, Sarawak, Malaysia.
- [7] Eddie Chi-Wah Lau, "Simple Bus Tracking System", Journal of Advanced Computer Science and Technology Research, Vol3, No.1, 2013
- [8] The Official Website of "Here comes the bus", a registered trademark of synovia solutionsTM, LLC. <https://herecomesthebus.com/>
- [9] The Official Website of "Trackschoolbus", a venture by Edsys.in for powerful transportation software application. <https://www.trackschoolbus.com/>
- [10]The Official Website of "CHAKRAVIEW" <http://chakraview.co.in/>
- [11]Ahmed ElShafee, Mahmoud EIMenshawi, and Mena Saeed, "Integrating Social Network Services with Vehicle Tracking Technologies", International Journal of Advanced Computer Science and Applications, Vol. 4, No. 6, 2013
- [12] Dr. (Mrs.) SayleeGcharge, Manal Chhaya, Gaurav Chheda, Jitesh Deshpande, NiketGajra, "Real Time Bus Monitoring System Using GPS", IRACST – Engineering Science and Technology: An International Journal (ESTIJ), ISSN: 2250-3498, Vol.2, No. 3, June 2012
- [13] The Official Documentation of Amazon Web services (AWS) <https://aws.amazon.com/documentation/>
- [14]Jinesh Varia, "Architecting for the cloud: Best practices", Amazon web services white paper Jan 2011.
- [15] The Official Documentation of Google Maps API's and its Pricing models <https://developers.google.com/maps/pricing-and-plans/>

Smart Ambulance System with Traffic Overriding

Chaitanya Zarekar¹, Sandeep Pardeshi², MayannkKumaar³, Lydia Suganya⁴

CMPN, Thakur College of Engineering and Technology, Kandivali-East, Mumbai, India

zarekarc@gmail.com¹, sandepp543@gmail.com², mayannk27@gmail.com³, lydia.jacob@gmail.com⁴

Abstract—Regular and deep insight Care Unit is where the patients who are seriously sick are look after for treatment. At this critical conditions the team of doctors need to have a real time update of patient's health related parameters like their blood pressure, heart pulse and temperature. It is very frustrating task to calculate it manually, and if patients are in large numbers then it becomes close to impossible. To avoid this type of circumstances the system based on IOT can be used because of its automation nature that can give updates to the Doctors all time using internet. Internet of Things (IOT) Based patients care system is an Arduino based system that takes all patient's information using few sensors. It uses Wi-Fi module to communicate this information to the internet. By using all these sensor of blood pressure and heart beat we can monitor patient health electrically but it has to be physically connected by the patient. When we press the button, the sensor detects the blood pressure and also the heart beat and these details will reach doctors system using internet. The temperature sensor detects the temperature of environment, when this sensor is close to the patient it measures the patient's body temperature. And IOT based automated traffic signal monitoring system mechanizes entire traffic signaling& also allow us for overriding the traffic signal over internet.

Keywords-IOT, Patient Monitoring System, Heartbeat Sensor, Automated traffic signal monitoring.

I. INTRODUCTION

Taking care for anyone after their own family members and other relative in today's fast moving days has become difficult. Keeping track of the health status of the patient at home is a difficult task. Especially for the patients who are very old in their age for them this is the most important work to check their health on time to time based at work. Thus we are trying to make an innovative system that automated this task with ease of use. Our system uses different type of sensors such as temperature as well as for a heartbeat sensing to keep track of patient health. The sensors which have connectivity with a microcontroller are used to monitor the status of patient by using the interface of an LCD display as well as Wi-Fi connection for transmitting alert message. If there is any changes in patient heartbeat or body temperature, the system will automatically send alert message to the user about the patient's status over IOT and also shows information of heartbeat and temperature of patient live on the internet. Thus IOT based patient health monitoring and tracking system effectively uses internet to monitor the patient health status and save precious lives on time.

And for the module of traffic overriding it aims to provide an effective solution for the traffic signal system to change the normal signals during emergency automatically. There is always an emergency situation for ambulance or fire brigade

where they can stuck in traffic; they require priority to go first. Also if there is high traffic at some area, then system uses an android application for controlling traffic and overrides the signals immediately giving green signal in the vehicle direction and red signal for all others. The density is measured in three ways low, medium and high according to which the timings are allotted for signals.[1] The timing overridden is done using Bluetooth technology in android device.

II. BACKGROUND

We are trying to monitor the patient's health from ambulance itself by using the values of various sensors like heartbeat, and body temperature. Hence all these data will be monitored and will be directly send to the doctor's web portal via Wi-Fi module form ambulance itself. When the ambulance arrives at patients door in the emergency it take time to reach hospital and it is possible that patients health can be in danger before reaching hospital then checkup takes place and then according to health conditions precarious are taken and there is also problem for ambulance to get stuck in traffic while reaching hospital. Above described problem can easily get solved, if we use smart ambulance system with traffic overriding. So we are proposing to make a Smart ambulance system with additional facility for ambulance to reach hospital as quick as possible by providing module of traffic overriding in which our idea is to create a system in ambulance itself that can measure the patient's heartbeat, and body temperature and sends this data to the doctor's web portal form ambulance itself, upon which doctor can make a list of prescribed medicine or make sure that all the required things are arranged before ambulance reaches hospital.

III. PROBLEM DEFINITION

There will be a system which includes two modules for making the ambulance smart and automated. One will makes sure that all the patient's health details like heart-beat and body temperature will reaches the hospital via Wi-Fi module the hospital. And the second will make sure that the ambulance does not stuck in the traffic and reaches the hospital as fast as it can by providing the application on the mobile of the driver by which the traffic signal can be overridden and ambulance can make its way to hospital.[2] This both modules are secure by authentication way which includes user-name and password facility to both of them. Which means no other then authorized party can use this system.

IV. SCOPE OF THE SYSTEM

The system collects vital information like body temperature, and heartbeat of the patient and regularly needs to be updatedtothe doctor. The doctor can monitor thedetails of patients' health on web portal now and then take all the required action

which are necessary before the ambulance reaches to the hospital as quick as possible. And android application will make sure that at any point there is no problem for the ambulance to get stuck in traffic.[3] The system can be extended by adding more features like checking blood pressure of the patient, interaction of doctor from hospital into ambulance and etc., Doctors can create awareness about diseases and their symptoms through information that they got on their web portal. From the evaluation and the result obtained from analysis the system is better for patients to reach hospital easily and for doctor to take decision earlier before ambulance reaches hospital.

V. APPLICATION OF THE SYSTEM

The paper has a variety of application such as system for fire bridged driver, police vans, sending health details using internet to hospital directly in serious conditions, overriding traffic by the authorized person in serious conditions and etc.[4]

The System has various impact such as doctors in hospital stays very alert and takes decisions by taking time, Patient health parameter data is reached to doctor before patient physically reaches in hospital. [5] So it is more beneficial as there can be precarious taken before time, In critical condition an emergency call is given to ICU and ambulance, Doctors can take all necessary action without visiting to patient, Highly sensitive.

VI. LITERATURE SURVEY

We have conducted a literature survey based upon which we have found the gaps in the existing system which can be seen in the table below (Table-I).

TableI: Literature Survey

Existing System Authors	Title of the paper	Gap Identified
Vikas Vippalapallandi Snigdha Ananthula [1]	Internet of things (IoT) based smart health care system	Harmful for patient body as sensors are placed on body, Sensors can get defect because of outside world conditions, And the cost of sensors used are very high.
Xiaofeng Ye, Bin Sun, Weiwei Lin [2]	A remote control system on traffic lights based on IEC61131-3 standard language	Setting of the traffic signal needs monitoring on fixed time delay, Battery need to change at regular time interval, Resolution of the screen used for monitoring purpose is very small.
Mukremin Ozkul, Elton Domnorri [3]	A traffic signal control system through anonymous messages	Not Compact, Message traffic due to unnecessary needed message, communication is based on short message, Real time monitoring of traffic can get incorrect, Wireless communication with traffic signal can get disturb due to connection issue.
	IoT based smart	

Punit Gupta, Deepika Agrawal, Jasmeet Chhabra[4]	healthcare kit.	Not Portable, Overhead due to carrying, Emergency message can get disturbed and unnecessary messages can be send.
B.Janani Saradha, G. Vijayshri, G. Vijayshri[5]	Intelligent traffic signal control system for ambulance using RFID and cloud	Speed is slow, Time required is more, Overhead due to cloud implementation.
Salman Ahmed, Sabrin Millat, Md. Aymanur Rahman[6]	Wireless health monitoring system for patients	Overhead due to Telecommunication technology implementation
Y M Jagadeesh, G. Merlin Suba, S Karthik[7]	Smart autonomous traffic light switching by traffic density measurement through sensors	Complex implementation because of Artificial Intelligence, There can be problem generated due to dynamic control of traffic signal.
Yujun Ma, Yulei Wang, Jun Yang[8]	Big Health Application System based on Health Internet of Things (IoT) and Big Data	Required huge amount of data storage for big data.

The Existing System uses Raspberry pi kit which increases the cost of the entire system and overhead as the existing system build a network, speed of execution is also slow and time taken is more, in our system we have proposed to use a Arduinokitwhichreduces cost as well complexity and No Overhead, our System contains One small sized system in which all components will be fixed and one more advantage is after prescription of doctor prescribed medicine will be taken before it reaches hospital and traffic problem is solved with the following components such as Bluetooth, Wi-Fi Module, Switch, Heartbeat, and Temperature Sensors, Power Supply Modules and Software like IOT Gecko libraries, and embedded language.

VII. PROPOSED WORK

Smart ambulance system with traffic overriding is very user friendly and in terms of handling also. [6] Smart ambulance system with traffic overriding can be needed to monitor and overriding it in emergency. Our system consist of two modules i.e.; patient health monitoring system (Fig 1.1) and traffic overriding system (Fig 1.2) It can detect and monitor the temperature and heartbeat of the patient and sends it to the hospital from the ambulance itself so that the crucial time which is needed to save the patient's life can be saved because it gives details to the hospital's management and doctor for taking the precautionary measures.

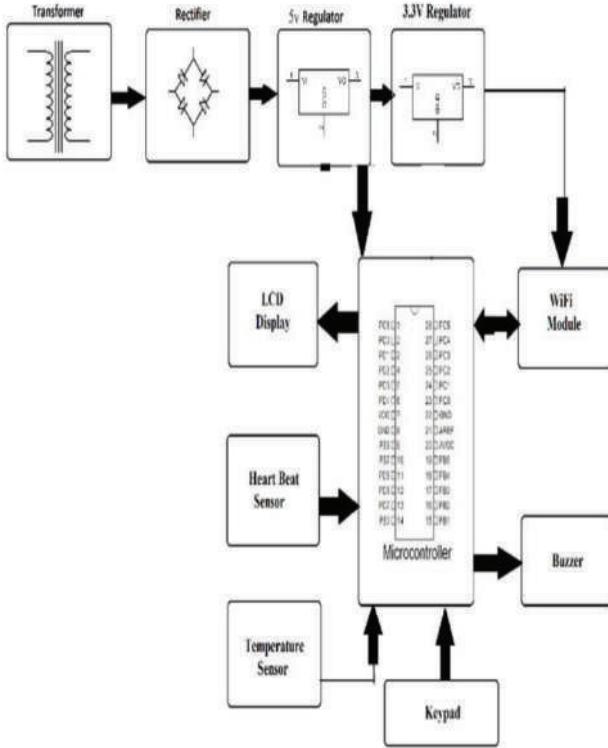


Fig 1.1: Block Diagram (Patient Health Monitoring Module)

When there is lot of traffic the density of the traffic will be measured by the sensors present on the board which will help in toggling the traffic light from red to green and vice versa. [7]

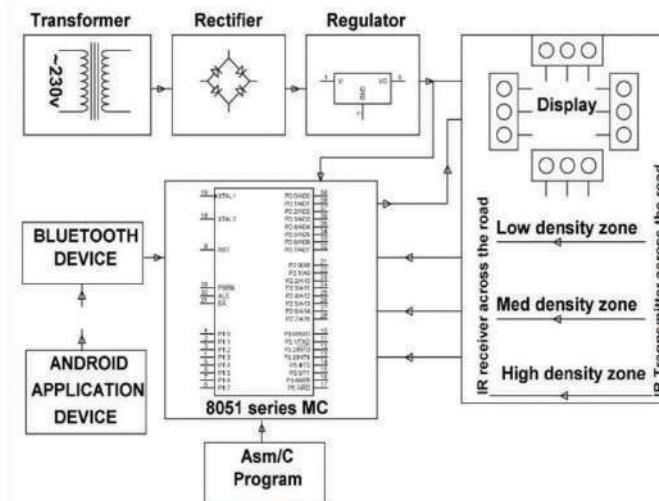


Fig 1.2: Block Diagram (Traffic Overriding Module)

Overriding of traffic can be used to make the ambulance to reach hospital as quick as possible. Most important thing is that the

measures which have been given by the doctor from hospital can be given to the patient so that his life can be saved. Data transfer speed given from the ambulance is quite high so that the doctor's precautionary measures given can be sent quickly

VIII. EXPECTED OUTCOME

The system will provide facility of monitoring patient's health using various types of sensors. Using the android application the driver of ambulance can avoid the traffic and makes its way to the hospital. [8] It will improve the efficiency of the health monitoring and traffic signal overriding with this system. Fig-2 Shows the system Architecture of Smart Ambulance System.



Fig 2: System Architecture for Smart Ambulance System

IX. SYSTEM WORKFLOW

A) Patient side (Arduino module):

Temperature sensor will sense the temperature, heartbeat sensor will give the pulse count of patient continuously through Arduino using Wi-Fi module. Monitoring will be done continuously. Thiskit will be near to the patient. The details of the patient will be transferred to the doctor side web portal.

B) Doctors Side Web Portal (PC):

Server will collect the data from patient side kit and will update it in data log. It will check if the condition is normal or critical, according to that action will be taken. If the condition is not normal the doctor will make prior arrangements before the patient arrives. If doctor gives any prescription to patient, it will send signal to secondary doctor present in the ambulance. And the secondary doctor will provide the medication to the patient in the ambulance.

C) *Traffic Signal overriding (android app):*

When there is lot of traffic the density of the traffic will be measured by the sensors present on the board which will help in toggling the traffic light from red to green and vice versa. Suppose there is a ambulance travelling from a lane which has traffic congestion so we can override the traffic signal with the help of this android application. This application will be handed over to an authorized person so that one cannot misuse or play around with traffic signal.

X. EXPERIMENTAL SETUP

The System is proposed with the aim of providing efficient healthcare to the patient for continuous monitoring and providing help in case of critical situation for which we have used Arduino, temperature sensor, heartbeat sensor, Wi-Fi module and LCD display as shown in figures listed from Fig 3.1 – Fig 3.4

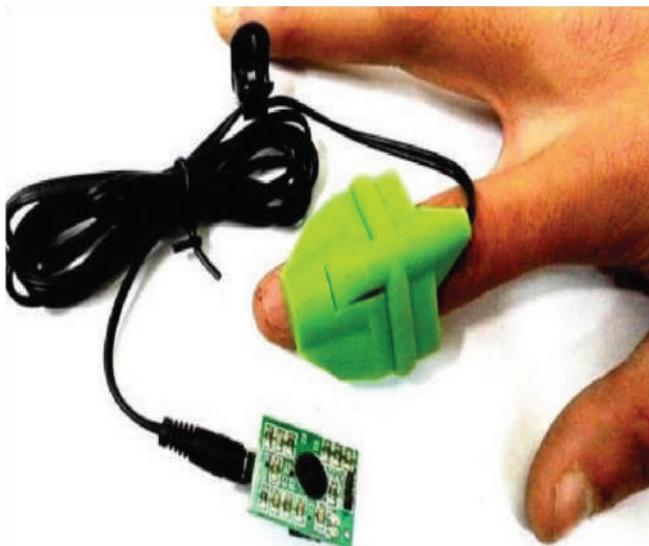


Fig 3.1:- Heartbeat Sensor (Measures the heartbeat of the patient)



Fig 3.2:- Temperature Sensor (Measures the body temperatureof the patient)

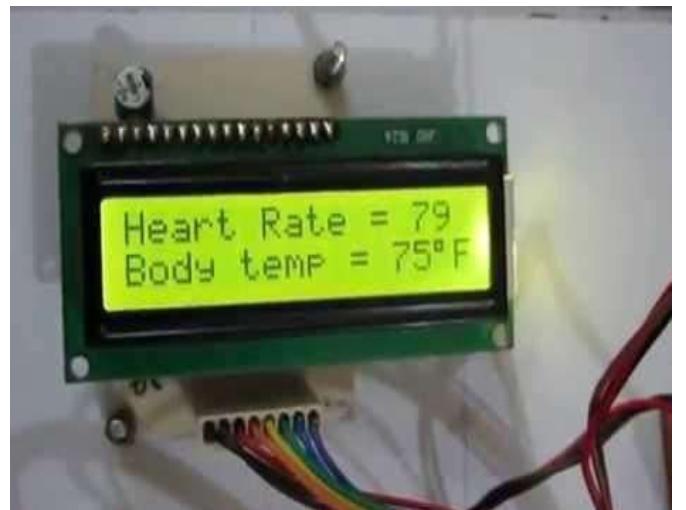


Fig 3.3:- LCD Display (Displays Result)



Fig 3.4:- Wi-Fi Module (Used for transferring of data)

XI. RESULTS

The User interface of the android application used in traffic overriding module proposed System is as shown in the fig 4: User Interface (Traffic Signal Overriding) and we have shown the comparison between the existing system and the proposed system with respect to certain parameters as shown in Table 2.

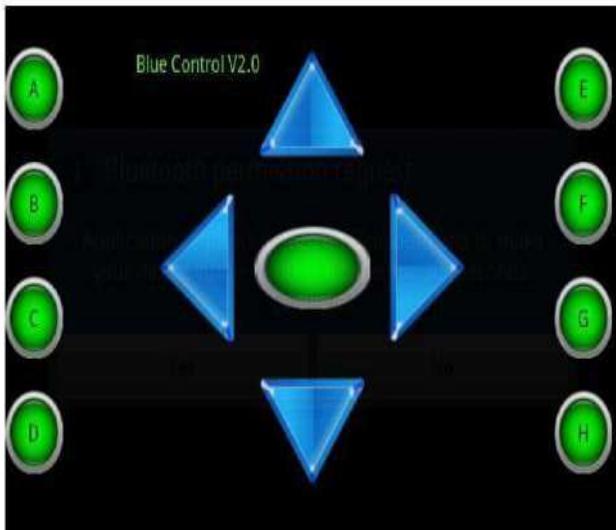


Fig 4: User Interface (Traffic Signal Overriding)

The table below shows the comparison between the existing system and the proposed system based on certain parameters as described below:

Table II: Existing vs. Proposed System

Parameter	Existing System	Proposed System
No. of Components used	Many	Medium(As compared to others)
Cost	High	Medium
Portable and compact	Portable and compact(due to wearable tags)	Portable and compact (due to size)
Type of Kit(Arduino /Raspberry pi)	Raspberry Pi	Arduino
Overhead and Speed/Time	Yes	No
User Interface	Complicated	User Friendly

XII. CONCLUSIONS

An idea is proposed for saving a precious life of patient's in a better and efficient way possible. It is beneficial for users in case of emergencies as it saves time. The smart ambulance provides necessary equipment required to track the patient's health status. The live feed data sent through the

ambulance to the hospital helps in keeping track of patient's health details. Sending patient's health information to the hospitals helps the hospital staff to get the necessary pre-requisites regarding the patient's treatment. Hence it reduces the time complexity and helps to provide faster medical services.

The traffic signal overriding module only gives the authority to the ambulance driver in order to have control over the traffic signal. The signal can be changed depending on the traffic condition. The signal is changed to green in case of heavy traffic. Thus the ambulance takes the patient to the hospital in a short period of time without any delay.

REFERENCES

- [1] VikasVippalapalli ETM Dept., GNITS, Hyderabad, India http: and SnigdhaAnanthula ETM Dept., GNITS, Hyderabad, India "Internet of things (IoT) based smart health care system" Published in: Signal Processing, Communication, Power and Embedded System (SCOPES), 2016 International Conference //ieeexplore.ieee.org/document/7955637/.
- [2] Xiaofeng Ye College of Metrology and Measurement Engineering, China Jiliang University, Hangzhou 310018, China; Bin Sun College of Metrology and Measurement Engineering, China Jiliang University, Hangzhou 310018, China; and Weiwei Lin College of Metrology and Measurement Engineering, China Jiliang University, Hangzhou 310018, China "A remote control system on traffic lights based on IEC61131-3 standard language" Published in: Intelligent Control and Automation (WCICA), 2010 8th World Congress on //ieeexplore.ieee.org/document/5553830/
- [3] MukreminOzkul Department of Computer Engineering, Epoka University, Tirana, 1039, Albania, Elton Domnori Department of Computer Engineering, Epoka University, Tirana, 1039,
- [4] Albaniahttp:"A traffic signal control system through anonymous messages Sign In or Purch" Published in: Intelligent Vehicles Symposium (IV), 2017 IEEE //ieeexplore.ieee.org/document/7995893/
- [5] Punit Gupta Jaypee University of Information Technology, Himachal Pradesh, India, DeepikaAgrawalJaypee University of Information Technology, Himachal Pradesh, India,
- [6] JasmeetChhabraJaypee University of Information Technology, Himachal Pradesh, India "IoTbased smart healthcare kit" Published in: Computational Techniques in Information and Communication Technologies (ICCTICT), 2016 International Conference //ieeexplore.ieee.org/document/7514585/
- [7] B. JananiSaradha Department of IT, Sri Sairam Engineering college, India, G. Vijayshri Department of IT, Sri Sairam Engineering college, India, T. Subha Department of IT, Sri Sairam Engineering college, India "Intelligent traffic signal control system for ambulance using RFID and cloud" Published in: Computing and Communications Technologies (ICCCT), 2017 2nd International Conference http://ieeexplore.ieee.org/document/7972257/
- [8] Salman Ahmed Department of Electrical and Electronic Engineering (EEE) American International UniversityBangladesh, Dhaka, Bangladesh, SabrinMillatDepartment of Electrical and Electronic Engineering (EEE) American International University-Bangladesh, Dhaka, Bangladesh, Md. AymanurRahmanDepartment of Electrical and Electronic Engineering (EEE) American International University-Bangladesh, Dhaka, Bangladesh, "Wireless health monitoring system for patients" Electrical and Computer Engineering (WIECON-ECE), 2015 IEEE International WIE Conference http://ieeexplore.ieee.org/document/7443888/
- [9] Y M Jagadeesh Department of Mechanical, Panimalar Engineering College, Chennai, India, G. Merlin Suba Department of Mechanical, Panimalar Engineering College, Chennai, India,
- [10] S KarthikDepartment of Mechanical, Panimalar Engineering College, Chennai, India "Smart autonomous traffic light switching by traffic density measurement through sensors" Published in: Computers,

- Communications, and Systems
(ICCCS), International Conference <http://ieeexplore.ieee.org/abstract/document/7562885/>
- [11] Yujun Ma Center of Information Development and Management, Nanyang Institute of Technology, Nanyang, China, Yulei Wang Center of Information Development and Management, Nanyang Institute of Technology, Nanyang, China,
- [12] Jun Yang School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China; "Big Health Application System based on Health Internet of Things and Big Data" Published in: IEEE Access (Volume: 5)
<http://ieeexplore.ieee.org/document/7782325/>

Android Application for Conference Alert

Deshmukh Ravish Zafar, Khan Darakhshan Zain, Shaikh Faizan, Ansari Bilal ZakirAli

Department of Computer Engineering, AIKTC, Mumbai, India

ravishkhan2603@gmail.com, khandarakhshanzain@gmail.com, fbi056@gmail.com, ansaribilal1a@gmail.com

Abstract—This Paper focuses on developing an android application for academic conferences alert such as IEEE paper conferences etc. This Application will provide the user with the information about the conferences with venue, dates and timings that will make it easier for the user to attend the conferences that are going to be held. Scrapping is also going to be done to get the data. This application will make it easier for the user to know about the conferences.

I. INTRODUCTION

Conference is a formal meeting in which group of people gather in order to talk about ideas or problems related to topic. Conferences can be of various types such as Business Conferences, Academic Conferences, Teacher-Parents Conferences, etc. Our Paper focuses on Academic Conferences such as IEEE Paper conferences, etc. As nowadays Academic Conferences have become very much important from student point of view as they get opportunity to showcase their talent in the field of their interest. Scrapping will also be used for the development of this project. Scrapping means getting the useful data from various websites for our own purpose.

The main purpose of developing this application is to make it easy for the user to get knowledge about the conferences that are going to take place with venues, dates and timings. As nowadays most of the people have smart phones with them it would become easy for the users to get notification directly on their phones instead of surfing various sites for getting the information about the conferences

II. SYSTEM ARCHITECTURE

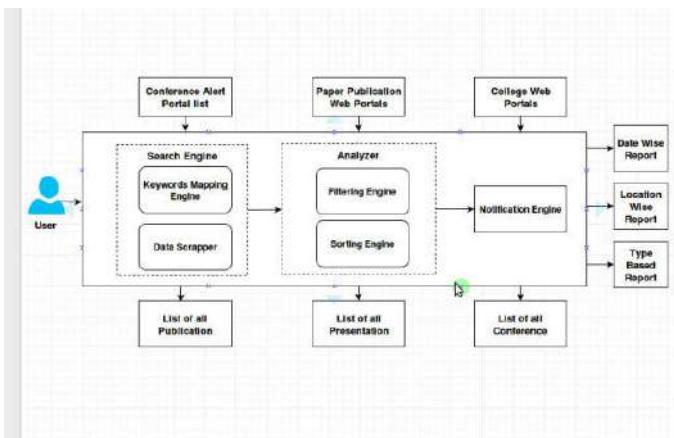


Fig 1. Scrapping

First, we will be doing scrapping by getting the data from various sites. Only the important data will be taken so that there won't be

any useless information. From various conference alert sites, paper publication portals as well from college web portals we will be scrapping the information. So after getting all the information we will be filter that data. Filtering means only the data that we going to provide to the users will be shown. Technology that is going to be used for scrapping is “beautiful soup”.

A. Urllib2

This is the library of Python that is used to fetch various URL. After fetching the further scrapping of data will be done by the python library named as “Beautiful soup”.

B. Beautiful Soup

Beautiful soup is a Python package for parsing HTML and XML documents (including having malformed markup, i.e. non-closed tags, so named after tag soup). It creates a parse tree for parsed pages that can be used to extract data from HTML, which is useful for web scrapping. It is available for python 2.6+ and python.

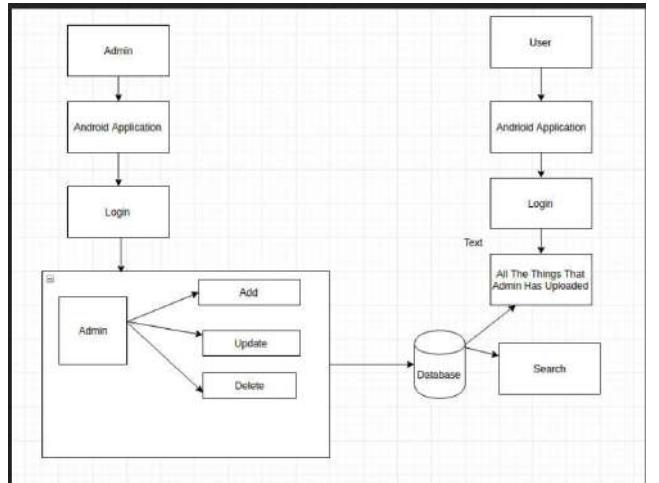


Fig 2. Working of Application

The working will be as follows:

All the operations will be done by the Admin. He will be logging in and all the data will be posted by him that is all the notifications about the conferences will be posted by him. Any updating, insertion or deletion will be done by him. All the notifications that the admin is going to post will automatically be stored in the database.

So when the user downloads the application after his login all the data that is stored in the database will be available to the user. He can search for whatever conferences he wants.

III. MARKET SURVEY

After surveying the market, we found that there are many conference websites but there is no application for such conferences so to make it easy for the users to get the knowledge about the conferences we are working to develop this application. As most of people have smart phones it people feel it boring surfing the websites. So this application could be beneficial from their point of view.

IV. ACKNOWLEDGEMENT

- 1.code red alert application
- 2.Time alerts
- 3.News Notifier

We would like to thank Professor Mukhtar Ansari for guiding us and making it possible for us making this paper. We would also like to thank our computer department for giving us the opportunity to showcase our capabilities. At last but not the least we would also like to thank our institute for giving us such a good platform and help us develop ourselves.

V. Preferences

Technical Approach for Agriculture

Musa Asmat Pawle, Nilofer Peerzade, Atif Thakur, Meenaz Gigani

Department of Computer, A.I.K.T.C School of Engineering, Panvel, India
musaarfa@gmail.com, nilofarpirjade@gmail.com, atifthakur5126@gmail.com, meenazgigani0310@gmail.com

Abstract—Today, convection agricultural is now improving to emphasise the productivity of the field. The agriculture, production becomes nowadays important in two terms quality and quantity. Under the Bureau of Indian Standards IS 15930(Part 1): 2010 requirements for good agricultural practices have been prescribed. From the research work of Mr S.K. Jadhav el at 2016, information like crop period, whether pruning is the foundation or forward pruning and appearance of infection of downy mildew etc., his team suggests various preventive measures and different pesticide treatments [5]. They conclude that the knowledge-based system will be helpful to an agricultural professional to take decision-related to the management of crop. Today's scenario was facing the problem of less space and more output. Overall expenditure of agricultural field reduced to half of its value. Use IOT based smart agricultural system give new research area in field. India's food deficient was changed to leading agricultural status. 21st-century the market is facing the main problem of "smart customer". In this plight, technology is playing a vital role to uplift the agricultural production. Due to boom explosion in population, there is a vast improvement in agricultural machines over last century. As humans are making more relevant them self with monitoring systems, GPS locators, maps and an electronic sensor, these technologies start taking the stand in the agricultural field. Agricultural engineers work on planning, supervising, and managing the building of dairy effluent schemes, irrigation, drainage, and flood and water control systems. They aim to conserve soil and water and to improve the processing of agricultural products. In this paper, there is a survey on the technical aspect of agriculture.

Keywords—agricultural, smart customer, technology etc.

I. INTRODUCTION

The agricultural field has a very vast history in the human race. India is well known as Golden Sparrow due to its production of agriculture in each era of the races. The complexity of the architecture is come due to the three different season of the environment and geographical structure of the nation. These give a very important position in the agricultural field in India.

R Gopalakrishnan and Dr YSP Thorat in collaboration with Tata companies have surveyed and concluded that agricultural growth was excellent till 2000 and then it has been slowly declined [1]. To which an urgent attention is required from the youth of the nation. As we know, the agricultural field provides employment to many youths.

Even though the share of agriculture in GDP has declined to one fifth from one-half at the time of independence, agriculture remains the predominant sector in terms of employment and livelihood provision for more than half of India's workforce engaged in it as the principal occupation [2]

Today every market in the world is facing the main problem with the concept of the Smart customer. To keep up with the requirement of this new Smart customer, agricultural field required new technique to improve quality and production. In this aspect, science and technology provide hands to this field. Scientific research provided data to improve the quality whereas technology approach provides control on the quality and production.

II. APPROACH TO SMART GREENHOUSE

With the help of smart greenhouse, we can able to have crops free from insecticide and pesticide that improve the quality of the crop. The functional block diagram for the smart greenhouse system is given in Fig 1.

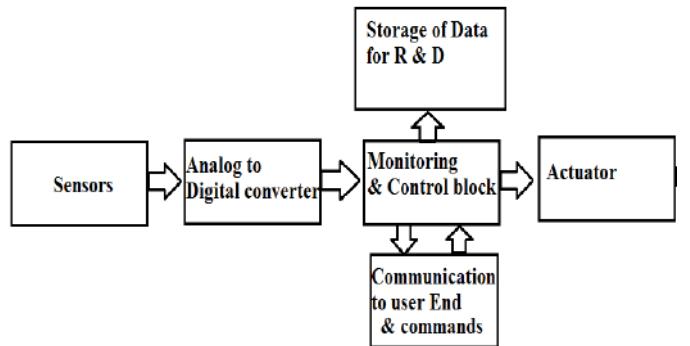


Fig 1: Functional block diagram for A Smart Green House System.

III. SENSORS

Sensors are the main key feature of the system. There are many sensors are available in the market to sense a different parameter of the environments. The Green House system has to take care of the following parameters:

- A. Humanity and Temperature of the system in Centigrade.
- B. CO₂ in PPM.
- C. Light in K Lux.
- D. Moisture from Air and Soil.

E. pH values of the crop.

A. Humanity and Temperature sensor:

For the temperature sensor, LM 35 is the best choice since it provides the precision integrated-circuit temperature devices with an output voltage linearly proportional to the Centigrade temperature. The LM35 device has an advantage over linear temperature sensors calibrated in Kelvin, as the user is not required to subtract a large constant voltage from the output to obtain convenient Centigrade scaling. The DHT11 Temperature & Humidity Sensor has features to measure temperature & humidity and provide a calibrated digital signal output. This sensor includes a resistive-type humidity measurement component and an NTC temperature measurement component, offering excellent quality, fast response, anti-interference ability and cost-effectiveness.

B. Carbon dioxide:

MQ135 are used in air quality control equipment for buildings/offices, are suitable for detecting of NH₃, NO_x, alcohol, Benzene, smoke, CO₂, etc. MQ-135 performs a good detection to smoke and other harmful gas, especially sensitive to ammonia, sulfide and benzene steam. Its ability to detect various harmful gas and lower cost make MQ-135 an ideal choice for different applications of gas detection. The detection range of CO₂ sensor has to set between 0~1000ppm through the pot. If the level of CO₂ crosses the 30% more than required, the system needs to take care of the production of CO₂.

C. Light in K Lux:

Usually, LDR is used as a light sensor. LDR is easy to use and quite reliable for the job. LDR stands for Light Emitting Diode. Its resistance varies according to the intensity of the light. A wide range of LDR is easily available in the market.

D. Moisture from Air and Soil:

This sensor can be used to test the moisture of soil when the soil is having a water shortage, the module output is at a high level, or else the output is at a low level. In soil, dielectric permittivity is a function of the water content. The sensor creates a voltage proportional to the dielectric permittivity of the surrounding medium. By using this sensor one can automatically water the flowering plant, or any other plants requiring automatic watering technique. Module triple output mode, the digital output is simple, analogue output more accurate, serial output with exact reading.

E. pH values of the crop:

A pH (potential of Hydrogen) probe measures the hydrogen ion activity in a liquid. At the tip of a pH probe is a glass membrane. This glass membrane permits hydrogen ions from the liquid being measured to diffuse into the outer layer of the glass, while larger

ions remain in the solution. The difference in the concentration of hydrogen ions (outside the probe vs. inside the probe) creates a VERY small current. This current is proportional to the concentration of hydrogen ions in the liquid being measured.

IV. ACTUATORS

These are the main block of function diagram. These actuators control the shutter, fan and water supply and other expected output of the system. Servo motor, Dc motor are commonly used for the mechanical support in shutter closing and opening. These motors are small in sizes and require less power to work on. There are many different types of gear motors available with different torques. It depended upon weight and the friction required we can select motors. Solenoids are used as a mechanical switch for the water supply. Fans are there to control the concentration of the gas CO₂.

V. MONITORING AND CONTROL BLOCK

The microcontroller is utilized as a core part of the framework in the system. The microcontroller is a marvellous discovery of the 19th century. Astounding improvement in microcontroller drives it to limitless lift to the electronic field. There exist quantities of vendors and their groups of microcontrollers in the market. AVR is a family of microcontrollers developed by Atmel in the beginning of 1996. These are modified Harvard architecture 8-bit RISC single-chip microcontrollers. AVR was one of the first microcontroller families to use on-chip flash memory for program storage, instead of one-time programmable ROM, EPROM or EEPROM utilized by different microcontrollers at the time. The AVR is a modified Harvard architecture machine, where program and data are stored in separate physical memory systems that appear in different address spaces, but having the ability to read data items from program memory using special instructions [3]. Presently, commercial off-the-rack are covering market colossally. The extremely well-known is Arduino Uno. The Arduino Uno is a microcontroller board based on the ATmega328 (datasheet). It has 14 digital input/output pins (of which 6 can be used as PWM outputs), 6 analogue inputs, a 16 MHz ceramic resonator, a USB connection, a power jack, an ICSP header, and a reset button. It contains everything needed to support the microcontroller [4].

VI. STORAGE OF DATA

Cloud computing is an internet-based computing service in which large groups of remote servers are networked to allow centralized data storage and online access to computer services or resources. Cloud Service Models are Infrastructure as a Service (IaaS), Software as a Service (SaaS) and Platform as a Service (PaaS). With the help of Infrastructure as a Service (IaaS), we can virtualize the computing resources. Software as a Service (SaaS) is a mechanism for delivering software that provides remote access to software as a web-based service. Platform as a Service (PaaS) allocates a platform and an environment which allows the developers to build applications and services through the internet. There are many cloud storage services available to Amazon Web Services (AWS), Google Cloud Platform (GCP), Windows Azure

Services Platform and ThinkSpeak etc. Cloud computing's one of the many key features is it, eliminates the capital expenses for setting up infrastructure. There are diff types of Clouds like Public Cloud, Private Cloud and Hybrid cloud available on the market today. In Public Cloud, a third party provides services to the customer. Private Could and Public Cloud are the analogous only difference is in Private Cloud is that the data and services are managed by an organization and it is more secure then Public Cloud.

VII. COMMUNICATION

There are Arduino shields, which can fix on board. We can use them externally also through ESP8266, GSM module and Zigbee. Zigbee devices can transmit data over long distances by passing data through a mesh network of intermediate devices to reach more distant ones. Zigbee has a defined rate of 250 kbit/s, best suited for intermittent data transmissions from a sensor or input device. GSM modules are used to send a message or miscall over a network. The ESP8266 is a low-cost Wi-Fi chip with full TCP/IP stack. It's a transceiver that can be added to any existing microcontroller based setups via UART (serial link) to enable the system to communicate over the Internet via Wi-Fi. Commands will be sent from an android application through the internet and to Wi-Fi network to ESP8266 with Arduino Mega. For data storing we can save it on Google drive.

As shown in the Figure 2, numerous parts are being utilized to build up this project. One of the fundamental segments is a sensor which detects the information from nature. Another vital segment is Arduino which chooses the fundamental stream of the framework. It takes the information from the sensor and chooses which actuators are to be enacted to meet the rural necessities

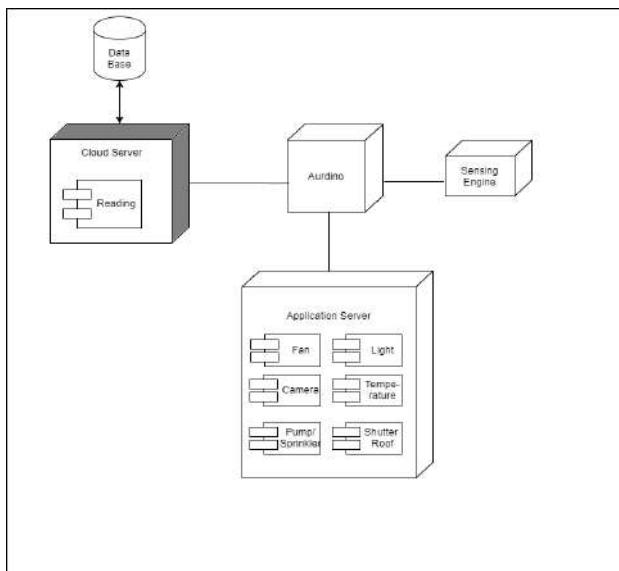


Fig 2: Deployment diagram for A Smart Green House System.

In the event that the additional assets are required to meet the necessities, it sends an activation signal to the respective actuator.

For instance, if the light is not present in the environment it will enact the light actuators in the framework to make the light accessible. This data is continuously sent to the Cloud Server for storing in Database. This Cloud Server is connected by the means of WiFi.

VIII. CONCLUSION

There are the ways to make a smart move toward the need for agriculture field; Technology of today's era is providing a hand in improvement. Agricultural professional and the management have to look into this matter. In conventional farming, we have to face lots of problems. With help of technology, we can able to produce insecticide and pesticide free crops and create a suitable environment for the proper growth of plants. Most important the Smart customers of today's era are directly connected to the farmer. So, the farmer knows the requirement of end users that improve the quality of crops.

REFERENCES

- [1] R Gopalakrishnan Dr YSP Thorat, What India can do differently in agriculture TATA companies (2015).
- [2] Planning Commission, Document of Eleventh Five Year Plan, 2007-12: 1. Inclusive Growth, Vol.1, (Chapter 1, p.1), and 2. Agriculture, Rural Development, Industries, Services and Physical Infrastructure, Vol. 4 (Chapter 1, p.1 and Chapter 4, p.79), Oxford University Press, New Delhi.
- [3] https://en.wikipedia.org/wiki/Atmel_AVR.
- [4] <http://www.arduino.org/products/boards/arduino-uno>.
- [5] Mr S.K. Jadhav, Dr R.D. Kumbhar, A Study of Prototype in the Rule-Based Expert System for the Management of Downy Mildew Disease in Grape Crop, Agricultural Situation in India, March 2016, pp.42-48

Cloud-based Core Banking Software for Small Finance Institutions

Manjiri Bhat, Nishant Shishodia, Vikram Rajpurohit, Narayan Mundhara, Anjali Yeole

Department of Computer Science, Vivekanand Education Society's Institute of Technology

Abstract - Credit societies are small institutions that work like banks for a limited number of people. These societies need a common platform for maintaining their database and transactions. The objective is to create a cloud-based system for these credit societies such that they can have a core banking application which can connect them with other small financial institutions. The project is developed with respect to the agile methodology. The software will provide high-level security and will be mobile and affordable.

Keywords - Banking applications, Cloud-based system, cloud-computing, Credit Society, Indian-client base, SaaS.

I. INTRODUCTION

Road As the world is moving towards digitization, a plethora of web-based solutions in the different aspects of life are emerging. Retail, finance, entertainment, etc. are all going digital. With digitization, comes the need for the development of a platform for these services. Particularly in the banking sector, a complex platform for the intricacies of the banking world is necessary. A new trend of using cloud-computing in the banking sector to reduce development costs is emerging.^[3]

Different solutions in the form of PaaS (Platform as a service), SaaS (Software as a service) and IaaS (Infrastructure as a service) which are different models of cloud-based technology are being developed.

Credit societies are basically small institutions established by a group of people which work like banks for a limited number of people. The number of members can range from 20 to 2000. Still, it is a small number to actually call it a bank but larger than a single account. Lending money, buying shares in the societies, forming groups to save money, depositing money for safekeeping are some of the characteristics of a Credit society, also called as "Patpedi". These credit societies also need a digital platform for maintaining their database and transactions. We aim to fulfill this need.

II. OVERVIEW

Sometimes referred to as "on-demand software"^[4], SaaS is typically accessed by users from a thin-client via a web browser. SaaS has become a common delivery model for many business applications. The cloud database distribution and storage, as well as the system infrastructure is managed by the service providers and the users have access to the application and their database.

One of the biggest constraints on small finance institutions to get automated software is the large investment costs involved in product development and the following continual maintenance costs. The other hidden costs for individual project development are as shown in the Figure 1. We aim to overcome this drawback by

creating a system which will be based on monthly billing and pay as per usage.

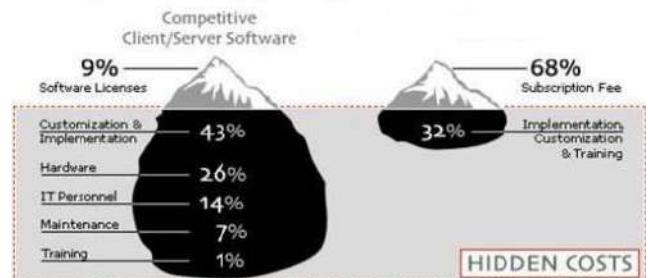


Fig 1: Hidden costs in existing system

III. PROPOSED SYSTEM

"MahCredit" is a cloud-based credit banking software for small financial institutions. The proposed system architecture is shown in Figure 2.

The feature for Devanagari display will also be incorporated in the system to make it easier to enter the data for people who are not comfortable with English. Also, standards for security and privacy of data in the system will be maintained.

A. Users

These are the owners or managers of the individual credit societies which avail the services provided by MahCredit.

B. GUI for SaaS

This is only part of the system in which user can enter features according to their requirements. GUI for SaaS is the graphical user interface provided by the system to its users. It is the topmost layer of the cloud-based system. It is necessary for user interaction. All the forms and tables are displayed or edited using the graphical user interface.

C. CCB Infrastructure

It is basically the connection between the database and the graphical user interface. Data processing, validation, authentication and conclusions are drawn in this layer of the architecture. This layer is vital to the proper working and functionality of the system.

D. DB

DB stands for databases created for different entities of the system. The first step when a user avails the services of MahCredit is his registration in the CS DB i.e. the credit society database.

IV. METHODOLOGY

We aim to develop this system using Free and Open Source Softwares(Foss). For the development of the client-side infrastructure Laravel 5.4[2], currently the most popular framework, is chosen. The database design will be created using PostgreSQL.

ASCII cannot be used to represent more than one script, therefore to store the data in Devanagari script Unicode is necessary. For Unicode to work in a website, the browsers should be Unicode enabled. Unicode is the way to store Indian script data in SQL databases. UTF-8 and UTF-16 are the two versions of Unicode available. [5] UTF-8 is required for the devanagari data storage for customer information in MahCredit.

V. CONCLUSION

This system has the Indian credit society community as its target demographic. It is developed using open source technologies which were absent in the previously studied system. MahCredit will

be developed using Laravel framework as it is the most popular framework with a lot of functionalities. We have chosen PostgreSQL because it is a FOSS. MahCredit once implemented will provide an effective online solution for the credit societies in India.

REFERENCES

- [1] Present credit banking system software implemented on individual systems.
- [2] Laravel 5.3 official documentation
- [3] Dr. Sheel Ghule, Ms. Rupali Chikhale, Mr. Kalpesh Parmar, “Cloud Computing in Banking Services”, International Journal of Scientific and Research Publications, Volume 4, Issue 6, June 2014 1 ISSN 2250-3153
- [4] Chitrali Agre, “Implementation of a Cloud in Banking Sector”, International Journal of Computer Science and Information Technology Research ISSN 2348-120X
- [5] Dr. ARD Prasad,” Creation of Digital Libraries in Indian Languages Using Unicode”, Workshop on Digital Libraries: Theory and Practice, March 2003, DRTC, Bangalore

A Scalable Hybrid Network Control Plane for Cloud Computing

Shreyash Sharma, Neha Kamble, Anand khandare

Computer Engineering, Thakur College of Engineering and Technology, Mumbai, India

Abstract—One aspect that many of us tend to forget about the cloud is that it's not completely digital. At one or the other place in the world, there has to be a data center or physical server that works like the backbone of cloud computing. SDN concentrated on separation of the control plane of the network, which decide about how packets should flow through the network from the data plane of the network, which actually moves packets from place to place. SDN indicates yet another step to a fully digital infrastructure for vendors as well as clients. Exploring the various missions and technology models of SDNs is difficult to position cloud services and realize advantages of cloud computing. For cloud users, who know their cloud providers' SDN plans, as well as the plans of private cloud software vendors, is the most critical element in assessing these providers' long-term value.

Keywords—software defined networking (SDN), backbone, switches, control plane, cloud networking

I. INTRODUCTION

The cloud computing model has been widely adopted in recent years. It provides a cost-effective solution for deploying new applications, removing the upfront hardware cost as well as the burden of system maintenance. The cloud computing model allows resources to be allocated on demand and provides the ability to elastically scale-out processing. SDN separate the network's control plane and data plane, and find complex control functions from network devices. It also supports a flow-based management, based on the OpenFlow protocol to enable highly programmable and flexible networks. Currently, almost all commercial switches support the Open Flow protocol. They consider the flow control to be a very important feature for supporting innovative applications, but also an inherent problem of flow-based SDN, because it introduces a great communication overhead between the data plane and the control plane, which limits the scalability. Many solutions to this problem have been explored. Some researchers design different control plane structures to stretched the control plane's processing ability. Consider some example, some studies construct flat control plane architecture to improve the control plane scalability and reduce the delay caused by geographical distance, as in HyperFlow, Onix and ONOS. But some studies build a centralized hierarchical control plane of SDN, in which the top layer controller is responsible for global applications service. Consider another example, Kandoo has a two-layer hierarchical architecture, in which the bottom layer controllers run local control applications based on the local network view, and the top layer controller runs global applications based on the global network-wide view. Logical xBar, on the other hand, introduces a recursive

building block to construct a centralized logical hierarchical SDN network. Though the above explained control plane architectures can improve the scalability of SDN networks, the flat and the centralized hierarchical architecture have limitations. Since routing is an essential control operation of a SDN network, they take the Dijkstra algorithm as an example to illustrate the problem. 1) The flat control plane architecture cannot solve the super-linear computational complexity growth of the control plane when a SDN network scales to large size. To illustrate this problem, they use the source IP address and the destination IP address together to identify a flow. They take an example to illustrate the problem. 2) The centralized logical hierarchical control plane architecture brings path stretch problem.

II. RELATED WORK

The SDN control plane is designed to receive connection requests either from the application layer explicitly or to monitor the traffic in the network and autonomously steer the bandwidth across the network based on the traffic characteristics. The control plane can scale in two directions: out or up. In the scale-out approach, the control plane functions are separated and distributed across physical or virtual servers. In the scale-up approach, the server's processing power is augmented by adding extra compute resources, such as x86 processors [1]. While doing survey we come to know that the scale-out and scale-up architectures, performance can be further enhanced by providing function-specific hardware acceleration [14]. Data centers and the terabits capacity of the DWDM network, this architecture can also be useful as a design model for large-scale data centers to limits due to energy consumption and network size through distribution of clusters over short distances. The control plane has capability to monitor traffic and collect the historical data, can improve the automation of network control and management. [10], hybrid cloud architecture is used that suited to the needs of social science researchers. They have described an implementation that expands and contracts a cluster adaptively to the current overall cluster workload as well as methods that are secure, efficient, and simple for storing data and providing secure networking [8]. An optimization problem was discovered to solve the problem of convex. They analyzed the channel utility, packet drop ratio, average transmission delay, and energy consumption to show the effectiveness of the hybrid MAC protocol, there are heterogeneous devices with different priorities [12]. Demonstrated how such a performance improvement leads to a scalable routing solution in the context of VANET environments. This have also shown how their main conclusions hold, even in the presence of location errors [9]. LazyCtrl can help reduce the workload of the central controller, improving the scalability of standard OpenFlow to a large extend [11]. The hybrid adaptive fuzzy controller could perform successful control without incorporating any lexical description into the design. Adaptive

fuzzy identification can be combined with, and can reinforce, adaptive fuzzy control [6]. The turning off procedure results in a degradation of the network performance in terms of latency and packet loss due to the increased utilization of some of the links [7]. Performance of the WiNoC with respect to other emerging NoC architectures for a specific system configuration. In ongoing and future Investigations, we come to know the Methodology to compare and contrast the performance of WiNoCs with the other alternatives by varying system Size, packet length, traffic patterns, and other relevant Parameters. More efficient and scalable techniques will also be developed for optimizing the WiNoC architectures, specifically in presence of complex transient workloads [16]. The performance of link state protocol and the Spanning Tree protocol for a grid topology and a representative metro topology. The advantages and shortcomings of both protocols, we proposed a hybrid approach for switching frames in layer 2 in the Metro Area Network. The hybrid approach was able to reduce the congestion ratio and delay as compared to the link state and Spanning Tree [17].

III. PROPOSED SYSTEM

We propose Orion, a hybrid distributed hierarchical control plane for large scale networks. The proposed architecture combines the advantages of flat and centralized hierarchical control planes, and addresses the two unresolved problems discussed above. For disaster resiliency of the control plane, we propose to design it as a virtual network, which can be solved using Virtual Network Mapping techniques. We select appropriate mapping of the controllers over the physical network such that the connectivity among the controllers (controller-to-controller) and between the switches to the controllers (switch-to-controllers) is not compromised by physical infrastructure failures caused by disasters.

IV. EXISTING SYSTEM

The Flood light OpenFlow controller provides a rich set of components. The central controller in control packet is implemented based on the existing Floodlight controller by introducing the following extensions. Encap action realizes packet encapsulation in edge switches by extending the existing OpenFlow v1.0 protocol. In the architecture, packet forwarding in the data plane replies on a packet like encapsulation. When a rule with this action is applied to a flow, the switch will encapsulate the packets with a new header targeting a given remote IP address.

V. DESIGN

We outline the design of Orion, hybrid hierarchical control plane architecture of SDN focusing on the intra-domain control and management of large-scale networks. Throughout this paper, a domain is a complete network which can be controlled and managed by one administrator. It can be divided into several areas, which are regions that can each be controlled by a single SDN controller.

Control Plane Scale-out Architecture: In the scale-out architecture, the basic platform is implemented with generic processors augmented by separate communications processors with

specialized hardware accelerators that can offload control plane functions. The control plane tasks are divided into sub-tasks, such as discovery, dissemination, and recovery, and are then distributed across the data center. Because the various tasks can execute on any server in the network or in the cloud, the scale-out architecture lends itself well to Software Defined Networking (SDN). Owing to its distributed arrangement, the architecture requires robust communications between the control plane and the data planes using APIs for the network protocol, such as OpenFlow [1].

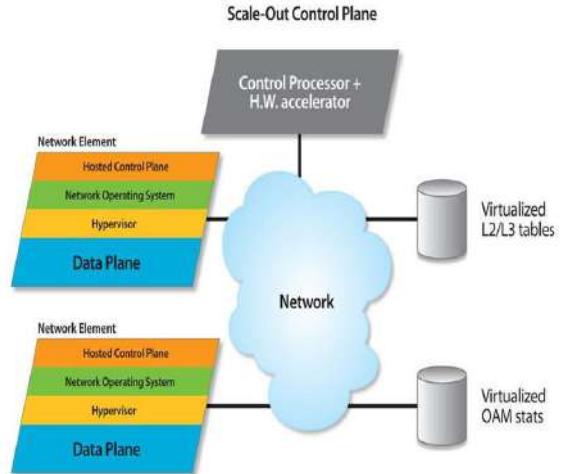


Fig 1: Architecture of Scalable Network

Protocol-aware communications processors are designed to handle specific control plane tasks and/or network management functions, including packet analysis and routing, security, ARP offload, IGMP messages, networking statistics, application-aware firewalling, QoS, etc. [1]

Control Plane Scale-up Architecture: In the scale-up architecture, the existing network control platforms are supplemented by additional and/or more powerful compute engines to help execute the network control stack. These supplemental resources free up server CPU cycles for other tasks, and result in an overall improvement in the network performance. Because general-purpose processors are not optimized for packet processing functions, however, they are not an ideal solution for the scale-up architecture. As with the scale-out architecture, performance can be improved dramatically using function-specific, protocol-aware communications processors [1].

VI. PERFORMANCE

The interconnect infrastructure is formed by connecting the cores in the subnets with each other and to the central hub through traditional metal wires. The hubs are then connected by wires and wireless links such that the second level of the network has the small-world property. The placement of the wireless links between a particular pair of source and destination hubs is important as this is responsible for establishing high speed, low-energy interconnects on the network, which will eventually result in performance gains.

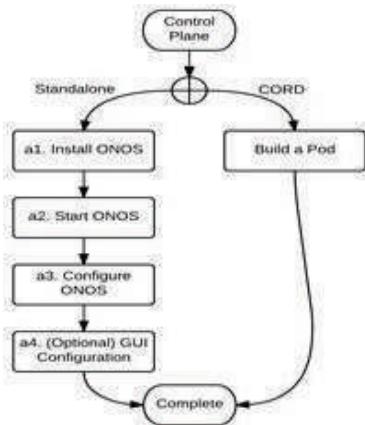


Fig 2: Scale Out Architecture

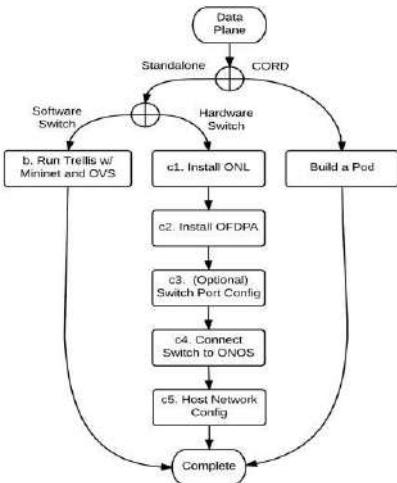


Fig 3: Scale In Architecture

Device Management Module: This module has two parts to deal with the area device information and domain device information.

- 1) **Area Device Management.** This sub-module obtains the host information through the ARP packet sent by the host. When a host sends an ARP packet, the switch that connects to the host sends a Packet-In message to the area controller. The sub-module decapsulates the PacketIn message acquires the host information and collects switch information.
- 2) **Domain Device Management.** In order to provide interarea host information, this module works as an ARP Proxy in Orion. In order to prevent a broadcast storm, we use an algorithm similar to Spanning Tree to avoid a broadcast loop. This sub-module also manages the global edge switches information.

VII. CONCLUSION

Therefore, in this paper, we have presented a novel architecture for A Scalable Hybrid Network Control Plane for Cloud Computing, which considers the concept of control and data plane separation. We formally model this disaster-aware control-plane

design and mapping problem, and demonstrate a significant reduction in the disruption of controller-to-controller and switch-to-controller communication channels using our approach. Further, we evaluate the effectiveness of Orion theoretically and experimentally. Our results show the efficiency and feasibility of Orion.

ACKNOWLEDGMENT

Sincere thanks to our guide Mr. Anand khadare who have guided to write this paper.Special thanks to PG coordinator Mrs. Harshali Patil for encouragement. Thanks to HOD PG Dr. R.R Sedumker for their support.

REFERENCES

- [1] <http://www.datacenterknowledge.com/archives/2012/08/30/virtualization-of-datacenters-new-options-in-the-control-and-data-planes>
- [2] <https://github.com/facebookarchive/scribe/wiki>.
- [3] www.wikipedia.com
- [4] www.google.com
- [5] N. McKeown et al., "OpenFlow: Enabling innovation in campus networks," ACM SIGCOMM Comput. Commun. Rev., vol. 38, no. 2, pp. 69–74, Apr. 2008.
- [6] Mehrdad Hojati and Saeed Gazor,"Hybrid Adaptive Fuzzy Identification and Control of Nonlinear Systems",IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 10, NO. 2, APRIL 2002
- [7] Anton N. Katov, Albena Mihovska, Neeli R. Prasad,"Hybrid SDN Architecture for Resource Consolidation in MPLS Networks",Aalborg University Aalborg, Denmark.
- [8] Steven Abramson, William Horka, Leonard Wisniewski,"A Hybrid Cloud Architecture for a Social Science Research Computing Data Center",1737 Cambridge St., Cambridge, MA 02138 USA"
- [9] Mohammad Al-Rabayah and Robert Malaney,"A New Scalable Hybrid Routing Protocol for VANETs",IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, VOL. 61, NO. 6, JULY 2012
- [10] Payman Samadi, Matteo Fiorani,Yiwen Shen,Lena Wosinska,"Flexible Architecture and Autonomous Control Plane for Metro-Scale Geographically Distributed Data Centers", "JOURNAL OF LIGHTWAVE TECHNOLOGY, VOL. 35, NO. 6, MARCH 15, 2017"
- [11] Kai Zheng, Lin Wang, Baohua Yang, Yi Sun, and Steve Uhlig,"LazyCtrl: A Scalable Hybrid Network Control Plane Design for Cloud Data Centers",IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. XX, NO. XX, SEPTEMBER XXXX".
- [12] D. H. Xu, Y. Z. Xiong, C. M. Qiao, and G. Z. Li, "Failure protection in layered networks with shared risk link groups", IEEE Netw, vol. 18, no. 3, pp. 36–41, May/Jun. 2004".
- [13] J. Sherry, S. Hasan, and C. Scott, "Making middleboxes someone else's problem: Network processing as a cloud service," in Proc. ACM SIGCOMM HotSDN, 2012, pp. 13–24.
- [14] Vaibhav Arora, Faisal Nawab, Divyakant Agrawal, Amr El Abbadi,"Janus: A Hybrid Scalable Multi-Representation".
- [15] CloudDatastore","DOI 0.1109/TKDE.2017.2773607, IEEE".

- [16] Amilan Ganguly, Kevin Chang, Sujay Deb, Partha Pratim Pande, Benjamin Belzer, and Christof Teuscher,” Scalable Hybrid Wireless Network-on-Chip Architectures for Multicore Systems”, “IEEE TRANSACTIONS ON COMPUTERS, VOL. 60, NO. 10, OCTOBER 2011”.
- [17] Minh Huynh and Prasant Mohapatra,” A Scalable Hybrid Approach to Switching in Metro Ethernet Networks”.

Optimizing AODV Protocol for Preventing Black Hole Attach on MANET

Saurabh Bagde

Computer Engineering

Thakur College of Engineering and Technology

Mumbai, India

drew.is4u@gmail.com

Abstract—A mobile ad-hoc network has multiple mobile network compatible devices connected to the network in a dynamic and a temporary method in which they can connect without any access point or a predesigned infrastructure. Firewalls and encryption software's protects the network from attacks. But many of them are not sufficient and efficient due to its restricted power and mobility. Mobile networks are more prone to attacks from malicious nodes in the network than the wired networks thus it needs more security. As wireless networks has more threats, the defense mechanism for it should be more intensive. Therefore, an effective system which can detect such attacks is important so it can avoid such attacks by detecting&isolating the problem created by such nodes and alert the system about such malicious node to other nodes as well. We propose a system which focus on analyzing and improving the security of AODV, which is one of the popular, routing protocols for MANET. Our system should ensure the security against BH attack. The proposed solution detects and removes BH nodes from MANET.

low cost path between the nodes and forward packets to the designated node in the network. As wireless ad-hoc networks don't have any get admission to point or base station so they're uncovered to various attacks, one such assaults are the Black hole attack. Inside the Black hole attack, a malicious node does now not permit packets to be forwarded to subsequent node. In one of these manner, packets are dropped within the community. The B.H. node takes advantage of the short comings present in direction discovery packets of the on-call for protocols, such as AODV. For link-ups connection establishment process by path discovery method of protocol, the nodes between the source and the sink nodes are susceptible to find a fresh path to the sink node from source node, sending discovery packets to the neighbour nodes[2]. Malicious nodes don't follow this process and alternatively, they without delay respond to the source node with solid records as although it has fresh enough course to the destination machines. therefore, supply node sends its data frame packets via the malicious node to the destination assuming it's far a real direction. B.H. attack may arise because of a malicious nodes which are disrupting the normal behaviour deliberately, as well as a broken node interface. however, nodes inside the network will try to discover a route for the receiving machine constantly, which makes the node devour its battery in addition to losing packets.

Keywords—Mobile Ad-Hoc Network, routing protocol, Black Hole Attack, AODV, MANET, Packet Delivery Ratio (PDR).

I. INTRODUCTION

Wireless ad-hoc net have self-dependant nodes that are self-managed. They do not need any predefined infrastructure. Ad-hoc mesh have a non-static topology such that client can easily connect or leave the network at any time. They have many potential applications, like, in armed services and rescue areas such as providing channels to soldiers on the field of battle to communicate or establishing a new network in a natural disaster where the existing network is collapsed like an earthquake. Ad-hoc networks are suitable for places where it is not easy to exercise set up a fixed infrastructure. Since the node converse with each other without a stable and predefined infra-structure, they provide the connectivity by forwarding data frames over themselves. To establish network link-ups or an uptime session on network, nodes uses routing protocols such as Ad-hoc On-Demand Length Vector (AODV), Dynamic Source Routing (DSR) and Destination-Sequenced Distance-Vector (DSDV) [1]. In such networks which are formed to achieve MANET, every node has also to play the role of a router as well to find asuitable and

Ad Hoc On-demand Vector Routing (AODV) protocol is a reactive routing protocol for MANET that creates routes dynamically among nodes which need to talk. The AODV routing protocol is derived from the DSDV set of rules. AODV is an development on DSDV because it commonly by means of creates routes on an on-call for foundation as a result minimizing the number of required proclaims, as opposed to maintaining a whole list of routes as within the DSDV set of rules[1]. The authors of AODV classify it as a natural on-call for route acquisition machine, as nodes that are not on a particular direction do not maintain routing facts. that means, the routing messages do now not include statistics about the complete course route, but simplest approximately the supply and the destination. therefore, routing messages do now not have an growing size. It makes use of destination series numbers to specify how sparkling a direction is (with regards to every other), which is used to grant loop freedom[1].

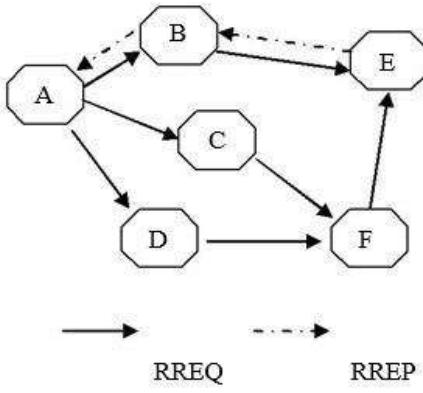


Fig 1: RREQ & RREP message exchange between A & E

on every occasion a node needs to ship a packet to a destination for which it has no „clean enough? direction (i.e., a valid path entry for the vacation spot whose related series range is at the least as excellent as the ones contained in any RREQ that the node has obtained for that vacation spot) it announces a direction request (RREQ) message to other member nodes in the network. each node that gets the broadcast units up a reverse direction in the direction of the originator of the RREQ(unless it has a latest 1). Whilst the supposed destination (or an intermediate node that has a „fresh enough? course to the vacation spot) receives the RREQ, it replies by using sending a course respond (RREP). it is essential to notice that the handiest mutable information in a RREQ and in a RREP is the hop rely (that's being monotonically expanded at each hop). The Route-Reply travels returned to the originator of the Route -Request (as a unicast). At every intermediate node, a path to the destination is set (again, until the node has a „more energizing? path than the one designated in the RREP). within the case that the RREQ is spoke back to by an intermediate node (and if the RREQ had set this selection), the intermediate node additionally sends a RREP to the destination. in this manner, it is able to be granted that the path route is being set up bi-directionally[3]. inside the case that a node gets a new direction (with the aid of a RREQ or with the aid of a RREP) and the node already has a path „as fresh? because the received one, the shortest one may be up dated. The supply node begins routing the information packet to the destination node via the neighboring node that first answered with an RREP. The AODV protocol is at risk of the famous black hollow attack. 1.

II. PROBLEM STATEMENT

An ad-Hoc network is self-organizing and self-configuring multi-hops wireless and decentralizes networks, where the country of the shape of the network modifications dynamically. this is specifically cause of the mobility of the nodes inside the community; nodes in the networks always strive to make use of the same random access wireless channel of the network, cooperating in an intimate way, so as to disguise themselves as clean part of the network, & be within the multi-hop forwarding. The cellular nodes within the community best does not acts as a hosts, but additionally as routers that route facts to from the others nodes in network. In cell ad-hoc networks there's no need of pre-existing infrastructure support as wi-fi networks, and on the

grounds that a vacation spot node might be out of range of a supply node to moving records packets; in order that there's need of a routing manner. that is always geared up to discover a path on the way to ahead the facts packets accurately between the source node and the destination node. inside a mobile, a base station can attain to all cellular nodes without the usage of a routing thru broadcast in not unusual situation of the networks. In ad-Hoc community every device or machine must be able to forward the frames & packets for other members in network. This continually creates extra troubles due to dynamic nature of topology that is unpredictable connectivity adjustments problems together with the nature of dynamic topology.

III. PROPOSED SOLUTION

The detection of black hole assault will paintings on specific stages. Packet shipping ratio test on destination node. on this step to start with we are able to set up the nodes within the network and make a community. The supply will start the verbal exchange and ship the course request packets to all neighbouring nodes and after receiving direction respond packets to all nodes send the statistics packets to all nodes, but after a while whilst destination node launch that the packets comes from source node very less. Than test the edge restriction and in step with this the threshold restrict is under 10-20 packets. the idea of this suspense the vacation spot node take a look at or we can say calculate the packet shipping ratio and attempt to attain the final end result. This packet shipping ratio tests if the whole packets rely less than 20 than the sink node check the packet transport ratio. The take a look at packet transport ratio we have the formulation that destination node use overall packets ship by the vacation spot via acquired through the sink node and we are able to discover the packet delivery. The check packet transport ratio we use the probabilities. The possibility checks by the vacation spot node on the idea of two-time slots. The reason in the back of that is the destination node examine two-time slot for come across the malicious node that t1 and t2. at the time of t1 first destination node take a look at what number of packets are effectively acquired out of overall packets. let's 2 out of 10 data packets are receive at time t1. the primary situation of possibility applies here. The destination node suspects about the malicious node on the basis of various parameter it is passed through and some inconsistencies ae found . This doubt clearance the destination node waits for t2 time slot. The time t2 once more start to test the a hit packets and behaviour of malicious node[3]. The time t2 the successful packets acquired four out of 10 packets but other packets are not acquired to vacation spot node. those analyses confirm the vacation spot node about the malicious node and apply the second conditions of opportunity right here. The task for packet supply ratio performs on the idea of those chances..

1. Supply node sends the route request messages to all of the nodes and gets respond and start communique.
2. Destination node receives packets less compare than a threshold limit and start check the packets counts.
3. destination node performs the feature of packet shipping ratio (pdr) with the total range of counts of supply node.
4. Now destination node reveals the pdr on the premise of opportunity or percent of black hollow node or not in community with two conditions. The vacation spot nodes assessments opportunity for exceptional two-time slots and on that basis the vacation spot node checks the

- step of pdr.
5. The source node checks the forward packets ratio of every node which send their message forward for the path with which node not send the forward message and calculates the results.
 6. Than verification the sequence number high of nodes. Any node has sequence number high than suspect but not sure about black hole.
 7. Now the source node sends the dummy packet with the message for new session creation with the two prime products of two nodes and asks to send that prime number.
 8. After receive the prime number of suspect node than try to divide that with the prime product of two nodes.
 9. If divide than genuine otherwise it will be malicious node.

IV. XPERIMENTAL RESULTS

In this model, a very simple and powerful way of imparting safety towards black hole assault by way of introducing some modifications to AODV is delivered. for this reason, with the aid of using AODV (with relied on nodes) as a routing algorithm in MANET, one can also make certain of not being liable to black hollow assaults. The device detects the malicious nodes and isolates it from the records forwarding and routing. We also infer that a greater particular research and with some add-on functions to AODV, will prevent the alternative safety threats can also be detected and to a sure extent save you just like the manner we've got proposed for the black hole assault. There might be an boom in throughput with the proposed method over conventional AODV algorithm. Time taken to discover and keep away from the malicious node might be minimum

The fig.3 shows the effect to the packet transport ratio (PDR) measured for the AODV protocol whilst the node mobility is improved. The result indicates both the instances, with the black hole assault and with out the black hole attack. it is observed in the experiments results after comparing & measuring that the packet shipping ratio dramatically decreases while there may be a malicious node inside the community[6]. for example, the packet delivery ratio is a hundred% while there may be no impact of Black hollow assault and whilst the node is shifting at the rate 10 m/s. but due to impact of the Black hole attack the packet shipping ratio decreases to eighty two %, because a number of the packets are dropped by means of the black hole node

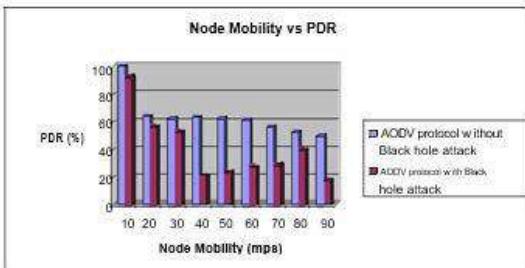


Fig 2: Impact of Black hole attack on PDR

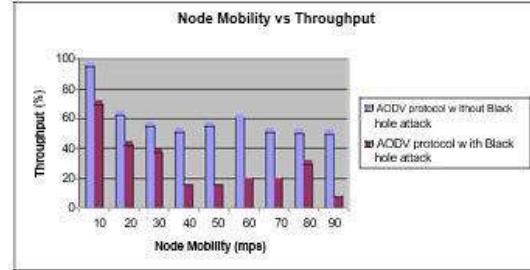


Fig 3: Impact of Black hole attack on Network Throughput

V. CONCLUSION

This paper discusses routing safety troubles of MANETs, the black hole assault that can be launched in opposition to a MANET and optimized AODV protocol as proposed solution for avoiding B.H. to assault. The proposed solution may be implemented to a) discover single and more than one black hollow nodes cooperating with each other in a MANET; and b) reveals at ease course from supply to destination via detecting and casting off black hollow nodes present inside the community. additionally, we confirmed that the impact of packet delivery ratio and Throughput has been detected with respect to the variable node mobility. there's discount in Packet shipping Ratio and Throughput. In Black hollow attack all network traffics are redirected to a particular node or from the malicious node inflicting extreme damage to networks and nodes as proven in the end result of the simulation. The detection of Black holes in advert hoc networks remains taken into consideration to be a difficult undertaking.

We simulated the Black hole assault inside the advert-hoc Networks and optimized the AODV protocol to mitigate the results of BH nodes with the aid of detecting and blockading the BH nodes from the advert-hoc network. We can also simulate other routing protocols and repurpose the equal algorithm used for AODV optimization for different MANET protocols. All routing protocols are anticipated to give unique consequences. For destiny work different routing protocols can be optimized and as each protocol provides extraordinary results, the fine may be decided on to avoid Black hole assaults.

Acknowledgements

The authors are grateful to the Head of the Department of PG and Dean of Academics, Dr. R. R. Sedamkar, of Thakur College of Engineering & Technology, Mumbai University, for providing the facilities to accomplish the present research work.

REFERENCES

- [1] Nisha P Joh, Ashly Thomas, "Prevention and Detection of Black Hole Attack in AODV based Mobile Ad-hoc Networks - A Review," International Journal of Scientific and Research Publications.
- [2] Meenakshi Patel, and Sanjay Sharma "Detection of Malicious Attack in MANET A Behavioral Approach" 2013 3rd IEEE International Advance Computing Conference (IACC)
- [3] Hizbulah Khattak and Nizamuddin, Fahad Khurshid," Preventing Black and Gray Hole Attacks in AODV using

- Optimal Path Routing and Hash" 2013 IEEE.
- [4] Seryvuth Tan, Keecheon Kim," Secure Route Discovery for Preventing Black Hole Attacks on AODV-based MANETs" in 7/2013 IEEE.
 - [5] Anand A. Aware, Kiran Bhandari "Prevention of Black hole Attack on AODV in MANET using hash function," 978-1-4799-6896-1/14/\$31.00 ©2014 IEEE.
 - [6] V. Keerthika, N. Malarvizhi, "Mitigate Black Hole attack using trust with AODV in MANET," 470978-9-3805-4421-2/16/\$31.00 c 2016 IEEE.
 - [7] Anand A. Aware, Kiran Bhandari , "Prevention of Black hole Attack on AODV in MANET using hash function," 978-1-4799-6896-1/14/\$31.00 ©2014 IEEE.
 - [8] V. Keerthika, N. Malarvizhi, "Mitigate Black Hole attack using trust with aodv in MANET," 470978-9-3805-4421-2/16/\$31.00 c 2016 IEEE.
 - [9] <https://courses.edx.org>

DESIGN OF SECURE ROBUSTNESS SPEAKER AUTHENTICATION SYSTEM

Amit Tiwari

Computer Engineering

Thakur College of Engineering and Technology

Mumbai, India

amitt2238@gmail.com

Abstract — With the use of biometric features for better security got more popular over a short period of time, challenges occurs for biometric authentication. The proposed system works on the problem of speaker identification and verification under noisy conditions where information about noise characteristics is not available. Speaker identification for authentication allows the user to identify the users by using their voices. A Mel Frequency Cepstrum co-Efficient (MFCC) based feature extraction method is used for human auditory filtering to identify different energy levels, change in frequencies etc. System in real time works on data sample capture between silent periods. Data samples are further processed with MFCC to match with the database entry pre-stored. The results are compared based on the accuracy, the speed and the number of MFCC.

Keywords—MFCC, Speaker recognition , Feature extraction , VQ

I. INTRODUCTION

The speaker recognition is a biometric approach that uses a person's voice for identification purposes. It depends on characteristics that are affected by the physical structure of the user's vocal tract and its behavioral characteristics. The speaker identification is different from other biometric methods in such a way that the voice samples are changing continuously for a short period of time (normally few seconds).

There are two major types of speaker recognition: a text dependent and a text independent .Text dependent is called as "Constrained" mode whereas Text Independent is known as "Unconstrained" mode. Since, there is no any advance information of the speaker's speech is available for Text Independent. It shows more difficult challenge than the text dependent systems whereas, a "text dependent" voice recognition system requires the person to speak a fixed phrase.

A. MFCC Process

MFCC is a technique based on human listening behavior that cannot recognize frequencies over KHz. MFCC depends on the difference of frequencies that can be distinguish by human ear. The signal is expressed in the MEL scale, this

scale is based on the perception of the pitches in an equally spaced intervals judged by observers. This scale uses a filter that is placed linearly at frequencies below 1KHz and above 1KHz for logarithmic spacing.

1.1 Pre-emphasis

In this process we emphasize the higher frequencies; this will enhance the energy in the signal at frequencies over 1KHz.Pre-emphasis method is used to increase signal Quality of a data transmission at the output.

B. Removal of noise

Having performed the normalization procedure, the magnitudes corresponding to the low frequencies affect the accuracy of the identification system. They can be considered as noise that needs to be eliminated or reduced. Therefore, all frequencies which are below a threshold value are eliminated i.e. their corresponding magnitudes are set to a value of zero.

C. Framing

Is the segmentation of the speech samples in boxes within the range of 20 ms to 40 ms. the voice signal is divide in terms of frames of N samples. Corresponding frames are separated by M where M is less than sample N. Different values used for M is M = 100 and N= 256.

D. Hamming windowing:

In signal processing, a window is used when a signal we are interested has a limited length. Indeed, a real signal has to be finite in time; in addition, a calculation is only possible from a finite number of points. To observe a signal in a finite time, we multiply it by a window function.

II. METHODOLOGY

This section describes the two main modules of a speaker recognition system which provides the various modules and some of the methods is used in this project work.

1. Feature Extraction: Feature Extraction converts the speech waveforms which are analog in nature into a set of features. For further analysis Feature vectors are used.

2. Feature Matching: In feature matching process the features extracted from the input voice are matched with the pre stored database and an identification decision is made.

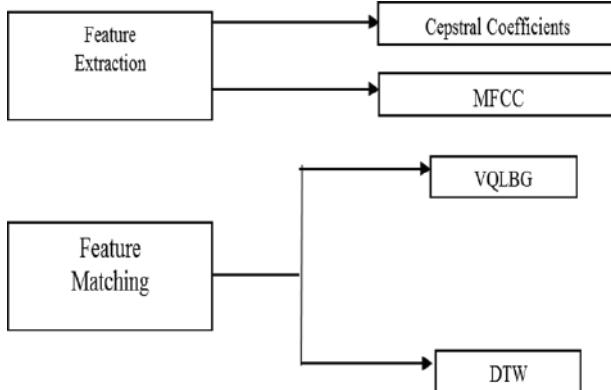


Fig 1: Modules of a speaker recognition system

A. Signal acquisition:

To record and convert the user's voice into digital form the Microphones and A/D converter are used. A vector representation of input signal is available at the end of this process. The duration of speech recording depends on the desired accuracy.

B. Speech signal pre-processing:

The speech signal $x(n)$ is not a static signal since the vocal tract is continuously changing and the model parameters are time-varying. But, it is generally admitted that these parameters are constant over sufficiently small time intervals. Generally, the signal is divided into frames of 20-25 ms denoted by $x_i(n)$. Due to this division into frames it tends to discontinuities in the temporal domain. To avoid this phenomenon, a Hamming window is used.

C. Feature extraction:

Based on the speech signal and pre-processing, features are extracted to design a model corresponding to the user. Generally, these extracted features must be robust to the user's voice, to noise and distortion. The methods consists of short-term spectrum features. We have chosen to extract MFCC (Mel-frequency cepstral coefficients) introduced by which reveal to be more robust and efficient in practice.

D. Speaker modelling:

After the completion of Feature extraction on each frame, the corresponding model or design requires a training phase. Again, we choose the VQ (vector quantization) method. It is based on the LBG (Linde Buzo and Gray) algorithm. This process allows, after clustering, to describe a voice sample by a model vector having a predefined fixed size, whatever the initial length of the signal.

E. Speaker recognition:

The above four steps represents the user training phase. For recognition or testing step we consider user authentication: in this process it is get identified that whether the user is owner or not. For VQ method, the identification part is generally performed through Euclidean distance computation between the reference sound sample and the new captured sample.

III. RELATED WORK

In the field of speaker recognition, researchers have worked upon few challenges on the various models of speaker identification. In [1], authors have used Short Term Fourier Transform (STFT) but the accuracy of identification was not achieved due to large dataset.

In [2], the author has applied first extracts MFCC features from the raw signal and then creates Information Set Features (ISF) by applying a fuzzy logic approach. ISF features reduce the size of the MFCC features and computes templates composed of only 12 floating-point numbers.

In [3], the authors have summarize the operation flow of the whole identification system and implement the speaker identification system with python.

SIS is currently in use with highly sensitive information such as forensic science, bank account transactions, medical applications, and password protection [4].

In [5], the algorithm used provides efficiency up to 40% at low Signal to Noise ratio (7dB) with very low computation/processing time.

The fusion feature vector of cepstrum with MFCC and formants gives good accuracy for emotion recognition. For speech recognition, the traditional classifiers used are SVM, HMM, kNN etc. But the simplest and the most effective with less complexity and great accuracy is SVM. [6].

IV. PROBLEM STATEMENT

In the case of speaker recognition the need is for more discriminative features between any two speakers rather than higher individual information on each vector present. Though the Cepstral coefficients perform quite well for speaker identification, its performance can further be improving by using the hidden features of the speech signal.

V . PROPOSED SYSTEM

Proposed system presents an effective method for speaker identification system. A Mel-Frequency Cepstrum Co-efficient (MFCC) based feature extraction method is used for human auditory filtering to identify different energy levels, change in frequencies etc. The results are compared based on the accuracy, the speed and the number of MFCC.

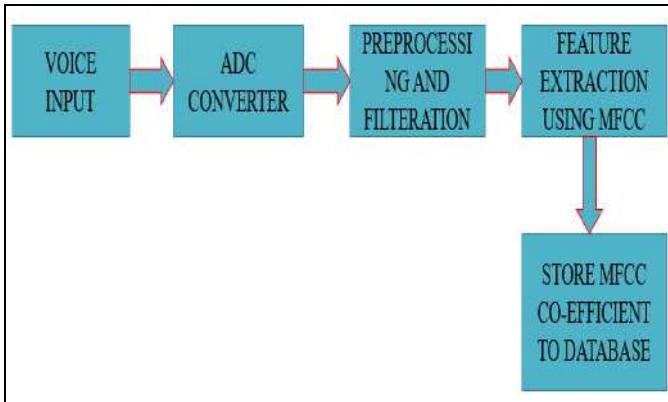


Fig 2: Training Set

In the Training Set the input to the set will be voice of user. After getting input voice ADC Converter is used to convert the analog signals coming from user voice to digital. Next, the Filteration and further preprocessing has been done. Feature Extraction can be done by using MFCC and co-efficient of MFCC is get stored in the Database.

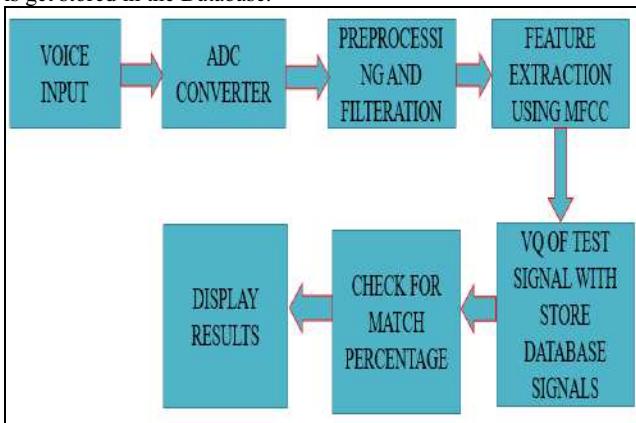


Fig 3: Testing (Recognition)

Now, in the testing or Recognition part first four steps are same as the training session. After extraction of features using MFCC the Vector Quantization of test signal can be performed from the database signals. Furthermore, calculation of match percentage is performed and result will get displayed on the screen whether the recognition has been done successfully or not.

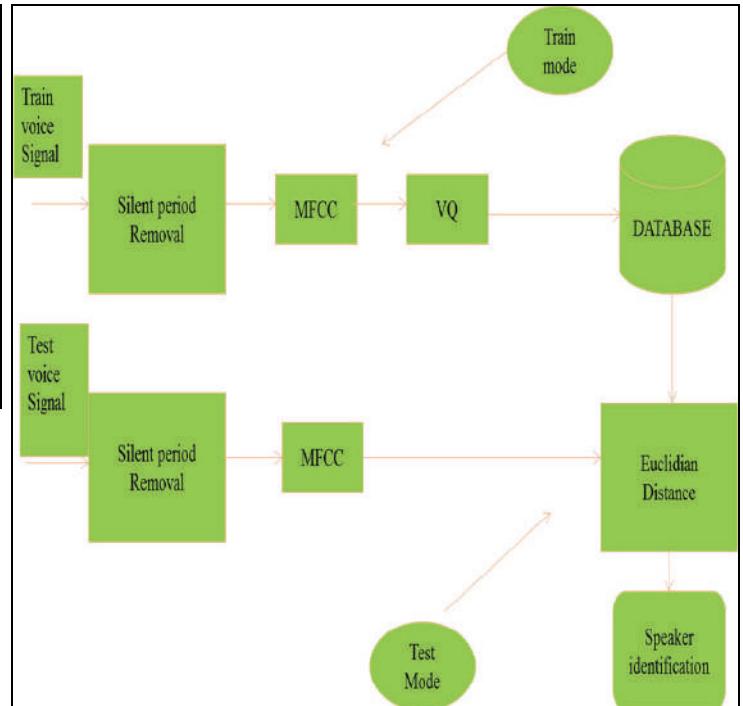


Figure 4. System model

VI. EXPERIMENTAL RESULTS

The database formation will consist of distinct speakers including both male and female speakers. It will also contain multiple sound files used for training and testing the Speaker Recognition module. New sound will be stored in real time for recognizing the Continuous Speech signal module under clean and noisy environments. Recognition rate of the trained VQ model is given by the formula:

$$RR = \frac{N_{correct}}{N_{total}} \times 100$$

In the above equation RR is the recognition rate, Ncorrect shows the number of correct (accurate) recognition of testing speech samples per digit, and the total number of speech samples per digit is represented by Ntotal

ACKNOWLEDGEMENT

We sincerely thank our guide Dr. Raghavendra R. Sedamkar for his guidance and constant support and teachings to better ourselves and the society.

REFERENCES

- [1] Dany Ishac, Antoine Abche“A Text-dependent Speaker-Recognition System”, Instrumentation and Measurement , 978-1-5090-3596-0/17, 2017 IEEE
- [2] Ruggero Donida Labati, Madasu Hanmandlu,” Text-Independent Speaker Recognition for Ambient Intelligence Applications by Using Information Set Features” Department of Computer Science and Engineering, 2017 IEEE
- [3] Fang-Yie Leu, Guan-Liang Lin,”An MFCC-based Speaker Identification System”, Computation and Language, Advanced Information Networking and Applications, 1550-445X/17, 2017 IEEE
- [4] Danoush Hosseinzadeh and Sridhar Krishnan,” Feature Fusion Techniques Based Training MLP for Speaker Identification System”, Proceedings of COLING 2014, IEEE 31st International Conference on Advanced Information Networking and Applications, pages 2335–2344, August 23-29 2014.
- [5] Kailash Shaw, Prajakta P. Dahake” Speaker Dependent Speech Emotion Recognition using MFCC and Support Vector Machine”, Automatic Control and Dynamic Optimization Techniques (ICACDOT), December 2016.
- [6] A.S.bhalerao,”Implementation of Automatic Speaker Recognition on TMS320C6713 Using MFCC”, Computer Communication and Informatics, December 2013.
- [7] Reza Aulia Sadewa, Tokorda Agung Budi Wirayuda,” Speaker Recognition Implementation for Authentication Using Filtered MFCC – VQ and a Thresholding Method”, Information and Communication Technology, December 2013.
- [8] Dr. Syed Akhtar Imam, Roma Bharti” Speaker Recognition using MFCC, shifted MFCC with Vector Quantization and Fuzzy”, Soft Computing Techniques and Implementations, November 2012.
- [9] <https://courses.edx.org>

Data Mining for Internet of Things

Deepali Joshi

Thakur College of Engineering & Technology

deepali88.dj@gmail.com

Abstract- Large amounts of data is generated by IOT. The tremendous increase in data and its collection in various fields compels us to do analysis and identify required knowledge. Data mining means identifying information from huge databases. In mining tremendous data is involved, and due to this computing with good performance is necessary for successful mining application. Computing in distributed way can be useful mechanism for Data Mining. In data mining the centralized mechanism is not beneficial since data is available on various databases and due to this the combining data mining, distributed computing will yield good results.

Keywords- IOT, Data Mining, Data Mining, Distributed

I. INTRODUCTION

The IOT consists of various devices, machines, objects, human beings that are related and will send data through the network without human being-to-human being or human being-to-computer interaction. IOT is a revolution of technology since the revolution of computer and Internet. The computers, the internet depends on human-beings for data and information. Around 50% of data that is on internet were captured and created by humans by entering the data through keyboard, taking a picture, scanning the data or by drawing the diagrams.

The problem that is faced today by the people is of limited time, and due to which it is difficult to capture data. If there are computers that can understand everything that should be known about things and using the data that is gathered, it would be beneficial. It will be predicted when replacement is needed or recall.

II. LITERATURE SURVEY

Data is an essential fact and is required in all fields. In order to retrieve the essential information mining can be used. IOT is connection of various components and these components need data in order to complete execution. The devices will gather and then process the data. IOT and mining will go hand in hand to carry out the operations.



Fig1.1: IOT

III. ARCHITECTURE

A. Architecture of Data Mining

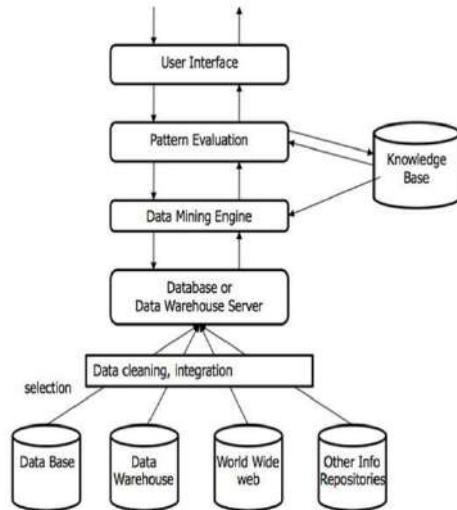


Fig1.2: Data Mining Architecture

- 1) *Databases, Data warehouses, www, various repository's containing information*

The data is available in the repositories. The operation such as cleaning, integration can be performed on the data.

- 2) *Database/ Data server*

The main function is to fetch required data, based on users request.

- 3) *Mining Engine*

It consists of various modules like classification, prediction, etc.

- 4) *Knowledge base*

This is knowledge which will guide or evaluate various patterns. This information is utilized to differentiate attributes in different hierarchies.

- 5) *Pattern Evaluation*

It communicates with system so that interesting patterns can be searched. It will be used with data mining module.

- 6) *User Interface*

It provides interface between and system and allows the communication through query or task. It will allow the user to search database, warehouse's structures and will display the patterns in various forms. [9]

B. IoT Architecture

It consists of five-layers. The various layers are:[1]

- The perception layer has sensors that will do sensing, gathering information. It senses objects.
- The transport layer will transfer the data to and from sensors through perception to processing
- The processing layer can also be called as middleware. It will store, analyze and will processes data from layer known as transport layer. It manages and

provides different services to the lower layers. It works on various technology's like databases, cloud computing, big data modules.

- The application layer will deliver service for specific application. It will specify applications of IoT
- The business layer will manage the entire IoT system, along with applications, business, profit models, users' privacy.

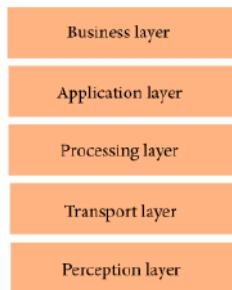


Fig1.3: IoT Architecture

IV. MINING FUNCTIONALITIES

The functionalities are classification, clustering, association, time series, and outlier. [9]

- Classification technique identifies models which describe, differentiate class, to predict unknown label. Example: Three classes sunny, windy, and rainy and we have to find to which class data will belong to.
- Clustering technique will analyze data based on similarity without a class model which is available. Example: The example which is given below will show the clustering technique. There are 10 objects of three different colors. Here the cluster of three different colors is to be created.



The objects of similar color will be grouped as specified:



- Association analysis will identify the association rules displaying attribute-value conditions that frequently occur in set of data. Example: buys(X, "computer") => buys(X, "software), X is representing customer
- Time series analysis consists of various methods, techniques for analyzing data to identify statistics, characteristics.

A. IoT and Clustering

Clustering means identifying data based on specific features, values. It is unsupervised learning because in other techniques such as data classification, we need to 'train' the system with data. Clustering is applicable on a data set without knowing much about it in advance. The clusters number is given as an input. Data clustering maybe not be used directly in IoT applications, but in various cases it will be an intermediate step for identifying patterns from the collected data.[2]

B. IoT and data classification

Classification is required for decision making. For an object, it will be assigned to predefined classes and this is known as classification. Classification is used when the data is associated to a different classes. It is not prediction. It is categorisation of new values. In the activity tracking example, the device you carry on senses motion using accelerometers and tilt sensors. It's a technique used by the device vendor to associate the sensor values with steps, stairs climbing. After clustering the steps and specifying into low, high or medium activity, the application can use again data classification that will determine overall value. [2]

C. Association Analysis

Association rule will focus on market basket method, and it will discover various rules for attribute value. Associations which occur many times and will help to generate knowledge which will aid in decision making.

D. IoT and Time series prediction

As the name it indicates an estimation of future data based on dataset that is already collected and analysed. The application used widely is meteorological and weather forecasting. [2]

V. CONCLUSIONS

The concept has arisen for managing, automating, exploring various devices. To take appropriate decisions, mining can be used with IoT. Mining helps in discovering required patterns apply algorithms to find hidden information. The mining is viewed in three different views: knowledge, technique, and application. In view known as knowledge, we review classification, clustering, association, time series, and outlier. In application view, mining application are viewed. The technique view is discussed with knowledge and application view. We need to understand various challenges.

REFERENCES

- [1] Research on Data Mining Models for the Internet of Things Shen Bin#, Liu Yuan*, Wang Xiaoyi*
- [2] Data Mining for Internet of Things: A Survey Chun-Wei Tsai, Chin-Feng Lai, Ming-Chao Chiang, and Laurence T. Yang, IEEE COMMUNICATIONS SURVEYS & TUTORIALS, VOL. 16, NO. 1, FIRST QUARTER 2014
- [3] A Research Direction on Data Mining with IOT, Foram Chovatiya, Purvi Prajapati, Jalpesh Vasa , Jay Patel, International Conference on Information and Communication Technology for Intelligent Systems 2017
- [4] Research on Data Mining Model in the Internet of Things Yan Chen, Ai-xia Han Cai-hua Zhang, International Conference on Automation, Mechanical Control and Computational Engineering (AMCCE 2015)
- [5] Data Mining for the Internet of Things: Literature Review and Challenges Feng Chen,^{1,2} Pan Deng,¹ Jiafu Wan,³ Daqiang Zhang,⁴ Athanasios V. Vasilakos,⁵ and Xiaohui Rong⁶, Hindawi Publishing Corporation International Journal of Distributed Sensor Networks Volume 2015, Article ID 431047, 14 pages
- [6] M. C. Domingo, "An overview of the internet of things for people with disabilities," Journal of Network and Computer Applications, vol. 35, no. 2, pp. 584–596, 2012.
- [7] G. Kortuem, F. Kawsar, V. Sundramoorthy, and D. Fitton, "Smart objects as building blocks for the internet of things," IEEE Internet Computing, vol. 14, no. 1, pp. 44–51, 2010.
- [8] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," Computer Networks, vol. 54, no. 15, pp. 2787–2805, 2010
- [9] H. Jiawei and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2011

Scalable Data Reduction Technique in Cloud Storage

Ankur Sinha, Anand Khandare

Thakur College of Engineering and Technology, Mumbai, India

Abstract—Data deduplication has gained increasing popularity as a space-reduction approach in cloud storage systems. One of the main challenges for data deduplication is the scalability of fingerprint-index search. In existing system, deduplication mainly focuses on backup system. In this paper, we propose a system that effectively exploits similarity and locality of data blocks to achieve high duplicate elimination and well balanced load at low RAM overhead in a real time environment. The main idea is to employ an algorithm by considering data chunks which are similar and eliminate duplicate chunks from the storage system, and then further enhance the resemblance detection efficiency based on frequency of their occurrences. This would further help us to enhance the data delivery capabilities in cloud storage system with reduced latency and high throughput using clustering method to store the data. We also propose a lightweight approach using the index & frequency information to further improve the data retrieval in a client-server architecture model, which is independent of file type and deliver higher deduplication ratio.

Keywords—component; Data Chunks, Data Deduplication, Cloud Storage

I. INTRODUCTION

Cloud storages provides a low-cost, scalable, location-independent infrastructure for data management and storage. The increasing data volume has led more and more people to pay attention to use the capacity of cloud storage than before. Cloud Storages provides many virtualized resources to users, as services across the entire Internet, while hiding platform and implementation details. GMAIL is one of the best examples of cloud storage which is used by most of us regularly. Cloud service providers provide highly available cloud database storage and slightly parallel computing resources at a comparatively low cost. One difficult challenge of cloud storage services is the management of the ever-increasing amount of data. To control data management in cloud computing, deduplication has been a well-known technique and has attracted more and mostly technology used in today's world. Data deduplication is removal of redundant data. Thus, reducing the amount of data in actual physical storage reduces a lot of storage requirements costs, infrastructure management costs and power consumptions as well.

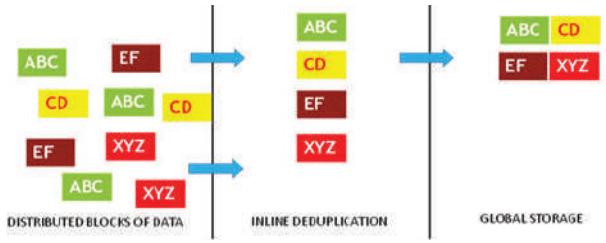


Fig. 1.1. Overview of data Reduction Technique

Though various deduplication techniques have been proposed and used, no single best solution has been developed to handle all types of redundancies. Considering performance and overhead, each deduplication technique has been developed with different designs considering the characteristics of data sets, system capacity and deduplication time. For example, if the data sets to be handled have many duplicate files, deduplication can compare files themselves without looking at the file content for faster running time. However, if data sets have similar files rather than identical files, deduplication should look inside the files to check what parts of the contents are the same as previously saved data for better storage space savings. Also, deduplication should consider different designs of system capacity. High-capacity servers can handle considerable overhead for deduplication, but low capacity clients should have lightweight deduplication designs for fast performance.

Deduplication can be divided based on granularity, deduplication place, and deduplication time. The main components of these three classification criteria are chunking, hashing and indexing. Chunking is a process that generates the unit of compared data, called a chunk. To compare duplicate chunks, hash keys of chunks are computed and compared, and a hash key is saved as an index for future comparison with other chunks.

Table I. Methods of Data Deduplication

Method based on granularity	Place	Time
File Level deduplication	Server	Inline
Fixed-size block level deduplication	Client	Offline
Variable-size block level deduplication	Network	

Deduplication is classified based on granularity. The unit of compared data can be at the file level or sub file level, which are further subdivided into fixed-size blocks, variable-sized chunks, packet payload or byte streams in a packet payload. The smaller the granularity used, the larger number of indexes created, but the more redundant data are detected and removed.

For place of deduplication, deduplication is divided into server-based and client-based deduplication for end-to-end systems. Server-based deduplication traditionally runs on high-capacity

servers, whereas client-based deduplication runs on clients that normally have limited capacity.

In terms of deduplication time, deduplication is divided into inline and offline deduplication. With inline deduplication, deduplication is performed before data are stored on disks, whereas offline deduplication involving performing deduplication after data are stored. Thus, inline deduplication does not require extra storage space but incurs latency overhead within a write path. Conversely, offline deduplication does not have latency overhead but requires extra storage space and more disk bandwidth because data saved in temporary storage are loaded for deduplication and deduplicated chunks are saved again to more permanent storage. Inline deduplication mainly focuses on latency-sensitive primary workloads, whereas offline deduplication concentrates on throughput-sensitive secondary workloads. Thus, inline deduplication studies tend to show trade-offs between storage space savings and fast running time.

II. PROBLEM DEFINITION

1. The existing system is limited to file level deduplication in real time scenarios, thus, leading to limitation in application level framework.
2. We propose a system to achieve data deduplication on block-level storage to achieve high performance in cloud storage with reduced overhead in real time system.
3. The proposed system will deliver efficient data access with reduced cost in terms of bandwidth usage in distributed environment.
4. The existing system is limited to de-duplicate only similar file format.
5. The scalability of storage nodes is difficult, which can be overcome using clustering method.

III. SCOPE OF PROJECT

The project is currently limited for analysis purpose for larger set of data with limited CPU and RAM. The clustering output delivers an efficient output to determine the position of elements in the simulated environment. The project doesn't modify any original data, however, using few mathematical assumptions in the data sets, we can determine the results from the output with respect to efficiency of clustering and data storage methods.

IV. OBJECTIVES

1. To achieve deduplication at low RAM overhead for index-lookup with agent.
2. To achieve near-exact efficiency of duplicate elimination on heterogeneous file format.
3. To reduce memory contention between bursty read and writes traffic.
4. To increase the throughput of storage system using density based information.
5. To obtain load balance among various storage nodes using frequency information.

V. LITRETURE REVIEW

Data deduplication is an efficient data reduction approach that not only reduces storage space by eliminating duplicate data but also minimizes the transmission of redundant data in low bandwidth network environments. In general, a chunk-level data deduplication

scheme splits data blocks of a data stream (e.g., backup files, databases, and virtual machine images) into multiple data chunks that are each uniquely identified and duplicate-detected by a secure SHA-1 or MD5 hash signature (also called a fingerprint). Storage systems then remove duplicates of data chunks and store only one copy of them to achieve the goal of space savings.

One of the well-known static chunking schemes is Venti [1]. Venti is a storage system using static chunking, where 160-bit SHA1 hash key is used as the address of the data. This enforces a write-once policy since no other data block can be found with the same address. The addresses of multiple writes of the same data are identical, hence duplicate data is easily identified and the data block is physically stored only once.

Another known content-defined chunking system is LBFS [2] which is a network file system designed for low bandwidth networks. LBFS exploits similarities between files or versions of the same file to save bandwidth. It avoids sending data over the network when the same data can already be found in the server's file system or the client's cache.

In [3], it is demonstrated how by using Rabinkarp fingerprinting and a trusted third party server, we can efficiently manage variable size block of data and randomized convergent keys in a cloud storage environment.

DEDE [4,5] is a decentralized deduplication system designed for SAN clustered file systems that supports a virtualized environment via a shared storage system. Each host maintains a write-log that contains the hashes of the blocks it has written. Periodically, each host queries and updates a shared index for the hashes in its own write-log to identify and reclaim storage for duplicate blocks.

Hiroaki Akutsu and Kazunori Ueda [6] proposed a data deduplication system using Multi-Stage Erasure coding to achieve redundancy with high performance by reducing the redundant amount of data by using parity based information. This also enabled more efficient methods for scalable storage systems.

CSPD [7] is a client based security proven deduplication system, which offloads the performance overhead on storage systems by performing deduplication operations on client end using agent based systems. This enables scalable systems to operate on lower bandwidth network which improves the efficiency of overall storage systems.

Zheng Yan, Lifang Zhang and Wenxiu Ding [8] proposed a heterogeneous data storage management scheme, which flexibly offers both deduplication management and access control at the same time across multiple Cloud Service Providers.

J. Xu [9] proposed growing need for secure cloud storage services and the attractive properties of the convergent cryptography lead us to combine them, thus, defining an innovative solution to the data outsourcing security and efficiency issues. The solution is also shown to be resistant to unauthorized access to data and to any data disclosure during sharing process, providing two levels of access control verification.

VI. RELATED THEORY

A. File-level deduplication

File-level deduplication uses file-level granularity, which is the most coarse-grained granularity. File-level deduplication compares entire files based on a hash value of a file, like SHA-1, to avoid

saving the same files. In this section, we demonstrate how file-level deduplication works and its implementation.

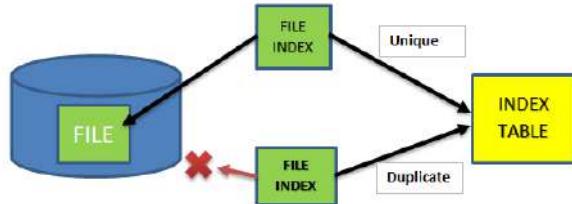


Fig. 6.1. File Level Deduplication

We begin by explaining how file-level deduplication works. As shown in Fig. 6.1., suppose we have two identical files. When we save the first file, deduplication computes an index that is a hash value using a one-way hash function. If the index is not found in the index table, the file is unique. In this case, the index and the file are saved to the index table and storage respectively. For the second file, the index of the file is found in the index table, so the corresponding file is not saved.

Algorithm:

```

DedupFile (string fileHashKey, string data)
{Boolean isUnique=false;
//check the bloom filter for index
If(existInBloomFilter(fileHashKey))
//check the inside chunk
{If (!isDuplicateInCache(fileHashKey))
{IsUnique=true ;}
Else {isUnique=true ;}
If (isUnique)
//save to storage
{SaveInCache (fileHashKey);
sm.setBufferData (fileHashKey, data); }
}
  
```

File-level deduplication can find redundancies of identical files but not redundancies within similar files. To find redundancies in similar files, fixed-size block deduplication has been proposed and uses fixed-size blocks for the granularity. In this section, we will understand how fixed-size block deduplication works and the implementation codes.

B. Fixed-size block-level

Fixed-size block-level deduplication separates a file into the same sized blocks and finds redundant blocks by comparing the indexes of the blocks. It runs fast because it only relies on offsets in a file to separate a file into blocks. However, fixed-size block deduplication has an issue when it comes to finding matching contents in similar files when the content at the beginning of the files is changed.

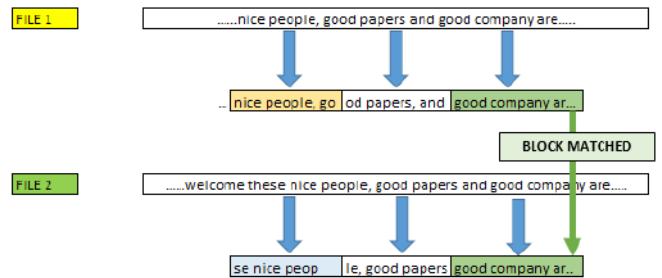


Fig. 6.2. Fixed Block Level Deduplication

For example, as shown in Fig. 6.2, suppose deduplication uses a 15 byte fixed-size block as granularity. When we save an original file File1, deduplication splits the file into 15 byte fixed size blocks. Likewise, when we save an updated file File2, in which we add the small text ‘welcome’ at the beginning of the original file, deduplication again splits the file into fixed-size blocks. However, blocks split from the updated second file are totally different from blocks split from the original first file. This is because the contents are shifted in the file; this is called the offset-shifting problem.

Algorithm:

```

SDedup::dedupBlock (string fileHashKey, string data, int blkSize){String blocks; Int numOfBlocks=0; String hashKey;
Int i;
//check duplicate file//
//A duplicate file does not need to be de-duplicated to blocks
if(isDuplicateInCache(fileHashKey)){return;}
else {saveInCache(fileHashKey);}
//check duplicate blocks
chunkWrapper.setAvgChunkSize(blkSize);
//set block size
Blocks = chunkWrapper.getBlocks (data, numOfBlocks);
//get blocks from data
For (i=0; i<numOfBlocks; i++){
HashKey=sha1Wrapper.getHashKey (blocks[i]);
//get hash key of block
If (!isDuplicateInCache(hashKey)){
saveInCache (hashKey);
//save an index of block
sm.setBufferData (hashKey, blocks[i]);
//save to storage}
Delete [] blocks;
//clear memory}
  
```

In pure fixed-size block deduplication, a file is directly split into blocks without checking whether the file itself exists, causing redundant processing overhead and memory overhead.

VII. PROPOSED DESIGN

Deduplication aims to find as many redundancies as possible while maintaining processing time. To reduce processing time, one typical technique is to check indexes of data in memory before accessing disks. If the data indexes are the same, deduplication

does not involve accessing the disks where the indexes are stored, which would reduce processing time. An index represents essential metadata that are used to compare data (or chunks).

The proposed system focuses on use of fixed block level deduplication technique to reduce the redundant chunks of data. The algorithm will read the input stream block by block from memory to identify unique chunks of data blocks along with frequency. An initial database will be constructed containing information of unique data chunks along with their frequency and physical address. The metadata from database will be used further to exploit the data distribution details with respect to their storage location, for further optimization. Each client in the environment will also run the same algorithm, thus, reducing the task at primary storage nodes. The metadata information from will be shared among clients to determine the density of data chunks with high frequency of hits to optimize the storage location. Based on frequency of data chunks, data will be clustered to further optimize the performance.



Fig. 7.1. Block Level Chunking

The algorithm for proposed methodology is below:

1. Start
2. Declare Variable
3. Initialize variable
4. Read block from Storage in tone iteration
 - 4.1 Read block till End of Data
 - 4.2 Generate HashKey using SHA-256
 - 4.3 Set the Frequency as 1
 - 4.4 If (Hashkey == True)
 - {++Frequency in bloom filter}
 - {Eliminate Duplicate Chunk}
 - 4.5. Store the Hashkey and Frequency in HashMap
5. Read the HashMap
 - {Using Scalable Clustering calculate the nodes}
6. Store High Frequency Chunks at Source
7. End

To compare redundant data, deduplication involves the computation of data indexes. Thus, an index should be unique for all data with different content. To ensure the uniqueness of an index, one-way hash functions, such as message digest 5 (MD5), secure hash algorithm 1 (SHA-1), or secure hash algorithm 2 (SHA-2) are used. These hash functions should not create the same index for different data. In other words, an index is normally considered a hash key that represents data. Indexes should be saved to permanent storage devices like a hard disk, but to speed up the comparison of indexes, they are prefetched in memory. An index table is a place where indexes are temporarily located for fast comparison. Such tables can be deployed using many different methods, but mainly they are built using hash tables, which allows comparisons to be made very quickly due to the time complexity of $O(1)$ with the overhead of hash table size.

To prevent an index table occupying memory as the number of indexes grows in the index table, a small summary vector, called a Bloom filter, is used to quickly check whether data are unique using small sized metadata. A Bloom filter is a bit array of m bits initially set to 0. Whenever same block (chunk) is found, bloom filter is used to update its occurrences. Thus, this enable us to track the frequency of block usage by various sources.

We also propose an approach for maintaining global copies of this frequency based metadata in a client-server based approach to offload the performance bottleneck at any specific node in given point in time. This also enable to scale-out the data in a clustered environment in real-time scenario with minimum computational efforts. The secured communication between the client-node also enables transfer of data to reduce overall network contention at any single node.

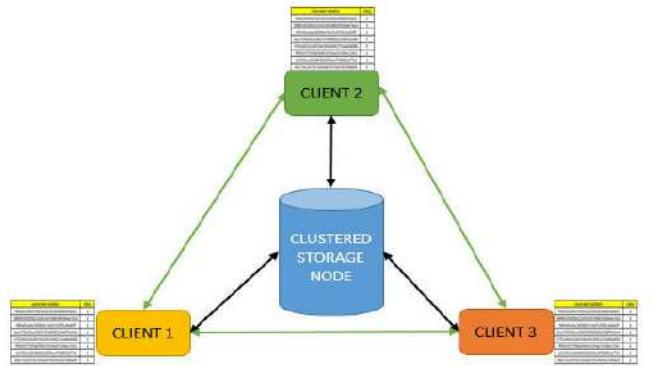


Fig. 7.2. Proposed Client-Server Architecture

VIII. IMPLEMENTATION & TEST RESULTS

The proposed system along with distributed dedup agents, will allow to load balance the data flow across the storage nodes in cloud environment. Thus, reducing the CPU utilization to some extent on primary storage node. The system will also provide fault tolerance to the SAN environment to some extent in case of node failure. With distributed metadata across the client nodes, the data request can be completed from client nodes itself, for the data blocks with high frequency. To demonstrate, we have saved some files at c:\test in the local memory of the machine. The proposed system is independent of the source data. We have considered 1MB of chunk size for scanning of the data. The size of window can be varied as per requirement which might change the overall performance and data storage capabilities.

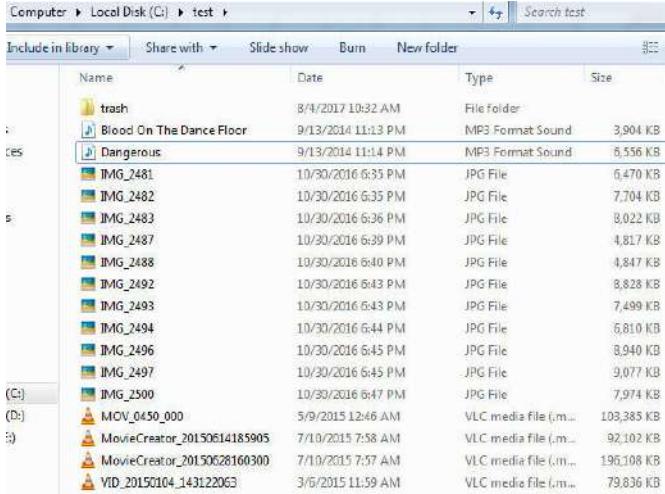


Fig. 8.1 Test Folder

The program initially scans the entire folder for the data and creates the frequency based table for the data retrieval procedure. Then based on the requirement, the number of clusters need to be specified to generate the clustering output. The runtime of the algorithm will vary depending upon the number of clusters.

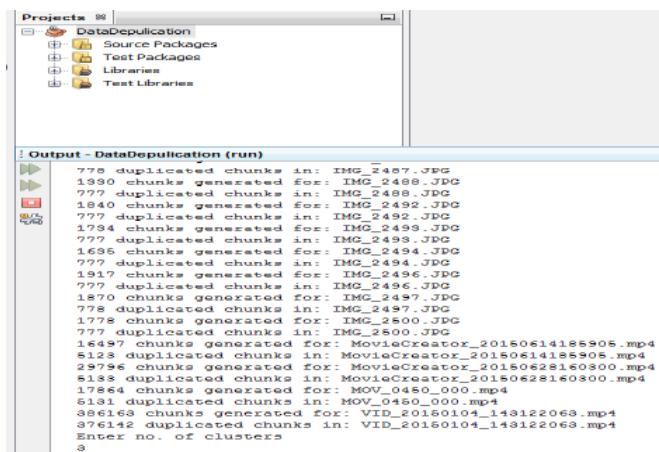


Fig. 8.2 Result of Scanning

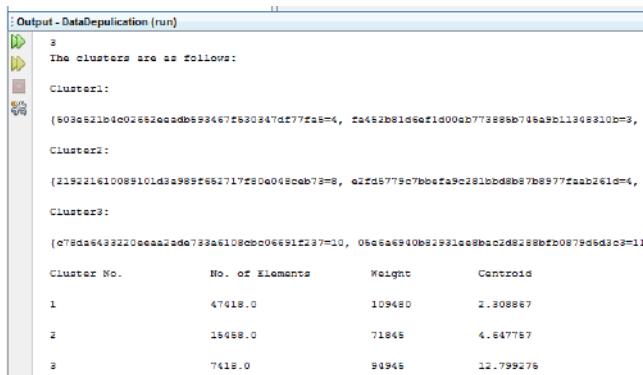


Fig. 8.3 Result of Clustering

IX. COMPARISION

Table II. Comparision Matrix

Parameters	Standalone Dedup	Scalable Dedup
Source data independent	No	Yes
Support multiple type/format of files	No	Yes
Granularity	Sub File level	Block Memory Level
Dedup ratio	Low	Very High
Data Scanning Complexity	$n^*O(m^*n)$	$n^*O(m+n)$
Data Retrieval Complexity	$n^*O(logm)$	$n^*O(1)$

IX. CONCLUSION

Data deduplication is a scalable and efficient data reduction technique for large-scale storage systems, which addresses the challenges imposed by the explosive growth in demand for data storage capacity. Data deduplication can reduce the number of disks used in the operation to reduce disk energy consumption costs. This paper provides a cloud based model for deduplication of large data. Apart from this, the proposed model also aims to provide security to the user data that has been stored in the Cloud Storage.

X. FUTURE SCOPE

The results shown displays significant improvement in the methodology of storing the data in the cloud environment. The efficiency and security parameters can further be improved by improving the encryption approaches at different silos. The hashing function used in determining the data blocks at the source can further optimized to increase the performance and reduce the RAM overheads. Similarly, the scalability of the clusters in real time environment in still not always recommended by many cloud vendors to avoid any corruption of data storages.

REFERENCES

- [1] S. Quinlan and S. Dorward, "Venti: a new approach to archival storage", Proceedings of the FAST 2002 Conference on File and Storage Technologies, (2002).
- [2] A. Muthitacharoen, B. Chen and D. Mazieres, "A low-bandwidth network file system", ACM SIGOPS Operating Systems Review, vol. 35, no. 5, (2001), pp. 174-187.
- [3] Haonan Su, Dong Zheng, Yinghui Zhang, "An Efficient and Secure Deduplication Scheme Based on Rabin Fingerprinting in Cloud Storage", 978-1-5386-3221-5/17 2017 IEEE.
- [4] F. Dougis and A. Iyengar, "Application-specific delta-encoding via resemblance detection", Proceedings of the USENIX Annual Technical Conference, (2003), pp. 1-23.
- [5] P. Kulkarni, F. Dougis, J. LaVoie and J. Tracey, "Redundancy elimination within large collections of files", Proceedings of the annual conference on USENIX Annual Technical Conference, USENIX Association, (2004).
- [6] Hiroaki Akutsu, Kazunori Ueda, Hideo Saito, "Network Optimized Multi-stage Erasure Coding for Scalable Storage Systems", 978-1-5090-5652-1/17 2017 IEEE.
- [7] Danping Li, Chao Yang, Chengzhou Li, "A client-based secure deduplication of multimedia data", 978-1-4673-8999-0/17 2017 IEEE.
- [8] Zheng Yan, Lifang Zhang, Wenxiu Ding, "Heterogeneous Data Storage Management with Deduplication in Cloud Computing", DOI 10.1109/TBDATA.2017.2701352, IEEE Transactions on Big Data.
- [9] J. Xu, E.-C. Chang, and J. Zhou, 2013, Weak leakage-resilient clientside deduplication of encrypted data in cloud storage.

- [10] Weiss, M.A.: *Data Structures and Algorithm Analysis in C++*, 3rd edn. Addison Wesley, Reading, MA (2005)
- [11] Xia,W., Jiang, H., Feng, D., Hua, Y.: SiLo: a similarity-locality based near-exact deduplication scheme with low RAM overhead and high throughput. In: USENIX Annual Technical Conference (2011)
- [12] Yan, F., Tan, Y.: A method of object-based de-duplication. *J. Netw.* 6(12), 1705–1712 (2011)
- [13] Zhu, B., Li, K., Patterson, H.: Avoiding the disk bottleneck in the data domain deduplication file system. In: Proceedings of the USENIX Conference on File and Storage Technologies (FAST) (2008)
- [14] Spring, N.T., Wetherall, D.: A protocol-independent technique for eliminating redundant network traffic. In: Proceedings of the ACM SIGCOMM 2000 Conference on Data Communication (2000)
- [15] Srinivasan, K., Bisson, T., Goodson, G., Voruganti, K.: iDedup: latency-aware, inline data deduplication for primary storage. In: Proceedings of the Tenth USENIX Conference on File and Storage Technologies (FAST) (2012)
- [16] Quinlan, S., Dorward, S.: Venti: a new approach to archival storage. In: Proceedings of the USENIX Conference on File and Storage Technologies (FAST) (2002)
- [17] Rabin, M.O.: Fingerprinting by random polynomials. Tech. Rep. Report TR-15-81, Harvard University (1981)
- [18] Min, J., Yoon, D., Won, Y.: Efficient deduplication techniques for modern backup operation. *IEEE Trans. Comput.* 60, 824–840 (2011)
- [19] Muthitacharoen, A., Chen, B., Mazières, D.: A low-bandwidth network file system. In: SOSP (2001)
- [20] Lillibridge, M., Eshghi, K., Bhagwat, D., Deolalikar, V., Trezise, G., Camble, P.: Sparse indexing: large scale, inline deduplication using sampling and locality. In: Proceedings of the USENIX Conference on File and Storage Technologies (FAST) (2009)
- [21] Liu, C., Lu, Y., Shi, C., Lu, G., Du, D., Wang, D.: ADMAD: Application-driven metadata aware de-duplication archival storage system. In: Fifth IEEE International Workshop on Storage Network Architecture and Parallel I/Os (SNAPI) , pp. 29–35 (2008)
- [22] Meyer, D.T., Bolosky,W.J.: A study of practical deduplication. In: Proceedings of the USENIX Conference on File and Storage Technologies (FAST) (2011)
- [23] Kim, D., Choi, B.Y.: HEDS: hybrid deduplication approach for email servers. In: 2012 Fourth International Conference on Ubiquitous and Future Networks (ICUFN) (2012)
- [24] Kim, D., Song, S., Choi, B.Y.: SAFE: structure-aware file and email deduplication for cloudbased storage systems. In: Proceedings of the 2nd IEEE International Conference on Cloud Networking (2013)
- [25] Debnath, B., Sengupta, S., Li, J.: ChunkStash: speeding up inline storage deduplication using flash memory. In: USENIX Annual Technical Conference (2010)
- [26] Dong,W., Douglis, F., Li, K., Patterson, R.H., Reddy, S., Shilane, P.: Tradeoffs in scalable data routing for deduplication clusters. In: Proceedings of the USENIX Conference on File and Storage Technologies (FAST) (2011)
- [27] Drago, I., Mellia, M., Munafò, M., Sperotto, A., Sadre, R., Pras, A.: Inside dropbox: understanding personal cloud storage services. In: Proceedings of the 2012 ACM Conference on Internet Measurement Conference (IMC), pp. 481–494 (2012)

Quantum Key Distribution: An Application of Quantum Cryptography

Yash Shah, Vidyadhari R. Singh
Lecturer, Assistant Professor
TCET, Mumbai

Abstract: Current cryptography techniques can be defeated easily by the use of ultra-speed computer. Quantum cryptography offers a robust alternative as it works on the principles of fundamental physics. It offers a system which exploits Heisenberg's uncertainty principle and different states of photons after polarization. Quantum cryptography offers a strong mechanism for key sharing in symmetric key cryptography called Quantum Key Distribution (QKD).

I. Introduction

In this modern age of Telecommunications and the Internet, the information has become a precious commodity. Sometimes it must therefore be kept safe from stealing. For example, the loss of private information to an eavesdropper. There are many features to security and many applications, ranging from secure commerce and payments to private communications and protecting passwords. One essential feature for secure communications is cryptography [1], which not only protects data from stealing or modification, but can also be used for user authentication. The main aim of cryptography is to protect data transferred in the likely presence of attackers. A cryptographic transformation of data is a procedure by which a plaintext data is encrypted, resulting in a modified text, called cipher text, that converts the message into an unintelligible message. The cipher text can be reverse-altered by a designated recipient so that the original plaintext can be recaptured. The techniques of cryptography are usually categorised as traditional or modern. Traditional techniques use operations of coding i.e. use of alternative words or phrases, transposition i.e. reordering of plaintext, and substitution i.e. modification of plaintext characters). Whereas, modern techniques use computers, and depends upon extremely long keys, convoluted algorithms, and intractable problems to achieve assurances of security. There are two main fields of modern cryptographic techniques: Public key Encryption [2] and Secret Key Encryption [1][2]. A public-key encryption, in which a message is encrypted with a recipient's public key. The message cannot be decrypted by anyone who does not possess the matching private key, who is thus presumed to be the owner of that key and the person associated with the public key. A secret key is an encryption key known only to the party or parties that exchange secret messages. The risk in this system is that if either party loses the key or it is stolen, the system is broken. The development of quantum cryptography was encouraged by the short-comings of classical cryptographic methods, which can be divided as either "public-key" or "secret-key" methods. Quantum cryptography is

an approach to a cryptography based on the laws of quantum physics.

II. Proposed work

Quantum Cryptography is an effort to allow two users of a common communication channel to create a body of shared and secret information. This information, which generally takes the form of a random string of bits, can then be used as a conventional secret key for secure communication. The Heisenberg Uncertainty principle and quantum entanglement can be exploited in as system of secure communication often referred to as "Quantum Cryptography". In quantum mechanics, the uncertainty principle, also known as Heisenberg's uncertainty principle or Heisenberg's indeterminacy principle, is any of a variety of mathematical inequalities asserting a fundamental limit to the precision with which certain pairs of physical properties of a particle, known as complementary variables. Each of the following elements included in quantum cryptography are explained with the help of an example with three actors viz. Alice (Sender), Bob (Receiver) and Eve (Eavesdropper)

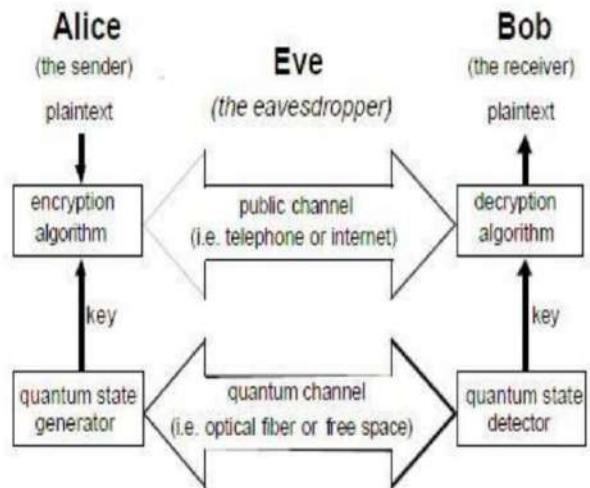


Figure 1. Actors in the illustrations

a) Key Distribution

Alice and Bob first agree on two representations for ones and zeroes. One for each basis used $\{\uparrow, \rightarrow\}$ and $\{\downarrow, \leftarrow\}$. This agreement can be done in public:

Table 1. Defining states of polarisation to 1s and 0s

1	0
↑	→
↗	↖

Alice sends a sequence of photons to Bob. Each photon in a state with polarization corresponding to 1 or 0, but with randomly chosen basis. Bob measures the state of the photons he receives, with each state measured with respect to randomly chosen basis. Alice and Bob

Table 2. Truth Table for eavesdropper

Alice's basis	Alice's bit	Alice's photon	Eve's basis	Correct	Eve's photon	Eve's bit	Correct
{↑, →}	1	↑	{↑, →}	Yes	↑	1	Yes
			{↗, ↖}	No	↗	1	Yes
	0	→	{↑, →}	Yes	→	0	Yes
			{↗, ↖}	No	↖	1	No
			{↗, ↖}		↗	0	Yes
			{↑, →}		↑	1	Yes
{↗, ↖}	1	↗	{↗, ↖}	Yes	↗	1	Yes
			{↑, →}	No	↑	1	Yes
	0	↖	{↗, ↖}	Yes	→	0	No
			{↑, →}	No	↖	0	Yes
			{↗, ↖}		→	0	Yes
			{↑, →}		↑	1	No

c) Eves Problem

The Heisenberg Uncertainty principle states that, it is impossible to determine the quantum state of any system without disturbing that system. The theory of photon polarization states that, an eavesdropper cannot copy unknown qubits i.e. unknown quantum states, due to no-cloning theorem which was first introduced by Wootters and Zurek in 1982. Eve has to re-send all the photons to Bob which will introduce an error, since Eve doesn't know the correct basis used by Alice. Bob will detect an increased error rate. Still possible for Eve to eavesdrop just a few photons, and

communicates via an open channel. For each photon, they reveal which basis was used for encoding and decoding respectively. All photons which has been encoded and decoded with the same basis are kept, while all those where the basis don't agree are discarded.

b) Eavesdropping

Eve has to randomly select basis for her measurement. Her basis will be wrong in 50% of the time. Whatever basis Eve chose she will measure 1 or 0. When Eve picks the wrong basis, there is 50% chance that she'll measure the right value of the bit.

Table 2. Truth Table for eavesdropper

hope that this will not increase the error to an alarming rate. If so, Eve would have at least partial knowledge of the key.

d) Detecting Eavesdropping

When Alice and Bob need to test for eavesdropping. By randomly selecting a number of bits from the key and compute its error rate. Error rate < Emax assume no eavesdropping. Error rate > Emax assume eavesdropping (or the channel is unexpectedly noisy). Alice and Bob should then discard the whole key and start over.

e) Noise

Noise might introduce errors. A detector might detect a photon even though there are no photons θ Solution: send the photons according to a time schedule. Then Bob knows when to expect a photon, and can discard those that doesn't fit into the scheme's time window. There also has to be some kind of error correction in the overall process.

f) Error Correction

Alice and Bob agree on a random permutation of the bits in the key. They split the key into blocks of length k. Compare the parity of each block. If they compute the same parity, the block is considered correct. If their parity is different, they look for the erroneous bit, using a binary search in the block. Alice and Bob discard the last bit of each block whose parity has been announced. This is repeated with different permutations and block size, until Alice and Bob fail to find any disagreement in many subsequent comparisons.

g) Privacy Amplification

Eve might have partial knowledge of the key. Transform the key into a shorter but secure key. Suppose there are n bits in the key and Eve has knowledge of m bits. Randomly chose a hash function where,

$$h(x): \{0,1\}^n \rightarrow \{0,1\}^{n-m-s}$$

Reduces Eve's knowledge of the key to $2^{-s/\ln 2}$ bits

Quantum key distribution (QKD) is a revolutionary security technology that exploits the laws of quantum mechanics to achieve information-theoretic secure key exchange. QKD enables two parties to “grow” a shared secret key without placing any limits on an adversary’s computational power and is unique in its ability to detect the presence of any third-party eavesdropping on the key exchange. Due to the fundamental laws of quantum mechanics, any third-party eavesdropping on the key exchange will introduce detectable errors. If the errors are below a defined threshold, an unconditionally secure key can be distilled. When QKD is used in conjunction with the one-time pad symmetric cryptographic algorithm, the result is an unconditionally secure cryptographic system. [5]

Quantum communication includes encoding information in quantum states, or qubits. Usually, photons are used for these quantum states. Quantum key distribution exploits certain properties of these quantum states to ensure its security. There are several different approaches to quantum key distribution, but they can be divided into two main categories depending on which property they exploit.

III. Security Implications

The hacking targets the vulnerabilities in the operation of a Quantum Key Distribution protocol or deficiencies in the components of the physical devices that are used in construction

of the system. In an unconditionally secure system, the private key from QKD is used as the key in a one-time pad. Since the key is information theoretically secure, so too is the encryption of the message: no computer, quantum or classical, will ever be able to decipher the encrypted message. There are challenges to this system, however. First, the one-time pad keys must be carefully stored and managed, as the double-use of one-time keys can seriously compromise security. Second, the Physical QKD systems cannot yet achieve sufficiently high key generation rates to be able to encrypt large messages with one-time pads in real time [6].

IV. Conclusion

QKD provides substantial advantages when compared to conventional key distribution. First, the security of QKD security rests on the foundations of quantum mechanics. This is in contrast to traditional key distribution protocols which rely on computational security, where the computational difficulty of certain mathematical functions is the foundation of security. Second, when using QKD, one can determine if an adversary is eavesdropping on the link because it will induce errors in the key exchange process. In contrast, traditional key exchange algorithms cannot provide any indication of eavesdropping or guarantee of key security.

Over the last 28 years, research in the QKD area has matured the technology and resulted in commercial QKD implementations. As the distance increases, the generated key rate drops. At long distances, QKD systems cannot generate enough key material to support bulk encryption using the OTP. However, commercial QKD systems such as the ID Quantique Cerberis6 system combine a conventional highspeed layer 2 encryption engine with the unconditional security of QKD technology. In this case, the key generated by the QKD system is used as the symmetric key for an Advanced Encryption Standard (AES) bulk encryptor. In this mode of operation, the AES key can be changed based upon the key generation rate of the QKD system. For example, the AES key could be changed once per minute if the QKD system is able to generate at least 128 key bits per minute. Finally, interest in QKD research continues to grow each year. The race is on to improve the quality of emitters, detectors, and fiber to enable QKD to operate over greater distances and at higher key rates. It is only a matter of time before you will encounter a QKD system in your security infrastructure. [5]

V. References

- [1] Quantum Cryptography: A Review, Seema S. Kute and Chitra G. Desai, Indian Journal of Science and Technology, Vol 10(3), DOI: 10.17485/ijst/2017/v10i3/110635, January 2017
- [2] Quantum Key Distribution Protocols: A Survey, Mohamed Elboukhari, Mostafa Azizi, Abdelmalek Azizi, International Journal of Universal Computer Sciences (Vol.1-2010/Iss.2) Elboukhari et al. / Quantum Key Distribution Protocols: A Survey / pp. 59-67
- [3] Quantum Key Distribution: A Revolutionary Security Technology, Michael R. Grimaila, Jeffrey Morris, and Douglas Hodson, 20 – ISSA Journal, June 2012
- [4] The Case for Quantum Key Distribution, Douglas Stebila, Michele Mosca and Norbert Lütkenhaus, Dec 2009

An Image Encryption And Decryption Using Chaotic Technique

Anita A. Lahane¹, Surabhi Singh², Pranjali Sankhe², Shruti Pimple²

Assistant Professor¹, UG Student²

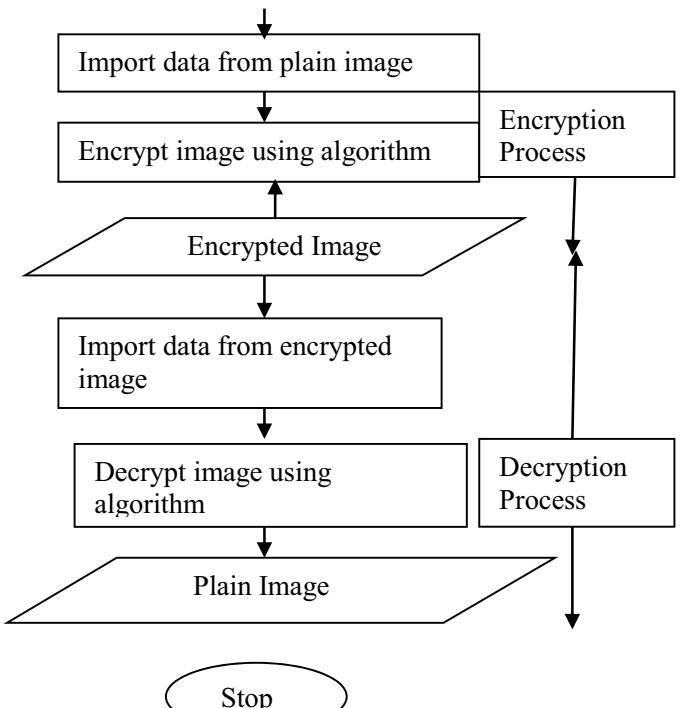
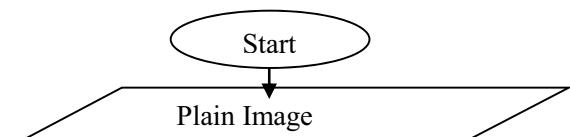
Department of Computer Engineering, Rajiv Gandhi Institute of Technology, Maharashtra, India
anitalahane@yahoo.co.in, sur1997.18@gmail.com, sankhep7@gmail.com, shrutispimple08@gmail.com

Abstract: In a fast digital world security plays a key role in transmission of images. To overcome these challenges Cryptography is used. Cryptography is the process of encryption and decryption of the data and used for making data, images secure and confidential. There is also another side where the attackers, unauthorized user gets an opportunity to reuse, retrieve, disturb the images. In this paper we will apply the chaotic encryption and chaotic decryption on an image. Chaotic Algorithm is a well-used method in real time secure image transmission system. Here the chaotic maps are used. Chaotic nature i.e., randomness property is present in both Henon map and Arnold cat map. Pseudorandom values generation plays an important key role in Henon maps and iteratively pixel shuffling is done in Arnold cat map. A sorting Technique is followed on key values produced by Henon map. By using those sorted positions, shuffle the pixel values generated by Arnold cat map iteratively. We will do some experimental analysis of the chaotic nature of the Henon map. It will be found that Henon map behaves chaotic in nature and encryption and decryption of image is successfully done. In this way the images are provided with high security for confidential transmission and shows a good resistance against brute-force attacks.

Keywords: Encryption and Decryption, Chaotic maps, Arnold cat map, Henon map, Pseudo-Random values, Pixel shuffling.

I. INTRODUCTION

Internet has evolved rapidly and thus the amount of data present on the internet is growing exponentially. The data shared in the form of text, images and other form may contain highly sensitive and confidential information. Image encryption and decryption process have been increased to meet the demand for real time secure image transmission over the internet. As everything has its pros and cons, the risk of data corruption, forging, data extraction, etc. has been a boon to the hackers. Hence to protect from the above the technology called cryptography was introduced. Cryptography means the method of storing and transmitting required data in a particular form so that legal users can only read and process it. Modern cryptography involves computer science algorithms and mathematical concepts. Cryptography techniques can be applied on any data like text, images, videos, etc., But in real time scenarios, as the methods of encryption and decryption has been low in speeds, which may result to significant latency.



Text Encryption and Image Encryption are different from their kind. Most existing encryption standards aim at image encryption compared with text encryption has its own characteristics and special features with many unique specifications. Hence we have chosen to implement the paper Chaos Encryption and Decryption using pixel Shuffling. It provides an efficient algorithm for encrypting images which takes shorter computational time and low computing power and gives high encryption strength. Our paper mainly concentrate on image encryption and decryption which involves key generation and pixel shuffling. In this approach we generate key using the random numbers. Randoms are classified into two types they are True Random Numbers(TRNGS) and Pseudo Random number generation (PRNGS).

II. PROPOSED SYSTEM

In this paper chaotic algorithm is used for encryption and decryption. Arnold cat map and Henon map technique is used. This encryption scheme is suitable for applications like internet image encryption and secure transmission of confidential information in the internet. proposed an image encryption scheme

which utilizes two chaotic logistic maps and an external key of 80-bit. The initial conditions for both logistic maps were derived using the external secret key. The first logistic map was used to generate numbers in the range between 1 and 24 and the initial condition of the second logistic map was modified by the numbers generated by the first logistic map.

III. METHODOLOGY

A. ARNOLD'S CAT MAP

- a) Arnold cat map was discovered by Valdimir Arnold in 1960. it takes the logics from linear algebra and uses them to change the pixel positions with respect to the original image. The Arnold Cat Map is a discrete system that stretches and folds its trajectories in phase space.
- b) Arnold cat map has a unique hyperbolic fixed point (the vertices of the square). The linear transformation which defines the map is hyperbolic: its eigen values are irrational numbers, one greater and the other smaller than 1 (in absolute value), so they are associated respectively to an expanding and a contracting eigenspace which are also the stable and unstable manifolds. The eigenspace are orthogonal because the matrix is symmetric.
- c) Defining the momentum variable $pt = qt - qt-1$, the above second order dynamics can be re-written as a mapping of the square $0 \leq q, p \leq N$
- d) Arnold cat Map is a transformation that can be applied to an image. The pixels of the image appear to be randomly rearranged, but when the transformation is repeated enough times, the original image will reappear.

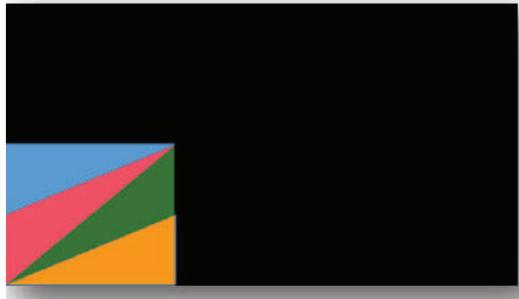


Figure : A Sample image showing the linear map

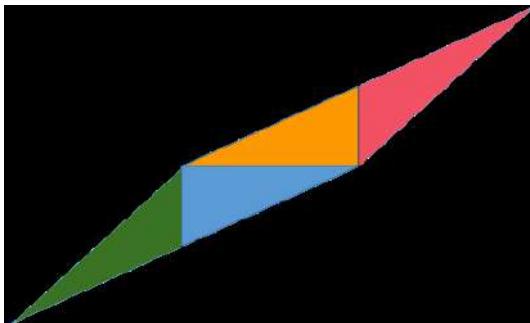


Figure An Arnold Map view

IV. HENON MAP

- a) The Henon map is a discrete time dynamic system introduced by michel henon.
- b) The map depends on two parameters, a and b, which for the

classical Henon map have values of $a = 1.4$ and $b = 0.3$. For the classical values the Henon map is chaotic. For other values of a and b the map may be chaotic, intermittent, or converge to a periodic orbit.

- c) These slopes arise from the linearizations of the stable manifold and unstable manifold of the fixed point. The unstable manifold of the fixed point in the attractor is contained in the strange attractor of the Henon map.
- d) It is a simplified model of coincare section of the lorenz model it is one of the most studied example of dynamically systems that exhibits chaotic behaviour it takes (x_0,y_0) in the plane and map it to a new point.
- e) FORMULA: $X_{n+1} = Y_n + 1 - a * x_n * y_n$ $Y_{n+1} = b * x_n$



Figure: Actual Henon Attractor

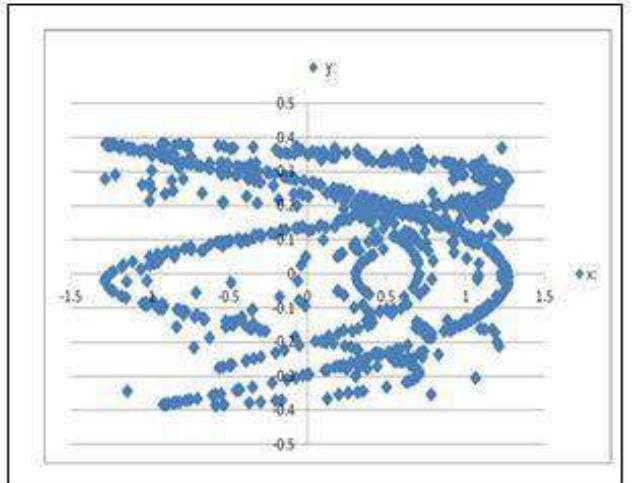


Figure: Actual Henon map generated through graph

V. PRESENT WORK

The image is taken for the process of encryption. It should be of .jpg or .jpeg, .png, .gif image format. The image pixels are extracted which is depending upon the dimension of the image i.e. image height and image width and store in an array. After that the pixel shuffling is done by using the Arnold's Cat map whose equations are given below:

$$xn+1 = (2 * xn + yn) \bmod 1$$

$$yn+1 = (xn + yn) \bmod 1$$

The pixel shuffling is done to change the pixel position and confuse the attacker, The new pixel value are now stored in a

new array.

After Arnold cat map we use henon map to generate the key values. The key values that are generated are repeated after a large iterations. The equations that are used are given below:

$$\begin{aligned} x_{n+1} &= y_n + 1 - a \times x_n \times x_n \\ y_{n+1} &= b \times x_n \end{aligned}$$

Where the values of a & b is 1.4 & 0.3 respectively.

If S is a given set of random numbers then the entropy of the given set H(S) can be calculated as:

$$H(S) = - \sum P(s) * [\log_2 P(s)] \text{ bits}$$

Where s belongs to S and P(s) is the probability of occurrences of s in sample space S. The number of iteration will decide the pixel position. As the key values we have got are all in the form of decimal format so we can't do Xor operation between a decimal value and an integer value, so we have used round method to make these decimal values to integer value by using this operation:

$$\text{Roundkey} = \text{Round} ((\text{key} * 65536) \bmod 256)$$

After getting the pixel values and the key values, the XOR operation is done between them and thus created a cipher image. Decryption is known as reverse of encryption. Now the process of decryption is done by using the process like pixel shuffling using Arnold's Cat map, generation of key values using the Henon map, and finally XOR operation between the rounded key values and pixel values of the encrypted image which brings the original image.

As we have used the chaotic maps i.e. ACM & Henon map, so we got the distorted image or the cipher image in the process of encryption by using the key value and simultaneously we got the original image in the process of decryption by using the same key value used at the time of encryption.

A. ALGORITHMS

Encryption algorithm:

1. The original image of .jpg or .jpeg format is chosen for the process of encryption.
2. Pixel extraction is done of the input image by taking the image dimension i.e. Height and Width of the image.
3. Pixel shuffling of pixels of the input image is done by using the Arnolds Cat map which is chaotic in nature.
4. Generation of Key values or the pseudo-random numbers using the Henon map which is chaotic in nature.
5. XOR operation is done between the pixel values generated from the input image and the key values generated by Henon map.
6. Cipher image or Encrypted image is done successfully and encryption process is over.

Decryption algorithm:

1. The cipher image which got from the process of encryption is chosen for the process of decryption.
2. Pixel extraction is done of the cipher image by taking the image dimension i.e. Height and Width of the cipher image.
3. Pixel shuffling of pixels of the cipher image is done by using the Arnolds Cat map which is chaotic in nature.
4. Generation of Key values using the Henon map which behaves chaotically.
5. XOR operation is done between the pixel values and the key values generated by Henon map.
6. Original image is brought back from the cipher image

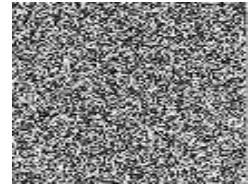
successfully and decryption process is over

B. EXPERIMENTAL ANALYSIS

Below are the sample output of the images, which are used to find the cipher and the decrypted image.



Original image



Encrypted image



Decrypted image



Decrypted image using wrong key

VI. CONCLUSION

In this paper, we have proposed a new encryption algorithm. The chaotic system is highly sensitive to initial values and parameters of the system. We used the randomness property of chaotic scheme i.e Arnold cat map and Henon map. The key value is generated using henon map and pixel shuffling is done using Arnold cat map. The process of decryption is same as encryption the same key is used for both the processes. This process will help to secure the data and information will be confidential and also would not be accessible to any unauthorized users.

VII. REFERENCES

- [1] Yung Dong; Jiasheng Liu; Canyan Zhu; Yiming Wang; 'Image encryption algorithm based on chaotic mapping' Computer Science and Information Technology (ICCSIT), 2010.
- [2] G.Chaitanya,B.Keerthi,A.Saleem, 'An image encryption and decryption using chaos algorithm' IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)eISSN:2278-2834,P-ISSN:2278-8735,Volume 10,Issue 2,Ver .11(Mar Apr 2015).pp 103- 108 <http://www.osrjournal.org>
- [3] Mayank Mishra ,Prashant Singh ,Chinmay Garg 'A new algorithm of encryption and decryption of image using chaotic mapping.' International Journal of InformationComputation Technology.ISSN 0974-2239 Volume4,Number7(2014) pp.741-746 International Research Publication House <http://www.ipphouse.com>
- [4] N.K. Pareek, V. Patidar, K.K. Sud, 'Image encryption using chaotic logistic map, Image and Vision Computing' Vol. 24, No. 9, 2006, pp. 926934.
- [5] Dr. Vivek Sharma,Hariom C. Agnihotri,Chetan Patil. 'An image encryption and decryption using two chaotic schemes' International journal of research in Advent Technology,Vol.2, No.2, Feb 2014.
- [6] Mintu "An Image Encryption & Decryption Approach Based on Pixel Shuffling Using Arnold Cat Map & Henon Map" by

Agyan Kumar Prusty, AsutoshPattanaik, Swastik Mishra
published in 2013 International Conference on Advanced
Computing and Communication Systems(ICACCS-2013),Dec19-
21,2013,coimbatore,India.

Multi Touch Technology Interface and Computer Interactions

Swastik Chava, Salonee America, Meher Dodhia, Mahima Jain

Department of Computer Engineering Thakur College of Engineering and Technology

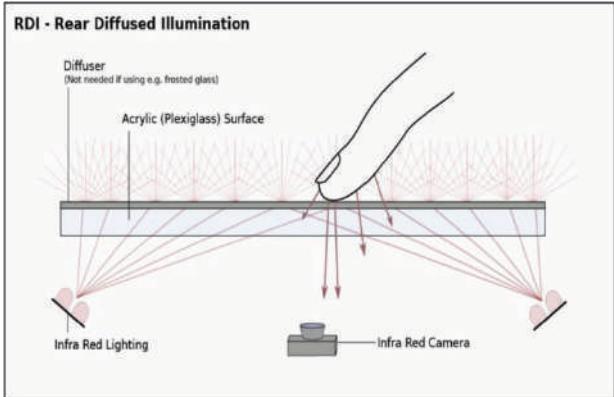
Abstract-This paper deals with the future technology which is already here, which allows the human interaction with the computer through recognizing touch. There are many touch technologies, however they are not used commercially due to many limitations. Due to the evolution in the past few years, it detects the inputs through tracking algorithm. With the common experience of touchscreens every day in life, it should focus on the quality for commercial use to access any system. Furthermore, FTIR and DI which are multi touch technology based on computer vision have been discussed. This challenge and type of interaction has many applications which can be used in shopping malls, restaurants, education and everyday use.

Keywords—multi touch technology, diffuser, blobs detection, gestures, ir camera, capacitive , acoustic, optical.

I. INTRODUCTION

A. Background:

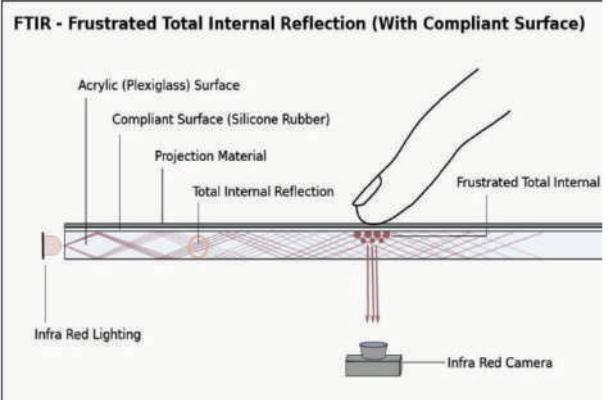
'Today everything is at the tip of one's fingers'. The invention of touch is evidence of this popular belief. Gone are the days of conventional technologies like keyboards and mouse. The user can access and interact with the computer screens just by using his fingers or for that matter just gestures. The intuitive and elegant nature of this technology makes it easy for everyone to use. Infinite applications can be implemented using the technology of multi touch. The popularity of smart phones has made it possible for the users to adapt to the touch technology. The technology has immense potential to be explored in countless ways and is still in an evolving stage.



B. Purpose:

The goal of this paper is to create the most cost effective and easiest techniques for multi touch screen applications.

The applications using gesture analysis bridges the gap between the hardware usage and touch interactions. Multi touch technology is a human-computer interaction technique and aims at reading several inputs simultaneously thereby syncing computation and communication. It is a user-centric and user-friendly interface which facilitates smooth performance, outstanding visual clarity, perfect and precise response time and eminent user experience. A multi touch screen consists of a software that reads the simultaneous inputs in



contrast to the single touch screens which senses one input at a time. There are two main features of multi touch. Firstly, one can have a direct interaction with the information displayed on the screen, wherever it is. Secondly there is no need of a hardware device like a mouse or a stylus to point on the screen which can merely be done by gestures. Suppose one needs to zoom in or zoom out, it can be done by using finger movements. Similarly one can

change, rotate, swipe or span across the entire screen without any intermediate device. A multi touch table or wall used at a supermarket or a hotel or even in industries increases productivity manifold as the number of users it interacts with increases. Thus this technology is advantageous in a million ways in data recovery, digital storyboard, geographical exploration and medicinal imaging.

II. MULTI TOUCH TECHNOLOGY

A. FTIR (Frustrated Total Internal Reflection):

In the phenomenon, an acrylic panel or sheet is used where IR led of the required wavelength are inserted in the edges of all sides of the panel. The IR led light is trapped in the acrylic sheet by internal reflection. As we touch the surface of the panel, the light is frustrated inside the panel which causes the IR led light to scatter where it is captured by the IR camera. When the blob is visible the IR camera senses the IR light which is not seen by naked eye. Generally, a silicon rubber is used to increase the sensitivity of the touch surface. While touching the acrylic one must press harder for the FTIR effect, but using silicon rubber no force is required and sensitivity is increased. *DI (Diffused Illumination)* :

In this phenomenon, the IR led are used above or below the projection surface with is called front and rear diffused illumination respectively. The surface is an acrylic panel with a diffuser on the projected surface. When we touch the screen, the blob created reflects extra light than diffuser and this light is sensed by the infrared camera. Thus, as the user touches the light, it can't pass through the finger and IR light hits the surface with extra light which is captured by the IR camera.

However, both these techniques have some pros and cons. While DI can recognize objects, the blobs and hovering can be calibrated easily. As FTIR is robust, it does not allow object recognitions easily. In FTIR, the calibration and blobs detection won't have to be adjusted every time using the touch interface as the light is always shining through the acrylic panel in the same way.



However, while implementing there are noises created by surrounding environment, quality of the hardware setup, etc. The calibration allows all the co-ordinate points to line up with the screen between the camera and projection.

TOUCH TECHNOLOGIES

There are two main types of multi touch screen technologies:

- 1) Overlay based-In this, the sensors are embedded in the screen which enable the touch features.
- 2) Perimeter based-In this, the sensors are found on the perimeter of the screen.

Overlay based touchscreen types

1) Capacitive touch technology-Indium tin oxide is used for coating the capacitive touchscreen panel which conducts electrical current throughout the screen. The sensor then becomes capacitive and a controlled field of electrons is formed along the horizontal and vertical axes. Capacitance is also exhibited by the human body as it has stored electrons. When the human finger comes in contact with the screen, the capacitive field of the screen is affected by an external capacitance which then sends a signal of a touch event to the processor for mathematical processing. Capacitive sensors can either be touched with a bare finger or with a conductive device being held by a bare hand. The Apple iPhone is an example of a product that uses capacitance touchscreen technology.

2) Resistive touch technology-In this there are two layers of transparent material.an air gap exists between them. On applying pressure to the outer layer, specific locations on the inner layer are touched. The panel then electrically acts similar to two voltage dividers with connected outputs. Due to this the electrical current changes and is recognized as a touch event. The controller then processes it. They are less sensitive. Contaminants affect the touch

Perimeter based technology types

1) Infrared touch technology-Beams of infrared lights are used for detecting touch event. There are different methods which are employed for the touchscreen panel. In the first method changes in the surface resistance is observed when there are thermal changes. However the drawbacks of this method are 1) it is slow 2) it requires warm hands. The second method uses interrupted light beam detected by horizontal and vertical sensors on the screen surface. 2) Surface acoustic wave touch-Ultrasonic waves are used to detect the event. When ultrasonic waves are passed over the panel a part of it is absorbed. The location of the touch event is registered and sent to the controller for processing. 3) Optical touch technology-It uses optical sensors, mainly CMOS sensors. This is a relatively versatile and scalable modern touchscreen technology. The sensors are located on the edges of the screen. Infrared backlights are placed in the camera's field of view on the other sides of the screen. A touch shows up as a shadow and each pair of cameras can then be triangulated to locate the touch.

Gestures

Gesture recognition is a human machine interaction technique which allows users to directly and naturally with the computer screens without any devices. The gestures are interpreted through mathematical algorithms. Gestures can be any bodily movement mostly involving the face or hand. The most basic gestures that are defined are select, Change, zoom and rotate. There are many techniques which have been used to interpret gestures.

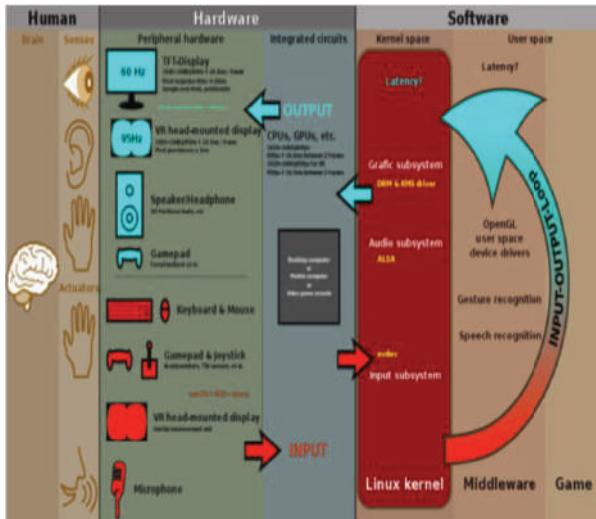
However the major issue is that no two individuals can have identical hand movements. Thus the screen should be efficient enough to understand the general meaning behind the gesture, provided it is close to the one that the screen actually recognizes. Only then will we be able to get the desired outputs.

We can categorize gestures in two types, Direct Manipulation Gesture and symbolic gestures pattern. In direct manipulation gesture we can use our fingers to translate, rotate and scale whereas in symbolic gesture we can make pattern like circle, triangle or some text.

Steps in converting input to multi touch gesture

- 1) Input layer-input is collected as video and electric signals.
- 2) Hardware abstraction layer -processing of data is done for the generation of image co-ordinates.
- 3) Transformation layer-calibration of the touch screen so converting of data to screen co-ordinates takes place.
- 4) Interpretation layer-a gesture is recognized at this stage. The three features of a gesture are start point, end point

and the motion happening between these points. 5) Widget layer- this layer gives the output. Zoom and pinch takes place here.



Applications of Multi Touch Screen Table:

The collaborative approach to these technologies gives rise to various applications which makes it more interesting and gives a unique touch to the technology. The main usage of these applications can be in museums, libraries, art galleries, schools, colleges and many other places. These applications can help improve the security of the organization. It can make a feature more interactive and collaborative. Day by day new ideas are improving the technology of multi touch screen table.

Shopping Billing System:

In these last months, virtual stores have been increasing all around the world. QR codes are used for shopping items instead of carts. Here, the items are directly received by the customer at the billing counter. But further this technology can be used for touch interface. While billing, sometimes the monitor hangs, freezes or QR scanner doesn't work which causes chaos. By using multi touch technology with raspberry pi + camera module techniques, it can be used for product tracking and item identification for marketing. Thus, just by keeping the item on the table, the product will be tracked and identified by the camera module and the information of the item will be directly stored into the system. While shopping if the customer wants to the details or health factors of a particular item they can just keep the item on the table. Here the items have different codes and the information is fed into the system. The customer will know about it eventually about the specifications and if they want to buy it. These approaches in the automated industries will have fast readability and customer satisfaction.

Collaborative Gaming:

Many players can come together and have a great time in the field of gaming. Players up to four in number can play games and have an interesting collaborative experience. The Multi Touch Screen Table provides features such as add new player, restart the game, quit the game. It also provides an ambiance for gaming which comprises of music that goes with the game aesthetics, the virtual effects and animation.

Face Creator:

For this application, one can create an art work in which the user can create a face and add on many features. Features such as nose,

eyes, hair color, sunglasses and many more. The image can then be saved and uploaded on any website. This is a unique way for the user to express their uniqueness in an art form. Many advancements can be made to this application so as to improve its functioning. The users can get as creative as they wish to be. Social media accounts of the user can be linked to this application so that it makes it easier for the user to share their creation.

Maps:

This feature of maps can be used extensively by people who travel often. This gives a different experience and makes it easy for the user to trace the route. Features such as find a location, calculating traffic time, finding out restaurants and cafes on the way for the destination can be provided.

Musical instruments:

Users can play the piano with ease and advanced technologies and techniques. The variations that can be brought in the sound field by this technology is immense. The improvements in the sound quality and the ease with which a musician can play the piano is commendable.

Art:

This application provides a platform for the user to draw and maximize his or her creativity. Features such as start a painting, stop, resume or continue, restart can be easily controlled by this application.

Education:

This technology can be used in schools, colleges and universities. It can be used by both teachers and students. Difficult concepts can be explained with total ease with the help of this technology. The students can get a different exposure and a different perspective to their way of learning. Food chains and restaurants:

The customers at the restaurants are able to order their food from their table itself. They can easily place, cancel, and update the order as per their wish. The Multi Touch Table will also detect the object that gets placed on the table. Details of the placed item can also be retrieved by this application. The bill information can also be fetched by this application.

Photo Editing:

This application plays a major role in the photo editing department because it makes it easier for the user to edit and enhance the minute details of the picture. Features include improve brightness, blur, add a text, add color and many more. Photo editing is widely used for magazines, newspapers.

Social Media:

Users can get an exclusive experience using this application for their social media accounts. The user friendliness of this technology will attract more users to use this application on a wide scale.

CONCLUSION

Touchscreens are 21st century interfaces. They are convenient as they replace buttons which occupy most of the space on smaller gadgets like smartphones, tablets etc. providing more screen space. Multi touch screen can be made using various techniques. The technology using FTIR is robust and very easy to implement. Different applications for multi-touch interfaces both exist and are being proposed. The use of multi-touch technology is expected to rapidly become common place.

REFERENCES

- [1] International Journal of Scientific and Research Publications, Volume 3, Issue 4, April 2020
- [2] International Journal of Scientific and Research

Publications, Volume 3, Issue 4, April 2013 1 ISSN 22503153 Multi
Touch: An Optical Approach (Comparison of
Various techniques) [3]<http://sethsandler.com/blog/>
[4]Advanced Computing: An International Journal (ACIJ
, Vol.2, No.6, November 2011 DOI:
10.5121/acij.2011.2611 113 MULTI-TOUCH SCREEN
INTERFACES AND GESTURE ANALYSIS: A
STUDY
[5]http://wiki.nuigroup.com/Compliant_surface
[6]<http://www.multigesture.net/projects/multi-touch-media-application-pro/>
[7] http://johnsear.com/multi_touch-table-applications/
[8] <http://www.military-technologies.net/2018/01/27/multi-touch-screen-market-global-industry-analysis-and-opportunity-assessment-2017-2022/>

Polymer Memory- Use of Plastic as Secondary Storage

Vipul Agarwal, Adwait Arayakandy, Ankit Gupta, Kalpana Gangwar

Computer Department, Thakur College of Engineering & Technology, Mumbai, India
vipagarwall@gmail.com, adwait2909@gmail.com, guptankit342@gmail.com, kalpana.iitm@gmail.com

Abstract—Polymer memory is currently a leading technology in memory device development. The core material used for the manufacturing of polymer memory is PEDOT, an electrically conducting polymer which is relatively denser and cheaper as compared to flash memory. This paper analyses the manufacturing, functionality, advantages & limitations of PEDOT based polymer memory as secondary storages.

Keywords—PEDOT; polymer; conductive; nano fibers; precipitate; morphology; photo-sensitive agents.

I. INTRODUCTION

The idea of integrated electronics in everyday objects is extremely attractive, but immediately beyond the cost structure inherent in silicon chips. Silicon solutions remain economically unattainable due to the high costs and processing materials. Digital memory has always been a close companion to all technological advances in information technology. Current memory technologies have many limitations. Dram is volatile and difficult to integrate, Ram is expensive and volatile, flash has slower scripts and fewer write/erase cycles than others. Experimenting with a polymeric material known as PEDOT, the University of Princeton researcher, Sven Moller, has come to a conclusion that while plastic conducts electricity at low voltages, it permanently loses its conductivity when exposed to higher voltages. Together with colleagues at HP Laboratories, he developed a method to exploit this property to store digital information, which can be stored as collections of ones and zeroes. The magical ingredient is not smaller transistors or an exotic material cooked up by the semiconductor industry. It is a plastic.

II. WHAT IS A PEDOT?

PEDOT - Poly (3,4-ethylenedioxythiophene) or PEDT, is a conductive polymer based on 3,4-ethylenedioxythiophene (EDOT monomer). It provides optical transparency in conducting state. It also has a moderate band gap and low redox potential. PEDOT coatings have a high stability in different charge and discharge cycles and can be electrogenerated directly in a conductive support.

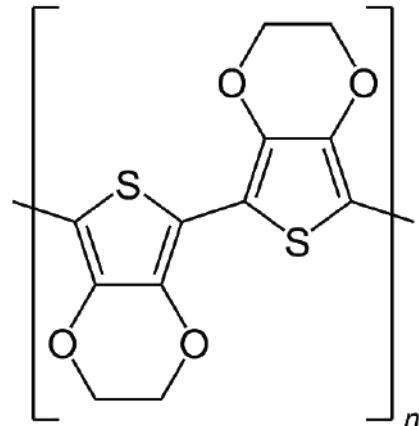


Fig. 1. EDOT Polymer

A. SYNTHESIS OF PEDOT

PEDOT nanofibers are produced from vanadium pentoxide nanofibers using a method named as nanofiber seeding method which is as follows:

1. EDOT is dissolved in an aqueous solution of camphorsulfonic acid (CSA) & a solution gel of vanadium Pentoxide nanofiber.
2. Radical cationic polymerization is initiated by addition of ammonium persulphate.
3. The resulting polymer precipitates from solution and has a general composition $(\text{PEDOT})(\text{CSA})_{0.11}(\text{HSO}_4)_{0.12}(\text{Cl})_{0.11}(\text{H}_2\text{O})_{0.19}$.
4. Washing the precipitate in a dilute HCl removes the vanadium compound from the composition.
5. The presence of the vanadium pentoxide seeds makes the difference between the formation of PEDOT nanofibers (100 to 180 nanometer diameter and one to several micro-meter long).
6. The formation of a more conventional granular morphology.
7. When applied to a solid substrate such as PET, PEDOT non-woven films have slightly lower optical transparency.

B. CONDUCTIVITY OF PEDOT

The PEDOT coating has a high stability during different charging and discharging cycles and can be electrogenerated directly on a conductive support. The material is a mixture of a negatively charged polymer called PSS- and one with a positive charge called PEDT +. It is used as an anti-static coating on the camera film and

on a photosensitive agent. Poly (3, 4-ethylenedioxythiophene): poly (styrenesulfonate) (PEDOT: PSS) is a conductive polymer that is very promising as the next generation of materials for the transparent electrode if it can achieve such high conductivity as that of the ITO (Indian Tin Oxide).

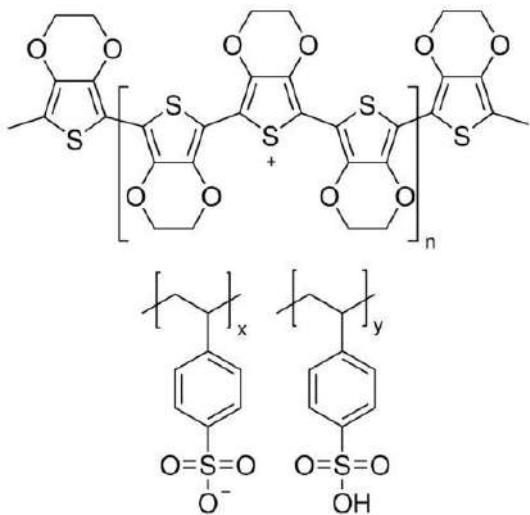


Fig. 2. PEDOT: PSS

III. ABOUT TECHNOLOGY

The device sandwiches a conductive polymer called PEDOT (polyethylene dioxythiophene). It can store a large amount of data (megabits) in a device of a square millimeter ten times denser than current memories. Converting the polymer to an insulator involves a permanent chemical change which means that the memory can only be written once.

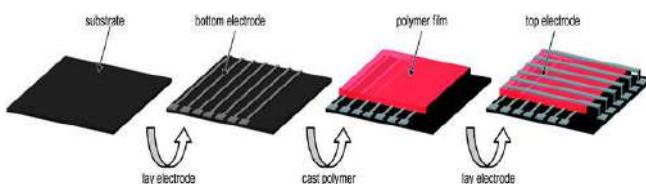


Fig. 3. Fabrication of Polymer Memory

IV. ARCHITECTURE

Each polymer is sandwiched between two electrodes which act as a memory cell. A voltage is applied between the bottom electrode. Memory is represented as space charges in the polymers that are detected using electrical pulses. The Thin Film Polymer system requires 0.5 million transistors per gigabyte of memory. On the other hand, a Silicon-based system requires between 1.5 to 6.5 billion transistors per gigabyte. The memory density can be increased by increasing the number of stacked layers without reducing the minimum feature size. To achieve multilevel stacking, the polymer used must be able to cross linked

so that it will not be dissolved upon deposition of additional layers.

A. STORING DATA

The polymer memory stores information in a completely different way than silicon devices. It stores data based on the electrical resistance of the polymer. An application of the electric field to a cell reduces the resistance of the polymer, thus increasing its ability to conduct the current; The polymer maintains its state until a field of opposite polarity is applied to increase its resistance to its original level. The different conductivity states represent information bits.

B. READ/WRITE DATA

For Writing: The data in the polymer memory use the cable an the diode surrounding the PEDOT blob to run a high or low current through it.

For Reading: The polymer memory flows the current through the upper cable and measures the current in the lower cable. The high current means that the bit is one and that the absence of current means that the bit is zero.

V. FEATURES OF POLYMER MEMORY

1. Data is stored by changing the polarization of the polymer between metal lines.
2. Polymer memory consists of zero transistor per bit of storage.
3. It has write speed faster than NAND and NOR flash memory.
4. It has simple processing and easy to integrate with other CMOS. It does not require cell standby power or refresh.
5. It has an operational temperature between -40°C and 110°C.

VI. COMPARISON (W.R.T SILICON-CHIPS)

1. Polymer memory can be stacked vertically in a product, producing a use of 3D space; Silicon chips can only be placed side by side.
2. It requires very few transistors, typically only 0.5M for 1GB storage compared to Silicon's 1.5 - 6.5B.
3. They cost about 5% of production compared to silicon-based memory.
4. The absence of moving parts or lasers offers a significant advantage in terms of speed compared to mechanical storage devices.
5. The polymer memory is not volatile, economical and has a fast read and write speed. On the other hand, RAM and DRAM are volatile, expensive and flash memory has a slower write speed.
6. Polymer memory manufacturing is easy compared to the production costs of the silicon-chip because the polymer is based on a solution and can be easily applied on large surfaces.
7. Polymers memory can be demonstrated as a better SSD than silicon based SSD (solid state unit).

VII. ADVANTAGES

1. Polymer memory layers can be stacked. This enables to achieve very high storage capacity.
2. Polymer memory is non-volatile in nature.

3. Polymer memory has a high speed of reading and writing capacity.
4. It is cheaper and has a very low cost per bit.
5. It has low power consumption.
6. It is easy to manufacture, uses an inkjet printer to spray liquid polymer circuits onto a surface.
7. Unlike a CD, reading data stored on this memory block does not involve any moving parts or laser, instead, it can be plugged directly into a circuit.

VIII. LIMITATION

1. It can be read many times but it can be write only once (WORM).
2. The biggest challenge in polymer memory is developing the production technique.
3. This technology is still under research, so it will take time to launch in the market.

IX. APPLICATION

1. Polymer Memory can be used as a WORM (Write Once Read Many) type memory applications.
2. It can be used in Holography Techniques due to its ability to store data permanently.
3. It can be used in personal Computers in replacement to the ROM Devices.
4. It can be used in Supercomputers to store a large amount data with the help of the memory's ability of stacking.



Fig. 4. Applications (Device Compatibility)

X. FACTS AND FICTION

1. A million bits of information could be stored in a square millimeter of material, or the thickness of a sheet of paper. A block of only one cubic centimeter can hold up to a thousand high-quality digital images.
2. According to the scientists, it would not require high temperature or vacuum chambers.
3. It is a very economical technology that offers an advantage over other technologies.
4. The PEDOT-based machine could solve the problem of hackers, who depend on the fact that they cannot afford to leave a trail for fear of being trapped by their dirty work. With PEDOT-based solutions, hackers could not delete their IP addresses.

XI. CONCLUSION

Polymer memory is considerably cheap and fast as compared to the silicon memory. This memory can be easily developed as the material required is easily available and the process of manufacturing is also simple. No huge investment is required as

compared to its counterpart. The power consumption is very less and the memory device is highly dense which can accumulate a large amount of data in small space. The data is maintained in memory even when the power is off. As the technology is still in development phase it does not enjoy large business in the market. It needs a lot of efforts by researchers and the marketing section to make this particular concept of memory popular and enjoy it is really worth.

REFERENCES

- [1] Technical Paper, "Memory Device Application Of a Conjugated Polymer: Role of Space Charge," Journal of Applied Physics, February 15, 2002.
- [2] Zhang, Xinyu; MacDiarmid, Alan G.; Manohar, Sanjeev K. (2005). "Chemical Synthesis of PEDOT nanofibers," Chemical Communications (42): 5328-30.doi:10.1039/b51129og. PMID 16244744.
- [3] A.Prakash, J.Ouyang, and J.Lin, "Polymer Memory Device based on Conjugated polymer and gold nanoparticles," J.Applied Physics, vol. 100, no.5, pp. 1-5, 2006.
- [4] MIT Technology Review 03/02- Improved Memory.
- [5] MIT Technology Review 09/02- Polymer Memory.
- [6] http://upload.wikimedia.org/wikipedia/commons/0/05/PEDOT_T.png
- [7] <https://www.intel.com>
- [8] Unfolding space, memory by N.Nagaraj Financial Daily from THE HINDU group of publications.
- [9] www.detnews.com
- [10] C. Perlov, W. Jackson, C. Taussig, S.Mo, and S. R. Forrest, "A polymer/ semiconductor write-once read-many times memory," *Nature*, vol. 426, November 2003.
- [11] <https://www.sigmadralich.com/catalog/product/aldrich/655201?lang=en®ion=IN>

Intelligent Traffic Light System(ITLS)

Akib Maredia, Rishabh Mishra, Ashutosh Mishra, Nitin Harane

Department of Computer Engineering, TCET, Mumbai 400101

akibmaredia@gmail.com, rishabhmisra220@gmail.com, ashu.mishra@gmail.com, nitinharane13@gmail.com

Abstract:Traffic is a popular problem in many cities of India including other countries. updatation of signals, poor law and bad traffic management has lead to traffic issue. One of the major problems with Urban Cities in India is that the existing infrastructure cannot be enlarged more, and thus the only option available is better management of the traffic. Traffic has a negative impact on economy, Hence this is the high time to effectively manage the traffic problem. There are various methods available for traffic management such as video analysis, infrared sensors, inductive loop detection, wireless sensor network, etc. All these methods are effective methods of smart traffic management. But the problem with these systems is that during installation time, the cost required for the installation and maintenance of the system is Very expensive. Therefore a new technology called Radio Frequency Identification (RFID) is used which can be coupled with the existing signal system that can act as a key to smart traffic management in real time. Another method is to use Wireless Sensor Network (WSN) and using new Avialable techniques for controlling the traffic flow. These techniques are very adaptive to traffic conditions on both ways intersections. A WSN is used as a tool to instrument and control traffic signals, while an intelligent traffic controller is developed to control the traffic

Keywords:Wireless sensor networks, intelligent traffic signal, traffic decongestion, real time traffic adaptation, traffic control , RFID ,.

I. INTRODUCTION

Traffic on road networks is nothing but slower speeds, increased trip time and increased queuing of the vehicles. When the number of vehicles exceeds the capacity of the road, traffic congestion occurs. In the metropolitan cities of India traffic congestion is a major problem. Traffic congestion is caused when the demand exceeds the available road capacity. This is known as saturation [1]. Individual incidents such as accidents or sudden braking of a car in a smooth flow of heavy traffic have rippling effects and cause traffic jams [2]. There are even severe security problems in traffic system due to anti social elements which also leads to stagnation of traffic at one place. In country like India, there is an annual loss of Rs 60,000 crores due to congestion (including fuel wastage). Congestion in India has also led to slow speeds of freight vehicles, and increased waiting time at checkpoints and toll plazas [3]. The average speed of vehicles on key corridors like

Mumbai-Chennai, Delhi-Chennai is less than 20kmph, while it is mere 21.35kmph on Delhi-Mumbai stretch. As per the transport corporation of India and IIM, India's freight volume is increasing annually at a rate of 9.08% and that of vehicles at 10.76%, but that of road is only 4.01%. This has resulted in reduced road space in accordance with the number of total vehicles [3].The average fuel mileage in India is only 3.96kmpl. The major reason for this is traffic congestion [3].India is the 2nd most populated country after China in Asia, thus with increase in population, the number of vehicles also increase [4].The economic growth has certainly has had an impact on urban traffic. As the income rises, more and more people begin to go for cars rather than two wheelers [5].Hence there is a need to manage traffic in a smart way as the management of traffic with the conventional way such as the signalling system is not having a major effect in curbing congestion of vehicular traffic.

II. EXISTING METHODOLOGY

A. Inductive Loop Detection

Inductive loop detection works on the principle that one or more turns of insulated wire are placed in shallow cut-outs in the roadway, a lead in wire runs from roadside pull box to the controller and to the electronic unit located in the controller cabinet. When a vehicle passes over the loop or stops, the induction of the wire is changed. Due to change in induction, there is change in the frequency. This change in the frequency causes the electronic unit to send a signal to the controller; indicating presence of the vehicle [6].Inductive loop detection is useful in knowing the vehicle presence, passage, occupancy And even the number of vehicles passing through a particular area [6, 7]. But there are few problems with this system. These include poor reliability due to improper connections made in the pull boxes and due to application of sealant over the cut-outs of the road. If this system is implemented in poor Pavement or where digging of the roads is frequent then the problem of reliability is aggravated [4, 5, 8].

B. Video Analysis

Video analysis consists of a smart camera placed which consists of sensors, a processing unit and a communication unit [9]. The traffic is continuously monitored using a smart camera. The video captured is then compressed so as to reduce the transmission bandwidth. The video analysis abstracts scene description from the raw video data. This description is then used to compute traffic statistics. This

Statistic includes frequency of the vehicles, average speed of the vehicles as well as the lane occupancy [9, 10]. The problems

associated with video analysis are – (a) the overall cost of the system is quite high (b) the system gets affected in case of heavy fog or rains (c) night time surveillance requires proper street lighting [6,8].

C. Infrared Sensors

Infrared sensors are used to detect energy emitted from vehicles, road surfaces and other objects. The energy captured by these infrared sensors is focused onto an infrared sensitive material using an optical system which then converts the energy into the electric signals. These signals are mounted overhead to view the traffic. Infrared sensors are used for signal control, detection of pedestrians in crosswalks and transmission of traffic information [11]. The basic disadvantages of infrared sensors are that the operation of the system may be affected due to fog; also installation and maintenance of the system is tedious [5, 8].

III. SMART TRAFFIC MANAGEMENT SYSTEM

A. Background

A Radio Frequency Identification (RFID) system consists of RFID controller and RFID tag.

1) RFID Controller:

The RFID controller consists of RFID interrogator. This interrogator is used for the communication with the RFID tag. The RFID controller then gets the signals/data received by the interrogator. Messaging interference is used to send commands and data messages from the controller components. Controller core is present inside the RFID controller. The controller core listens to the interrogators and depending upon the configuration; the controller core can perform read/write operations upon the RFID tag or can do both listening and performing operations [5]. The RFID controller can have serial interface through which external GSM/GPRS devices can be interfaced with it to make a dual radio device.

2) RFID Tag:

RFID tags are wireless devices which make use of radio frequency electromagnetic fields to transfer data, which is used for identifying and tracking of the objects. RFID tags are of two types: Active and Passive [12]. Active RFID has a battery installed, which the passive RFID doesn't have. Passive RFID has to depend on external source for working. Tags information can be stored in a non-volatile memory. Tag consists of a Radio Frequency transmitter and receiver. Each tag can be assigned a unique serial number [13].

B. Relevant Algorithm

Input:

Max_red denotes the maximum time for which the signal can be red.

Max_green denotes the maximum time for which the signal can be green.

Min_freq_count denotes the minimum frequency of vehicles passing per second stored statically in controllers.

Act_freq_count denotes the actual frequency of the vehicles passing per second = \sum vehicles/second.

Timer denotes the actual timer count.

Algorithm:

```

1. When the signal turn green.
While (Timer<Max_green and Timer is not 0) do
    If (Act_freq_count>Min_freq_count)
        Keep the signal green.
        Decrement timer count by 1.
    Else if (Act_freq_count<=Min_freq_count)
        Goto 2.
End
2. Make the signal red. Turn the adjacent signal green. Goto 1.
Desired Output: Effective congestion management
```

C. System Overview

D.

Each vehicle can be installed with a RFID tag. This RFID tag would store all the information regarding the vehicle such as the vehicle number, etc. RFID tags can be used in identifying each vehicle uniquely and also help the driver to receive some traffic messages. The existing signaling system can be coupled with the RFID controller. As described in figure 1, each signal can have the information regarding every vehicle that passes by it. Thus when a vehicle passes by a signal, the signal can automatically keep the count of the vehicles passing by it, and help in detection of traffic congestion. Each signal should be stored with a threshold value for which it should be red and green. Now depending upon the frequency of the vehicles passing by the signal per second, the timer can be dynamically controlled. Each controller of the signal should be stored with a value of minimum frequency of the vehicles passing by the signal. As soon as this minimum frequency is reached, the controller should send a command to the signal to turn red. Thus the signal is controlled dynamically. For example, suppose for a signal, maximum time for which a signal can be red is set to be 30 seconds and maximum time for which the signal can be green is set as 20 seconds. The controller is stored with the value of minimum frequency of vehicles passing by it per second as 5. Now suppose the signal turns green, the timer starts with a maximum value of 20. Initially the frequency of the vehicles passing the signal per second is 10, after 10 seconds this frequency reduces to 5, and then automatically the RFID controller sends a command to the signal to turn red. Thus the signal turns red and its adjacent signal in that junction turns green. This process continues in a cycle. Thus dynamic controlling of the signal helps in reducing the wastage of time. This also helps in avoiding traffic congestion as priority is given to a high vehicular traffic road. This system helps in detection of traffic congestion. If the frequency of the vehicles passing the signal per second remains higher than the value

set	even	though
-----	------	--------

the maximum value of the timer is reached, then the congestion has occurred at that point. Once the congestion has been detected,

the RFID controller can send a message to its preceding signal's controller notifying it to temporarily stop traffic along that stretch.

After receiving the message from its successor signal the RFID controller will put ON the red signal for that stretch towards that congested crossing point for a predefined time period. When the congestion is released at the crossing, the respective signal's controller will send another message to its earlier controller indicating to resume the traffic flow again in that direction. Accepting this message the controller of the preceding signal put the red light OFF and green signal ON and restart the signal cycle as before.

IV. APPLICATIONS

A. Detection and Management of traffic Congestion

In addition to the earlier method of traffic congestion detection, one more method can be used. A server can be maintained which can receive certain crucial data calculated by the Controller of the signals. The main aim is to implement a system that would trace the travel time of individual cars as they pass the roadside controllers and compute an average trip time using a rule-based system to decide whether the area is congested or uncongested. If congestion is sensed then system would control traffic signals / generate automatic re-routing messages to selected approaching vehicles.

C. Automatic Billing of Core Area / Toll Charges

Automatic toll collection and automatic —core area charge collections are also done using the same framework. Controller unit will be placed at toll-booth and along the motor able roads around the core area which will detect each individual vehicle uniquely within its zone by capturing their device ids and will keep records of the time during which the vehicle was seen by those Controllers within its reading zone. This information will be sent to a main server. Accordingly the main server will calculate the charges and raise bills against the vehicle ids [14].

V. CONCLUSION AND FUTURE WORK

The proposed work focuses on Smart Traffic management System using RFID which will eliminate the drawbacks of the existing system such as high implementation cost, dependency on the environmental conditions, etc. The proposed system aims at effective management of traffic congestion. It is also cost effective than the existing system. Furthermore, the study presents the problems in metropolitan areas all over the world caused by congestions and the related sources. Congestions developed to a problem, which affects economies worldwide. Particularly metropolitan areas are worst hit under these conditions. Congestions have a negative impact on the financial situation of a country, on the environment and hence the overall quality of life. The proposed system can be enhanced by using any other powerful communication network other than GSM. penalty will be calculated in the server and billed monthly to the vehicle owner [14].

REFERENCES

- [1] 21st Century operations Using 21st Century Technologies. U.S Department Of transportation Federal Highway Administration.2008-08-29. Retrieved 2008-09-25. <http://www.ops.fhwa.dot.gov/aboutus/opstory.htm>
- [2] William Beaty. Jan 1998. Traffic Waves —Sometimes onedriver can vastly improve traffic! <http://www.amasci.com/amateur/traffic/traffic1.html>
- [3] Dipak K Dash, TNN May 31, 2012. —India loses Rs60,000 crore due to traffic congestion: Study. Times Of India. http://articles.timesofindia.indiatimes.com/2012-05-31/india/31920307_1_toll-plazas-road-space-stoppage
- [4] Azeem Uddin, Draft, 23 March 2009. Traffic congestion in Indian cities: Challenges of a Rising power. http://www.visionwebsite.eu/UserFiles/File/filedascaricare/Scientific%20Partners,Papers%28Kyoto%29/Draft_ko_c_Azeem%20Uddin.pdf
- [5] FHWA-HRT-06-108. October 2006. Traffic DetectorHandbook: Third Edition—Volume I. http://www.fhwa.dot.gov/publications/research/operations/its/06108/US7245220_B2. Jul 17, 2007. Radio frequency identification (RFID) controller. <http://www.google.com/patents/US7245220>
- [6] Ali, S.S.M.Indian Inst. of Technol. Madras, Chennai, India, George, B.; Vanajakshi L.: A simple multiple loop sensor configuration for vehicle detection in an undisciplined traffic Sensing Technology (ICST), 2011 Fifth International Conference21568065. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6137062&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D6137062
- [7] FHWA-RD-96-100. July 1995. Detection Technology: IVHSVolume1.<http://ntl.bts.gov/DOCS/96100/index.html> Bing-Fei Wu. Dept of Electr. And control eng, Nat. Chiao Tung Univ, Hsinchu, Taiwan. A new Approach to Video-based Traffic surveillance using fuzzy hybrid Information Inference Mechanism.March2013.IEEE IntelligentTrafficSociety.<http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6264098>
- [8] Xianbin Cao ; Sch. of Electron. & Inf. Eng., Beihang Univ., Beijing, China ; Changxia Wu ; Jinhe Lan ; Pingkun Yan : Vehicle Detection and Motion Analysis in Low-Altitude Airborne Video Under Urban Environment Circuits and Systems for Video Technology, IEEE Transactions on (Volume:21 , Issue: 10) 1051-8215http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5955106&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D5955106
- [9] Hussain, T.M. ; Dept. of Electr. Eng., City Univ. of New York, NY, USA ; Saadawi, T.N. ; Ahmed, S.A.: Overhead infrared sensor for monitoring vehicular traffic:Vehicular Technology, IEEE Transactions on (Volume:42 , Issue: 4) 0018-9545
- [10] <http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=26076>
- [11] <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=260764>
- [12] Bichlien Hoang,Ashley Caudill: EEE Emerging Technology portal, 2012 <http://www.ieee.org/about/technologies/emerging/rfid.pdf>
- [13] http://en.wikipedia.org/wiki/Radio-frequency_identification
- [14] Wang Hongjian ; Chongqing Special Equip. Quality Safe Inspection Center, Chongqing, China ; Tang Yuelin ; Li

Zhi RFID Technology Applied in Highway Traffic Management 978-1-4244-8683-0. Optoelectronics and Image Processing (ICOIP), 2010 International

Conference on (Volume: 2)
http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5663108&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D5663108

Study on types of Trojans in Security

Veena Kulkarni

Thakur College of Engineering and Technology, Mumbai, India
veena.kulkarni@thakureducation.org

Abstract— the world is utilizing digitization in different fields like ecommerce, communication, social media, preservation of library media including the audio, video and text, artificial intelligence etc. The future step of this is to secure the generated data from attackers with malicious intentions. There are various attacks conducted on the modifying/impacting the integrity, privacy and authenticity of data. Trojan is a virus which gets downloaded by any rogue program and silently sits in the computer, coming up for an opportunity to attack the machine/software. This paper studies the range of attacks caused by Trojan virus, its impact and the ways of prevention of this virus.

Keywords—virus, Trojan, security attack.

I. INTRODUCTION

‘Trojan horse’ is the technique used in the past to attack the city of Troy. The name is given to the virus as it uses the same method to attack any computer. Trojan horse is considered as one of the most severe threat to computer security. A Trojan is an executable file which is downloaded and resides in the operating system. They can be used by cyber-criminals as well as hackers to try and gain access to users' systems. Users are naturally tricked into downloading and executing malicious files containing Trojan virus on their systems. Once they are installed and start operating, they can help hackers to spy on the user, steal their important data, and get backdoor access to their system.

Two main types of Trojans are identified:

1. Time bomb and logic bomb- they get executed at specific dates and when a specific event/logic occurs.
2. droppers- they are appearing as legitimate programs however they install Trojans/virus in the system.

Trojan horses get downloaded on your computer through various means. Some of them are listed below:

1. **Game downloads:** Games downloaded from some compromised sites can cause Trojans to enter your computer. The user is unaware as it does not immediately show any impacts.
2. **Social engineering:** The users are tricked to believe that the nasty program is legitimate and they have to download / open the files.
3. **Attachments in Email:** The attacked can send spam mail which consists of content which tries to catch the attention of the user, like prize declaration, discounts, lottery etc, which tempt the user to open the file and thus provide entry for the Trojan.

4. **Exploits:** The patches sent by the owner of the code to fix the bugs can be compromised and send Trojans.
5. **Loopholes in the code:** if the loopholes in the code /Operating Systems are not identified and fixed they might make the computers vulnerable to attacks.
6. **The Trojan can hijack a session by using IP spoofing** or it can perform a man – in – middle attack by sniffing the packets and then inserting malicious packets in between a secure communication.

II. WHAT CAN TROJANS DO

Trojans can affect /impact the computer in different ways:

1. Trojans can delete / modify/ copy or block the data residing in the computer
2. It can open the FTP port no 21 and allow the hacker to penetrate the computer using the file transfer protocol.
3. It can help the attacker to include more computers to create a network of bots by sending the email address from the computer
4. It can steal the sensitive data i.e. the login ids, password, credit card numbers, email addresses from the victim and provide to the attacker.
5. They might disturb the performance/ working of computers and/or networks.
6. Trojans can affect the systems performance and ask for ransom to the users for restoring to the previous performance of the system.
7. They can track the data entered by the user using the keyboard, also take screen shots or get a list of running applications.
8. They can help in performing a DDOS attack by sending spam mails to the victim computer thus overwhelming it with mails, so that the important messages cannot be received.
9. They can activate a system's webcam and record videos. Also if a microphone is there they can record conversations.

A Remote Access Trojan or (RAT) is a Spyware or type of malware, which is used to control Windows OS remotely. RATs are available commercially (e.g. Dark Comet, Poison

Ivy, Havex etc) and can be unknowingly installed on endpoints using phishing and spear-phishing methods. [6]

Trojan horses can be written in all programming languages, which include MS-DOS batch programs, Delphi, BASIC and other languages. The goal is to write a Trojan which can install itself cautiously without being detected and install itself on the victim's machine. The two mostly used programming languages for writing RATs are C,C++ and Delphi, because both languages can create small programs that can be stored in a single executable file. There are many RATs which are popular for damaging the system, some are Back Orifice, Netbus and SubSeven or Sub7.

A. Securing against trojans

The subsequent steps can be taken for avoiding Trojans:

1. Keep antivirus software updated from time to time
2. To steer clear of downloading/ Installing programs or downloading /opening attachments that aren't from a trusted source.
3. At the administrative level, steps can be taken for blocking the idle ports, turning off the unused services. If the port number matches a Trojan port, and the port is open there are chances of a RAT infection.
4. Installing a firewall.
5. Install an anti-Trojan horse programs which contain a database of signatures that are unique to certain Trojan horses.
6. Users should not open links coming as attachment for a website, as they can be targeted for a phishing attack.
7. User should have regular backups of the file/data to ensure minimum loss of data.
8. User should enable pop up blockers as they are the most common methods to install the Trojan.
9. Network administrators should constantly monitor outgoing traffic, even when systems are not in use.

III. HOW TO DETECT RAT INFECTIONS

There are many RATs available in the market, the most commonly used are Back Orifice , Subseven and Netbus.

When user system gets compromised it may show following cases:

1. If the computer is running slowly or constant pop up advertisements appear when user opens the browser, or there is any change noticed in homepage.
2. Someone has logged in to the user system and changed the password, thereby locking the user out of the system. Fake messages and emails are sent from the user accounts to the clients / people registered with the user.
3. The company is targeted for DDOS attack; it receives spam mails in large numbers from various places and thus cannot send any important mails.

4. The outgoing traffic increases, even when no one is working on the machine.
5. The database is compromised and client data/ email addresses get deleted.

Back Orifice is actually a program which is designed to expose the security flaws of Microsoft's Windows operating systems[2].It was created by a group of hackers, they are called Cult of the Dead Cow. Back Orifice allows controlling of any remote computer, running Windows. It can steal passwords, record the keystrokes, and access a file system and much more, while remaining undetected to the user. It can be used as remote administration tool as well as for hacking purposes. Figure 1 shows home page for the Back Orifice.

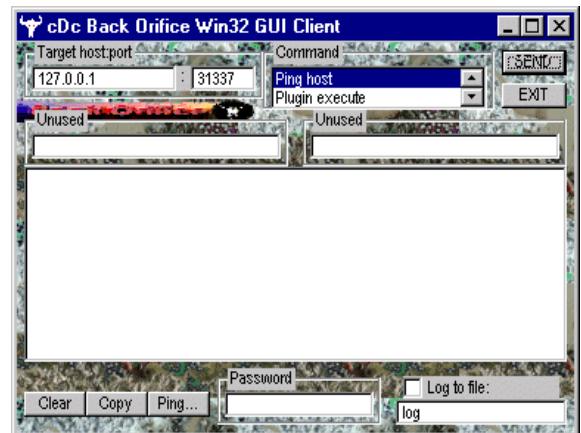


Fig. 1. Back Orifice page[7]

Netstat can be used to detect if Back office virus is present on your computer. Netstat is an utility that comes inbuilt with many operating systems, including Windows. It displays all listening and active ports.

Open a DOS command prompt and type netstat -a command

```
C:\Administrator>Administrator: C:\Windows\system32\cmd.exe
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. All rights reserved.

C:\Users\Administrator>netstat -a
```

Fig. 2. Example of netstat command

This command will give LISTENING, ESTABLISHED and TIME WAIT connections. If the results reveal that there is a

connection with a port numbered 31337 then Back Orifice is present on that machine. The Netstat command however has a weakness, it does not tell which programs or files are initiating the activity to find which files are initiating the activity a port enumerator can be used. The process of tracing an open port to its causative agent is called port enumeration (or port mapping)[3], thus by using a port enumerator the process can be mapped to an open port. The Netstat utility present in Windows XP includes a new -o parameter which displays the process id associated with the port. The process id can be identified in the Task Manager with the specific program.

Sub Seven or Sub7 as shown in figure 3, is a Trojan program which is a malware designed by mobman. It performs all operations to be performed by a Trojan e.g., help in conducting a back door attack, deleting files, controlling a machine etc. Sub 7 requires its server program to be installed on the client machine. Once the connection is established that sub7 can continue with its operations.

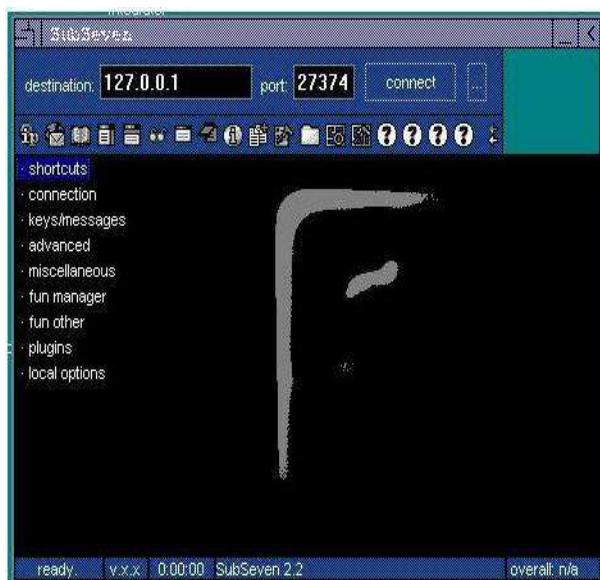


Fig. 3. Sub Seven User Interface [5]

The Netbus trojan is another trojan which works in same way as backorifice does. NetBus Trojan is used for controlling remotely a computer system over a protected network. NetBus was created in Delphi by Carl- Fredrik Neikter.

Netbus can do the following [8]

- Keystroke logging
- Keystroke injection
- Screen captures

- Program launching
- File browsing
- Shutting down the system
- Opening / closing CD-tray
- Tunneling protocol

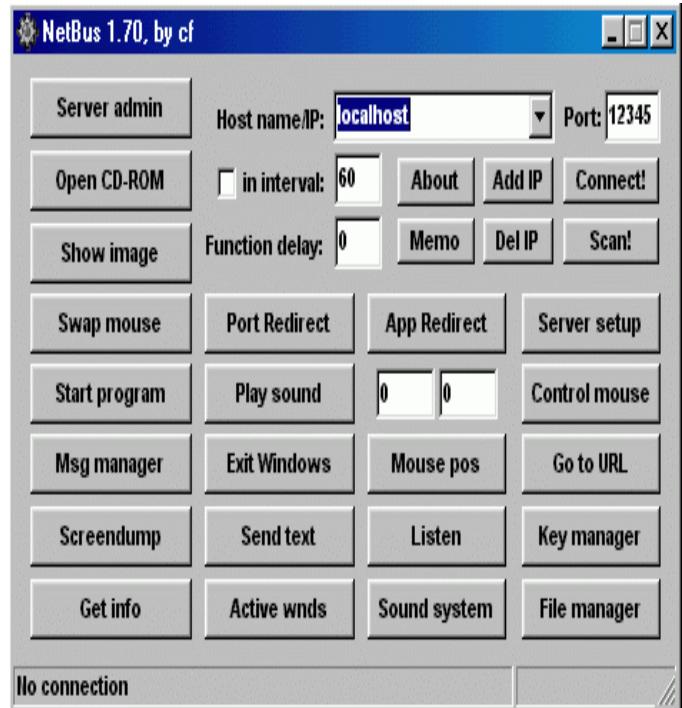


Fig. 4. Netbus User Interface [9]

The figure 4 shows the user interface for a Netbus Trojan. In the UI the options are given for performing the type of operation.

NetBus uses Transmission Control Protocol (TCP) for communication, and it always uses ports 12345 and 12346[10] for listening any connections. The Netstat command can tell whether NetBus is installed.

IV. CONCLUSION

Trojan horses attack/enter the system under cover in disguise without being detected. They allow the hackers to access the system and control the system. Some Trojans install new programs in the systems while some change the programs as per their requirement. Both types of virus can cause damage to the victim system as well as the user data. Thus the user has to be further careful about the authenticity of the file/program which they are downloading /installing in their systems. In this paper three Trojans are studied.

V. REFERENCES

- [1] <https://itstillworks.com/do-trojan-horses-computer-6526732.html>
- [2] <http://searchmidmarketsecurity.techtarget.com/definition/Back-Orifice>
- [3] <http://www.itprotoday.com/security/11-port-enumerators>
- [4] <https://www.kaspersky.co.in/resource-center/threats/trojans>
- [5] https://www.symantec.com/security_response/writeup.jsp?docid=2001-031420-1227-99&tabid=2
- [6] <http://www.trusteer.com/en/glossary/remote-access-trojan-rat>
- [7] <https://www.f-secure.com/virus-info/v-pics/bo.gif>
- [8] <https://en.wikipedia.org/wiki/NetBus>
- [9] http://techgenix.com/the_netbus_trojan/
- [10] <http://www.omninett.net.au/advise8.htm>

Billing system using raspberry-pi

Shivam Mishra¹, Ritesh Mishra², Bharat Joshi³, Rashmi Thakur⁴

Department of Computer Engineering, Thakur College of Engineering and Technology, Mumbai, India

Abstract - Billing systems are developed to reduce day to day manual works and helps the shopkeeper to manage sales, incoming and outgoing stocks. Such a software's are usually developed on desktop systems which many times suffer with overhead parameters like cost as they require the computer systems for users to work on. Also they require a huge space for storage. EBM's takes less space but they are not much efficient as they lack memory and processing speed. Thus we are aiming to develop a billing system using python and raspberry-pi in order to reduce cost and storage space with high efficiency. We are also adding some useful features like Zeroise function, receipts and issues entry.

Keywords : *EBMs, Python, Raspberry-pi, Zeroise function, receipts, issue.*

I. INTRODUCTION

In today's fast moving world data management is a very important task. Each and every organization needs their data not only to be stored in a proper format but also wants to analyze the data in a way to get to some conclusion. When raw information is stored in a proper format then it is called as data and when that data is processed and some conclusion is derived then that conclusion is called knowledge. In today's world knowledge is the most important factor to grow in business. Every business has to store some records in some form. All these records helps owner in many things like - Knowing the availability of product, Verifying the customer's bill (in case of return or changing product). Understanding the pattern of sales of particular product, Understanding the area of scope, To calculate future requirement of products.

A. Background

It all started with keeping all the records in writing in form of books but that was not the most efficient way of keeping records and making bills. After invention of computer, the scenario changed and it gave the efficient method of keeping bills. In last few years the technology of billing is became even better by analyzing the records to give the probability of selling a particular product. In this domain, with time the human efforts required to make bills or understand pattern of sales or understand records etc. is reduced hugely. Recently Amazon launched a just walk out technology which uses Computer Vision, Deep learning algorithms & Sensor Fusion to make first mall which works automatically without any human help. All the part of billing, storing the bills, receipt and cancellation is done automatically through customers amazon account.

The project "Billing system using Raspberry-pi" is a solution to all those problems. It automate the process of ordering and billing

of a store. This application is designed considering the need of small scale stores. The main focus in development of project is to provide system which can do all the work needed with very friendly interface and in a budget of our targeted users. This application also administrates its users and maintains records of product and customers. It is cheap, easy to understand and efficient

B. Problem Definition

Many small scale businesses do not use digital way of storing data and billing. They follow traditional way of keeping records due to which their management of business is not efficient. Some uses EBMs (Electronic Billing Machines) which is faster and better than traditional method. Generally these system does not analyze the data hence results in restricted growth of a business. Our billing system stores the data in a representable manner which can later be used for analysis which will be beneficial for the shop. It avoids the chances of data loss, human error, improper verification etc.

II. EXISTING TECHNIQUES

There are various techniques used by different organizations. POS(Point Of Sale) systems are used in this domain of accounting.

TABLE I.: VARIOUS TECHNIQUES USED BY DIFFERENT ORGANIZATIONS.

Company	Software	Features
D-Mart	R-Soft(Retail Software) by Yash System[1]	Multiple user id, storing data online, cash register tracking.
McDonald's	NP6(New Pos 6) by Savista[2]	Touchscreen, time sensitive with offering meals, multiple user id.
BEST (MSRTC)	BTM102[3]	User id, selecting route, RFID system to scan the passes
Burger King	MICROS by Oracle [4]	Same as R-soft with numerous languages

There are so many others POS systems with different features as per requirement of client. Some of the POS are TouchBistro, Zomato Base, Clover, ShopKeep, eZeeBurrP!, AccuPOS, Epos Now, Ordyx, Cake, Vivonet, Brink POS, PixelPoint POS.

III. LITERATURE SURVEY

TABLE II.: LITERATURE SURVEY OF PAPERS

Sr No	Author Name	Year	Title of the Paper	Summary/Conclusions	Gap Identification
1	Nada Eissa and Andrew Zeitlin with SaahilKarpe and Sally Murray	November 2014	Incidence and Impact of Electronic Billing Machines for VAT in Rwanda[8]	This paper analyzes the adoption rate and tax compliance impacts of an innovative program in Rwanda, which introduced Electronic Billing Machines to strengthen VAT compliance.	It made the process of billing faster but data was not storing data, so process of returning or refunding the product cannot be verified with proper resources.
2	Victor Steenbergen	May 2017	Reaping the benefits of Electronic Billing Machines: using data-driven tools to improve VAT compliance[9]	Using examples from Rwanda, this paper argues EBMs can have transformative impacts for VAT compliance, but only when combined with data analytics and receipt audits to enforce EBM receipt issuance. This can be done using a three-step approach: initial ‘benchmark audits’ can establish firms’ true sales patterns; data analytics then flag firms deviating from this pattern, which trigger automatic ‘mystery shopper’ audits to verify non-compliance and sanction the firm.	It gave the problem with the EBMs and explained the need of area but did not give the exact method to achieve that in budget.
3	Sonali S. Lagu & Prof. Sanjay B. Deshmukh	June 2015	Raspberry Pi for Automation of Water Treatment Plant[10]	This paper focuses on an innovative and intelligent control and monitoring system for Water Treatment Plant by using “Raspberry Pi” as an effective alternative to PLCs for the automation of small water treatment plants. Raspberry Pi is a minicomputer which has an ability to control the system comes with advantages like low cost and compact size.	This project was focused on water treatment and we learned that we can use similar technology to solve the problem of Billing Domain.
4	ANIL KUMAR ANJANA ,BHAIYA LAL ANJANA & SAURABH JAIN	May 2016	Billing System[11]	The project “Billing system” is an application to automate the process of ordering and billing of a “Departmental store”. This web based application is designed considering the chain of departmental store which is located in various cities. This application also administrates its users and customers.	It is a web based system so need a internet support always in order to provide efficient services.

A. Some case studies on Electronic Billing machines :

- In a Case Study on KSRTC by the World Bank[12] it shows the great effect in terms of economy as well as ease of using when Karnataka state started using the ETMs (Electronic Ticketing Machines) provided by company called MicroFX. They started with the model BTM100 which was simple ticketing machine but with time they learned a lot and made BTM101, BTM102 & BTM103. BTM102 had a cutter with printer shears the ticket after printing. In BTM103 GSM and GPRS modules were introduced & RFID to check the cards.[13]
- In a Case study on MSRTC[14] by CIO it shows that Maharashtra government saved 25 lakhs per day by using the ETMs.
- From 2009 San Francisco International Airport(SFO) started using Paperless boarding pass system and employed the barcode scanning system which reduced the cost effectively and also saved the paper.
- Rajasthan Electronics & Instruments Limited(REIL) also initiated the use of Billing machines by the name of SBMs(Spot Billing Machines) with High Speed CPU ,256 KB program memory , 4 MB data memory , Real Time Clock , LCD display unit (4 rows x 20 character) with backlight ,30 Keys soft silicon rubber Keypad , High speed serial port / program port (115 Kbps) , Built-in 24 column Impact printer , Built in 2.2 AH Lithium-Ion or 1.7 AH Ni MH battery , ABS plastic case housing , Programmable as per user requirement , Light weight, versatile and affordable in June 2016.[15]

IV. PROPOSED WORK

The above block diagram shows the different components of the system and their connections.

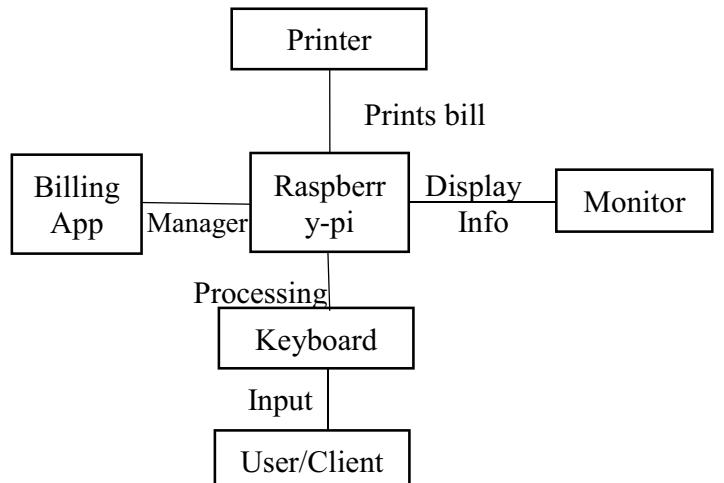


Fig 1. different components of the system and their connections.

The main goal of this project will be to develop an easy, cheap and effective desktop application with very friendly interface.

The basic steps to use this application are as follows:

1. Store the detail credentials(Name, Address, Contact)
2. Store the tax structure(define tax rates)
3. Set the hierarchy of users (Owner, Manager, Clerks etc.)
4. Set passwords for different users
5. Store product details (product id, availability etc.)

The basic functionality this application will have:

1. Generate the bill for customers(using product id)
2. See all stored bill records(Tabular & graphical representation)
3. See all products availability(Tabular & graphical representation)
4. Zeroise function (Day-End button).

A. Technology :

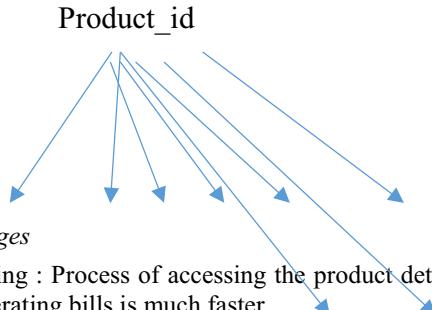
Raspberry Pi:-The Raspberry Pi is a low cost, credit-card sized computer that plugs into a computer monitor or TV, and uses a standard keyboard and mouse. It is a capable little device that enables people of all ages to explore computing, and to learn how to program in languages like Scratch and Python. It's capable of doing everything you'd expect a desktop computer to do, from browsing the internet and

playing high-definition video, to making spreadsheets, word-processing, and playing games.[5]

Python 2.7 :- It's a dynamic object oriented programming language. It has efficient high-level data structures and a simple but effective approach to object-oriented programming. Python's elegant syntax and dynamic typing, together with its interpreted nature, make it an ideal language for scripting and rapid application development in many areas on most platforms.[6]

JSON :-JSON (JavaScript Object Notation) is a lightweight data-interchange format.. JSON is a text format that is completely language independent but uses conventions that are familiar to programmers of the C-family of languages, including C, C++, C#, Java, JavaScript, Perl, Python, and many others. These properties make JSON an ideal data-interchange language.[7]

Some structures of storage system is shown below:



C. Advantages

- Fast billing : Process of accessing the product details and generating bills is much faster.

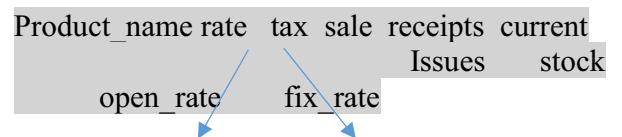


Fig 2: Schema for product

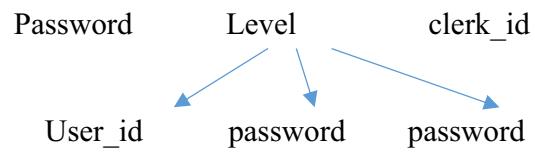


Fig 3 Schema for use

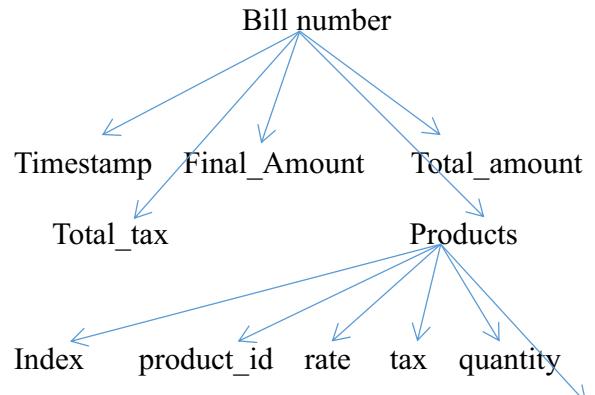


Fig 4 Schema for Bill Number

B. Project Impact Analysis

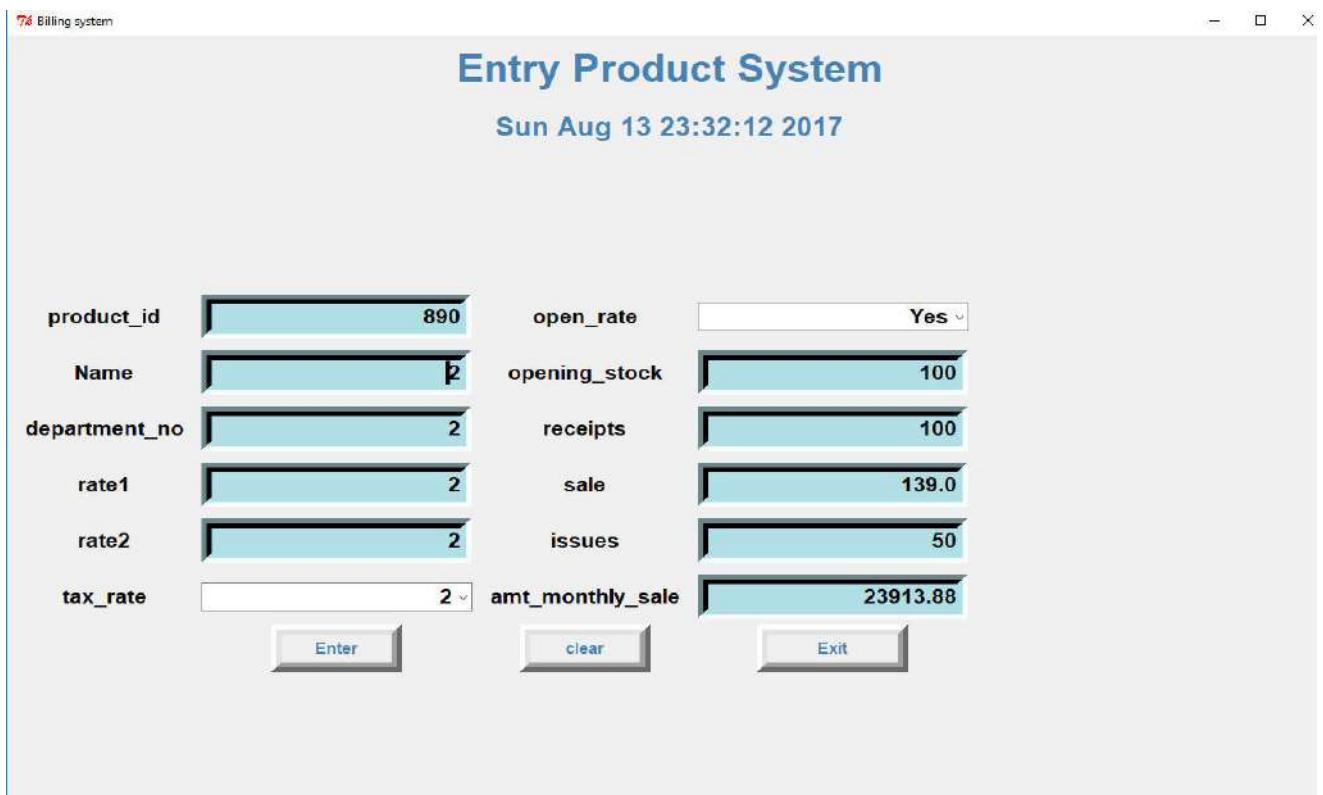
- Social Impact : The project will give chance to many small scale businesses to grow, and it will also increase the digitalization.
- Ethical/Legal Impact : The project will result into proper storage and mapping of data which will reduce the chances of fraud and also help government verifying the taxes.
- Environmental Impact : The project will result into less use of paper.
- Financial Impact : The developed project will be cheaper than the present similar systems will be feasible for small scale industries.
- Enable shop owner to keep record of products as well as sale : All the details of products as well as records of sales are stored on the system.
- Enable to find pattern of sales :An algorithm is build to predict the pattern of sales.

- Easy to maintain : All the feature and properties are very easy to access.
- Basic system doesn't require internet connectivity
- Provide daily financial reports : At the end day it gives the full report of product sale, issues,receipts,amount earned etc.
- Zeroisefunction :This function changes the value of parameters which are responsible for storing the data day-wise. Like Amount_earned, Availability, receipts, issues, opening_stock etc. the value in available parameter is given to opening stock for next day. For better understandability of zeroise function let us take an example there is a Chinese restaurant, generally in any billing system the end of the daily calculation is

calculated at the 12 AM. But in this kind of system the problem which occurs is, if the restaurant is open till late night after 12 AM the calculation will be done till 12 am only and the calculation after that will be counted in next day which can create problem for the restaurant to maintain the record. So to overcome this problem we are using zeroise function, with the help of zeroise function we have the option to close the day whenever we want. This will lead to smooth calculation and there will be no problem in maintaining the record.

- Easily portable: As the whole system is light and portable it can be used at all places.

V. RESULTS& DISCUSSION



The screenshot shows a Windows application window titled "Billing system". The main title of the application is "Entry Product System" and the current date and time displayed are "Sun Aug 13 23:32:12 2017". The interface contains several input fields and calculated values:

product_id	890	open_rate	Yes
Name	2	opening_stock	100
department_no	2	receipts	100
rate1	2	sale	139.0
rate2	2	issues	50
tax_rate	2	amt_monthly_sale	23913.88

Below the table are three buttons: "Enter", "clear", and "Exit".

Fig 5. Entry system

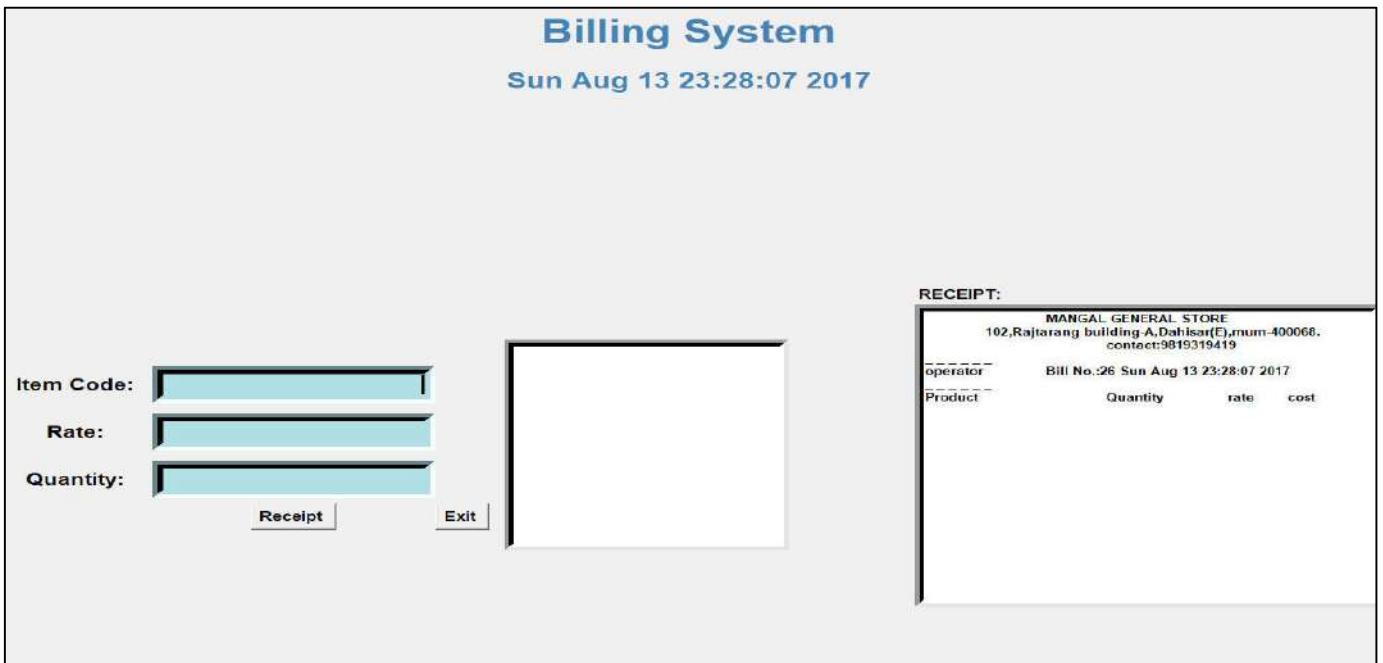


Fig 6. Billing system

BILLING SUMMARY				
<input type="text"/> <input type="button" value="Search"/>				
BILL DATA				
Bill no.	Date	Total_Tax	Total_Amount	Final_Amount
11	Thu Jul 06 02:07:18 2017	41.4	230.0	271.4
12	Thu Jul 06 23:53:20 2017	0.24	68.0	68.24
13	Thu Jul 06 23:55:47 2017	0.72	4.0	4.72
14	Thu Jul 06 23:56:08 2017	1.08	6.0	7.08
15	Thu Jul 06 23:57:40 2017	9.72	54.0	63.72
16	Fri Jul 07 00:01:12 2017	49.14	273.0	322.14
17	Fri Jul 07 00:01:55 2017	8.28	46.0	54.28
18	Fri Jul 07 00:06:28 2017	10.56	52.0	60.56
19	Fri Jul 07 00:08:13 2017	13.5	75.0	88.5
20	Fri Jul 07 00:09:49 2017	4.5	25.0	29.5
21	Fri Jul 07 00:11:45 2017	38.52	214.0	252.52
22	Fri Jul 07 00:14:03 2017	28.5	160.0	188.8
23	Sat Jul 08 00:00:00 2017	3029.88	20165.0	23195.88
24	Sat Jul 08 12:09:26 2017	0.28	46.0	54.28
25	Sun Jul 09 21:28:17 2017	9.0	50.0	59.0

Fig 7. Billing summary

VI. FUTURESCOPE

- As this project is mainly focused on reducing cost and provide interface that can be used by local stores and dairies, so lot of advancements are possible like adding barcode scanner, communication between devices etc.
- The thing that we have planned that will be useful and also in budget of local shopkeepers. We will make a website and connect all the machines to cloud. Every shop owner have to create their profile on our website and then link their system to our server. We will analyze the data on big scale (data from all the

shopkeepers from all the regions) on various parameters like products demand based on region, season, cost etc. we will also suggest the shopkeepers to introduce the new products which are in demand in market.

- Adding an algorithm based on the nature of sales which will calculate the number of items shopkeeper should purchase in advance.
- Not suited for large scale businesses: As large scale industries have data in a huge amount which can be managed more efficiently by distributed systems. Our system is standalone system for now, so handling data in a very huge amount is not feasible.

- Doesn't learn the pattern of ordering (in current version): In advanced billing systems the system itself suggests the product according to sales pattern. For

example there is a sporting goods store, where the number of cricket balls sold is much higher than other sporting goods. So, whenever shopkeeper starts billing, the moment he/she hit the 'B' button the first suggestion will be of balls rather than other sporting goods like bat, baseball, basketball, badminton etc.

VII. CONCLUSION

We have tried to develop a system that can be a great help for the owner of the small scale industries to receiving bill from the customer. Thus we have implemented our project "Billing System using Raspberry-pi" which is better than traditional register system as it is simpler, cheaper and more efficient. This project will provide automated taxation and calculation, bill generation, summary report, bill and stock report.

REFERENCES

- [1] <http://www.yashsoftware.in/>
- [2] <https://www.expertmarket.com/pos/mcdonalds-pos-system>
- [3] <http://www.microfx.co.in/products.html>
- [4] <https://www.oracle.com/corporate/acquisitions/micros/index.html>
- [5] Raspberry Pi User Guide Book by Eben Upton and Gareth Halfacree
- [6] Learn Python the Hard Way: A Very Simple Introduction to the Terrifyingly Beautiful Book by Zed Shaw
- [7] Beginning JSON, Book by Ben Smith
- [8] Incidence and Impact of Electronic Billing Machines for VAT in Rwanda by Nada Eissa and Andrew Zeitlin with Saahil Karpe and Sally Murray in November 2014
- [9] Reaping the benefits of Electronic Billing Machines: using data-driven tools to improve VAT compliance by Victor Steenbergen in May 2017.
- [10] Raspberry Pi for Automation of Water Treatment Plant by Prof. Sanjay B. Deshmukh in June 2015.
- [11] Agrawal, P. (2008) 'Performance Improvement Planning: Developing Effective Billing and Collection Practices', Field Note 44119, Water and Sanitation Program - South Asia, World Bank
- [12] <http://bestpedia.in/2015/07/electronic-ticketing-systems-1/>
- [13] <https://www.ssatp.org/sites/ssatp/files/publications/Toolkits/Fares%20Toolkit%20content/case-studies/karnataka-india.html>
- [14] <http://www.cio.in/case-study/msrtc-saved-rs90-crores-electronic-ticketing-system>
- [15] <http://www.microfx.co.in/products.htm#a>

Enhanced Classification Framework

Rishi Shah¹, Chandrashekhar Bora², Amit Mehra³, Yash Kaul⁴, Harshali Patil⁵

Computer Engineering, Thakur College of Engineering and Technology, Mumbai, India

Abstract— Now-a-days social networking sites are thriving, so large amount of data is generated. Millions of people are sharing their views daily on micro blogging sites, since it contains short and simple expressions. In this paper, we will discuss about a paradigm to extract the sentiment from a micro blogging website, Twitter, where users post their opinions for everything. In this paper, we will discuss the existing psychiatry of twitter dataset with data mining approach such as use of Sentiment analysis algorithm using machine learning algorithms. An approach is introduced that by design classifies the sentiments of Tweets taken from Twitter dataset as in [1]. These messages or tweets are classified as positive, negative or neutral with respect to a question term. This is very useful for the companies who want to know the feedback about their product brands or the customers who want to search the opinion from others about product before purchase. We will use machine learning algorithms for classifying the sentiment of Twitter messages using distant management which is discussed in [8]. The preparation data consists of Twitter messages with emoticons, acronyms which are used as deafening labels discuss in [4]. We examine sentiment analysis on Twitter data. The aim of this review paper are: (1) we use Parts Of Speech (POS)-specific prior polarity features. (2) We also use a tree kernel to prevent the need for monotonous feature engineering

Keywords— Micro blogging, Twitter, Sentiment, Classifiers, Sentiment Analysis.

I. INTRODUCTION

We know that there are almost 111 micro blogging sites. Micro blogging websites are nothing but social media site to which user makes short and frequent posts. Twitter is one of the famous micro blogging services where user can read and post messages which are 140 characters in length. Twitter messages are also called as Tweets. We will use these tweets as raw data. We will use a method that automatically extracts tweets into positive, negative or neutral sentiments. By using the sentiment analysis the customer can know the feedback about the product or services before making a purchase. The company can use sentiment analysis to know the opinion of customers about their products, so that they can analyze customer satisfaction and according to that they can improve their product. Sentiment analysis has become one of popular research area in computational linguistics, because of the explosion of sentiment information from social web sites (i.e., Twitter and Facebook), online forums, and blogs as in paper [10]. We are going to use three models explicitly unigram model, tree kernel model and feature based model. Sentiment Classification has been researched for better result. Traditionally, Sentiment classification concentrated for classifying larger pieces of text which includes reviews or feedback. Twitter includes

tweets which are different from reviews. Both Twitter and reviews are differentiated by their purpose. Tweeter's emotion or feeling on particular topic can be express by using tweets. While, summarized thoughts of authors are represented by reviews. On the other hand, tweets are more casual with the limited 140 characters text in length. In paper [1], there is use of two resources : 1) a hand annotated dictionary for emoticons 2) an acronym dictionary gathered from web. The approach is the use of different machine learning classifiers and feature extractors. Naive Bayes, Maximum Entropy (MaxEnt), and Support Vector Machines (SVM) are the machine learning classifiers.

The feature extractors are 1.Unigrams

2. Bigrams
3. Both unigrams and bigrams,
4. Unigrams with part of speech tags .

In paper [1] and [2], one of the best uses of Sentiment Analysis is that the organization knows their own business progress by user's feedback.

Sentiment Analysis is exceedingly domain centered; the submission developed for twitter can't be used for facebook. When looking at Twitter, it is particularly problematic. For example: "The meal was awesome but the service was terrible". In this case, computer gets confused for the result of sentiment.

A. Machine Learning Methods:

There are three different machine learning algorithms who achieved great success for text categorization as in paper [3] which are as follows:

1. NAIVE BAYES:

Naive Bayes model is a simplest model. For the classification of the text this model works well. This statement is called class conditional autonomy. As in [6], it is made to simplify the computation and in this sense considered as "Naive".

Class c^* is assigned to tweet which is denoted by d ,
Where, $c^* = \arg\max_{c \in C} P(c|d)$

$$P(c|d) = \frac{(P(c) \sum_{f=1}^m p(f|c)p(f|c)n_i(d))}{P(d)} \quad (1)$$

From [2]

Parameters $P(c)$ and $P(f|c)$ are obtained through maximum estimates, and add-1 smoothing is utilized for unseen features.

2. MAXIMUM ENTROPY (MAXENT):

This model is Feature based model. MaxEnt do not make any autonomy assumption for its features, consequently MaxEnt is diverse than Naïve Bayes. MaxEnt can handle features overlapping problems better than Naïve Bayes. Stanford classifier is used for classification in MaxEnt model. In practical scenarios different types of problems can be resolved by MaxEnt easily as compared to Naïve Bayes.

3. SUPPORT VECTOR MACHINES (SVMs):

Support Vector Machines are theoretically well motivated algorithms and has been developed from statistical learning theory since the 60s. The class of algorithms called SVMs which are used for pattern recognition. They are effective and famous categorization learning tool. Support vector machines represent an extension to nonlinear models of the generalized representation algorithm industrial by Vladimir Vapnik. The SVM algorithm is based on the statistical learning conjecture and the Vapnik-Chervonenkis (VC) dimension introduce by Vladimir Vapnik and Alexey Chervonenkis. A few methods were devised and analyzed because of centrality of the SVM optimization problem which are discussed in [9].

Using these machine learning algorithms, three models are developed in Weka namely Unigram Model, tree kernel model and feature based model. These models will be used for feature extraction.

As in paper [11] which presents SentiView tool. It is an interactive visualization system and it focuses on analysis of public sentiments for popular topics on the Internet. Uncertainty modeling and model-driven adjustment is combined in SentiView, it mines and models the changes of the sentiment on public topics, by searching and correlating frequent words in text data.

II. PROPOSED WORK

Sentiments are the vocabulary or sentences that symbolize view or estimation that is held or uttered that can be positive, negative or neutral. We are going to propose a novel hybrid approach involving both corpus-based and dictionary-based techniques, which will find the semantic orientation of the sentiments words in tweets. We will also consider features like emoticons, neutralization, negation handling and capitalization as they have recently become a huge part of the internet language.

The proposed Sentiment Analysis on twitter data is based on two vital parts viz Data Extraction, pre-processing of extracted data and categorization.

To uncover the sentiments, we will first extract the opinion words from tweets and then we find out their orientation, i.e., to decide whether each sentiment word reflects exaggerated and self-indulgent feelings of tenderness, sadness, or nostalgia.

Paper [10] shows the overall feature about the background of the classification framework.

The following steps will expound the process of the proposed system which is discussed in paper [2] and [6] shown in fig [1]:

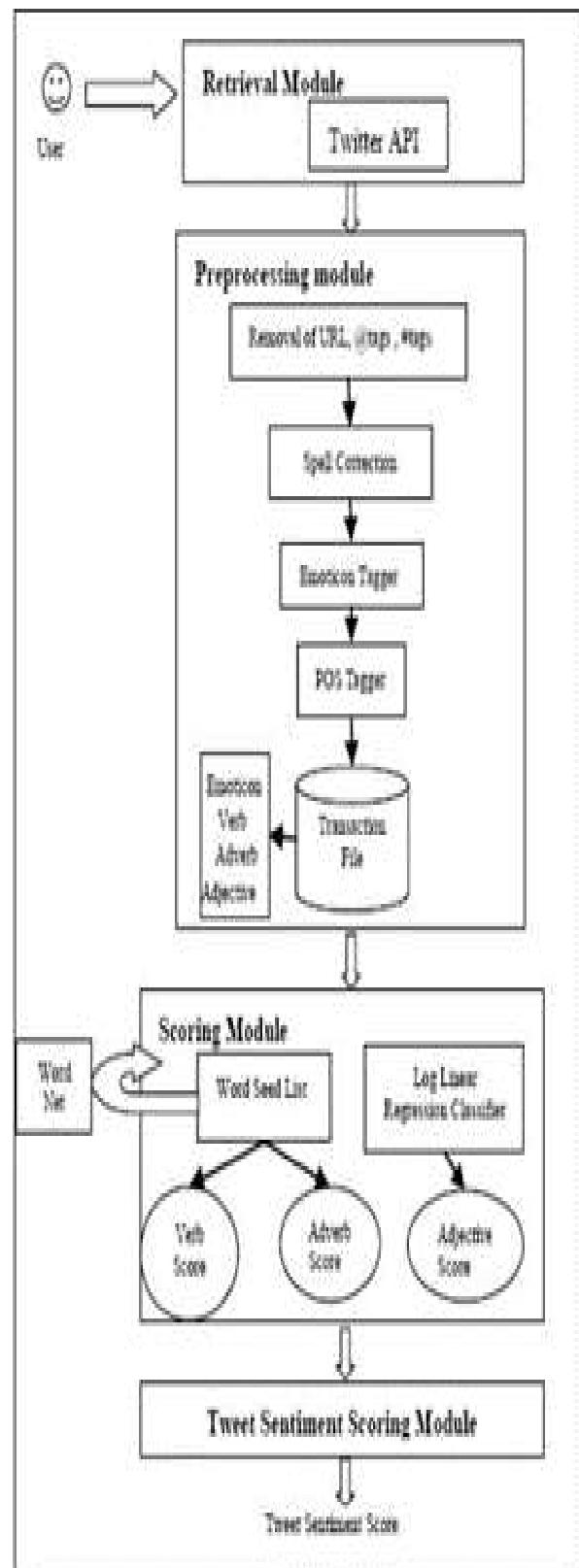


Fig I. Analysis of sentiments

- Retrieval of tweets
- Pre-processing of extracted data
- Parallel processing
- Sentiment scoring module
- Output sentiment

These steps are explained below

A. Retrieval of tweets :

As twitter is the most exaggerated part of social networking site, it consists of various blogs which are related to various topics worldwide. Instead of taking whole blogs, we will rather search on particular topic and download all its web pages then extract them in the form of text files by using mining tool i.e. Weka which provides sentiment classifier.

B. Construction of n-grams:

Negation words such as “no”, “not” is attached to a word which follows or precedes it. For Instance: “I do not like remix music” has two bigrams: “I do+not”, “do+not like”, “not+like remix music”. So the precision of the cataloging improves by such procedure, because negation plays an vital role in sentiment analysis. Paper [3] represents that negation needs to be taken into account, because it is a very common linguistic construction that affects polarity.

C. Parallel processing:

Training of classifier data is the main motive of this step. Every database has hidden information which can be used for managerial. Classification and prediction are two forms of data analysis which can be worn to extract models unfolding imperative data and future trends. Classification is process of finding a set of models or function that portray and tell apart data classes or concepts, for the principle of being able to use the model for predicting the class of objects whose class label is unidentified.

The derived model is based on the analysis of a set of training data. Training data consists of data objects whose class labels are known. The resultant model can be represented in a range of forms, such as organization (IF-THEN) rules, decision trees, mathematical formulae, or neural networks.

Classification process is done in a two step process. First step is Model Construction in which we will build a model from the training set. And step2 is Model Usage in which we will check the accuracy of the model and use it for classifying new data.

D. Sentiment scoring module:

Prior polarity of words is the basic of our number of features. The dictionary is used in [1] in English language words.

E. Output sentiment:

Based on the dictionary assignment of score, the proposed system interprets whether the tweet is positive, negative or neutral.

III. DATA DESCRIPTION

Twitter is a social networking and microblogging service that allows users to post real time messages, called tweets. Tweets are short messages, restricted to 140 characters in length. Due to the scenery of this micro blogging service (quick and short messages), people use acronyms, make spelling mistakes, use emoticons and other typescript that express special meanings. Following is a brief vocabulary allied with tweets. Emoticons: These are facial expressions pictorially represented using punctuation and letters; they express the user’s mood. objective: Users of Twitter exploit the “@” representation to refer to other users on the microblog. Referring to other users in this manner automatically alerts them.

We acquire 26,975 manually annotated Twitter data (tweets) from a profitable source. They have made part of their data publicly available. For information on how to obtain the data, see Acknowledgments section at the end of the paper. They collected the data by archiving the real-time stream. No language, location or any other kind of restriction was made during the streaming process. In fact, their collection consists of tweets in foreign languages. They use Google translate to convert it into English before the annotation process. Each tweet is labeled by a person annotator as optimistic pessimistic, unbiased or scrap. The “junk” label funds that the tweet cannot be implicit by a human annotator. A manual analysis of a arbitrary example of tweets labeled as “junk” not compulsory that many of these tweets were those that were not translated well using Google translate. We abolish the tweets with scrap label intended for experiment. This leaves us with an unbalanced sample of 8,753 tweets. We use stratified sampling to get a balanced data-set of 5127 tweets (1709) tweets each from classes positive, negative and neutral).

IV. RESOURCES AND PRE-PROCESSING OF DATA

We bring in two new resources for pre-processing twitter data: 1) an emoticon dictionary and 2) an contraction dictionary. We prepare the emoticon dictionary by labeling 170 emoticons listed on Wikipedia with their emotional state. For example, “:)” is labeled as positive whereas “:=(“ is labeled as negative. We allocate each emoticon a label beginning the subsequent set of labels: Extremely-positive, Extremely-negative, Positive, Negative, and Neutral. We compile an acronym dictionary from an on-line resource. The dictionary has translations for 5,184 acronyms. For example, lol is translated to laughing out loud.

We pre-process every single one the tweets as follows: a) reinstate all the emoticons with a their sentiment polarity by looking up the emoticon dictionary, b) re-place all URLs with

a tag jjU jj, c) replace targets (e.g. “@John”) with tag jjT jj, d) replace all negation (e.g. not, no, never, n’t, cannot) by tag “NOT”, and e) replace a sequence of repeated characters by three characters, for example, convert cooooooooool to coool. We do not replace the sequence by only two characters since we want to differentiate between the regular usage and emphasized usage of the word. TABLE I. ACRONYM AND ENGLISH EXPANSION

Acronym	English expansion
gr8, gr8t	great
lol	laughing out loud
rofl	rolling on the floor laughing
bff	best friend forever
(Rishi Shahm, 2017)	

We present a quantity of preface figures about the data in Table 3. We use the Stanford tokenizer (Klein and Manning, 2003) to tokenize the tweets. We use a stop word dictionary³ to identify stop words. All the other words which are found in WordNet (Fellbaum, 1998) are counted as English words. We use the standard tagset defined by the Penn Treebank for identifying punctuation. We record the occurrence of three standard twitter tags: emoticons, URLs and targets. The remaining tokens are either non English words (like coool, zzz etc.) or other symbols.

http://en.wikipedia.org/wiki/List_of_emoticons

²<http://www.noslang.com/> ³<http://www.webconfs.com/stopwords.php>

TABLE II: RECORDS ABOUT THE STATISTICS WORN FOR OUR EXPERIMENT

Emoticon	Polarity
:-) :o) :] :c:	Positive
:D C:	Extremely-Positive

:-(:(:c:[Negative
D8 D; D= DX v.v	Extremely-Negative
:j	Neutral

Number of tokens	79,152
Number of stop words	30,371
Number of English words	23,837
Number of punctuation marks	9,356
Number of capitalized words	4,851
Number of twitter tags	3,371
Number of exclamation marks	2,228
Number of negations	942
Number of other tokens	9047

In Table II we see that 38.3% of the tokens are stop words, 30.1% of the tokens are found in WordNet and 1.2% tokens are negation words. 11.8% of all the tokens are punctuation marks not including exclamation marks which build up for 2.8% of all tokens. In total, 84.1% of all tokens are tokens that we expect to see in a typical English language text. There are 4.2% tags that are specific to Twitter which include emoticons, target, hastags and “RT” (retweet). The remaining 11.7% tokens are either words that cannot be found in WordNet (like Zzzzz, kewl) or special symbols which do not fall in the category of Twitter tags.

IV. PRIOR POLARITY SCORING

A numeral of our features are based on preceding schism of words. For obtaining the prior polarity of words, we take motivation from work by Agarwal et al. (2009). We use Dictionary of Affect in Lan-guge (DAL) (Whissel, 1989) and extend it using WordNet. This lexicon of about 8000 English idiom words assign every word a appeal score (2 R) between 1 (Negative) - 3 (Positive). We first normalize the scores by diving each score my the scale (which is equal to 3). We judge words with polarity less than 0.5 as negative, higher than 0.8 as positive and the rest as neutral. If a word is not directly found in the dictionary, we retrieve all synonyms from Wordnet. We then look for each of the synonyms in DAL. If any synonym is found in DAL, we allocate the original word the same amiability score as its synonym. If none of the synonyms is present in DAL, the word is not associated with

any prior polarity. For the given information we directly found preceding polarity of 81.1% of the words. We find polarity of other 7.8% of the words by using WordNet. So we find prior polarity of about 88.9% of English language words.

V. EXPECTED OUTCOME

In this section, we present experiments and outcome for two taxonomy tasks: 1) optimistic versus pessimistic and 2) optimistic versus pessimistic versus unbiased. For each of the classification tasks we present three models, as well as results for two combinations of these models:

- Unigram model (our baseline)
- Tree kernel model
- 100 Senti-features model
- Kernel plus Senti-features
- Unigram plus Senti-features

For the unigram plus Senti features model, we present feature analysis to gain impending about what kind of features are adding together most value to the model. We also present learning curves for each of the mod-els and compare learning abilities of models when provided limited data.

Experimental-Set-up: For all our experiments we use Support Vector Machines (SVM) and account averaged 5-fold cross-validation test results. We tune the C parameter for SVM using an embedded 5-fold cross-validation on the training data of each fold, i.e. for each fold, we first run 5-fold cross-validation only on the training data of that fold for different values of C. We pick the setting that yields the best cross-validation error and use that C for determining test error for that fold. As usual, the reported accuracies is the average over the five folds.

VI. CONCLUSION

We presented results for sentiment analysis on Twitter. We use previously proposed state-of-the-art unigram model as our baseline and report an overall gain of over 4% for two classification tasks: a binary, positive versus negative and a 3-way positive versus negative versus neutral. We presented a comprehensive set of experiments for both these tasks on manually annotated data that is a random sample of stream of tweets. We investigated two kinds of models: tree kernel and feature based models and demonstrate that both these models outperform the unigram baseline. For our feature-based approach, we do feature analysis which reveals that the most important features are those that combine the prior polarity of words and their parts-of-speech tags. We reticently conclude that sentiment analysis for Twitter data is not that unusual from sentiment analysis for supplementary genres.

In future work, we will explore even richer linguistic analysis, for example, parsing, semantic analysis and topic modeling.

REFERENCES

- [1] Luciano Barbosa and Junlan Feng. 2010. vigorous sentiment discovery on twitter from biased and noisy data. dealings of the 23rd International Conference on Computational Linguistics: Posters, pages 36–44.
- [2] Adam Bermingham and Alan Smeaton. 2010. Classify-ing sentiment in microblogs: is brevity an advantage is brevity an advantage? ACM, pages 1833–1836.
- [3] C. Fellbaum. 1998. Wordnet, an electronic lexical database. MIT Press.
- [4] Michael Gamon. 2004. Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis. Proceedings of the 20th international conference on Computational Linguistics.
- [5] Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment categorization using distant supervision. Technical report, Stanford.
- [6] David Haussler. 1999. Convolution kernels on discrete structures. Technical report, University of California at Santa Cruz.
- [7] M Hu and B Liu. 2004. Mining and summarizing customer reviews. KDD.
- [8] S M Kim and E Hovy. 2004. Determining the sentiment of opinions. Coling.
- [9] Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. Proceedings of the 41st Meet-ing of the Association for Computational Linguistics.
- [10] Rishi Shah, Yash kaul, Chandrashekhar Bora, Amit Mehra: Enhanced Classification Framework:IJRASET 2017. (Rishi Shahm, 2017)

Enhanced System for Delivery And Efficient Path Finding

Pratiksha Chourasia¹,ShivamPandey²,Simran Malawat³,Shradha Mhatre⁴,Harshali Patil⁵

Computer Engineering, Thakur College of Engineering and Technology,Mumbai, India

Abstract- The existing delivery/courier system is manual though it could be automated. In the current system, all the data are maintained in books and it is difficult to modify data. Usually, for updating data in current system, user either has to delete or overwrite the data and similarly, to maintain the order, delivery, employee details and other important operations in courier/delivery management requires a lot of manpower and time. Since the system is manual, data is prone to human errors. It becomes practically impossible to do the operations without any error. Nowadays, customers prefer deliveries at their convenient time with no compromises in quality. So, to manage all this becomes very difficult for the employees. The existing system is highly intangible in terms of reliability and ability to perform the promised services accurately with security of records.

Nowadays, 50% of companies of the world use the services of various delivery systems. Today's consumer have become increasingly demanding. All the consumer, employee, management people needs would be fulfilled by the proposed, which is an online software designed for stakeholders of courier/delivery system .This software will enable customers to receive delivery from a source as well

as send delivery to required destination and also track the delivery. We are presenting an application system for efficient Delivery and Maintenance. The application primarily focuses on reducing the burden and discomfort in managing and delivering all the products and services. This will reduce the complications of all the stakeholders such as Delivery Manager, Customer and Delivery Boy for providing services to the concerned personnel. It will provide customers with fast and assured delivery. It will enable the courier delivery boy to locate the fastest and most convenient route for delivery by considering the current traffic situations.The courier manager can evaluate the performance of each of his courier delivery boy through feedbacks from the customer and amount of products delivered per day.

To design this system, we are using various technologies such as web technologies like HTML, PHP for developing web application, Android and Google API for developing application for delivery boy. An android application will be created for helping delivery boy to reach the destination as quickly as possible.

I. INTRODUCTION

In this project, we are presenting an application system for efficient Delivery and Maintenance. The need to design this system is because of the customer's increasing demands of faster delivery and better quality. Also it's a challenge for the distributors to match the speed of the increasing demands and provide efficient service by using the traditional system. So the solution to this is the ENHANCED SYSTEM FOR DELIVERY AND EFFICIENT PATH FINDINGthat modifies the traditional courier delivery system to an automated delivery system. This will reduce the cumbersome of all the stakeholders such as Delivery Manager, Customer and Delivery Boy for providing services to the concerned personnel. One of the biggest challenge is fast and on-time delivery. The system solves this problem by scheduling the delivery route for the delivery boy and locates the fastest and most convenient route for delivery taking into consideration the traffic parameters, location of delivery, etc. The system has been built using web technologies such as HTML,PHP (for developing web application) and Android , Google API (for mobile application for delivery boy).The application primarily focuses on reducing the burden and discomfort in managing and delivering all the products and services. The system will maintain all the records without any ambiguity and can be accessed and modified only after authentication of the user. It will also provide the courier manager to evaluate the performance of each of his delivery boy through feedbacks from the customer and amount of products delivered per day.

II. LITERATURE SURVEY

According to Wikipedia, a courier service company is a company that offers to deliver messages, packages and mails. Couriers are distinguished from ordinary mail services by features such as speed, security, tracking, signature, specialization and individualization of express services, and swift delivery times, which are optional for most everyday mail services. But this system is not efficient, as a messenger can only be allocated to one user. Therefore it is under-utilization of messenger's labour services and other resources.

In this wide and huge world, in order to send and receive goods the most convenient way of transportation is the delivery service. The owner sending goods need not worry much about its parcel's delivery since the delivery company takes care of the parcel's delivery. Though, the delivery company needs to go through a

round of hurdles to transfer one parcel to another efficiently and on time. According to the paper on A Google-Map-Based Arterial Traffic Information System (2007), With the help of live traffic updates, the process of delivery is enhanced and the delivery company can strategize as to where the delivery should be done first in order to save time and deliver more number of goods with respect to traffic parameters. Also, the customer can easily track its goods based on the traffic updates at the comfort of his/her home.

According to Richard E. Korf, path-finding search occurs in the presence of various sources of knowledge such as heuristic evaluation functions, sub goals, macro-operators, and abstractions. With the help of efficient path finding we can deliver the parcels faster and also reduce the time of searching the exact location of the customer. This in turn helps the distributors to deliver more and more number of goods to the intended customers.

According to the paper on Data Mining: Evaluating Performance of Employee's using Classification Algorithm Based on Decision Tree (2014), classification maps data into predefined groups of classes. Performance Analysis helps us model and evaluates a particular employee's performance which in turn can help in deciding the incentives and added bonus along with deciding the best performer of the month and also the profits a particular employee is contributing to the company. This motivates the employees to do better in the future.

III. GAP IDENTIFICATION

There exists a general uncertainty and issue about whether a document is delivered/collected on time.

Mostly delivery companies keep the records in delivery books or files which becomes very difficult to maintain and secure. Maintaining a separate record for each entity is difficult and hence deals with data ambiguity and human errors. Another issue with current system is scheduling the delivery, assigning delivery boy, maintaining bills and accounts, etc. Whereas all this can be computerized which will save human time and efforts. Also, most companies find it very hard to effectively monitor the task of their delivery as it is practically impossible to monitor the delivery time because of the variables involved such as traffic jams, identifying delivery points, weather conditions etc.

TABLE I. CURRENT SYSTEM VS EXISTING SYSTEM

Current System	Existing System
<p>1. The current system will compensate for uncertainties relating to effectively deciding and monitoring the task of their delivery considering traffic jams, accidents and weather conditions.</p> <p>2. It will provide the courier manager to evaluate the performance of each of his courier delivery boy through feedbacks from the customer and amount of products delivered per day.</p> <p>3. It will provide customers with fast and assured delivery.</p>	<p>1. The application primarily focuses on reducing the burden and discomfort in managing and delivering all the products and services.</p> <p>2. Performance evaluation was done manually through the product delivered per day.</p> <p>3. It does not guarantee fast and assured delivery to customers.</p>

IV. PROPOSED WORK

It is very difficult to manage bills and updating of customer information and products because it is done manually. Because of that, many problems occurred such as managing huge amount of data having large amount of data and different data items. Since we are handling large amount of data, the possibility of losing data is very high. If in current system, any problem occurs, then it becomes very difficult to find the solution to that problem and also it takes lots of time for recovery. When sending courier/parcel/product through any courier services, people want to know where their courier/parcel/product is currently and whether it is shifting to the right places or not, if not then how much time it will take to get to the right place and where it is now. Taking all this information manually can be very difficult for the user/admin and keeping this information safe is very difficult. All these operations can be very time consuming. Handling such information requires a lot of paper work and management processes. Most companies find it very hard to effectively decide, schedule and monitor the task of their delivery. Practically it is impossible to monitor the delivery time because of the various factors involved in it such as traffic jams, identifying delivery points, weather conditions etc. The current system will compensate for all such uncertainties relating to effectively deciding and monitoring the task of their delivery considering traffic jams, accidents and weather conditions. The application primarily focuses on reducing the burden and discomfort in managing and delivering all the products and services. The proposed system will be highly reliable in terms of performance and accuracy. It will be error free with the records, tracking details. The system will be user friendly and secure. The overall maintenance of the system will be significantly low compared to the existing system. It will not create any ambiguity in the customer's mind regarding the delivery.

The proposed system will be able to reduce all the disadvantages of current system. It will be able to give efficient tracking details as well as data security. System will be secured as compared to previous system. Data updation will become very easy. Proposed system consist of application for delivery boy from which admin and customer can track delivery boy, admin will be using web application for managing all the

product and customer. In existing system it is very difficult to prepare weekly or monthly report, but proposed system will generate the report automatically on the basis of ratings given by customer so it becomes very easy for admin to classify delivery boy.

V. OBJECTIVES

The objective of this project is:

- To develop an automated software solution for delivery services that provide efficient, error free and a secure system.
- To provide timely delivery of products and enable the customers to check the delivery status through alerts via email, SMS etc.
- To generate the reports as per the requirements.
- To estimate incentive based salaries of employees as per the performance analysis.

A. Abbreviations and Acronyms

SDS- Smart Delivery System

API-Application Programming Interface

UI-User Interface

ESFDAEPF- Enhanced System For Delivery And EfficientPathFinding

VI. SCOPE

This system is developed considering the requirements gathered during the initial phase of the project. The scope includes automating the process and increasing ease of use. The system shall provide customers an ability to track consignment.

System interface shall be user friendly; in an attempt to increase the app usage a “User Guide” shall be available for customers to support in case of any queries related to functionalities. The system shall be able to conduct performance analysis of employees and company. This will help management in deciding career growth and awarding monetary benefits to deserved employees. It will also help in the overall growth and analytics of the distributor company.

VII. EXPECTED OUTCOME

The expected outcome of the system should be as follows:

- The system should be user friendly
- Low cost and easy to maintain the system
- The reports should consist less errors
- The map should give an optimum route to the destination
- Tracking details should be accurate
- To reduce the difficulties of stakeholders of the system
- To provide faster and assured delivery to the customer

VI. IMPLEMENTATION

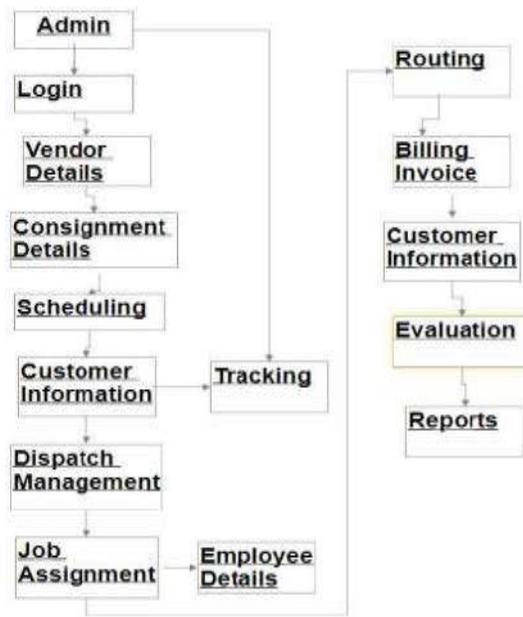


Figure 1 Working

How it works:

1. Admin is a power user in this system. Admin and Employees have login rights. Admin can login in web application whereas Delivery boy can login in android application.
2. When admin logs in, he/she enters vendor details/consignment details and customer information with addresses where delivery needs to be made.
3. The addresses are fetched by the system in a map (which is an android application for delivery boy) format wherein scheduling is done.
4. After the scheduling is done, automatically an invoice is generated and printed.
5. Printed invoice are then attached to each consignment to make it easily deliverable by delivery boy.
6. Then Admin assigns delivery job to the respective delivery boy after dispatch management.
7. Once the packages are dispatched, an automated message is sent to customer with a link. The message contains tracking and feedback link.
8. Admin also has an ability to track the consignment. Billing invoice is generated at the time of label printing.

9. At the end of the day/month an evaluation is been generated containing a list of customer who logged their feedback and waiting for the same.

VII. CONCLUSION

Enhanced System For Delivery And Efficient Path Finding is a solution that has been proposed to eradicate glaring logistic issues that occurs in delivery services. Any unaccredited access to the system would be denied. Using the map application for delivery boy, fast delivery is assured. With the help of feedback mechanism, the system provides an efficient way to evaluate employee's performance which in turn helps the management to be decisive in providing career growth opportunities.

REFERENCES

- [1] E. J. Seymour, B. Miller, "Google-Based Mapping System for Traffic Management Center Web Navigation", presented at the 86th Transportation Research Board Annual Meeting, January 21–25 2007.
- [2] OPTIMAL PATH-FINDING ALGORITHMS*, Richard E. Korf Computer Science Department University of California, Los Angeles Los Angeles, Ca. 90024
- [3] Data Mining: Evaluating Performance of Employee's using Classification Algorithm Based on Decision Tree
- [4] Suo Qi. Database Marketing Research Based on Data Mining [J]. Technology and Market, 2007(2):53-54
- [5] Han Jing, "Application of Fuzzy Data Mining Algorithm in PerformanceEvaluation of Human Resource". Economic and Management Institute, YanTai University, Yantai, Shandong, P. R. China, 264005
- [6] <https://en.wikipedia.org/wiki/Courier>
- [7] [https://en.wikipedia.org/wiki/Delivery_\(commerce\)](https://en.wikipedia.org/wiki/Delivery_(commerce))
- [8] Enhanced System For Delivery Ms. PratikshaChourasia , Mr. ShivamPandey , Ms. SimranMalawat , Ms. ShradhaMhatre , Thakur College of Engineering and TechnologyMumbai, India

Canteen Management System Using Payment Gateway

Dhruvi Bhagat¹, Garima Choudhary², Ashish Kothari³, Rashmi Thakur⁴

Department of Computer Engineering, Thakur College of Engineering and Technology, Mumbai, India

Abstract—The project is intended to give easy access to the students for canteen therefore making it convenient for both students and working staff of canteen. The canteen management system will result in time saving of the students. Our purpose of this project is to reduce the manual work and make canteen more computerized thereby making it more efficient, faster and user friendly. Our attempt is to increase the sales of the canteen by using this software. The existing system is highly manual. It is sluggish and consumes a lot of time causing inconvenience to costumers. The system provides for user-id validation, hence unauthorized access is prevented. Details of all the items is displayed along with the prices and special offers. The students can select whatever items they wish to. The total amount of all the items is displayed before confirmation of order. On selecting the items to be ordered, then confirming the order, the customer will be given an estimated time required for his/her order to be served. The customer can then choose to thus place the final order or not. The system allows the customer to cancel existing order. The ultimate goal of the project is to build a database that integrates the process of ordering food in canteen.

Keywords—easy access, user ID validation, review, automation, payment gateway.

I. INTRODUCTION

In today's age of fast canteen automation in the canteen, many canteens have chosen to focus on quick preparation and speedy delivery of orders. Until now, most of the delivery of orders were placed over the phone, but there are many disadvantages to the system, including the troublesomeness of the customer demanding to have a physical copy of the menu, lack of a visual confirmation that the order was placed correctly, and the necessity for the canteen to have an employee answering the phone and taking orders.

The main advantage of an online ordering system is that it greatly simplifies the ordering process for both the customer and the canteen. When the customer visits the ordering web page, they are presented with an interactive and up-to-date menu, complete with all available options and adjusting prices based on the selected options.

After making a selection, the item is then added to their order, which the customer can review the details at any time before checking out. This provides instant visual confirmation of what was selected. This system also greatly lightens the load on the canteen's end, as the entire process of taking orders is automated.

Once an order is placed on the web page, it is entered into the database and then retrieved, in pretty much real-time, by a web-based application on the canteen's end. In this application, all items in the order are demonstrated, along with their consistent options and delivery details, in a brief and easy to read manner. This allows canteen employees to quickly go through the orders as they are placed and produce the necessary items with minimal delay and confusion..

II. PROBLEM DEFINITION

In the existing system, the most common payment method is through cash. However, when a student orders something for say 67 Rs and gives a 100 Rs note, the canteen has to give back exact 33 Rs in return. In many cases, its possible that the canteen owner might not have enough change to give back the exact amount. In such a scenario, canteen owner maintains a notebook in which he keeps a track of each student's money record. Such notebook maintenance is extremely tedious and inefficient.

RFID has been around for quite a lot of time but since the cost of RFID is high it is restricted to high capital businesses. Although the costs are falling, RFID is still more expensive compared to available alternatives. RFID can have problems with metal and water and can cause malfunction. RFID has security issues. Unauthorized devices may be able to read and even change data on tags without the knowledge of the person who owns the object.

III. LITERATURE SURVEY

In “canteen management thesis Chapter 2”[6] Taylor postulated that many organizations provide canteens as an additional subsidized facility for their employees as a welfare measure. The practice of the canteen by the employees depends on numerous dynamics like centralized food distribution center with instantaneous and meek in operations. Canteen design provides a user-friendly system that facilitates quick and efficient operations to cover a larger section of employees within a specified time. There are canteens that are even paperless and have almost cashless transactions. This system is also appropriate in an environment where multiple caterers operate in a closed environment or at multiple locations. Advantages cited are the following removal of the manual system of maintaining the coupons instant information at ones fingertips for immediate use support for both prepaid and postpaid methods of accounting implementation of computerized management in an organization where the employees have cards any type of identification cards for unique identification preset daily limits on the amount of usage of canteen facility use of coupons for transactions reduction of wastage of

food items advanced notification on the usage of the canteen facilities by the employees on any particular day or shift and freedom on the conduct of different types of reports with respect to employee consumption and with respect to daily sales by the canteen. The challenges included in case of multiple locations, required canteen consolidated utilization which is very tedious or almost impossible in the case of a manual process multiple items management effective inventory management reordering of items to be managed in order to restock at proper time and subsidized food tracking.

A. “Cloud Based Canteen Management System”[2] research paper by Tazeen Khan, Daniel Yunus. Today everything around us is automated. Manual and paperwork has been replaced by automation. Automation decreases errors and helps in increasing efficiency. As canteens are a daily part of students, employees, staffs. The existing system is paper based where problem are faced by both the customer and owner. This paper proposes a system to make the whole process easier for both the canteen and for the customers by automating the process on Cloud. With the help of Cloud Computing technology no hardware components and installations are required. Hosting on cloud is cost effective than owning individual components. This proposed system bridges the gap between canteen and its usage. We have offered Radio Frequency Identification based card that will be used for the transactions of each customer at the counter. Also, to reduce the queues another method to order and pay is through mobile based app where deductions are carried out directly from the customer’s account. Account can be recharged through online transfer and e-wallets. Both the web and mobile applications will be hosted on cloud..

TABLE I :COMPARATIVE STUDY OF DIFFERENT MODELS I.E TRADITIONAL MODEL, MODEL USING RFID

	TRADITIONAL SYSTEM	RFID SYSTEM	CURRENT SYSTEM
Manpower required	More	Less	Less
Hardware cost	Expensive	Expensive	None
Manual calculation	Required	Not required	Not required
Role of higher officials	None	Required	Required
Payment gateway	Not available	Not available	Available
Time consuming	Yes	No	No
Security issues	High	Moderate	Less

B. “In-time billing process for canteen management system”[1] research paper by B Muniraja, J Rajanikanth. In colleges, schools or Inside an University campus canteen facility is provided. Students, college staff or university staff use this facility. In this canteen students or employees pay their bills by cash. In previous system, cash payment is the only option for making the payment. This is the case for small canteens. In rare cases some big canteens credit card facility is provided. The main drawback of cash payment system is that user always needs to carry the cash with him/her. And he/she needs to pay the exact amount otherwise there is problems for the remaining amount. One more problem is that in colleges the students are going to have food in the same canteen throughout the month. In such cases an account of students is maintained in a notebook. Canteen vendor inscribes order details of students on day-to-day basis and at the end of the month total is calculated. This method has limitation and draw backs of maintaining paper based records. These paper based records can get spoiled or damaged and data of student account might get lost. To give an effective solution for these problems we have proposed system called “In-Time billing process for canteen management system”. For this system canteen owner or canteen administrative person will give a RFID card to the user. This user can be a student in case of collages and an employee in case of an organization or company. This system is very effective. Advantage of this system is that it is really help full and effective for those people who have to go to canteen on daily basis. They need not carry cash. A card is assigned to each user and user can recharge this card with a certain sum.”

IV. PROPOSED MODEL

Each customer will know about the most ordered items and about the food items he/she has ordered the most. Hybrid algorithms are used to provide most ordered items.

They will get unique receipt number for each order placed and hence the delivery will be more organized and systematic thereby avoiding manual errors.

Previous systems used RFID cards. Even though they’re easy to use, to optimize the hardware we opted for payment gateway.

Customers can come cashless and still order for food items and pay using any payment gateway or using the balance in their personal account. If the balance isn’t sufficient enough, an error will be generated and the order will not be placed.

The ultimate goal is to increase the sales of the canteen.

Our proposed Canteen Management System is targeted to colleges and large companies. It has a very simple and easy to use UI, unlike other cogent Canteen Management System (CMS). It uses payment gateways and initial balance loading system in the accounts of the customer for payment process. As the system is online and easily accessible it enhances the working of the system and defines the purpose of the proposed system. Every canteen owner will have his/her login for customization of the menus and the weekly offers offered by the canteen. Our system will be having an online site and a mobile adaptable site for ordering the food. Our system will recommend and provide the customer with the hot selling food of the day i.e a food which has been ordered

by most of the customers. This is done using the hybrid algorithms. Our system will replace the traditional paper based systems in the canteen and also the RFID based systems. It is a huge enhancement over both these systems.

The system will keep a track of the total number of orders of the day or months, the most popular dish, an algorithm to predict and show what ingredient or food supply the canteen needs to restock and when will it be exhausted, the spending on restocking the supply, and it also calculates the profit/loss of the system. Due to the reason of the system being online the chances and manipulation of data is very low. The system has various users like the customer, admin, canteen employees. Every user of the website is provided a secure login. The system will bring about huge advancement in the way how canteen works

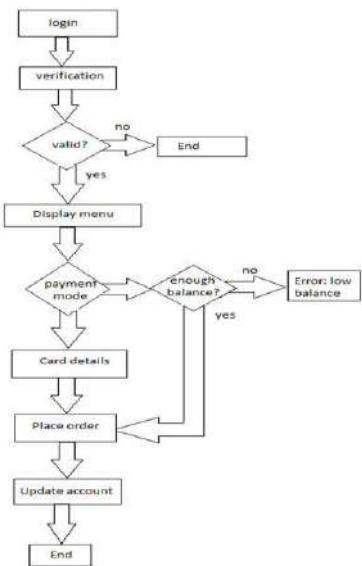


Fig. 1. Working of the system

Working:

Initially, the customer will come and register himself to the manager. He will have to provide his name, mobile number, department, class. As soon as he provides the mobile number, the user will get OTP on his number registered, and through the OTP received his mobile number will be verified.

The next step the customer will have to give certain cash amount to the manager so that he has a non-zero balance. Using the balance he can order food items from the canteen every time he wants and without having to give cash each time and the price of the food items ordered will be automatically deducted from the user's account until it is zero.

The other payment method is through payment gateway, where in if the customer forgets about his balance and his balance is zero and at the same time he is cashless, so he can use payment gateway to pay the required amount. When the customer has ordered food the manager will provide with a unique receipt number to the customer and the specific customer can collect when it's ready.

Now when the customer wants to order food, the customer does not need to scroll through the whole menu rather he can just order from the hot deals tab that will be provided on the website or from the tab which will show most frequently ordered food items by the user himself

TABLE II : DESCRIPTION

SCREEN NAME	DESCRIPTION
Login	Log into the system
Menu	Display the menu
Time	How much time it will take to serve the ordered food
Canteen Service	Create order, modify order, view order, cancel order
Customer balance	Showing the available balance
Administer	Availability and rates
Administer User	Create account; change password
Administer Meals	Create, modify, and delete meal items and prices
Payment	Accept payment for food

Pseudo code

Step 1: Enter login credentials

Step 2: Verification through email

Step 3: If account exists go to step 4 else go to step 9

Step 4: Display menu

Step 5: If the user orders, go to step 6 else go to step 9

Step 6: Payment mode: wallet or card

Step 7: Wallet

7.1: If the balance is enough go to step 7.2 else go to step 8

7.2: Place order and Pay

7.3: If the order is placed then go to step 9

Step 8: Card

8.1: Enter card details

8.2: Verify card details.

8.3: If the card details are valid go to step 8.4 else go to step 4

8.4: Pay

Step 9: Exit

V. RESULT AND PERFORMANCE

As a result the whole system will be less manual and more computerized which in turn will reduce labor cost and may require qualified staff to manage the reception desk. The system will be faster and efficient. The customer can order food and take away the delivery without having to pay cash or use card because the amount will be deducted from their account itself. Also the customer will know about the frequently ordered items by him/her and hence saving the time to scroll through the whole menu. The system will also show the most ordered items in the canteen using hybrid algorithms. We know that preparing each food item

approximately takes more than 7 minutes to be prepared, so the customer can place their order and can collect it later based on their unique receipt number



Fig. 2. :Snapshot of the homepage



Fig. 3. Snapshot Of Payment Gateway

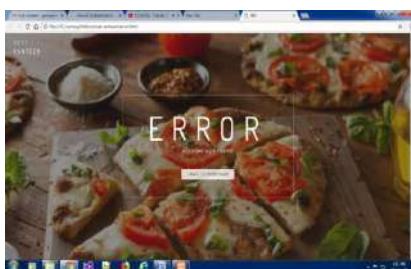


Fig. 4. Snapshot of error page during login

VI. FUTURE SCOPE

- This system can also make use of NFC (Near Field Communication) payment methods, using the NFC feature easily.
- Working:
- The College can provide each student with an NFC readable id-card
- If the e-wallet does not have enough balance or the credit/debit card is blocked or cannot be read at that instant then the NFC enabled payment seems to be best possible option
- Barcode can be installed in the system which can be read by any barcode scanner application using mobile phones.
- If some of the safety checks fail but user is able to verify the transaction with all verification steps the payment

may still go through, so that user doesn't have to initiate the transaction again. In this way ease of use factor is maintained.

VII. CONCLUSION

The Canteen Management System is an attempt to ease the process of ordering food for the convenience of the customers. It saves great amount of time and maintains security of each customer's balance. To encourage online food ordering, special offers are also offered. Once the order is placed, a unique receipt number is generated through which the customer can have his order. After each order, the balance will be updated and if the balance is zero, no order shall be placed.

The system will also provide hot deals available for today and the food item each customer orders more often.

VIII. ACKNOWLEDGMENT

I am grateful and privileged enough to get constant backing, care and guidance from our project guide of CMPN department who helped us in successfully completing our project work. Also, I would like to extend our sincere regards to Thakur College of Engineering and Technology for their timely support. I would also like to extend my thanks to the laboratory assistants of the CMPN department for their aid in offering me the resources in running the project. Lastly, I wish to thank my parents for their support and reassurance throughout my study.

IX. REFERENCES

- [1] G. E B Muniraja , J Rajanikanth, "In-Time Billing Process for Canteen Management System" International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 200-203 (2015).
- [2] Tazeen Khan, Daniel Yunus," International Journal for Research in Engineering Application & Management (IJREAM)" Final year B.E (Computer), Rizvi College of Engineering Mumbai, Maharashtra, India.
- [3] Shweta Shashikant Tanpure, Priyanka R. Shidankar, .Madhura M. Joshi," Automated Food Ordering System with Real-Time Customer Feedback" International Journal of Advanced Research in Computer Science and Software Engineering. Software Engineering Volume 3, Issue 2, February 2013.
- [4] K. Kamarudin, et al., "The Application of Wireless Food Ordering System", MASAUM Journal of Computing, vol. 1,pp. 178-184, 2009.
- [5] Taylor, "Canteen Management Thesis Chapter 2"
- [6] J. Purnama, et al."Application of Order Management System in Restaurants" Seminar Nasional Aplikasi Teknologi Informasi 2007, Yogyakarta, 16 June 2007 (SNATI 2007) ISSN: 1907-5022.
- [7] J. Mustafa,et al., "Touch &Dine: A multi-touchable dining system", UAEE International Journal of Computer Science and its Applications, volume 1, issue 1.ISSN 2250-376

Smart Automated Computer Education Management System

Ravishankar S. Singh¹, Pratik S. Singh², Swaraj S. Kasliwal³, Mrs. Veena Kulkarni⁴

Department of Computer Engineering, Thakur College of Engineering and Technology, Mumbai, India

Abstract-The project is non – profitable web based system. It is focused on equipping the indigent and rural students, who find it difficult to pursue their studies in their own various interested field. The proposed project strives to educate the students by strengthening or developing the fundamentals with minimal cost. The proposed system various modes of learning. The students can have access to video lectures, conferences, pdf notes etc. The students can also solve their doubts using virtual assistant.

The system proposed is not just a simple course system like any other but also the one consisting of notification and updation system. With regards to this, the one with appropriate access will be able to amend or modify certain, though limited, aspects which are not easily provided in other systems. From paper management to schedule adjustment and from that to portfolio assessment, the system provides an immense array of provision making the system more versatile and uniquely capable.

This system is focused on removing the barrier caused due to the financial condition of the students and also the lack of modes of interaction. Being a web based application, it also gives the rural students some free courses. This is a very light weight system. The proposed system would be very easy to use so any age group students can use it. The service of the proposed system can be avail easily.

I.INTRODUCTION

The project presented is an application based non-profitable system which is to be developed for certain category of students, especially indigent students and rural students. Since so many students manage education by engaging in rather external activities, system tries to provide a grandeur environment for student to grow in fields apart from college curriculum. The application system provides a platform where students who can learn about new technology in various field both theoretical and practical aspects.

II.MOTIVATION

There is a need for such application which will provide an easy access to different courses from beginner's level in a simple and interactive manner. Also the application will help students of other streams who want to learn computer courses and students can learn advance courses which are not provided in college.

III.PROBLEM DEFINITION

Due to the lack of money and bad financial condition so many students not able to get knowledge of computer or join institutes. If the above mentioned decides to join some institute the latter may encounter some experts that don't answer their basic questions. So this application is going to help those students to get education of computer from experts and they can start their education from very scratch.

IV.OVERVIEW

This project is an application based non-profitable system which is to be developed for students specially indigent students and rural students. Many Websites and applications are already available which provides education but not from scratch and on high cost. If students need any suggestions regarding to courses they can directly contact through the course instructor which can give them suggestions regarding to the course. For using this application new student have to sign up and old student can register themselves for new course. This application system will recommend courses based on their fields of student interests. Voice based searching will be available for students by which students can search every information related to the courses. By using interactive GUI students will be able to check catalogue of every courses. The system provides a platform to conduct webinars and to publish papers in various publications.

V.CURRENT SYSTEM

The current system are all web based application which provides courses at high prices. Some application provides courses at low prices but the content is not good or it is so advance that a beginner cannot understand. Some application provides good content but they do not have interactive interface.

In current system the main problem is if anyone has any query the management system is not that interactive. There are platforms which provide courses but there platform give recommendation of courses it doesn't provide any suggestion for a naive user who doesn't know from where he / she should start or how he should approach a particular courses.

The other problem with this system is to have an active internet otherwise you can't work or access the courses, where as the main purpose of application is to provide courses to rural people so if they don't have internet than also they can access the course material.

In the previous systems if student have some queries he has to wait for replies of experts and if his queries are irrelevant to experts then students do not get replies from experts. If some students have language problem previous

system cannot solve and it cannot provide any suggestions regarding the courses.

So the identified previous systems are lagging in solving these predicament of students. The system is focused on Indian students basically who are facing difficulties with English. The project goal is to provide computer education to indigent students so integration online and offline platforms is considered.

VI. PROPOSED SYSTEM

In this project; after careful and relentless consideration; team have come to conclusion that system would be proceeding with SDLC, specifically pre-emitive SDLC or waterfall model. The reason to choose this specific model is due to its merits and current dynamics with the external institute the respective project is building for.

As per previous discussion with Onsite advisor of that institute, team was provided with required specification and overalls of the project and accordingly the task to develop the project was allocated.

This represents the structural design of system where the system is divided into two initial aspects; staff record only accessible by restricted personals and public data which can be accessed by student and other concerned peoples.

The structure suggest that system consist of four level; the first level is concerned with data accessibility in system and the second level is more concerned with the involved modules and the particular roles they play in next stage.

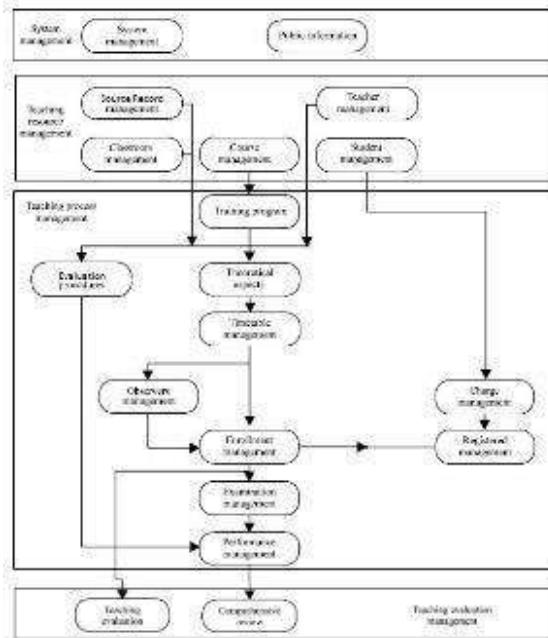


Figure 1: Block Diagram

The third level is the one which shows basic functionality of system and involving different modules as per their roles. And the final level is the evaluation stage based on functioning of involved parties in the previous level.

A. Flow of the System

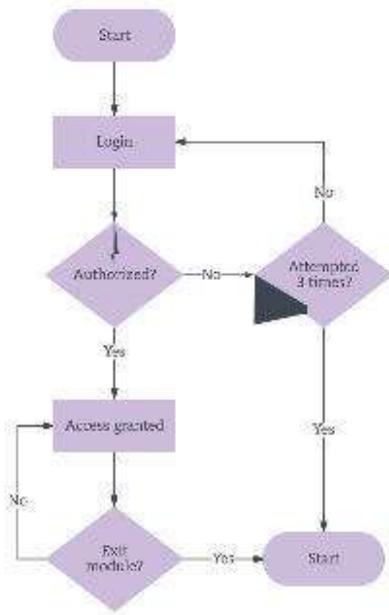


Figure 2: Flowchart of Login [8]

The initial stage is the login stage. Here user has to login with his unique id and password, if any of the above thing is wrong the system will ask again, and the system will give user three chances.

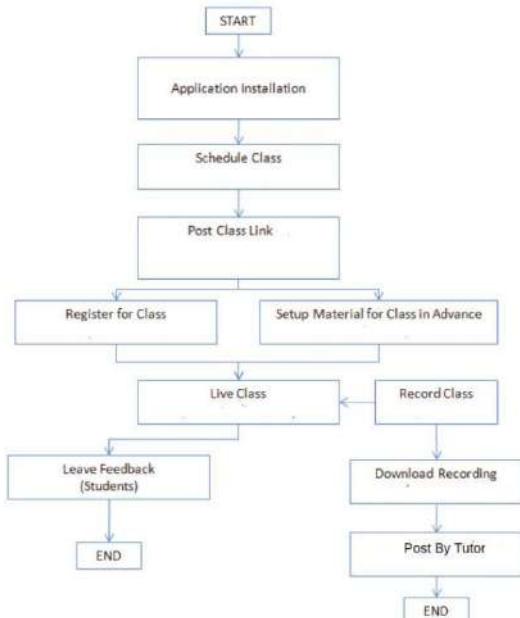


Figure 3: Flowchart of Course Upload [9]

This flowchart depicts how instructor will upload different assignment and course materials. Each course will be given unique id.

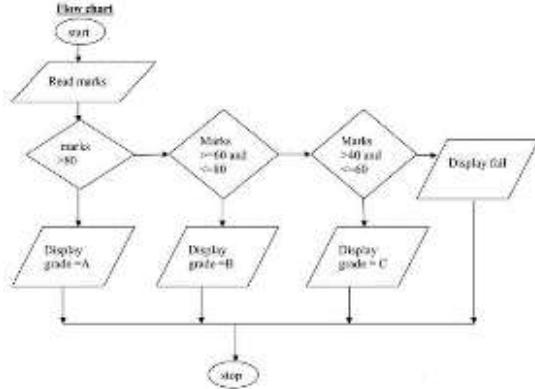


Figure 4: Flowchart of Grading [10]

In this stage grading of student assignment and test will done. Based on how student has performed grades will be given.

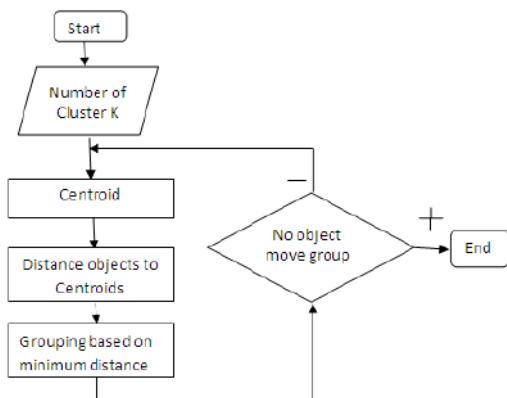


Figure 5: Flowchart of K-means Algorithm (Recommendation) [11]

The flowchart is of the algorithm we will use in our recommender system. The work of the recommender system is to recommend subjects to students based on their previous choices.

B. Recommender System

There are plans for implementation of clustering algorithm after completion of initial phase of project. The algorithm will be used for recommending courses based on student interest and answering their basic question. The specific algorithm which will be implemented for this will be k mean algorithm. It is an unsupervised learning algorithm.

The goal of algorithm is to find groups in the data with number of groups represented by K. The algorithm works iteratively to assign each data point to one of the group or cluster based on the features. Data points are clustered based on similarity.

The algorithms starts with initial estimates for the K centroids, which can either be randomly generated or randomly selected from the data set. The algorithm then iterates between two steps:

In the data assignment step, each data point is assigned to its nearest centroid, based on the squared Euclidean distance. More formally, if c_i is the collection of centroids in set C , then each data point x is assigned to a cluster based on

$$\underset{c_i \in C}{\operatorname{argmin}} \operatorname{dist}(c_i, x)^2$$

Where, “ $\operatorname{dist}(\cdot)$ ” is the standard Euclidean distance. Let the set of data point assignments for each i th cluster centroid be S_i .

Whereas, in the centroid update step, the centroids are recomputed. This is done by taking the mean of all data points assigned to that centroid's cluster.

$$c_i = \frac{1}{|S_i|} \cdot \sum_{x_i \in S_i} x_i$$

The algorithm iterates between steps one and two until a stopping criteria is met (i.e., no data points change clusters, the sum of the distances is minimized, or some maximum number of iterations is reached).

This algorithm is guaranteed to converge to a result. The result may be a local optimum (i.e. not necessarily the best possible outcome), meaning that assessing more than one run of the algorithm with randomized starting centroids may give a better outcome. [12]

VII. CONCLUSION & FUTURE SCOPE

A. CONCLUSION

In this project, the team was able to develop an institute management system where the system handles certain specific areas. Areas like the administrating staff region where from the schedule calendar to student notification, lecturers coordination and other aspects are handled. Then there is lecturer region where from test arrangement to reports of every student under them showing their performance analyses are handled by the respective lecturer.

Also, one takes a look at student region where every student is provided with very necessary aspect like the scheduled lecture updation and frequent exams. Also made available are the video lecture and other source material to read on. The system structure can be said as three tier system where administrating staff, students involved and lecturers all have separate rights to channel through and fulfill the necessary role.

Many students from a backward background are not able to get the level of education to break the barrier. Even students of suburbs suffer from lack of in depth knowledge of the context. Here is where the institute will be able to make a difference. By providing a well-equipped system to a promising cause there will be significant amount of impact on the society's culture and direct it towards a new direction of modernization. Under an institute which focuses of indigent people education and betterment, the system plays a

very crucial role to provide them with a stable backbone that can help them to get more organized with.

B. FUTURE SCOPE

In the starting phase the application will be deployed in some computer institutes under the NGO after success of first phase more number of computer institutes and individual experts can join by themselves.

Project will be deployed in all over in India under Pragati Computer Institute and आज़ाद शिक्षण अवं सेवा समिति

The future scope of this application is to provide different courses other than computer courses.

In its future version it will be fully automated so that expert panel is not required anymore.

An android application is to be developed of this application.

Fully automated system is scheduled to be implemented in the next phase where input of admin will be lower.

VIII. References

- [1] “Small Business Data Storage Solutions” research paper by Synology A New NAS Experience, Synology America Corp. 2899 152nd Ave NE Redmond, WA USA 1-425-818-1587.
- [2] D B Heras, D. Otero, and F. Arguello, ”An eco feedback system for improving the sustainability Performance of universities,” in Proc. 2011 IEEE International Conference on Virtual Environments Human –Computer Interfaces and Measurement Systems, Ottawa, ON 2011 , pp. 1 – 6
- [3] S. Jeyalatha, B. Vijayakumar, and G.S. Wadhwa, “Design and implementation of web based application for relational data maintenance in an university environment,” in Proc. 2011 International Conference and Workshop on Current Trends in Information Technology , Dubai,pp. 105-112
- [4] Rohan Bhirangi, Smita Bhoir “Automated Question Paper Generation System” International Journal of Emerging Research in Management & Technology ISSN: 2278-9359 (Volume-5, Issue-4), April 2016
- [5] Veerle Van den Eynden, Louise Corti, Matthew Woppard, Libby Bishop and Laurence Horton “MANAGING AND SHARING DATA” UK Data Archive University of Essex Wivenhoe Park Colchester Essex CO4 3SQ ISBN: 1-904059-78-3, e 16 March 2011.
- [6] Sawal Tandon Computer Science and Engineering, Lovely Professional University “Application of Incoming SMS to a Website, Control the Website to Send Bulk SMS” (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (3) , 2012,4287 – 4289, ISSN:0975-9646
- [7] Hannah Bast and Claudius Korzen “The Iccite Research Paper Management System”, Department of Computer Science, University of Freiburg, Germany, X. Lin et al. (Eds.): WISE 2013, Part II, LNCS 8181, pp. 396–409, 2013. Springer-Verlag Berlin Heidelberg 2013.
- [8] <https://www.lucidchart.com/pages/examples/flowchart/system-flowchart-template>
- [9] https://www.researchgate.net/figure/260165534_Figure-1-Flowchart-for-conducting-WizIQ-virtual-tutorials-Note-ITS-Information-Technology
- [10] <https://sillycodes.com/c-program-to-calculate-grade-of-student/>
- [11] https://www.researchgate.net/figure/269800441_Fig-1-K-Means-Clustering-Flow-Chart
- [12] Learn data Science, Machine Learning “Introduction to K-means Clustering” Andria Trevino 12.06.16 data science https://googleweblight.com/i?u=https://www.datascience.com/blog/k-means-clustering&grqid=qjW3m_1w&hl=en-IN

Graphical Approach For Frequent Pattern Mining

Anupriya Babbar¹, Vipul Babbar², Tanushsree Bhattacharjee¹

Thakur College of engineering & tech¹, AscenaRetail Inc².

Anupriyababbar.86@gmail.com,Vipul.babbar@gmail.com,tanu.b.4488@gmail.com

Abstract: Data mining is a method to extract useful information from large databases. It performs many tasks such as classification, clustering, prediction, association analysis [1]. One of the most important researched fields of data mining is frequent pattern mining, which plays vital role in all the above mentioned tasks.

One of the major drawback frequent pattern mining is that it requires multiple database scans to drill out the frequent patterns and may produce large number of frequent patterns especially with long patterns, the refined solution of the above problems is Maximal Frequent Pattern (MFP) it is the smallest representative set for frequent pattern generation, MFP's are the frequent patterns whose superset cannot be frequent [2].

This paper proposes a graphical method to produce MFP which will generate frequent patterns. This method introduces two new properties; a graph structure called as Prime graph and a PG-Miner algorithm. Prime graph is a simple graph structure by traversing it by one scan can produce frequent patterns as the graph itself captures the whole information about the transactions by using an optimizing data transformation technique which uses prime number theory. PG-miner is the proposed algorithm which traverses the prime graph and prunes the infrequent items. The efficiency and compactness of proposed method is proved with the help of experimental results.

Keywords: Data mining, frequent pattern mining, graphs, prime numbers

I. INTRODUCTION

With the increase in the size of database there is a need of developing a tool which can drill down the useful information from the database with ease.

Knowledge Discovery of Data (KDD) is a process to extract useful patterns from the database KDD Process shown with the help of figure. Data mining is an important step of KDD, which is used to extract useful information and can be implemented in many areas like data bases, artificial intelligence, knowledge discovery in neural networks etc.

Frequent pattern mining is the one of the most important tool of data mining, which is used to extract frequent patterns based on minimum support or confidence value. Association rule mining is based on extracting interesting co-relations, Frequent pattern mining is the first step of association rule mining in this the patterns which satisfy the threshold is frequent otherwise infrequent [14]. Many algorithms have been devised to mine frequent patterns.

They basically fall in two categories: (1) Mining frequent patterns with candidate generation Mining frequent patterns without candidate generation. Methods with candidate generation like

Apriori [16], partition [21], incremental based [17] [19], suffers from many problems like multiple database scans and candidate generation. Many extensions are made to the previous algorithm but still it encounters the above problems. And method without candidate generation like Pattern Growth [20] or FP-growth is an improvement over candidate generation algorithms. They require two database scans to drill out the frequent patterns from the database; several optimizations are made to reduce the number of database scans and lessen the time taken and the search space to produce frequent patterns. Maximal Frequent Pattern (MFP) is a reasonable solution for the above mentioned problem, as it is a smallest representative set to produce frequent patterns; it reduces the number of frequent set generation [3]. This paper proposes a graphical method based on MFP to produce frequent patterns, this graphical approach can be extended to all data mining tasks. However, most of the times some changes are made in graph structure, pruning or traversal technique. This method uses simple graph structure to keep the transaction information and a graph miner algorithm to traverse the graph to find the frequent patterns and prunes the infrequent patterns. This method uses data transformation technique to convert data into prime number format which reduces the size of data sets significantly, then construction of prime-graph takes place and with the help of prime minor algorithm frequent patterns can be mined and it prunes all in frequent patterns from the data set, in one database scan, as all the useful information related with the transaction is stored in prime-graph, by traversing the graph once only frequent patterns can be mined. Various experiments have been performed on the web log data set to prove the efficiency and the correctness of the proposed method. Organization of the paper is as follows Section 2 introduces the problem and reviews some efficient related works. The proposed method is described in section 3. The experimental results and evaluation show in section 4. Finally, section 5 contains the conclusions and future works. Problems and related work. Problems of FPM Algorithms In a data set the items which satisfies user defined threshold are frequent otherwise infrequent. In many real applications it is infeasible to find out all frequent patterns especially when data set is highly populated. It is a tedious task to decide the threshold value as low threshold may produce large number of patterns destroying the accuracy of mining and the high threshold will only produce very less patterns leaving even some of the frequent item sets. Algorithm with candidate generation may generate large number of candidates to produce frequent patterns which require more space and database scans and make complete process expensive. The major problems with this algorithm are of multiple database scan and the search space [6].

II. PROBLEMS AND RELATED WORK.

A. Problems of FPM Algorithms

In a data set the items which satisfies user defined threshold are frequent otherwise infrequent.

1. In many real applications it is infeasible to find out all frequent patterns especially when data set is highly populated.
2. It is a tedious task to decide the threshold value as low threshold may produce large number of patterns destroying the accuracy of mining and the high threshold will only produce very less patterns leaving even some of the frequent item sets.
3. Algorithm with candidate generation may generate large number of candidates to produce frequent patterns which require more space and database scans and make complete process expensive.
4. The major problems with this algorithm are of multiple database scan and the search space [6].

B. Related works

To overcome the problem of previously proposed algorithms many extensions are being made to reduce database scans and the number of candidate generation like Aclose [10], CHARM [8], Cobbler [11], Carpenter [11], AFOPT [12] and etc, are the extensions of Apriori which is a method based on candidate generation. FP-growth [20] is a method based on without candidate generation have been proposed which requires two database scans in first scan item sets are arranged according to the frequency descending order and during second scan it mine all frequent items. It is advancement over prefix tree. FP-tree merges the links which have same value. It compacts the data and enhances the performance by increasing the speed. It requires large memory space for parsely populated data set where common path is very low. There is another method known ELCAT [7] which uses vertical data format rather than horizontal data format, it prove much more efficient then Apriori as it uses Boolean power set lattice theory which requires less space to store information about the transaction. The refined solution proposed in our method is to derive frequent patterns from MFP, many algorithms have been devised which generates frequent patterns from MFP and prune infrequent item sets. There are two pruning techniques used

1. Subset pruning mining: the all subsets of anyfrequent pattern are pruned because they cannot be maximal frequent pattern.

2. Superset infrequency pruning: the all supersets of any infrequent pattern are pruned because they cannot be frequent pattern. But still they still require two database scans like Pincer search algorithm [13] [4]. It makes use of both top-down and bottom-up traversal to mine MFP. Depth project is another method to mine MFP which uses depth first traversal [15] and both pruning techniques and moves in lexicographic order to traverse. This is an efficient method to mine frequent patterns. The extension of depth project is MAFIA [5]. Rymon's set enumeration is used by above methods which avoid counting the support of all frequent patterns [18].

Flex [20] is a method based on vertical data format, it moves in lexicographic order which is based on first test than generate which ensures that generated nodes are frequent. It is the one of the most efficient method to mine MFP. But the major drawback

is it needs the huge amount of memory to store the information about item sets.

III. PROPOSED METHOD

A. Data Transformation technique

Data pre-processing is an essential step of data mining as shown in figure. It comprise of data cleaning, data reduction and data transformation [22]. In our method data transformation is used to reduce the size of data set significantly. In this method the web log dataset is transformed with prime based compaction which reduces the size of dataset. Each complete transaction is transformed into PMV (Prime Multiplied Value) a positive integer. During Prime graph construction transaction given $T=(\text{Pid}, Z)$ where Pid is the ID of transaction and $Z = \{a_1, \dots, a_m\}$ is the item set of Z . Prime Multiplied Value Pid is computed with the help of equation 1

$$\text{Mod}[(\text{PMV}, P_r)]$$

Where P_r is the number of item set of Z .

With the help of above equation can be transformed into contracted form. In fact data transformation is an abstracted form of transactions. This is explained with the help of an example in table 1 there shows eight transaction of website login and page number. In which page number is then transformed into prime numbers and then prime multiplied value is calculated.

When this transformation is applied to the real web log data result will be in drastic compaction. It reduces the size of data set more than half. This process is independent of size and type of data set, any data set can be reduced like $P=(4, \{8, 7, 12, 11\})$ and $P_0 = (4, \{8884, 990, 7123, 1234\})$ are transformed to the same value 770.

Table I The transaction database DB and its Transaction Values

TID	Page No.	Transformed	TV
1	8, 5, 11, 20, 6	2, 3, 5, 7, 11	2310
2	8, 5, 11, 20, 9	2, 3, 5, 7, 13	2730
3	8, 5, 6	2, 3, 11	66
4	8, 11, 20	2, 5, 7, 11	770
5	11, 20, 9	5, 7, 13	455
6	8, 11, 20, 9	2, 5, 7, 13	910
7	8, 11, 20	2, 5, 7	70
8	11, 20, 9 5, 7, 13		455

B. Prime Graph Construction

Graph structure enhances the performance of mining by data compression and by reducing search space by using pruning techniques. Thus the graph structures have been considered as a good option in previous data mining researches. This research introduces a simple graph structure called Prime Graph (Prime-number Compressed Graph). Prime graph uses the concept of prime number theory for transformation. This method improves the performance by data compressing and pruning techniques. A Prime Graph includes number of nodes which consist of prime

number allotted to the item set of transaction ($P_{1...n}$) and on the other hand some nodes consist of Prime multiplied value i.e. $PMV_{1...m}$. The node structure consisted mainly of several different fields: value, local-count, global-count, status and link. PMV is getting stored in the value field. During insertion of current PMV local-count field set by 1 if function $[mod(PMV_m, P_{1...n})] = 0$ or no remainder. The global-count field registers support of pattern P which presented by its PMV. Global-count register is used to store the support information of all frequent and non-frequent items which can be further used for mining according to the user defined threshold. The status field is use to keep track of traversing when a new node is visited its value changes from 0 to 1. The link field is to form inward and outward edges to and from the nodes. Fig. 1 & 2 shows the construction of Prime Graph for transactions shown in table 1. The construction operation mainly consists of insertion of nodes PMV(s) and $P_{1...n}$ into Prime Graph based on definitions below:

Definition1: Links through PMV and P_n will be connected depending upon the formulated equation $[mod (PMV_n, P_{1...n})]=0$ or 1. Each and every value of PMV get modulo divided by P. If there is no remainder or 0 that means PMV is completely divisible by P, then there will be a link form between from that P directed towards PMV and local-count increased by 1.

Definition 2: Link from one PMV to other PMV is formed when one PMV is completely Divisible by other PMV.

Definition 3: A self loop to a node of PMV is form when same value of PMV is repeated more than one time i.e. same subset of itemset is been repeated more than once in a whole set of transaction.

1. Working Of Prime Graph For Elementary Transaction

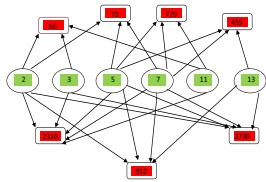


Fig 1: PC-Graph Construction For elementary Transaction.

The count of the edges which directed from P towards PMV's or the out degree of a P is the total frequency of the appearance of P in complete set of transaction. It is shown below with the help of a figure.

2. Working Of Prime Graph For Subset Transaction

The count of the edges which directed from one PMV towards other PMV or the same PMV is the total frequency of the appearance of particular subset of itemset in a complete set of transaction. It is shown below with the help of figure.

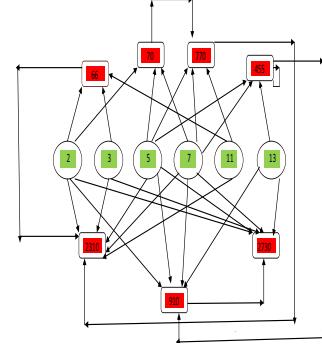


Fig2: PC-Graph Construction For Subset Transaction.

IV. PRIME_GRAPH MINER ALGORITHM

As explained in previous section, during construction of Prime graph following registers maintained by storing the updated value

- a) count- which stores the frequency of particular items.
- b) Local-count- Keeps the value of current PMV.
- c) Global-counting-keep track on the frequency of frequent and infrequent items.
- d) Status

Step1. Traverses the graph in top-down direction

Step2. Compare the frequency of the elementary itemset (pages) to the user defined threshold

Step3. Prunes the infrequent itemsets

Step4. Matches the subset of the transactions with one another with the help of PMV

Step5. Compares the frequency of repeated subset transaction with the user defined threshold.

Step6. Results gives the frequent elementary itemset of the frequent page numbers and the frequent subsets of the transactions that are same set of pages repeated in more than one transaction.

The Prime_Graph Miner algorithm traverses the completed Prime Graph to discover the MFP in top-down direction. There is no need to database scan again, because all information about items and patterns are stored in the Prime Graph itself. The miner algorithm prunes the infrequent itemsets. As a result the search space is reduced, which dramatically reduces the computation and memory requirement and enhances the efficiency. Table 2 shows the item frequency and considered prime number for transaction database

Table II Frequency of pages and proposed prime number
Page no. Prime number Item Frequency

8	2	6
53	3	
11	5	7
207	7	
611	3	
913	4	

IV. Experimental Results

All experiments were performed in a time-sharing environment in a 2.4 GHz PC. All the algorithms are implemented using Matlab.

In first experiment we use synthetic web log sparse datasets .The number of transactions are 50, the average transaction Login is 12 and the number of transaction increased from 50 to 100 to evaluate how the data transformation technique can compact the size of dataset. Fig.4.1 shows comparison of the size of original dataset with the size of transformed dataset using proposed data transformation technique.

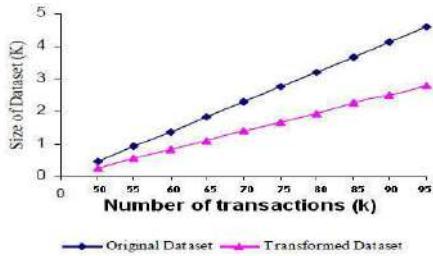


Fig3. Size comparison of data

Second experiment is to compare the performance of the Prime Graph with the FP-Tree on the web log dataset. To allow a fair comparison of algorithms, firstly plot all transactions using Prime Graph and FP-Tree separately. Time taken by six random sets of 50 transactions of 12 logins are recorded to plot a comparative graph between tree construction and prime-graph. Then all frequent patterns are generated by same procedure run in cached mode. Fig. 4.4 shows the PC_Graph outperforms the performance of FP-Tree.

Hence, this is proved by the experiments that proposed method is an improvement of previous method that has been proposed for maximal frequent pattern mining and it requires only one database scan. The experimental result verifies the compactness and the efficiency of Prime Graph method.

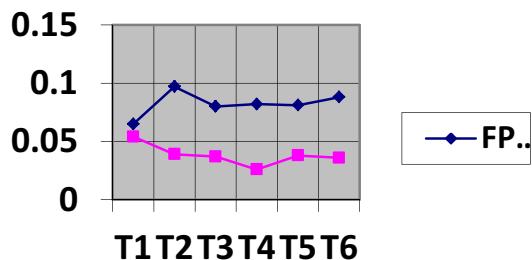


Fig4. comparative analyses of FP-Tree and prime-Graph

V. CONCLUSION AND FUTURE WORK

This Proposed method concludes that Prime graph method is a technique based on without candidate generation so it does not produce large number of frequent candidates to generate further frequent patterns. It requires only one database scan to mine the frequent patterns as all the useful information about the transaction stores in the Prime graph itself. It requires less search space as Miner algorithm Prunes the infrequent itemsets by using combination of subset and superset pruning techniques which reduces the size of dataset up to an extent. It is time efficient; as time required in constructing FP-Tree is much greater than the time needed to plot a Prime graph with the same set of

data. The experimental result verifies the compactness and the efficiency of Prime Graph method. comparative analyses of FP-Tree and prime-Graph 5 Conclusion and Future work This Proposed method concludes that Prime graph method is a technique based on without candidate generation so it does not produce large number of frequent candidates to generate further frequent patterns. It requires only one database scan to mine the frequent patterns as all the useful information about the transaction stores in the Prime graph itself. It requires less search space as Miner algorithm Prunes the infrequent itemsets by using combination of subset and superset pruning techniques which reduces the size of dataset up to an extent. It is time efficient; as time required in constructing FP-Tree is much greater than the time needed to plot a Prime graph with the same set of data. This method is an improvement over FP-Tree method in terms of time, space and speed. It is independent of size of dataset, whatever be the size of transaction it can be transformed into prime number and it gives frequency of both elementary itemsets as well as subsets [2]. Our proposed method, is simple to implement, easy to understand and does not includes any complex structures. This graphical method can extended up to wide applications for enhancing performance of the particular like it can be applied in incremental mining of frequent patterns where database transactions are updated regularly. In addition, it can also be used for interactive mining of frequent patterns where minimum support threshold can be changed to find new correlation between patterns. This method can be used for large graph structures with unique nodes and can be applied to large databases to find out the particular subset repetition of the transaction which can be useful to avoid frauds as well as can be useful in discovering knowledge for artificial intelligence based applications.

REFERENCES

- [1] 1.ThashmeeKarunaratne, “Is Frequent Pattern Mining Useful In Building Predictive Models?” Stockholm University, Forum 100, Se-164 40 Kista, Sweden.
- [2] 2. Norwati Mustapha, Mohammad-Hossein Nadimi-Shahraki, Ali B Mamat, Md. Nasir B Sulaiman “A Numerical Method For Frequent Patterns Mining Journal Of Theoretical And Applied Information Technology”. Journal Of Theoretical And Applied Information Technology, 2009.
- [3] 3. Nadimi-Shahraki M.H., N. Mustapha, M. N.B.Sulaiman, And A. B. Mamat, “A New Method For Mining Maximal Frequent Itemsets”, Presented At International Ieee Symposium On Information Technology, 2008. Itsim 2008., Malaysia, 2008, Pp. 309- 312.
- [4] 4. Bayardo Jr R. J.,” Efficiently Mining Long Patterns From Databases”, AcmSigmod International Conference On Management Of Data, Pp. 85-93, 1998.
- [5] 5. Burdick D., M. Calimlim, And J. Gehrke, “Mafia: A Maximal Frequent Itemset Algorithm For Transactional Databases”, 17th International Conference On Data Engineering, Pp. 443-452, 2001.
- [6] 6. Jiawei Han ,Jian Pei , Iwen Yin , “Mining Frequent Patterns Without Candidate Generation: A Frequent-Pattern Tree Approach”, Received May 21, 2000; Revised April 21, 2001.
- [7] 7.Zaki M. J., “Scalable Algorithms For Association Mining”, Ieee Transactions On Knowledge And Dataengineering, Pp. 372-390, 2000.
- [8] 8. Bart Goethals, “Memory Issues In Frequent Item Set Mining” Sac’04, March 14–17, 2004.
- [9] 9. Pei J, Han J, Mortazavi-Asl B, Pinto H, Chen Q, Dayal U, Hsu M-C, “ Prefixspan: Mining Sequential Patterns Efficiently By Prefix-Projected Pattern Growth”. In: Proceeding Of The 2001international Conference On Data Engineering (icde’01), Heidelberg, Germany, 2011,Database Technology, Valencia, Spain., 1998.

School Management ERP

AnandKhandare^[1], Raj Shah^[2], Vishwa Solanki^[3], Aakash Patel^[4], Abhishek Anand^[5]

Thakur College of Engineering & Technology

Kandivali (E)

Abstract-School ERP is a package web based and Android App School Management System which gives school to use and operate many of integrated interrelated modules and manage the administration of school and students efficiently. Due to its ever growing and competitive nature, the education sector has always been in need of a quality solution to manage and serve the school resources efficiently. IT sector is giving number of solutions to schools like smart classroom,digital learning solutions and school management system to make learning easier and manage school administration effectively.

Keywords:Smart Classroom, Efficiency, Admission management ,Campus, Quality, Facilities, Business Intelligence

I. INTRODUCTION

We offer school ERP which simplifies and automates schools administration process. The School Management System is accurate and reliable and can be conveniently accessed from school intranet as well as from the public internet. It is fully browser and android application based school management system which also includes virtual campus which can be linked with school portal and contains powerful online access to bring parents, teachers and students on a common interactive platform. Yet another advantage of the ERP system is that it runs on minimal hardware and easily fits in the budget of schools and for there customers. Where in ERP users have role based access rights which tightly models existing school's hierarchy. School ERP is totally customizable according to the needs of school's administration and students and their parents needs.

IT sector is giving number of solutions to schools like smart classroom, digital learning solutions and school management system to make learning easier and manage school administration effectively. Today educational institution is not limited to imparting education alone, but it is adapting latest trends in IT for improving the quality of education and handling various activities of school including admissions, class management, library management, logistics, inventory, fee management, alumni, accounts etc. We offer school ERP which simplifies and automates schools administration process.



Fig 1:Start Screen

Benefits to School Administration:

CREATE ACCOUNT GROUPS			
S NO	GROUP NAME	UNDER GROUP	ACTION
1	STUDENTS	PRIMARY	
2	STAFF	PRIMARY	
3	LABOUR	PRIMARY	
4	EVENT	PRIMARY	
5	TECHNICAL STUDENT	STUDENTS	
6		PRIMARY	

Fig 2:Account Screen

II. OVERALL DESCRIPTION

1. Admin Management System

User Management

- Individual login of authorized personnel with access rights to manage operations and resources in smooth manner

- Module wise access permission & restriction management.

Admission/Registration Management

- End-to-end admission module which takes care of complete registration system
- Suggestions about fees school-wise, class-wise, etc. and provision for deductions/concessions applicable individual cases.
- Shows particulars of pending fees, late charges, total amount collected and other info

Fees Management

- Defines time of fees to be collected
- Admission forms issued / submitted
- Student enrolment details
- Issues important documents, letter of grants and scholarships to students

Inquiry Management

Manages each inquiry and provides accurate information to respective students.

Attendance Management

- Student attendance & leave management module.
- Teacher & staff attendance & leave management.
- Attendance & leave management directly through access card or biometric system integration

Accounts Management

- All type of payment transaction management is done using this module
- Several report related to accounts of a student

Appraisal Schedule

- Daily/Weekly/Monthly appraisal schedule report
- Appraisal scheduling

Batches & Course Management

- Manage & Upload courses
- Manage batches & section for a particular academic year
- Manage subjects for each batch
- Students' result & promotion management.
- Import previous batch's subjects.
- Manage HOD, class teacher & other subject teachers depending on the courses, class etc.

Certification Management

- Certificate generation process
- Certificate Request Posting
- Certificate Printing

Leads Management

- Online lead management
- Lead generation Report
- Lead feedback report

Student Transfer Management

- Student centre transfer
- Student batch transfer
- Student transfer report

Education Delivery Management

- Faculty Performance Report
- Faculty Training Schedule
- Faculty Feedback

Feedback Management

- Student feedback report
- Student feedback form
- Student feedback entry

Timetable Management

- Automatic & Diligent Time Table Generation for Year/Month/Week
- It provides details of availability of particular faculty with respect to the timetable and subject.
- Substitute/ Proxy Faculty Management
- Class wise period analysis
- Daily class activity tracker
- Substitution Alert
- Subject & Class wise teacher setting

Profile Management

- Helps to manage profile of students, faculty and staff
- Provides an easy and quick access to all the information for students, faculty and staff.
- Up to date and accurate data which saves lot of man hours of the administration

Library Management

- Bar code enabled tracking list of books
- Reports of books issued & received
- Integrated with Fees & dues management

Hostel Management

- Monitor entry/exit
- Attendance of inmates
- Disciplinary Logs
- Fess Collection
- Records of maintenance/repairs of rooms
- Maintains Guest /Visitor log
- Maintains daily expense sheet of the hostel
- Mess Bill Collection
- Hostel Electricity Usage record

Canteen/Cafeteria Management

- Monthly collection of payment from the mess
- Easy Generation of bill
- Material Management
- Menu Design
- Special norms can also be integrated within system
- Details of the daily/weekly/monthly consumption of meal by students
- Flexibility to define canteen fees priory in the fee structure or depending on the consumption

Class Room Management

- Maintains total classroom supplies of books, writing equipments, stationeries and other inventories for particular classroom along with monitoring sitting arrangements, lighting and other essential facilities.
- Provides direct support to necessities of particular classroom that creates healthy classroom environment.

Examination Management

- Examination Name (Term wise / Semester wise)
- Automatic generation of exam timetable , seating arrangement, hall tickets for students
- Collection of question papers from faculty
- Delegation of invigilation (name of teacher, class room no., subject)
- Allotment of paper correction to professors
- Grades/Marks obtained by students
- Result Manager (declaration of result).

2. Student Management System

Online student portal enables students to communication and collaborate online. It's like a cloud and collaborative portal designed to help student to collect any information online whenever, whatever, wherever they wanted to. It is a user friendly dynamic website from where a student can get to see his/her status of the course, performance in the course, and any updated information.

Student Management

- Student Management
- Student Performance Management
- Student Feedback Overview Management
- Announcement Management
- Page Management System.
- Gallery Management System to upload and manage photos.
- Notice Management System to display the notice on the website.
- Member Management to add/modify/delete the member from the website.
- News Management system.
- Project posting management.
- Photo Gallery Management.
- Video Gallery Management.
- Unlimited Color changing options.
- Banner Style Changing Features.
- Banner contains management.
- Video upload manager.
- Banner management system.
- User review management.
- Visitor query management.
- Admin General Settings.

Student Login Features

- Student course status
- Student Certificate Status
- Student Feedback
- Student Performance
- Chat with expert
- Batch Status
- Online Fees Payment
- Program Change
- Batch Change
- Cloud Collaborative Portal
- Student Dashboard
- Chat with your batch-mate
- Online Notice
- Browse the site (Up to 15 Dynamic Pages).
- 100% Dynamic.
- Category wise page management.
- Video clip showing.
- Photo Gallery.
- Simple SEO Features.
- Visitor can Comment.
- Portfolio Page.
- Browse Gallery.
- Write review.
- Send queries.
- Latest Post from the blog.
- Student testimonials.
- Join mailing list for new customer.
- Follow on Face book, Twitter, ETC.

3. Staff Management System

Staff management

- Log employee time attendance automatically

- Detect breaks and idle time automatically
- Alternatively enter hours manually through a web form
- Private mode available to employees for privacy on breaks or when completing personal tasks
- Takes screenshots at regular intervals
- Link hours and attendance information to payroll software
- Shows the online status of all team members
- Real time screen monitoring available for Quality Assurance
- Email monitoring can store all emails sent by employees
- Generate reports on employee application and document usage
- Ability to import and export data.

III. EXTERNAL FEATURES

Attendance Management

- Online attendance marking
 - Employee self service
 - Absence management
 - Real time reporting
 - Easy integration
 - Overtime management
 - Multi-level approval
 - Team view
 - Shift module
- Biometric device integration option.

Leave Management System

- Leave Policy
 - Multiple Leave Schemes
 - Leave Transactions
 - Year end Processing
 - Leave management
 - Multi-level approval process
 - Public/company holidays
 - Produces reports highlighting absenteeism patterns and trends.
 - Calculate accurate leave balances
 - Track leave approval
 - Assign work to other during leave period
 - Anytime leave ledger
- Biometric device integration option.

IV. OTHER NONFUNCTIONAL FEATURES

School Bus Tracking:

This application gives most focus to safety & security of students. It provides access to customizable GPS school bus tracking system. Guardians, school administration as well as the transport contracting agency can track & get updated with the route, time schedules, pickup and drops of buses.

This application will serve all the purpose of an advance school management system with a very competitive & affordable price.

School management had never been so easy and flexible like this with this unique application.

V. MAJOR BENEFITS

- Transparency in school operations.
- Availability of real time up to date information.
- Better management of school academic process & administration.
- Instant access to required information and activities of the school.
- A centralized system which makes reporting possible at one place.
- School ERP is a web based Application.
- Easy access through a web browser having an Internet connection.
- Keeps parents up to date with their ward's progress in school.
- Integrated with Bulk SMS
- Software for sending specific SMS Alerts.
- Online Registration, easy follow up and Admission.
- Better interaction between parents, teachers & school management.
- Efficient computing centralized storage, high memory and fast process.
- Integration of Academic Calendar, School Notices and other activities.
- High level security at application level user level and program level.
- Savings from phased out legacy systems
- Streamlines Education Processes
- Automates Important Functions Increases Productivity and Efficiency
- Big Savings in Person
- Improves data and process integrity
- Enhances operational security
- Reinforces accountability and transparency
- Single system to manage all School related information from anywhere in the School
- Reliable and secure system
- Complete Automation of operations
- More Time to focus on Strategic Tasks
- Better informed decision making for management.
- Parents have access to all academic information about their wards through internet.

Support:

School is backed by excellent service support} through.
Dedicated Help Desk.{ Personal Visit.{ Telephonic Support.{ Email.{ Data transfer through Internet.{

REFERENCES:

- [1] ICMT conferences-2016
- [2] Partner Management document schools
- [3] School Management system- Addls Ababa
- [4] ERP system for college-IJST

Roadways Safety Suggestions

Deep Suchak, Nishant Shishodia, Vikram Rajpurohit, Kunal Singhal, Prachi Janrao

Thakur College of Engineering & Technology, Mumbai University, India

deep.suchak96@gmail.com, shishodia2021@gmail.com, vikramrajpurohit888@gmail.com, kunalsinghal96@gmail.com, prachi.janrao@gmail.com

Abstract - There are a lot of vehicles driving on the roadway every day, and traffic accidents could happen at any time anywhere. Some accident involves fatality, means people die in that accident. We propose to build a project that analyses the accidents data set to give safety suggestions. In this project we apply statistics analysis and data mining algorithms on a Fatal Accident dataset as an attempt to address this problem. The relationship between fatal rate and other attributes including vehicle type, time of day, surface condition, light condition, casualty severity, and age of casualty are investigated. Association rules are discovered by Apriori algorithm and FP Growth algorithm[2]. Certain safety driving suggestions will be made based on statistics, association rules, classification model, and clusters obtained.

Keywords - data mining; apriori; classification; road safety.

I. INTRODUCTION

Road and traffic accidents are uncertain and unpredictable incidents and their analysis requires the knowledge of the factors affecting them. Road and traffic accidents are defined by a set of variables which are mostly of discrete nature. The major problem in the analysis of accident data is its heterogeneous nature. Thus heterogeneity must be considered during analysis of the data otherwise, some relationship between the data may remain hidden[4]. Although, researchers used segmentation of the data to reduce this heterogeneity using some measures such as expert knowledge, but there is no guarantee that this will lead to an optimal segmentation which consists of homogeneous groups of road accidents[5]. Therefore, classified analysis can assist the segmentation of road accidents. Classification which is an important data mining technique can be used as a preliminary task to achieve various goals.

The objective of this system is to provide safety suggestions to government transportation agencies and to normal citizens based on the data of the past accidents and its location. As to evaluate our proposed method we perform data processing and association rule mining to provide us with association between accidents and variables related to accidents.

II. PROPOSED FRAMEWORK

Roadway traffic safety is a major concern for transportation governing agencies as well as ordinary citizens. In order to give safe driving suggestions, careful analysis of roadway traffic data is critical to find out variables that are closely related to accidents like speed of the vehicles, severity of damage, climatic conditions, etc. In this project we apply statistics analysis and data mining algorithms on a Road Accident dataset as an attempt to address this problem. For proper analysis of the accidents following strategies will be considered:

A. Data Preprocessing

Data preprocessing is one of the important tasks in data mining. Data preprocessing mainly deals with removing noise, handle missing values, removing irrelevant attributes in order to make the data ready for the analysis. In this step, our aim is to preprocess the accident data in order to make it appropriate for the analysis.

B. Association Rules

Association rule mining is a very popular data mining technique that extracts interesting and hidden relations between various attributes in a large data set. Association rule mining produces a set of rules that define the underlying patterns in the data set. The associativity of two characteristics of accident is determined by the frequency of their occurrence together in the data set. A rule $A \rightarrow B$ indicates that if A occurs then B will also

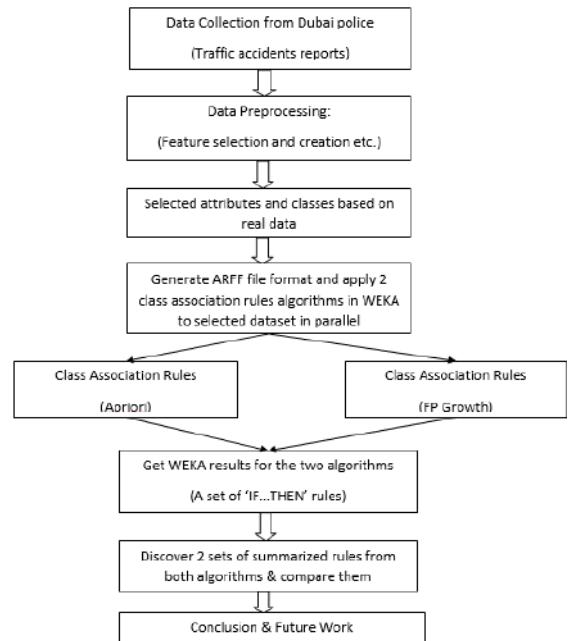


Fig 1: Proposed Framework for analysis[2]

occur.

- Apriori algorithm^[3] was proposed by Agrawal and Srikant to find the association rules from large datasets. The pseudo-code for traditional association rule mining algorithm for frequent itemset generation in figure 2.

```

Lk = {Frequent itemset of size k}
Ck = {Candidate itemset of size k}
L1 = {frequent 1 itemsets};
k=1;
while (Lk - 1 ≠ ∅) then
    Ck+1 = candidates generated from Lk
    For each transaction t ∈ D do
        Increment the count of candidates in Ck+1 that also contained in t
        Lk+1 = candidates in Ck+1 with minimum support
        k=k+1;
    Return UkLk
```

Fig 2: Apriori algorithm

An association rule is considered as a strong rule if it satisfies a minimum confidence and support. A minimum support S of a rule $A \rightarrow B$ indicates that in $x\%$ of all transactions A and B together occurs; whereas a confidence C of a rule indicates that in $C\%$ of all transaction when A occur then B also occurs.

$$\text{Support} = P(A \cap B)$$

$$\text{Confidence} = P(A|B) = \frac{P(A \cap B)}{P(A)}$$

- The FP-Growth (Frequent Pattern) Algorithm, proposed by J. Han, is a method for mining the complete set of frequent patterns by pattern fragment growth, using an extended prefix-tree structure for storing compressed and crucial information about frequent patterns named frequent-pattern tree (FP-tree). compressed and crucial information about frequent patterns named frequent-pattern tree (FP-tree).

For so much it uses a divide-and-conquer strategy. The core of this method is the usage of a special data structure named frequent-pattern tree, which retains the itemset association information.

In simple words, this algorithm works as follows:

- It compresses the input database creating an FP-tree instance to represent frequent items.
- It divides the compressed database into a set of conditional databases, each one associated with one frequent pattern.
- Finally, each such database is mined separately. Using this strategy, the FP-Growth reduces the search costs looking for short patterns recursively and then concatenating them in the long frequent patterns, offering good selectivity.

C. Data set description

An Accident data for this research were obtained from the United Kingdom Government website. The data set consists of 555 road accidents for the year 2016, in UK. After preprocessing of the data, 9 variables were identified satisfactory for the research (as shown in Table: II.1). The data set comprised of accident characteristics time of day, road and lighting conditions, number of injured victims, victims age and gender, type of vehicle, road surface and locality, etc.

Table1: Dataset Description

S.No	Attribute	Code
1.	Time of Day	Morning Afternoon Evening Midnight
2.	1 st Road Class & Number	M62 A641 U A629 A58 A6025 A681 A646 A6033 A644 B6114 B6112 A649 A643 A6036 A6026 B611 B6147 A647 A672 A6029 A624 A626 A62 B6113 B6138 A1502
3.	Road Surface	1 = Dry 2 = Wet / Damp 3 = Snow 4 = Ice / Frost
4.	Lighting Conditions	DayLight StreetLights

		Darkness
5.	Casualty Class	1 = Driver 2 = Vehicle 3 = Pedestrian
6.	Casualty Severity	1 = Fatal 2 = Serious 3 = Slight
7.	Sex of Casualty	1 = Male 2 = Female
8.	Age of Casualty	Childhood Teenage Youngster Midage Senior
9.	Type of Vehicles	Car Taxi 2wheel Goods Vehicle Other Bus

III. PROPOSED FRAMEWORK

A. Classification

The basic requirement for classification of data is to identify the constraints required to classify data into several classes. After identifying the constraints, we may apply simple classifier algorithm to segment the accident data set. After getting appropriate segmentation of the data set, our next task is the characterization of each class. Thus, we characterize our obtained classes as: fatal accidents, serious accidents and slight accidents.^[1]

Table 4 Best Rules for Slight class using Apriori Algorithm

Rule No.	Best Rule
1	Time (24hr)=afternoon Road Surface=1 Sex of Casualty=1 80 ==> Lighting Conditions=daylight 77 <conf:(0.96)>
2	Time (24hr)=afternoon Sex of Casualty=2 Type of Vehicle=car 88 ==> Lighting Conditions=daylight 84 <conf:(0.95)>
3	Lighting Conditions=daylight Casualty Class=1 Sex of Casualty=2 87 ==> Type of Vehicle=car 82 <conf:(0.94)>
4	1st Road Class & No=U Sex of Casualty=2 98 ==> Type of Vehicle=car 90 <conf:(0.92)>
5	Time (24hr)=afternoon Sex of Casualty=2 97 ==> Type of Vehicle=car 88 <conf:(0.91)>

Table 5: Best Rules for Fatal class using Apriori Algorithm

RuleNo.	Best Rule

1	Road Surface=1 Lighting Conditions=streetlights 4 ==> Time (24hr)=evening 4 <conf:(1)>
2	Lighting Conditions=daylight 4 ==> Road Surface=1 Casualty Class=1 4 <conf:(1)>
3	Time (24hr)=evening 5 ==> Road Surface=1 5 <conf:(1)>
4	Sex of Casualty=1 Type of Vehicle=2wheel 4 ==> Casualty Class=1 4 <conf:(1)>
5	Time (24hr)=evening Road Surface=1 Type of Vehicle=car 4 ==> Lighting Conditions=streetlights 4 <conf:(1)>

Applying Apriori algorithm on Slight class accidents (Table: III.2) shows us that most of the accidents happen in daylight. Also, male accidents occur more during afternoon when road surface is Dry.

Apriori algorithm when applied on Sever class accidents (Table: III.4), it shows that most accidents happen during daylight when vehicle is two-wheeler and the road surface is Dry.

Apriori Rule mining on Fatal class accidents (Table: III.3) indicates that most number of fatal accidents occur during evening. Rules show that male accidents occur more when they are on two wheelers. Many number of accidents have occurred during evening when the streetlights were on and the road surface were dry.

All these segments are further analyzed using association rule mining to find the correlation among different attributes in the data.

B. Association Rule Mining

We have applied Apriori algorithm and FP Growth algorithm on each class to generate association rules. In order to generate association rules a minimum 20 % support values is selected. These rules are also evaluated on the basis of confidence and lift measures. The strong rules with high lift value are considered for analysis. The strong 5 rules for each cluster have been shown in this paper.

Table 6: Best Rules for Severe class using Apriori Algorithm

Rule No.	Best Rule
1	Road Surface=1 Lighting Conditions=daylight Type of Vehicle=2wheel 11 ==> Casualty Class=1 11 <conf:(1)>
2	Time (24hr)=afternoon Road Surface=1 Type of Vehicle=car 10 ==> Lighting Conditions=daylight 10 <conf:(1)>
3	Lighting Conditions=streetlights Casualty Class=2 Type of Vehicle=car 10 ==> Sex of Casualty=1 10 <conf:(1)>
4	Time (24hr)=midnight Lighting Conditions=streetlights 12 ==> Sex of Casualty=1 11 <conf:(0.92)>
5	Road Surface=1 Casualty Class=1 Sex of Casualty=1 Type of Vehicle=2wheel 11 ==> Lighting Conditions=daylight 10 <conf:(0.91)>

Table 7: Best Rules for Fatal class using FP Growth Algorithm

Rule No.	Best Rule
1	[Casualty Class=1_binarized=1, Time (24hr)=evening_binarized=1]: 4 ==> [Road Surface=1_binarized=1]: 4 <conf:(1)>
2	[Type of Vehicle=car_binarized=1, Time (24hr)=evening_binarized=1]: 4 ==> [Road Surface=1_binarized=1]: 4 <conf:(1)>
3	[Type of Vehicle=car_binarized=1, Lighting Conditions=streetlights_binarized=1]: 4 ==> [Road Surface=1_binarized=1]: 4 <conf:(1)>
4	[Road Surface=1_binarized=1, Lighting Conditions=streetlights_binarized=1]: 4 ==> [Time (24hr)=evening_binarized=1]: 4 <conf:(1)>
5	[Type of Vehicle=car_binarized=1, Time (24hr)=evening_binarized=1]: 4 ==> [Road Surface=1_binarized=1, Lighting Conditions=streetlights_binarized=1]: 4 <conf:(1)>

Applying FP Growth algorithm on Fatal accidents (Table 5) tell us that during evening where the streetlights are on and the road surface is dry.

Table 8: Best Rules for Slight class using FP Growth Algorithm

Rule No.	Best Rule
1	[1st Road Class & No=U_binarized=1, Time (24hr)=afternoon_binarized=1]: 81 ==> [Lighting Conditions=daylight_binarized=1]: 79 <conf:(0.98)>
2	[Road Surface=1_binarized=1, Type of Vehicle=car_binarized=1, Time (24hr)=afternoon_binarized=1]: 109 ==> [Lighting Conditions=daylight_binarized=1]: 104 <conf:(0.95)>
3	[Time (24hr)=afternoon_binarized=1, Age of Casualty=midage_binarized=1]: 77 ==> [Lighting Conditions=daylight_binarized=1]: 72 <conf:(0.94)>
4	[Casualty Class=1_binarized=1, Sex of Casualty=2_binarized=1]: 104 ==> [Type of Vehicle=car_binarized=1]: 97 <conf:(0.93)>
5	[Sex of Casualty=2_binarized=1, 1st Road Class & No=U_binarized=1]: 97 ==> [Type of Vehicle=car_binarized=1]: 89 <conf:(0.92)>

FP Growth algorithm on Sever class (Table 7) results that most of sever accidents take place during midnight even though there are street lights were on.

FP Growth algorithm when applied on Slight class accident data (Table 6) indicate that most of the accidents occur in car, especially during afternoon when the lighting conditions is daylight.

IV. CONCLUSION

The project developed would shape the future. The whole purpose of this system is to provide safety suggestions people based on the data of the past accidents and its location. As to evaluate our proposed method we perform data processing and various data mining methods such as association rule mining to provide us with association between accidents and variables related to accidents. We found out that FP Growth algorithm outperforms Apriori algorithm as compared through results. Unfortunately, FP Growth is computationally expensive, especially when a large number of patterns exist. As a part of future work we can create a mobile application the can track user's location to provide safety suggestion to the user around him. Table 9: Best Rules for Severe class using FP Growth Algorithm

Rule No.	Best Rule
1	[Road Surface=1_binarized=1, Time (24hr)=afternoon_binarized=1]: 15 ==> [Lighting Conditions=daylight_binarized=1]: 15 <conf:(1)>
2	[Lighting Conditions=daylight_binarized=1, Type of Vehicle=2wheel_binarized=1]: 14 ==> [Casualty Class=1_binarized=1]: 14 <conf:(1)>
3	[Road Surface=1_binarized=1, Lighting Conditions=daylight_binarized=1, Type of Vehicle=2wheel_binarized=1]: 11 ==> [Casualty Class=1_binarized=1]: 11 <conf:(1)>
4	[Time (24hr)=midnight_binarized=1]: 13 ==> [Lighting Conditions=streetlights_binarized=1]: 12 <conf:(0.92)>
5	[Road Surface=1_binarized=1, Age of Casualty=seniors_binarized=1]: 12 ==> [Lighting Conditions=daylight_binarized=1]: 11 <conf:(0.92)>

REFERENCES

- [1] Liling Li, Sharad Shrestha and Gongzhu Hu. Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques. 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA).
- [2] Amira A El Tayeb, Vikas Pareek, and Abdelaziz Araar. Applying association rules mining algorithms for traffic accidents in Dubai. International Journal of Soft Computing and Engineering, September 2015.
- [3] Sachin Kumar and Durga Toshniwal. Analysing road accident data using association rule mining. In Proceedings of International Conference on Computing, Communication and Security, pages 1–6, 2015.
- [4] KMA Solaiman, Md Mustafizur Rahman, and Nashid Shahriar. Avra Bangladesh collection, analysis & visualization of road accident data in Bangladesh. In Proceedings of International Conference on Informatics, Electronics & Vision, pages 1–6. IEEE, 2013.
- [5] Heydar Jafarzadeh, Rouhollah Rahmati Torkashvand, Chamran Asgari and Amir Amiry Provide A New Approach for Mining Fuzzy Association Rules using Apriori Algorithm.

Content Management System Using Web Technology

Saurabh Singh¹, Ajaykumar Verma², Shubham Vishwakarm³, Rahul Singh⁴, Sheetal Rathi⁵

Department of Computer Engineering, Thakur College of Engineering and Technology, Mumbai, India

saurabhasingh1997@gmail.com, ajayverma1244@gmail.com, vshubum@gmail.com, rs315@gmai.com, Sheetal.rathi@thakureducation.org

Abstract-The CMS is an asset to the company. But it has been overlooked to be as a source of a competitive advantage. The CMS is been considered as a platform, and it could help organizations use the web which the easiest way to reach users in challenging times and even give them a competitive advantage over competitors.

A CMS is considered as a growing platform for the relevant staff to manage a web experience, without needing much of technical skills. It brings built-in compliance and security standards to allow the following:

Content creation and contribution

Management of content and Administration

Content publishing and serving

Both the enterprise content management and also the web content management systems have two major components: the content management application (CMA) and content delivery application (CDA).

The CMA is a known graphical user interface (GUI) that allows the user to control the product creation, modification and the removal of content from a website without needing to know anything about web technology. The backend services are provided by CDA, that also supports management and the delivery of the content once it that has been created in the CMA.

I. INTRODUCTION

A content management system or (CMS) is the software or set of interrelated programs that is used for creation and managing digital content over web. CMS's are typically used for the Content management of Enterprise (ECM) and also the web content management (WCM) whereas it has other uses as well. But ECM and WCM are the most important aspects due to which CMS are in use nowadays. An ECM provides facilities as such collaborative effort in the workplace which can be achieved just by the integration of document management, and also digital asset management over web which is of great importance to the company and records functionalities for retention, and also provide end users with access to the organization important digital assets which is role based. Whereas WCM provides facilities such as collaborative effort in authoring for websites and others as well which are of great importance to the

enterprise using the CMS for their company. ECM software mainly includes a WCM functionality for publishing which is an add-on with other features, but the ECM pages usually remain behind firewall of the organization.

There are various business benefits of using CMS.

The business benefits include:

- Reduce the site maintenance costs
- Basically increase on content publish efficiency
- Very-less long-term dependencies on specialist skill sets
- Greater or increase in visual consistency
- Distributing the content authoring to a wider team with minimal training

II. BACKGROUND

- Different features of CMS vary from CMS to CMS, but the underlying principle of operation remains the same such as indexing, searching, retrieving, managing different formats and publication.
- Indexing, retrieving and searching features index all data so that they can be easily accessed by search functions and let users to search usually done by keywordssuch as dates, attributes or author name.
- Format management simplifies the process of transforming scanned documents and electronic documents into HTML or other format documents such as PDF.
- Revision features deals with periodic updating and revision of the content after initial publication. Revision control keeps track of changes made by user.
- Publishing functionality permits users to use a provided template or a set of templates authorized by the organization, as well as wizard and other corresponding tools for creating or modifying content.
- Other popular features of CMS's include:
 - i. Search Engine Optimization supported website URLs
 - ii. Online help facility is integrated along with the privilege of discussion forums.
 - iii. Permission modules based on groups.

- iv. Supports customized & non-customized templates.
- v. Simple to install (like wizard installation process)
- vi. Admin community with bi-lingual support.

III. PROBLEM IDENTIFICATION

There exist many advantages of using Web based CMS technology over other existing systems or techniques in corporate world. Companies find it very pleasing and profitable to use due to the advantages listed below. These basically sums up what actually a web based CMS can provide to the existing Systems. These are the benefits which leads to enforcing a company to use a CMS system. Content management systems are basically often considered as a silver bullet which will eventually solve all your content problems. In reality existing/other CMS is not enough as it comes with its own limitations which is been improved upon or being solved in our web based CMS.

The following is the top limitations which is been solved in our web based CMS.

- a) *A lack of an editorial control:* It has consequence such as lack of central control which will ensure the quality and accuracy of copy produced. Which will eventually lead to contradictions and difference in styles of writing across the site.

Our Solution: Get an editor to control editing

- b) *Reputation of contributors:* The greatest and biggest selling points of having a content management system is that they eventually allow anybody to post to your website which is one of its selling points and it comes with a disadvantage. Unfortunately, just because your staff can edit the site, does not guarantee that they will definitely do so and there is a danger of repudiation as well.

Our solution: Recognize the importance of the web

- c) *Poorly trained authors:* Whenever an organization buys out a new content management system they almost always offer some form of training which is pretty essential with the complexity of existing CMS to use for the members of the organization. However, in many cases it is not enough.

Our solution: We will provide video training material

- d) *No clear calls to action:* As we have already said, almost all the content providers are focus on conveying the information to others rather than meeting user needs. Until the user is ready to take those major steps they are left to wander around the website. However, they are also neglecting the business needs too. Even when content providers are thinking about calls to action, they are focusing on the big actions such as "contact us."

Our solution: Always guide the user to the next action

- e) *No sense of community:* Increasingly, content management systems come with some great community tools. They have forums, comments and integrate with everything from Facebook to Twitter. However, great technology does not build great communities.

IV. LITERATURE SURVEY

It is clear from the society survey and feasibility study of the system that this idea will majorly attract the people who are into their own startup firm, blogger, etc. And those who are not into development i.e. not having that much technical skills or do not want to invest their time into development can use this web based CMS. These are the target audience which we have aimed through web based CMS. In CMS there is no involvement of technical tools and it is very easy to deploy so startup idea can be converted into firm with no delay [1].

Any development of a system is directly or indirectly related to society so Society Survey is an essential part of the development phase [2]. The development of a system may either affect the society in a positive way or a negative way.

- *Moving from CaaS to EaaS:*
There was a rage of headless CMS tools in 2016, especially among enterprise architects who were looking for maximum front-end flexibility [1]. But the future trend involve that the intelligent WCM architectures will supply EaaS as well. It sounds easy but it is not easy as it sounds.
- *Shifting of Marketplace towards Complexity:*
This is current trend in the world of CMS but it will really boast in the upcoming future as digital increasingly becomes an arena of experts [4].
- *Entire Cloud shift towards PaaS:*
The current trend clearly shows that cloud has hit WCM. Nowadays most organizations are looking to implement CMS technology purely based on CLOUD [6].
- *WCM vendors offering Marketing Data Warehousing:*
The scope of WCM is a debatable topic. One key question that arises that "Should our WCM vendors should offer marketing or Not?" But, in reality, almost every enterprise will want to keep their data and content and also their engagement tiers distinct and separate which will eventually offer more sort of flexibility to the organization using the CMS [3][4].
- *Growing Emphasis on Sharable Micro Experience:*
It can be said to be a sort of prediction but in current trends in enterprise world regarding CMS.

If the development is not acceptable to the society then there is not much scope of it being accepted in the world or at larger scale so on basis of which it can be concluded whether to undergo development phase or not. Thus, a survey must be

carried out before hand of what they feel of this development and take their considerations into account before starting the project development. Hence, we too asked people in our neighborhood about our project whether they would like to use a system which will help them to exhibit their products and services to the end users and users will get notifications of the same and the response was positive. In today's world, where majority of people use mobile applications for their day to day needs, the application for advance marketing will be a success.

V. PROPOSED MODEL

Our proposed model will overcome the following drawbacks as we have studied in the literature survey. Our proposed model is moving from CaaS to EaaS which is a very big advancement over the existing CMS Systems. The current trend in the world of CMS but it will really boast in the upcoming future as digital increasingly becomes an arena of experts.

It is clear from the society survey and feasibility study of the system that this idea will majorly attract the people who are into their own startup firm, blogger, etc. And those who are not into development i.e. not having that much technical skills or do not want to invest their time into development can use this web based CMS. These are the target audience which we have aimed through web based CMS. In CMS there is no involvement of technical tools and it is very easy to deploy so startup idea can be converted into firm with no delay. These observations from the Literature Survey can be used as the features that we will incorporate in our model. Our model that is our website is focusing on incorporating the features we found to be unavailable in the current systems but may be able to see it in near future. Our proposed model will offer a great user experience for both the users using it and the client using our CMS to build their own site which is our goal. With further release we are focusing to incorporate the future features which are not available currently in any CMS or if available then only in few of them which will be our selling point. Or the point which will make the entrepreneurs who are not willing to invest a much in setting up the online business.

We will also provide the basic or core features provided by most of the Content Management Systems. Which include,

- i. Search Engine Optimization supported website URLs
- ii. Online help facility is integrated along with the privilege of discussion forums.
- iii. Permission modules based on groups.
- iv. Supports customized & non-customized templates.
- v. Simple to install (like wizard installation process)
- vi. Admin community with bi-lingual support.

Figure 1 is the block diagram of our Content Management System,



Figure 3: Blog Dummy Website based on our

The figure 1 is a block diagram gives a gist of information about the proposed model. To be true it acts as a skeleton to our Content Management System.

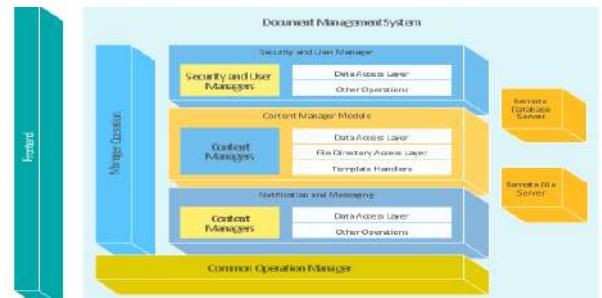


Figure 1: Block Diagram of Content Management System

VI. RESULTS AND DISCUSSION

There exist many advantages of using Web based CMS technology over other existing systems or techniques in corporate world. Companies find it very pleasing and profitable to use due to the advantages listed below. These basically sums up what actually a web based CMS can provide to the existing Systems. These are the benefits which leads to enforcing a company to use a CMS system.

- 1) It Improves Digital Marketing.
- 2) Reduces the time required and money as well.
- 3) Provides different access levels to various users.
- 4) Convenient and easy to use and content can be updated regularly without the involvement of the development team
- 5) Ability to make design changes.
- 6) CMS web design enables easy customizations and enhancements
- 7) Ability to add new features via plug-ins.
- 8) Enables workflow management.

The end product will be almost same as the proposed model. But more likely that new features will be added as soon new technology is been introduced in the market. But our main objective/ focus is to provide the clients a free experience of Content Management System with great user experience.

The greatest and biggest selling points of having a content management system is that they eventually allow anybody to post to your website which is one of its selling points and it comes with a disadvantage. Unfortunately, just because your staff can edit the site, does not guarantee that they will definitely do so and there is a danger of repudiation as well.



Figure 2: Dummy Blog Website

But our System does provide ability to control and make the employees of the stakeholders to make changes or updates to the System, so that the system can be up-to-date.

The following are the images of a dummy blog website been built on our Content Management System,

- The first image is the dashboard of the Blog website been built on our CMS.
- The image clearly shows that it provides various features.
- The second image (figure 3) is the image which clearly shows that our system also provide profile view feature which is rarely found in a CMS.
- It also provides basic information about the user and activity performed by that particular user.

VII. CONCLUSION

A content management system (CMS) is a software application or set of related programs that is used to create and manage digital content over web. CMS's are typically used for enterprise content management (ECM) and web content management (WCM) whereas it has other uses as well. But ECM and WCM are the most important aspects due to which CMS are in use nowadays.

While of all the existing CMS WordPress seem the best CMS platform compared to other CMS such as Drupal and Joomla platforms, there still are situations wherein you might not find WordPress a worthy choice.

Our proposed system stands out at this point. It offers both continuity and reliability which is rarely available together. But the success of any project considering the fact depends on human factors, proposed CMS is undoubtedly the ideal choice as a reliable CMS platform. After all, it is very straightforward and extremely easy to operate and thereby reduces the chances of the human bottleneck and on the other hand it increases the productivity to a great extent.

VIII. FUTURE SCOPE

Based on our Literature Survey the CMS can see a variety of change in near future.

- *Moving from CaaS to EaaS:* There was a rage of headless CMS tools in 2016, especially among enterprise architects who were looking for maximum front-end flexibility. But the future trend involve that the intelligent WCM architectures will supply EaaS as well. It sounds easy but it is not easy as it sounds.
- *Shifting of Marketplace towards Complexity:* This is current trend in the world of CMS but it will really boast in

the upcoming future as digital increasingly becomes an arena of experts

- *Entire Cloud shift towards PaaS:* The current trend clearly shows that cloud has hit WCM. Nowadays most organizations are looking to implement CMS technology purely based on CLOUD
- *WCM vendors offering Marketing Data Warehousing:* The scope of WCM is a debatable topic. One key question that arises that “Should our WCM vendors should offer marketing or Not?” But, In reality, almost every enterprise will want to keep their data and content and also their engagement tiers distinct and separate which will eventually offer more sort of flexibility to the organization using the CMS.
- *Growing Emphasis on Sharable Micro Experience:* It can be said to be a sort of prediction but in current trends in enterprise world regarding CMS. This will eventually present two major challenges to the existing web based CMS systems in order to prevail and sustain in this trend. Which can be summarized as
 - a) Opening up new platforms on their own for the other non-enterprise contributors.
 - b) Providing omni-channel capabilities to preview the marketplace

IX. References

- [1] <https://www.trainingindustry.com/glossary/content-management-system-cms/>
- [2] <http://www.accuracy.com/web-content-management-systems-intranet>
- [3] <https://snap360.ca/website-design-and-development/content-management-systems-giving-you-control/>
- [4] <https://www.smartbugmedia.com/blog/4-guidelines-for-choosing-the-best-content-management-system-cms-for-a-business-website>
- [5] <https://pimcore.com/nl/resources/blog/why-you-should-be-using-a-content-management-system-cms-a981>
- [6] <https://www.makeuseof.com/tag/10-popular-content-management-systems-online/>
- [7] <https://www.everyinteraction.com/definition/content-management-system-cms/>

Real Time Monitoring And Control In Irrigation System Using IoT

Vishal Rajak¹,Ninad Virkar², Prakash Shukla³, Anurag Shetty⁴, Veena Kulkarni⁵
Computer Engineering, Thakur College of Engineering and Technology, Mumbai, India

Abstract-The population of our country is increasing day by day, after a few decades there will be a serious problem of food if we do not change the traditional way of growing crops. The agricultural needs to be modernized by using the new technologies. The main objective of this paper is to automate the irrigation process which not only saves time and money but also reduces the effort thereby helping the farmers. The system uses temperature, humidity and moisture sensors. The sensors continuously monitors the surroundings and provide this information to the user in real time. The system proposes a soil moisture sensor at each place where the moisture has to be scrutinized. Once the moisture reaches a threshold value, the system takes applicable steps to normalize or even stop the flow of water. The system consist of a GSM module which can help to send notifications to our users easily. Sensors are used to automate the irrigation process in agriculture. To provide the smart irrigation further smart sensors can be implanted in the farmland soil that can easily analyses the degree of moisture. The paper focuses on providing an economical yet efficient product for farmers and other entities. The proposed system will be to easy to use for all types of people.

Keywords-*IoT, sensors, GSM, arduino*

I. INTRODUCTION

A. BACKGROUND

Two of the most common problems with manual farming system process is with watering schedules. Irrigation scheduling and planning is subtle answers to the queries of "Best time to on the water pump?" and "By what time I should water?" Initializing an irrigation phase prematurely and/or running a watering process extensively is considered surplus watering. At the very least this practice wastes water and money. Overwatering the crops is harmful and damages them. Furthermore, it can have a serious impact on their production. Likewise, starting an irrigation cycle lately or not administrating the system for a long enough period of time is considered under watering and can be the reason of abridged harvests and deprived crop quality affecting the rates. Highlights on to such concerns is the key to curtail their fiscal and concrete effect on crops. Approaches to deal with these complications is to introduce an analytical object detection system comprising of soil moisture based control systems. These machineries embedded with sensing devices and moisture sensors estimates the tangible soil moisture. This routine is characteristically provides genuine and accurate results because it is measuring the moisture level of the soil. Soil moisture control systems indicates the grower when to initiate the watering course and also state him when the soil moisture level extents field capacity.

B. PROBLEM IDENTIFICATION

To satisfy the need for increasing demand of food, it is necessary that we revolutionize the agricultural process using advanced technologies. To meet the demand for food agriculture is the only source. Agriculture plays an important role in developing country like India. The farmers use irrigation to water their crops which is essential for their proper growth. Irrigation is the process of giving water to crops. The amount of water required by plants maybe same or different depending upon them.

II. PROPOSED METHODOLOGY

In our system, the sensors will collect the information from the farm and then they will pass this information using the GSM and GPRS module which will be displayed to the user in the app/website. The user can then control by switching the water pump on or off. Arduino microcontroller is the brain of our system which uses logic to turn on/off the pump. The output of soil sensor circuit is directly connected to digital pin D7 of Arduino. A LED is used at the sensor circuit, this LED's ON state indicates the presence of moisture in the soil and OFF state indicates the absence of moisture in the soil. GSM module is used for sending SMS to the user. Here we have used TTL SIM800 GSM module, which gives and takes TTL logic directly (user may use any GSM module). A LM317 Voltage regulator is used to power the SIM800 GSM module. LM317 is very sensitive to voltage rating and it is recommended to read its datasheet before use. Its operating voltage rating is 3.8v to 4.2v

III. PROPOSED SYSTEM

The system will monitor the soil moisture present in the soil using a soil moisture sensor and accordingly turn the pump on/off for the irrigation thereby automating the process. The system uses soil moisture sensors which are placed at appropriate positions. These sensors monitor the soil moisture value and compares it with the threshold value. If the current value is less than threshold, it turns on the pump. The system consist of a GSM module which can help to send notifications to farmers. To solve the problems faced in the traditional irrigation process, smart irrigation uses modern technologies to reduce wastage of time and effort of farmers thereby using the required water in amore effective way.

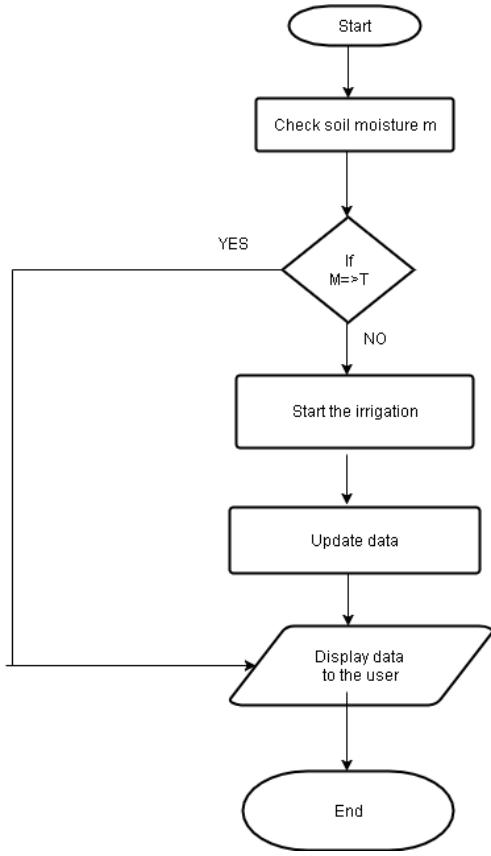


Figure 1. Implementation

IV. CONCLUSION

One of the most drastic problem faced by farmers in manual irrigation system is the approach of timely watering the farmlands. The reason to such arising situation is due to lack of knowledge to farmers about watering patterns of specific soil variety. Controlled watering patterns and analytical data retrieved from the sensors overshadows the approach to manual irrigation. This ideology provides ease to farmers by facilitating them with remote access system , precise analysis of moisture and other determining parameters as well as scheduled water dripping provision. Here is an idea that helps farmers to water crops in an effective way, reduce wastage of water and automates the process of irrigation which reduces wastage of time of farmers.

IV. FUTURE SCOPE

Our system is easily scalable. It can be used on a small scale to water plants in our house, gardens and public parks. The system can be used to water stadiums thereby reducing loss of water.

REFERENCES

- [1] Cuihong Liu WentaoRenBenhua Zhang ChangyiLv, The application of soil temperature measurement by LM35 temperature sensors, International conference on Electronic and Mechanical Engineering and Information Technology (EMEIT),2011
- [2] G. Yang, Y. Liu, L. Zhao, S. Cui, Q. Meng, and H. Chen, "Automatic irrigation system based on wireless network", Proc. of 8th IEEE International Conference on Control and Automation, Xiamen, pp. 2120-2125, Jun. 2010
- [3] Jiachun Li WenteiTu, An Irrigation Control System Based on C8051F and ARM, International Conference on Digital Manufacturing and Automation (ICDMA), 2010
- [4] Zhiyu Li, Qiangui Xiao. "Design of temperature control system based on LM35 temperature sensor," Micro computer information, 20(9), 2004, pp. 51-52
- [5] N. Shah and I. Das, "Precision Irrigation Sensor Network Based Irrigation", a book on Problems, Perspectives and Challenges of Agricultural Water Management, IIT Bombay, India, pp. 217-232, April 2008
- [6] Daniel K.Fisher and HirutKebede, a Low Cost Microcontroller-Based System to Monitor Crop Temperature and Water Status, pp. 168-173, 2010
- [7] P. Corke , T. Wark , R. Jurdak , H. Wen , P. Valencia and D. Moore "Environmental wireless sensor networks", Proc. IEEE, vol. 98, pp.1903 -1917 2010
- [8] Y. Kim, R. G. Evans, and W.M. Iversen, "Remote sensing and control of an irrigation system using a distributed wireless sensor network", IEEE Transactions on Instrumentation and Measurement, Germany, vol. 57, issue 7, pp. 1379-1387, Jul. 2008.
- [9] Shen Jin, Song Jingling, Han Qiuyan, Wang Shengde, and YangYan, "A Remote Measurement and Control System for Greenhouse Based on GSM-SMS", IEEE 8th International Conference on Electronic Measurement and Instrument, 2007,pp. 45-82.
- [10] Pranita A. Bhosale, Prof. V. V. Dixit," International Journal of Scientific & Technology" vol 1, pp118- 123,2012 [11] Y. Zhou, X. Yang, L. Wang, and Y. Ying, "A wireless design of low-cost irrigation system using ZigBee technology", Proc. of IEEE Conference on Networks Security, Wireless Communications and Trusted Computing, Hubei, vol. 1, pp. 572-575, April 2009.

Agile in Automotive Manufacturing

Tasneem Mirza, Archana Patankar

TSEC, Mumbai

tmm1207@gmail.com, patankararchu@gmail.com

Abstract - The automotive industry is a very traditional industry. But over the past decade the automobile industry is buzzing with advancements from electric cars to vehicle to vehicle connectivity to self-driving cars. Customers want a luxurious driving experience and no longer see the car just as mode of travel. And with this growth there is a need to change. As the number of cars on the roads increases, the automotive manufacturers must simplify and shorten the product development process. One model which is heavily used in the software industry is the agile model. This paper focuses on the challenges the automotive industry currently faces and reviews the applicability of agility with respect to automotive industry i.e. Agile Project Management, Agile Product development and Agile Supply Chain design.

I. INTRODUCTION

The idea of agile was originally coined by the manufacturing industry, but it was the software industry that actually adopted the agile practices and principles and took advantage of it. Since the release of the Agile Manifesto, software industries have successfully adopted the agile principles and practices [1]. The automotive industry is a very traditional industry. But over the past decade the automotive is growing and evolving. And with this growth there is a need to change Today, automotive buyers want their vehicles to be equipped with all of the latest technologies. Customers are demanding state-of-the-art cars that ensure a safe, comfortable, and luxurious driving experience. Cars are becoming smarter, more connected and easier to use. Customers no longer accept standardized products but want products that satisfy their individual requirements. Also, automotive manufacturers are facing tough competition from its competitors to offer the latest technologies to its customers in minimum possible time. Automotive manufacturers and suppliers are confronted with increasing complexity as a result of increasing numbers of products and options, shorter technology cycles and increasing pressure to innovate. The challenge automotive manufacturers are facing currently is whether they have the expertise and tools to deal with these changes.

II. CHALLENGES IN AUTOMOTIVE INDUSTRY

The global automotive industry is growing and with the growth there is a need to change. The typical challenges faced by OEMs (Original equipment Manufacturers) today are:

A. Increasing complexity in a bid to offer the latest technology

In the past automotive manufacturers had lesser number of attributes and longer production cycles. But in recent times for automotive manufacturers to remain competitive they have to offer the customer a variety of product attributes keeping in mind the latest technologies like internet of things which adds to the complexity of the industry[2].

B. Shorten the production cycle

The production life cycle of any car is 3+ years. But the automotive industry is innovating every 2-3 years. At this rate any automotive manufacturer will not be able to survive in the market. Hence what is needed is to shorten the production life cycle of a car[3].

C. Increasing pressure to innovate

Currently there is a wave of innovation in the automotive industry. The car industry is in a race as to who can deliver the latest innovations. Innovations in terms of fuel efficiency, connected cars, in-car infotainment to self-driving cars.

D. React quickly to fast changing consumer demands.

For the young generation today, a car isn't just a way to travel from one location to another. For many it is a status symbol, a piece of luxury, fast service and they also expect their cars to be environment friendly. Nowadays, the customers' expectations and their satisfaction levels are manifold. Thus, automotive manufacturers should be able to quickly respond to the volatile market.

The question here is that, can the OEMs remain competitive in this turbulent and volatile market. The answer is they can but for that new production concepts and mindset have to be developed to manage the complexity in car production. As the number of cars on the roads increases, the automotive manufacturers can shorten the product development process by using concepts of the Agile model. One model which is heavily used in the software industry is the agile model which could help to overcome the challenges faced by the OEMs today. Internet of Things is reshaping almost all industry sectors. Applications like Smart Home, Smart Manufacturing, Smart Healthcare, Smart City, Smart Farming, Connected Cars and Wearables amongst many others are transforming the Businesses and enhancing the Customer experience. This paper identifies the innovative applications of IoT in Automotive industry in the areas of Connected Car services/applications, Vehicle Communications, IoT in Intelligent Transportation, IoT based SupplyChain Management in Automotive Industry and New Generation Cars.

III. AGILE MODEL

Agile is a new way of thinking. Agile software development in the field of software engineering comprises of a set of management practices and values centered around the customer and uses iteration and incremental development. Here the user requirements and the solution to the problem domain evolve as a result of collaboration between self-organizing and cross functional teams. Since the past 12 years software industry is taking the advantage of agile principles and concepts. The question here is whether agile can be applied to manufacturing industries. Technological innovation has brought major changes in

the automotive market. As a result, customers are becoming more demanding and there are frequent changes in the product design. Agile hence can be applied as a response to deal with volatility in the market [4].

Agile as a framework can be applied to a number of areas like project management, product development and supply chain. Fig III 1 shows the different possible dimension of Agile.

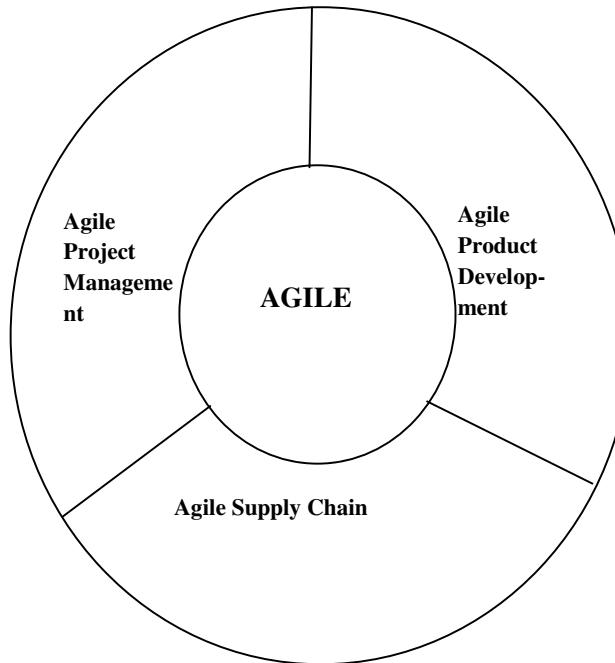


FIG.1: DIMENSIONS OF AGILE W.R.T AUTOMOTIVE MANUFACTURING:

A. Agile Project Management

Project planning and control is very challenging for industries involved in new product development and technologies. Project management is a set of processes that helps organizations to improve the quality of projects by including process descriptions, roles, responsibilities, work breakdown structures[4]. Traditional project management is suitable for projects where the requirements are very well defined and the rate at which the requirements change is very low.

Agile project management in contrast has the following characteristics

1. There is a high level of uncertainty with respect to the project goal and the customer requests are unpredictable or incomplete and may change during the duration of the project.
2. There is a close and frequent collaboration with the customer.
3. This approach lacks documentation.
4. Project plan is iterative

5. Focuses on self-organizing and cross functional teams

The project's goal is defined in less detail at the start and a project implementation schedule is roughly prepared. The project is divided into iterations where each iteration lasts 1 to 4 weeks. The most important functions are undertaken at the start and the least important ones are considered in the end. The implementation plan for each iteration is made by the project team and not the project manager which includes the detailed specification i.e. the tasks, hours of work etc. also taking into account the client's new wishes or the ideas of the developers. Daily stand up meetings are held where each developer updates the entire team about the progress of his/her work[5].

Despite the enormous popularity of agile project management in the field of software development, there is a reluctance in adopting this type of project management in non-IT sector since frequent changes are too expensive and partial deliverables or prototypes might not be used.

B. Agile Product development

Agile Product development focuses on engineering activities associated with the project's product eg. an automotive product. One of the most important benefits of Agile is that the cost of change is very low hence creating an environment for adaptable products and projects. For agile product development(hardware) to be successful, the cost of change in fabrication needs to be reduced. This is where 3 D printing technologies boost the agile values. In the absence of 3 D technology, traditional machining or fabricating techniques are generally costly and have lead times that are too lengthy to apply to a short agile iteration. 3 D printers have mobilised the production stage of automotive industry from prototyping to development of tools to the manufacturing of parts. One of the benefits of 3 D printing is rapid prototyping. With the use of 3 D printer's organizations can deliver a working prototype in a record time enhancing innovation and creativity.

Use of 3 D printers in automobile manufacturing can give the following benefits [7]:

1. RAPID PROTOTYPING:

3 D printers can be used in rapid prototyping in the pre-manufacturing phase by the automotive industry. This helps to realize the product faster and improving the product using customer feedback.

2. LOW TURNAROUND TIME:

Use of 3 D printers reduces the time to prototype hence the turnaround time of manufacturing across all subsequent stages. This leads to added business value and lowers the costs.

3. LOW CONSUMPTION AND WASTAGE:

Use of 3 D printers in automotive manufacturing leads to low consumption and wastage as compared to the traditional approaches.

4. INCREASED FLEXIBILITY:

3 D printers allows organizations to experiment with multiple options in the development stage. This helps the organization to be

more agile to make changes on the fly and helps to them to stay in tune with the market requirements.

5. REDUCING INVENTORY:

Currently millions of components are kept in the warehouse waiting to be supplied to the customers. One of the biggest advantages of 3 D printing is to store parts digitally and built them on demand hence reducing the inventory.

C. Agile Supply Chain

Automotive Supply chain is quite different from the supply chain of other industries like pencils, cricket bat or aerated drinks. There are automotive manufacturers like Ford, Mahindra, Toyota etc. called as OEMs (Original equipment manufacturers). Though they produce some original equipment's, their real strength is in marketing the cars, ordering parts from the various suppliers and final assembly of the car. Most OEMs create 30-35% of the product in house and the rest of the product is produced outside[8][9]. Companies that produce parts and supply directly to the OEMs are called Tier 1 suppliers. Tiers 2 suppliers are organizations that supply components to Tier1 suppliers. Tier3 organizations produce raw materials that firms in the supply chain require to make products and components. The complexity of the automotive supply chain can be understood from the fact that a typical car contains about 20,000 components[10]. The automotive supply chain consists of a large number of Tier1, Tier 2, Tier3 suppliers and also a number of dealerships. In short, the automotive supply chain is very complex. In addition, nowadays buyers demand for certain features and specific configuration which needs a quick response from the automotive supply chain. Also, as customer's buying patterns are changing rapidly hence there is a change in supply chain management. The solution to the various supply chain challenges is Agile supply chain. Agile supply chain is not an alternative but a necessity for the success of any organization in today's volatile market. Agile supply chain is an approach to manage the supply chain network flexibly in order to keep up to the customer demands.

To be an agile supply chain, a supply chain should have the following characteristics[11]:

1. VIRTUAL INTEGRATION

Virtual Integration means information about real market arising from customer demands is shared among all the members in the supply chain. Based on the requirements of the market, every member in the supply chain responds with its capability to fulfil the demand. This virtual integration results in end to end visibility and helps to identify bottlenecks in the supply chain network.

2. MARKET SENSITIVITY

Today the automotive supply chain is market driven. Forecasting demands of the market from past trends an obsolete way to sense the demand in such a volatile market. Nowadays organizations are focussing on future, hence they need to use best techniques to forecast the demand based on daily Point of Sale. This means that an agile supply chain should be demand driven and not forecast driven.

3. PROCESS ALIGNMENT

Nowadays as more and more organizations outsource every function, it is important to have a trust-based relationship between the suppliers and buyers. Transparency of information between the members of the supply chain network helps the organizations to survive in turbulent markets.

4. NETWORK BASED

There has to be a structure, co-ordination and better relationship management between all parties in the supply chain. The whole supply chain should be viewed as a single entity to respond to the customer demand and reduce the total cost of the supply chain.

IV. AGILE IN AUTOMOTIVE R & D^[6]

There is a growing scope of applying agile in automotive R & D. Today's automotive industry is facing a transformative change. Here mechanical centred industries need to compete with the digital industry, increasing customer demands, and intensive competition from its competitors. Currently automotive R & D is all about creativity, risk taking which includes a lot of undefined factors and unpredictability. The products designed by the R & D team have to be innovative, with a shorter time to market and more complex in nature. The existing processes used are highly structured and works only for predictable environments. In the wake of these conditions agile approach works for automotive R & D systems. Agile approach uses shorter iterations and emphasizes using a self-organizing and cross functional teams and changes requested by the customer can be incorporated easily since agile approach has a scope to incorporate variations. Hence agile approach can be successfully adopted by embedded and mechanics development projects also.

V. CONCLUSION

The automotive industry is buzzing with growth. Customer expectations are changing and so is the technology. Hence the need of the hour is to change the development process to overcome the challenges which includes shorten the product development life cycle of an automobile and quickly respond to the volatile market. One model which is used heavily in the software industry to respond to unpredictable customer requirements is the Agile model. An agile mindset is a new way of thinking. Agile. Agile as a framework can be applied to project management, product development and supply chain with respect to the automotive industry.

Agile Project Management focuses on requirements specification, project schedule(dividing the project into short iterations), team work and team collaboration(self-organizing and cross functional teams) and client collaboration. Agile product development focuses on the engineering activities associated with building the product i.e. the use of 3 D printers for producing prototypes, parts and tools. Agile supply chain focuses on making the supply chain network of an organization flexible and responsive in turbulent markets. Also, it is researched that there is a scope of applying agile to automotive R & D since automotive R & D is all about creativity, risk taking and unpredictability.

REFERENCES

- [1] <https://www.forbes.com/sites/stevedenning/2012/09/24/how-manufacturing-can-learn-from-software-to-become-agile/#52107466bd6b> [Last Accessed 11/12/2017]
- [2] H. Schleich, J. Schaffer, L. F. Scavarda, "Managing complexity in automotive production", 19th International Conference on Production Research
- [3] Ing. Dušan Sabadka," Impacts of Shortening Product Life Cycle in The Automotive Industry", Grant project VEGA no. 1/0102/11
- [4] Milos Jovanovic, Bojan Lalic, Antònia Mas, Antoni Lluís Mesquida," The Agile approach in industrial and software engineering project management", Journal of Applied Engineering Science 13(2015) 4,331
- [5] Aljaž Stare,"Agile Project Management in Product Development Projects", Procedia-Social and Behavioral Sciences 119(2014)295–304.
- [6] Achim Kostron, Andreas Brauchle, Dr. Bastian Hanisch,, "Rise of Agile in Automotive R&D?" (white paper)
- [7] <http://www.divbyz.com/industries/automotive> [Last Accessed 23/12/2017]
- [8] Intaher M. Ambe, Johanna A. Badenhorst-Weiss," An Automotive Supply Chain Model for a Demand-Driven Environment", Journal of Transport and Supply Chain Management; Vol 5, No 1 (2011), 1-22. doi: 10.4102/jtscm.v5i1.18
- [9] Ambe, I. M., Badenhorst-Weiss, J. A., "Strategic supply chain framework for the automotive Industry". African Journal of Business Management Vol. 4(10), pp. 2110- 2120, 18August, 2010
- [10] Souresh Bhattacharya, Dr. D. Mukhopadhyay, Dr. Sunil Giri," Supply Chain Management in Indian Automotive Industry Complexities, Challenges and Way Ahead", International Journal of Managing Value and Supply Chains (IJMVSC) Vol.5, No. 2, June 2014, DOI: 10.5121/
- [11] Kispersk-Moron, D.and Swierczek, A.," The Agile Capabilities of Polish Companies in the Supply Chain: An Empirical Study", International Journal of Production Economics (2009) 118(1): 217-224

ACCIDENT DETECTION & ALERT SYSTEM

Prashant Kapri, Suraj Maurya, Shubham Patane, Abhimanyu Singh, Arul Shalom

Department of Computer Engineering, Thakur College of Engineering & Technology, Mumbai, India

Abstract - Nowadays, the deaths caused by road accidents in urban areas are increasing considerably due to various factors. These Death rates can be considerably reduced by providing medical assistance immediately. The major elements that delay the provision of medical help are traffic congestion, lack of ambulance services, no network connectivity and negligence. To resolve these factors, an automated response system is required. Smartphone with their in-built sensors are excellent platforms for building such automated systems. In our project we introduce the Automatic Accident Detection and Alert System (ADAS) that will minimize the time gap and ensure medical assistance immediately. This System comprises of two main components—the server and the ADAS software. Various sensors in the Smartphone will help in identifying the site of the accident. The ADAS system will send the message to the ambulance in case of accident. The ADAS client system will identify the location with the support of in-built sensors in the Smartphone and will alert the nearest medical assistance provider. This will enable the response team to arrive at the accident site and provide medical support to the victim on time. This project will help to decrease the action time and thus reduce the death tolls.

Keywords—Smartphone, Medical Assistance, Sensors, Automation, Software.

I. INTRODUCTION

Every year in India around 1214 road accidents occur and about 377 casualties happen every day [8]. Maximum of the accidents result in deaths as ambulance is not called immediately and as people do not inform the ambulance to avoid police interrogation. The accident might occur at an isolated location where people are not present to report the accident. Recent technologies in vehicles have inbuilt hardware modules to spot and report accidents. Such systems are expensive and non-portable. Not all cars have such systems, only luxury cars have such facility. Hence we introduce Accident Detection and Alert System (ADAS) which will identify the accident with the help of sensors in the Smartphone. Since many Smartphone have the basic required sensors and good computing power, they could be employed to detect accidents and request response. As compared to hardware add-ons, Smartphone are portable - we could carry them in any vehicle we are driving or even travelling in. The way we would use their sensors will make this system inexpensive and lifesaving. The processes to detect accidents could be updated easily and has more scope for forthcoming enhancements. As we are using Smartphone for communication we could use multiple ways of communicating with server, i.e. if the internet connectivity is not available the SMS could be used to converse with the server for help. The principal objective of ADAS is to successfully detect accidents and communicate the same to ensure that the medical assistance can reach the accident location on time. The data from this system

could be used to analyze and study the acceleration waveforms generated during the accidents.

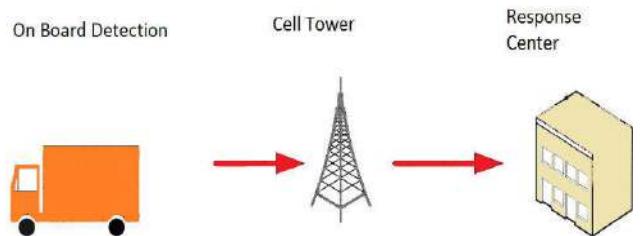


Fig 1: Traditional Accident Detection Systems

II. LITERATURE SURVEY

Earlier researches and works of various authors on automatic accident detection system are discussed as follows:

- Automatic Accident Detection and Ambulance Rescue with Intelligent Traffic Light System [6]. It is a full hardware based system including the likes of microcontroller, modem, drivers, GPS etc. It includes three basic units - Vehicle unit, Ambulance control unit, and Traffic unit. The author has developed a hardware system which in case of an accident notifies ambulance with appropriate data about the accident location. And ITLS system will help the ambulance to reach the hospital at earliest by manipulating the traffic signals. This can be efficiently carried out with the help of Smartphone with their in-built sensor and Google maps. Besides that, the GSM modem used in the system creates a delay while sending messages to the ambulance control unit (since it is a queue based technique) also the maintenance or upgrading process of this ITLS system is quite expensive.

- An Efficient Vehicle Accident detection using Sensor Technology [3].It has been suggested that with the support of sensors like vibration sensor, MEMS (micro electrical mechanical system), GPS and GSM we can develop an efficient accident detection system. The requirements to successfully implement this system are the sensors, which can be easily accomplished with the help of Smartphone and their in-built sensors. There may be delay because of external GSM used – which is a queue based technique. Also the maintenance of the sensor will be costly.

- Utilizing the Emergence of Android Smartphone for Public Welfare by providing Advance Accident Detection and Remedy by 108 Ambulance [4] here they have developed an android application that is used to identify the accident using variation in acceleration parameters. After detecting the accident application spontaneously generates the topographical information by GPS and send pre-recorded voice message to emergency response service. The crucial theory behind the working of this application is that the mobile phone should not be kept with the

driver who is driving the car. It must be attached inside the vehicle. The biggest shortcoming or the loophole in this system is that the phone may tilt or fall inside the vehicle accidentally without having a real time accident thus generating false positives.

- Providing Accident Detection in Vehicular Network through OBD-II Devices and Android based Smartphones [1], here the researcher develops an accident detection and report system that chain Smartphone with vehicle through a second generation On-Board-Diagnostics (OBD-II) that works as an interface to accomplish smart vehicle modeling, providing the user emergency facilities. The researchers have established an android application that deploys an SMS to pre-stored address with related information about the accident location. Also a call is made to the emergency service. The only prerequisite to reach the goal of this system is OBD-II standard. The OBD-II standard is made compulsory from 2001 in U.S and it is also a European, and Japanese variant of this standard, thus this solution is limited to these countries only. Besides that the maintenance as well as upgrading procedure of the system is pretty expensive.

- Accident Detection Depending on the Vehicle Position and Vehicle Theft Tracking, Reporting Systems [5], here the researcher introduces a new system with different algorithm that sense the accidents with the help of accelerometer sensor's tilt direction and other various hardware like GSM modem and GPS. The researches have also developed an android application which will display the accident location in case it happens. The vital components on which the system is solely dependent are the 3-axis accelerometer sensor and GSM modem, which can be replaced with a single device i.e. 'Smartphone' as it comes with the entire mentioned sensor above pre-built in it. In addition to this system uses GSM modem, which can create a delay while sending the emergency message to the user as it is a queue based technique. Beside that the maintenance of the hardware system is quite expensive.

- Car Accident Notification System based on Internet of Things [10]; here the researchers introduce an emergency call notification system using Internet of Things and Cloud computing. The researches have implemented the proposed system using XBee Wi-Fi module, XBee Shield, GPS module, Seeeduino and crash sensors. The basic idea is to detect the accident with the help of crash sensor and trace the exact co-ordinates of the accident spot via cloud using XBee Wi-Fi to the nearest hospital. The main aim was to propose a system allowing global interconnect with the Internet of Thing and Cloud. Despite the limitation the system is a step forward in the field of Internet of Things and with the help of Cloud the information can be transmitted to a long distance. Furthermore the system can be improvised by programming the system to immediately notify the family members of the victim.

- Assistance through Communication Technologies and Vehicle [11], proposed a prototype architecture called as e-NOTIFY which will help in increasing the chance of survival for passengers involved in car accidents. The proposed system offers automated detection, reports and assistance to the victims exploiting the capabilities of vehicular communication technologies. The goal of the system is to provide an architecture that allows 1. Automatic sending of data files containing information about the incident to the control unit. 2. Assessment of the damage done to the vehicle and its occupants, based on the data received from the incident. According to the reported

information and the preliminary accident estimation, the system will alert the required rescue organization to optimize accident assistance.

III. PROPOSED SYSTEM ARCHITECTURE

In the proposed methodology the approach to the solution is with the help of in-built sensors in the Smartphone and physical context information to detect accidents. The ADAS system comprises of two key components—the ADAS server and the ADAS software that are described below:

The ADAS software will be installed in Smartphone and it will perform the task of detecting the accident as well as sending information to the server part. The ADAS software will behave as a sensing device and it will also act as an interface for the third-party observers to contribute information to the accident report. The ADAS software will provide mapping functionality with the assistance of Google maps on the device. This map will allow other motorist to plan their route intelligently around an accident, hereby reducing the congestion. The ADAS client software can access to the data from phone database (such as a contact list) to designate emergency contacts. The detection will completely depend on the in-built sensors in the smart phone such as accelerometer sensor, GPS receiver, Microphone. Fig 2 shows the ADAS System Architecture

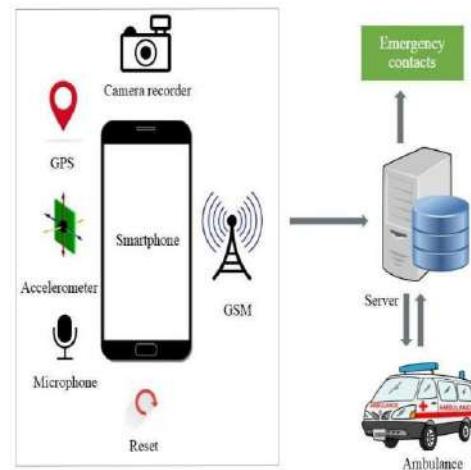


Fig 2: ADAS Architecture

Let us comprehend the functions of different interoperated components of the software.

1. SMARTPHONE ACCELEROMETER SENSOR:

The software will continuously monitor or extract accelerometer sensor information to identify the accident with the help of G-force and deviations in speed of vehicle. Vehicle speed is used to increase the probability of detecting accidents.

2. SMARTPHONE GPS TRANSMITTER:

The software will extract the GPS data and will help in determining the exact position of the vehicle on the globe.

3. SMARTPHONE MICROPHONE:

The microphone will detect high-decibel acoustics event like sound created when an airbag is opened or when vehicles collide. Microphone will increase the possibility of detecting accidents.

4. SMARTPHONE CAMERA:

The camera will also help in increasing the probability of detecting accident. The Smartphone camera of the client as well as of the observers can be used to record and send videos or photo's to the server part so that an emergency dispatch can be deployed to the accident location.

The ADAS server will statically analyze the data received from the software and will notify or contact the client family, friends or emergency responder like ambulance about the accident. The ADAS server will provide an emergency dispatch of the nearby ambulance available to the accident location. The ADAS server will also perform the functions like map and data hosting as well as the task like multimedia hosting (such as videos and photos). Fig 3 describes the flow of ADAS System. It will allow client to share accident parameters (such as acceleration, route and speed of vehicle) for analysis.

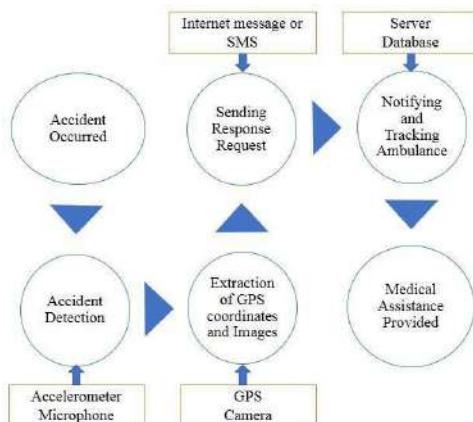


Fig.3: ADAS flowchart

IV. CONCLUSION

Accident detection operation is not an easy task to handle; it can be an extremely complicated process when it comes to real time applications, which is the main reason why it is not implemented yet on a large scale. The proposed system will help to improve the present scenarios.

- Although in-vehicle accident detection system provides emergency responders with essential information as fast as possible but unavailability of this system are restricted by their non-portability and costs, whereas Smartphone provides a promising platform with same sensors at cheap price and portability benefits. Smartphone can surpass the functionality of conventional in vehicle accident detection system.

- Various Smartphone based accident detection systems are exposed to false positive readings. In the proposed system various sensors and features are introduced to increase the accuracy of the system. It is also have additional key feature of resetting the alarm which is not present in the referenced system.

Our system will impressively decrease the redundancy found in other accident detection systems.

- ADAS Smartphone interface will have activation/deactivation button which will allow the user to start or stop medical response in a time window of 30 seconds, after which the medical assistance will be called by default. The user will also have option to call back the medical service even when the alarm is not initiated.

- ADAS software will allow the uninjured client and the observers to take images, videos of the accident and send them to emergency responders to report the accident.

- ADAS software will quickly notify the friends or family of the client. The notification will be sent to pre-registered emergency contacts and it will provide the exact position of the accident site

V. EXPECTED OUTPUT

In comparison with the work done in automatic detection and response the proposed system will be able to overcome various shortfalls and enhance the automated accident detection and response. The table ... depicts the results that is expected on the successful implementation of this project.

Ref.	Limitation	Solution
[1]	<ul style="list-style-type: none"> Works only with OBD II Only available in USA and European countries 	<ul style="list-style-type: none"> ADAS could be easily available and installed in any Smartphone.
[3]	<ul style="list-style-type: none"> Bulky hardware components GSM module have slow speed 	<ul style="list-style-type: none"> ADAS use Smartphones inbuilt sensor
[4]	<ul style="list-style-type: none"> Phone must be docked inside the vehicle Unnecessary use of network 	<ul style="list-style-type: none"> ADAS could be used on body as well as kept on the dashboard
[5]	<ul style="list-style-type: none"> Non portable Tilting of vehicle is considered hence can give false positive when on body 	<ul style="list-style-type: none"> Can travel with the user. Combination of accelerometer and microphone used
[6]	<ul style="list-style-type: none"> Unnecessary traffic lights are controlled Bulky Hardware 	<ul style="list-style-type: none"> One click installation Highly Portable and Available

Fig 4: Expected Outcomes

VI. SCOPE FOR IMPROVEMENT

The entire project is dependent on the application that is installed on the smartphone. This project can be integrated to the virtual systems of the vehicle that can

enhance the performance of this system. This project can also be carried forward and used to predict and issue a warning to the driver in case of any irregularity or issues noted in the driving or the vehicle itself. Thus, preventing accidents from happening. Though there are many ways to improve the notification system, it is completely dependent on the hardware that is installed on the vehicle and therefore we conclude that the limitations of the implementation of the project are;

- The damage of the hardware that is installed on the vehicle or the hardware carried by the user.
- The connectivity of the hardware with the control tower has to be maintained.

REFERENCES

- [1] Jorge Z, Carlos T. , Juan C. and Pietro M. , “Providing Accident Detection in Vehicular Network through OBD-II Devices and Android based Smartphones”, Proceedings of the IEEE 36th conference on Local Computer Networks, Washington, DC, USA, PP.813-819,October 2011.
- [2] A Review Paper on “Accident Detection System Using Intelligent Algorithm for VANET” by Saad Masood Butt. Faculty of Computer Science & Software Engineering, University of Malaya, MalaysiaJournal of Software Engineering and Applications November 2016.
- [3] An Efficient Vehicle Accident detection using Sensor Technology”by P.Kaliuga Lakshmi1, C.Thangamani2, Research scholar1, Assistant Professor2, P.K.R Arts College for Women, Gobichettipalayam. IJARCET Volume 5 Issue 3, March 2016
- [4] Patel K.H., “Utilizing the Emergence of Android Smartphones for Public Welfare by providing Advance Accident Detection and Remedy by 108 Ambulance”, International Journal of Engineering Research and Technology, Vol 2, Issue 9, PP 1340-1342, September 2011.
- [5] ‘Accident Detection Depending on the Vehicle Position and Vehicle Theft Tracking, Reporting Systems’Modugula Ravikanth Reddy †, J.Tulasi ♀ Electronics and Communication Engineering Department, St.Mary’s Group of institutions ,Guntur, India IJSETR Volume 3, Issue 9,
- [6] September 2014
- [7] “Automatic Accident Detection and Ambulance Rescue With Intelligent Traffic Light System” by Mr.S.Iyyappan 1, Mr.V.Nandagopal2P.G Scholar, Dept. of EEE, Ganadipathy Tulis’s Jain Engineering College, Vellore, India1 Assistant Professor, Dept. of EEE, Ganadipathy Tulis’s Jain Engineering College, Vellore, India2IJAREEIE Vol. 2, Issue 4, April 2013
- [8] “Automatic Vehicle Accident Detection and Messaging System Using GSM and GPS Modem”C.Prabha 1, R.Sunitha 2, R.Anitha3DOI: 10.15662/ijareeie.2014.0307062
- [9] NDTV Website: sites.ndtv.com/roadsafety/important- feature-to-you-in-your-car-5
- [10] Priyal Raut and Vandhana Sachdev (2014) “Car Accident Notification System based on Internet of Things ”. International journal of Computer Applications, Volume 107.
- [11] Manuel Fogue, Piedad Garrido, Francisco J. Martinez, Juan-Carlos Cano, Carlos T. Calafate, and Pietro Manzoni (2012) “Assistance through Communication Technologies and Vehicle”, IEEE VEHICULAR TECHNOLOGY MAGAZINE.

News Validation System

Sufiyan Pawaskar, Nooralam Shaikh, Pradyush Rane and Prof. Tabrez Khan

Department of Computer Engineering

*Anjuman-I-Islam's Kalsekar Technical Campus
University of Mumbai*

Abstract—*News is an information of particular interest that was not known in the past, the information implied by news should be accurate and reliable. Currently there exists no resource that could validate whether any news article is giving valid information or not. This paper proposes a comparative study of various news analysis methods and system that would validate news. The comparative study gives an insight of methods that would be more feasible, accurate and reliable out of all methods present in news validation. Based on comparative study, our system will go with hybrid approach. Hybrid approach is combination of both linguistic and network approach, this approach is developed by integrating advantages of linguistic and network approach and by eliminating their drawbacks. It has components like authentic source identification, date validation, data collection, comparison and support analysis. All these parameters will conclude in well-formed results. The system will feed upon normal text of information in the context of news articles, the output of the system will be a probability percentage, that would tell at which extent the information is reliable and valid. Our work in news validation system can help cut down potential misleading information in the news and we can make sure that everyone gets reliable, dependable, authentic and accurate information.*

Keywords—news validation, fake news detection, hybrid approach, network approach, linguistic approach, sentimental analysis, semantic analysis.

I. INTRODUCTION

News is nothing but information about current events. The arrival of the web and the social web brings with it a tremendous number of new sources. The accessibility of these news sources generates a large amount of information which can often times be contradicting and confusing. Facebook, for example, can be seen as a social platform that allows individuals and groups of individuals to freely exchange thoughts and opinions. When this information travels the social web, it is difficult to distinguish between valid and unsupported news. News verification aims to implement a technology that can identify fake news. Fake news detection is defined as a news which is intentionally altered. The problem of fake news detection is more challenging and complicated task than detecting deceptive news, since the political language on TV interviews, post on Facebook and Twitter are mostly short statements.

There are various approaches that can be used to develop news validation or fake news detection system. Majorly there are two types of approaches, linguistic approach and network approach. In Linguistic approach, some liar uses their language skill to avoid being caught guilty. There is some leakage of words from which we can identify that whether they are saying truth or not. The goal in the linguistic approach is to look for such words or leakages. Network approach is innovative and varied, using network properties and behavior are ways to complement content-based approaches that rely on deceptive language and leakage cues to predict deception. Hybrid approach is the combination of both network approach as well as linguistic approach. In our system we are going to use hybrid approach because individually network or linguistic approach is not too accurate

to increase efficiency and for better results we shall be using hybrid approach.

II. LITERATURE REVIEW

A. What is News Validation System?

The objective of News Validation system is to detect the chances of a particular news article (news report, editorial, expose, etc.) being intentionally altered by someone [5]. It is also observed that nowadays the common digital methods for gathering news is social media. In social media, articles are often decontextualized from the source fact and can mix freely with fiction which leads to rumors or fake news [2].

A fake news detection system is a system that can suggest you that whether this news or data is reliable or not. This modern world has started declining use of printed paper [2]. In a survey of America, it was found that people under age of 30 years don't read the newspaper and also 70% of American people above age of 30 years also avoid reading the newspaper and they basically depend on social network for the news [2]. A research on big data shows that, big data will have revolutionary effect on knowledge and understanding of world, there is also chances of false information. [5]

B. Who Produces Fake News?

Fake news articles originate on several types of websites. For example, some sites are established entirely to print intentionally fabricated and misleading articles. The names of these sites are often chosen to resemble those of legitimate news organizations. For example, In India Dainik bhaskar is trust worthy and Dainik bharat is not trust worthy comparatively [4].

C. What Approach can be taken in fake news detection?

For fake news detection we can use various approaches like Linguistic approach, in which the news manipulator uses their language strategically to avoid being found guilty or get caught. In spite of what they are saying language leakage occurs with certain verbal aspects such as patterns of pronoun, conjunction and negative emotion word usage [1].

Network approach is innovative and varied, using network properties and behavior are ways to complement content-based approaches that rely on deceptive language and leakage cues to predict deception.

III. COMPARATIVE STUDY

The arrival of the web and the social web brings with it a tremendous amount of news sources. The accessibility of these news sources generates a large wave of information which can often be contradicting and confusing. Facebook, for example, can be seen as a social platform that allows individuals and groups of individuals to freely exchange thoughts and opinions. When this information travels the social web, it is difficult to distinguish between valid and unsupported news. fake news has played a big role.

Fake news can be used for various reasons: gaining political, financial, religious influences are among those reasons. The social web is one of the fastest and easiest platform for fake news to be spread easily and reach millions of people.

D. Linguistic Approach

The Linguistic Approach can be described as a method where the content of an item gets extracted and analyzed regarding language patterns. [1] producers of fake information tend to use language in a strategic manner. Researchers describe that sometimes a "leakage" occurs, meaning that a break in the pattern is observable. Following methods can be used to identify those leakages Data representation, Deep syntax, Rhetorical structure and discourse analysis.

1) Data Representation:

To represent text "Bag of words" approach can be used. This method generates unigram and bigram model of by analyzing repetition of each word. This can later be used for sentimental analysis. In this bag of words model, individual words or sentences are assigned subjectivity score. That subjective score can be used for sentiment analysis of a paragraph. Based on subjective score each paragraph can be assigned a positive or a negative score. The drawback of this approach is it only perform lexical analysis on a text it does not consider the semantics of text hence accuracy is not obtained.

2) Deep Syntax:

Here the sentence structure is analyzed by using probability context free grammars (PCFG). A PCFG translates text into rewrite rules converting into a parse tree. This method provides accuracy of 85-91% according to Conroy, Rubin and Chen [5], but this method alone is not sufficient to detect deception in a paragraph.

3) Stop Words:

As stated by Christopher D. Manning, Prabhakar Raghavan and Hinrich Schuetze [34] Some words do not add value to the context of a text, thus those words do not need to be processed by the system. Those words are described as common terms or stop words [3]. Nonetheless, it is also stated that in some circumstances the use of stop words are essential in order to keep the context [3]. An example used here is 'Prime minister of India' which contains one stop word. When removing those stop words, the leftover words do not convey the original meaning, possibly leading to a false interpretation by the system [3]. However, due to the usage of the bag of words approach, a loss in meaning does not lead to an effect the outcome negatively.

4) Stemming and Lemmatization:

Articles and other forms of text use different forms of a word due to grammatical reasons [34]. The process of Stemming connotes that the affixes are removed from a word, meaning the ends of words are cut off potentially resulting in the base form of a word. [3] Lemmatization replace informal leymen words to formal words in order to reach the goal of finding the base of a word.

5) Vectorization:

In order to convert textual data into representable information for the classifier, vectorization is needed.

- Count Vectorizer: Converts textual data into token counts, meaning it creates a matrix that describes the frequency of occurrences of words. Such a matrix is called a document-term matrix. However, a Count Vectorizer uses a vocabulary dictionary, which strains the memory of the computer.

- Tf-idf Transformer: Can be described as a transformer, that transforms a count matrix into TF (term frequency) or TFIDF (term-frequency times inverse document-frequency) representation. TFIDF goal is to provide each token with a weight for its occurrence frequency, meaning that if a word occurs often in a lot of different documents the word gets a low value, indicating that it is not of value.

- TFIDF Vectorizer: TFIDF vectorizer is the combination of the Count Vectorizer and the Tfidf transformer.

E. Network Approach

As stated by Conroy, Rubin, and Chen [1] the Network Approach is complementary to the content based Linguistic Approach. Especially due to the rise of news websites, which allow real-time updates, the Network Approach has become more important. It uses information such as metadata or structured knowledge about news in order to uncover fake news.

1. Linked Data: According to N. J. Conroy, V. L. Rubin, and Y. Chen [1] collecting knowledge from networks are possibility to achieve scalable computational fact-checking information about any news. Collecting information from various news websites allows a system to detect truthfulness of particular news. Use of scrapped data as a corpus to algorithm for training purpose trains plays a vital role in validation of news.
2. Social network behavior: There are millions of news posted, commented, shared on social media daily. Analysis on reaction of users on various posts (news) can majorly contribute to validate any news. Analysis on comments, shares, etc. on various post can contribute majorly to validate any news, but monitoring various posts is very complex task and it is a drawback of this techniques.

F. Hybrid Approach

As of now two approaches are discussed linguistic and network approach both have advantages and disadvantages. Hybrid approach is combination of both linguistic and network approach, this approach is developed by integrating advantages of linguistic and network approach and by eliminating their drawbacks. There are several different techniques implemented to obtain accuracy

1) Authentic Source Identification:

Data from authentic and authorized source plays very important role in news validation. Authenticity of source and author of training data set can increase the accuracy of results. This technique will be implemented in Source and author check sub-module in validator module of news validation system.

2) Date Validation:

Sometimes it happens like old news get recreated and shared intentionally or some old news is shared without checking its date, so to tackle this problem date validation is very important. there are chances that some old news is updated by positive results but not mentioned in news, hence it is very important to check date. This technique will be implemented in date check sub-module in validator module of news validation system.

3) Data collection and comparison:

Collection of articles from authentic and genuine source and analyzing that plays a very important role in news validation. Data is collected based on keyword extracted from heading and paragraph of target news article and comparison is done between scrapped articles and target news articles. Matching rate of both articles is inversely proportional to chances of news to be fake,

more match rate can result to less chance of news article to be fake. This technique will be implemented in content check and pattern match sub-module in validator module of news validation system. Grammatical leakage is also analyzed in this module.

4) Support Analysis:

Reaction of users on a news also plays an important role in defining its validity. Reaction will be analyzed using sentimental analysis on comments of users on related news articles fetched from genuine news websites

IV. SYSTEM ARCHITECTURE

The system architecture is mainly divided in two sub modules, the first sub module consists of client side and the main module is news validation server.

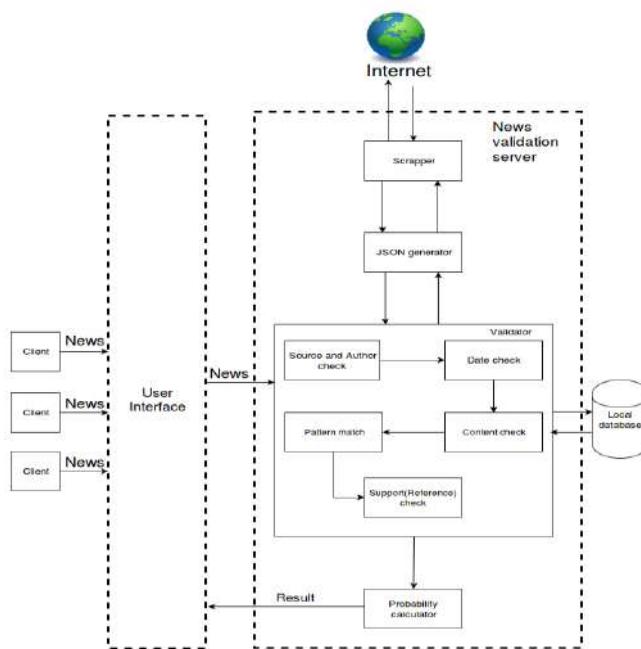


Fig. 1. System Architecture of proposed Hybrid News Validation System

A. 1.Client

The client will be connected to the main module via the user interface. The client will give input to the system, that will go via the user interface to the news validation server. The server will perform operations and will return the output to client via the user interface.

B. User Interface

The user interface will act as a bridge to communicate between client and the main server. The User Interface will take input from the client and provide that input to the Validator module, also it will take input as result from probability calculator and provide that to client.

C. Validator

The user interface will act as a bridge to communicate between client and the main server. The User Interface will take input from the client and provide that input to the Validator module, also it will take input as result from probability calculator and provide that to client.

- 1) Source and Author Check: Data from authentic and authorized source plays very important role in news validation. Source and author check module authenticate and authorize the source to

fetch data from internet. Authentication and Authorization is done on the basis of data stored in database which is updated repeatedly. Authenticity of source and author can increase the accuracy of results.

- 2) Date Check: Sometimes it happens like old news get recreated and shared intentionally or shared without
- 3) Checking its date, so to tackle this problem date validation is very important. there are chances that some old news is updated by positive results but not mentioned in news, hence it is very important to check date. This technique will be implemented in this sub module.
- 4) Content Check and Pattern Match: Content check sub-module check for linguistic leakages in the content. These leakages are detected by implementing natural language processing algorithms. Pattern match sub-module compare both articles fetched from internet and our source articles and check for common keywords in both of them, High match rate represents high chances of news to be true.
- 5) Support (Reference) Check: There are various websites that provides users to comment opinions on any news. Those opinions can be analyzed by implementing sentimental analysis on that comments and result can help in validation of news.
- 6) Scrapper: Scrapper module will scrap related news articles over the internet, Beautiful soup package is used to develop this module.
- 7) JSON Generator: JSON Generator module will convert unstructured data into structured data i.e. text format into JSON format, since structured data is very easy to analyze.
- 8) Probability Calculator: Probability calculator module will calculate the probability of news to be fake by collecting result from all sub-modules of validator module and will return a single probability value to UI module.

V. CONCLUSION

Information in internet should be correct and reliable. But currently our society lacks a system that can check whether if some information implied in news is true or not. Developing such system will make a difference in society and will help to maintain peace.

REFERENCES

- [1] N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," Proceedings of the Association for Information Science and Technology vol. 52, no. 1, pp. 1–4, 2015.
- [2] Josef Moucachiein "Developing a News Aggregation and Validation System" 2017.
- [3] C. D. Manning, P. Raghavan, and H. Schuetze, "Introduction to Information Retrieval". Cambridge University Press, 2008.
- [4] Hunt Alcott, "Social Media and Fake News in the 2016 Election". Stanford University, 2017.
- [5] Victoria L. Rubin, Yimin Chen and Niall J. Conroy, "Deception Detection for News: Three Types of Fakes". University of Western Ontario, 2015.

Exam Cell Automation System and Timetable Generator

Prof.Dnyaneshwar Dhangar^[1], Aishwarya Sontakke^[2], Jhanvi Shah^[3], Namrata Bodas^[4], Yash Shah^[5]
dnyaneshwar.dhangar@mctrgit.ac.in, aishwarya.sontakke@gmail.com, jhanvishah307@gmail.com,
namratabodas96@gmail.com, yashbshah96@yahoo.co.in

Abstract- Exam Cell Automation System is developed for the college or institute to computerize the traditional way of conducting exams and to simplify examination hall allotment and seating arrangement. Also, automating the timetable generating process will help get rid of various anomalies caused due to human error. Mostly staff faces many problems in assigning exam halls depending upon the availability of classes, capacity of classes, student details, allotment of duties to the respective teachers, etc. And the manual system of preparing time table in colleges is very time consuming and usually ends up with various classes clashing either at same room or with same teachers having more than one class at a time. To overcome all these problems, an automated system is proposed.

Keywords- Exam Cell Automation System, Timetable Generation System, Constraints, Genetic algorithm, optimal solution, etc.

I. INTRODUCTION

Exam cell allotment is a process of complete allotment of students and faculty to their respective classrooms for given examination. It involves various sub-processes like managing the time allotment for exams of different courses from different academic years on the same day (in case of internal exams) or over a given period and scheduling the faculty according to their workload, duties, etc. A college timetable is a temporal arrangement of a set of lectures and classrooms in which all given constraints are satisfied. Creating both of these systems manually is complex and time-consuming process. The proposed software is used to overcome the entire problem which they are faced, and making complete automation of manual system to computerized system.

A. PURPOSE

A completely automated output will be generated by the final system which will save a lot of time and effort of an institute administration. The system should be able to handle all the user defined constraints and it should achieve ease of use for user of system so that he/she can make automatic time table. The system should focus on the optimized use of resources. We also aim to make the system generic so that it can work equally well for different institutes and colleges.

B. SCOPE

The scope of the project is the system on which the software is installed, i.e. the project is developed as a third-party general-purpose software, and it will work for a particular institute which can efficiently generate optimal solutions.

II. SYSTEM ANALYSIS

A. EXISTING SYSTEM

Existing system is based on manual paperwork and manual calculations. The amount of work makes this process very slow and tedious, making it inefficient. The system requires maintaining a lot of documents. Handling of such important documents is again a tedious work. Also, retrieval of important facts and statistics will consume time. Since, there is a lot of manual work involved in current system, mistake in one detail can lead to wrong generation of page. This system faces a huge problem as there is no proper collection of requirements. This system is to enhance manual work and also more energy is wasted to allocate the seating arrangement.

B. DISADVANTAGES OF EXISTING SYSTEM

- The retrieval of data is very slow and data is not maintained efficiently.
- More calculations are required to generate the report so it is generated at the end of the session.
- All calculations for generating reports are done manually which may lead to errors.
- Existing system requires lots of paper work. Loss of even single register/record can lead to difficult situation because all the papers are needed to generate the reports.

Report generation is very time consuming as the work done is manual and reports cannot be generated in the middle of the session.

II. PROPOSED SYSTEM

Enterprise-oriented software is to be designed to automate both exam hall allocation and generate timetable. The system will take various inputs like information of students, number of available classrooms, faculty, courses, etc. Depending upon the inputs it will generate a possible output, making optimal utilization of all resources that will best suit any of constraints or college rules. The project showcases an automated system which ensures the reduction of the tediousness, more effective work and systematic management. This system will help the college in saving the extra time spent in manual work, avoiding mistakes due to human error, will increase efficiency and save time, and will allow neat handling of data rather than error prone records.

The proposed system includes

- Interface for input
- Database Capabilities
- Processing Capabilities
- Search Panel

A. CHARACTERISTICS OF PROPOSED SYSTEM

- User friendly: The retrieval and storage of data is fast and data is maintained efficiently because the proposed system is user friendly. The GUI is provided in the proposed system that enables user to deal with the system very easily.
- Easy report generation: Reports can be easily generated in the proposed system. So that the user can generate the report as per the requirement or in the middle of the semester.
- Very less paper work: The proposed system requires very little paper work. All the data is entered into the computer promptly and report can be generated through computers. Moreover, work becomes very easy because there is no need to keep data on paper.
- Computer operator control: Computer operated control will be present so that there is no chance of errors. Moreover, storing and retrieving of information is easy so work can be done quickly and in time.

Boosts enterprise accessibility.

III. ARCHITECTURE DIAGRAM

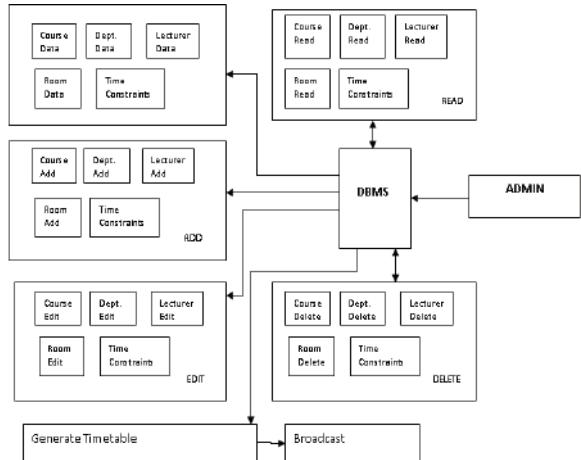


Figure 1: Timetable Generation

The system will take various inputs from the user as well as various rules, facts and constraints, which are stored in knowledge base. It serves as input to our timetable generator algorithm residing on the server machine. After the representation of Knowledge Base is standardized, design of the timetabling algorithm will be done.

The input data contains:

1. Professor: Data describes the name of lecturers along with their identification number.
2. Subject: Data describes the name of courses in the current term.
3. Room: Data describes the room number and their capacity.
4. Time intervals: It indicates starting time of the lecture along with duration.

System Constraints for Timetable Generation are divided into 2 categories:

- 1) Hard Constraints: The timetable is subjected to the following four types of hard constraints, which must be satisfied by a solution to be considered as a valid one. A student should have only one class at a Time.
 - a Teacher should have only one class at a time.
 - b A room should be booked only for one class at a time.
 - c Some classes require classes to have particular equipment. For instance, audio visual equipment, projectors etc.
- 2) Soft Constraints: These are the constraints that are of no great concern but are still taken into contemplation. They don't need to be satisfied but the solutions are generally considered to be good if they are satisfied.
 - a Courses must be eventually distributed.
 - b Students should not have any buffer time between two classes on a single day.
 - c Scheduling of teachers should be evenly divided throughout the week.

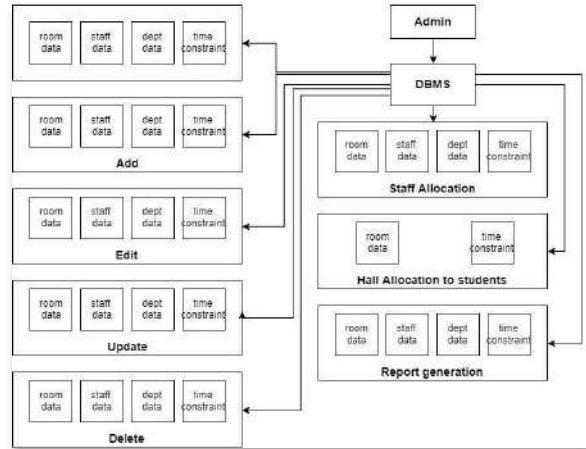


Figure 2: Exam Cell Automation System

The system will provide a way to allocate exam hall for each student without clash. The system will store the list of classes and its capacity where exam can be conducted and list of students appearing the exam on each day. The system provides facility to generate report of the hall allocation.

The input data contains:

1. Professor: Data describes the name of invigilators along with their identification number.
2. Duties: Data describes the name of the invigilators along with number of duties to be assigned to them.
3. Subject: Data describes the name of courses in the current term.
4. Room: Data describes the room number and their capacity.
5. Time intervals: It indicates starting time along with duration of the exam.
6. Student: It indicates the total number of students, their roll numbers and the divisions to which they belong.

System Constraints for exam cell automation system are divided into 2 categories:

- 1) Hard Constraints: The exam cell automation is subjected to the following three types of hard constraints, which must be satisfied by a solution to be considered as a valid one.
 - a) A student should have only one exam at a Time.
 - b) A Teacher should have only one exam invigilation duty at a time.
 - c) A room should be booked only for one exam at a time.
- 2) Soft Constraints: These are the constraints that are not the real concern but are yet taken into consideration. They don't need to be satisfied but the solutions are generally considered to be good if they are satisfied.
 - a) Course teacher should not be assigned exam invigilation duty for the same subject for the same class.
 - b) Students should have some free time between two exams on a day.
 - c) Scheduling of teachers should be well spread over the examination period.

IV. METHODOLOGY

The search space of the given problem is too vast, many solutions exist in the search space and few of them are not feasible. Feasible solutions here mean those which do not violate hard constraints and as well try to satisfy soft constraints.

There are currently many different solution generation algorithms in existence. These are major algorithms that have appeared in the literature on generating solutions. The fundamental solution generation techniques are the following:

1. Integer Programming/Linear Programming
2. Constraint Based Programming
3. Genetic and Evolutionary Algorithms
4. Simulated Annealing
5. TABU Search

Using Genetic Algorithm, a number of trade-off solutions, in terms of multiple objectives of the problem, could be obtained very easily. Furthermore, each of the obtained solutions would be much better than a manually prepared solution which is in use.

A. GENETIC ALGORITHM

Genetic algorithms are general search and optimization algorithms inspired by processes and normally related with natural world. Genetic algorithm mimics the process of natural selection and can be used as a method for solving complex optimization problems which have large spaces. They can be used as techniques for solving complex problems and for searching huge problem spaces. Unlike many heuristic schemes, which have only single optimal solution at a time, genetic algorithms maintain many individual solutions in the form of population. Individuals (parents) are selected from the population and are then mated to form a new individual (child). The child is further mutated to introduce variety into the population. Rather than starting from a single point inside the search space, Genetic Algorithm is initialized to the population of guesses. They are usually random and will be spread throughout the search space.

A typical algorithm uses three operators, selection, crossover and mutation, to direct the population towards convergence at global optimum. A Genetic Algorithm, requires a process of initializing, breeding, mutating, choosing and killing. It can be said that most methods called Genetic Algorithms have at least the following

elements in common: Population of chromosomes, Selection according to fitness, Crossover to produce new offspring, and random mutation of new offspring.

B. GA OPERATORS

1. Chromosome representation: Chromosome is a collection of parameters which represents a proposed solution to the problem that the genetic algorithm is trying to achieve.
2. Initial population: A possible solution to the given problem is given by each chromosome of this population.
3. Selection: Chromosomes are selected from the population for reproduction by this operator. The fitter the chromosome, the more times it is likely to be selected to reproduce.
4. Crossover: Cross over is a process of taking more than one parent solutions and producing a child solution from them. There are methods for selection of the chromosomes. This operator randomly selects a locus and exchanges the subsequences prior to and after that locus between two chromosomes to create two offspring.
5. Mutation: Mutation changes one or more gene values in a chromosome from its beginning state. In mutation, the solution may change completely from the solution obtained before. Hence Genetic Algorithm can come to better solution using mutation.

Fitness Function: The fitness function is defined on the genetic representation and measures the quality of the given solution.

V. SOFTWARE ENVIRONMENT

A. FRONT END

- A software development platform written in Java namely, NetBeans. The NetBeans Platform allows applications to be developed from a collection of modular software components called modules. The NetBeans IDE is primarily meant for development in Java, but also supports other languages, in particular PHP, C/C++ and HTML5. It is a cross-platform and runs on Microsoft Windows, Mac OS X, Linux, Solaris and other platforms supporting a compatible JVM.

B. BACK END

- MySQL is a database generation system (DBMS) for relational databases, a database being a collection of interrelated data, be it text, numbers, or binary files, that are stored and kept organized by the DBMS. MySQL was selected to develop the database for this system because, MySQL offers excellent performance, portability and reliability, with moderate learning curve at little to no cost because MySQL is the world's most popular open source database. Besides that, another reason for it being chosen is PHP has good support for MySQL.

VI. CONCLUSION

Ultimately the result of the implementation of this project will lead to reduced workload of the staff. The result would be a fully-fledged working Automated Exam Cell Automation and Timetable Generation System. There won't be any need to use multiple different systems for different activities. The processes will be covered by the proposed system. Various slot combinations can be acquired so that another timetable is generated as of need. The project reduces time consumption and the pain in framing the timetable manually. This will reduce the tediousness of the manual processes and give a chance for efficient, flexible and automated process.

REFERENCES

- [1] Michael R.G and David S. Johnson. Computers and intractability: a guide to the theory of np-completeness. 1979.
- [2] BP Romero. Examination Scheduling in a Large Engineering School: Computer Assisted Participative Procedure, volume 12 of 17-23. Industrial Organization Department Industrial Engineering Technical School, Madrid, Spain, April 1982.
- [3] Bhaduri a university timetable scheduling using genetic algorithm. Advances in Recent Technologies in Communication and Computing, 2009. ART Com '09. International Conference
- [4] Dipti Shrinivasan automated time table generation using multiple context reasoning for university modules Published in: evolutionary computation, 2002. cec'02. proceedings of the 2002 congress on (volume:2)
- [5] Silberschatz, H. Korth and S. Sudarshan, Database System Concepts, 4th Edition, McGraw-Hill International, 2002.
- [6] Solving Timetable Scheduling Problem by Using Genetic Algorithms Branimir Sigl, Marin Golub, Vedran Mornar Faculty of Electrical Engineering and Computing, University of Zagreb Unska 3, 10000 Zagreb, Croatia
- [7] Leon Bambrick Supervisor Dr B Lovell Lecture Timetabling Using Genetic Algorithms Available : <http://secretgeek.net/content/bambrilg.pdf>
- [8] Sanjay R. Sutar , Rajan S. Bichkar University Timetabling based on Hard Constraints using Genetic Algorithm Available : <http://research.ijcaonline.org/ volume42/number15/ pxc3877964.pdf>
- [9] Eng. Ahmed Hamdi Abu ABSA, Dr. Sana'a Wafa Al-Sayegh, Elearning Timetable Generator Using Genetic Algorithms Available : https://uqu.edu.sa/files2/tiny_mce/plugins/filemanager/files/30/papers/f18_9.pdf
- [10] Xamclick Exam Cell Automation System by Harsha Khutafale, Hardik Mate, Vaishnavi Sabnavis, prof. Nilesh Patil, International Journal of Advanced Research of Computer Science and Software Engineering- 2016
- [11] Automated Examination Support System by Bondre Rutuja Avinash, Durgi Varsha Vijaykumar, Mohite Pradnesh Rajeev, Parkar Vishal V, International Journal of Current Engineering and Technology- 2015

Decision Making in Agriculture Using Data Analytics

Sudha Shankar, Madhuri Rao

Department of Information Technology, Thadomal Shahani Engineering College Bandra, Mumbai University, Fort, Mumbai, India
sshankar27@gmail.com, my_rao@yahoo.com

Abstract - Data Analytics has a key role to play in Decision Making. It is possible to make the right policy decisions in the Agricultural sector by using Data Analytics technologies and Information Communication Technology.

Keywords -Security, Multimedia, Cryptography, DES (Data Encryption Standard), AES (Advanced Encryption Standard).

I. INTRODUCTION

The need to make informed decisions in agriculture is necessary to improve economic returns. The agricultural sector now has a lot of data owing to the advances in mobile and remote measurement systems. This requires high level analytical capabilities. In this paper we will be investigating the ways and means of integrating Data Analytics in Decision Making in Agriculture

II. BIG DATA ANALYTICAL TECHNIQUES

The following technologies are available that can also be used in the Agriculture sector [2], [3], [6].

A. Data Integration and Storage with Unified Integrated User Interface:

In this technology the data can be of different forms such as semantic, statistical, geographical or operational.

B. Dashboard tool –

This tool is used to display several complex data points quickly to the user.

C. Social Network Aggregation and Single Window Display

Data by aggregating social data from several social networking sites such as Facebook, twitter, You Tube and others pertaining to a point of discussion and making them available in a single window for analysis.

D. Hot Topic Sensing & Topic Summarization

A tool used in the e-government domain to extract and summarize content appearing in social sites and internet forums. This content is meant to be of great public importance and hence is frequently discussed.

E. Community Feedback Platform

A tool used to get feedback from society on a issue.

F. Visualization Tools

Data could be social data or statistical data. The Visualization tool helps in gaining insights and gaining the right information from it.

G. Outgoing Multichannel Social Media Single Window Messaging

H. Opinion Maps –

A tool that uses geo-referencing to voice opinions of people located at different geographical locations.

I. Opinion Mining Solutions / Sentiment Analysis –

This tool helps in the classification of emotions expressed within a review as positive or negative and representing it in a form a user understands it.

J. Collaborative Tools –

These tools are used by workers in a team who are geographically dispersed but need to work on a project together. The collaboration tool makes it easier and faster for the workers to exchange ideas, get feedback, send approvals and collaborate in a professional manner.

K. Simulation and Impact Visualization –

This tool is used to model a real phenomenon and express it as a set of mathematical formulae.

L. Visual Fuzzy Cognitive Maps [12] –

Acknowledge-based tool used to model and simulate dynamic systems using a graphical representation with concepts such as states, variables and entities.

M. Crowd sourcing [13] –

This is a process of taking services or funding from a crowd and outsourcing it to a crowd of people.

N. Serious Gaming and Persuasive technologies –

This tool deals with ICT solutions that can induce behavioral change.

O. Machine Learning

These tools use algorithms to parse data, train the data and then use it to draw inferences without any human intervention. [8]

P. Statistical Analysis [10], [11]

This is the process of drawing inferences from quantitative data. There can be different methods used such as summarizing data, averaging methods, measuring the spread of data, regression analysis, sample size determination and hypothesis testing.

III. BIG DATA ANALYTICAL TECHNIQUES

Information and Communication Technology (ICT) has supported the process of decision making for several years.

- Drones
- Wireless Technologies – the use of this technology eliminates the need for installation of coaxial cables^[14]
- Global Positioning Systems (GPS)
- Computer Controlled Devices – Automated
- Radio Frequency Identification (RFID)^[7] – it has been used in identifying and tracking animals. In the food chain it is used for traceability control.
- Sensors^[15] – sensors can be used for various applications in agriculture. A sensor for scanning farm fields from the sky, sensor for connected cows can do away with a farmer actually going to the field for grazing his cows, sensors for monitoring pH level of soils, sensors which are hidden in grains for controlling the conditions of stored grain.

IV. INTEGRATING DATA ANALYTICS IN DECISION MAKING IN AGRICULTURE

A. Policy Makers' Level

Decision Making in Agriculture would be discussed with reference to the traditional Policy Cycle^[1]

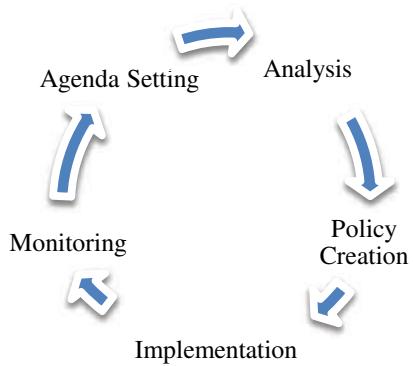


Fig.Fehler! Kein Text mit angegebener Formatvorlage im Dokument..1: Policy Cycle

Each of the stages is described below.

- Agenda setting: defining what the problem to be addressed is for a policy to be formed or changed.
- Analysis: the different challenges and opportunities for the agenda item are analyzed to produce a draft policy document.
- Creating the policy: the actual policy is formed after processes of formal consultation, risk analysis, pilot studies and implementation plan design.
- Implementing the policy: this involves activities such as legislation, regulation, guidance, and creation of a delivery plan.
- Monitoring the policy: in this phase the policy in action is evaluated and reviewed considering research evidence and

views of users. At this stage there is a possibility of going back to stage 1 with possible changes

1. TECHNOLOGY ASSIGNMENT

Each phase is split into subtasks. Ideally the Agricultural department should consider the approach given in the table below.

Table 1: Technology Assignment to Policy Cycle

Phase	Subtasks	Data Analytics Technology
Agenda Setting	Construction of Policy Statement Validation through Experts	a) Policy Indicator Dashboard b Incoming Social Networking Aggregation and Single Window Display c) Visual Social Data Analysis d) Hot Topic Sensing and Topic Summarization e) Visualization of Statistical Data f) Opinion Maps
Analysis	Definition of Values and Goals Gather evidence and knowledge from sources Gather opinions from society and citizens Developing a range of options	a) Visualization of Statistical Data b) Incoming Social Networking Aggregation and Single Window Display c) Opinion Maps d)Outgoing Multichannel Social Media Single Window Messaging e) Visualization Tools f) Visual Fuzzy Cognitive Maps g) Community Feedback Platform h) Simulation and Impact Visualization
Policy Creation	Discussion and Debates with Stakeholders Summarize debates Policy Actions Weighing in the Policy Context Drafting proposals	a) Visual Social Data Analysis b) Community Feedback Platform c) Outgoing Multichannel Social Media Single Window Messaging d) Hot Topic Sensing and Topic Summarization
Implementation	Implementation tasks Policy Enforcement	Various schemes which may use some Data Analytics techniques
Monitoring	Impact Evaluation Administrative and Judicial evaluation Monitoring of Key	a) Simulation and Impact Visualization b) Policy Indicator Dashboard

	indicators Simulation as Forecast	c) Visualization of Statistical data
--	--------------------------------------	---

In the above table the overlapping of some technologies in more than one phase of the Policy Cycle is seen [2], [3].

Some of the above-mentioned technologies have been described in the context of agriculture.

2. POLICY INDICATOR DASHBOARD

This is a powerful tool that can be used by policy makers when discussing a particular issue. It can be used to set perspective on important attributes of interest so as to be able to monitor the policy at management level [2].

Crop	Season	Pesticide	Fertilizer
Rice	Kharif	P1	F1
Wheat	Rabi	P1	NA
Pulses	Rabi	P2	F2

↓
Drill Down and Visualizations

Fig.2: Policy Indicator Dashboard

3. SOCIAL NETWORK AGGREGATION AND SINGLE WINDOW DISPLAY

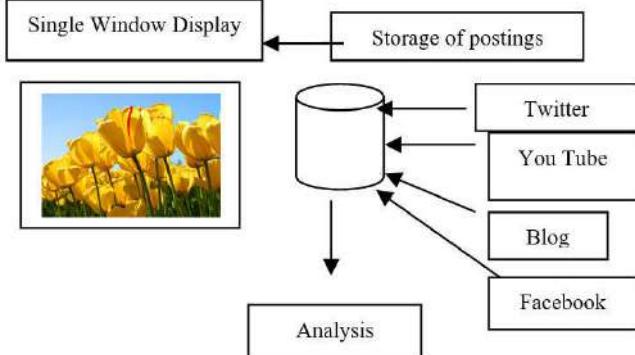


Fig.3 Social Network Aggregation and Single Window Display

Decision Maker receives postings from various social sites that are pulled into a single location. These postings are displayed on a single window [2]. This tool helps decision makers gather public opinion at one window that can be further analyzed.

4. VISUALIZATION TOOLS [9]

- Demographics Visualization – with the help of these tools the stake holders and the policy makers get a clear idea of the data and its trends over time. Thus, there is no need to struggle with large amount of numbers.
- Legal Arguments Visualization
- Discussion Arguments Visualization
- Geo Visualization

5. TECHNOLOGIES THAT CAN BE USED AT FARMER LEVEL

This level deals with farmer related issues, prices and Economic data. It also covers those technologies which would

help the farmer see and understand what is happening or will happen on his farm. Mentioned below are some of the technologies that can be used:

1. Simulation
2. Social Network Aggregation and Single Window Display
3. Outgoing Multichannel Social Media Single Window Messaging

A farmer wishes to post some information of use to social sites. The functionality is just the reverse of Social Network Aggregation and Single Window Display as discussed in Fig.3.

6. 4. OPINION MAPS

Opinion maps are used for geo referenced interaction. So farmers at different locations can voice their requests or issues while being logged into this interactive map that can be easily integrated into any internal or external website [2]. Thus, farmers can interact with the decision makers.



Fig 4:Opinion Maps

5. Machine Learning [8]

a) In Accurate Disease Diagnosis – images of the diseased plant can be taken using technologies such as satellites, land based rovers, pictures from smart phones or using Unmanned aerial vehicles (UAVs). These pictures are then uploaded and the machine learning software diagnoses the disease and indicates a management plan.

b) Forecasting weather conditions – with the availability of past data ranging over several years, these powerful algorithms can provide weather forecasts. This is useful to the farmer in making decisions as to whether he should water his field or wait for the rain.

c) In plant breeding machine learning algorithms are being used to help in creating more efficient seeds. This results in seeds which are genetically modified and more healthy, adaptive and productive.

6. ICT for private farm specific data collection

a) Using drones to collect crop data is extremely useful in the agriculture industry [5]. Drones along with small computer-operated aircrafts, allow farmers to record readings of crop yield, plant height, groundwater basin health, soil nutrient loads and

other parameters that would be difficult to collect periodically across an entire farm. Satellite or a manned aircraft can be used to do the same says Charles Neal, president of Aero Systems West, a company that uses drones to collect data on farms.

b) Using GPS on the farm [8] – a farmer can pinpoint a crop in his field to monitor its condition in terms of its drought resistance, insect attack pressure or lack of nitrogen. Once the exact co-ordinates of the crop are known the correct kind of attention and care can be administered to it.

c) The connected tractor [8] – this modern technology enables tractors to communicate and share important information on the field. For instance, a tractor can communicate to another about where it has already applied fertilizer or water. This feature will help in reducing tractor emissions, conserving energy and avoiding unnecessary overlap.

B. Technologies that can be used at Landscape Level

This level considers weather, soil and other bio physical data. Mentioned below are some of the technologies that can be used:

1. Machine Learning

2. Statistical Analysis

3. ICT for public data like weather, climate, physical data about a location, prices and economic data. There are Farming apps that can track weather, planting or sowing data and soil data. With this the farmer gains information on ways to make his crop healthy and survive.

V. CONCLUSION

The different ways of integrating Big Data Analytics in Decision making in the Agricultural sector have been discussed. These Data Analytics technologies and ICT if adopted would be

useful in bringing about policy improvements and supporting productivity gains in a positive way.

REFERENCES

- [1] Jones, C.O.: An Introduction to the Study of Public Policy. Brooks/Cole Publishing Co. (3rd Edition 1984) (1970).
- [3] <http://cordis.europa.eu/docs/projects/cnect/9/287119/080/deliverables/001->
- [4] D219PolicyDesignProcessandFUPOLcomponentIntegrationfinal.pdf
- [5] 0204F01_International Research Roadmap on ICT tools for Governance and Policy Modeling
- [6] <http://www.foodandagpolicy.org/content/using-big-data-inform-agricultural-decisions>
- [7] https://motherboard.vice.com/en_us/article/9amb45/the-most-useful-tool-on-the-modern-farm-is-data
- [8] <https://egov20.wordpress.com/2013/05/10/how-can-technology-improve-public-policy-come-and-find-out-in-dublin/>
- [9] https://en.wikipedia.org/wiki/Information_and_communications_technology_in_agriculture
- [10] <https://modernag.org/innovation/machine-learning-changing-modern-agriculture/>
- [11] http://crossover-project.eu/Portals/0/0205F01_International%20Research%20Roadmap.pdf
- [12] <https://www.skillsyouneed.com/num/simple-statistical-analysis.html>
- [13] <https://www.bigskyassociates.com/blog/bid/356764/5-Most-Important-Methods-For-Statistical-Data-Analysis>
- [14] <http://www.site.uottawa.ca/~rfalc032/isfuros2017/tutorials/gonzalo.pdf>
- [15] <https://dailycrowdsourcing.com/training/crowdsourcing/what-is-crowdsourcing>
- [16] https://en.wikipedia.org/wiki/Information_and_communications_technology_in_agriculture
- [17] <http://www.innovation-hub.com/society/types-of-smart-sensors-applied-to-agriculture/#1>

SHARED NEAREST NEIGHBOUR APPROACH FOR CLUSTERING DOCUMENTS

KalpanaGangwar

Assistant Professor, TCET,Mumbai, India

Kalpana.iiitm@gmail.com

Abstract— Clustering techniques are used to group objects (documents) such that objects (documents) in the same cluster represent same theme or topic. This grouping of objects (documents) into cluster is based on information contained in it. Each document is taken as an object having its well-defined characteristics and relationships between documents are found by measuring these well-defined characteristics. Objective is to increase the similarity within a cluster as it results in greater difference between clusters; leading to better clustering. Mainly two types of clustering are used based on the problem statement Hierarchical or K-means. Both these algorithms assume that an object(document) belongs to a cluster if it is closer to at least one object(document) in that cluster; then the other objects in other clusters. Hierarchical clustering is sensitive to outliers and not suitable for large datasets; while in k-means initial seed has a strong impact on final results. To overcome this, we are implementing “Clustering using shared nearest neighbour”. Documents are first represented in vector-space model; similarity matrix is computed from that and then a shared nearest neighbour graph is constructed; from which the clusters are formed.

Keywords: Clustering, hierarchical clustering, k-means, outliers, vector-space model

I. INTRODUCTION

Clustering is an unsupervised learning technique in the area of Data Mining or Machine Learning. In data mining or even in data science world, the objective of an unsupervised learning task is trying to find hidden structure in unlabeled data. Since the examples given to the learner are unlabeled, there is no error or reward signal to evaluate a potential solution [1]. In other words, we can say that it's an arrangement of data items which are “similar” between them closely and “dissimilar” to data items in other clusters. Once we have understood what clustering is, the next step is to determine what are the steps required for the process.

The first important step is to determine similarity or dissimilarity measure. Let us assume that the two objects are denoted by x and y and we are using similarity as the measure, then

Similarity (x, y): is large

Dissimilarity (x, y): is small

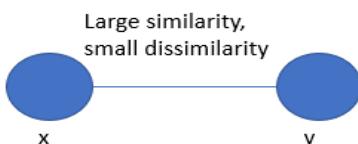


Fig 1: Distance as a measure of similarity

The next step is to identify a criterion function that will evaluate clusters we obtained in the previous step using a similarity or dissimilarity measure. An Internal Criterion Function produces a clustering solution which optimizes “criterion function” defined over the documents that are part of a cluster. E.g. a “criterion function” maximizes the sum of average pairwise similarities between documents assigned to each cluster, weighted according to the size of each cluster. If we use cosine function to measure the similarity between documents, then we want the clustering solution to optimize the following criterion function (T_1): [2]

$$T_1 = \sum_{r=1}^k \frac{\|D_r\|^2}{n_r}$$

Where $\|D_r\|^2$ = cosine similarity between the two documents

N_r = number of documents in the cluster

Shannon's entropy is another criterion function which can be used for evaluation of results. Entropy is generally used to measure the disorder between objects of a cluster. For each cluster, the category distribution of data is calculated first i.e. p_{ij} is the probability that a member of cluster j belongs to category i.

Then the entropy of each cluster j is calculated as

$$Ej = - \sum_i p_{ij} \log(p_{ij})$$

The total entropy is calculated as the sum of the entropies of each cluster weighted by the size of each cluster:

$$Een = \sum_{j=1}^m n_j * Ej/n$$

Where m is the total number of clusters, n_j is the size of j^{th} cluster and n is the total number of documents.

We have two main algorithms used for the purpose of clustering i.e. K-means and Hierarchical. Both of them work measure whether an object belongs in the cluster or not by identifying similarity with some object in the same cluster. Due to this factor k-means has a lot of impact on final clusters when we start with a particular set of seeds or initial points. Similarly, hierarchical clustering may not lead to clear clusters being formed at the end of the process.

Clustering is one of the widely used unsupervised learning tool that has many applications in computer science field. It can be used for clustering documents according to the theme represented by document content while a simple search result will only see for the keywords in a set of documents to term them as similar. Clustering is also used for organizing a large set of documents

according to a common theme. In many applications there is a need to find duplicates or near-duplicates in a large number of documents. Clustering is employed for plagiarism detection, grouping of related news stories and to reorder search results rankings (to assure higher diversity among the topmost documents) [3].

II. RELATED WORK

Our focus here is how clustering is used for topic or theme related document clustering. In [4] the large dimensionality of a document is overcome with the identification of frequent itemsets present in a document before the documents are given as an input in the clustering process. A study on Topic Identification using k-means clustering algorithm for big vs. small documents is researched in [5]. The results in [5] highlighted the need for more robust methods for documents with more number of terms. They emphasized on more robust methods like semantic based methods for big document collections. Application of document clustering for identification of keywords from the clusters and then using these keywords for generating text summaries is discussed in [6]. Document clustering applied to stream of documents to incorporate temporal properties of incoming documents in a repository is discussed in [7].

Quality and accuracy of identifying document themes is still a challenging task because of the large number of words contained in a document. In this shared nearest neighbor approach, we will be trying to derive better quality of results and finding out what percentage of documents have been accurately labelled.

III. SHARED NEAREST NEIGHBOR APPROACH

We begin by calculating the document similarity matrix, i.e., the matrix which gives the cosine similarity for each pair of documents. Once this similarity matrix is calculated, we find the first n nearest neighbors for each document. (Every document is considered to be its own 0th neighbor). In the nearest neighbor graph, there is a link from document i to document j, if i and j both have each other in their nearest neighbor list.

In the shared nearest neighbor graph, there is a link from i to j if there is a link from i to j in the near neighbor graph.

The strength of this link is equal to the number of shared near neighbors of i and j. Now we will apply the threshold value and take all connected components from shared nearest neighbor graph as final clusters.

While doing this we need to take care of two issues:

- i) if the threshold value is set too high, a natural cluster may be split into smaller clusters.
- ii) if the threshold value is set too low, clusters may be formed even though patterns in the data are not very significant.

We have identified two parameters for our approach:

- a) One related to strength of the links
- b) One related to the number of links

Strength of the link can be above a threshold for being labelled as a strong link. It can be strong enough to be merged i.e. above merging threshold.

Number of links can again be having two thresholds. One is called the noise threshold, below which will not be considered for our final results. There will be a topic threshold according to which the topic will be decided or clusters will be formed.

IV. METHODOLOGY

For this process we are using a principle that assumes that documents in the same cluster take words from small number of specialized vocabularies and few words may be taken from a general vocabulary.

We are having a pre-processing step which removes all the irrelevant content from the documents and then apply the clustering process.

Step1: Pre-processing of Documents (for each document)

- a) Convert all the words to lowercase.
- b) Remove punctuations and numbers as it is not giving any information for clustering.
- c) Remove all the stop words occurring in the document.
- d) Perform document stemming.
- e) Remove all the whitespaces.

Step2: Constructing the graphs

- a) Find the term-document matrix for each the documents. Let us denote it by TDM_i where i is in $\{1, 2, 3, \dots, k\}$
- b) From the TDM calculate cosine similarity for each pair of the matrix.
- c) Prepare a list of nearest neighbours for every document i.
- d) We construct a nearest neighbour graph. Each document is represented by a node. Any two nodes i and j will have an edge if i is on j's nearest neighbour list and vice versa.
- e) From the nearest neighbour graph, we will construct shared nearest neighbour graph. The strength of link is equal to number of shared nearest neighbours for i and j.

Step 3: Clustering algorithm

- a) We will have two parameters associated with the links.
- b) First, we calculate for each document i, number of strong links. (to be a strong link, strength of link should be above a certain threshold)
 - c) If the number of strong links for a document is lesser than noise threshold, it is not considered for clustering.
 - d) If the number of strong links is greater than topic threshold, it is made to represent the neighbourhood.
 - e) For any pair of documents i and j, if strength of link is greater than merge threshold, the documents will appear together in the clustering.

While the probability of a document belonging to a class different from its nearest neighbour's class may be relatively high, this probability decreases as the two documents share more and more neighbours. This is the main idea behind the algorithm. [7]

V. RESULTS AND DISCUSSION

For discussion we are considering 10 text documents from various categories.

A snapshot of TDM matrix for 10 documents:

Terms	Docs									
	1	1	1	1	1	1	1	1	1	1
account	2	0	0	0	0	0	0	2	0	0
adjust	1	0	0	0	0	0	0	0	0	0
advert	1	0	0	0	0	0	0	0	0	0
advertis	2	0	0	0	0	0	0	0	0	1
alexand	1	0	0	0	0	0	0	0	0	0
already	1	0	1	0	0	0	0	0	0	1
also	2	1	0	0	0	1	1	1	0	1
amount	1	0	0	0	0	0	1	0	0	0
analyst	1	0	0	0	0	0	1	0	0	0
aol	7	0	0	0	0	0	0	0	0	0
around	1	0	0	0	0	0	0	1	0	0
asid	1	0	0	0	0	0	0	0	0	0
back	1	0	1	1	0	0	0	0	0	0
benefit	1	0	0	0	0	0	0	0	0	0

Fig 2: TDM Matrix

Cosine similarity matrix for 5 documents is given below:

	1	2	3	4	5
1	1.0000000	0.049672006	0.14653242	0.08518655	0.08426366
2	0.04967201	1.000000000	0.04285714	0.05339685	0.03929887
3	0.14653242	0.042857143	1.21428571	0.09737072	0.05239850
4	0.08518655	0.053396848	0.09737072	1.0000000	0.04896351
5	0.08426366	0.039298875	0.05239850	0.04896351	1.0000000
6	0.05402064	0.044389676	0.06103580	0.02439983	0.03561615
7	0.10864648	0.064851907	0.04716502	0.05962833	0.07298284
8	0.09692677	0.042420347	0.07878064	0.10392871	0.05001216
9	0.05938686	0.028466193	0.08223567	0.06676100	0.11891253
10	0.08656746	0.007780276	0.14393510	0.03079160	0.08204455

Fig 3: Cosine Similarity Matrix

Once the cosine similarity is calculated, we will construct the nearest neighbor graph.

For that we need to find out the nearest neighbor list. We will set our first threshold value to pick only the most similar documents for the nearest neighbor list. We have constructed a nearest neighbor graph from the nearest neighbor list. Thickness of the edge represents similarity between two documents. The nearest neighbor graph for 10 documents is as follows:

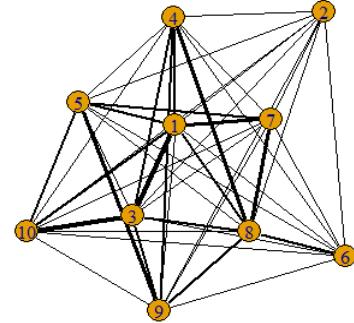


Fig 4: Nearest Neighbor Graph

Next, we calculate the strong links. Then for each document number of strong links, followed by the labelling step and finally the cluster formation.

VI. CONCLUSION AND FUTURE SCOPE

The size of nearest neighbor list here is an important factor. If the size is 1 we will find a pair of documents having each other on the list. And if the size is 60 then we can have a list containing neighbors ranging from 1 to several 100's. Not all the shared nearest neighbors contribute equally, as when a document which is shared is at a higher place in both the documents nearest neighbor list, it should contribute more to the similarity between two documents. This could be implemented as a future addition to the algorithm discussed in the paper.

REFERENCES

- [1] Supervised and unsupervised learning (Sept, 2014), Retrieved from: <http://dataaspirant.com/2014/09/19/supervised-and-unsupervised-learning/>
- [2] Criterion Functions for Document Clustering: Experiments and Analysis (2001), Ying Zhao and George Karypis, Retrieved from <http://glaros.dtc.umn.edu/gkhome/node/165>
- [3] Hierarchical Document Clustering Using Frequent Itemsets, Benjamin C.M. Fung, Retrieved from <https://www.cs.sfu.ca/~ester/papers/FWE03Camera.pdf>
- [4] Hsi-Cheng Chang Chiun-Chieh Hsu, Using Topic Keyword Clusters for Automatic Document Clustering, Proceedings of the Third International Conference on Information Technology and Applications (ICITA'05)
- [5] K. I. M. Kono, K. O. Youngjoong, and S. E. O. Jungyun, 2003, "Topic keyword identification for text summarization using lexical clustering", IEICE transactions on information and systems 86.9: 1695-1701.
- [6] Ankur Agarwal, Hurtado, Jose L., and Xingquan Zhu, 2016, "Topic discovery and future trend forecasting for texts", Journal of Big Data 3.1: 7.

- [7] Information Retrieval and Clustering, W. Wu, H. Xiong and S. Shekhar(Eds.), 2002 Kluwer Academic Publishers, Finding Topics in Collections of Documents: A Shared Nearest Neighbour Approach.

AUTOMATED VILLAGE - A WAY TO PROGRESS

Anand Khandare^[1], Pratiksha Singh^[2], Tanvi Sanzgiri^[3], Shaheen Sayyed^[4], Tanya Talwar^[5]
*Thakur College of Engineering & Technology
Mumbai*

Abstract— There is a quite good number of people in India staying in villages. If we want to improvise the condition of our country, we should commence in order to improve condition of our villages. The condition of Indian villages is not appreciable, people are not well settled and major issue is the source of earning which is mostly farming. In village, technology acts as a catalyst for development, enabling local business opportunities, education and improving employability.

We thought of a system (web portal) which will help Gram Panchayat to collect and analyze data to draw conclusions and process in the direction of development especially at financial level. The main concern about the system is to collect the data of the population of village and classify them on the basis of earning criteria. Data consist of information like (head of the family, earning members, no of children, whether children are taking education or not), this data will help them to analyze economy level of the village.

Portal will consist of modules which will store, maintain, monitor, and update data. In the present scenario data of villagers is manually maintain and Gram Panchayat official need to go through entire record diary of the records for data access.

Keywords: farming, earning members, head of the family, financial level.

I. INTRODUCTION

A. Background

Gram sabha is generally formed in the village with population at least exceeding 1500 people . Gram sabha plays a vital role in making sanitary arrangement ,development of employment, etc. Gram sabha also maintains information of villagers (head of the family, earning members, no of children, whether children are taking education or not). In the current scenario all this information is managed manually recorded in registers. All the Gram Panchayat have already been supplied desktop, laptop, printer, MS office package etc. out of 12th finance Commission grants. So we came up with a web portal /system for the automated management of data.

B. Problem Definition

Our web portal is divided into three modules .Data entering module, Data retrieval module and Data update module .In data entering module the villager information will be entered by the gram sabha official .Each family will have a f-id as reference number to the family. All the data will be saved in database in a cluster form and clustering will be based on the f-id. Rows with same f-id will belong to the same cluster. Data retrieval and data access module is for easy data access. For data retrieval, f-id is given as input in a query

that will run in the background which will fetch the cluster. Similarly for updating data, f-id will be used. Our proposed system will replace the current manual working of gram sabha official and will lead to better management of data and time. This one initiative may contribute to help the rural areas of our country at the local level. This may also support the campaign of our Prime Minister, A digitalization process in every corner of our country.

II. RELATED WORK

^[1]There are many villages in Maharashtra which are yet to be developed completely. Few officials has taken initiative to make the rural community digitized .In Maharashtra, Arisal village which is known as malnutrition capital' has taken into consideration by Chief Minister Devendra Fadnavis giving hope to make the village the first digital village. Addressing Microsoft Future Unleashed: Accelerating India' expo, Mr. Fadnavis said, "With help from Microsoft, the first digital village is coming up at Arisal, which is the capital of malnutrition in the country,' adding that the state plans to have 50 digital hamlets. However this was just one step for betterment of small villages. There is a need to create such more than 100 villages of our country.

^[2]In Gujarat, the Akodara village (Sabarkantha district) has bagged the tag of becoming the first digital village of India. The total Population of village is 1,191 people and about 250 houses uses cashless system for payments. Most of the transaction in the village are done through digital modes like SMS, net-banking or debit cards.

^[3]Common Services Center Gram Panchayat Mansai is the portal which provide the e-governance services to villagers where they have put the Panchayat into e-system under digital India programme which has make the manual method of services of the village into the digitalized method. Working Plan by CSC: Boost the economy by put new and better techniques. Common Services Centers(CSCS) are a strategic important node of the Digital India Programme. They are ACCESS POINT for providing various services to different villages in India.

^[4]Under the Digital India Mission, 'Digital Gram Panchayat, Nani ' was the India's first Android App.

This mobile application is of Village Nani, Post Sihot Chhoti, Via Phagwala, Dist Sikar (Rajasthan) INDIA PIN 332001.The App facilitates in getting the Information of elected PRI Representatives like Member of Zila Parishad, Member of Block Samiti, panchayat wise Sarpanch Detail & Panchayat wise/Ward wise detail of Panches. user can make direct Call to the desired representatives.

[5]It is a manual prepared under a Joint Ministry of Panchayati Raj-UNDP Programme, on village planning.

The 73rd and 74th Amendments to the constitution mandated the establishment of Panchayats at various levels of village. It also stated that Panchayats will form a structure to develop economically. It includes capacity development for local governance and presents the village development in easily understandable format to achieve it successfully.

[6]The Vagrancy Act 1824 (5 Geo. 4. c. 83) is an Act of Parliament of the United Kingdom that makes it an offence to sleep rough or beg. Anyone in England and Wales found to be homeless or to be trying to cadge subsistence money can be arrested. A law was enacted to deal with the increasing numbers of homeless and penniless urban poor in England and Wales following the conclusion of the Napoleonic Wars in 1815. Contemporary critics condemned the Act for being a catch-all offence because it did not consider the circumstances as to why an individual might be placed in such a predicament.

[7]Every Person who under the Provisions of this Act shall have been convicted as an idle and disorderly Person, or as a Rogue and Vagabond, shall be deemed to be actually chargeable to the Parish, Township, or Place in which such Person shall reside; and such Person shall be liable to be removed to the Parish of his or her last legal Settlement, by the Order of Two Justices of the Peace of the Division or Place in which such Person shall reside.

[8]Indian Government, at all levels, announces Welfare Schemes for a cross section of the society from time to time. These schemes could be either Central, State specific or a joint collaboration between the Centre and the States. In this section, we have attempted to provide you an easy and single point access to information about several welfare schemes of the Government and their various aspects including eligible beneficiaries, types of benefits, scheme details etc.

[9]A drift from physical power based labor to IT software works, permanency of jobs to project based organization, Government reforms that are moving from pro labor to pro capitalist has led to impact on labor system in India. Liberalization of Indian economy and increasing globalization and foreign companies interested in Investment in India , Global market uncertainty making corporations unsure of the workforce they would need has led to rise in contract based labor system and out sourcing. The article tries to portray the change of process of hiring; impact of contract labor on society and companies, concerns for employers, influence of these economic conditions on Contract labor, Act, 1970 is discussed in this article. Faced with economic onslaught unleashed by multinationals and global giants in the wake of introduction of contemporary policies of globalization and economic liberalization, Indian employers have been clamoring for more and more flexibility in their dealings with employees. Further, an introduction of Information technology and variety in product range has added the need to have workforce only on temporary basis, most commonly known as contract labor.

[10]VMS is not just a single program, but a suite of management tools including everything from accounting, compliance, collections, and delinquency control, to interactive web modules allowing residents, property managers, board members and vendors to interact and communicate in real time. Find out what thousands of homeowners associations, condominium associations and apartment/rental

management companies across the country already know...Village Management Software can and will increase the efficiency and productivity of your property management business, giving your company the competitive edge it needs to thrive and grow.

[11]Vagrants in Calcutta are taught to take a new direction in life. Every year, the state Vagrancy Department organizes the Bhabaghure Mela (vagrants' fair). At the week-long event last month, stocks had to be replenished on the third day. By the sixth and last day, the department had made over ₹ 50,000 from sales. The government policy has brought around 70 percent of the vagrants to the homes have turned into small-time traders. With some guidance, they've become carpenters, weavers, printers, bookbinders, coir-goods makers, metal smiths, even sculptors. The products are eventually sold on the open market.

[12]The Government of India has proposed the problem of the beggary in the Ministry Of Social Justice And Empowerment Lok Sabha. The Office of Registrar General & Census Commissioner, India has informed that according to the Census 2011, the total population of Beggars, Vagrants etc. in India is 413670. The States are responsible for taking necessary preventive and rehabilitative steps. As per available information, 20 States and 2 Union Territories have either enacted their own Anti-Beggary Legislation or adopted legislations enacted by other States/UTs. Presently, there is no scheme of the Central Government on Beggary.

[13]As Digital India Initiative, The Government of India has created Open Government Data (OGD) Platform. The portal is intended to be used by Government of India Ministries/ Departments their organizations to publish datasets, documents, services, tools and applications collected by them for public use. It intends to increase transparency in the functioning of Government and also open avenues for many more innovative uses of Government Data to give different perspective. It has also maintained the data of vagrants in India, including their basic information and type of work.

III. PROPOSED METHODOLOGY

This proposal is meant for providing one solution to help the zero income segment of population that is beggars, unskilled people, in village. Since we are familiar with the Agenda and Requirements, this portal will be developed using waterfall software development model.

Objectives are as follows-

- (1)Automatic Management of village's data
- (2)Online Portal for helping needy and people
- (3)Reduce Poverty
- (4)Maintain Database for villagers
- (5)Literate the villager's child
- (6)Improve financial condition of village

Process Model is-

PHASE I

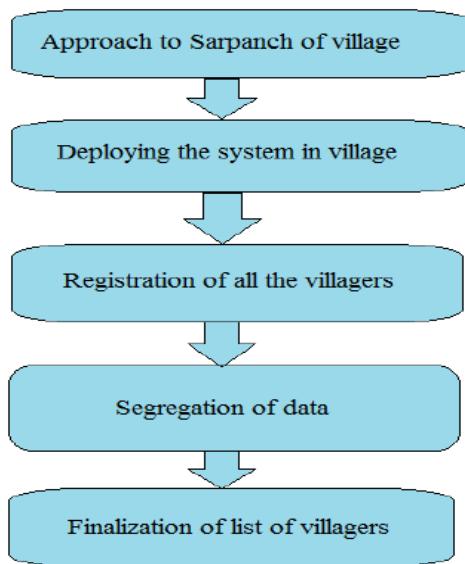


Fig. 1: Data Collection & Segregation

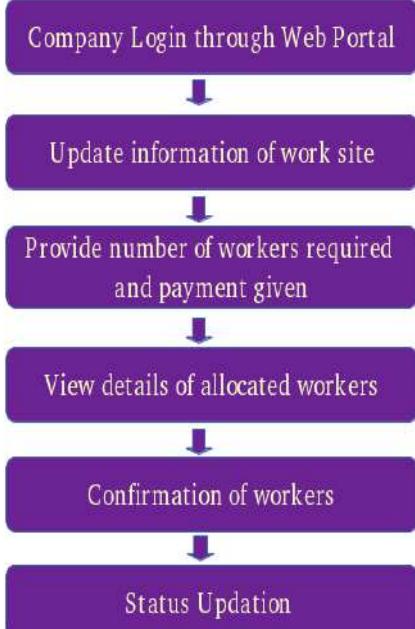
PHASE II

Fig. 2: Work Allotment & Updating

Technology used is-
FrontEnd: HTML, BOOTSTRAP, CSS, JS
Back End: PHP
Database: Mysql

IV. RESULTS AND DISCUSSION

Fig. 3: Front Page



Fig. 4: Insertion & Updating Data Modules



Fig. 5: Data Management Modules

V. CONCLUSION & FUTURE SCOPE

Under Panchayat raj system, the role of Gram Panchayat is very important. The progress and development of village depends on it. It is called the pillar of "Gram Swarajya" or "Gram Suraj". All the Gram Panchayat have already been supplied desktop, laptop, printer, MS

office package etc. out of 12th finance Commission grants. But, still retrieving the contents is a manual work. Our country being in the advancement process, we need to get free of manual era and move on to getting automated.

The overall efficiency of the village is improved and accelerated. In the rapidly expanding technological wonderland, that is the 21st century, a digital village is a cherry on the cake of the country. This project would help in reducing paper work. The data to be accessed would be well organized, easily searchable and shareable, can be reorganized. The entire system is fast, reliable, repeatable, reusable and comprehensive. Security a major issue will also be tackled by taking backups. Therefore, the village will put a step ahead in development making India truly a developing nation.

REFERENCES

- [1] <https://www.ndtv.com/india-news/araisal-to-be-maharashtras-first-digital-village-devendra-fadnavis-1240514>
- [2] <http://akodara-digitalvillage.in>
- [3] <http://www.digitalmansai.in/>
- [4] <https://play.google.com/store/apps/details?id=suren.grampanchayatnami>
- [5] <http://www.undp.org/content/dam/india/docs/DG/preparation-of-manual-for-planning-integrated-village-development.pdf>
- [6] https://en.wikipedia.org/wiki/Vagrancy_Act_1824
- [7] http://www.legislation.gov.uk/ukpga/1824/83/pdfs/ukpga_18240083_en.pdf
- [8] <https://www.sarkariyojna.co.in/central-government/>
- [9] <http://www.legalservicesindia.com/article/article/perspectives-on-contract-labour-act-1452-1.html>
- [10] http://www.isprs.org/caravan/documents/GIC_AV_Database.pdf
- [11] <http://indiatoday.intoday.in/story/vagrants-in-calcutta-are-taught-to-take-a-new-direction-in-life/1/244230.html>
- [12] <http://164.100.47.190/loksabhaquestions/annex/7/AU1831.pdf>
- [13] <https://data.gov.in/keywords/vagrants>

Augmented Situation Awareness and Intelligence Using Data Mining

Rohit Shashi Dubey, Akash Mishra, Sunil Yadav, Devyashwar Raj

Department of Information Technology, Mumbai, India

rohititboyy@gmail.com, aaloknath.mishra@gmail.com, sunil.yadav@slrtce.in, devyashwarraj@gmail.com

Abstract—Geospatial Data has an important role for advancing sustainable and development goals. Geospatial technology can highly benefit the traditional data in the field of administration, statistics and business analytics. In this report, we have present geospatial approach towards criminology. Geospatial provides an alternative to existing method of approach to analyze the crime pattern geographically

Keywords—Kriging Model, Geospatial Data, Crime, Prediction

I. INTRODUCTION

Data mining is a popular technological method that converts chunks of data into useful knowledge and information that can help the data users make intelligent choices and take efficient steps for their own benefit. Specifically, data mining used for hidden patterns amongst huge sets of data that can help to understand, predict, and guide future behavior. Data Mining is the set of methodologies used in analyzing data from various dimensions and perspectives, finding previously unknown hidden patterns, classifying and grouping the data and summarizing the identified relationships.

There is an enormous dataset that allow academics and practitioners to identify and map different types of land use (e.g. residential, commercial, industrial, leisure and Public services) different types of neighborhood (e.g. Disadvantaged inner cities, affluent suburbs, Student areas), together with street networks and major transport routes. Information such as this, referred to as “Geospatial data”, can potentially be very useful in understanding crime patterns and in explaining why crime hotspots occur in particular places. However, the extent to which beneficiaries are aware that such data exists is largely unknown and the added value that such data might bring is largely unexplored. An important component of this project is to investigate geospatial data for crime mapping and crime analysis.

II. SPATIAL PREDICTION – KRIGING ALGORITHM

Spatial prediction involves some component of randomness. This is crucial with geostatistics when you’re making inferences on a data set kriging weights are estimated from the variogram. The quality of the estimate surface is reflected in the quality of the weights. You want weights that give an unbiased prediction and the smallest variance.

Kriging is an advanced geostatistical procedure that generates an estimated surface from a scattered set of points with z-values. Unlike other interpolation methods in the Interpolation toolset, to use the Kriging tool effectively involves an interactive investigation

of the spatial behavior of the phenomenon represented by the z-values before you select the best estimation method for generating the output surface. Ordinary kriging based on centroids of administrative units to produce a surface of homicide rates and to identify clusters. However, recent advances in geostatistical methodology, such as area-to-area (ATA) and area-to-point (ATP) kriging and Poisson kriging, have opened up new opportunities.: Variography

Fitting a model, or spatial modeling, is also known as structural analysis, or variography. In spatial modeling of the structure of the measured points, you begin with a graph of the empirical semi-variogram, computed with the following equation for all pairs of locations separated by distance h :

$$\text{Semi variogram}(distance) = 0.5 * \text{average}((value - value)^2)$$

Each pair of locations has a unique distance, and there are often many pairs of points. To plot all pairs quickly becomes unmanageable. Instead of plotting each pair, the pairs are grouped into lag bins. For example, compute the average semi-variance for all pairs of points that are greater than 40 meters apart but less than 50 meters. The empirical semi-variogram is a graph of the averaged semi-variogram values on the y-axis and the distance (or lag) on the x-axis.

A. Mathematical Model

SPHERICAL

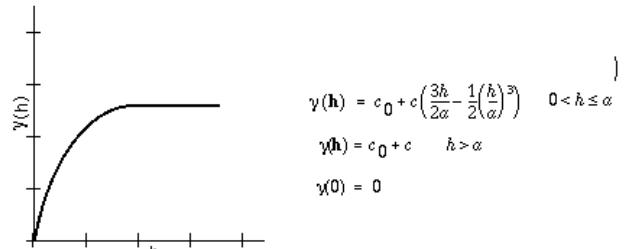
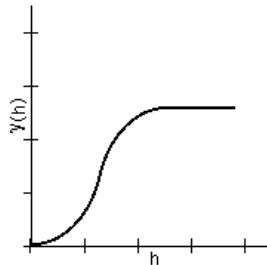


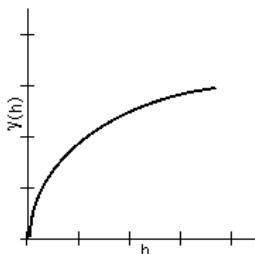
Fig (a): Spherical Model

GAUSSIAN

$$\gamma(h) = c_0 + c \left(1 - \exp\left(-\frac{h^2}{r^2}\right)\right)$$

$$\gamma(0) = 0$$

Fig (b): Gaussian Model

EXPONENTIAL

$$\gamma(h) = c_0 + c \left(1 - \exp\left(-\frac{h}{r}\right)\right) \quad h > 0$$

$$\gamma(0) = 0$$

Fig (c): Exponential Model

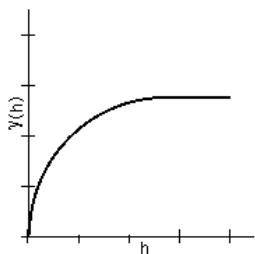
CIRCULAR

Fig (d): Circular Model

B. Syntax

```
Kriging_3d(in_point_features, z_field, out_surface_raster,
semiVariogram_props, {cell_size}, {search_radius},
{out_variance_prediction_raster})
```

Kriging example 1 (Python window)

```
import arcpy
from arcpy import env
env.workspace = "C:/data"
arcpy.Kriging_3d("ca_ozone_pts.shp", "OZONE",
"c:/output/krigout",
```

$$\text{"Spherical", 2000, "Variable 12"} \alpha + \beta = \chi. \quad (1)$$

C. Kriging example 2 (stand-alone script)

This example inputs a point shapefile and interpolates the output surface as a Grid raster.

```
# Name: Kriging_3d_Ex_02.py
# Description: Interpolates a surface from points using kriging.
# Requirements: 3D Analyst Extension
# Import system modules
```

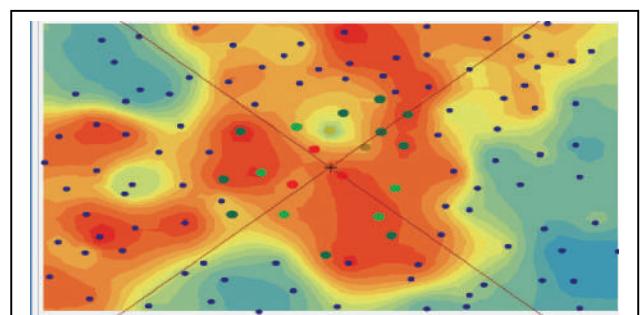
```
import arcpy
from arcpy import env
```

```
# Set environment settings
env.workspace = "C:/data"
```

```
# Set local variables
inFeatures = "ca_ozone_pts.shp"
field = "OZONE"
outRaster = "C:/output/krigoutput02"
cellSize = 2000
outVarRaster = "C:/output/outvariance"
kModel = "CIRCULAR"
kRadius = 20000
```

```
# Check out the ArcGIS 3D Analyst extension license
arcpy.CheckOutExtension("3D")
```

```
# Execute Kriging
arcpy.Kriging_3d(inFeatures, field, outRaster, kModel, cellSize,
kRadius, outVarRaster)
```

III. REVIEW OF METHODS**VARIOGRAM ANALYSIS**

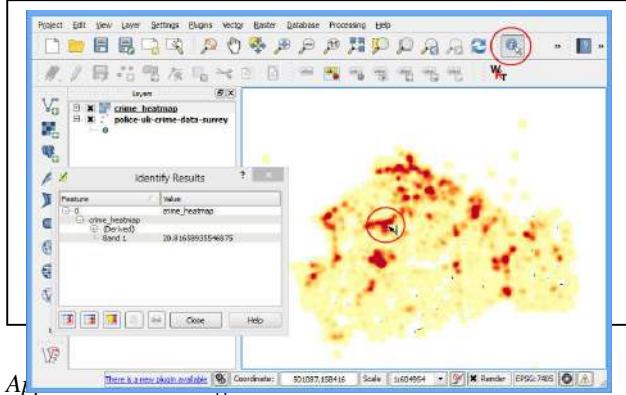
Relies on the semi-variogram. in simple terms, semivariograms quantify autocorrelation because it graphs out the variance of all pairs of data according to distance. chances are that closer things are more related and have small semi-variance. while far things are less related and have a high semi-variance. but at a certain distance (range), autocorrelation becomes independent. where that variation levels off, it's called (sill). this means there is no longer any spatial autocorrelation or relationship between the closeness of your data points.

$$\alpha + \beta = \chi. \quad (1) \quad (1)$$

HOTSPOT-ANALYSIS

We have been able to successfully apply some spatial techniques which include the use of location, the development of kernel surface estimation algorithms, and Local Indicators of Spatial Association such as Getis and Ord Gi. Some researchers have also recognized that a simple spatial concentration of crime can be valuable. A number of similar programs is available which will not give a specific shape but would produce a hotspot surface map called heatmap, rather than similar to the temperature that we see in the weather report. These maps actually show kernel density surface map of the crime intensity also, this process involves estimating the density of crime across an entire two-dimensional study area, based on the known locations of. KDE begins with overlaying grid on top of study area and calculating a density estimate based on the center points of each grid cell. Each distance between an incident and the center of a grid cell is then weighted based on a specific method of interpolation and bandwidth. The heat map process using QGIS and shows a number of parameters that must be considered before a density estimate can be produced. These parameters include the grid cell size, the method of interpolation, and the bandwidth.

QGIS provides an ease of implementation and interpretation. It is directly automating an algorithm to draw the data's cluster and process it.



The system will be designed to predict the different situations leading criminal activities and will create awareness by using data mining prediction techniques. We will use the vast amount of data which is available locally and if available in social media such as twitter and by using these data we will try to predict the outcome of the different criminal situation. Moreover, with this project we could also compare different algorithms and use other methods in social data mining to predict the outcome.

1) GIS aids crime analysis by.:

- a) Identification and mapping crime incidents and events that may require further investigation;
- b) Supporting pattern and analysis across multiple jurisdictions;
- c) Improving the implementation of various policing methodologies to reduce overall crime and disorder.
- d) Integrating traditional and nontraditional law enforcement data to improve overall analysis.
- e) Educating the public with visual information to clarify crime concerns and enlist community action.

f) Providing tools and techniques to capture crime series and forecast future crime occurrences.

g) *Tactical, strategic, or administrative problem:* Any data containing location information can be displayed and analysed using geographic information system (GIS) technology. GIS is an essential part of a crime analyst's (Police, Law Enforcement officers etc.) toolkit – means of creating valuable information for the concerned officers (on-duty) in the field. By incorporating traditional law enforcement data with data such as demographics, infrastructure, and offender tracking, user can use GIS to transform information into actionable intelligence. GIS will also help to improve critical decision making in a rapidly changing environment and have a direct impact on the safety of the on-duty officers and the citizens they are serving for. Every crime problem is related to some location, whether it's an address, street, ZIP Code, or district. GIS can help user to leverage the locational aspect of the data to analyse, understand, and build solutions to the problems user foreseen.

h) Violence against women in India:

Crimes which are directed specifically against women and in which only women are victims are characterised as crime against women. It is equally important to clarify the concept of 'Violence against Women'. Violence is also known as abuse and includes any sort of physical aggression or, misbehaves. When violence is committed at home it becomes domestic violence and involves family members such as children, spouse, parents or servants. Domestic violence may involve different means such as hitting, kicking, biting, shoving, restraining, and throwing objects. In broad terms, it includes threats, sexual abuse, emotional abuse, controlling or domineering, intimidation, stalking, passive/covert abuse and economic deprivation, rape, abduction, kidnapping, murder (all cases of criminal violence, dowry death, wife battering, sexual abuse, maltreatment of a widow and for an elderly women (all cases of domestic violence) and eve-teasing, forcing wife/daughter-in-law to go for feticide, forcing a young widow to commit sati, etc (all cases of social violence), are issues which affect a large section of society. The United Nations (UN) defined 'Violence against Women' in 1993 in declaration on the 'Elimination of Violence against Women'. It defines it as any act of gender-based violence that results in, or is likely to result in, physical, sexual or psychological harm or suffering to women, including threats of such acts, coercion or arbitrary deprivation of liberty, whether occurring in public or private life..

IV. CONCLUSION

In this paper, spatial prediction of crime location is analyzed and predicted. It is used to make short-term forecasting of property crime for one city of Maharashtra. The result shows that geospatial model fits the data well and makes higher accurate forecasting. This

work is proved to be very helpful to the local police stations and municipal governments in improving the efficiency of decision-making and emergency management.

ACKNOWLEDGMENT

This work is based on data collected as part of the project “Augmented Situation awareness and Intelligence using Data Mining,” by Bachelor of Engineering students of Shree L.R Tiwari College of Engineering, under the supervision of MR. Sunil Yadav, Department of Information Technology, Shree L.R Tiwari College of Engineering, Thane, Maharashtra.

REFERENCES

- [1] ↗ ShyamVaranNath, “Crime Pattern Detection Using Data Mining”, IEEE, 2006.
- [2] ↗ Messner, S. F., Anselin, L., Baller, R. D., Hawkins, D. F., Deane, G. and Tolnay, S. E. (1999) The spatial
- [3] Patterning of county homicide rates: An application of Exploratory Spatial Data Analysis, Journal of
- [4] Quantitative Criminology, 15(4), 423–450.
- [5] Kyriakidis P. A Geostatistical Framework For Area-To-Point Spatial Interpolation. Geographical Analysis. 2004;36:259–89.
- [6] ↗ Gupta R., Rajitha K., Basu S., Mittal S.K., 2012. Application of GIS in Crime Analysis: Gateway to Safe City.
- [7] Patterns: Frontiers of practice (pp. 77–85). Thousand Oaks, CA: Sage.
- [9] ↗ Johnson, C.P., 2000. Crime Mapping and Analysis Using GIS. Geomatics 2000:
- [10] ↗ Leigh, A., Read, T. and Tilley, N. (1996) Problem-oriented policing., Police Research Group: Crime
- [11] ↗ Detection and Prevention Series, 75.

Artificial Limbic Brain Models using Neural Networks for Emotional Analysis

Anushri Arora, Akanksha Joshi

Btech Computer Science Department

Mukesh Patel School Of Technology Management and Engineering, Mumbai, India

anushriarora@gmail.com, akanksha276@gmail.com

Abstract- Artificial intelligence has steadily been gaining popularity and the attention of scientists and engineers in the past few decades. AI was built using mathematical techniques based on probability and statistics allowing machines to give their best guess for any solution. However, emotional incorporation in AI still remains vastly uncharted. This paper is in line with this vision. It puts forth an architectural representation for such an emotionally astute intelligence by discussing the major setbacks faced while creating a computational model for emotions. It does so by describing the various ways in which emotions are mapped artificially as well as by describing biologically inspired predictive models that learn from these mappings to generate vectors that allow for accurate predictions. Apart from this citations have been made to animal cognition for prosodic nuances of speech which may result in higher sophistication of existing systems to detect emotions.

Keywords- artificial intelligence; cognition; limbic brain; emotional intelligence; canine neurology.

I. INTRODUCTION

Neuroscience and psychology have long been treated as sciences that are rather isolated from the world of computing. However, recent advancements in machine learning algorithms and computational hardware has caused scientists to bridge the gap between these fields [1]. This has caused an evolutionary change in the way traditional computer programs are written.

In order to solve real world problems, programs are now making use of probabilistic models often termed as artificial neural networks to account for randomness and variability. These models are loosely designed on the working of the human brain.

While a number of well renowned scientists have been able to remodel a large aspect of this glorious form of engineering, yet one of the most crucial and highly prized aspect of humans remains starkly distinguished from the machines we have created, our emotions[3][4].

Emotions are fuzzy, they're complicated and this kind of unpredicted, uncategorized, highly nested nebula of labels is a perfect fit for such predictive models. With this in mind we decided to review the nature of probabilistic predictive models

when applied to this subtly differentiated plethora of human emotions.

To do so we have divided this review paper into six distinct sections. Section I being the Introduction, Section II elaborates on the Background, Section III expands on the Proposed Methodology which further sub-contains a Sensory Stimulation and its Preprocessing and B. Emotional Models, Section IV discusses the results, Section V concludes the paper and Section VI consists of the references cited in this paper.

II. BACKGROUND

It is difficult to understand the concept of consciousness and creativity which sets humans apart. To impart these features to machines, we first need to understand their root cause in humans themselves. This problem of imparting an abstraction of characteristics to generate knowledge, which can then be used for further understanding of human emotional psychology, can distinctively be seen as a four-part process consisting of 1. Data collection and validation, 2. Pre-processing, 3. Learning based on artificial emotional models and 4. Predicting in previously unseen scenarios. Based on our study of state of the art publications, we identify emotional models to play a crucial role in generating efficient results.[2][5][11]

Our brain consists of the cortical region(rational thinking) and the limbic region (emotional hub). All rational decisions are made by the cortical region, however the strongest and most crucial decisions are not only influenced by rational understanding of a situation but are also greatly affected by the emotional and environmental factors of the situation and that's where the limbic system comes into play. The limbic system lies right under the cerebrum, and on either side of the thalamus. It is a complex collection of different brain parts which together function to give the emotional judgment [2][11].

We aim to partially reproduce the same characteristics as the biological system into a computational model inspired by this neurological behavior.

III. PROPOSED METHODOLOGY

Based on this literature review an architectural model is presented for such an emotionally adept system. This can be visualized through fig 1.

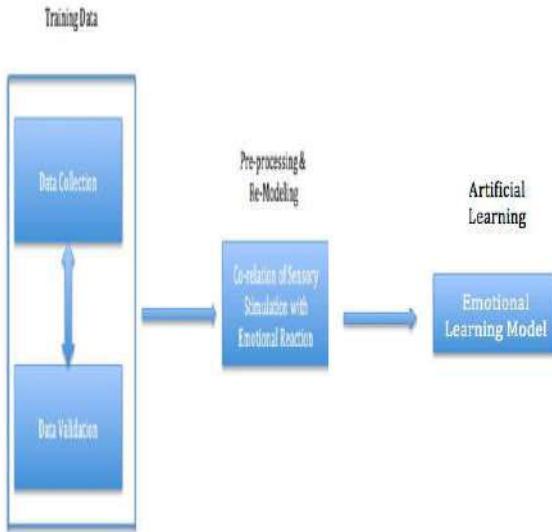


Fig. 1. Diagram representing architectural view of emotional model.

A. Sensory stimulation and preprocessing

Literary studies on emotion and cognition indicate their complexity and the inadequacy to subjectively quantify them. Hence the first key step is to create an accurate system of measurement. This is done by describing emotional reactions as a function of measurable sensory cues. Subsequently, a number of studies have made use of various quantifiable features to observe patterns and relations between these features and the emotions experienced [1][3][4][7].

Chung K Lee and his colleagues [3] attempted to measure the variation of emotions as a function of change in the Autonomic Nervous System (the ANS refers to that portion of the human nervous system that functions without any conscious thought), by measuring the ECG and GSR signals of the sample humans. The emotions were mapped to a two-dimensional space known as the valence-arousal plane, where valence indicates the nature of the emotion (positive or negative) and arousal indicates the amount experienced. These results were then used to train a Multilayer Perceptron (MLP) neural network, which in itself would then represent a particular emotion. The experiment was performed on a group of six individuals consisting of three males and three females and an estimating rate of 80.2% [3] correct prediction was achieved by using this method, indicating that constructing relations between the describing variables of the ANS and emotions provides reliable and efficient outcomes. The summarized results of the experiment can be seen in the table below.

Table I. Prediction Percentage of Trained Neural Network.

Emotion \ Test	a	b	c	d
a	84.6	13	2	0.4
b	5.6	77	11.6	5.5
c	0.9	11.3	71.7	16.1
d	3.3	4.1	5.2	87.5
Total accuracy				80.2 %

a: Sad, b: Calm pleasure, c: Interesting Pleasure, d: Fear. [3]

While the results of the designed classifier are categorically efficient it is primarily a static model and hence is not practically scalable for real time emotion detection. Apart from this, the model accounts for only four categories of emotions and thus the efficiency levels are unclear in case of larger number of classes or for differentiating between degrees of the same emotion.

One particularly interesting way to account for a larger degree of classification is done by regenerating the input data in a manner that specifically highlights its critical components which will directly affect emotional understanding. This form of pre-processing not only provides features that can be mapped to emotions but also allows for elimination of unwanted noise thereby increasing accuracy of prediction.

In recent years the ability to draw unbiased conclusions from data has shown tremendous advancements.

Meysam Asgari and his colleagues [4] introduce a method of prosodic and acoustic feature extraction, including shimmer and jitter from speech by employing harmonic models over the standard existing methods. This was used for screening adolescents with depression based on speech and spoken utterances. Correspondingly, the condition of the subjected adolescents were clinically benchmarked via the LIFE (living in family environment) coding system. Speech being a non stationary signal was framed using the Hamming window technique and voiced sequences were statistically determined. Features such as pitch frequency, HNR, H12, jitter shimmer and harmonic coefficients were approximated apart from the standard features which were used in case of the unvoiced segments of speech. These values were then used to obtain a per-utterance 192 dimension feature vector by applying standard summary statistics to each frame. This resulted in a 74% accuracy which is sufficient for preliminary screening of depressed individuals.[4] The summarised results are shown in Table II.

Table II. Comparison of performance of SVM classifier using different features for classifying clinical depression of adolescents [4].

Features	Classification Accuracy
Chance	52.4
Text	65.4
<i>openSMILE</i>	64.7 [†]
<i>openSMILE+Text</i>	68.0 [†]
Harmonic Model	68.7 [†]
Harmonic Model+Text	74.0 [†]

While these methods of generating qualitative feature spaces have been proven to be efficient, they are unable to account for one of the most intrinsic properties of human emotion, the continuous nature of its intensity. This implies that there are no number of finite dimensions which can be used to depict the infinitesimal variations of paroxysms of human emotion. In a recent study by Jing Han and his colleagues [10] a mathematical model is proposed where emotions are described in the form of states and their transference is derived based on both external stimuli as well as and more importantly an emotional intensity attenuation function, which in turn is defined in terms of emotion type, stimulus intensity and personality.

E: The intensity factor of the external

T Cognition reappraisal factor

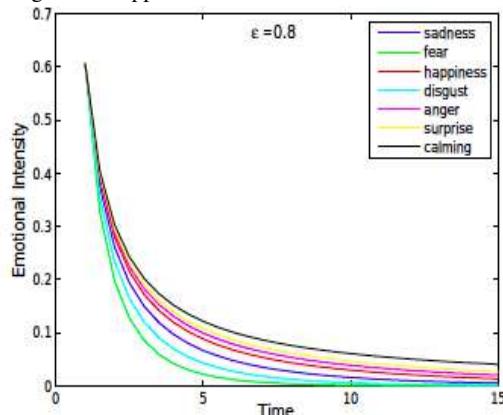


Fig. 2. Simulation result of Emotional Intensity Attenuation Trend under different emotional States [10].

Based on Gross emotion regulation model, Finite State Machine (FSM) and the psychological energy theory, the emotional state transference process is described. The initial values of emotional state transition probability matrix are calculated by the interaction between emotions. These are corrected by the emotional attenuation function which is a function of the cognitive reappraisal. This mathematical hypothesis was simulated and the results as shown below validate that the hypothesis can efficiently visualize the effect of the emotional

intensity on the transitional probability of emotional states.[10]

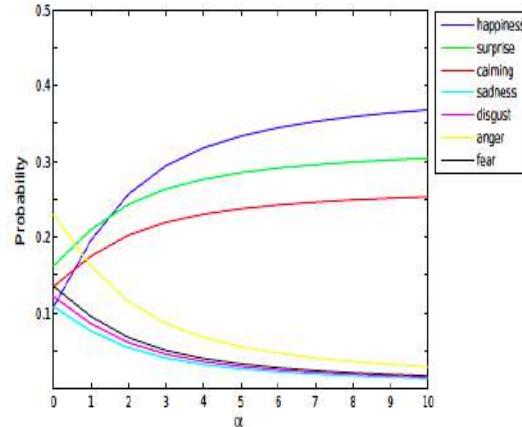


Fig. 3. Simulation result of Transition Probability curve with the external emotional stimuli being positive ($\alpha > 0$) [10].

B. Emotional models

Emotional experience has two distinct components in human beings: ‘automatic’ and ‘attended’. Automatic is based more on the ventral and limbic regions of the brain whereas the attention part is concerned with cognitive aspects of experience, and involves more of the analogous side of an invertebrate(dorsal).The possible dissociation between the cognitive and emotional components can be studied by first gathering data. This can be done by using a variety of methods including brain imaging and controlled experiments which analyze emotions compared to their neutral inputs. This data is then put into a model engineered for emotional analysis. Parietal and dorsal prefrontal sites create control signals to achieve attention focusing on a specific input. Using this data the model reverse engineers the parietal cortex and produces the attention signals which modulate the motor areas. An extension of this model involves a limbic brain network having representation for salience and valence. A limbic based neural network is thus created for better feedback mechanism. A feedback which has both rational and emotional thinking [2].

The limbic system is a complex system of different sub parts, each playing a vital role in emotional reaction and motivation. Amygdala is one such organ. It is responsible for emotional evaluation of the stimuli which is used as a basis for emotional states, reactions and is used to signal attention and form long-term memories [2][5][11].The amygdala and the orbitofrontal cortex are connected with two pathways involving the basal ganglia.The direct pathway that has an excitatory influence on cortical activity and the indirect pathway that has an inhibitory influence on the cortex.These connections could provide a means by which either positive or negative emotion could affect the judgments of the critic if present, which may influence the ideas influenced and retrieved [5][6].

They use a two pronged computational model consisting of the amygdaloid part and the orbit frontal cortex. The amygdaloid part receives inputs from the thalamus and from cortical areas, while the orbital part receives inputs from the cortical areas and the amygdala only. A reinforcing signal is also received by the system. The amygdaloid part learns to predict and react to a given reinforcer. This subsystem can never unlearn a connection

i.e. once learned, it is permanent. This gives the system the ability to retain emotional connections for as long as required. The orbitofrontal system checks for mismatches between the base systems predictions and the actual received reinforcer and will inhibit the system output in proportion to the mismatch. Basic simulations are run to test features of acquisition-extinction-reacquisition, simple blocking and conditioned inhibition. We will discuss acquisition-extinction-reacquisition and simple blocking.

C. Acquisition

In this, the model is expected to associate a stimulus with a reward/reinforcer, if the reinforcer is absent the stimulus should disassociate, and then re-associate them again.

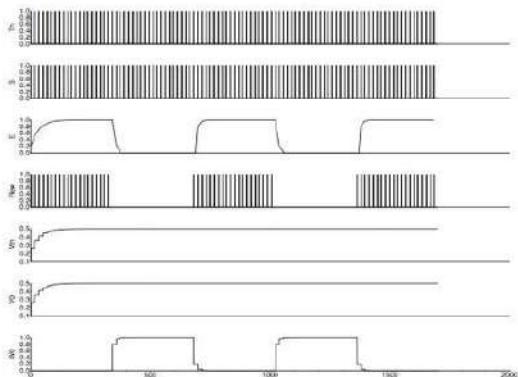


Fig. 4. Acquisition [5]

From Fig 4 we can see that the stimulus Sand the thalamic input Th occur simultaneously and with the same intensity, resulting in Vth and V0 sharing the responsibility for the association to Rew as can be seen in the output E. When the reinforcer disappears, the amygdaloid weights are not affected, instead the orbitofrontal weight W0 rapidly increases and inhibits the output. As soon as the reinforcer reappears, W0 decreases to zero, allowing the amygdala to express the previously learned association.

D. Blocking

In this blocking simulation we show the ability of the model not to associate a stimulus with the reinforce if there is

already an established association.

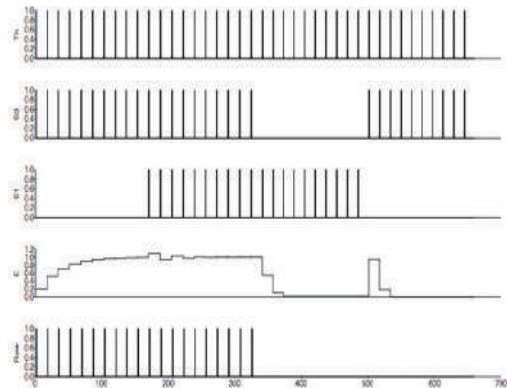


Fig. 5. Blocking [5]

This blocking schedule is run in three phases: first we associate S0 with the reinforcer, then present both S0 and S1 together with the reinforcement, and last, test S1 to see whether it has been associated with the reinforcer. We see there is no response to S1 in the output E. This is because when the system is presented with S1 and Th in the absence of a reward, the orbitofrontal part will learn to inhibit a response through the connection weight for S0.

These simulations give results as expected and indicate that the model contains the basic features needed for associative learning. However, the model is not a complete learning system. The two most important missing parts are a context model and some form of motor learning system that can use the output of this model. The context model would mostly be replacing the orbitofrontal inputs. A motor learning system would use the output of this model as a reinforcing signal for learning motor sequences.[5]

While emotional cognition of humans has been highly researched and replicated, animal cognition and neuroscience has been relatively ignored. A number of recent studies have pointed towards highly sophisticated and developed systems of auditory cognition and correlation of these signals with emotions especially in canines. [8][9].

As a result of thousands of years of domestication, dogs and humans have shared a similar environment. In a comparative study between primate and non-primate species, both dogs and humans were presented with the same set of vocal and nonvocal stimuli. Findings revealed that acoustical cues related to vocal emotional valence are processed similarly in the dog and human auditory cortex.[8] Recent advancements in canine neurology has shown that domestic dogs can perceive dog and human emotions from both auditory and visual inputs. In a cross modal experiment dogs were shown either human or dog faces with different emotional valences (happy / playful against angry / aggressive). This was paired with a single vocalization with either a positive or negative valence or Brownian noise from the same individual. Results showed that dogs looked significantly longer at the face whose expression were compatible with the valence of vocalization. Dogs performed in this way without any prior training and hence shows that these emotional signals are intrinsically important.[9]

IV. RESULTS AND DISCUSSION

As per the above review we find the following comparison highlighting the techniques which are most satisfactory in each category.

Table III. Comparative study of proposed architectures

Architectural Component	Brief Description	Statistical Results
1.Data Collection	As a supervised model of learning is imparted at a later stage this phase of data collection and classification becomes extremely vital. Emotional cues can be obtained through various sources such as speech, images, text, biological sensors etc. While a single source can be used, multiple source almost always tend to improve efficiency of prediction.	This paper focuses primarily on speech/ sound signals as test data. Apart from this using biological sensors to obtain signals like ECG and GSR which have high correlations with emotion turns out to be extremely advantageous majorly due to their non-maskable nature.
2.Biological basis of validation	The LIFE coding technique proves to provide more valuable and justifiable results in a multitude of settings. It also provides a larger degree of emotional understanding based on not only verbal but non-verbal cues as well as compared to traditional valence-arousal based emotional modeling.	LIFE coding technique provides 27 content codes and 10 affect codes. This methodology is implemented in a manner that replicates daily activity and thus provides higher levels of accuracy as compared to projection and classification of emotions in valence arousal plane.
3. Pre-processing	Ensures extraction of valuable features from the data and thereby elimination of unwanted components such as noise. This can be accomplished through a variety of ways each offering their own set of advantages.	As compared to traditional methodologies speech signals can be regenerated using harmonic models which act as a filter allowing only prosodic and acoustic features of speech to pass through. Apart from this state models can be implemented with their transition being governed by an attenuation function. This provides promising results as indicated above and accounts for the temporal nature of

		emotions.
4. Learning model	Data is subject to a model engineered for emotional analysis. A computational model having the amygdaloid part and the orbitofrontal part is used to receive inputs. The amygdala learns to predict a system and never unlearns a connection whereas the orbitofrontal part inhibits the mismatches between the base systems predictions and the actual received reinforcer.	Simulations for testing features of acquisition-extinction-reacquisition, simple blocking and conditioned inhibition indicate that the model contains the basic features of associative learning. The output of this model can be used by a motor learning system as a reinforcing signal for learning motor sequences.

V. CONCLUSION

This paper clearly outlines the major components involved in the designing of an emotionally appropriate artificial intelligence. Emotions even now remain to be highly convoluted and their continuous nature with no definite upper or lower bound of intensity make them highly challenging to be designed in a binary system of a computational model. However, highly efficient methods have been demonstrated throughout the course of this paper indicating that emotions have to be represented in terms of quantifiable sensory variables and efficiency has been highly increased if these variables are non maskable [1][3][4][12]. Also prosodic nuances can be regenerated allowing for higher degree of sensitivity[4][7]. Another key property of emotions is their time varying nature which has been accounted for by state models and by extending particle filter tracking to sample transition models with higher contextual probability[7][10]. This paper further highlights biologically inspired emotional learning models and the amygdaloid-orbitofrontal system whose two process model constructs a functional model of emotional processing as part of a general learning system.

An extension has also been proposed to regenerate cognitive models as seen in highly sensitive animals like dogs which may result in better coordinate emotions with the temporal, spatial, acoustic and prosodic components of speech[8][9].

REFERENCES

- [1] Rosalind W. Picard, Elias Vyzas and Jennifer Healey, "Toward Machines with Emotional Intelligence", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume-23, Issue-10, October 2001.
- [2] Tanushree, Shilpa yadav, and Yashika arora, "Limbic-based artificial emotional neural networks", International Journal of Research (IJR) Vol-1, Issue-9, October 2014.
- [3] ChungK Lee, SK Yoo, YoonJ Park, NamHyun Kim,KeeSam Jeong, and ByungChae Lee, "Using Neural Network to Recognize Human Emotions from Heart Rate Variability and Skin Resistance", IEEE,Shanghai,China,2005.

-
- [4] Meysam Asgari, Izhak Shafran, and Lisa B. Sheeber, "Inferring Clinical Depression From Speech And Spoken Utterances", IEEE, REIMS, FRANCE, 2014.
 - [5] Jan More'n and Christian Balkenius,"A Computational Model of Emotional Learning in the Amygdala".
 - [6] Daniel S. Levine and Mihai Nadin, "Towards a neural network theory of emotional influences on creativity",The 2013 International Joint Conference on Neural Networks (IJCNN), 2013.
 - [7] Ozgun Cirakman and Bilge Gunsel, "Online Speaker Emotion Tracking with a Dynamic State Transition Model", 23rd International Conference on Pattern Recognition (ICPR), Cancún Center, Cancún, México, December 4-8, 2016.
 - [8] Attila Andics, Ma'rta Ga'csi, Tama's Farago', Anna Kis, and A'da'm Miklo'si, "Voice-Sensitive Regions in the Dog and Human Brain Are Revealed by Comparative fMRI", Current Biology 24, 574–578, March 3, 2014.
 - [9] Natalia Albuquerque, Kun Guo, Anna Wilkinson, Carine Savalli, Emma Otta, and Daniel Mills, "Dogs recognize dog and human emotions", The Royal Society, 13 January 2016.
 - [10] Jing Han, Lun Xie, Xin Liu and Zhiliang Wang, "Emotional state transfer model based on FSM", International Multisensor Fusion and Information Integration for Intelligent Systems (MFI),28-29 September 2014.
 - [11] Sascha Fruehholz, "On the Role of the Limbic Brain System in Recognizing Emotions From Paralinguistic Speech Features", Speech Communication 12. ITG Symposium, 5-7 October 2016.
 - [12] Einat Liebenthal, David A. Silbersweig and Emily Stern, "The Language, Tone and Prosody of Emotions: Neural Substrates and Dynamics of Spoken-Word Emotion Perception",Frontiers in Neuroscience, 8 November 2016.

How to explore the Art of Computer security

Drashti Shrimal,Vinitta Sunish., Veena Kulkarni

*Department of Computer Engineering, Thakur College of Engineering & Technology,
Mumbai, India*

Abstract-The continuous development of computer network system brings both a great experience and convenience but new security threats for users. Computer security problem generally includes network system security and data security. Specially, it refers to the reliability of computer, confidentiality, integrity and availability of information information within the machine. Network security issue is at every layer of computer network, some might say the OSI model. Also Network security aims atmaintaining the authenticity, integrity, availability, confidentiality, dependability, and audit-ability of network.This paper mainly aims to discuss about the need for Computer security, implementations of security mechanisms, algorithms, policies and latest threats of various systems that are upcoming today.

Keywords-Goals, Cryptography, Cryptanalysis, Access Control Lists, Mechanisms, Bell-LaPadula, Biba.

I. INTRODUCTION

Computer security should be a basic management task in any firm. It is an extension of the duty to prevent the organization's assets against misuse or loss.

Also, the information extracted, manipulated, stored and processed by computers is the most significant asset of most organizations. (Technical experts to call the term information security to describe the process of protecting computing. It plays a vital role in compiling an organization's ability to survive as what the law calls a going concern. Moreover, maintaining this process will comprises of ensuring that the corporation is complying with relevant statutory and regulatory enterprise necessities.)

Information is inevitable in all kinds of entrepreneurial activities, and must be therefore protected as assets. Information security is assured in diverse approaches, inclusive of related guidelines, strategies, techniques, organizational structures, software program packages and hardware equipment capable of take away many resources of protection jeopardizing together with espionage, laptop fraud and deceit, sabotage, vandalism, hearth or water.

Computer security is the safety of computing structures and the information that they keep or get admission to.

What number of attacks do you suspect that takes place on regular basis?

Heaps of attacks in keeping with minute bombard our campus network.

An unprotected laptop can grow to be inflamed or compromised within a few seconds after it is linked to the community. A compromised machine is a chance of threat to every person else, too - not simply to you. [1]

II. LITERATURE SURVEY

A. Goals of Security

Computer security rests on Availability Integrity, Confidentiality and that is CIA. The translation of these elements range, as do the contexts wherein they rise up. The meaning of factor in a given situation or environment is dictated by using the needs of the people, customs and laws of particular businesses. But we could define it in a general way as follows-

1. CONFIDENTIALITY

Confidentiality is the concealment of information or resources. The need of preserving facts secret arises from the usage of computer in sensitive fields including government. Ex-military,banks.

2. INTEGRITY

Integrity refers back to the correctness and dependability on trustworthiness of data or assets and it usually phrased in terms of stopping mistaken or non-authorized change. Integrity includes integration of data and integrity of source.

3. AVAILABILITY

Availability refers back to the potential to use the information or resource wanted.

Computer security specialists usually cope with three commonplace challenges to availability:Denial of service (DoS) due to intentional assaults or due to undiscovered flaws in implementation (for example, a software written through a programmer who is blind to a flaw that would crash this system if a positive surprising input is encountered). loss of records machine skills because of natural screw ups (fires, floods, storms, or earthquakes) or human movements (bombs or strikes).[2]

B. Threats

A threat, within the boundaries of computer security, refers to something that has the capability to purpose serious harm to a laptop system. A chance is something that can or may not show up, however has the capability to cause severe harm. Threats can result in assaults on computer systems, networks and greater.

C. Cryptography

Cryptography means secret writing. Essentially it is writing text in secret form such that it is not understandable to attackers. Cryptanalysis is the breaking of codes. The primary thing of cryptography is Cryptosystem.

D. Policies:

1. Security Policies

A security model is a version that represents a selected coverage or set of regulations. A model abstracts info applicable for analysis. Analyses hardly ever talk precise guidelines; they normally cognizance on particular traits of guidelines, because many rules show off those characteristics; and the extra regulations with those traits, the more useful the evaluation. Through the HRU result, no single nontrivial evaluation can cowl all rules, however restricting the magnificence of security policies sufficiently permits significant evaluation of that magnificence of guidelines.[3]

2. Confidentiality Policies

Confidentiality is one of the elements of privacy, an issue diagnosed within the legal guidelines of many government entities (which include the privacy Act of the USA and similar rules in Sweden). Aside from constraining what information a government entity can legally obtain from individuals, such acts place constraints on the disclosure and use of that information. Unauthorized disclosure can result in penalties that include jail or fines; also, such disclosure undermines the authority and respect that individuals have for the government and inhibits them from disclosing that type of information to the agencies so compromised.

3. The Bell - LaPadula Model

The handiest type of confidentiality classification is a fixed of security clearances organized in a linear (general) ordering . these clearances represent sensitivity levels. The better the security clearance, the more sensitive the facts (and the extra the want to preserve it private). a subject has a security clearance. In the figure, Claire's security clearance is C (for CONFIDENTIAL), and Thomas' is TS (for TOP SECRET). An object has a security classification; the security classification of the electronic mail files is S (for SECRET), and that of the telephone list files is UC (for UNCLASSIFIED). (when we refer to each difficulty clearances and item classifications, we use the term "classification.") The purpose is to prevent read get admission to objects at a classification of security higher than the subject's clearance.

The Bell- LaPadula security model combines mandatory and discretionary get entry to controls. Following is, "S has discretionary read (write) access to O" way that the access control matrix access for S and O corresponding to the discretionary to manage issue incorporates a read (write) proper. In different words, were the required controls no longer present, S would be capable of read (write) O[3].

TOP SECRET (TS)	Tamara, Thomas	Personnel Files
SECRET (S)	Sally, Samuel	Electronic Mail Files
CONFIDENTIAL (C)	Claire, Clarence	Activity Log Files
UNCLASSIFIED (UC)	Ulalee, Ursula	Telephone List Files

Fig 1: Bell - LaPadula Model classification[3]

C. Integrity Policies

1. Biba Integrity Model-

Biba studied system integrity nature in 1977. His model, a system includes a hard and fast S of subjects, a set O of gadgets, and a set I of integrity levels. The levels in his model are ordered. The relation = ? $I \times I$ holds and the second one integrity degree dominates or is similar to the first. The function $i:S \cup O \rightarrow I$ returns the integrity level of an object or a subject [3].

2. Clark Wilson Integrity Models

In 1987, David Clark and David Wilson developed an integrity model radically different from previous models. This model uses transactions as the basic operation, which models many commercial systems more realistically than previous models. One important subject of a industrial environment, as discussed above, is the integrity of the information in the system and of the actions executed on that statistics. The facts is said to be in a steady kingdom (or regular) if it satisfies given properties. as an instance, allow D be the amount of money deposited to this point today, W the quantity of money withdrawn up to now nowadays, YB the quantity of cash in all bills on the end of the day before today, and TB the amount of money in all debts so far today. Then the consistency property is $D + YB - W = TB$ Before and after each action, the consistency conditions must hold.

A nicely-fashioned transaction is a chain of operations that transition the machine from one constant country to every other consistent state. stem from one consistent state to another. For example, if a depositor transfers money from one account to another, the transaction is the transfer; two operations, the deduction from the first account and the addition to the second account, make up this transaction. Each operation may leave the data in an inconsistent state, but the well-formed transaction must preserve consistency. [3]

III. IMPLEMENTATION - I

Implementing Computer security techniques fall under following types:

A. Cryptography

The art or science encompassing the standards and techniques of reworking an intelligible message into one that is unintelligible, after which re-remodeling that message again to its original shape.[4]

The classical Cryptosystem consists of following types-

1. Transposition Cipher

In this, it is a way of encryption by means of which the positions held through devices of simple textual content are shifted. The cipher text consists of a permutation of the obvious textual content.

2. Substitution cipher

In this, according to a set gadget, it is a way of encrypting by using units of simple textual content are changed with cipher text, according to a set gadget; the "gadgets" may be single letters, pairs of letters, triplets of letters, combinations of the above, and so on.

3. Vigenère cipher

It is a way of encrypting alphabetic text by using a chain of interwoven Caesar ciphers, which is based totally at the letters of a key-word .

4. One time pad

In this, a plain text is paired with a random mystery key. It is also known as a one-time pad. Then, every bit or character of the apparent text is encrypted by combining it with the corresponding bit or individual from the pad the use of modular addition. If the secret's truly random, is as a minimum as long as the apparent textual content, is by no means reused in entire or in component, and is stored completely secret, then the ensuing cipher text might be impossible to decrypt or damage.

5. Public key Cryptosystem

PKC works in way illustrated in following figure.

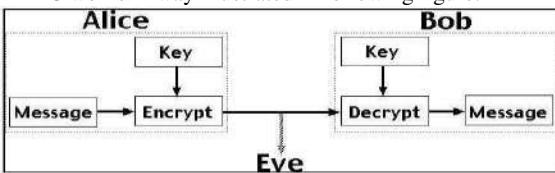


Fig 2: Working of Public key Cryptosystem[5]

There are 2 types of PKCs as follows:

A. Diffie-Hellman

It was the first PKC proposed. It is based on Discrete Logarithm Problem.

B. RSA

It is an exponential cipher. This type of cipher is even used today.

(Note: Algorithms to be followed in section)

IV. IMPLEMENTATION - II

Protecting cryptographic keys may sound simple: just put the key into file and use operating system access control mechanisms to protect it. But as we know in a number of ways these mechanisms can be compromised leading to keys getting invaded. In this section we discuss some mechanisms to prevent keys. Following are some key management techniques.

A. Kerberos

Kerberos offers a centralized authentication server whose feature is to authenticate customers to servers and servers to users. in contrast to maximum different authentication schemes described in this e book, Kerberos is based exclusively on symmetric encryption, making little need of public-key encryption.

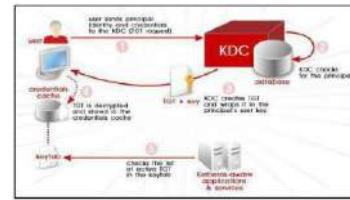


Fig 3: Working of Kerberos[6]

B. Key escrow

It (also called a “ truthful ” Cryptosystem) is an arrangement in which the keys needed to decrypt encrypted facts are held in escrow in order that, below positive circumstances, an authorized 0.33 birthday party may also gain get admission to to those keys . Key escrow is a records safety degree in which a cryptographic key's entrusted to a 3rd birthday party (i.e., saved in escrow). underneath normal circumstances, the key isn't released to someone other than the sender or receiver without proper authorization.

For the above key management techniques, various authentication techniques are used for verifying the user authenticity. Techniques include following:

1. Passwords
2. Challenge-Response (OTP)
3. Biometric
 - a. Fingerprint
 - b. Face recognition
 - c. Retina Scan
 - d. Face scan
 - e. Voice recognition

A combination of above three techniques is used for authenticity of users.

Access Control is a set of controls to restrict access to certain resources. If we think about it, access controls are everywhere around us. A door on your room, the guards permitting you to go into the office constructing on seeing your get entry to card, swiping your card and scanning your arms at the biometric device, a queue for meals at the canteen or getting into your credentials to access facebook, all are examples of numerous forms of get admission to manipulate. here we attention handiest at the logical get admission to control mechanisms.

1. Discretionary Access Control (DAC)

Discretionary access controls are based on get entry to rights at the identification of the difficulty and the identification of the item involved. identification is the key; the proprietor of the object constrains who can get right of entry to it by means of permitting simplest particular subjects to have get right of entry to. The owner states the constraint in phrases of the identification of the subject, or the proprietor of the challenge. example: assume a toddler maintains a diary.

The kid controls get entry to the diary, because she will permit someone to study it (provide read access) or now not permit a person to study it (deny study get right of entry to). the kid permits her mom to study it, but nobody else. that is a discretionary access

manipulate because get right of entry to to the diary is based totally on the identification of the challenge (mother) requesting examine get right of entry to to the item (the diary).

2. Mandatory Access Control (MAC)

When a system mechanism controls get admission to to an object and an man or woman user cannot modify that get admission to, the manager is a obligatory get entry to manage (MAC), on occasion known as a rule- based get admission to manage. The operating machine enforces obligatory get right of entry to controls. Neither the challenge nor the owner of the item can determine whether get entry to is granted. Normally, the device mechanism will take a look at information associated with both the situation and the item to determine whether or not the issue has to get entry to the object. Regulations describe the conditions beneath which access is allowed.

EXAMPLE: The regulation permits a court docket to access using data without the owners' permission.

3. Role Based Access Control (RBAC)

RBAC is the buzzword across corporations these days. in this model the access to a aid is governed based totally at the position that the issue holds within an agency. RBAC is likewise known as non-discretionary access control due to the fact the user inherits privileges which can be tied to his function. The user does now not have a control over the function that he will be assigned. each of the above get right of entry to models has its own benefits and drawbacks. the choice of the precise get right of entry to version by using an business enterprise should be carried out by way of thinking about various factors including type of business, no of customers, enterprise's safety coverage and so on.

4. Access Control lists(ACLs)

An obvious variation of the access manage matrix is to keep every column with the object it represents. As a result, every object has associated with it a set of pairs, with every pair containing a subject and a fixed of rights. The named situation can get admission to the related item using any of these rights. More formally:

Let S be set of subjects, and R be set of rights, of a system. An access control list (ACL) l is a set of pairs l = { (s,

r) : s ∈ S, r ⊆ R }. Let acl be a function that determines the access control list l associated with a particular object o. The interpretation of the access control list acl(o) = { (si, ri) : 1 ≤ i ≤ n } is that subject si may access o using any right in ri .

V. STEPWISE EXPLANATION OF

ALGORITHMS

A.Diffie-Hellman

ALICE	BOB
Public Keys available = P, G	Public Keys available = P, G
Private Key selected = a	Private Key selected = b
Key generated = x = G ^a mod P	Key generated = x = G ^b mod P
Exchange of generated keys takes place	
Key received = y	Key received = x
Generated Secret key = k _a = y ^b mod P	Generated Secret key = k _b = x ^a mod P
<i>Algebraically it can be shown that k_a = k_b</i>	
<i>Users now have a symmetric secret key to encrypt.</i>	

Fig 4: Deffie - Hellman Steps[7]

2. RSA

* Generating Public key:

Select two prime no's. Suppose A = 53 and B =

59. Now Initial part- Public key : n = A*B= 3127.

We need a small exponent say e :

But e Must be An integer.

Not be a factor of n.

1 < e < Φ(n) [Φ(n) is discussed

below], Let us now consider it to be 3.

* Generating Private Key :

We need to calculate Φ(n) :

Such that Φ(n) =(A-1)(B-1)

so, Φ(n) = 3016

Lets calculate Private Key, d : d = (k*Φ(n) + 1) / e for some integer k

For k = 2, value of d is 2011.

Now we are done with our – Public Key (n = 3127 and e = 3) & Private Key(d = 2011)

*Encryption:

Now lets encrypt "HI" :

Convert letters to numbers : H = 8 and I =

9 Thus Encrypted Data c = 89e mod n.

Thus our Encrypted Data comes out to be 1394

Now we will decrypt 1349 :

Decrypted Data = cd mod n.

Thus Encrypted Data is 898 = H and I =

9 i.e. "HI". [7]

V. ALGORITHM COMPARISON

Table I: Comparison of RSA and Diffie-Hellman[8]

No.	Parameters	RSA	Diffie-Hellman
1.	Encryption Cost	Cheaper Public key is smaller to encode.	Expensive
2.	Public key encoding	Less robust(1024 bits).	Public key is bigger to encode.
3.	Strength	Depends on difficulty of Integer Factorization.	More Robust(1024 bits).
4.	Security	Performs to only sender.	Depends on difficulty of Discrete Logarithm.
5.	Authenticati on	Extremely difficult.	Performs for both Sender and Receiver.
6.	Key generation	Common modulus and cycle attack.	Easier.
7.	Type of attacks possible		Man in the middle attack.

VI. LATEST RISKY THREATS

A famous approach used by internet site operators to study the keystrokes, mouse moves and scrolling conduct of site visitors on web pages is fraught with chance, according to researchers at Princeton's middle for records era coverage.

The method offered by a number of service carriers uses scripts to seize the activity of a tourist on an internet web page, shop it on the provider's servers, and play it back on call for for a website's operators. The idea at the back of the practice is to offer operators insights into how users are interacting with their web sites and to identify orphan and ambiguous pages. Let us see a few threats that are upcoming since the past few years:

A. Peeping Scripts

However, the quantity of statistics accrued by means of the scripts far exceeds consumer expectations, in step with researchers Steven Englehardt, Gunes Acar and Arvind Narayanan.

Textual content typed into bureaucracy is accrued before a user submits the shape, and specific mouse moves are stored -- all with none visual indication to the user, they stated in an online submit. what's greater, the data cannot be fairly predicted to be stored anonymous.

"In contrast to regular analytics services that provide mixture statistics, these scripts are meant for the recording and playback of character browsing periods, as though a person is looking over your shoulder." that means that whether or not a tourist completes a shape and submits it to the website or now not, any data keyed in on the internet site can be visible by way of the operator.[9]

B. chaiOS

software developer Abraham Masri claimed to have located the computer virus, called "chaiOS,"The so-called "textual content bomb" commonly causes an iPhone to crash and, in some instances, restart. Sending a message which incorporates the hyperlink to Masri's code would be all it takes to prompt the trojan horse — although the recipient did no longer click on at the hyperlink. Meanwhile, On a Mac laptop, the security flaw changed into determined to crash the Safari browser, as well as causing other slowdowns.[10]

C. Ransomware

Ransomware conserving companies information for ransom has surged up these days at an outstanding rate. And SonicWall reviews that ransomware tries have swelled up from 2.8 million in 2015 to 638 million final 12 months. The enterprise's file additionally confirms that as plenty as \$209 million became paid in 1Q of 2016 alone. for that reason the amount paid says a lot about malware.[11]

D. Internet of things Botnets

In past due 2016, whilst an considerable DDoS assault became launched on DNS service issuer called DYN, the assault proved that many service carriers were unwell-equipped to deal with the scope of the today's cyber attacks. Mirai Botnet changed into located to be the perpetrator and this instance bowled over the complete commercial enterprise community which otherwise notion that safety in IoT devices become just secondary. So, IoT botnets are actually status second at the threats list. And Gartner

expects that around 8.four billion of things gets linked to the net on this year- possibly a variety of problem may be in keep in destiny.[11]

E. Phishing and whaling attacks

'Phishing' is a concept in which hackers ship fraudulent emails from depended on debts to goal corporations via individual group of workers individuals. while an innocent staff member clicks on the email, then attachment that's tagged to the email begins functioning freeing a malware capable of stealing records. 'Whaling ' takes the above stated cyber assault strategy to next degree with the aid of focused on excessive worth people, frequently CIOs or CEOs of a firm. FBI has warned all corporates operating inside and outside of America about this rip-off and showed that hackers have succeeded in making \$3 million from such fraudulent transactions last 12 months.[1]

F. Business Process Compromise Attacks

Trend Micro has defined this concept of cyber attack as a fairly new phenomenon where hackers are the use of techniques to manipulate the day after day operations of a business in their prefer. For instance, in the 12 months 2013 drug traffickers from South the USA controlled to intercept the community of an Antwerp to song the motion and region of bins. This helped the traffickers to retrieve the cargo at a secluded location before the naval police attempted to arbitrate their operations. So, in this case, hackers were applied to compromise the commercial enterprise procedure of a government firm to stay away from regulation enforcement forces and for monetary profits.

G. Machine Learning enabled attacks

It looks as if the technology of Artificial Intelligence appears to be serving each the best and terrible people. In line with a current Intel safety report, device learning is getting used to release social engineering assaults. Like, if hackers benefit get right of entry to to publicly available facts, they can use complicated evaluation gear to pick out targets more precisely and with a extra stage of achievement. As an instance, in UK, hackers are gaining access to databases associated with tax filing to launch ransomware related assaults on individuals who've filed for the highest IT returns. This proves that the facts available on public structures may be used to launch attacks on individuals for minting money.

VII. FUTURE WORK

Our old preconceptions about IT security and its relationship with modern society are melting away rapidly.The fear of computer security threats has so far surpassed the reality of Cyber security failures. The chief candidates for a catastrophic failure are world financial markets, military commands and control systems and critical infrastructure systems.The key to managing threats is understanding them- the key to understanding them is to find a way to map them against specific behaviors or events, the activities which help provide this definition and mapping represents the core of threat management.

Some threat predictions are the following:

Prediction 1: Denial of Service(DoS) will disappear as a mainstream threat by 2021.

Prediction 2: The world's first entirely cyber war will be fought by or before 2021. It will likely be undeclared and referred to as something else.

Prediction 3: Despite remarkable efforts and expense to secure global infrastructures, they will remain vulnerable. Cyber adversaries retain their lead.

VIII. CONCLUSION

Statistics and a lot of research study shows that data theft and abuse are becoming a profitable business worldwide. Perfect computer systems pose a significant barrier to illegal activities, yet there is always a chance to hack and misuse a system. Organizations together with ISO, IEC, OECD and IEE have consequently organized a huge variety of standards, guidelines and commands on how to enforce Information Security management, e.g.:

- a) ISO/IEC Guide 73: 2002 Risk management. Vocabulary. Guidelines for use in standards.
- b) ISO/IEC 13335-1: 2004 Information technology security techniques.
- c) ISO/IEC 27002: 2007 Information technology. Security techniques. Code of practice for information security management.
- d) Management of information and communications technology security. Part 1: Concepts and models for information and communications technology security management.
- e) ISO/IEC 15408-1: 1999 Information technology. Security techniques. Evaluation criteria for IT security. Part 1: Introduction and general model.
- f) ISO/IEC 15489-1: 2001 Information and documentation. Records management. Part 1: General.
- g) OECD Guidelines for the Security of Information Systems and Networks. Towards a Culture of Security. 2002 ISO/IEC TR 18044 Information Technology.
- h) Security Techniques. Information security incident management. [12]

REFERENCES

- [1] <https://its.ucsc.edu/security/training/intro.html>
- [2] <http://www.pearsonitcertification.com/articles/article.aspx?p=2218577&seqNum=3>
- [3] Computer Security by Matt Bishop
- [4] <https://www.eng.tau.ac.il/~yash/crypto-netsec/classical.htm>
- [5] <http://pajhome.org.uk/crypt/rsa/intro.html>
- [6] https://access.redhat.com/documentation/en-us/red_hat_enterprise_linux/6/html/managing_smart_cards/using_kerberos

[7] <https://www.geeksforgeeks.org/implementation-diffie-hellman-algorithm/>

[8] <http://accentsjournals.org/PaperDirectory/Journal/TIS/2016/1/5.pdf>

[9] <https://www.technewsworld.com/story/84973.html>

[10] <https://www.cnbc.com/2018/01/18/apples-latest-security-flaw-is-a-text-bomb.html>

[11] <https://us.norton.com/internetsecurity-how-to-top-ten-cyber-security-predictions-2017.html>

[12] <https://www.degruyter.com/downloadpdf/j/rput.2012.20.is-sue-sn/v10186-012-0019-0/v10186-012-0019-0.pdf>

Text Preprocessing and Polarity check on Social Media Data

Prajakta N. Samant, Ashuta K. Thakur, Harshali Patil
Computer Engineering, Thakur College of Engineering Mumbai, India

Abstract—With increasing ease of micro blogging use of social media has significantly increased in day to day life. People use Social Media platform like Instagram, Facebook, Twitter, for expressing views, opinion or emotions which can be their sentiments and opinions about particular thing. Expressing views or opinion known as sentiments and analyzing this sentiment referred as sentimental analysis. Polarity of sentiment can be categorized in three category positive sentiment, negative sentiment or neutral sentiment. Sentiment analysis is also Known as Opinion Mining. Opinion mining is useful for effective decision making. Corpus is being used for analyzing sentiment. Several machine learning techniques are being implemented like naïve Bayes, support vector machines, maximum entropy. However, data in social media is unstructured and required some preprocessing. Currently various techniques for polarity check have been developed but they differ in accuracy and performance. This paper focuses on various methods of preprocessing of social media data and survey of different technique used in sentiment analysis and understands the work level

Keywords—data preprocessing, sentiment analysis, opinion mining, social media.

I. INTRODUCTION

Huge amount of data is shared on Social media. People express views, opinions, knowledge and thoughts with the world by using Social Media like Facebook, twitter, Instagram. Private blogs, forums, review sites, tweets are mediums of microblogging. In microblogging people share their views for example movie reviews are shared on microblogs like ratings, comments, likes and dislikes. However, this review is unstructured and needs to be preprocessed before analyzing it. They utilize tweets finishing off with positive emojis like ":" "-)" as positive and negative emojis like ":(":-(" as negative. Reviews can be in the form of images and emojis. Sentiment is being extracted from User reviews and categorized as positive, negative or neutral.

Classification of sentiment would be helpful in business intelligence. Behaviors of reviewer could be judge through polarity. Opinion mining and sentiment analysis is based upon three approaches first is machine learning, second is lexicon and third is hybrid. Machine learning based approaches mainly use supervised learning, where a piece of text is compared with human developed list of sentiment bearing words. Second type of techniques are based on proper grammatical check on the text using various methods of Natural Language Processing. Hybrid techniques use combination of above mentioned. This paper is

discussed as follows: section 2 describes related work and approaches, section 3 focuses on Some challenges related to sentiment analysis fields. Finally, we conclude this paper in Section 4.

Below diagram is steps for sentiment analysis on social media data

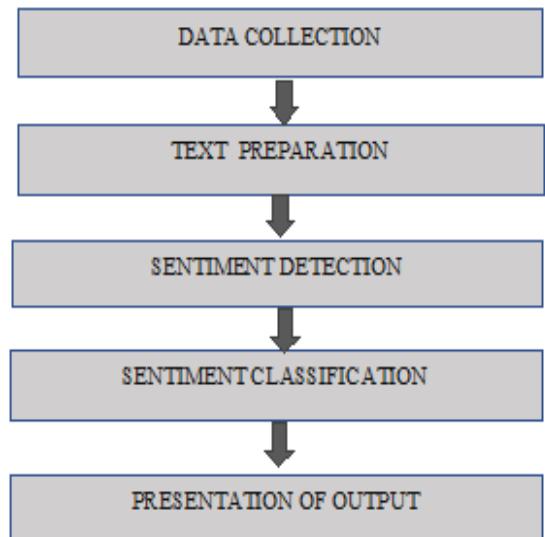


Fig 1: Steps for sentiment analysis

- Collecting data: social media content is being collected in first step. User expressed these data in different ways by using different vocabularies. Analyzing data manually is almost impossible. In this manner, to separate and characterize information content examination and characteristic dialect handling is utilized
- Preparation of content: comprises in cleaning the removed information before investigation. Non-printed substance and substance that are unimportant for the investigation are recognized and dispensed with.
- Detecting notion: the removed sentences of the audits and assessments are inspected. Sentences with subjective articulations (feelings, convictions and perspectives) are held and sentences with target correspondence (certainties, truthful data) are disposed of;
- classifying sentiment: characterizing sentences as positive, negative, great, awful; like, hate, yet various focuses can be considered in arrangement principle point is to change over disrupted content into significant data.

- output: main aim is to convert disorganized text into meaningful information. Final output is displayed in form of graphs like line graphs pie chart and, bar chart

II. LITERATURE REVIEW

Various models are proposed for sentiment analysis by researchers. In this paper we are discussing brief review about some of previously proposed system.

Authors of [1] proposed system in Weka an open source information mining apparatus has been utilized in order to perform supposition grouping on motion picture survey dataset. Here, objective is to arrange dataset into positive and negative and shape the consolidated lexicon of Twitter dataset and online survey dataset Combined expression of twitter dataset and online audit dataset frames a word reference. As in the wake of characterizing each word likelihood as positive, negative and nonpartisan. Think about the likelihood for each word and order each word into three distinct lexicons in view of most noteworthy extremity of each word. and categorize each word into three different dictionaries based on highest polarity of each word. we analyses the dataset based on accuracy given by naïve Bayes multinomial. Online review dataset accuracy around 94.968% and for twitter its around 82.695%. Results show that we get better accuracy for online review as compared to twitter tweets as online review are clearer and in detail compare to twitter tweets

According to [2] in sentiment analysis on social media images, Web pictures in light of both picture highlights and logical social media information. this paper provides a way of using both of them, and formulate sentiment analysis problem in two scenarios: supervised and unsupervised. In supervised method new framework proposed in this paper was named RSAI (Robust Sentiment Analysis for Images) which is efficient and effective. RSAI blends the low-level visual highlights to the recognized mid-level protests and guide them to a lexicon. RSAI concludes sentiment of image by factorizing an info picture highlights matrix into three factors corresponding to sentiment-features, image-term and term sentiment. If images are unlabeled then unsupervised method is used. Unsupervised assumption investigation for online networking pictures with literary data has two difficulties that are the means by which to utilize printed data to empower unsupervised assessment examination for web-based social networking pictures and how model systematically interact with image and textual information. For above challenges this paper proposed novel Unsupervised Sentiment Analysis (USEA)framework, which performs sentiment analysis for social media pictures in an unsupervised manner. These two frameworks provide better performance the baseline method.

Authors of [3] concentrated on retweeting Structure-mindful Approach for wistful examination which utilizes model that can mine conclusions concerning retweeting tree structures and retweeting remarks. The propose framework to mine opinions in microblog domain and then build a real-time analysis system to monitor the sentiment propagation. Opinions are abstracted as triples, based on an opinion mining method and opinion summary is generated. Author build a microblog-oriented sentiment lexicon and proposed a lexicon-based sentiment analysis algorithm to classify sentiments.

Dynamic administration piece component utilized for assessment investigation via web-based networking media data. Paper [4] proposed a quality model to survey nature of social data. (SAaaS) structure separate slants from social data and change into important data to convey as an administration. Previous approaches didn't consider different types and characteristics but SAaaS considers different properties such as data size, type. Author [4] proposed a paper on domain of disease surveillance, spatio-temporal properties and sentiment analysis is used to identify the locations of disease. This paper considered small set of data but real word social information is exponential so this is drawback of this paper.

Authors of [6] proposed paper on understanding engineering students' learning experience on Twitter posts. They initially directed a subjective investigation on tests taken from around 25,000 tweets related to building understudies' school life. They discovered building understudies experience issues, for example, inadequacy of social engagement, overwhelming investigation load, and rest lack. In view of these results, we completed a multi-stamp Naïve Bayes order calculation. Workflow to bridge and integrate a qualitative research methodology and large-scale data mining techniques are major contribution in this paper. This gives a work process to breaking down online networking information for instructive purposes that conquers the significant constraints of both expansive scale computational examination of client created literary substance and manual subjective investigation.

It is workable at the stock cost of a few organizations to be anticipated with a normal precision as high as 76.12%. They proposed a procedure to dig Twitter information for answers to the inquiries like if the cost of a determination of 30 organizations recorded in NASDAQ and the New York Stock Exchange can really be anticipated by the given 15 million records of Twitter messages [7]. Methods used in this paper are to predict stock market movement, propose an effective method based on LDA model to clean microblogs, and a user-group model to predict stock market movement.

Authors of [8] have shown that the sentiment analysis results produced by emotions of user. lexicon knowledge extraction algorithm is used for extracting domain knowledge. Implement it within an end-to-end social media analysis system. lexical-based methods have lack of fine-grained sensing capability.

III. TEXT PREPROCESSING METHOD

Preprocessing technique assumes a critical part in content mining and opinion investigation. It is the initial phase in the content mining process [5].

Need of Text Preprocessing in NLP System:

1. To lessen indexing (or information) record size of the Text reports
 - i) Stop words accounts 20-30% of aggregate word checks in a specific content archives
 - ii) Stemming may diminish ordering size as much as 40-half

2To enhance the proficiency and viability of the IR framework

i) Stop words are not helpful for looking or Text mining and they may confound the recovery framework

ii) Stemming utilized for coordinating the comparative words in a content archive

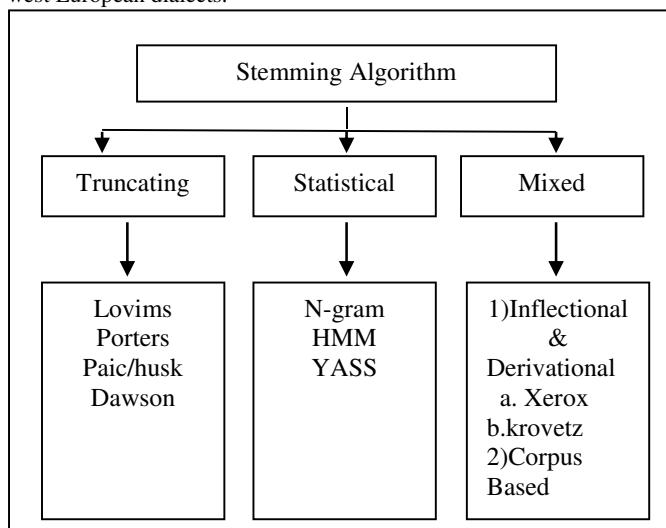
In this paper, we talk about the three key strides of preprocessing specifically, stop words expulsion, stemming and TF/IDF calculations

Table I: Preprocessing techniques

Preprocessing Technique	Applied option
Weighting scheme	TF-IDF
Stemming	Snowball stemmer
Stop word removal	Rainbow list
Tokenization	1. Unigram 2. Bigram 3. 1-to-3-gram
Feature selection	1.All 2.InfoGainAttributeEval/ Ranker-IG>0 3.ClassifierAttributeEval-Random Forest / Ranker - top 70%

TF-IDF weighting scheme: It is a standard approach to feature vector construction. TF-IDF stands for the “term frequency-inverse document frequency” and is a numerical measurement that reflects how imperative a word is to an archive in a corpus.

Stemming: Stemming is utilized to distinguish the foundation of a word. For example, the words computing, computed, computing, computations all can be stemmed to the word “compute”. The motivation behind this technique is to expel different postfixes, to decrease the quantity of words, to have precisely coordinating stems, to spare time and memory space. In stemming, interpretation of morphological types of a word to its stem is finished accepting everyone is semantically related. There are two focuses are considered while utilizing a stemmer: Words that don't have a similar significance ought to be kept separate. Morphological types of a word are expected to have a similar base importance and consequently it ought to be mapped to a similar stem. These two principles are great and adequate in content mining or dialect preparing applications. Stemming is normally considered as a review upgrading gadget. For dialects with generally basic morphology, the energy of stemming is not exactly for those with a more mind-boggling morphology. A large portion of the stemming tests done as such far are in English and other west European dialects.



Errors in Stemming: There are mainly two errors in stemming.

1. over stemming

2. under stemming

Over-stemming is when two words with various stems are stemmed to a similar root. This is generally called a false positive. Under-stemming is when two words that should be stemmed to a comparable root are unquestionably not. This is otherwise called a false negative.

A. Extraction This technique is utilized to tokenize the document content into singular word.

B. Stop Words Elimination Stop Words Elimination Stop words are a division of regular dialect. The intention that stop-words ought to be expelled from a content is that they influence the content to look heavier and less essential for experts. The most widely recognized words in content reports are articles, relational words, and star things, and so on that does not give the importance of the records. These words are dealt with as stop words. Case for stop words: the, in, an, a, with, and so on. Stop words are required to remove from documents because these words are not measured as keywords in text mining applications [4].

C. Stop word removal methods Four types of stop word removal methods are followed, the methods are used to remove stop words from the files [4].

1. The Classic Method: The great strategy depends on expelling prevent words acquired from pre-arranged records.

2. Strategies in light of Zipf's Law (Z-Methods): notwithstanding the great stop list, we utilize three stop word creation techniques moved by Zipf's law, including: expelling most regular words (TF-High) and evacuating words that happen once, i.e. singleton words (TF1) [4]. We likewise consider expelling words with low backwards archive recurrence (IDF).

3. The Mutual Information Method (MI) The common data technique (MI) is a regulated strategy that works by processing the shared data between a given term and a record class (e.g., positive, negative), giving a recommendation of how much data the term can tell about a given class. Low shared data proposes that the term has a low separation control and thusly it ought to be expelled from content minningapplication.

4. Term Based Random Sampling (TBRS) This technique was first proposed by Lo et al. (2005) to physically distinguish the prevent words from web records. This strategy works by repeating over isolated pieces of information which are haphazardly chosen. It at that point positions terms in each lump in light of them in-arrange esteems utilizing the Kullback-Leibler uniqueness measure as appeared in Equation $dx(t) = Px(t).log2 Px(t) ??(t)$ (1) Where $Px(t)$ is the standardized term recurrence of a term t inside a mass x , and $P(t)$ is the standardized term recurrence of t in the whole accumulation. The final stop list is then constructed by taking the least informative terms in all chunks, removing all possible duplications

Tokenization: This setting splits the documents into words/terms, constructing a word vector, known as bag-of-words. We propose NGramTokenizer to compare word unigram, bigram and 1-to-3-gram

Feature selection: It is a process by which the number of attributes is decreased into a better subset which can bring highest accuracy. The benefits of performing this option on the data are the limitation of over fitting, the improvement of accuracy and the reduction in training [4].

Challenges in Tokenization Challenges in tokenization rely upon the kind of dialect. Dialects, for example, English and French are alluded to as spacedelimited as an extensive segment of the words are isolated from each other by blank areas.Dialects, for example, Chinese and Thai are alluded to as unsegmented as words don't have clear limits. Tokenizing unsegmented dialect sentences requires extra lexical and morphological data. Tokenization is likewise influenced by composing framework and the typographical structure of the words. Structure of dialects can be assembled into three classifications:

- Isolating: Words don't partition into littler units.
Illustration: Mandarin Chinese
- Agglutinative: Words partition into littler units.
Illustration: Japanese, Tamil
- Inflectional: Boundaries between morphemes are not clear and uncertain regarding syntactic significance.
Case: Latin.

E. Convert text to lowercase This strategy is utilized to stay away from recognize words essentially on case.

F Remove Number Numbers might possibly be pertinent to our investigations. Normally it doesn't convey any significance in supposition examination

G. Remove PunctuationExpel Punctuation can give syntactic setting which bolsters understanding. For pack of words-based slant examination accentuation does not include esteem. For instance: " . , , ? "

H. Lemmatisation – change to word reference base frame i.e., "produce" & "produced" become "produce"

I. Sparse terms – We are regularly not keen on occasional terms in our records. Such "scanty" terms ought to be expelled from the record term network.

IV. POLARITYCHECKING TECHNIQUES AND MODELS

A. Polarity checking techniques

1. Sentiment analysis based SENTIWORDNET

The approach portrayed in this paper depends on Senti WordNet, a lexical asset for conclusion mining. In Senti WordNet, to every synset of WordNet, a triple of extremity scores is relegated i.e., an energy, cynicism and objectivity score. The total of these scores is dependably 1. For instance, the triple {0, 1, 0} (energy, cynicism, objectivity) is appointed to the synset of the expression "terrible". The entirety of all scores of this synset is 1. SentiWordNet has been made consequently by methods for a mix of semantic and measurement classifiers. It has been connected in various feeling related errands, i.e. for assessment investigation with promising outcomes.

2. Bigrams

Bigrams are utilized as a part of request to build the exactness of the classifier.

The impact of past word on current word assumes real part in opinion examination thus we consider bigrams as opposed to unigrams. When all is said in done, going before word will indicate more impact on the present word as opposed to the succeeding word subsequently we think about the extremity of going before word. For instance, consider the sentence "The workmanship demonstrates the way of life and social issues

winning by then of time."Bigrams should be possible as takes after "The workmanship", "craftsmanship appears", "demonstrates the", "way of life and", "and social", "social issues", "issues winning", "winning at"and so on

3. SES algorithm for preprocessed data to calculate sentiment

Step 1:Assign a weight to each word from the SentiWordNet lexicon **Step 2:**Sentence level extremity is figured as consider the sentences to ascertain the normal score

Step 3:check (sent_sentim_word + 3) and (sent_sentim_word - 3) for Modifier from modifier_dict if word found as modifier at that point figure general weight.

Step 4:If there is invalidation word (Not, Never, N't, Doesn't, Cannot, Nor, Don't, Wouldn't, No) close to the N, Check (N+3) and (N-3) at that point turn around its extremity. e.g. (OW=+0.8 ? OM= - 0.8)

Step 5:Check the modifier word in the sentence, if exists at that point recalculate the extremity alluding the weightage lexicon a similar procedure will be rehashed that score of which assessment word will be influenced. For e.g., in the sentence the sentence "the staff were extremely pleasant and agreeable", in this sentence the very is upgrade the heaviness of the closest assessment word i.e., decent

Step 6: Certain sort of things influence the sentence extremity, so recalculate the extremity if such kinds of word happen.

A. Polarity checking FEATURE MODELS

1. WORD N-GRAMS FEATURES MODEL

Word n-grams highlights are the least difficult element for Twitter estimation examination. Scientists report cutting edge execution for opinion examination on Twitter information utilizing a unigram demonstrate. In this work, word unigram and bigram highlight (alluded to N-grams display) is one of the element models.

2. PRIOR POLARITY SCORE FEATURE MODEL

An earlier extremity score is a vocabulary-based supposition highlight, and some methodologies utilize it as a conclusion include for tweet feeling examination. We utilized theAFINN4 dictionary and extended it utilizing Senti-Wordnet to get the earlier extremity score of the tweet.The earlier extremity score of a tweet is the sum of the sentiment score of each word in the tweet The sentiment score of each word is computed by measuring the PMI ,PMI stands for point-wise mutual information between the term and the positive or negative category of the tweet using the formula: Sen Score(w)=PMI(w, pos)-PMI(w, neg) Where is a term in the lexicon, PMI(w, pos) is the PMI score between w and the positive class, and PMI (w, neg) is The PMI score between w and the negative class. Therefore, a positiveSenScore(w) recommends a more grounded relationship of word with positive slant and the other way around. In this work, the prior polarity score feature (referred to Prior polarity model) is another feature model.

3. TWITTER SENTIMENT CLASSIFIERS

To assess the effect of pre-processing on sentiment classification, we used four popular supervised classifiers in the literature of sentiment analysis, Support Vector Machine (SVM, parameter c is 100, gamma is 0.5, kernel is linear, other parameters are the default values), Naive Bayes (NB), Logistic Regression (LR, default parameters), and Random Forest (RF, parameter max_depth is 30, estimators is 4000, other parameters are set to the default values). This paper uses the Grid Search search for these

parameters as the optimal parameters and uses scikit-learn library to perform the classifier.

4. BASELINE AND EVALUATION CRITERIA

Two classification tasks are performed: a parallel assignment of grouping opinion into positive and negative classes and a 3-waytask of ordering assumption into positive, negative, and impartial classes. The binary task is performed on all five datasets and the 3-way task is performed on four datasets using SVM, NB,

V. COMPARATIVE ANALYSIS

Table II: Techniques for Sentiment Analysis

R*	Approach	Tools/Techniques	Experiment	Language Dependency	M/C	Data Scope	Data Source
1	Online Review Opinion predication (2015)	Weka (open source data mining tool	sentiment classification on movie review dataset	Yes	ML*	Movie reviews	IMDB
2	<i>RSAI</i> (Robust Sentiment Analysis for Images) Framework analysis (2015)	Supervised learning, unsupervised learning K means	Analyzing sentiments on image	No	LB*	Instagram	Instagram Post
3	sentiment classification	opinion mining method	Opinion mining for tweets	Yes	LB*	Twitter	Tweets
4	Dynamic Service Composition Approach (2017)	Sentiment Analysis as a Service	To extract sentiment from multiple social information service and transform into useful information	Yes	LB*	Social networking sites	Social media data
5	Sales prediction (2012)	Sentiment PLSA (S-PLSA ARSQA, an Autoregressive Sentiment and Quality Aware model	To Predict Sales Performance	Yes	ML*	Movie reviews	IMDB
6	Qualitative analysis and large-scale data	Naïve-Bayes multi-label classification algorithm	Social Media Data provide opportunities for student to understand user learning experience	Yes	ML*	Facebook	Post
7	Predicting Stock Market Movement (2014)	NLP techniques	Predicting the movement in Stock Market based on analysis	Yes	ML*	Twitter	Tweets

The effect of text pre-processing is evaluated by the gain or loss of accuracy and F1- measure. In the binary task, the F1-measure is the average of positive and negative classification F1-measure. In the 3-way task, the F1-measure is average of the positive, neutral,

LR, and RF classifiers. The baseline method is the classic method (C-Method) respectively using the N-grams and the Prior polarity model, which was applied all six pre-processing methods, including removing URLs, removing stop words, removing numbers, reverting words that contain repeated letters to their original form, replacing negative mentions, and expanding acronyms to original word. The accuracy and F1-measure are utilized to assess the overall sentiment classification performance.

and negative classification F1-measure. We use 10-fold cross validation, which is a technique that is useful to evaluate a classification algorithm for a given corpora, splitting and evaluating the training set several times.

VI. CONCLUSION

Social Media is wide platform for tracking Sentiment Analysis where large number of user share their feelings. Social media is one of the greatest stages where vast number of messages are distributed each day which makes it a perfect hotspot for catching the feelings towards different Realtime topics, such as politics, products, goods or celebrities, etc. The main goal of this paper is to give an overview of updates in sentiment analysis and various methods of preprocessing of social media data.

VII. APPLICATION

A. Risk Management

Polarity Checking Technique can significantly expand the capacity to relieve chance, empowering complete administration of thousands of sources and petabytes of content reports, and giving the capacity to connect together data and have the capacity to get to the correct data at the perfect time.

B. Customer Care Service

Content investigation is utilized to give a fast, mechanized reaction to the client, drastically decreasing their dependence accessible as needs be focus administrators to tackle issues.

C. Fraud Detection Through Claim Investigation

Fraud recognition through cases examination Insurance organizations are exploiting content mining advancements by consolidating the aftereffects of content investigation with organized information to avoid fakes and quickly process claims.

D. Contextual Advertising

Computerized promoting is a modestly new and developing field of utilization for content investigation. Here, companies such as [Admantx](#) have made text mining the core engine for contextual retargeting with great success

E. Business Intelligence

This procedure is utilized by expansive organizations to maintain and bolster basic leadership. Here, content mining truly has the effect, empowering the investigator to rapidly seize the ANSWER EVEN while breaking down petabytes of interior and open source information.

F. Social Media Data Analysis

Online networking information examination Today, online person to person communication is a champion among the most productive wellsprings of unstructured information; associations

have paid heed. Online networking is progressively being perceived as an important wellspring of market and client knowledge, and organizations are utilizing it to break down or foresee client needs and comprehend the view of their image.

VIII. FUTURE SCOPE

Researchers can recognize better approaches for grouping the literary information into different inclinations, for example, bliss, sharpness, assurance, and smoothness. Many more abbreviation for particular text are being used for which polarity check algorithm cab be more developed more efficiently.

REFERENCES

- [1] Rajni Singh, and Rajdeep Kaur, "Sentiment analysis on social media and online review," International Journal of Computer Applications, Volume 121 – No.20 July 2015.
- [2] Yilin Wang and Baoxin Li, Sentiment analysis on social media Images, IEEE 15th International Conference on Data Mining Workshops, 2015.
- [3] Lu Lin, Jianxin Li, Richong Zhang, Weiren Yu and Chenggen Sun, "Opinion Mining and Sentiment Analysis in Social Networks: A Retweeting Structure-aware Approach", IEEE/ACM 7th International Conference on Utility and Cloud Computing, 2014.
- [4] AkriwiKrouskas, Christos Troussas, Maria Virvou, "The effect of preprocessing techniques on Twitter Sentiment Analysis," IEEE Information, Intelligence, Systems & Applications (IISA), 2016 7th International Conference on, 2016.
- [5] V. Srividhya, R. Anitha, "Evaluating Preprocessing Techniques in Text Categorization," International Journal of Computer Science and Application Issue 2010.
- [6] Emma Haddia, XiaohuiLiua , Yong Shib, "The Role of Text Pre-processing in Sentiment Analysis," 2013 International Conference on Information Technology and Quantitative Management.
- [7] Kashif Ali1, Hai Dong, AthmanBouguettaya, AbdelkarimErradi, Rachid Hadjidj, "Sentiment Analysis as a Service: A social media-based sentiment analysis framework," IEEE 24th International Conference on Web Services, 2017.
- [8] Shulong Tan, Yang Li, Huan Sun, Ziyu Guan, Xifeng Yan, JiajunBu, Chun Chen, and Xiaofei He, "Interpreting the Public Sentiment Variations on Twitter", IEEE trans. on knowledge and data engineering, vol. 26, no. 5, May 2014.
- [9] Xin Chen, MihaelaVorvoreanu, and Krishna Madhavan, "Mining Social Media Data for Understanding Students' Learning Experiences" IEEE trans. on learning technologies, vol. 7, no. 3, July-September 2014.
- [10] LI Bing, Keith C.C. Chan, Carol OU, "Public Sentiment Analysis in Twitter Data for Prediction of a Company's Stock Price Movements", 2014 IEEE, 11th International Conference on e-Business Engineering

Database Access for Non-Technical Users Using NLP and Voice Recognition

Rashmi Thakur¹, Aishwarya Manjalkar², Anuradha Maurya²

Assistant Professor¹, M.E. Scholar²

*Dept. of Computer Science Engineering, Thakur College of Engineering and Technology
thakurrashmik@gmail.com, aishwarya.manjalkar.am@gmail.com, anuradha.maurya102@gmail.com*

Abstract - The field of “Natural Language Processing” (NLP) has seen a dramatic shift in both research direction and methodology in the past several years. In the past, most work in computational linguistics tended to focus on purely symbolic methods. Recently, more and more work is shifting towards hybrid methods that combine new empirical corpus-based methods, including the use of probabilistic and information theoretic techniques, with traditional symbolic methods. The main purpose of Natural Language Query Processing is for an input sentence(s) to be interpreted by the computer and appropriate action taken; asking questions to databases in natural language is very convenient and easy method of data access, especially for casual users who do not understand complicated databases such as SQL. This project proposes the architecture for translating input sentence(s) into SQL query using Semantic Grammar with added feature of Voice Recognition.

Keywords - NLP, NLAP, natural language processing, transliteration, voice recognition, speech recognition

I. INTRODUCTION

GUI plays an instrumental role in the success of any browser. GUI enables user control and improves interaction with the browser. GUI is considered a part of the user trust model for all types of software including browsers. GUI in browsers includes notification bars, status bars, address bars, download dialog boxes, HTTP authentication dialog, and browser objects such as frames, buttons, etc. Users interact with the GUI components in their routine life jobs. GUI flaws are considered design errors in which an attacker can circumvent the normal functioning of the browser by running malicious JavaScript. Primarily, GUI bugs in browsers are mostly exploited by spoofing [1], clickjacking, likejacking, password manager attacks. Spoofing attacks are those kinds of attacks that tamper the UI component of software in order to fool users into performing false operations by exploiting their ignorance. Clickjacking attacks are under the categorization of UI redressing attacks in which an attacker embeds a hidden UI object such as buttons, frames, etc. For example, an attacker can easily place a hidden button over the real button in a browser window that executes a malicious function when a user clicks it. Likejacking attacks fall into the UI redressing attacks in which an attacker embeds a hidden UI object in the form of likes, dislikes, add to lists options etc., which on-click to this can execute some malicious functions. Password manager attacks can consist of SQL injections on the look-alike GUI website used with different server connectivity. This can lead to the hacking of sensitive password of the sensitive sites like e-banking, industrial and educational login system, mail login password, important pins and other information.

Such attacks can take place through different servers’ connectives’, which are hard to recognize in a look-alike GUI. Therefore, this gives information about various attacks which are not easy to recognize in GUI.

II. PROBLEM STATEMENT

A. A rule-based approach for NLP based query processing.

Information is playing an important role in our lives. One of the major sources of information is databases. Databases and database technology are having major impact on the growing use of computers. In order to retrieve information from a database, one needs to formulate a query in such way that the computer will understand and produce the desired output. Generally, query processing is handled by the Structured Query Language (SQL). But the non-IT people cannot be able to write SQL queries as they may not be aware of the SQL as well as structure of the database.

1. ADVANTAGES

- When time is a crucial factor, machine translation can save the day. You don't have to spend hours poring over dictionaries to translate the words. Instead, the software can translate the content quickly and provide a quality output to the user in no time at all.
- Confidentiality is another matter which makes machine translation favorable. Giving sensitive data to translator might be risky while with machine translation your information is protected.

2. DISADVANTAGES

- Accuracy is not offered by the machine translation on a consistent basis. You can get the documents, but machine translation only does word to word translation without comprehending the information which might have to be correct manually latter on.
- Systematic and formal rules are followed by machine translation, so it cannot concentrate on a context solve ambiguity and neither makes use of experience or mental outlook like a human translator can

B. Improved voice activity detection for speech recognition system.

An improved voice activity detection (VAD) based on the radial basis function neural network (RBF NN) and continuous wavelet transform (CWT) for speech recognition system is presented in the paper. The input speech signal is analyzed in the form of fixed size window by using Mel-frequency spectral coefficients (MFCC). Within the windowed signal, the proposed RBF-CWT VAD algorithm detects the speech/ non-speech signal using the RBF NN.

1. ADVANTAGES

- There is no training.
- Relieves burden of learning syntax.

2. DISADVANTAGES

- Requires clarification dialogue
- May require more keystrokes.
- May not show context.
- It is unpredictable.

C. Developing a system for machine translation from Hindi language to English language.

Many research organizations in India and abroad have started developing translation systems for the Indian languages recently using conventional approaches like ruled-based or exemplified-based or hybrid. Very few have tried to identify universality of government and binding (GB) theory, which emphasizes common phrase structure for all the languages. In this paper, a machine translation system based on ruled-based theory is proposed. The system takes Hindi as source language and English as target language.

1. ADVANTAGES

- Increase productivity.
- Can help with menial computer tasks, such as browsing and scrolling.
- Can help people who have trouble using their hands.
- Can help people who have cognitive disabilities.
- Has long term benefits for students.

2. DISADVANTAGES

- Can be hacked with pre-recorded verbal messages.
- Has an initial period of adjusting to each user's voice?
- Less accurate when there is background noise.
- Can be distracting in a cubicle environment.

III. SYSTEM DESCRIPTION

Following is the process followed for Library Management:

- A user enters a simple English sentence.
- This sentence is broken down into individual words, called tokens.
- Each token is assigned a unique token ID.
- All the useless tokens, which are not required for forming a query, are discarded.
- Various combinations of useful tokens are made.
- An appropriate query is formed and fired in the database
- Following is the process followed for Library Management:
- A user enters a simple English sentence.
- This sentence is broken down into individual words, called tokens.
- Each token is assigned a unique token ID.
- All the useless tokens, which are not required for forming a query, are discarded.
- Various combinations of useful tokens are made.
- An appropriate query is formed and fired in the database.

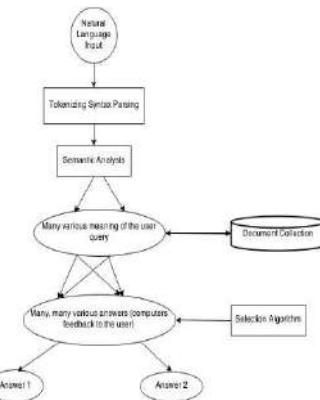


Fig.1: Stages in Natural Language Interpretation Process

The tuples used in the database are shown below:

- Book_ID.
- Book_Name.
- Publication.
- Book_Author.
- Book_Price.
- Copies_Available

IV. IMPLEMENTATION DETAIL

A. Method Used

1. MORPHOLOGICAL ANALYSIS:

The given sentence is split into tokens and assigned the respective types with the help of the dictionary. Each word is assigned its respective type such as noun; verb etc. with the help of the data dictionary that has been created and populated with domain specific and commonly used words. Individual words are analyzed into their components and non-word tokens such as punctuations are separated from the words.

2. SYNTACTIC ANALYSIS:

The Syntax analyzer checks the lexically correct input for grammatical correctness. The keywords are extracted and passed on to either the query analyzer. The output of the Syntax analyzer will be whether the given input is syntactically or grammatically correct as per the given rules. This process is also called as Parsing. English grammar rules are used for checking the syntax. The Syntax analyzer has been formed based on the grammar for English language [3].

3. SEMANTIC ANALYSIS:

Parsing only verifies that the program consists of tokens arranged in a grammatically valid combination. The main intention is to get the meaning of the sentence.

4. TRANSLATION:

After the semantic analysis SQL query is generated by using Semantic Grammar. After the creation of SQL query this query is applied on database to retrieve the data. The SQL statement and the NLP statement to access the database would result in the same output the only difference being, a normal person who doesn't know anything about SQL can easily access the database.

5. SEMANTIC GRAMMAR:

Systematic and formal rules are followed by machine translation, so it cannot concentrate on a context solve ambiguity and neither makes use of experience or mental outlook like a human translator can

There are two main parts of a semantic grammar.

The first is a lexicon that stores all the possible words the grammar is aware of. A simple entry in the lexicon might look like this:-

(customer → customer patron member)

(customers → customers patrons members)

The other part of the semantic grammar involves rules to combine the terminal symbols in the lexicon to form phrases or sentences in a specific way. For example, the rule (AIT_TAPE_CUSTOMER → RENTED by customer

AIT_Number)^[1]

B. Implementation Logic Used

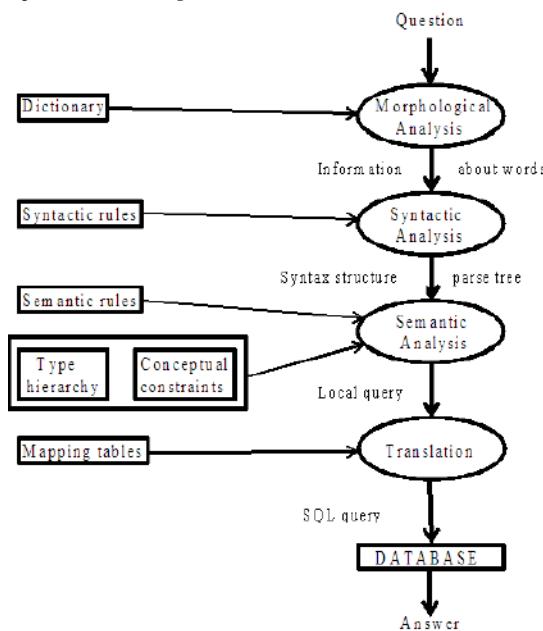


Fig.2: Implementation of NLQP

The Above NLQP is designed for Library Management System. The working of the entire system can be explained as follows: A Librarian in Library types a Question to NLQP engine in a Natural Human English Language. This question gets fed into NLQP engine is broken down in its First Stage

1. MORPHOLOGICAL ANALYSIS (TOKENIZATION)

The given input sentence is split into tokens and assigned the respective types with the help of the dictionary.

2. SYNTACTIC ANALYSIS

The Syntax analyzer checks the lexically (relating to words) correct input for grammatical correctness.

3. SEMANTIC ANALYSIS

Parsing only verifies that the program consists of tokens arranged in a grammatically valid combination. The main intention is to get the meaning of the sentence.

4. TRANSLATION

After the semantic analysis SQL query is generated by using Semantic Grammar. After the creation of SQL query this query is applied on database to retrieve the data.

5. VOICE RECOGNITION

It is a biometric technology used to identify a particular individual's voice. Using techniques such as Mel Frequency Cepstrum Coefficients (MFCC) Vector, Quantization and Hidden Markov Model, we are able to convert the given speech to text

6. SPEECH TO TEXT

The system will identify the words in spoken language and will translate it into text (Query).

7. LOCAL LANGUAGE TO ENGLISH CONVERSION

The system will identify local language (words) from dictionary and convert into English sentences or words the scroll down window on the left of the MS Word Formatting toolbar.

V. RESULTS

The first screenshot shows the text input. A user enters input in simple English sentence, which, on clicking “GET”, is fired in the database. We get an appropriate result as follows

NLQP FOR LIBRARY MANAGEMENT SYSTEM						
Enter Statement						
Generated Query						
Book_ID	Book_Name	Publition	Book_Author	Book_Price	Category	Author
1	Infinite Theory	Wiley	Jacob Abraham	450	3	
2	Big Data	Wiley	John Maclean	350	5	
3	Digit Image Pro	Harper Collins	Garrison Keillor	250	10	
4	AI	Wiley	David J. Deamer	200	20	
5	Fundamentals of Physics	Pearson	Lorraine Randall	140	18	
6	Electronics	Prentice Hall	Dennis Roddy	470	21	
7	Information Tech.	Wiley	Jack T. Matherne	540	18	
8	Computer	Wiley	John G. Kemeny	200	20	
9	Natural Language	Prentice Hall	Steve Asbell	450	24	
10	Wireless & Mobile	Pearson	David J. Deamer	720	18	
11	ECommerce Fun	Wiley	Henry Gair	350	15	

Fig.3: Text input: “show me all book”

NLQP FOR LIBRARY MANAGEMENT SYSTEM						
Enter Statement						
Generated Query						
Book_Author	Author					
1	Jacob Abraham					
2	John Maclean					
3	Garrison Keillor					
4	David J. Deamer					
5	Lorraine Randall					
6	Dennis Roddy					
7	Jack T. Matherne					
8	John G. Kemeny					
9	Steve Asbell					
10	David J. Deamer					
11	Henry Gair					

Fig.4: Speech input: “show all author”

The above screenshot shows a speech input. A user enters a simple sentence with the help of a microphone. This sentence is taken as an input and fired in the database. Thus, we get the following results

VI. DESIGN

A. Main Input Form



Fig.5: Main Input Form

B. Typed Input Form



Fig.6: Typed Input Form

C. Speech Input Form



Fig.7:Speech Input Form

D. Output Form



Fig.8: Output Form

VII. SCOPE OF THE SYSTEM

The scope of the proposed system is as follows:

- Working with DBMS and RDBMS related applications require basic knowledge of MS-SQL, Oracle, etc.
- User enters input in simple English language.
- Regional languages like Hindi and Marathi can also be used.
- Voice recognition can be used for interacting with the system.

A limited data dictionary is used for particular words; which must be regularly updated.

VIII. CONCLUSION

The system “Natural Language Processing” for SQL will be able to facilitate both the users and software engineers in terms of generating SQL queries automatically. The task of novel user will be simplified by providing an easy interface which will be more familiar and well known to the user.

The following things will be accomplished:

- User writes the requirements in simple English in a few statements and the system will have the ability to analyze the given script.
- For disabled users, we have provided a facility of speech input.
- After the analysis and mining of associated information, the system generates the intended SQL queries that will run directly.
- The system will have the robust ability to create code automatically without the external environment.

The system will provide a quick and reliable way to generate SQL queries to save time and money of both the user and system analyst

IX. FUTURE WORK

The proposed project can be used for:

- For building parameterized applications for systems with variable inputs.
- To implement NLQP for online access of data.
- Automation (home/office, etc.).
- Diseases recognition.
- Robotics.

REFERENCES

- [1] GauriRao, Dr. S. H. Patil, “NATURAL LANGUAGE QUERY PROCESSING,” in International Journal of Advances in Engineering & Technology, vol 2, Maharashtra, 2009, pp. 01-06.
- [2] Arati K. Deshpande, Prakash. R. Devale, “NATURAL LANGUAGE QUERY PROCESSING USING PROBABILISTIC CONTEXT FREE GRAMMAR”, in International Journal of Advances in Engineering & Technology, vol 3, Pune, May 2012, pp. 568-573.
- [3] Rukshan Alexander, PrashanthiRukshan, and Sinnatham by Mahesan, “NATURAL LANGUAGE WEB INTERFACE FOR DATABASE (NLWIDB)”, in Proceedings of the Third International Symposium, Sri Lanka, July 2013, pp. 01-08.
- [4] F.Siasardjahantighi, M.Norouzifard, S.H.Davarpanah, M.H.Shenassa, “USING NATURAL LANGUAGE PROCESSING IN ORDER TO CREATE SQL QUERIES”, inProceedings of the International Conference on Computer and Communication Engineering, Kuala Lumpur, May 2008, pp. 600-604.
- [5] P H Rathod, M L Dhore, R M Dhore, “HINDI AND MARATHI TO ENGLISH MACHINE TRANSLITERATION USING SVM”, in International Journal on Natural Language Computing, vol 2, Pune, August 2013, pp. 55-71.
- [6] Kashyap Patel, R.K. Prasad, “SPEECH RECOGNITION AND VERIFICATION USING MFCC & VQ”, in International Journal of Emerging Science and Engineering, vol 1, Pune, May 2013, pp. 33-37.
- [7] Suma Swamy and K.V Ramakrishnan, “AN EFFICIENT SPEECH RECOGNITION SYSTEM”, in Computer Science & Engineering: An International Journal, vol 3, Chennai, August 2013, pp. 21-27

Multimedia Recoloring Technique for Protanopic CVD

Nazneen A. Pendhari

*Thakur college of Engineering & Technology
Kandivali (E)*

mailnazneen@gmail.com, rr.sedamkar@thakureducation.org

Abstract-Color Blindness is a result of absence or fault in one of the 3 cones in the human eye or it is hereditary, Because of this deficiency the Color-Blind Users also called as Color Vision Deficient(CVD) face many problem in day to day life like they cannot understand traffic signal, road maps consisting of various colors, electrical wires etc. Since now everything is almost digital and the information is exchanged through the digital media a lot, the CVD are less fortunate to understand this information. The RGB to LMS algorithm is used to recolor those pixels in images and videos which are unperceivable to the users. After recoloring they will be able to distinguish the colors and it will be more understandable. We have focused on the Protanopia (Dichromacy) i.e. Red Green Color deficiency. It is fast and it maintains color consistency among frames.

Keywords-CB (Color Blind), CBP (colorblind Perceivable), CBU (Color Blind Unperceivable), color vision deficiency (CVD), Long Medium Short(LMS).

I. INTRODUCTION

Color blindness is a deficiency of unrecognizing certain colors. The person cannot distinguish between certain colors such as red, green and sometimes blue. A human can perceive colors because of cones and rods present in the eye. The cones are responsible for identifying colors in high light levels and the rods are responsible for vision in low light levels[1].

There are 3 types of cones in the eye: short wavelength (Red color), Medium Wavelength (Green color) and Long Wavelength(Blue). The brain accepts the input through these cones for color perception[2]. The color blindness is caused due to the absence or defect in one of the cones or it can be hereditary. There is different form of colorblindness, some people can perceive colors normally in bright lights while they have problem in dim light and in other cases they are unable to perceive in any light[3].

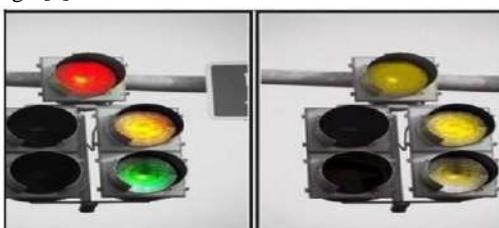


Figure 1 Traffic signal perceived by normal user and

Colour Blind User

There are 3 forms of colour Blindness viz. Monochromacy – This is complete colour blindness, here the person can see all the objects in grey scale i.e. black and white shades. This is due to malfunctioning or absent of cones.

Dichromacy- This involves person having any 2 cones in eyes and there is absence of one cone completely. There are 3 types of dichromacy

Protanopia- In this case, the cone that is sensitive to red colour is absent completely. Hence, this is also called as “Red weakness”. This type of colour blindness affects 1% of men and 0.02% of women.

Deuteranopia – This is also called as “Green weakness”, here the cone sensitive to green light are absent. It affects 1% of men and 0.01% of women.

Tritanopia- This is also called as “Blue Yellow deficiency”, here the cone that is sensitive to blue light is absent. It affects about 0.002% of men and 0.001% of women.

Anomalous Trichromacy- Here any 1 cone is not completely absent but they malfunction. The eyes use 3 cones in different proportions from the normal person. Hence, they have difficulty in discriminating and detecting similar shades. This is also of 3 types
Protanomaly- Malfunctioning in red cones. Deuteranomaly – Malfunctioning in green cones.
Tritanomaly- Malfunctioning in blue cones

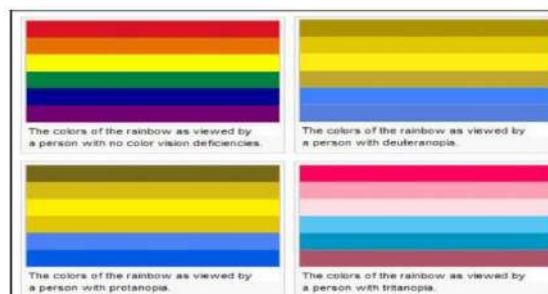


Figure 2 The Colour Band seen by different type of CVDs
faulty cones but sometimes by a fault in the pathway from the cone to the brain. People with normal color vision have all three types of cone/pathway working correctly but color blindness occurs when one or more of the cone types are faulty. For example, if the red cone is faulty you won't be able to see colors

containing red clearly. Most people with color blindness can't distinguish certain shades of red and green. Color blindness is a usually a genetic (hereditary) condition (you are born with it). Red/green and blue color blindness is usually passed down from your parents. The gene which is responsible for the condition is carried on the X chromosome and this is the reason why many more men are affected than women.

II. PROPOSED SYSTEM

Requirements of the proposed algorithm are:

- 1) The difference between the CBU and CBP colors increases by improving the color perceptibility of the image.
 - 2) The TCC has to be maintained.
 - 3) To enhance the local contrast in the region of the CBU color.
- The exact physical causes of color blindness are still being researched but it is believed that color blindness is usually caused by

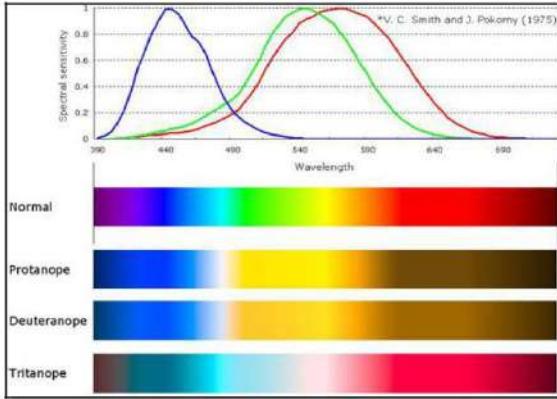


Figure 3 Comparative View of RGB Wavelength of normal and CVD user

The working of the proposed method is as follows:

A. CBU Color Detection

CBU color detection is the process of selecting the colors to be modified[5]. The most involuntary approach for CBU color detection is to select the colors which have large difference between the original color and its CBS color. However, it is difficult to maintain the TCC because the color to be converted varies at each frame[6]. We proposed a new CBU color detector. To prevent the color to be reversed, we change the reddish color extraction function $\square(X^0)$ as

$$\square(X^0) = 1, \text{ if } R^0 > G^0 \\ 0, \text{ otherwise}$$

The proposed method yield the clearer boundary in the CBU color region.

B. Color Mapping Strategy Selection

At each frame, we select one of two strategies for color mapping according to the temporal color difference (DT). Let DT be the pixel-value distance between the pixel to be converted for color mapping and the pixel

in the previous frame at the same image coordinate, C^0 . Then,

$$DT = \omega(Xt^0, Xt^{-1})$$

Where Xt^0 and Xt^{-1} are respectively, the color vectors at the same image coordinate, C^0 , in the current and previous frames. Based on DT , a color mapping strategy (CMS) is selected as follows:

$$CMS = CMSs, \text{ if } DT > \tau_2$$

$$CMS_t, \text{ otherwise}$$

If $DT > \tau_2$ for large motion, color rotation is performed by using the spatial color mapping strategy ($CMSs$). Or, if $DT \leq \tau_2$ for small motion, the temporal color mapping strategy (CMS_t) performs fast color mapping by using the mapping result of the previous frame.

C. The Spatial Color Mapping Strategy

In the $CMSs$ region, each CBU color, X^0 , is updated to X^0 by color rotation with angle $\Delta\theta$. In the proposed method $\Delta\theta$ is defined as a sum of two angles

$$\Delta\theta = \theta^G + \theta^L$$

Where θ^G and θ^L are, respectively, the global rotational angle and the local rotational angle. Here, we utilize CS to assign different θ^L and add it to the angle for rotating the CBU color. Thus, the contrast can be enhanced not only at the border of CBU region but also inside the CBU region. As a result, we obtain θ^L at the pixel position C^0 as

$$\theta^L = GCSP(Bt, C^0)$$

where is a scale factor to satisfy $0 \leq \Delta\theta = \theta^G + \theta^L \leq \pi/2$ and $GCSP(Bt, C^0)$ is the CS profile generator which produces the CS profile at C^0 , by using the blue channel, Bt , of the current frame, Ft . Since the blue channel is more perceptible than the other channels to the CB, we utilize the blue channel as an input image of the CS profile generator.

D. Temporal Color Mapping Strategy

In this section we introduce how to utilize the spatiotemporal constraint to reduce the computational cost while preserving the TCC. For this purpose, CMS_t finds the pixel that has the most similar color in the previous frame and performs the color prediction process with a simplified operation. First, the similar color searching (SCS) finds X^{*-1} that is the most similar color in the previous frame, $Ft-1$ around the same position, C^0 , of the input color Xt^0 as

$$Xt^{-1} = \arg \min (X^0, Xt^{-1}), \\ XeRS(ct-1)$$

where Rs ($ct-1$) is the search range of the C^0 in $Ft-1$. In the proposed method, we set a radius of Rs to 2. Second, the linear color prediction (LCP) is used to find Xt^0 using the RGB components of Xt^0 , Xt^{-1} , X^{*-1} as follows:

$$Xt^0 = \begin{bmatrix} R_t^0 \\ G_t^0 \\ B_t^0 \end{bmatrix} = \begin{bmatrix} LP(R_{t-1}^0, R_{t-1}^0, R_{t-1}^0) \\ LP(G_{t-1}^0, G_{t-1}^0, G_{t-1}^0) \\ LP(B_{t-1}^0, B_{t-1}^0, B_{t-1}^0) \end{bmatrix}$$

where $LP(\cdot)$ is a linear prediction function defined as
 $LP(R^0_t, R^0_{t-1}, R^*_{t-1}) = (R^0_t - R^0_{t-1}/R^*_{t-1} - R^0_{t-1}) R^0_{t-1} + (R^*_{t-1} - R^0_{t-1}) / (R^*_{t-1} - R^0_{t-1}) R^*_{t-1}$

For the computational efficiency, when (X^*_{t-1}, X^0_t) is than τ_3 ($\tau_3 < \tau_2$), we terminate the SCS process and directly map X^*_{t-1} , X^0_t without the linear prediction method.

III. EXPERIMENTS

In this section, we compare the performance of our technique with the conventional techniques. In the implementation of the proposed algorithm, we convert the floating point operation to the integer operation. In addition, time-consuming operators such as trigonometrical function and the square root function are calculated in advance and tabulated.

The color accessibility in the image-level is evaluated. The NAT indicating how much distortion occurs through the recoloring process is calculated by,

$$NAT = \frac{1}{n(O_1O_2)} \left[\sum_{x \in O_1} \|X^0 - X^{\hat{0}}\|_2 \right]$$

where $n(A)$ is the number of elements in the set A . The smaller NAT indicates the better preservation of the color information.

TABLE I
NUMERICAL COMPARISON OF IMAGE-LEVEL QUALITY METRICS

Image	Resolution	Processing time(sec)	NAT
Brain	235 * 215	7.103870	19.8017
Gaugin	369 * 294	12.130848	19.4673
Snow hockey	320 * 158	10.079801	14.9938
Vegetable	670*273	15.159256	57.3348

Here, table shows numerically computed value of NAT. It indicates that the recolored images are natural enough in comparison with original image. The results are obtained by using laptop with corei7 and 12GB RAM

a) Image Recoloring Results

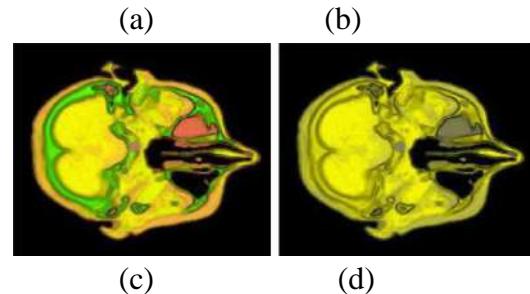
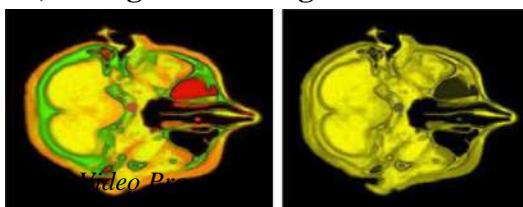


Figure 4 : (a) Original image Brain.png
(b) Original image seen by CVD (c) Recolor image
(d) Recolored image seen by CVD.

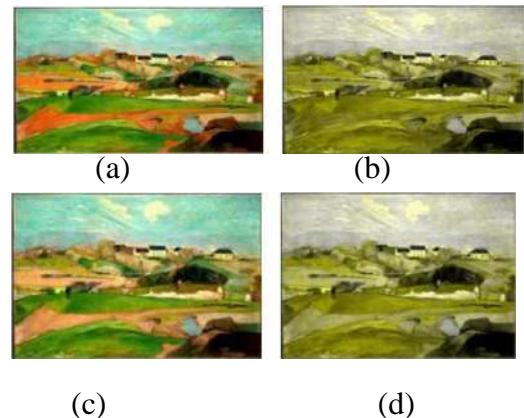


Figure 5 : (a) Original image Gaugin.png (b) Image seen by CVD (c) Recolor image (d) Recolored image seen by CVD

TABLE II
NUMERICAL COMPARISON OF COMPUTATIONAL COST

Video	Resolution	Duration (sec)	Processing time (sec)	FPS
Cricket	1280*720	7	1011.5	30
Basketball	640*360	10	404.2	30
Django unchained	640*360	10	345.5	30
football	640*360	30	195.9	5

We have calculated the computational complexity of the proposed method. Comparison results are represented in table II .

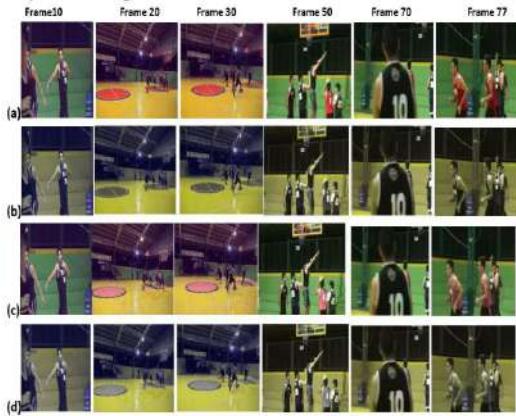


Figure 6 :Testing performed on the Basketball.avi video using proposed method.(a) original video frames,
(b) shows original Frames seen by CVD,(c) shows recoloured Frames,(d) recoloured Frames seen by CVD.

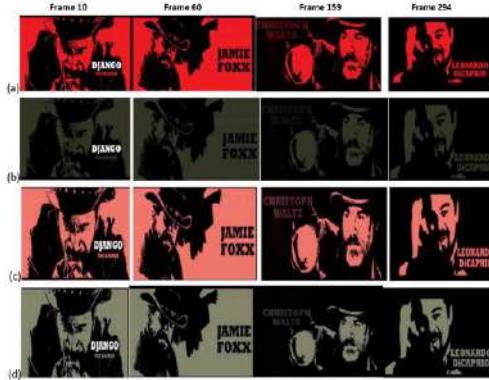


Figure 7 :Tesiting performed on the Django Unchanged.avi video using proposed method.(a) original video frames(b) Original frames seen by CVD(c) Recolored Frames(d) Recolored Frames seen by CVD

IV.CONCLUSION

The algorithm that considers floating point and uses matrix calculation and also the algorithm that uses neural networking concept are time consuming. The clustering algorithm are suitable only for some colors. The proposed method is fast color modification method for colorblind. RGB to LMS algorithm is used to maintain temporal color consistency of the re-colored video sequence in addition to spatial. The proposed method has low computational complexity. TCC is maintained, local color contrast is also maintained. This will be very useful for the people who are suffering from color blindness as they will have access to information as well as enjoy the video.

REFERENCES

- [1] DeepTiS.Khurge,BhagyashreePeshwani, “Modifying Image appearanceto ImproveInformation content for ColorBlind viewers,” published in IEEE 2015.Hideaki Orii, Hiroshi Maeda, HideakiKawano, and Takaharu,” ColorConversion Algorithm for Color Blindness using Self-Organizing Map,” in IEEE 2014.
- [2] Behzad Sajadi, Manuel M. Oliveira, Ramesh Raskar, AditiMajumder, and Rosa'lia G. Schneider, ”Using Patterns to EncodeColorInformation forDichromats,” published inIEEE transactions on visualization and computer graphics, vol. 19, no. 1, january 2013.
- [3] Jae-Yun Jeong, Yeo-Jin Yoon, Hyun-Ji Kim, Young-Hyun Kim, and Sung-Jea Ko, ”Color Modification for Color-blind Viewers Using the Dynamic Color transformation,” published in 2012 IEEE International Conference.
- [4] Jae-Yun Jeong, Tae-Shick Wang, Hyun-Ji Kim, Sung-Jea Ko and Yeo-Jin Yoon, “An Efficient Re-coloring Method with Information Preserving for the Color-blind” published in IEEE 2011.
- [5] Anagnostopoulos, G. T sekouras, and C. Kalloniatis, I. Anagnostopoulos,”Intelligent modification for the daltonization process of digitized paintings,”published in International Conference on Computer Vision Systems, 2007.
- [6] P. Doliotis, V. Athitsos, C. Anagnostopoulos, and G. Tsekouras, ”Intelligent modification of colors in digitized paintings for enhancing thevisual perception of color-blind viewers,”in Proceedings of the 5th International Conference 2009.
- [7] Er. Nazneen Pendhari, Mazgaonkar Raheen, Tole Sania, Dabhoiwala Misbah,“An Overview on Various Re-Colouring Methods for the Colour Vision Deficient,”(IJAFRSE)Volume 1, Issue 11, April 2015.
- [8] Mazgaonkar Raheen, Dabhoiwala Misbah, Tole Sania , Er. Nazneen Pendhari, “Video and Image Re-colouring for the ColourVision Deficient,” in International Journal of Global Technology Initiatives ,March 2015.
- [9] Er. Nazneen Pendhari, Abul Hasan Ali,Aamir Shoeb Alam Khan, Akash AshokPandey, “Feature distinction in visual media for the anomalous trichromats,”published in IEEE International Conference (ICETECH) in March 2016.

MACHINE LEARNING ALGORITHM ANALYSIS

Megharani Patil^[1], Ishita Sharma^[2], Manisha Sharma^[3], Sakshi Singh^[4], Sruthi Harish^[5]

*Thakur College of Engineering and Technology
Mumbai, India*

Abstract— Machine learning is the subject area that provides computers the capability to learn without being explicitly programmed. The biggest techgiants like Netflix, Facebook and Google that stand invincible for a couple of years from now have established a strong foundation of themselves in Machine Learning. All these things use a big amount of data, so machine learning is used to handle a large amount of data to showcase the desired result. To get most value from machine learning we must know how to pair the best algorithm with right tools and processes. So here we present a paper which compares and analyzes the various algorithms such as SVM, KNN, K-Means Clustering and Decision Tree. Recently SVM gained attention as a learning technique to attack classification problem. It solve classification problem by learning from training example .The K-mean algorithm operates on a given data set through pre-defined number of clusters k. The output of K Means algorithm is k clusters with input data partitioned among the clusters. kNN is a memory based model defined by a set of objects known as examples for which the outcome are known. In decision analysis a decision tree is used visually and explicitly to represent decisions and decision making. It uses a tree like model of decisions which is a commonly used tool in data mining for deriving a strategy to reach a particular goal.

Keywords— Artificial Intelligence, Clustering, Support Vector Machine, Data Mining, Reinforcement Learning.

I. INTRODUCTION

Machine learning is a type of artificial intelligence that learns as it defines new patterns in data, allow data scientist to explicitly point revenue opportunities and create strategies to enhance individual experience using information concealed in large data sets. Determining the right algorithm is the key motive of any machine learning project and because we have to select them among many such algorithm it becomes essential to understand their strengths and weaknesses. There are many types of machine learning some of them are supervised learning, unsupervised learning, reinforced learning. In Supervised Learning we are provided with an input variable which is X and an output variable which is Y and we use an algorithm to learn the mapping function from input to output. The equation is $Y=f(X)$. Supervised learning algorithm try to model relationship and dependencies between the target prediction output and the input features such that we can predict the output values for new data based on those relationship which it learned from the previous data sets. Algorithm in Supervised Learning:- Nearest Neighbor, Naïve Bayes, Decision Tree, Linear Regression, Support Vector Machine(SVM).

II. RELATED WORK

If instances are provided with known labels then the learning is said to be supervised in contrast to unsupervised algorithms .By applying unsupervised algorithms, researchers hope to discover unknown, but useful classes of items[1].Another kind of machine learning is reinforcement learning[2]. A brief review of what ML includes can be found in[3].A historical survey of logic and instance based learning can be found in [4]. Instance based learning algorithms are known to be lazy–learning algorithms [5],as they delay the induction or generalization process until the classification is performed. Lazy-learning algorithms require less computation time than the eager-learning algorithms but require more time during classification process[6].Aha [7] and De Mantaras[4] presented a review of instance based learning classifiers. Murthy[8] provided an overview of work in decision trees and a sample of their usefulness to newcomers as well as practitioners in the field of ML. The most well-known algorithm in literature for building decision trees is the C4.5[9].Comparison between decision tree and other learning algorithms has been done in[10].[11] proposed Rainforest, a framework for building fast and scalable algorithms to construct decision trees that gracefully adapt to the amount of main memory available. To sum up, one of the most useful characteristics of decision trees is their comprehensibility. People can easily understand why a decision tree classifies an instance as belonging to a specific class[6]. An excellent survey on SVMs can be found in[12] and a more recent book is by [13].In SVM the selection of an appropriate kernel function is important ,since the kernel function defines the feature space in which the training set instances will be classified. Genton[14] described classification of several classes of kernel Training. Training the SVM is done by training the Nth dimensional QP problem, where N is the number of samples in the training dataset. Solving this problem in the standard QP methods involves large matrix operations, as well as time-consuming numerical computations, and is mostly very slow and impractical for large problems. Sequential Minimal Optimization (SMO) is comparatively much simple algorithm as it can solve the SVM QP problem faster without the need of any extra matrix storage and without using numerical QP optimization steps at all [15].

III. METHODS OF LEARNING

A. SUPERVISED LEARNING

Supervised learning is a data mining task of assuming a function from a given labeled training data. The training data consist of a set of training examples. In supervised learning, each example is a pair consisting of an input object (typically a vector) and the desired output value (also called the supervisory signal).-

X y (pre-classified Training Examples) Given an observation x , what is the best label for y . A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples. An optimal scenario will allow the algorithm to correctly determine the class labels for unknown instances. It is thus essential for the learning algorithm to conclude from the training data to unknown conditions in a analytical way.

B. UNSUPERVISED LEARNING

Unsupervised machine learning is the machine learning task of inferring a function to describe hidden structure from "unlabeled" data (a classification or categorization is not included in the observations). Since the examples provided to the learner are unlabeled, there is no analysis of the authenticity of the structure which is the output by the relevant algorithm—which is one way of distinguishing the learning algorithm.

C. REINFORCED LEARNING

Reinforcement Learning is a type of Machine Learning, and thereby also a branch of Artificial Intelligence. In order to maximize its performance, reinforced learning allows machines and software agents to automatically determine the ideal behavior within a particular environment. Simple reward feedback is required for the agent to learn its behavior; this is known as the reinforcement signal. There are many different algorithms that tackle this issue. Reinforcement Learning is described by a specific type of problem, and all its solutions are accounted as Reinforcement Learning algorithms. In the problem, an agent is supposed decide the best action to select based on his current state. When this step is repeated, the problem is known as a Markov Decision Process.

IV. MACHINE LEARNING ALGORITHMS

A. K-NEAREST NEIGHBOUR

K-nearest neighbor algorithm (KNN) is a method used for classification and regression. It does not use any parameters. In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression. In k-NN classification, the output is a member of the class. An object is classified by a majority vote of its neighbors, the object is allocated to the class most common among its k nearest neighbors (k is a positive integer, typically small). The output is an attribute value of its k nearest neighbors in the KNN regression. The value obtained is the average of the values of its k nearest neighbors. KNN is a model of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is postponed until classification.



Fig 1: Voronoi diagram

The KNN algorithm is simplest among all machine learning algorithms, both for classification and regression. It is a useful technique which can be used to assign weight to the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the far ones. For example, a common weighing procedure consists in assigning each neighbor a weight of $1/d$, where d represents the distance to the neighbour. The neighbours are taken from a set of objects for k-NN classification or the object property value for k-NN regression is known. This can be thought of as the training set for the algorithm. An irregularity of the KNN algorithm is that it is sensitive to the local structure of the data. The algorithm is not to be confused with k-means, another popular machine learning technique. The training examples each along with a class label are termed as vectors in a multidimensional feature space. The training phase of the algorithm consists of storing the feature vectors and class labels of the training samples. In the classification phase, k is a user-defined constant, and an unlabeled vector termed as a query or test point is classified by assigning the label which is most recurrent among the k training samples nearest to that query point. Euclidean distance is most commonly used distance metric for continuous variables. The best choice of k relies on the data; generally, larger values of k reduce the effect of noise on the classification, but make boundaries between classes less distinct. The special case where the class is predicted to be the class of the closest training sample (i.e. when $k = 1$) is called the nearest neighbor algorithm. The accuracy of the k-NN algorithm can be drastically reduced by the presence of noisy or irrelevant features.

Nearest Neighbors: Pros and Cons

Pros:

- It is very simple to implement
- It has a high flexibility to feature / distance choices
- It is capable of efficiently handling the multi-class cases

Cons:

- It is time consuming to search for nearest neighbors
- It is necessary to have a meaningful distance function

B. DECISION TREE

A tree has many applications in real life like for searching, compression, parsing, scheduling of processes and turns out that it has influenced a wide area of machine learning, covering both classification and regression. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. As the name goes, it uses a tree-like model of decisions. It is a commonly used tool in data mining for deriving a strategy to reach a particular goal and is also widely used in machine learning. Consider a very basic example that uses data set for predicting whether a passenger will survive or not. In the figure 4.2.1, 3 features from the data set, namely sex, age and sibsp(number of spouses or children along) are used.

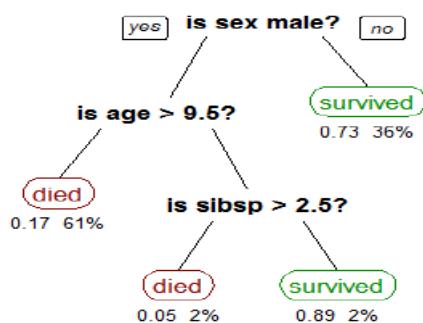


Fig 2: Decision Tree

Even though a real dataset have many features and this will just be a branch in a much bigger tree, but the simplicity of the algorithm cannot be neglected .This methodology is widely known as learning decision tree from data. The above tree is called Classification tree as the target is to classify passenger as survived or died. Regression trees are represented in the same manner, similarly as they predict continuous values like price of a house. Decision Tree algorithms are also referred to as Classification or Regression Trees. Growing a tree involves deciding on which features to choose and what conditions to use for splitting, along with knowing when to stop. As a tree generally grows randomly, there is a need to trim it down for it to look precise.



Fig 3: Recursive binary splitting

In the recursive binary splitting all the properties are considered and the point where they are splitting into different branches are tested using cost function. The split with the best is that the point with the lowest cost is selected.

In figure 4.2.2 the tree is learned from dataset. "In the root, all attributes are considered and the training data is divided into groups based on the split. We have three features, so we will have three candidate splits." The split that costs least is chosen, which is in our example is sex of the passenger. "This algorithm is recursive in nature as the groups formed can be sub-divided using same strategy. Due to this method, this algorithm is also known as the greedy algorithm", as we want to lower the cost of this method. This makes the root node as best predictor/classifier. The maximum height of tree is the length of longest path from root node to leaf node.

Advantages of CART

- It is simple to understand, interpret and visualize.
- Decision trees absolutely perform feature selection.
- It is capable of handling both numerical and categorical data as well as multi-output problems.

- It requires relatively less effort from users for data preparation.
- The tree performance is not affected by nonlinear relationships between parameters

Disadvantages of CART

- Decision Trees is not proper for continuous variables and it result in insecurity .
- It is easy to use when compared with other decision making configuration but creating large decision trees that contain several branches is a complex and time consuming task.
- Decision tree machine learning algorithms consider only one attribute at a time and might not be best suited for actual data in the real life.

C. SUPPORT VECTOR MACHINE

"Support Vector Machine" (SVM) is a supervised machine learning algorithm which is utilized for both classification and regression challenges. Even though, it is mostly used in classification problems. In this algorithm, each data item is plotted in n-dimensional space .The value of each feature is the value of the particular co-ordinate. Then, the appropriate hyper-plane is obtained by finding out the plane that segregates the two classes very well.

SCENARIO 1

In figure 4.3.1, there are three hyper-planes (A, B and C). In order to find the right hyper-plane that classifies the star and circle the thumb rule has to be applied. The thumb rule to choose the right hyper-plane is to select the hyper-plane which segregates the two classes better. In the figure 4.3.1 hyper-plane "B" has excellently performed this job.

Scenario 1:

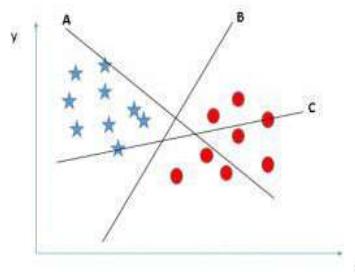


Fig 4: Scenario 1

Scenario 2

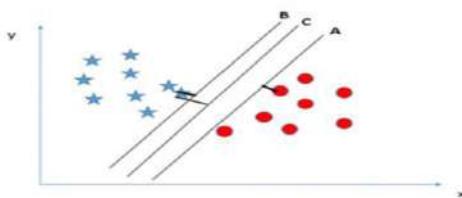


Fig 5: Scenario 2

In figure 4.3.2, there are three hyper-planes (A, B and C) and all are segregating the classes well. In this scenario as the thumb rule is well obeyed by all the 3 hyper-planes. So, in this case the appropriate hyper-plane is obtained by checking the distances between nearest data point (either class) and hyper-plane and then selecting the hyper-plane that has maximum distance between the nearest data point of each class and hyper-plane. This distance is called as Margin. In the figure 4.3.2, margin for hyper-plane C is higher as compared to both A and B(for both the classes). Hence, the appropriate hyper-plane is C. Another important and valid reason for choosing the hyper-plane with higher margin is robustness. If the hyper-plane having low margin is selected then there is high chance of miss-classification.

Scenario 3

In figure 4.3.3, hyper-plane B has higher margin in comparison to hyper-plane A. But, in this case, selection of that hyper-plane takes place which segregates the classes more accurately prior to maximizing margin. Here, hyper-plane A has more accurately classified the two classes. Hence, hyper-plane A is the appropriate hyper-plane.

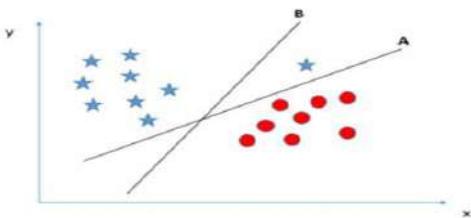


Fig 6: Scenario 3

Scenario 4

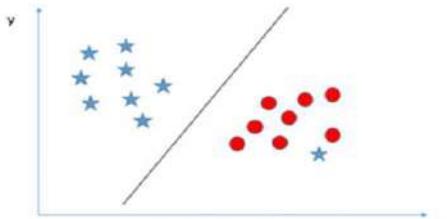


Fig 7: Scenario 4

In the figure 4.3.4 ,the segregation of the two classes is not possible using a straight line, as one of star lies in the territory of other(circle) class as an outlier . So one star at other end is like an outlier for star class. SVM has a feature

to ignore outliers and find the hyper-plane that has maximum margin. Hence, SVM is robust to outliers.

Scenario 5

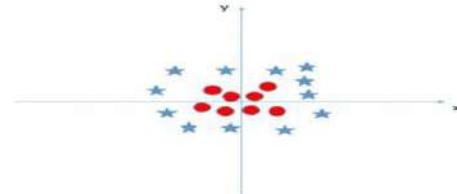


Fig 8 Scenario 5

In the figure 4.3.6, a linear hyper-plane cannot enable the segregation between the two classes. SVM solves this problem by introducing a new additional feature .On plotting the data points on axis x and z:

Scenario 6

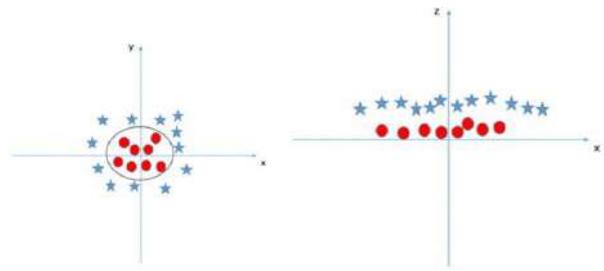


Fig 9

Fig 10

For these set of data points, all values of z is positive , as z is the squared sum of both x and y. In the original plot, red circles appear close to the origin of x and y axes, leading to lower value of z and star relatively away from the origin resulting in the higher value of z. These are those functions which take low dimensional input space and transform it to a higher dimensional space i.e. it converts a not separable problem to a separable problem. These functions are called kernels. It is very useful in non-linear separation problem. Basically, it does some extremely complex data transformations and finds out the process to separate the data based on the labels or outputs defined by user.

PROS AND CONS OF SVM

Pros:

- It works really well with clear margin of separation
- It is effective in high dimensional spaces.
- It is effective in cases where number of dimensions is greater than the number of samples.
- SVM is memory efficient as it uses a subset of training points in the decision function (called as support vectors).

Cons:

- SVM is inefficient when the data set is very large as higher training time is required.
- It also doesn't perform very well, when the target classes are overlapping i.e. data set has more noise.
- Probability estimates are calculated using an expensive five-fold cross-validation(using SVC method of Python scikitlearn) as these probability estimates are not directly provided by SVM.

D. K-MEANS CLUSTERING

Clustering is a process of categorizing objects into groups whose members are identical in some way. In short a cluster is collection of objects which are identical among them but different from the objects of other cluster.

Classification of clustering algorithm:

- Exclusive clustering
- Overlapping clustering
- Hierarchical clustering
- Probabilistic clustering

K-means clustering is a type of unsupervised learning, which is used when you have unlabeled data (i.e., data without defined categories or groups).

The purpose of implementing this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided.

- Data points are clustered based on feature similarity.
- Basically K-Means runs on distance calculations, which uses “Euclidean Distance” for this purpose.
- Euclidean distance calculates the distance between two given data points using the following formula

The formula given below represents the distance in 2-Dimensional space but the same useful in multi-dimensional space.

$$d^2(\mathbf{x}, \mathbf{m}_k) = \sum_{n=1}^N (\mathbf{x}_n - \mathbf{m}_{kn})^2$$

[16]

K-MEANS CLUSTERING ALGORITHM:

Given the cluster number K , the K-means algorithm is carried out in three steps after initialization.

Initialization: set seed points (randomly)

- 1) Assigning each object to the cluster of the nearest seed point measured with a specific distance metric.
- 2) Compute new seed points as the centroid of the clusters of the current partition (the centroid is the centre i.e. mean point of the cluster)
- 3) Go to Step 1, stop when no more new assignment (i.e., membership in each)

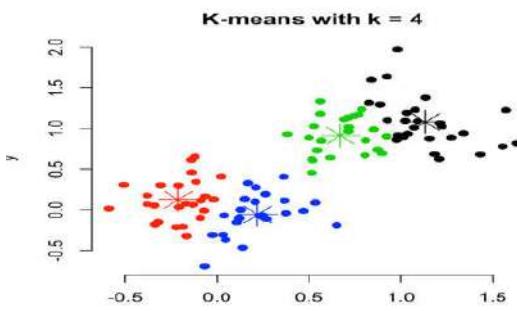


Fig 11: K-Means clustering

APPLICATIONS

1. DONOR SEGMENTATION

For a Non-Profit Organization, a segmentation was to understand level and response of people who will give donation and continue to contribute for a cause.

Some of the dimensions used for K Means clustering were

- Salary
- Gender
- Number of times and Value of contributed in the last 6/12 months
- Age
- Participation in Events and Type of Events (e.g. conferences, sports, environment, animal rights, women empowerment etc)

2. COLOUR BASED IMAGE SEGMENTATION USING K-MEANS CLUSTERING

- Color image segmentation is an upcoming topic of the research for researchers in image processing.
- There is a vital need for better segmentation approach because of its utmost importance in the technique known as image processing.
- Clustering is mostly used methodology for the segmentation of images. This content is a combined approach for the segmentation of images.
- Here, we are combining K-mean Clustering Technique and Water-shed segmentation algorithm.
- Color space has most starring impact on segmentation process, we have chosen [a*b* color space in our approach.
- When we perform segmentation process or fetching of the clusters from an image, noise arises in that image.
- To eliminate that noise filters are being used. First, we will check the type of noise and then we will decide what kind of filter is to be used.

V. CONCLUSION

This paper describes best-known ML techniques in relative detail. The key question when dealing with ML classification is not whether a learning algorithm is superior to others, but under which conditions a particular method can significantly outperform others, on a given application problem. After a better understanding of the strengths and weaknesses of each method, the possibility of integrating two or more algorithms together to solve a problem should be investigated. The objective is to utilize the strengths for an instance, SVMs tend to perform much better when dealing with multi-dimensions and continuous features. In contrast, logic-based systems like decision trees tend to perform better when dealing with discrete features. For SVM, a large sample size is required in order to achieve its maximum prediction accuracy. Most decision tree algorithms cannot perform well with problems that require diagonal partitioning. Similarly though training time varies according to the nature of the application task and dataset, lazy learning algorithms like kNN require zero training time because the training instance is simply stored. Univariate decision trees are also quite popular to be fast-at any rate, comparatively much faster than SVMs. The basic kNN require great deal of storage space for the training phase, and its execution space is at least as big as its training space whereas, the non-lazy learners, execution space is much smaller than the training space. Decision trees are considered very easy to interpret, whereas SVMs have notably poor interpretability. Even, kNN have a poor interpretability because, unstructured collection of datasets is far from readable. K-means

Clustering is the simplest and hence the most convenient unsupervised learning algorithm. It can be well utilized to solve clustering issues and the cases that include a large dataset. Thereby we conclude that no single algorithm can uniformly outperform other algorithms over all datasets. The simplest approach is to estimate the accuracy of the candidate algorithms on the problem and to select the one that appears to be the most accurate.

REFERENCES

- [1] S.Kotsiantis,P.Pintelas,"Recent Advances in Clustering:A Brief Survey",WSEAS Transactions on Information Science and Applications,Vol. 1,No 1 pp73-81,2004
- [2] Gosavi, A., "Simulation-Based Optimization: Parametric Optimization Techniques and Reinforcement Learning, Series: Operations Research/Computer Science Interfaces Series", Vol. 25, Springer,2003
- [3] Dutton, D. & Conroy, G. 1996,"A review of Machine Learning, Knowledge Engineering Review ":" pp 341-367 ,1996
- [4] De Mantaras & Armengol E.," Machine Learning from examples: Inductive and Lazy Methods .Data & Knowledge Engineering ":" pp 99-123,1998
- [5] Mitchell,T ."Machine Learning",Mcgraw Hill,1997
- [6] S.B. Kotsiantis"Supervised Machine Learning :A Review of Classification Techniques"
- [7]Aha,D.,Lazy Learning. Dordrecht:Kluwer Academic Publishers,1997
- [8]Murthy,"Automatic Construction of Decision Trees from Data:A Multidisciplinary Survey,Data Mining and Knowledge Discovery 2": pp 345-389,1998
- [9]Quinlan,J.R.1993.C4.5: Programs for Machine Learning. Morgan Kaufmann,San Francisco.
- [10]Tjen-Sien L.,Wei-Yin L.,Yu-Shan, " A Comparison of Prediction Accuracy, Complexity,Trainig Time of Thirty-Three Old and New Classification Algorithms.Machine Learning", 40: pp 203-228,2000
- [11]Gehrke,J.,Ramakrishnan, R. & Ganti, V. 2000,Rainforest-A Framework for Fast Decision Tree Construction of Large Datasets,Data Mining and Knowledge Discovery, Vol. 4,Issue 2-3, Jul 2000, pp127-162
- [12]Burges,C. ,"A tutorial on Support Vector Machines for Pattern Recognition.Data Mining and Knowledge Discovery".2(2):pp 1-47,1998
- [13]Cristianini, N. & Shawe-Taylor, J.".An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods",Cambridge University Press,Cambridge,2000
- [14]Genton, M.."Classes of Kernels for Machine Learning: A Statistics Perspective. Journal of Machine Learning Research": pp 299-312,2001
- [15] Platt, J. , "Using sparseness and analytical QP to speed training of support vector machines. In Kearns ,M., Solla,S. & Cohn, D.ed.Advances in neural information processing systems". MIT press.,1999
- [16]Website
<http://syllabus.cs.manchester.ac.uk/ugt/2017/COMP24111/materials/slides/K-means.ppt>, Ke Chen, Reading : [7.3,EA], [9.1,CMB] , pp 3

Handwritten Pattern Recognition System Using Artificial Neural Network

Ruchit Tripathi,Sudarshan Dasarathy,Vishal Singh,Jenil Shah,Rekha Sharma

Department of Computer Engineering, Thakur College of Engineering & Technology,Mumbai, India

Abstract—Artificial Intelligence is proving to be vital in the 21st century. Difficult tasks are being performed with the help of many applications. But the areas like computer vision lacks accuracy to recognize patterns and letters. The accuracy to recognize handwritten pattern is still less than the expectation. Handwritten digit recognition has been a major problem in the computer vision. Present algorithms are inaccurate to make it a perfect smart system. This is an important field because of many day-to-day life applications where large number of documents with handwritten digits should be entered and converted into text form such as bank cheque analysis, insurance paper analysis and other handwritten forms etc. So, our main aim is to develop a system program which can tackle with handwritten digits and letters and makes it easy to read by the computer system. We are going to use Machine Learning algorithm such as Backpropagation Neural Network Algorithm to build a system which can deal with the problem of computer vision with higher accuracy. Additionally, we are going to implement Multiplier Feed Forward Neural Network and Support Vector Machine algorithms to recognize various other patterns. In order to achieve 98% accuracy, we are going to focus on different Image processing techniques also. Large number of handwritten datasets will be used to train our system. After that, our trained system will be able to understand unknown complex handwritten patterns.

Keywords—Handwritten digit and letter recognition, Patterns, Artificial neural network,Machine Learning

I. INTRODUCTION

In the field of image processing and pattern recognition, handwriting recognition has one of the hottest and demanding directions in the recent years. Day by day new technologies and innovative methods have been proposed continuously. With the development of the smartphone operation system, the application in handwritten recognition has aroused more and more attention from researchers. It contributes immensely to the advancement of an automation process and can improve the interface between man and machine in numerous applications. A Handwritten pattern recognition system with a good recognition performance needs to maintain a very high recognition rate, and at the same time, to obtain a very high reliability, or a very low error rate. Recent developments on classifiers and feature extraction have significantly increased the recognition accuracy of handwritten digit recognition systems.

The neural networks have been successfully used to yield comparably high recognition accuracy levels. Several applications including mail sorting, bank processing, document reading and postal address recognition require handwriting recognition

systems. As a result, the handwriting recognition continues to be an active area for research towards exploring the newer techniques that would improve recognition accuracy. Here our aim is to provide the most enhanced and accurate system which exhibits all the above mentioned features

II. LITERATURE SURVEY

Previous work on handwritten pattern recognition system has been done in many researches. Our main aim is to tackle the difficulties in the previous work and to provide the most accurate system. The following survey table classifies the scientists based on the parameters like method, accuracy, processing time, type of text, text/digits or both and size of database.

Following are the work done by the scientist on the handwritten pattern recognition system:

Table I- Handwritten pattern recognition		
Reference Number	Advantages	Disadvantages
[1]	High recognition rate (95.19%)	Certain digits causes issue in recognition (3,8,9,7)
[2]	Process time was quick, higher accuracy in few methods (97.36%)	Less accurate in few methods (ceparmimethod)
[4]	High recognition rate and can be used for only arabic (above 95%)	Performance decreases, less accurate in case of Arabic
[5]	Robust detection takes place between noise and partial occlusion	Accuracy was not good (72.5% only)
[6]	Provides accurate information than previous methods (98% accurate)	Not suitable for digits with straight lines eg. 10 ,19
[2]	Deals only neighborhood pixels (clustering)	Overfitting on dataset increases.

III. PROPOSED METHOD

The selected model for our project is incremental model. Iterative and Incremental development is any combination of both iterative design or iterative method and incremental build model for software development. The combination is of long standing and has been widely suggested for large development efforts. The basic idea behind this method is to develop a system through repeated cycles (iterative) and in smaller portions at a time (incremental), allowing software developers to take advantage of what was learned during development of earlier parts or versions of the system. Learning comes from both the development and use of the system, where possible key steps in the process start with a simple implementation of a subset of the software requirements and iteratively enhance the evolving versions until the full system is implemented. At each iteration, design modifications are made and new functional capabilities are added.

In our project, as explained in incremental project, we are going to break our whole work into small modules. After breaking into modules, we are going to work on these small modules independently which will increase time efficiency and faster execution of our project. Each module will pass through three phases of development (Design & development, Testing, Implementation). In Development and design phase, Planning, requirement gathering and development will be done. After development of each module, testing of modules will be done to ensure that there are no bugs in the modules. At the end implantation of each module will be done independently and after successful implementation, all modules will be merged and the whole project will be released for customer feedback. Customer will give new requirements related to our product and based on that changes will be done in iteration. Since it is an incremental model, iteration of each phases will take place based on customers' requirements. Every time a new version of the product will be released with new upgrades.

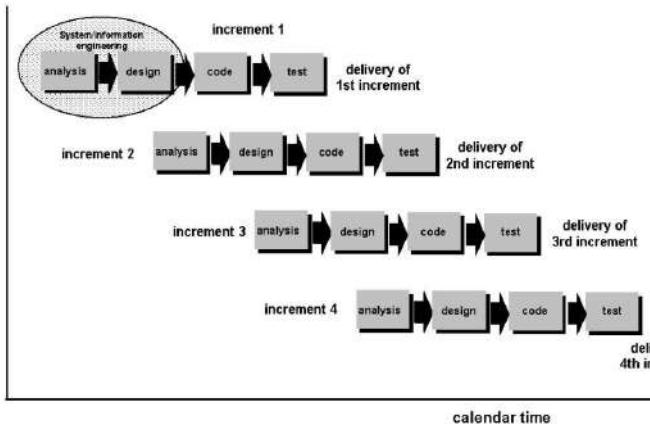


Fig 1: Phases of Incremental Model of our Project

The proposed method will be useful for recognizing handwritten digits and letters with the accuracy 98%. We are going to use Machine Learning algorithm such as Backpropagation. Neural Network Algorithm to build a system which can deal with this computer vision problem with higher accuracy. In addition to that we are going to implement Multiplier Feed Formal Neural Network and Support Vector Machine algorithms to recognize different other patterns. In order to achieve more accuracy, we will

focus on different Image processing technique which will help our algorithm to produce more accurate results.

Large number of handwritten datasets will be used to train our system. After that, our trained system with the help of Different Neural Networks Algorithm and other machine learning algorithms will be able to recognize future complex handwritten problems.

The classification stage is the decision making part of a recognition system and it uses the features extracted in the previous stage. A feed forward back propagation neural network having two hidden layers with architecture of 54-100-100-38 is used to perform the classification. The hidden layers use log sigmoid activation function, and the output layer is a competitive layer as one of the characters have to be identified. The feature vector is denoted as X , and defined as $X = (f_1, f_2, \dots, f_d)$, where f denotes features and d is the number of zones into which each character is divided. The number of input neurons is determined by length of the feature vector d . The total numbers of characters n determines the number of neurons in the output layer. The number of neurons in each hidden layer is obtained by trial and error.

The output of i^{th} layer is given by

$$a^i \log \sigma(w^i a^{i-1} + b^i) +$$

Where,

$$i = [1, 2, 3] \text{ and } a^0 = \mathbf{P}$$

w^i = weight vector of i^{th} layer

a^i = output of i^{th} layer

b^i = bias vector for i^{th} layer

IV. BLOCK DIAGRAM

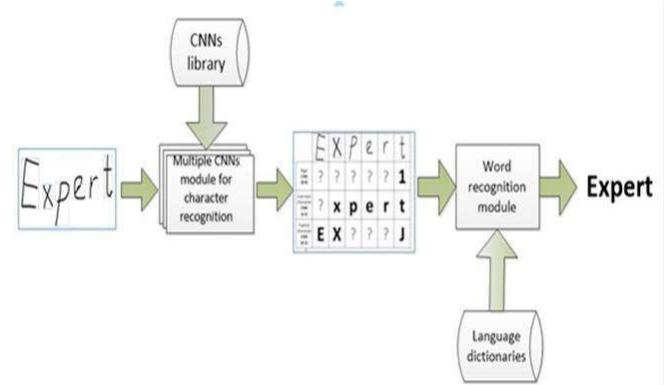


Fig 2: Block diagram proposed character recognition system

Our system consists of preprocessing, segmentation, feature extraction, classification and recognition, and post processing stages. The first important work is done by the scanners. They will scan the handwritten documents such as letters, claim papers etc. and segmentation and feature extraction work will be done. Preprocessing followed by segmentation and feature extraction. Preprocessing includes the steps that are required

to shape the input image into a form suitable for segmentation in the segmentation, the input image is segmented into individual characters and then, each character is resized into $m \times n$ pixels towards the extracting the features. The multi-class classification and the backpropagation algorithms will be used to recognize the input data given by the previous stages.

V. IMPLEMENTATION

Based on the preceding researches, literature survey and by using the neural network algorithms we have executed implementation of our software.

1	4	9	9	5	3	2	8	6	1
6	7	1	1	9	7	2	3	6	5
2	4	9	5	6	1	9	2	1	0
1	0	7	0	3	1	1	2	0	8
3	1	0	4	0	5	2	9	3	9
3	7	8	4	4	7	4	2	5	6
5	7	1	4	9	8	4	1	8	3
2	3	6	6	2	9	9	2	2	4
5	1	5	4	7	3	4	7	4	4
9	3	4	9	6	9	2	0	5	0

Fig 3: Software Implementation

The above table is the dataset of digits. We have obtained this dataset through preprocessing. Using this dataset we have trained our system.

```

81 - pred = predict(Theta1, Theta2, X(rp(i),:));
82 - fprintf('\nNeural Network Prediction: %d (digit %d)\n', pred,
83 -
84 - % Pause
85 - fprintf('Program paused. Press enter to continue.\n');
86 - pause;
87 - end
88
89
```

Command Window

New to MATLAB? See resources for [Getting Started](#).

Program paused. Press enter to continue.

Loading Saved Neural Network Parameters ...

Training Set Accuracy: 97.520000

Fig 4: Snapshot-1

The above snapshot gives the accuracy of our system. Using the dataset we have achieved 97.52 % accuracy as per our goal.

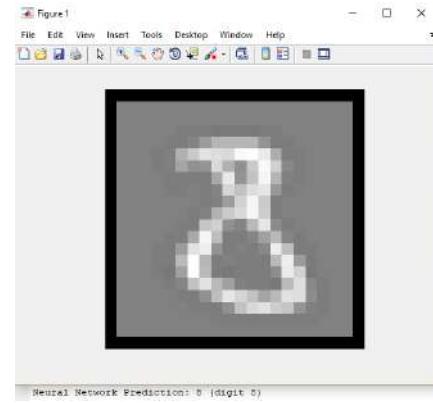


Fig 5: Snapshot-2

After training our system, we have worked upon the test dataset. We ran the test for digit 8 and we were able to achieve the desired result successfully. The above snapshot shows the test for digit 8

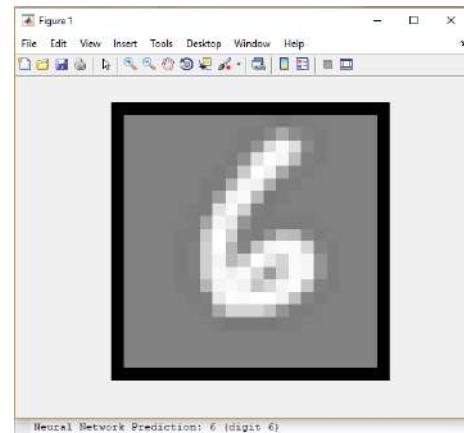


Fig 6: Snapshot-3

Similarly, our software was able to provide the accurate result for the digit 6 as well. The above snapshot shows the test for digit 6.

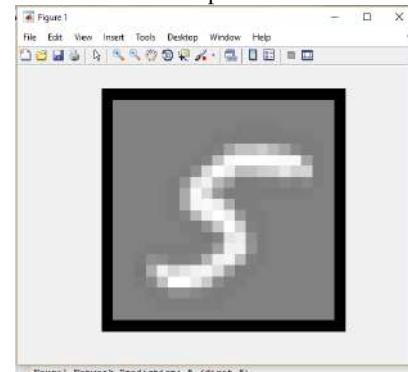


Fig 7: Snapshot-4

Also, our system provided accurate result for the digit 5 too. The above snapshot shows the test for digit 5.

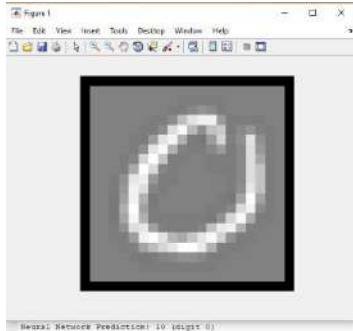


Fig 8: Snapshot-5

The above snapshot shows the test for digit 10. For the digit 10 our system was inaccurate and gave the result as 0. We are currently working on rectifying this bug in our software.

VI. CONCLUSIONS AND FUTURE SCOPE

We have reviewed various research papers and done a literature survey on various papers and found out the gaps in the existing

system. We have also developed various documents required for the implementation of the project.

Most of the private and government sectors still use the traditional way of checking and verifying the handwritten patterns. This method decreases the work efficiency and consumes more time. Introduction of handwritten pattern recognition system in these sectors will help to convert the handwritten document into file format and it will be easy to verify the documents. This will boost time as well as work efficiency. If we connect our system to a dictionary server of different languages, our system would be able to predict the meaning of scanned texts in its respective language.

REFERENCES

- [1] A. Gattal, "Isolated Handwritten Digit Recognition Using oBIFs and Background Features", 12th IAPR Workshop on Document Analysis Systems, at 305-310, April 2016.
- [2] R. BABU, "Novel Technique for Handwritten Digit Recognition Image Feature Extraction for Recognition", IJVIPNS-IJENS Vol: 15 No: 06.
- [3] Bouchra El jacimy, "Handwritten digit recognition based on DCT features and SVM classifier", 10.1109/ICoCS.2014.7060935, 16 March 2015.
- [4] Mamta Garg, "A Novel Approach to Recognize the off-line Handwritten Numerals using MLP and SVM Classifiers", IJCSET13-04-07-091
- [5] J Pradeep, "MLP (Multi-layer Perceptron)", P. E. Ajmire, IJPRET, 2015; Volume4(4):1-10
- [6] Bouchra El jacimy, "Handwritten digit recognition based on DCT features and SVM classifier", IEEE:14985664, 16 March 2015.

Handwritten Pattern Recognition System Using Artificial Neural Network

Ruchit Tripathi,Sudarshan Dasarathy,Vishal Singh,Jenil Shah,Rekha Sharma

Department of Computer Engineering, Thakur College of Engineering & Technology,Mumbai, India

Abstract—Artificial Intelligence is proving to be vital in the 21st century. Difficult tasks are being performed with the help of many applications. But the areas like computer vision lacks accuracy to recognize patterns and letters. The accuracy to recognize handwritten pattern is still less than the expectation. Handwritten digit recognition has been a major problem in the computer vision. Present algorithms are inaccurate to make it a perfect smart system. This is an important field because of many day-to-day life applications where large number of documents with handwritten digits should be entered and converted into text form such as bank cheque analysis, insurance paper analysis and other handwritten forms etc. So, our main aim is to develop a system program which can tackle with handwritten digits and letters and makes it easy to read by the computer system. We are going to use Machine Learning algorithm such as Backpropagation Neural Network Algorithm to build a system which can deal with the problem of computer vision with higher accuracy. Additionally, we are going to implement Multiplier Feed Forward Neural Network and Support Vector Machine algorithms to recognize various other patterns. In order to achieve 98% accuracy, we are going to focus on different Image processing techniques also. Large number of handwritten datasets will be used to train our system. After that, our trained system will be able to understand unknown complex handwritten patterns.

Keywords—Handwritten digit and letter recognition, Patterns, Artificial neural network,Machine Learning

I. INTRODUCTION

In the field of image processing and pattern recognition, handwriting recognition has one of the hottest and demanding directions in the recent years. Day by day new technologies and innovative methods have been proposed continuously. With the development of the smartphone operation system, the application in handwritten recognition has aroused more and more attention from researchers. It contributes immensely to the advancement of an automation process and can improve the interface between man and machine in numerous applications. A Handwritten pattern recognition system with a good recognition performance needs to maintain a very high recognition rate, and at the same time, to obtain a very high reliability, or a very low error rate. Recent developments on classifiers and feature extraction have significantly increased the recognition accuracy of handwritten digit recognition systems.

The neural networks have been successfully used to yield comparably high recognition accuracy levels. Several applications including mail sorting, bank processing, document reading and postal address recognition require handwriting recognition

systems. As a result, the handwriting recognition continues to be an active area for research towards exploring the newer techniques that would improve recognition accuracy. Here our aim is to provide the most enhanced and accurate system which exhibits all the above mentioned features

II. LITERATURE SURVEY

Previous work on handwritten pattern recognition system has been done in many researches. Our main aim is to tackle the difficulties in the previous work and to provide the most accurate system. The following survey table classifies the scientists based on the parameters like method, accuracy, processing time, type of text, text/digits or both and size of database.

Following are the work done by the scientist on the handwritten pattern recognition system:

Table I- Handwritten pattern recognition		
Reference Number	Advantages	Disadvantages
[1]	High recognition rate (95.19%)	Certain digits causes issue in recognition (3,8,9,7)
[2]	Process time was quick, higher accuracy in few methods (97.36%)	Less accurate in few methods (ceparmimethod)
[4]	High recognition rate and can be used for only arabic (above 95%)	Performance decreases, less accurate in case of Arabic
[5]	Robust detection takes place between noise and partial occlusion	Accuracy was not good (72.5% only)
[6]	Provides accurate information than previous methods (98% accurate)	Not suitable for digits with straight lines eg. 10 ,19
[2]	Deals only neighborhood pixels (clustering)	Overfitting on dataset increases.

III. PROPOSED METHOD

The selected model for our project is incremental model. Iterative and Incremental development is any combination of both iterative design or iterative method and incremental build model for software development. The combination is of long standing and has been widely suggested for large development efforts. The basic idea behind this method is to develop a system through repeated cycles (iterative) and in smaller portions at a time (incremental), allowing software developers to take advantage of what was learned during development of earlier parts or versions of the system. Learning comes from both the development and use of the system, where possible key steps in the process start with a simple implementation of a subset of the software requirements and iteratively enhance the evolving versions until the full system is implemented. At each iteration, design modifications are made and new functional capabilities are added.

In our project, as explained in incremental project, we are going to break our whole work into small modules. After breaking into modules, we are going to work on these small modules independently which will increase time efficiency and faster execution of our project. Each module will pass through three phases of development (Design & development, Testing, Implementation). In Development and design phase, Planning, requirement gathering and development will be done. After development of each module, testing of modules will be done to ensure that there are no bugs in the modules. At the end implantation of each module will be done independently and after successful implementation, all modules will be merged and the whole project will be released for customer feedback. Customer will give new requirements related to our product and based on that changes will be done in iteration. Since it is an incremental model, iteration of each phases will take place based on customers' requirements. Every time a new version of the product will be released with new upgrades.

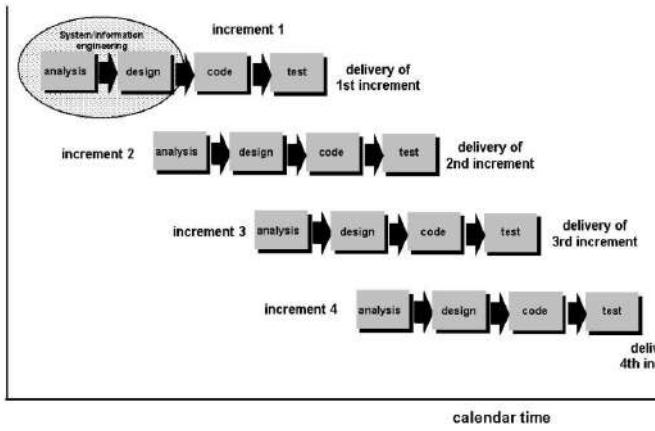


Fig 1: Phases of Incremental Model of our Project

The proposed method will be useful for recognizing handwritten digits and letters with the accuracy 98%. We are going to use Machine Learning algorithm such as Backpropagation. Neural Network Algorithm to build a system which can deal with this computer vision problem with higher accuracy. In addition to that we are going to implement Multiplier Feed Formal Neural Network and Support Vector Machine algorithms to recognize different other patterns. In order to achieve more accuracy, we will

focus on different Image processing technique which will help our algorithm to produce more accurate results.

Large number of handwritten datasets will be used to train our system. After that, our trained system with the help of Different Neural Networks Algorithm and other machine learning algorithms will be able to recognize future complex handwritten problems.

The classification stage is the decision making part of a recognition system and it uses the features extracted in the previous stage. A feed forward back propagation neural network having two hidden layers with architecture of 54-100-100-38 is used to perform the classification. The hidden layers use log sigmoid activation function, and the output layer is a competitive layer as one of the characters have to be identified. The feature vector is denoted as X , and defined as $X = (f_1, f_2, \dots, f_d)$, where f denotes features and d is the number of zones into which each character is divided. The number of input neurons is determined by length of the feature vector d . The total numbers of characters n determines the number of neurons in the output layer. The number of neurons in each hidden layer is obtained by trial and error.

The output of i^{th} layer is given by

$$a^i \log \sigma(w^i a^{i-1} + b^i) +$$

Where,

$$i = [1, 2, 3] \text{ and } a^0 = \mathbf{P}$$

w^i = weight vector of i^{th} layer

a^i = output of i^{th} layer

b^i = bias vector for i^{th} layer

IV. BLOCK DIAGRAM

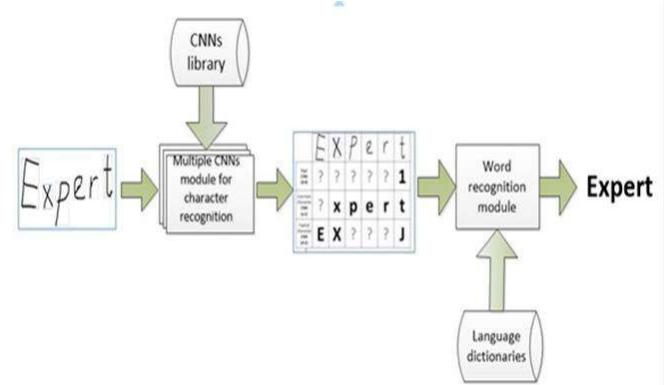


Fig 2: Block diagram proposed character recognition system

Our system consists of preprocessing, segmentation, feature extraction, classification and recognition, and post processing stages. The first important work is done by the scanners. They will scan the handwritten documents such as letters, claim papers etc. and segmentation and feature extraction work will be done. Preprocessing followed by segmentation and feature extraction. Preprocessing includes the steps that are required

to shape the input image into a form suitable for segmentation in the segmentation, the input image is segmented into individual characters and then, each character is resized into $m \times n$ pixels towards the extracting the features. The multi-class classification and the backpropagation algorithms will be used to recognize the input data given by the previous stages.

V. IMPLEMENTATION

Based on the preceding researches, literature survey and by using the neural network algorithms we have executed implementation of our software.

1	4	9	9	5	3	2	8	6	1
6	7	1	1	9	7	2	3	6	5
2	4	9	5	6	1	9	2	1	0
1	0	7	0	3	1	1	2	0	8
3	1	0	4	0	5	2	9	3	9
3	7	8	4	4	7	4	2	5	6
5	7	1	4	9	8	4	1	8	3
2	3	6	6	2	9	9	2	2	4
5	1	5	4	7	3	4	7	4	4
9	3	4	9	6	9	2	0	5	0

Fig 3: Software Implementation

The above table is the dataset of digits. We have obtained this dataset through preprocessing. Using this dataset we have trained our system.

```

81 - pred = predict(Theta1, Theta2, X(rp(i),:));
82 - fprintf('\nNeural Network Prediction: %d (digit %d)\n', pred,
83 -
84 - % Pause
85 - fprintf('Program paused. Press enter to continue.\n');
86 - pause;
87 - end
88
89
```

Command Window

New to MATLAB? See resources for [Getting Started](#).

Program paused. Press enter to continue.

Loading Saved Neural Network Parameters ...

Training Set Accuracy: 97.520000

Fig 4: Snapshot-1

The above snapshot gives the accuracy of our system. Using the dataset we have achieved 97.52 % accuracy as per our goal.

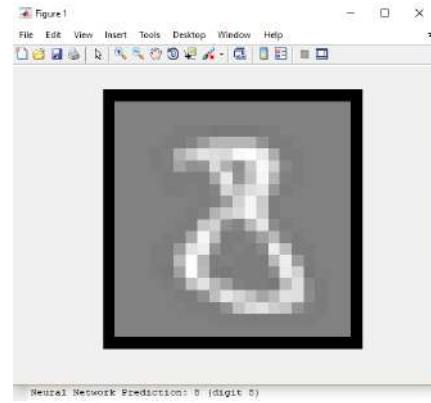


Fig 5: Snapshot-2

After training our system, we have worked upon the test dataset. We ran the test for digit 8 and we were able to achieve the desired result successfully. The above snapshot shows the test for digit 8

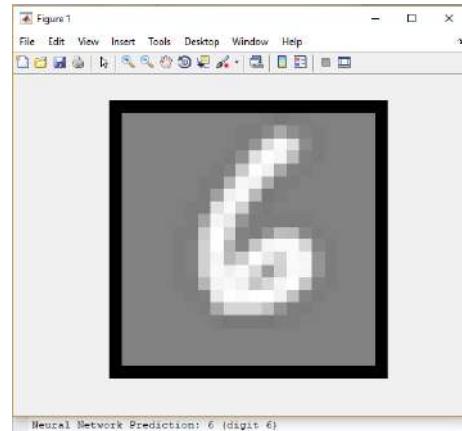


Fig 6: Snapshot-3

Similarly, our software was able to provide the accurate result for the digit 6 as well. The above snapshot shows the test for digit 6.

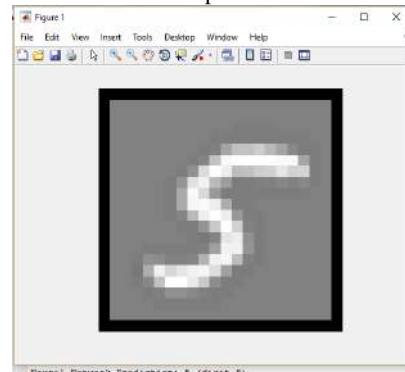


Fig 7: Snapshot-4

Also, our system provided accurate result for the digit 5 too. The above snapshot shows the test for digit 5.

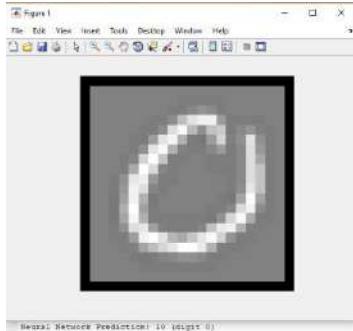


Fig 8: Snapshot-5

The above snapshot shows the test for digit 10. For the digit 10 our system was inaccurate and gave the result as 0. We are currently working on rectifying this bug in our software.

VI. CONCLUSIONS AND FUTURE SCOPE

We have reviewed various research papers and done a literature survey on various papers and found out the gaps in the existing

system. We have also developed various documents required for the implementation of the project.

Most of the private and government sectors still use the traditional way of checking and verifying the handwritten patterns. This method decreases the work efficiency and consumes more time. Introduction of handwritten pattern recognition system in these sectors will help to convert the handwritten document into file format and it will be easy to verify the documents. This will boost time as well as work efficiency. If we connect our system to a dictionary server of different languages, our system would be able to predict the meaning of scanned texts in its respective language.

REFERENCES

- [1] A. Gattal, "Isolated Handwritten Digit Recognition Using oBIFs and Background Features", 12th IAPR Workshop on Document Analysis Systems, at 305-310, April 2016.
- [2] R. BABU, "Novel Technique for Handwritten Digit Recognition Image Feature Extraction for Recognition", IJVIPNS-IJENS Vol: 15 No: 06.
- [3] Bouchra El jacimy, "Handwritten digit recognition based on DCT features and SVM classifier", 10.1109/ICoCS.2014.7060935, 16 March 2015.
- [4] Mamta Garg, "A Novel Approach to Recognize the off-line Handwritten Numerals using MLP and SVM Classifiers", IJCSET13-04-07-091
- [5] J Pradeep, "MLP (Multi-layer Perceptron)", P. E. Ajmire, IJPRET, 2015; Volume4(4):1-10
- [6] Bouchra El jacimy, "Handwritten digit recognition based on DCT features and SVM classifier", IEEE:14985664, 16 March 2015.

Commodity Search Engine for Online Shopping Using Web Datasets

Saurabh Rai, Varun Shah, Drishti Sisodiya, Riddhi Vaidya, Shiwani Gupta

Department of Computer Engineering, Thakur College of Engineering and Technology
saurabhrai251@gmail.com, varunshah1218@gmail.com, drishti255@gmail.com,
riddhivaidya16@gmail.com, shiwani.gupta3005@gmail.com

Abstract—With the popularity of Internet and e-commerce, the number of shopping websites has rapidly increased on the Internet, and this enables people to shop easily through the Internet. Consumers spend a lot of time searching commodity, because they need to filter and compare search results data by themselves. In recent years, there is some growing parity websites helping consumers to buy cheaper commodity. Although these websites can help consumers get the parity price of commodities, the search results are not so ideal. Because these websites may occur problems about the difference commodity between search results and consumers want to search, or the difference commodity price between search results and commodity web page. Therefore, this study attempts to use Keyword Searching and datasets as a basic approach. This study proposes a novel commodity search system to track consumer demand, and that is, when the commodity price of any website is lower than the consumer price conditions, the system will proactively notify consumers. This study results indicate that the novel commodity search system could assist consumers to search commodity and provide historical price information of commodity for consumers to decide. The aim of the project is to provide hassle free shopping based on available products recommendation to the user at efficient cost from the available datasets of products.

Keywords—Chat-Bot, Commodities, Consumer, Web Datasets.

I. INTRODUCTION

Online shopping is a form of electronic commerce which allows consumers to directly buy goods or services from a seller over the Internet using a web browser. Consumers find a product of interest by visiting the website of the retailer directly or by searching among alternative vendors using a shopping search engine, which displays the same product's availability and pricing at different e-retailers. With the popularity of Internet and e-commerce, the number of shopping websites has rapidly increased on the Internet, and this enables people to shop easily through the Internet.

Consumers spend a lot of time searching commodity, because they need to filter and compare search results data by themselves. Consumers find a product of interest by visiting the website of the retailer directly or by searching

among alternative vendors using a shopping search engine. Once a particular product has been found on the website of the seller, most online retailers use shopping cart software to allow the consumer to accumulate multiple items and to adjust quantities, like filling a physical shopping cart or basket in a conventional store. A "checkout" process follows in which payment and delivery information is collected, if necessary. Some stores allow consumers to sign up for a permanent online account so that some or all of this information only needs to be entered once. The consumer often receives an e-mail confirmation once the transaction is complete. In recent years, there is some growing parity websites helping consumers to buy cheaper commodity. Although these websites can help consumers get the parity price of commodities, the search results are not so ideal. Because these websites may occur problems about the difference commodity between search results and consumers want to search, or the difference commodity price between search results and commodity web page. Therefore, this study attempts to use Keyword Finding and Web Datasets technique as a basic approach.

This study proposes a novel commodity search system to track consumer demand, and that is, when the commodity price of any website is lower than the consumer price conditions, the system will proactively notify consumers. This study results indicate that the novel commodity search system could assist consumers to search commodity, and provide historical price information of commodity for consumers to decide whereas II. LITERATURE REVIEW

At current there are System, for example Haptik where the user logs-in chat with the system and at the backend there is a person sitting at another end who replies to all the queries whereas here the reply will be given as per the availability and datasets and in case of haptik there should be a person available 24x7 to reply queries but in our case there is no such requirement^[1] work on opinion mining of mobile phone reviews on E-Commerce site: Flip kart. The paper presenters have also done a feature-based classification of reviews. The objective is to benefit the customers and assist them in choosing the right product. This in case^[2] attempts to use Semantic Analysis, Ontology, and A.I as a basic approach. Which will analyze commodity prices, provide

the relevant information to consumer's reference, will aid consumers to search commodity, long term track for consumer demand, and proactively notify track results. In this the [3] presents an original commodity information search model, which integrates the semantic similarity computation based on semantic vector-space model.[4] The search model not only realizes semantic retrieval of commodity information, but also has the commodity selection and comparison mechanism, so as to provide the intelligent information search services for customers' online shopping.[5] This study mainly focuses on Online Product Recommendation affect consumer shopping efficiency, how consumer shopping efficiency influence consumers' loyalty and how product type moderates the relationship between quality of Online Product Recommendation and consumer shopping efficiency.[6] This paper presents our work on opinion mining of mobile phone reviews on E-Commerce site:Flipkart .A feature based classification of reviews has been done by us. The objective is to benefit the customers and assist them in choosing the right product. As future work we propo to offer a summary of reviews for more than two products and also automatically rank products based on the features that the user is interested in. [7]. The recommendation system is designed to recommends a clothing combination suitable to a specified situation and the user preference under the condition that every clothing item should be recommended with equal frequency by using Bayesian networks and the feedback of recommendation output evaluation by the user.

III. PROPOSED METHODOLOGY

At first, we will train our system by uploading dataset onto our system. The dataset will be containing data like Type of product, category, size, Age, Price. Based on these Factors the Systems will be trained. The System will be trained using Classification algorithm like NAIVE BAYES or ID3, etc. The Classification algorithm will generate a rule set onto the System. The User will provide input to the System like his product of his choice, category, size, age. Based on the product the System will predict the product and notify the User. Along with the Predicted Result i.e., Predicted the product, the System Will also suggest the related products or product of same variety etc. As our system also has Chabot which will interact with the user, it we'll have its own datasets which will be setting up at the file level.

5.1 SCOPE

This project mainly focuses on searching a commodity and finding best prices possible and enabling users to make quality decisions. The primary objectives of the project are as stated above, and the project aims at achieving all of them. With personalized search engines like Google and AltaVista which focused on searches based on keywords

this project follows the same principle of searching based on keywords. It is basically a single software with multimillion use, this can be used to compare prices, get better deals or to know more about the product on a single platform. Through the study it is seen that ecommerce business is growing rapidly than a conventional business, which helps to optimize costs and a hassle-free shopping experience. It can be kept updated with all the new products or commodities advancing in the market. The more data collected on the user, the more personalized result will be. As in today's time where everyone has less time to search over other websites, this will lead to less consumption of time for searching of product.

5.2 OBJECTIVE

The primary objective of the project is to create a one stop solution to find across all the ecommerce platforms and enhance the shopping experience of all the current users. Nowadays all the commodities are being searched before being purchased, therefore it is necessary that the search engine is well informed of all the products that the consumers want which is done using web mining. One of many objectives of the project is to reduce the time and effort of the consumer that he invests for searching a commodity on different platforms and provide a single platform to compare all the prices and quality across different ecommerce websites. Consumer gets an experience of hassle free shopping using a search engine. It makes use of web mining to search most relevant commodities according to a keyword. One of the most critical benefits is to improvise the quality of decision consumer makes. Using personalized search allows faster product and service discoveries for consumers and reduces the amount of necessary advertisement money spent to reach those consumers.

5.3 Process Model:

In our project as explained in incremental model, we are going to break to whole work into small modules. After breaking into modules, we are going to work on these modules which will increase the efficiency and faster execution of our project. Each module will pass through three phases of development (design, development, testing, and implementation). Planning, requirement gathering, and development will be done, after development of each module, testing of modules will be done to ensure that there are no bugs in the module. At the end implementation of each module will be done independently and after successful implementation all modules will be merged and the whole project will be released for customer feedback. Customer will give feedback and accordingly changes will be done if needed or else it will be released maintenance update from time to time. As we are implementing incremental model it will be easy to add any module in future. But before adding it will have designed, developed, tested and implemented. Since it is an incremental model,

iteration of each phase will take place based on customer's requirement. Every time a new version of the product will be released with new upgrades.

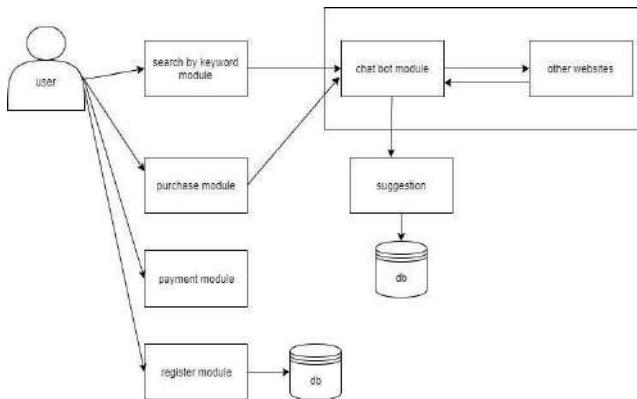


Fig 1. Working of System

5.4 Algorithm:

Algorithm for User side:

- User enters the statement.
- → if statement is a casual talk like hello,hi; then Chatbot will reply with a casual answer like hi,hello,etc.
- else if the statement is related to the products; the chatbot will tell the statement, process it and find the keywords related to the product.
- depending upon the keywords received after processing of statement, the Chatbot will search for the products related to those keywords.
- if products related to the keywords are found then it will be displayed on the front page of website.
- else the chatbot will show "No such products available".
- User can add the products to cart by clicking on add to cart button on any product whichever he wants to add to the cart button on any product whichever he wants to add to the cart.
- When the user adds all the products desired in the cart then from there he can further checkout with those products and place that order.
- When order is placed the mail will be triggered to the admin as well as the user(confirmation mail).

Fig 1.2 Algorithm for User Side

Algorithm for Admin side:

- Admin will enter the email id and password on login page.
 - the email id and password will then be checked with the entries in the database.
 - if the credentials are valid, admin is allowed access in the admin part.
 - else
 - admin will be asked to reenter the credentials
- Admin can now add the products which are to be displayed to the user for selling.
 - for adding the products, the admin will be provided with a form where he has to enter all the products details and submit the form.
 - on successful submission of the form the product will get added to the database.
- Admin will be provided with the page (view products)
 - on the view products, all the products from the database will be fetched using the MySQL query.
 - the admin has the rights to edit or delete the products before checking out.
- When admin clicks on the edit button in view products,
 - admin will be taken to the page which will have the form which was used for adding the products. Using this page admin can edit whichever parameter that needs to be changed.
- If admin clicks on delete on view products,
 - delete query will be executed on the product which will delete the product entry from database on successful deletion.
- Admin can view all the orders placed
 - all the orders placed by the user will be fetched from the order table database and will be displayed here.
- When the admin clicks on logout all the session will be destroyed and the admin will be out directed to the login page

IV. EXPECTED RESULTS

The User will provide input to the System like his product of his choice, category, size, age. Based on the product the System will predict the product and notify the User. Along with the Predicted Result i.e., Predicted the product, the System will also suggest the related products or product of same variety etc. Hence the System will generalize the searched commodity by the consumer and filters out the resultant product. The chatbot system based on artificial intelligence senses the product searched and pitches the related field products, consumable items to the user.

V. CONCLUSION

The following project will give us the overview of the different commodities searched online with the help of AI chat-bot. It also gives us the related key searches of the commodities online based on the web datasets. The search model not only realizes semantic retrieval of commodity information, but also has the commodity selection and comparison mechanism, to provide the intelligent information search services for customers' online shopping. The experimental results show that the algorithm presented in the paper is superior to the traditional algorithm of semantic information retrieval and is more suitable for customers' shopping needs. Based on present research work, we will improve the search model and make it to provide more accurate information search services for customers' shopping in the internet.

VI. REFERENCES

- [1] Venkata Rajeev P, Smriti Rekha V: Recommending Products to Customers using Opinion Mining of Online Product Reviews and Features ;2015.
- [2] Ming-Hsiung Ying*, Yeh-Yen Hsu: A Commodity Search System for Online Shopping Based on Ontology and Web Mining; 2014, International conference on software Intelligence Technologies and Application.
- [3] Lin Yu-Chu Yuusuke, Kawakita Etsuko, Suzuki Haruhisa Ichikawa Personalized Clothing-Recommendation System based on a Modified Bayesian Network; 2012 IEEE/IPSJ 12th International Symposium on Applications and the Internet.
- [4] Lin, Q. X., Chang, C. H., & Chen, C. L. (2010) "A Simple and Effective Closed Test for Chinese Word Segmentation Based on Sequence Labeling" Computational Linguistics and Chinese Language Processing, Vol. 15, No. 3-4, pp.161.-180
- [5] J. Liu, G. Z. Hu*, Y. Yu, W. J. Yi*, L. L. Zuo*: Research on the Effect of the Recommendation System on Customer Online Shopping Experience.;2010, International conference on software Intelligence Technologies and Application.
- [6] Ziming Zeng: Commodity Information Search Model of E-Commerce Search Engine Based on Semantic Similarity and Multi-Attribute Decision Method; Vol. 5, No. 7; July 2010 136 ISSN.
- [7] Zhang gong-jie & Huang sui. (2008). Research and Implementation of Semantic Indexing based on Ontology. Computer and Engineering and Design, 29(8): 2078-2080.

Monitoring And Assessment of Daily Activities Using IOT and Machine Learning

Aditya Sarvankar¹, Jonathan Kumar², Omar Dalvi³, Shailesh Sangle⁴

Department Of Computer Engineering, Thakur College Of Engineering & Technology, Mumbai, Maharashtra

sarvankaraditya@gmail.com, jonathankumar.chelsea@gmail.com, omardalvi@gmail.com, sss.sangle@gmail.com

Abstract— The numbers of fit bands and other IoT devices such as sleep trackers etc. have risen exponentially. With the amount of data now available through the means of these devices about people from all walks of life has risen greatly too. All of the daily activities of a person amount to something so there needs to be a pattern to the amount of data collected by the means of these various devices such as a sleep tracker and a fitness band.

Currently there are applications which collect the data for the user. We are making an application which will use this data and assess it and find patterns in it using ID3 decision tree algorithm in supervised learning. The assessed data will be collected in a database and use it as training data and tests will be run on it to find patterns, so that the machine knows under what circumstances a user is classified as being healthy, and under what circumstances, a user is classified as being not healthy. We further aim to suggest certain changes that the user can make to his daily lifestyle, be it sleep time, steps walked, or calories consumed and burned, so that he lives a more healthy lifestyle. Such an application and assessed data can be useful to various institutions, fitness companies etc, or even to any individual to lead a better life.

Keywords—Machine Learning, clustering, supervised learning, patterns, prediction, IoT.

I. INTRODUCTION

The main aim of this project is to develop an application which uses the collected data and assesses it to find patterns in it. Mental stress is one of the growing problems of the present society. The number of people experiencing mental stress is increasing day by day. Stress is a response of our body to prepare itself to face difficult situations. When a person goes under stress, his nervous system responds by releasing stress hormones. These hormones make our body ready for emergency actions. In certain situations it becomes dangerous and can put a person in serious mental disorder. Long term effects of stress can be chronic. Chronic effect of the stress causes health problems like hypertension [1], cardiovascular diseases [2] and memory problems. The sense of loneliness and hopelessness may lead people to suicide. People might be a little less likely to notice whether they are under any kind of stress or may be generally less sensitive to stress. Stress detection technology will help people understand and relieve stress by increasing their knowledge of unusually high or low

heartbeats that would otherwise go undetected. For this objective we have designed a smart band device in order to detect different conductance levels of the skin and predict whether the person is under stress or not. But skin conductance alone cannot accurately predict the stress level in everyday activities. Physiological responses caused by stress can also be provoked by physical activities like running, lacking of sleep etc. In order to accurately measure the stress level, classification should be made [3]. The fit band is capable of detecting heartbeat by analyzing different parameters in accordance with skin conductance like activities tracking, sleep quality etc. The collected data is then transmitted to user's smart phone via Bluetooth and upload to web from where it is accessed to find patterns to further ease the user experience[8]. The main idea of this assessment is that it will help the user to know whether he is indeed living a healthy lifestyle or not, as living a healthy lifestyle is a very vital need of the present. If the user is not being classified as healthy, he will be provided certain changes that he can make, so as to come closer to a healthy lifestyle. Even if the user is classified as healthy, he will get certain possible improvements, so as to live a healthier lifestyle.

II. PROPOSED METHODOLOGY

A. Skin Conductance

Skin conductance is considered as a biomarker for stress in which eccrine sweat activity that is controlled only by sympathetic nervous system is measured. Variation in skin conductance is in relation with sweat secretion. When a person is under stress, this stress puts the sympathetic nervous system into action. Since sweat glands are controlled by sympathetic nervous system so it activates the activity of the sweat glands. Sweat secretion from sweat glands reduce the skin resistance and increase the skin conductance level. Thus skin conductance acts as indicator for sympathetic activation due to stress reaction and can be used for stress measurement. Along with SC, heart rate and skin temperature can also be used as indicator for sympathetic nervous system activities[4]. The heartrate measured by the fitbit device is usually pretty accurate, especially at rest. There might be a little errors, but it is nothing major that it will produce different results.

B. Tracking sleep

Any fitbit device uses accelerometers to track your movements, including the speed and direction of your motion. This is how they track your activity during the day, and how they tell when you're asleep. When you set your device to "sleep mode," it monitors your movements. When you sync your device the next morning, the software will translate those movements into sleep data. If you may have ever edited your sleep times the next day, you'll have seen a glimpse into how this works—data about your movement during "awake" time gets re-analysed when you notify the software that you were indeed asleep then, and you then see it charted as sleep data. This method of using a device to track any kind of movements in order to measure sleep is called actigraphy. Actigraphy is often done in sleep studies using an "actigraph" device—similar to a Fitbit or Jawbone UP, it's usually a device worn on the wrist that tracks movement while you're sleeping. Software will then translate these different movements into periods of awake and periods of sleep. For people with any kind of sleep disorders or general sleep disruptions, actigraphy is a very easy way to have their sleep patterns studied by a doctor without having to sleep in the clinic, or any kind of lab. An actigraph device can be worn anywhere, including at home, and isn't usually too uncomfortable—most of them look like a small lightweight watch. An actigraph can also be worn 24/7 and track sleepiness during the whole day, based on different movements. Thus, actigraphy might be used for sleep studies where convenience is important. An example of tracked sleep time is shown below:



Figure 1 : Example of sleep data

C. Collecting Training Data

The first step will be to collect a large sample of data. The data will be extracted from the web using any fitbit. A large amount of data is available online for use. This training data will include a large variety of users, of different age, weight, height. All of their daily activities will be collected and stored. The data of each user

will be aggregated and there will only be one tuple for each user in the database. The daily activities of each user will vary, depending on his lifestyle. The activities will include heart rate, steps walked, calories burned, calories eaten, and sleep time. The machine will then be trained using ID3 algorithm. The machine will know based on what parameters a user is generally classified as healthy or not healthy. The larger the training data set, the more accurate the results will be for each user. Therefore it is much better to use as big of training data set as possible. Also the training data set should cover different types of people. It should include young, middle aged and old people. It should include people playing sports and people doing desk jobs. It should include people sleeping for 8-10 hours, and people struggling to sleep even 4-5 hours. Only when the data is vast and varied, the data for any new user can be predicted more accurately.

C. Assessing Monitored data

The first step for assessing the data would be scraping the Fitbit Site: for "sleep quality" data. If you try to log in to the Fitbit website, you will see that they indeed allow exporting data related to sleep duration, steps walked and some other data. However, crucial data like heart-rate during activities, number of movements during the night, duration of restless sleep, etc are completely missing. However, we can use the logic to scrape other sites in a similar fashion since we will be using our login credentials. Aggregating downloaded data: We may also download data that is freely available on the internet itself and then join them together, selecting only the data that we need. This step is very important because in the real life, data is rarely found in a single place or a single repository. Data cleansing, derived variables and other processing steps will need to happen in this section.

Hypothesis testing: In this section, we will try to gain knowledge about what factors affect sleep quality in general. Does it actually depend on any movements during the night, is there better sleep on different days, say weekend nights, etc.? Does working out more increase sleep quality?

We may combine the data scrapped from the web, heartrate and steps datafiles. We will have data for say, 2 months regarding the following variables:
 -sleep duration / start/ end time, sleep quality
 -number of movements during the night, number of times awake, duration of both.
 -Calories_burnt/ day, number of minutes performing light/ moderate/ heavy exercise.
 -weekday, date, month.

Below is an example of extracted data:-

	date	sleepDuration	awakeCount	restlessCount	awakeDuration	restlessDuration
1	2016-11-01	459	1	23	1	31
2	2016-11-02	511	3	30	13	50
3	2016-11-03	452	1	27	1	37
4	2016-11-04	447	0	1	0	1
5	2016-11-05	86	3	0	3	0
6	2016-11-05	250	1	9	2	12
7	2016-11-06	338	9	0	10	0
8	2016-11-06	155	0	0	0	0
9	2016-11-07	511	1	25	1	34
10	2016-11-08	433	1	20	4	29
11	2016-11-08	132	0	5	0	6
12	2016-11-09	292	6	0	6	0

Figure 2: Example of extracted sleep data

database	Date	Calories.Burned	Steps	Distance	Floors	Minutes.Sedentary	Minut
10	2016-11-08	11/8/2016	1543	5806	2.81	0	1365
11	2016-11-08	11/8/2016	1543	5806	2.81	0	1365
12	2016-11-09	11/9/2016	1860	4346	1.95	1	804
13	2016-11-10	11/10/2016	2119	4632	2.08	0	768
14	2016-11-11	11/11/2016	2310	11513	5.31	6	574
15	2016-11-12	11/12/2016	2353	10492	4.72	6	642
16	2016-11-12	11/12/2016	2353	10492	4.72	6	642
17	2016-11-13	11/13/2016	1728	2946	1.33	0	704
18	2016-11-14	11/14/2016	1993	5085	2.28	0	766
19	2016-11-15	11/15/2016	1981	3937	1.77	1	751

Figure 3: Example of combined data

Serial No.	Age	Weight(kg)	Height(cm)	Body Mass Index(BMI)	Sex	Steps walked	Calories consumed	Calories burned	HeartRate	Sleep time(hours)
1	18	60	160	23.4	Male	4500	2200	2000	60	7
2	30	70	165	26.1	Male	5500	2500	2300	70	8
3	22	65	163	24.5	Male	7500	2800	2650	75	7
4	21	58	156	21.8	Female	4200	1830	1850	68	5
5	22	75	175	24.5	Male	8500	2500	2650	72	7.5
6	24	56	158	22.4	Female	7650	1800	1850	65	6.5
7	28	58	168	21.5	Female	5850	2050	1750	75	7
8	40	80	175	26.1	Male	4500	2100	1950	78	9.5
9	19	50	165	23.4	Male	8500	1900	2200	68	6
10	18	52	165	23.4	Female	7800	1600	2150	79	5.5

Figure 4: Daily data of each user aggregated

The main aim of this project is to develop an application which uses the collected data and assesses it to find patterns in it. This can be done by collecting a large sample of data and using it as training data by using supervised learning methods such as ID3 decision tree algorithm. The ID3 algorithm begins with the original set S as the root node. On each iteration of the algorithm, it iterates through every unused attribute of the set S and calculates the entropy H(S) (or information gain IG(S)) of that attribute. It then selects the attribute which has the smallest entropy (or largest information gain) value. The set S is then split by the selected attribute (e.g. age is less than 50, age is between 50 and 100, age is greater than 100) to produce subsets of the data. The algorithm continues to recurse on each subset, considering only attributes never selected before. Recursion on a subset may stop in one of these cases:

- Every element in the subset belongs to the same class (+ or -), then the node is turned into a leaf and labelled with the class of the examples
- There are no more attributes to be selected, but the examples still do not belong to the same class (some are + and some are -), then the node is turned into a leaf and labelled with the most common class of the examples in the subset
- There are no examples in the subset, this happens when no example in the parent set was found to be matching a specific value of the selected attribute, for example if there was no example with age ≥ 100 . Then a leaf is created, and labelled with the most common class of the examples in the parent set.

Throughout the algorithm, the decision tree is constructed with each non-terminal node representing the selected attribute on which the data was split, and terminal nodes representing the class label of the final subset of this branch.

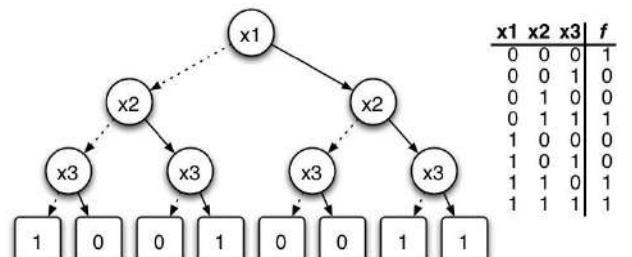


Figure 5 : ID3 Algorithm

ID3 does not guarantee an optimal solution; it can get stuck in local optima. It uses a greedy approach by selecting the best attribute to split the dataset on each iteration. One improvement that can be made on the algorithm can be to use backtracking during the search for the optimal decision tree. ID3 can overfit to the training data. To avoid overfitting, smaller decision trees should be preferred over larger ones. This algorithm usually produces small trees, but it does not always produce the smallest possible tree. ID3 is harder to use on continuous data. If the values of any given attribute is continuous, then there are many more places to split the data on this attribute, and searching for the best value to split by can be time consuming.

III. RESULTS AND DISCUSSION

A basic application would require a large sample of data, collected over long periods, and with a vast variety of users, ranging from different age groups, and different genders, and different height and weight. Once we have collected this sample data, we can use ID3 to find patterns in it. Unless this data is large, ID3 might not provide a very accurate result. The learning algorithm requires a prior specification of the number of cluster centers. Randomly choosing of the cluster center cannot lead us to the fruitful result. Using the above data, we use hypothesis testing

methods (ID3) to understand different kinds of patterns in the data that we have collected and stored. Once we run the code, we expect to observe the following results or something very similar to it:-

-Number of times awake increase when daily steps are between four thousand to seven thousand steps.
 -Weekends do not necessarily equate to better sleep, even though duration of sleep may be higher.

The user will be classified as healthy or not healthy, depending on his input data, which would include the sleep time, steps walked, calories eaten, calories walked and also his personal data such as age, gender, weight, height and BMI index, which is calculated using weight and height. We will also classify each user as being healthy or not. Once the machine is trained using this training data, the new user can enter his data, and the machine will be able to predict if the user is healthy or not. And the machine will also give certain changes the user can make to his lifestyle so that he lives a more healthy lifestyle

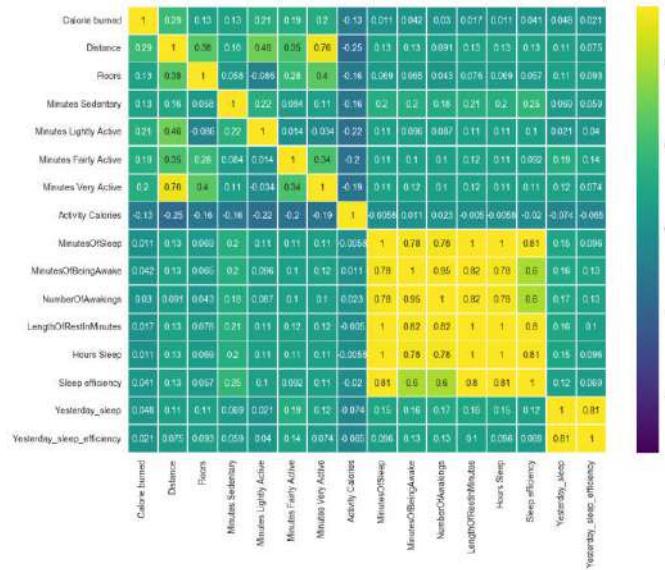


Figure 6 : Correlation Matrix

IV. CONCLUSION

Once this data is collected using the smartband and stored in the application using IoT, the algorithm is applied, we can find patterns, which will show that all the input data, i.e. the number of steps walked, the amount of sleep, the calories burned, and the heartrate will have a relation of some sort between them. This relation will be found out by using the ID3 decision tree algorithm. All this information and data and results will be stored and the user can view them anytime. According to the results, the user can change or alter his input data, so that he can gain favorable results

V. REFERENCES

- [1] A. Ste Ptoe, "Invited review: The link between stress and illness," *J Psychosom Res*, Vol. 35 no. 6.
- [2] C. Schubert M. Lambertz, R. Nelesen, W. Bardwell, J.-B. Choi, and J. Dimsdale, "Effects of stress on heart."
- [3] K. Elissa, "Mental stress detection using physiological soft computing techniques".
- [4] R. K. Dishmans, Y. Nakamura, M Dunn, and S. N. Blair, Heart vascularity and diseases.
- [5] Book by Christopher Bishop , "Pattern recognition and machine learning".
- [6] Muhammad Zubair, Changwoo Yoon, Jisu Kim, Jinsul Kim, "Smart Wearable devices for stress detection" in IEEE conference.
- [7] Book by John Slavio, "Machine Learning for beginners".
- [8] Book by Adrian McEwen and Hakim Cassimally, "Designing the internet of thing

Home Automation System using Raspberry Pi

Uday Menon, Rahul Yadav, Shubham Singal, Kaustubh Shah, Vikas Singh

Department of Computer Engineering, Thakur College of Engineering and Technology, Mumbai, India

udaymenon@gmail.com, rahul.y494@gmail.com, shubhamrsingal@gmail.com, kp.shah4595@gmail.com,
vikas.singh@thakureducation.org

Abstract—Due to the exponential increase in population there is an increase in consumption of electricity, there is a dire need to conserve electricity in every way possible. The reason why people are unable to manage energy efficiently is that they are not privy to the energy that they consume. Most of the wastage of electricity is caused due to negligence. The inability to access and control the appliances from remote locations is one of the reasons for unnecessary energy wastage. A web or an android application is used by the users to manage and control the systems.

Keywords— Energy management; Raspberry Pi; Mobility; Electronic appliances; Smart Home Automation; Web application; Android application

I. INTRODUCTION

Houses of the 21st century will become extremely smart and safe, especially when deployed in a private home. A home automation system is a means that allows a user to control the power supply to any electric appliance. Many of the existing home automation and smart home systems in the market are mostly based on wired communication. This does not cause any issue if the system is planned well in advance and the incorporated into the electrical structure of the building during the physical construction of the building. But for pre-existing buildings the implementation cost goes up exponentially. In contrast, wireless systems can be of great advantage for automation systems as well as internet enabled systems. With the advancement of wireless technologies such as Wi-Fi, cloud computing and cellular network in the recent past, wireless systems are used on a daily basis and implemented everywhere.

Electronic and Electrical surroundings with relation to this context is any environment that consists of appliances like fans, television sets, air conditioners, motors, heater, lighting systems, etc. A remotely accessible environment is an environment within which every aspect of it can be remotely managed using the software as an interface. The software is a combination of both web application and android application. Such remotely accessible systems are already on the market, however they have a variety of drawbacks. This paper will provide the proposed system that is an improvement on the existing systems in the market.

II. LITERATURE SURVEY

In the technical paper Vinay Sagar and K N, Kusuma S M, "Home Automation Using Internet Of things", International Research Journal of Engineering and Technology (IRJET), they have written about the project of HAS using IoT wherein they have used Intel Galileo Microcontroller. They have used

Wi-Fi as a medium for internet connectivity. They have even created experimental setup of HAS where they have controlled two lights, a cooler, and alarm for a gas leak. They have even connected gas, light, temperature, and a motion sensor and they have set a threshold for each sensor's reading using which they control the appliances. For example- A cooler will turn on when the room temperature exceeds the set threshold and in turn reduces the room temperature. As a frontend part, they have created a HTML-based Web server page using which the user can handle the entire system.

Whereas, only switching on and off the appliances is just not enough for reducing the power consumption of the system. "Design and Implementation of a Wi-Fi based Home Automation System" by Ahmed Elshafee and Karim AlaaHamed, International Journal of Computer, Electrical Automation, Control and Information Engineering is Arduino based distributed HAS system which consists of a server, hardware and interface modules. The webserver controls hardware, a single interface module and can be configured to handle more than one hardware interface module. The hardware interface module controls all alarms and actuators. They have used Wi-Fi as a medium for connecting to the internet. In this the user can create group events i.e. when a certain trigger is activated, the device will perform a series of tasks that are preset by the user. These macros can be activated manually or as a reaction for certain trigger light motion sensors and surveillance cameras. Since they have used Arduino as the base of the system while adding more hardware interface modules we require number of Arduino chipset to be installed in the system which results in the increases of complexity and the overall cost of the system and also eventually increase the power consumption of the entire system.

The paper "Android based Home Automation Using Raspberry Pi" by Shaiju Paul, Aswathy B and Ashlin Anthony is very close to the paper we are proposing. They have created Home the HAS wherein they have simply connected the home appliances to the Raspberry Pi using relay circuit and they have used an android application for a user interface. Their system consists of three main components a Raspberry Pi board, Wi-Fi transceiver and a Relay circuit. Wi-Fi is used as a communication channel between android phone and the Raspberry Pi board. They have masked the complexity of the technicalities involved in the HAS by including them in a simple yet comprehensive set of related concepts. In this paper, they have not used sensors, due to which the project doesn't provide the base for an automated system. Only toggling the power supply to the appliances is just not enough for reducing the power consumption of the system.

The paper "Home automation using Raspberry Pi" by Monika M Patel, Mehl A Jajal and Dixita B Vataliya. In the paper, they have made use of a Raspberry Pi model B as their central processor, a Wi-Fi dongle for their internet connectivity, a relay circuit for connecting appliances to the system. A WebIOPi for IoT framework. Their relay circuit operates at a 5-12V output.

WebIOPi is an integrated IoT framework for Raspberry Pi. The system in this paper, when turned on, will first check the current situation of rooms and change in the state of lights. This system provides the user with remote control of various lights and appliances within their home.

In this system, the sensors are not used which gives us the information about the conditional intensity with which the appliances should work. They are not using the Android based application which is the current trending topic. Also, they are not manipulating the intensity with which the appliances will work, which could have reduced the power consumption greatly.

III. PROPOSED SYSTEM

A. System Architecture

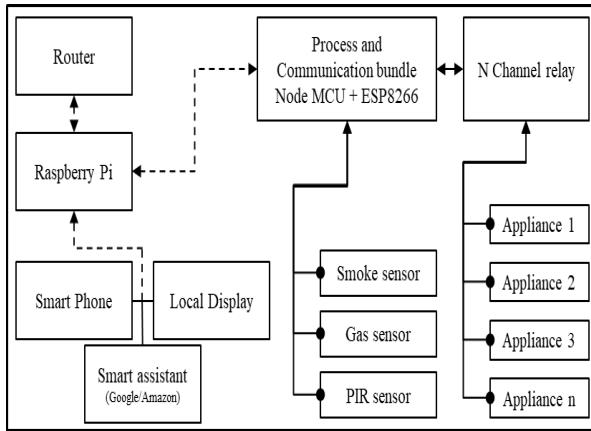


Fig 1: System Architecture

The proposed model of Home Automation System using Raspberry Pi is as shown in the figure above. The primary brain of the system is a Raspberry Pi, it is responsible to make decisions according to the input from the user or according to pre set parameters . The model consists of different sensors like gas sensor, temperature and humidity sensor and motion sensor. Initially the Raspberry Pi connects to the internet through Wi-Fi. When the connection is established it will start reading the parameters of the sensors. The threshold levels for the required sensors are set as per user preferences. The sensor data is updated in the table maintained in the web server. The data can be accessed and analyzed at any point in time. If the sensor parameters are greater than the threshold level then the user is notified so that the decision for required actuation is done for the controlling of the parameters. In the proposed model the temperature, gas leakage, motion in the house is monitored. The temperature and the motion detection are stored in cloud for analysis. If the temperature exceeds the threshold level then the cooler will turn on automatically and it will turn off when the temperature reaches the user set parameter. Similarly when there is a leakage of gas in the house, alarm is raised giving the alert sound. Lights are turned on/off automatically by detecting the intensity of light outside the house. The user can also monitor the electric appliances through the internet via web server. If the lights or any electrical appliances are left on, the user will be notified and they can turn it off remotely through the mobile application.

The system is also capable of working with modern virtual assistants, like the Google assistant and Amazon Alexa. These devices provide the user with a great degree of convenience. This

integration makes the whole operation of toggling of appliances as natural as a conversation. The integration also lets the user set their own custom routines, such as activating the coffee maker in the morning when they wake up. The audio feedback from the smart device adds to the convenience of not needing to recheck the status of a device.

B. Functions

The proposed home automation system has the capability to control the following appliances monitor the following alarms in the user's house:

- a) Temperature and humidity
- b) Motion
- c) Fire and smoke
- d) Light fixtures
- e) Power outlets
- f) Appliance connected to smart socket

The system can be controlled using multiple methods:

- a) Mobile application
- b) Web application
- c) Local display
- d) Voice command, through a virtual assistant like Amazon Alexa or Google assistant

IV. IMPLEMENTATION

The system is a combination of both hardware and software. Since the processing is done by a Raspberry Pi that is connected to the internet, updating of the system can be done via OTA (Over the Air) updates.

A. Hardware

The primary component is the Raspberry Pi. The Pi will contain the code that connects to the webserver to update the database with the sensor values as well as the current condition of the switches. If there are any changes made to the values in the table, the Pi takes the appropriate action. The Pi will also be connected to a display that will act as a digital switch board as well as a central console for the house.

The Pi is connected to the Node MCU Wi-Fi transceiver, this acts as a communication bridge between the Pi and the relay board. The MCU is responsible for controlling the relay board as well as sending the sensor data to the Raspberry Pi. The relay board has mechanical switches that carry out the physical switching. The existing wiring of the switch is connected to the relay board. The sensors are connected to the Raspberry Pi directly onto the board. The sensor values are updated at regular intervals to the webserver. These values are used to alert the user if they cross a certain threshold.

B. Software

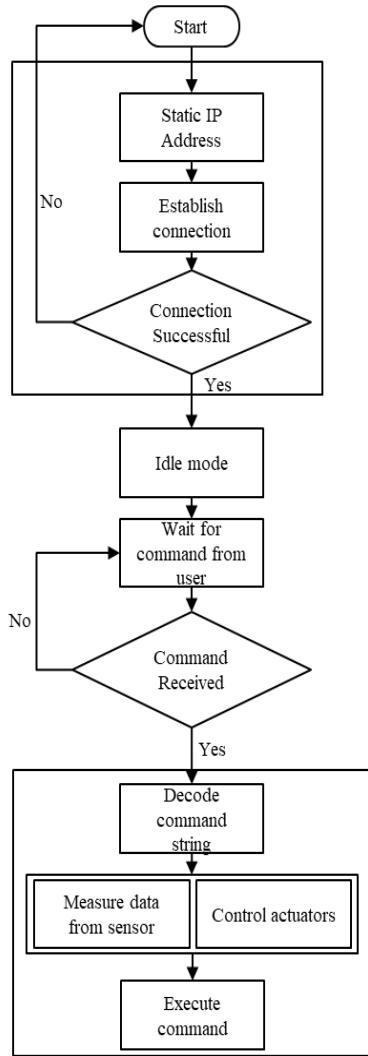


Fig 2: Flowchart

The above flowchart shows the steps involved in the initialization of the device as well as the steps involved in execution of a command.

The system after powering up will first establish a connection with the webserver. The user will then login and the Raspberry Pi will then access the particular user's database, where the sensor as well as switch values are stored and update them with the current values of the environment.

The system will then go into idle mode. It will wait for a command from the user, be it from a local network or through the internet. Once the command is received, the sensor data is measured for any clash with pre-set user parameters, if there is no clash then the command is executed.

The sensors values are updated in the webserver in regular intervals. If the values are out of the range set by the user, then the user is alerted so that the appropriate action can be taken to bring the situation under control. The PIR sensor is used to detect any type of motion, this is useful in tracking any unauthorized access when the user is not at home, thus improving the safety of the

house. The monitoring of movement also helps in curbing electrical wastage, if a particular room has no occupant but appliances are active, then the user is notified about the same.

The user can access the system via the web application, mobile application or via the touch console, which is an optional extra. On first use of the application, the user has to register themselves. On logging in they will have to then pair their account to the device. Once paired they can log into their account from anywhere and make changes to the system.

V. RESULT AND DISCUSSION

The device is connected to the existing wiring system. Once it is powered up it'll first connect to the user's mobile device. Then the user will be prompted to register the device to their account. Once connection has been established, the device will then take the current condition of the environment and update the database.

Once the user enters the mobile application, they will be welcomed with a login screen. After successful authentication the will then be brought to the home screen, this screen consists of the devices that the user has added to their profile. It can be configured to display the readings from the various sensors in the system like gas sensor, temperature and humidity sensor and any other additional sensors that the user opts to install.

The user can then see the status of their switches and the environment of their house. They can control the switches by simply tapping on the icons representing the respective appliance. The on-board current and voltage sensors send data periodically to the raspberry Pi. This data is stored on the database, this data can also be used to generate a rough estimate of the electricity bill and provide a detailed consumption analysis.

VI. CONCLUSION

The Home Automation using Raspberry Pi has been analytically proven to work adequately by connecting appliances to it and the appliances were successfully controlled remotely through internet. The designed system not only monitors the sensor data, like temperature, gas, motion, current and voltage consumption but also actuates a process according to the requirement, for example switching on the light when it gets dark or deactivating an outlet that has been active but idle. It also stores the sensor parameters in the webserver in a timely manner. This will help the user to analyse the condition of various parameters in the home anywhere.

REFERENCES

- [1] Baris Yuksekaya, A. Alper Kayalar, M. Bilgehan Tosun, M. Kaan Ozcan, and Ali Ziya Alkar "A GSM, Internet and Speech Controlled Wireless Interactive Home Automation System", 2006, IEEE Transactions on Consumer Electronics, Vol. 52(3) , pp. 837 - 843.
- [2] Rozita Teymourzadeh,Salah Addin Ahmed,Kok Wai Chan and Mok Vee Hoong , "Smart GSM Based Home Automation System", 2013, IEEE Conference on Systems, Process & Control, Kuala Lumpur, Malaysia.
- [3] Vinay Sagar K N, Kusuma S M, "Home Automation Using Internet OF things", International Research Journal of Engineering and Technology (IRJET), Volume: 2, Issue:03 June -2015, pp 1965 - 1970.
- [4] Ahmed Elshafee, Karim AlaaHamed, "Design and Implementation of a Wi-Fi based Home Automation System", International Journal of Computer, Electrical Automation,Control and Information Engineering Vol:6,No:8,2012,pp 1074 - 1080.

- [5] Shaijupaul, Ashlin Anthony, Aswathy B, "Android based Home Automation Using Raspberry Pi, IJCAT - International Journal of Computing and Technology, Volume 1, Issue 1, February 2014, pp 143 - 147.
- [6] Pawansingh,Krupachotalia,SanketPingle, " Review paper on SmartGSM based Home Automation System".International Research Journal of Engineering and Technology (IRJET), Volume: 3, Issue :04 April - 2016, pp 1838 - 1843.
- [7] Monika M Patel, Mehl A Jajal, Dixita B vataliya, "Home automation using Raspberry Pi, "International Journal of Innovative and Emerging Research in Engineering Volume 2, Issue 3, 2015.
- [8] Rahul Godha, Sneh Prateek, Nikhita Kataria "Home Automation: Access Control for IoT Devices" International Journal of Scientific and Research Publications, Volume 4, Issue 10, October 2014 ISSN 2250-3153
- [9] Vishwajeet H.Bhide "A Survey on the Smart Homes using Internet of Things (IoT)"
- [10] Faisal Baig, Saira Baig, Muhammad Fahad Khan, "Controlling Home Appliance Remotely through Voice Command", 2012, International Journal of Computer Applications, Vol. 48(17), pp.1 - 5.
- [11] Bader M. O. Al-thobaiti, Iman I. M. Abosolaiman, Mahdi H. M. Alzaharani, Sami H.. Almalki, Mohamed S. Soliman, "Design and Implementation of a Reliable Wireless Real-Time Home Automation System Based on Arduino Uno Single-Board Microcontroller", 2014, International Journal of Control, Automation and Systems, Vol. 3(3), pp. 11 - 15.
- [12] Rajeev Piyare "Internet of Things: Ubiquitous Home Control and Monitoring System using Android based Smart Phone" International Journal of Internet of Things 2013, 2(1): 5-11DOI: 10.5923/j.ijit.20130201.02.

Real Time Drowsiness Detection System

Bharat Rathore¹, Rishabh Rathod², Aakashkumar Yadav³, Abhishek Yadav⁴, Mr. Shailesh Sangle⁵

Department of Computer Engineering
Thakur College of Engineering & Technology

Abstract-The proposed paper describes how efficiently a non-intrusive computer vision based concept be used for detecting drowsiness of the Driver. A System will be developed which aims to improve the road safety using advanced technology. The proposed system will use a basic Web-cam interfaced with Raspberry-Pi that points directly towards the driver's face and monitors the driver's eyes and Face in order to detect fatigue. If Symptoms of Fatigue such as closed eyes or head lowering occurs a warning signal is issued to alert the driver. The product prototype proposed is unique to the road safety purpose. It uses the concept of image processing. Open-CV will be integrated with the python and deployed on raspberry-pi. Harr classifier algorithm used will help to determine if the eyes are open or closed. The algorithm developed is unique to any currently published papers in terms of its application with Raspberry pi, the primary objective of the project was to make the system portable and hence can be applied in existing vehicles to improve the safety by giving driver a feedback when he feels drowsy. The algorithm uses image binarization to mark the edges of the face. Once the face area is found, the computation of the horizontal averages in the area helps us locating the eyes. Eye regions in the face shows great intensity changes, the eyes are located by tracing the intensity changes on the face. Once the eyes are located, distance is measured between the intensity changes in the eye area which helps us to determine whether the eyes are open or closed. Large distance computed corresponds to eye closure. If the eyes are found closed for few consecutive frames, the system draws the conclusion that the driver is falling asleep and issues a warning signal. There is also a provision of monitoring the head region in the frame, similarly if for few consecutive frames if the head region is out of the frame also triggers a warning condition. The system is also able to detect when the eyes cannot be found and work under reasonable lighting conditions [5].

Keywords— *Face detection, eye detection, image binarization, Harr classifier and drowsy detection*

I. INTRODUCTION

Driver fatigue is a major factor for a large number of road vehicle accidents. Recent statistics reports that 1,200 deaths and 76,000 injuries are reported annually due to fatigue. The technology advancements for detecting and preventing drowsiness behind the wheel has been a major challenge in the domain of accident prevention systems. Because of the losses that drowsiness causes on the roads, methods are needed to be developed for

counteracting its effects [1]. The objective of the project is to come out with a prototype of drowsiness detection system. The aim will be to design a system that will monitor the real-time status of the driver's eyes accurately. By systemic monitoring of the eyes and head position symptoms of fatigue will be detected beforehand and early enough that would help avoiding a car accident. For detection of fatigue, analysis of image sequence of a face is done, and the movement of eye and blink patterns are determined.

Face images analysis is a research area with many applications such as face recognition, human identification security systems and virtual tools [7]. In This project we will mainly focus on the extraction of the eye region, this involves considering the entire image of the face, and then to determine the eye region by using a self-developed image-processing algorithm and Once the position of the eyes are located, the system will determine whether the eyes are opened or closed along with the position of the head and accordingly detect fatigue.

While developing a driver monitoring system, two issues such as driver fatigue measurement and distraction detection should be solved. And our proposed system solves both by monitoring the state of eyes we can trigger warning if driver has fallen asleep. By monitoring the head region in particular area algorithm can monitor if driver is alert or not.

II. LITERATURE REVIEW

Researchers have worked in recent times for detection of driver inattention, mainly focused on drowsiness. There have been lots of research on sleep in the field of psychology and medicine and accurate indicators of sleep have been developed [4]. Electrical change in brain is represented by Electroencephalograms (EEG) [5] with the help of electrode placed on scalp. Small voltages produced in the brain cortex are detected by these electrodes. Waves of serval frequencies known as alpha, beta, and gamma are formed due to potential which linked to various cognitive processes, drowsiness and sleep stages. With the help of electrooculography (EOG), brain studies couple EEG, which helps to detect eye movements, and electromyogram (EMG) monitors muscular tone. These are best data for drowsiness detection and many drowsiness detection systems make use of it. But they require electrode to place on drivers' head which can be annoying hence problematic.

EEG requires external help and takes a few minutes, and medical equipment is always expensive. Recent research has introduced some contact-less readings, but no remarkable results have been achieved so far. Nonetheless, physiological measures such as EEG have been used in some projects [6], and are frequently used as the ground truth for testing other, less invasive methods.

A driver's state of attention can also be characterized using indirect measurements and contact-less sensors. Lateral position of the vehicle inside the lane, steering wheel movements and time-to-line crossing are commonly used, and some commercial systems have been developed. These systems do not monitor the driver's condition, but its driving pattern and gives a result which is not so accurate. Hence, we have proposed this technique to monitor the driver's condition over driving pattern.

III. METHODOLOGY

Different types of methodologies has developed to find out driver drowsiness so as to prevent road accidents.

A) Physiological level approach- This technique is an intrusive method wherein electrode are used to obtain pulse rate, heart rate and brain activity information. ECG is used to calculate the variations in heart rate and also detect different conditions for drowsiness. The correlation between different signals such as ECG (electrocardiogram), EEG (electroencephalogram), and EMG (electromyogram) are made and then the output is generated whether the driver is drowsy or not.

B) Vehicle based approach- This technique continuously monitors the position of car in lane, steering wheel position and pressure on acceleration pedal. By measuring all this parameters system indicates whether driver is drowsy or not.

C) Behavioral based approach- In this technique eye blinking frequency, head pose, etc., of a driver is monitored through a camera and the driver is alerted if any of these drowsiness symptoms are detected.

IV. PROPOSED SYSTEM

Practically, Physiological based approach is not suitable for drowsiness detection as it is an intrusive method which is not user friendly for driver and for drowsiness detection. Vehicle based approach is based on monitoring the car instead of driver and hence it is difficult to determine whether driver is feeling sleepy or not.

The proposed system consist of three major components:-

A. Capturing frames: Camera mounted on the dashboard captures the images of driver's face including eyes and passes this data to processing component.

B. Processing and detecting component: Captured facial image is used to determine whether drivers' eye is closed or open. The driver's current eye state is determine by using HARR classifier and Viola Jones algorithm which is use for object and face detection.

C. Signaling: After processing the eye blink frequency decision will be made whether to give alarm to driver or not. If eyes is closed on more than 22 frames then alarm will be given

to driver. For signaling a buzzer will interfaced with Raspberry pi.

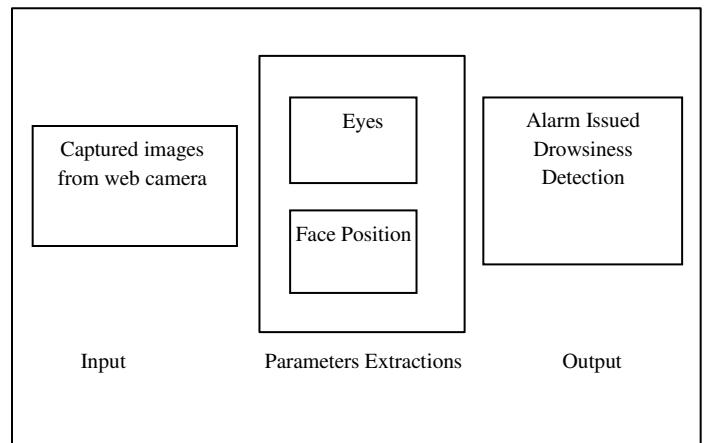


Fig.1. Block Diagram of the Proposed System

A. Block Diagram

System proposed by us will have open source 5 megapixel digital camera which will capture real time images of driver .We will send these captured images to a Raspberry- pi system board. . The Raspberry-pi system is loaded with Raspbian OS and Python packages for Open CV (Computer Vision

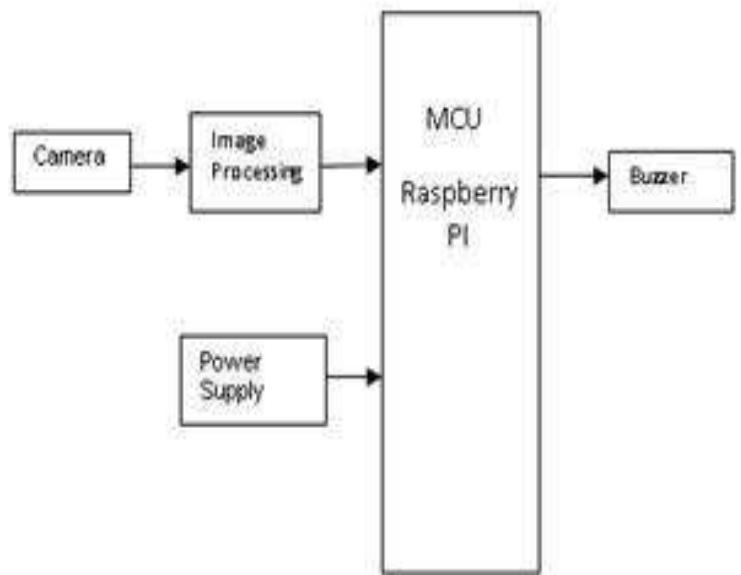


Figure 2. Block diagram of system

). As mentioned above we are more focusing on eye part of driver hence we will use HARR features. For edge detection of pupil and iris Hough transform is used.

We have a threshold value for pupils and iris area and compare it with current value to detect drowsiness. If it is more than the threshold value then condition for drowsiness is detected and warning is issued with help of buzzer. It is connected to PIC 16F controller and PIC controller is serially interfaced with Raspberry-pi system board.

B. Flowchart

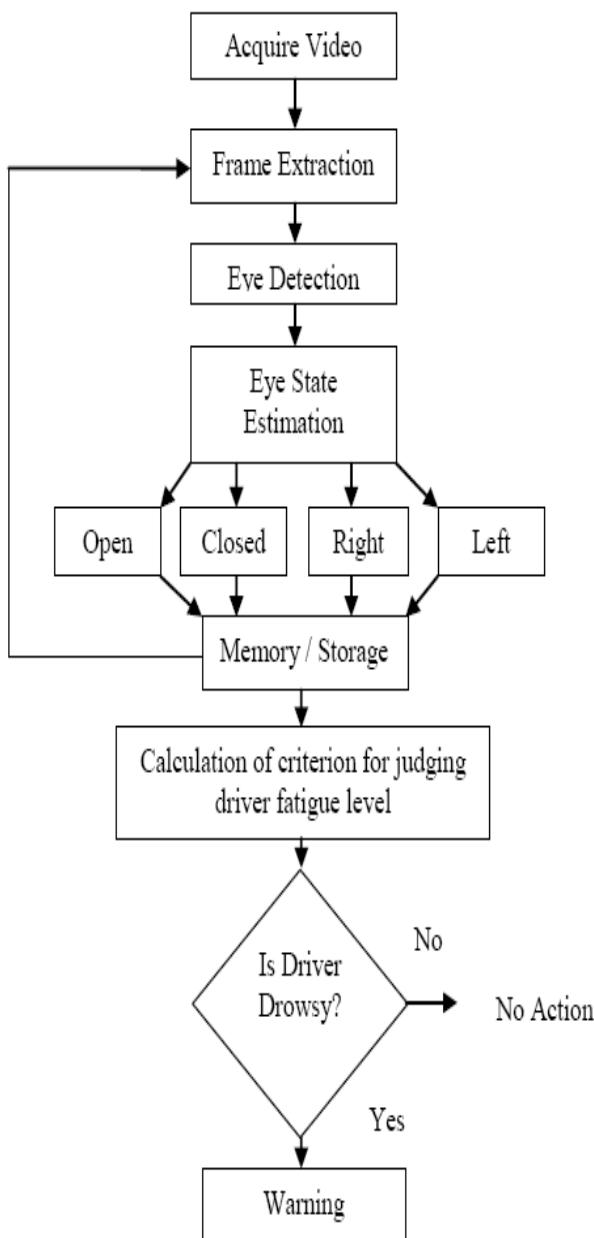


Figure 3. Flowchart[8]

V. SYSTEM REQUIREMENTS

- A. *Hardware Requirement*
 - I. Raspberry pi 3B
 - II. Buzzer
 - III. Camera
- B. *Software Requirement*
 - I. OS: Linux
 - II. Programming language: Python, open CV

VI. FUTURE SCOPE

In today's scenario we are discarding near about 20% of frame in estimation process. In this 20% frame neural network used for eye estimation do not achieve same state in both eyes hence these 20% frames are not used in drowsiness detection. We are currently using video segment at 30 fps and while using video segment we can accurately able to detect drowsiness but if we try to slow it down then it is not working accurately more short of hit and miss. That's why our future scope is to increase the frame rate so that we can make use of all the video frame segment. Each person has different eye hence this system can be developed as much reliable eye state database. Our aim is to detect drowsiness and in future this system will be integral part of safety system in vehicles and used to save many lives.

VII. CONCLUSION

We use eye closer rate to identify drowsiness. It acts as an indicator for drowsiness. Video captured by camera is used to extract frames and these frames are used as input to extract eye region with the help of eye region extractor. Eye region extractor help us to find eye region which are gray scaled, resized and histogram equalized. After this step, using neural network left and right eye regions are merged to estimate the eye state of subject. After completion of this step each video frame is tagged as open (0), semi-closed (0.5), closed (1) and no valid estimation .For performing valid estimation of eye state we have to take mean of eye states. For generating alert message to driver average value of eye state must be greater than threshold value.

REFERENCES

- [1] U.S. Dept. of Transportation, "Traffic Safety Facts 2006: A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System," tech. report DOT-HS 810 818, Nat'l Highway Traffic Safety Administration, 2006.
- [2] <http://www.lboro.ac.uk/departments/ssehs/research/behavioural-medicine/sleep-research-centre/>
- [3] Y. Liang, "Detecting driver distraction," Ph. D thesis, University of Iowa, 2009.
- [4] M. Bayly, B. Fildes, M. Regan, and K. Young, "Review of crash effectiveness of intelligent transport system," TRAffic Accident Causation in Europe (TRACE), 2007.
- [5] E. Rogado, J. Garcia, R. Barea, L. Bergasa and E. Lopez, "Driver Fatigue Detection System," Proc. IEEE Int. Conf. Robotics and Biomimetics, 2009.
- [6] S. Abtahi, "Driver Drowsiness Monitoring based on Yawning Detection," MS thesis, University of Ottawa,

- 2012.
- [7] IOSR Journal of VLSI and Signal Processing Vol. 4 – Jan2014. <http://sine.ni.com/cs/app/doc/p/id/cs->
IOSR Journal of VLSI and Signal Processing Vol. 4 – J2014.
- [8] <http://sine.ni.com/cs/app/doc/p/id/cs-13515->

Secured ATM Management Using Image Steganography

Mr. Piyush Akoliya¹, Mr. Sujay Salvi², Rushikesh Shinde³, Sakshi Singh⁴, Ruta Pathak⁵

Department of Computer Engineering,

Thakur College of Engineering & Technology,Mumbai,India

Abstract— Steganography originates from the Greek word "steganos"-covered or secret and "graphy" -writing or drawing. Steganography technology uses modern data compression as well as cryptography to write the secret message in a cover file to hide the existence of that data. Steganography in future could lead to be one of the most important technologies to provide internet security and privacy of data in an open environment such as the Internet. Steganography research has been encouraged over time due to the strength that the cryptographic systems lack and the desire to increase the level of secrecy in an open environment. Image Steganography is the technique of writing image in such a way that only the intended receiver and sender can suspect the very existence of the message.. Our proposed System will be using 3 level security that is after login with the 4 digit ATM password a 8-digit OTP would be sent to the user on his Registered mobile number once the user enters the 8 digit OTP the this OTP is encrypted and hidden in a Image. Thus, hiding its very existence. This helps to overcome the vulnerability of the 4 digit ATM pin being hacked using Brute Force attack by 9999 attempts. Thus the main objective of the proposed system is to provide more security, hence it gives more priority to security over time required for execution of a transaction. This way we can have a secure ATM transaction.

Keywords—Steganography, Advance Encryption Standard, Automated Teller Machine, One Time Password.

I. INTRODUCTION

In this paper we presented Least Significant Bit based Image Steganography. Least Significant Bit based image Steganography is a good method of embedding sensitive information behind some cover media. As the time required for a ATM transaction to complete should be minimal hence the combination of AES encryption along with Least Significant Bit based image steganography could prove to be a really good security model. AES is preferred over DES due to its simplicity and its speed. Security of money, data is the most important issue in the world today. Our project is based on security of ATM. The steganography and cryptography differ in the way they are evaluated: Steganography helps to hide the confidential message away from the enemy in such a way that the attacker does know that the message exist and cryptography helps to prevent attacker from reading the text by encoding it. In this project we will stimulate an ATM machine where dynamic computerized generated password are provided to user transaction. As soon as

user enter their account number a new dynamic computer generated password will be sent to user's mobile phone on the number which is registered at the bank. The user then enters the OTP is the computer generated code into the machine this code is then encrypted using AES encryption and then is encrypted data is then hidden inside an image which is then used to perform the transactions or is transmitted over the network. After the OTP is verified successfully the user may access his/her account. Once the user logs out of his account the OTP is destroyed. Thus the main objective of the proposed system is to provide more security, hence it gives more priority to security over time required for execution of a transaction.

This way we can have a secure ATM transaction. In this paper, we proposed the combination of cryptography and steganography has been achieved by using the AES algorithm and LSB technique. Advance encryption standard (AES) is used to encrypt the confidential data and the LSB technique is used to hide encrypted secret image into cover image. To yield better imperceptibility the proposed method provides a higher level of similarity between the cover and stego pictures as a result. The combination of Image Steganography along with Encryption proves to be a good security between two parties in case of secret communication and to prevent the attraction of eavesdroppers. Finally we can conclude that the proposed technique is effective for secret data communication.

II. COMPARATIVE ANALYSIS OF EXISTING TECHNOLOGY

Various systems are available for information hiding in an image, but they have some drawbacks i.e., they either do not encrypt the message or use a very weak algorithm in order to perform cryptography. They use the same key for encryption and decryption making it easy for the intruder to get access of the information. In some other cases the technique used may not be very efficient that is, the original image and the resulting image will be easily distinguishable by naked human eyes. For example DES algorithm, an encryption algorithm, used keys of smaller sizes (64 bit key) hence it was easy to decode it using computations. Algorithms using keys of these sizes are easily cracked by any intruder. So it is better if one goes for algorithms using keys of larger size which are difficult to decrypt and provide better security. Where stitching is concerned, multiband blending, gain compensation, automatic straightening makes the image smooth and more realistic.

The old system consists of four pin static password, eavesdropping is a major issue there. This issue is very efficiently solved in the proposed system since it uses dynamic generation of password. Problem of spoofing is eradicated since the dynamic passwords are valid only for a single transaction process.

In existing systems, using brute force we can crack the password in 9999 attempts whereas in the proposed system the same will require 99999999 attempts which is extremely difficult. The existing systems are using only 1 level security that is using AES (Advance Encryption Standard) encryption. The proposed system will be using 3 level security

- 8-pin dynamic generated password or OTP.
- AES Encryption to encrypt the OTP.
- Image Steganography to hide the encrypted OTP.

Image Steganography is a technique used to hide the data within an image. Use of image steganography for the proposed project helps to hide the very existence of the ATM transactions information. Thus the use of image steganography helps to make the existing systems more secure.

III. PROPOSED METHOD

The Data Encryption Standard (DES) shall consist of the Data Encryption Algorithm (DES) given below. These devices shall be designed in such a way that they may be used in a computer system or network to provide cryptographic protection to binary coded data. The method of implementation will depend on the application and environment. The devices shall be implemented in such a way that they may be tested and validated as accurately performing the transformations specified in the following algorithms. Our proposed System will be using 3 level security that is after login with the 4 digit ATM password a 8-digit OTP would be sent to the user on his Registered mobile number once the user enters the 8 digit OTP the this OTP is encrypted and hidden in a Image. thus hiding its very existence. This helps to overcome the vulnerability of the 4 digit ATM pin being hacked using Brute Force attack by 9999 attempts. Steganography is a technology where modern data compression, information theory, spread spectrum, and cryptography technologies are brought together to satisfy the need for privacy on the Internet. Image Steganography is the art and science of writing image in such a way that no one apart from the intent receiver and sender can suspect the existence of the message.

The Modules which is used to serve the Authentication of Bank customer is done in two phases.

- Data Embedding phase
- Authentication phase

A) DATA EMBEDDING PHASE:

The “Data Embedding phase” uses career file, which is an image file, in which the information related to customer is embedded for embedding purpose. In the encryption phase the data is embedded into the image using “Least Significant Bit algorithm”(LSB) by which the least significant bits of the secret document are arranged with the bits of carrier file such as image, Such that the message bits will merge with the bits of carrier file. In this procedure LSB algorithm helps for securing the originality of image. The encryption pattern depends on the type of

encryption we use. Least significant bit (LSB) insertion is a common, simple approach to embedding information in a cover image . In simple terms, the information will be first embedded and then the information will be extracted and the app will be made primarily using an IDE and Embedded Language for back end development and Java programming language to design User Interfaces.. In the proposed design, the following factors are chosen, and embedded:

- *Account Number*: This is a unique number generated by bank to each customers. The Account number constitutes of Type of account, Branch number, then the unique number for each account in that branch.

- *Customer Name*: Name of the customer holding that ATM card.

- *Address*: This field have the address of that particular customer.

B) AUTHENTICATION PHASE:

Day-by-day With the increasing amount of trustworthy and confidential information being accessible or transferred over the network the need for keeping unauthorized persons from access to this data. Nowadays, it is very easy to steal someone's identity and misuse it in the field of computing world and that is the reason some special verification methods had to be used. These methods will ensure that the person or system requesting the service is really who they say they are. The authentication phase starts when the client is ready and establishes a secured connection with the server. The server's response contains the *DName* as well as a PIN form into which the user at the client side will enter his PIN. Other simple methods of file encryption-decryption concepts, which are readily available in Java examples are easily captured by middle way itself.

This project helps to send a file from one place to another in a secured manner. The system provides authentication for transaction to 1 customer with a valid debit card at a time. In the Authentication phase, user has to carry two image files along with general information stored -.A cover image, which is a image of passport sized photo of Bank customer in which the information stated in Embedding phase is hidden. In any form, Consider a ATM card having a memory chip, consisting these two image files. Now when you need the authentication of a customer in a core banking the Share image file stored in the chip of ATM is retrieved and Stacked with the particular share. Authentication is done by comparing the hash value in the image as well as the Hash value Calculated by the information retrieved by the Cover image file. The first important work is done by the scanners.

They will scan the handwritten documents such as letters, claim papers etc. The multi-class classification and the back propagation algorithms will be used to recognize the input data given by the previous stages.

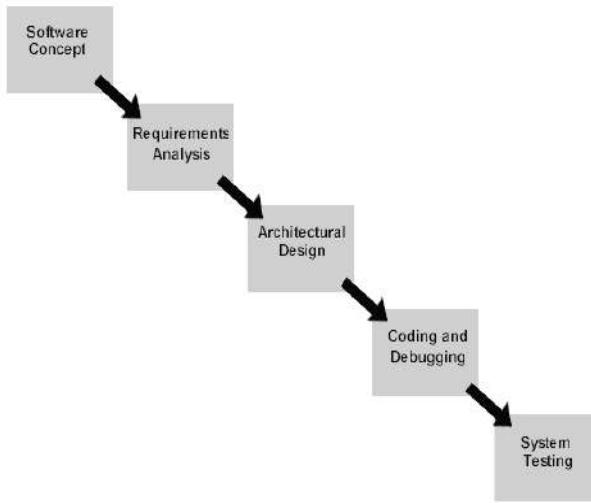


Figure 1 Development Stages

IV. FUTURE SCOPE

We have to carry out both “ATM client” and “server” and also reproduce the basic transactions of the “Automated teller machine”.

- “ATM simulator and server” project is divided into two phases. First phase will be focusing on developing the back-end server, the centralization of all the transactions. Second phase will be to develop the front-end ATM.
- To use ATM, client places his/her card into card reader he provides the PIN number as the input . The session starts when customer places his/her cart into card reader and finish when client presses reject button on ATM to get the ATM card back.

ATM client should be able to support both console and graphical interface. For efficient work, the project should

have the re-usability, it follows the OO principal and keep close for modification and open for extension in mind”.

- 1) *Social impact:* The project developed will provide a user-friendly environment as
- 2) *Ethical or legal impact:* No ethical or legal impact
- 3) *Environmental impact:* No direct environmental impact.
- 4) *Financial impact:* The project developed will minimize the efforts of business to market their product.

The project will provide a one stop solution for showcasing the offers. This project will overcome the traditional methods of marketing and hence save a lot of financial capital invested in resources like printing of pamphlets, man power required for distributing pamphlets etc.

V. CONCLUSIONS

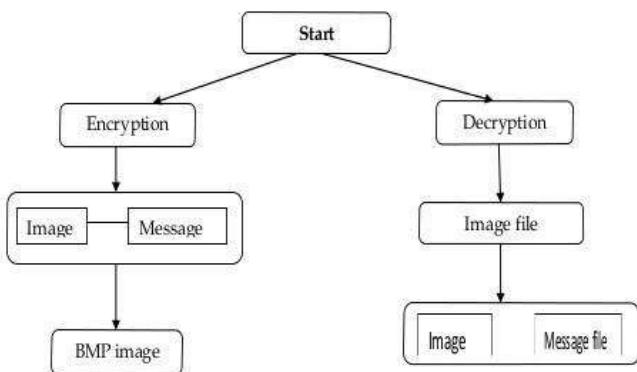
In this paper we have used LSB based Image Steganography. It is a good method where crucial information is rooted behind some cover media. LSB based steganography in combination with AES provides a good security model in order to hide data. AES is in demand over DES since it simple and has a better speed. Steganography technology which is an important part in the future parts of internet security and also for the secrecy in open systems for instance the Web. Steganography examination is mainly driven by the lacking strength in the cryptographic systems on their own and the need to have complete a cryptic open systems environment. Steganography is seen as a high-level encryption. It is used in information security within institutions. This use in institutions will provide a mechanism to perform two of the five key pillars of information security namely secrecy and integrity.

The confidentiality of the hidden message is protected and is not recognized in its hidden and encrypted form.[Both in the place of storage and during transmission while the encrypting of the concealed message protects the integrity of the data]. Since some of the main image stenographic techniques known about, one can see that there exists a large selection of approaches so that the information can be hidden in images. Almost all image file formats have different methods of hiding of messages, with variation in strong and weak points respectively. And our proposed project will solve the purpose of providing increased level of security to the ATM management system.

REFERENCES

- [1] <http://ieeexplore.ieee.org/document/6828087/>
- [2] <http://ieeexplore.ieee.org/document/7542298/>
- [3] <http://research.ijcaonline.org/rfem2016/number1/rfem45108.pdf>
- [4] https://www.ijarcse.com/docs/papers/Volume_6/6_June2016/V_616-0290.pdf

Figure 2 Block diagram for Steganography



- [5] http://ictactjournals.in/paper/IJCT_V7_I1_paper4_1245_1254.pdf
- [6] Pahati, OJ (2001-11-29). "Confounding Carnivore: How to Protect Your Online Privacy". AlterNet. Archived from the original on 2007-07-16. Retrieved 2008-09-02.
- [7] Petitolas, FAP; Anderson RJ; Kuhn MG (1999). "Information Hiding: A survey"(pdf). Proceedings of the IEEE (special issue). 87 (7): 1062–78. doi:10.1109/5.771065. Retrieved 2008-09-02.
- [8] "Polygraphiae (cf. p. 71f)" (in German). Digitale Sammlungen. Retrieved 2015-05-27.
- [9] Zarrelli, Natalie, "Knitting as an espionage tool", Atlas Obscura

Natural Language Processing Based SQL Engine

Tanushree Bhattacharjee, Abhijeet Kumar Saha, Jaineel Parmar, Pranit Patil

*Department of Computer Engineering, Thakur College of Engineering and Technology, Mumbai, India
abhijeetkrssaha@gmail.com, parmarjaineel@gmail.com, patilpranit@rediffmail.com*

Abstract— Natural Language Processing is a technology that assesses relationship of words, for example, activities, elements or events included inside unstructured messages, sentences and paragraphs found in a variety of text based documents. Natural language processing seek is characteristic dialect preparing innovation that particularly tackles issue of discovering answers to an inquiry which can be asked by basically going into a hunt interface. Utilizing this system, the control of writings is gained by knowledge extraction. The main concept of Natural Language Query Processing is to be done for an English sentence so that it can be understood by the computer and proper move to be made. In this project the separation concept of data keywords is implemented as tokens. The synonyms of the Token's names are put away in the database. During content parsing the keywords are isolated and the word coordinate which are deciphered will put in the database.

Keywords—Natural Language Processing, Unstructured Text, Tokens, Text Parsing

I. INTRODUCTION

Natural language processing is becoming one of the most active areas in Human-computer Interaction. The objective of NLP is to empower communication amongst individuals and PCs without falling back on retention of complex commands and procedures [1]. As such, NLP is a method, which can make the computer understand the languages naturally used by humans.

While natural language may be the easiest symbol system for people to learn and use, it has proved to be the hardest for a computer to master [2]. In spite of the difficulties, natural language processing is broadly viewed as a promising and basically vital undertaking in the field of computer research.

Unlike keyword search in Google or Yahoo, Natural Language Processing Question Answering Search particularly enables clients to make inquiries in their natural language and after that recovers the most important answers within seconds [3]. The standard search process requires the execution of different catchphrase mixes that at that point drive the searcher to tap on joins, which are gone by too much of the time, in the event that he doesn't locate any agreeable answer, at that point the way toward seeking proceeds until the point when the client discovers something or surrenders [4]. In Natural Language Processing Search, there is no additional work and no compelling reason to look through different connections, bringing about gigantic time investment funds. Entering a question is simple for the user even though the technology behind the scenes is highly complex.

In this project, the framework will acknowledge client's inquiry as an information. The program will check whether the inquiry is substantial or not utilizing Reed-Kellogg syntax function. This function will generate tokens by performing the division of the question clause. The token from the question clause is compared with clauses already stored in the dictionary. At that point the calculation examines the conditions and tries to discover articulations most like example by contrasting linguistic structure and semantics [6]. If both syntax & semantics match, the algorithm starts building the syntax fragment common for both utterances. More syntax nodes have been matched, higher is the matching score. As a result the best answer is shown.

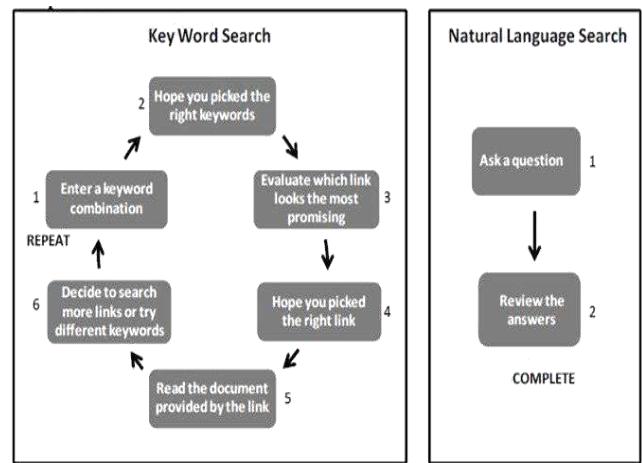


Fig 1: Comparison of search process

If no utterances are found matching with the pattern then system will display "No answer found". While parsing the input, if the input is not a valid question, the system will display message "Cannot parse question".

II. LITERATURE SURVEY

The historical backdrop of characteristic dialect preparing portrays the advances of normal dialect handling. There is some cover with the historical backdrop of machine interpretation and the historical backdrop of simulated intelligence. In 1950, Alan Turing distributed his well known article "Registering Machinery and Intelligence" which proposed what is presently called the Turing test as a basis of insight [5].

In 1957, Noam Chomsky's Syntactic Structures upset Linguistics with 'all-inclusive sentence structure', it was a manage based arrangement of syntactic structures [7].

The Georgetown explore in 1954 included completely programmed interpretation of more than sixty Russian sentences into English. The creators asserted that inside three or five years, machine interpretation would be a tackled issue. In any case,

genuine advance was much slower, and after the ALPAC report in 1966, which found that ten years in length inquire about had neglected to satisfy the desires, subsidizing for machine interpretation was drastically decreased. Minimal further research in machine interpretation was directed until the late 1980s, when the main factual machine interpretation frameworks were produced [9].

Some eminently effective NLP frameworks created in the 1960s were SHRDLU, a characteristic dialect framework working in limited "squares universes" with confined vocabularies.

In 1969 Roger Schank presented the calculated reliance hypothesis for normal dialect understanding. This model, in part affected by crafted by Sydney Lamb, was broadly utilized by Schank's understudies at Yale University, for example, Robert Wilensky, Wendy Lehnert, and Janet Kolodner [8].

In 1970, William A. Woods presented the expanded progress organize (ATN) to speak to regular dialect input. Rather than express structure rules ATNs utilized a proportionate arrangement of limited state automata that were called recursively. ATNs and their more broad arrangement called "summed up ATNs" kept on being utilized for various years. During the 70's numerous software engineers started to compose 'theoretical ontologies', which organized genuine data into PC reasonable information [12].

Up to the 1980s, most NLP frameworks depended on complex arrangements of transcribed tenets. Beginning in the late 1980s, in any case, there was an upheaval in NLP with the presentation of machine learning calculations for dialect preparing. This was expected both to the unfaltering increment in computational power coming about because of Moore's Law and the progressive reducing of the strength of Chomskyan hypotheses of etymology, whose hypothetical underpinnings demoralized the kind of corpus semantics that underlies the machine-learning way to deal with dialect preparing [13].

III. EXISTING SYSTEM

A. LUNAR:

LUNAR (Woods, 1973) included a framework that addressed inquiries regarding rock tests brought once more from the moon. Two databases were utilized, the compound investigations and the writing references. The program utilized an Augmented Transition Network (ATN) parser and Woods' Procedural Semantics. The framework was casually exhibited at the Second Annual Lunar Science Conference in 1971.

B. LIFER/LADDER

It was one of the principal great database NLP frameworks. It was outlined as a characteristic dialect interface to a database of data about US Navy ships. This framework, as portrayed in a paper by Hendrix (1978), utilized a semantic syntax to parse inquiries and inquiry a disseminated database. The LIFERILADDER framework could just help straightforward one table questions or various table inquiries with simple join conditions.

C. CHAT

The framework CHAT-80 is a standout amongst the most referenced NLP frameworks in the eighties. The framework was actualized in Prolog. The CHAT-80 was a very amazing, proficient and modern framework [11].

IV. RECENT DEVELOPMENTS

Charniak (1998) brings up that 90% precision can be gotten in appointing some portion of ordinary dialect tag to an inquiry word by applying basic factual measures. Jelinek (2001) is a generally referred to source on the utilization of measurable strategies in NLP, particularly in inquiry handling. Rosenfeld (2005) audits factual dialect models for dialect preparing and contends for a Bayesian way to deal with the combination of semantic hypotheses of data [10].

Trouble of outlining a legitimate inquiry can be eased through programmed Query Reformulation (QR) - an in the engine activity for reformulating a client's question with no extra contribution from the client. The QR framework (2013) works by advancing a client's pursuit question with certain particular extra terms drawn from the most noteworthy positioned relics recovered in light of the underlying inquiry. The critical point here is that these extra terms infused into a question are those that are considered to be "close" to the first inquiry terms in the source code based on positional proximity [14].

V. PROPOSED SYSTEM

Our undertaking focuses on question change with morphological, syntactic and semantic investigation.

In the first place, the inquiry (from the upper field) is parsed to get a Reed-Kellogg tree language structure chart. At that point the diagram is changed into its immediate answer frame. For instance Question Clause linguistic structure hub is supplanted with a Clause sentence structure hub. The subsequent diagram is utilized as a linguistic structure lexical example. At that point the calculation checks the content in the second field and tries to discover articulations, most like the example.

In the first place, it thinks about a linguistic structure hub from the example with a language structure hub from target content. On the off chance that grammar hubs coordinate, it thinks about the implications of words on the hubs. To think about word implications it just analyzes the Lexemes. On the off chance that both punctuation and implications coordinate, calculation goes down the linguistic structure trees and assembles the grammar piece basic for the two expressions. The more sentence structure hubs have been coordinated, the higher is coordinating score. The best answers are appeared accordingly [15].

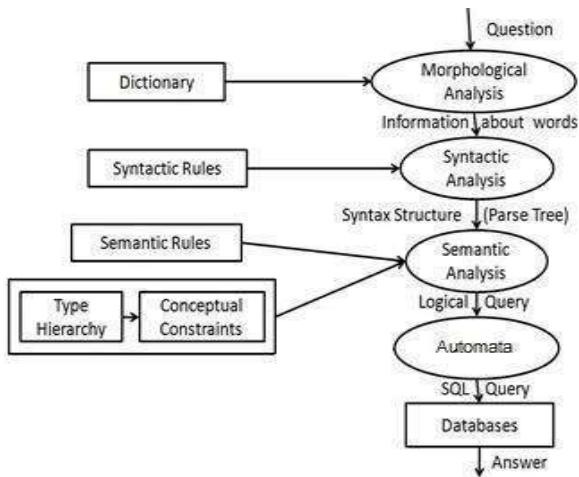


Fig 2: Flowchart proposed system

In the event that inquiry has an inquiry word, the device guarantees that inquiry word is constantly coordinated. The hub in the appropriate response chart, which coordinates the inquiry word, is the short answer [16].

The present task is being created to meet the desires demonstrated in the advanced age. An endeavour has been made through this undertaking to do all work ease and quick.

VI. ALGORITHM

The working of algorithm is explained with below example.

Text file

The longest river in the world is Nile.

The second longest river in China is Huanghe.

The Yangtze River is the longest river in Asia.

Strawberries contain no fat.

There is nearly no fat in strawberries.

Step1:

Input from the User to Natural Language Query Processor

Eg. What is the second longest river in China?

Step2: Once given input question, the program actually checks for a valid question using reed-kellogg syntax function it checks whether we have a proper question.

And this is the output of the code

```

▶ [QuestionClause]
  ▶ ::subject
    river (noun + singular)
      ▶ ::adjectiveModifier
        longest (adjective)
  ▶ ::verb
    is (verb + verb_base_form + auxiliary_verb)
      ▶ ::subjectComplementNoun
        -?->What (pronoun)
      ▶ ::adverbModifier
        second (adverb)
      ▶ ::adverbModifier [PrepositionalPhrase]
        ▶ ::preposition
          in (preposition)
        ▶ ::determiner
          China (noun + singular)
  
```

Step 3:

If there are no utterances in sentence then it will display “No answer found”

Step4:

Syntax graphs are matched

This is syntax graph for the question “What is the second longest river in China?”

Step5:

If syntax nodes match, then meanings of words associated with syntax nodes are compared

Here meanings of the words mean “Lexemes” and lexemes mean, if river is noun in questions, and if it is noun in input text then river lexeme is matched.

Step6:

If both syntax and meanings are equal, and if the utterance are considered to be equal, then matching score is incremented.

The more the matches of lexemes, the more the score and the more score gets the output answer. Once it has found matches we will have the output.

VII. CONCLUSION:

The main objective of this paper was to throw some light on basic architecture Natural Language Processing, its use in SQL query generation. We planned our project and developed the feasibility report. The project is an economically and technically feasible project.

The system developed will be beneficial to the many users which are not familiar with SQL language but want to access the data from the database.

REFERENCES

- [1] Turing Test From “Artificial Intelligence by Stuart Russel & Arnold.
- [2] Yucong Duan, Christophe Cruz (2011), Formalizing Semantic of Natural Language through Conceptualization from Existence. International Journal of Innovation, Management and Technology(2011)
- [3] Christopher D. Manning, Hinrich Schütze: Foundations of Statistical Natural Language Processing, MIT Press (1999)
- [4] Sisman.B .Kak, A.C. Publication Year: (2013)Page(s):309 - 318 ,ISSN :2160-1852 ,Print ISBN:978-1-4799-0345-0 ,INSPEC Accession Number:13828005 ,Conference Location :San Francisco, CA DOI:10.1109/MSR.2013.6624044 Publisher:-IEEE
- [5] R. K. Ando and T. Zhang. A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research (JMLR)*, 6:1817–1953, 2005.
- [6] R. M. Bell, Y. Koren, and C. Volinsky. The BellKor solution to the Netflix Prize. Technical report, AT&T Labs, 2007. <http://www.research.att.com/~volinsky/netflix>.
- [7] Y. Bengio and R. Ducharme. A neural probabilistic language model. In Advances in Neural Information Processing Systems (NIPS 13), 2001.
- [8] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle. Greedy layer-wise training of deep networks. In Advances in Neural Information Processing Systems (NIPS 19), 2007.
- [9] Y. Bengio, J. Louradour, R. Collobert, and J. Weston. Curriculum learning. In International Conferenceon Machine Learning (ICML), 2009.
- [10] L. Bottou. Stochastic gradient learning in neural networks. In Proceedings of Neuro-N'imes. EC2,1991.
- [11] L. Bottou. Online algorithms and stochastic approximations. In David Saad, editor, *Online Learningand Neural Networks*. Cambridge University Press, Cambridge, UK, 1998.
- [12] L. Bottou and P. Gallinari. A framework for the cooperation of learning algorithms. In *Advances in Neural Information Processing Systems (NIPS 3)*, 1991.
- [13] L. Bottou, Y. LeCun, and Yoshua Bengio. Global training of document processing systems using graph transformer networks. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 489–493, 1997.
- [14] J. S. Bridle. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In F. Fogelman Souli e and J. H'erault, editors, *Neurocomputing: Algorithms, Architectures and Applications*, pages 227–236. NATO ASI Series, 1990.
- [15] P. F. Brown, P. V. deSouza, R. L. Mercer, V. J. D. Pietra, and J C. Lai. Class-based n-gram models of natural language. *Computational Linguistics*, 18(4):467–479, 1992a.
P. F. Brown, V. J. Della Pietra, R. L. Mercer, S. A. Della Pietra, and J. C. Lai. An estimate of an upper bound for the entropy of english. *Computational Linguistics*, 18(1):31–41, 1992b.

Business Intelligence and Business Rating

Ali Akbar Nathani, Shaikh Mehboob Siraj, Aavez Kazi, Shaikh Sharukh, Prof. Sameer Panwala

Department of Computer Engineering, Anjuman-I-Islam's Kalsekar Technical Campus, University of Mumbai

Abstract - A rating is an evaluation or assessment of something, in terms of quality, quantity or combination of both. The main objective of rating the existing businesses is to make easy for the investors whether investing for a particular business is worth or not. Companies use Business Intelligence to adapt to the amending environment in order to be maintain in the market. Rating system is not only beneficial for the investors but also for the deserving companies. We will be getting the companies' liabilities and assets which is credited or debited from the companies' capital and generate financial statements. According to their financial statements generated, we will then rate them according to multiple parameters like profit and loss per year, amount of capital used in marketing, etc. As when the company keeps using BI portal for managing their financial profile, we will be updating their ratings eventually. Let's consider that a company named 'XYZ' is using this BI portal. This company will then be rated according to the financial statements that are generated after using the BI system. This rating is visible to all the top investors which will help them to choose company that is worth investing.

I. INTRODUCTION

What do you understand when we first hear the word Business Intelligence, Business intelligence (BI) is a technology-driven process for analyzing data and presenting actionable information to help executives, managers and other corporate end users make informed business decisions? BI encompasses a wide variety of tools, applications and methodologies that enable organizations to collect data from internal systems and external sources; prepare it for analysis; develop and generate reports and statistics for further business decision. Nowadays we can see that every product in the market is judged by the Quality and Durability of the product, so likewise the businesses should be judged according to some criteria so here the Business Rating System (BRS) comes in existence to freely exchanges thoughts and opinions. When this information travels the social web, it is difficult Facebook and Twitter are mostly short statements. There are various approaches that can be used to develop news validation or fake news detection system. Major there are two types of approaches, linguistic approach and network approach. In Linguistic approach, some liar uses their language skill to avoid being caught guilty. There is some leakage of words from which we can identify that whether they are saying truth or not. The goal in the linguistic approach is to look for such words or leakages. Hybrid approach is the combination of both network approach as well as linguistic approach. In our system we are going to use hybrid approach because individually network or linguistic approach is not too accurate to increase efficiency and for better results we shall be using hybrid approach.

II. PROPOSED FRAMEWORK

A. What is Business Intelligence System?

Business Intelligence(BI) is to integrate, collect and analyze company's data to analyze business information to re-build business processes and technologies used for being stable in the current market. The BI System also generates reports and statistics according to the conclusion derived from the business analysis. Business Intelligence is all about automating the decision-making accurately essential for the business processes rather than just guessing and taking risk for the business. The guessing done by the business persons can be inaccurate and can harm the business. The purpose of BIS is to support better business decision making. Essentially, BIS are Data-Driven Decision Support System, companies rely on business intelligence tools such as data warehouse, data mining, and data modeling to collect, manage, store, and organize business information.

B. What is a Rating System?

There are many different rating systems for individual opinions of content, nowadays we can see that everything is based on the rating system every product we buy from the market depend on the rate what is given to the product. The rating of the product can be in many ways some of can be done on the scale of 1-10, Some of the rating can be on the star or some on the percentage from 1-100. One of the most common example is the motion picture rating system for the movies. The rating system basically gives you Quality, Integrity and Results.

C. How do software tools optimize Business Intelligence?

Business Intelligence software systems provide historical, current, and predictive views of business and determining analytic information about the business using data mining or data Mart. Software elements support reporting, interactive "slice-and-dice" pivot-table analyses, visualization, and statistical data mining. Applications tackle sales, production, financial, and many other sources of business data for purposes that include business performance management. Information from different businesses from same domain are gathered to analyze the target business. Benchmarking is a valuable data analysis tool business use to grow their productivity, revenue, and overall success related to that competitors. Businesses often struggle by benchmarking, but BI simplifies the process, providing actionable information and reports that can be easily understood. Benchmarking reports offer accurate, timely data which will allow businesses to evaluate their performance against that partners and competitors, providing valuable insight into the organizations current state. Business intelligence software comprises a range of data analytics tools designed to analyze and manage data related to your business operations. BI rich visualization capabilities allow businesses to monitor sales, logistics, and even productivity—and provides extensive data analysis using richly visualized, intuitive reports

III. COMPARATIVE STUDY

What we can see that if the rating system is not available then there will be a great choice between the user that what product should we buy from the market. One of the best way to compare the Quality of the product is by the review from the customer regarding the product. In this approach we are going to rate the Business of a particular company / Individual depending on the market hold, customer Reviews, Profit and loss statement and their various standard procedure on which the system is going to rate the particular Business. The Business Intelligence (BI) not only will rate your system but also help the company to see the pros and cons of the company which will help the company to generate a great amount of Revenue for company.

IV. SYSTEM ARCHITECTURE

The system architecture is mainly divided in two sub modules, the first sub module consists of the Business Intelligence System (BIS) and the other module consist of the Business Rating System (BRS). The first module shows how the Business Intelligence System help for the development of the Business by taking the input from the company, the input could be in the form of the financial statement. The financial statement which are generated by the company are entered in the business intelligence system according to the financial statement provided by the company the business intelligence will generate a report which will give the brief information about the current status of the company what are the pro and cons and also what should the company do in order to withstand all the problem and overcome all the deficiencies Example : If a company is investing a large amount of money in the advertisement then the business intelligence system will give a report that the company is investing a large sum of money in the advertisement.

The second module consists of the Business Rating System (BRS) the BRS will rate your business according to the standard procedure which has been laid out by the company. The standard procedure could be the location of the company, the amount of worker which are working in the company, the instrument which are used in the company, the qualification of the employee etc. The company will be rated using the rating system a particular company will be rated and it will get a certain rate. Then this rating will be displayed on the website and there will be investor which are willing to invest in this company can contact the company. The BRS system are very essential for the company which has a very good scope, but they are unable to get the funding.

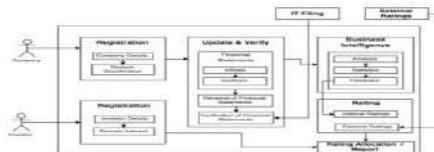


Fig.1: Business Intelligence System Architecture

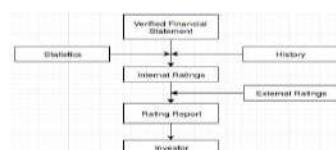


Fig.2: Business Rating System Flowchart

1. COMPANY:

The company will be connected to the BI module. The company will give input to the system in terms of financial statements, that will go to the BI System. The BI will perform operations and will return output.

2. INVESTOR

The investor is the main component who request for company ratings, statistics and full details of company. The BI module will take the request as an input and provide the company rating, statistics and full detail report of company as an output to the investor.

3. FINANCIAL STATEMENTS

Financial Statement are reports that every business organization must prepare to be able to indicate to its stakeholders and operators, the complete picture of the organization's health.

4. BUSINESS INTELLIGENCE(BI)

The BI model will take the input from company in terms of financial statements Analyze those inputs and generates 3 things

A. Management Information System(MIS)

MIS is a means to generate regular and timely reports that help management take critical decisions that shape the business, and help it improvise on several aspects.

B. Statistics:

Statistics: Statistics are generated on the bases of the financial statement which are provided by the company which are further enhanced using the Business Intelligence System.

C. Business Process Re-engineering(BPR):

It involves the good of core business process to archive good improvements in productivity and quality. In BPR companies start with a blank sheet and rethink existing process to deliver more value to the customer.

D. Business Rating(BR):

This module will Analise the input provided by BI module i.e MIS, statistics, BPR and according to the external and internal criteria it will generate the ratings

E. IT Filing:

It is the form in which an assessee file information about his Income and tax thereon to Income Tax Department.

V. CONCLUSION

Those companies are deserving they should get proper investment from the great investors, so our system will help those companies to get proper investments, investor can also see in which company should they invest so they can get more profit.

REFERENCES

- [1] <https://financesonline.com/purpose-business-intelligence-business>
- [2] https://en.wikipedia.org/wiki/Business_intelligence
- [3] <https://www.oddup.com/>

Security of Multimedia data using Hybrid Model

Gautam Jha

Department. of Computer Science Engineering, Thakur College of Engineering and Technology
gautamjha229@gmail.com

Abstract - In the recent trend, data are increasing enormously through every second of life. Securing data is one of the main challenges nowadays, as data travels through different medium where the medium is not secured. To prevent secret data from an unauthorized access there are a number of approaches. The multimedia files can be protected by the approach of encryption and decryption, which is the methods of symmetric key cryptography and asymmetric key cryptography. This paper focuses on a proposed technique used for encrypting and decrypting multimedia data through the hybrid proposed model based on the assemblage of Advanced Encryption Standard and Deffie Hellman Key Exchange Algorithm. The proposed method generates a key by the sender using Deffie Hellman algorithm and passed to the AES algorithm which treat as a key for encrypting multimedia data. The encrypted data will be decoded by the receiver key using Deffie Hellman and AES. The benefit of this proposed hybrid model is that it will provide the security to the multimedia data and also fill the gap of storage, speed and security. It is even used for societal for data communication over the network.

Keywords - Security, Multimedia, Cryptography, DES (Data Encryption Standard), AES (Advanced Encryption Standard).

I. INTRODUCTION

In a communication system, security of information is a prima concern nowadays. There are lots of techniques like hashing, steganography, cryptography that make available as for the safety of essential information. Confidential data can be secured by the application of Encryption and Decryption techniques as it takes an important part. In a gone times, through the study of past events the data were made unreadable by scrambling the contents of information to keep it secret. Cryptography algorithms like DES, AES, Blowfish, RSA and Deffie Hellman provides a higher level of security by using a set of keys for encryption and decryption.

AES is a symmetric key algorithm where the plain text size is a block size of 128,192,256 bits and the length of the key is same as the size of the plain text i.e. 128,192,256 bits. It involves basically four steps as follows:

- Byte Substitution
- Shifting of rows
- Mix columns
- Add a round key

It has total rounds of 10, 12 or 14 to encrypt data looking on the plain text size of 128, 192 and 256 respectively.

Initially bits are converted into bytes and ordered in a matrix form. After which, data are substituted from the substitution box and are arranged again in a matrix form. In the further step shifting

of rows take place depending on the offset used, i.e. offset 0, offset 1, offset 2 and offset 3. Offset 0 means data will remain on the same position there will not be any shifting of data whereas offset 1 means data shifted by one position and respectively. In the succeeding step columns are mixed in a similar approach as of rows shifting. At the last, Addround key is added after the mixing of columns and the process repeats till the last rounds depending on the plain text size.

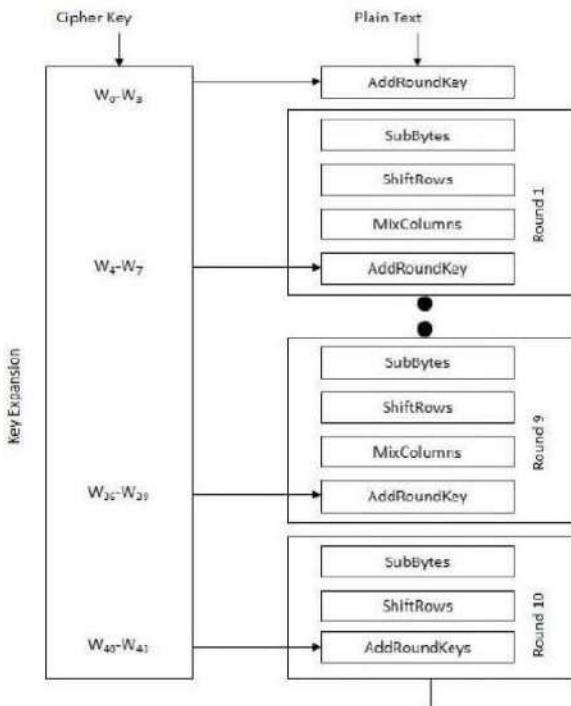


Fig.1: AES process [8]

Deffie Hellman key exchange algorithm, in which is keys are used for communication to exchange data in a public network. It involves basically five steps as follows:

- Global Public elements
- User 'A' Key generation
- User 'B' Key generation
- Generation of secret key by user 'A'
- Generation of secret key by user 'B'

Initially the two global public elements are used, one is a prime number element, and another is a primitive root of that prime number. In the next step, user A selects a random number which is smaller than the prime number and it is used as a private key. In the next step user B selects a random number which is smaller than the

prime number and it is used as a private key. In the fourth step user B shares his private key with user A and then user A generates a secret key for exchanging data. Similarly, user B generates a secret key for exchanging data. In the last two steps we get the same result which is used as a shared secret key.

The result provided by Advanced Encryption Standard is much better and faster than any other algorithm like Data Encryption Standard. Diffie Hellman Key Exchange algorithm exchanges their private key and public key to prepare a common shared key for exchanging the data. The proposed hybrid algorithm is designed by combining AES and DHM to encrypt the multimedia data by converting it into a bit equivalent to reduce the overall time constraint and provide a high level of security.

II. MOTIVATION

The security of multimedia data in an efficient manner to manage the risks associated with it and maintains the security requirement. Security of data to avoid, the loss of information by most convenient methods in a more secure way.

III. OBJECTIVES

- The objective of this work is to enhance the security of multimedia data.
- It also fills the gap of storage, speed and security along the communication of data over the network.

IV. LITERATURE SURVEY

A. A Multiple layer Text security using Variable block size Cryptography and Image Steganography

In the cryptography large amount of research has been done to keep the data secured. The techniques of cryptography and steganography are used to secure a text by multiple layer approach. In this approach, the text is converted into cipher text using cryptography methods of variable block size and that cipher text is hidden in an image using steganography approach. These methods deal only with the image files. [2]

B. DES, AES and Blowfish: Symmetric Key Cryptography Algorithms Simulation Based Performance Analysis

DES (Data Encryption Standard) is a symmetric key algorithm, where the same key is used for encryption and decryption of data and it started to demoralize due to its natural limitations. It uses 64-bit plain text message along with 56 bits of key for encrypting it into a 64-bit cipher text. Due to the usage of small key length, it makes it vulnerable to brute force attack as cipher is simple and straightforward.

These limitations were overcome using Triple DES which means three times DES is carried out by a pair of key of size 112 or 168-bit. The encryption process of converting plain text into cipher text in 3DES takes relatively more time as compared to simple DES. To overcome the limitations of DES, a new cryptographic proposal AES or subset of Rijndael cipher was given by two Belgian cryptographers, Joan Daemon and Vincent Rijmen. In AES depending on the number of rounds i.e. 10,12 or 14 rounds it uses a single key of size 128, 192 or 256-bit which is same as the size of the plain text. AES has been classified into four steps of Byte substitution, shifting or rows, mix columns and adding a round key for each round except the last round to provide the corresponding cipher-text. DES is faster in performance comparison than AES, but

it lacks in terms of security. The asymmetric key-based encryption algorithms use the pair of keys; a public key and a private key which is mathematically bounded to each other. [3][4]

C. A Method for Obtaining Digital Signatures and Public-Key Cryptosystems

RSA is an asymmetric key based encryption technique which is used widely and for sharing the information through the communication channels it uses 1024-bit key stream. Message confidentiality in RSA is exclusively dependent on large prime number and discrete logarithm problem. It is highly secured and provides security to online transaction data over the communication network. In a research article by Boneh, specified the limitations in RSA and the various possible attacks on an RSA cryptosystem. It is easy to identify the private key if planned properly by using a timing attack.

In just 10 iterations an author of paper "On the Power of Simple Branch Prediction Analysis" has claimed to obtain 508 bits of 512 bits RSA [5]

D. Steganography for Inserting Message on Digital Image Using LSB and AES Cryptographic Algorithm.

In this data are encrypted with the help of AES algorithm. The data which is to be encrypted are in text format. It uses Least Significant Bit method for inserting data on a digital image to make it more secure. The system is not much secure as it uses only Least Significant Bit. It deals only with the image files. [6].

E. Securing Data in Cloud Using AES Algorithm

Data in file format are encrypted by encryption technique. The files are encrypted with the AES Algorithm. The proposed system will work only when there is a stable internet connection. It follows the steps involved in AES Algorithm, store data on the cloud in the encrypted form which can be downloaded or decrypted with usage of concepts of keys. The proposed system can be applied in various fields like Voice communication, Network appliances and Virtual Private Network. [7]

F. Advanced Encryption Standard (AES) Algorithm to Encrypt and Decrypt Data

In this paper, it describes all the processes of Advanced Encryption Standard. The procedure used for encrypting data along with decrypting data for making data more secure and in efficient manner. It provides better security than DES and 3DES. [8].

V. RESEARCH GAPS

Table.1: Research Gaps

Sr No	Title	Description	Gap Analysis	Future Scope
1.	Multiple Layer Text Security using Variable block size Cryptography and Steganography	This method hides encrypted message inside an image	It deals only with the text and image	Security for Multimedia files like image, audio and video can be done

2.	DES, AES and Blowfish: Symmetric Key Cryptography Algorithms Simulation Based Performance Analysis	This method provides the result using simulation model using symmetric key Cryptography like DES, AES, Blowfish	They have used DES for encryption. In this the Key size is very small that is vulnerable to attack.	Increase the Key size for better result
3.	A Method for Obtaining Digital Signatures and Public-Key Cryptosystems	It uses only public key along with RSA for obtaining a Digital Signature	Compare to the private key, public key are not much secured	Increase the Usage of both Public Key and Private Key
4.	Steganography for Inserting Message on Digital Image Using LSB and AES Cryptographic Algorithm	The data in the text format are encrypted with the help of AES algorithm. It uses LSB method for inserting data on a digital image to make it more secure.	It is not much secure as it uses Least Significant Bit. It only deals with text.	Security for Multimedia files like image, audio and video can be done.
5.	Securing Data in Cloud Using AES Algorithm	It uses Advanced Encryption Standards algorithm for securing the data in cloud.	It only works on file which contains data in text formats.	It can be implemented for other data like images, audios etc.

VI. PROBLEM DEFINITION

After studying the limitations of the current existing system, and sensing that there is a need of a more effective system to secure multimedia data in various terms as follows:

Time Complexity, as in many approaches the system consists complex parts which increases the time complexity, but the simple systems have less time complexity. So, the compromise between two is must needed.

Memory Efficiency, the text encryption method requires less memory as compared to the multimedia data encryption. So, the compromise between two is must needed.

Input supported, the system used till now only deals with the text in the direct form or taken from some files. Multimedia files uses image, audio, graphical contents require large memory for storage and the time required for the purpose of encryption and decryption is also high.

Information is playing an important role in our lives. One of the major sources of information is databases. Databases and database technology are having major impact on the growing use of computers. In order to retrieve information from a database, one needs to formulate a query in such way that the computer will understand and produce the desired output. Generally, query

processing is handled by the Structured Query Language (SQL). But the non-IT people cannot be able to write SQL queries as they may not be aware of the SQL as well as structure of the database.

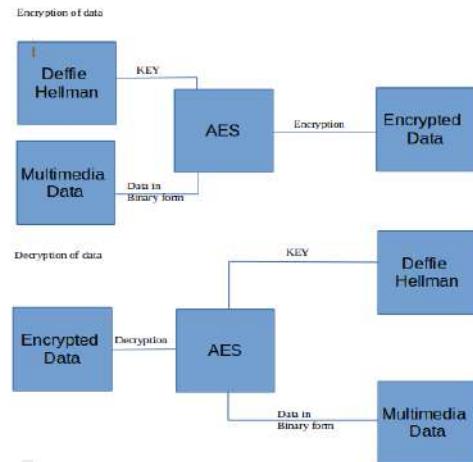


Fig.2: System Model

The proposed system model provides security to the multimedia data by compounding two algorithms viz. Deffie Hellman Key Exchange algorithm and Advanced Encryption Standard.

Deffie Hellman algorithm create a shared secret key by exchanging between the two users and it is used as a key for AES algorithm for encrypting the multimedia data which in binary form. The multimedia data in binary form will go for encryption and provide the encrypted file and process will be reversed to get the original data.

VII. IMPLEMENTATION AND RESULTS

This paper proposes a hybrid system for securing the multimedia data type like any file, images, audios etc.



Fig.3: AES Encryption



Fig.4: AES Decryption

VIII. CONCLUSION

The proposed hybrid technique which is prepared by combination of two cryptography algorithms it provides result faster and in efficient manner. The algorithms of Deffie Hell Man Key exchange and Advanced Encryption Standard will provide better results by providing high level of security to the multimedia data then the current existing system. It will be used to secure the

multimedia data in the form of a file, image, audio etc. and also provide better memory efficiency.

REFERENCES

- [1] GauriRao, W.Stallings, "Cryptography and Network Security: Principles and Practices, fourth edition", pp.592, November 2005.
- [2] Shivani Chauhan, Jyotsna, Janmejai Kumar, Amit Doegar, Multiple layer Text security using Variable block size Cryptography and Image Steganography, 3rd IEEE International Conference on "Computational Intelligence and Communication Technology" (IEEE-CICT 2017).
- [3] Daemen, Joan, Rijmen, Vincent. (March 9), AES Proposal: Rijndael. National Institute of Standards and Technology 2003; p. 1. Retrieved 21 February 2013.
- [4] Jawahar Thakur, Nagesh Kumar. DES, AES and Blowfish: Symmetric Key Cryptography Algorithms Simulation Based Performance Analysis. International Journal of Emerging Technology and Advanced Engineering December 2011; vol 1(2), p.6-12.
- [5] R.L. Rivest, A. Shamir, and L. Adleman. A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. Communications of the ACM 1977; p. 120-126.
- [6] Nurhayati, Syukri Sayid Ahmad, Steganography for inserting message on digital image using least significant bit and AES cryptographic algorithm. IEEE Conferences 2016.
- [7] Prasoon Raghav, Rahul Kumar, Rajat Parashar, Securing Data in Cloud Using AES Algorithm. International Journal of Engineering Science and Computing, April 2016; Volume 6 Issue No. 4. ISSN 2321 3361 © 2016 IJESC.
- [8] Ako Muhammad Abdullah, Advanced Encryption Standard (AES) Algorithm to Encrypt and Decrypt Data; Research Gate.
- [9] wikipedia.org,"Deffie Hellman Key Exchange Algorithm", https://en.wikipedia.org/wiki/Diffie%20Hellman_key_exchange
- [10] wikipedia.org,"Advanced Encryption Standard", <https://en.wikipedia.org/>

Proposed Improved Text Summarization Method Using Adaptive Neuro Fuzzy Inference System

Pratiksha D. Warule ,Sudhir D. Sawarkar, Archana Gulati

Datta Meghe college of Engineering Engineering, Airoli, India

pnsawane128@gmail.com,sudhir_sawarkar@yahoo.com,gulatiarchana81@gmail.com

Abstract— Nowadays data present on the World Wide Web is growing exponentially. People use search engines like Google, Bing, and Yahoo etc for retrieving required information. But as the information present on the web huge it is necessary for user to make the summary of this information. User can easily understand the large volume of data with the help of summary, and does not require spending so many for analyzing the collected information. Text summarization is the process of condensing the large source information into a shorter version. While summarizing the text one should preserve its information content and overall meaning. It is very difficult and time consuming process for humans to summarize large documents. So text summarization is used using different techniques like fuzzy logic, genetic algorithm, neural network. In our proposed system we have combined the neural networks and fuzzy logic using Adaptive neuro fuzzy inference system along with lexical features. So it will overcome the drawbacks of both neural network and fuzzy logic. This proposed system will take the learning ability of neural network and uncertainty data handling of fuzzy system. In addition to this it also adds the lexical feature calculation of input data so that we can get precise summary of input text.

Keywords—Neural Network, Fuzzy logic, ANFIS, Lexical Analysis

I. INTRODUCTION

A. TEXT SUMMARIZATION APPROACHES

The main idea behind text summarization is to identify the essential content from the text, understanding its meaning clearly and generate the short text from it. Here understanding of text is Natural Language Processing problem. Natural language processing uses different techniques such as semantic analysis, inferential interpretation, discourse processing etc. for text summarization.

1) Linguistic Approach

These methods use linguistic knowledge so that computer can analyze the text semantically and according to that it selects the sentences for summary by considering the verb, subject, noun etc. so these methods are difficult than statistical methods. This method finds the term relationship between the documents by using part of speech tagging, grammar analysis, and thesaurus usage and selects the meaningful sentences.

2) Statistical Approach

Statistical approach summarizes the text without understanding the meaning of the words. It depends on the

statistical distribution of some features. In this method the sentences are selected based on the word frequency, indicator phrases without considering the meaning of the words. There are different methods for finding the key sentences for example, the title method, the aggregation similarity method, the location method, the frequency method etc.

3) Hybrid Approach

Hybrid method uses the combination of statistical and linguistic method for generating meaningful and short summary.

B. Text Summarization Categories

1) Based on the Approaches

In extractive summary the important sentences are extracted from the source text and added to the summary whereas in abstractive summary novel sentences are generated for summary.

2) Based on the type of details

In an indicative summary the long text document is analyzed and then it produces only the main concept from the long text. These summaries are small and they encourage user to read the long text. Informative summary is the substitution for the original text. It provides the concise knowledge about the original text to the user.

3) Based on the type of content

Generic summary does not depend on the subject of the original text and can be used by any type of user. Query based summary is the answer of the users query. It gives the user's view. So it's not user specific as it cannot be used by any type of user.

4) Based on limitation

In Generic specific summary, it can accept only special type of input like stories, newspaper articles, manuals etc. whereas in Domain independent summary it can accept different type of input text. Domain independent summary is not dependent on the domain and it can be used by any type of user.

5) Based on the number of input documents

In single document summarization only single document is taken as input for generating the summary. So this approach is easy compared to multi document summarization. In multi document summarization multiple documents are taken as input for generating the summary. This approach is difficult to implement.

6) Based on language

In Mono lingual system it takes input of documents with only specific language and summary is generated based on that language only, whereas in Multi lingual system it takes input of documents with different languages and gives summary based on that different languages.

7) Based on the knowledge of the user.

A background summary assumes the reader's prior knowledge of the general setting of the input text(s) content is poor, and hence includes explanatory material, such as circumstances of place, time, and actors. Adjust-the-news summary contains just the new or principal themes, assuming that the reader knows enough background to interpret them in context.

II. LITERATURE SURVEY

In previous work done in text summarization based on different techniques like neural network, fuzzy logic, Genetic Algorithm, clustering etc.

Taeoh Jo proposed text summarization using KNN (K Nearest Neighbor). He computed the similarity between feature vectors by considering the similarity among features as well as among values. This approach provides more reliability and represents data items more compactly. In text pre-processing step it encodes the text into numerical vector. Then the similarity between the two vectors is calculated by considering feature similarity. Then it applies revised KNN algorithm which considers both feature value and feature similarity for classifying the sentences as summary sentences and non-summary sentences.

Saif alZahir, Qandeel Fatima and Martin Cenek proposed new graph based technique for text summarization. It generates multi-edge-irregular-graph that states the number of times the word has occurred in the sentences of the input text. This graph is then transformed into symmetric matrix from where it can get the ranking of sentences. Using this data along with the threshold value it can generate summarized text. This method is fast and it can be used for real time text summarization. Sebastian Suarez Benjumea and Elizabeth Leon Guzman developed genetic clustering algorithm for extractive text summarization. It clusters the sentences which are close representation of the text topics. It clusters the sentences with the help of fitness function which is based on redundancy and coverage. Then it applies the scoring function which selects the most relevant sentences of each topic which will be included in text summary.

Arman Kiani -B M. R. Akbarzadeh -T proposed a new technique for text summarization which uses both Genetic Algorithm and Genetic programming for optimizing the rule sets and membership functions of the fuzzy system. In this method Genetic Programming is used for rule sets whereas Genetic programming is used for membership functions. This method extracts the important sentences in the input text by decreasing the redundancy of the data.

Khosrow Kaikhah proposed text summarization using neural networks and Rhetorical Structure Theory. Here the network is trained according to the style of a human. The individual user can train the network as per their style.

A. TEXT SUMMARIZATION FEATURES

1) Title feature

The sentence that contains the word(s) in the document title will be given high score. Occurrence of words from the document title in a sentence indicates that the sentence is highly relevant to the document. This can be computed by

counting the number of matching characters between the words in sentence and the words in the document title.

$$f1 = \frac{(\text{Number of title words in sentence})}{(\text{Number of word in document title})} \quad (1)$$

2) Thematic word

This feature is used to determine the commonness of a term. A term that is used frequently is probably related to the topic of the document. We consider the top 10 words as the maximum number of frequent semantic terms.

$$\begin{aligned} f2 \\ = & (\text{Number of frequent terms in sentence}) \\ & \div \text{Max}(\text{Number of frequent terms}) \end{aligned} \quad (2)$$

3) Term weight

The importance or weight of each word in the document can be computed. The weight W_i of word i can be calculated by the traditional tf.idf method. We adopted this method as tf.isf (term frequency, Inverse sentence frequency):

$$Wi = tfi \times isfi = tfi \times \log N ni$$

Where tfi is the term frequency of word i in the document, N is the total number of sentences and ni is number of sentences in which word i occurs. Using Equation the term weight score for a sentence can be computed as follows:

$$f3 = \sum_{i=1}^k Wi(S) \div \text{Max} \sum_{i=1}^k Wi(S) \quad (3)$$

Where $Wi(S)$ is the term weight of word i in sentence S and k is the total number of words in sentence S .

4) Sentence Length

This feature is useful to filter out short or long sentences. Too short or long sentence is not good for summary. This feature computation uses minimum and maximum length threshold values.

$$\begin{aligned} SL = 0 & \text{ if } L < \text{Min} \text{ or } L > \text{Max} \\ & \text{Otherwise} \\ SL = & \text{Sin}((L-\text{Min}) * ((\text{Max}\Theta - \text{Min}\Theta) / (\text{Max} - \text{Min}))) \end{aligned} \quad (4)$$

Where, L = Length of Sentence

Min = Minimum Length of Sentence

Max = Maximum Length of Sentence

$\text{Min } \Theta$ = Minimum Angle (Minimum Angle=0)

$\text{Max } \Theta$ =Maximum Angle (Maximum Angle=180)

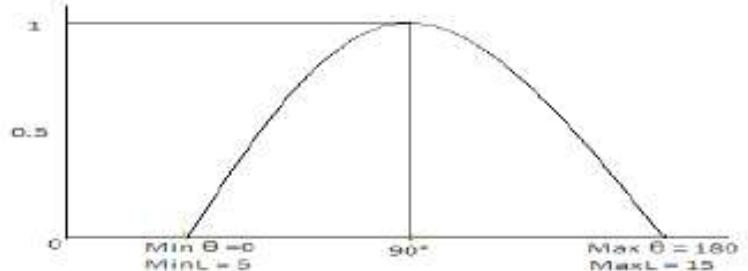


Fig. Sentence Length

5) Sentence Position

Position of the sentence in the text, decides its importance. Sentences in the beginning defines the theme of the document whereas end sentences conclude or summarize the document. In this threshold value in percentage defines how many sentences in the beginning and at the end are retained in Summary whose weight is given here ,SP = 1.

Remaining sentences, weight is computed as follows

$$Sp = \text{Cos}((P - \text{MinVal}) * ((\text{Max}\Theta - \text{Min}\Theta) / (\text{MaxVal} - \text{MinVal}))) \quad (5)$$

Where TR = Threshold Value

MinVal = N*TR (Minimum Value of Sentence)

MaxVal = N*(1 - TR) (Maximum Value of Sentence)

N = Number of sentences in document

Min Θ = Minimum Angle (Minimum Angle=0)

Max Θ = Maximum Angle (Maximum Angle=180)

P = Current Position of sentence

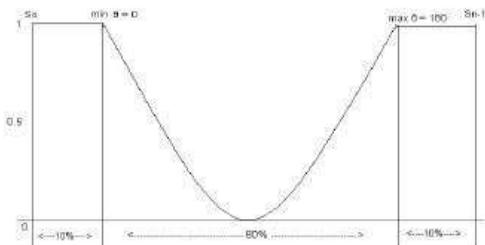


Fig. Sentence Position

6) Numerical Data

The Sentence that contains numerical data is important and it should be included in the summary.

$$\begin{aligned} ND &= 1, \text{ Digit exist} \\ &0, \text{ Digit does not exist} \end{aligned} \quad (6)$$

7) Sentence to Sentence Similarity

This feature finds the similarity between the sentences. For each sentence S, similarity between S and every other sentence is computed by the method of token matching. The two dimensional matrix is formed of the size [N] [N] where N is Number of sentences in the document. In this matrix diagonal elements should be assigned 0 values as sentence should not get compared with itself.

$$\text{Sim}(S_i, S_j) = \text{TM}[(t_i)^{n_1}, (t_j)^{m_1}]$$

where TM is token matching method. The score for sentence to sentence similarity is calculated as ratio of summary of similarity of sentence S with every other sentence over the maximum summary.

$$F7 = \frac{\sum [\text{Sim}(S_i, S_j)]^{N_1}}{\text{Max}} = \frac{\sum [\text{Sim}(S_i, S_j)]^{N_1}}{\sum [\text{Sim}(S_i, S_j)]^{N_1}} \quad (7)$$

8) SVO Qualification

SVO is a type of sentence structure. English language has SVO structure. In SVO structure first subject comes before verb and then object comes. SVO is used in active voice. Here tagging algorithm is used where based on the tags assigned first noun word is taken as a subject of the sentence. Then whole sentence is parsed up to its end and if the last word is object then sentence is qualified as SVO. Only those sentences which are marked as SVO structure will be taken for further processing of text summarization. SVO Qualification

SVO (Si)=1 , SVO Qualified

=0, SVO Not Qualified

B. Abbreviations And Acronyms

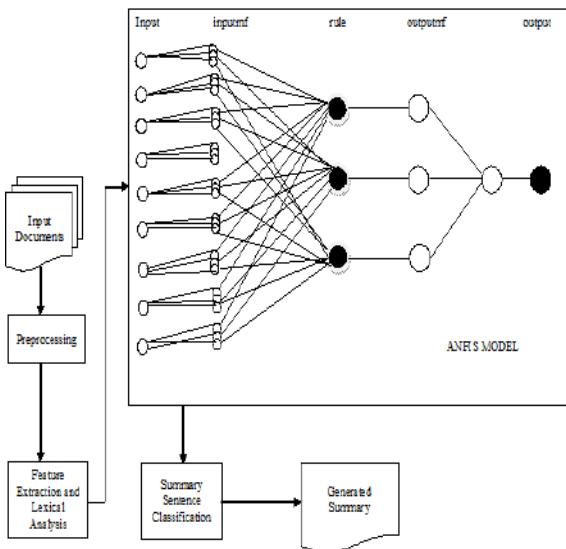
ANFIS (Adaptive Neuro fuzzy inference system)

POS (Parts of Speech)

III. PROPOSED METHOD

In our proposed method we will combine the fuzzy logic with neural network. This method will use the ANFIS (Adaptive Neuro Fuzzy Inference system) method for combining these two approaches. This proposed method will take the uncertainty handling of fuzzy logic and learning capabilities of neural network. It overcomes the limitations of neural network and fuzzy logic.

In our proposed method ANFIS model is used for classifying documents sentences as summary sentences and non-summary sentences. In this proposed method preprocessing step will include the lexical chains along with other preprocessing work. This method takes the features such as title, sentence length, proper noun, thematic word, term weight, sentence position, numerical data, and sentence to sentence similarity for calculating the feature values for every sentence from the training set. When the training of model is finished it can predict the summary sentences from the given input documents. In this method no human or linguistic expert will be needed for generation of fuzzy rules, as rules are automatically generated using subtractive clustering method. Subtractive clustering method maps the input output data and automatically estimates the cluster number and cluster centers. Here each instance will be assumed as a potential cluster center and if the instances have a value that is in the range of the first cluster then it will be included in the first cluster otherwise it will create a new cluster. This process will be repeated until all instances are included in the cluster. Least-square estimation and back propagation gradient descent method will be used for training the model. The trained ANFIS model will be used for classifying the sentences as summary sentence and non-summary sentence. It will use the binary classifier where class 1 will represent summary sentences and class 0 will represent non-summary sentences. For classifying predicted output threshold value will be used.



Architecture of Proposed system

IV. FUTURE SCOPE

This proposed method can overcome the drawbacks of neural network and fuzzy logic by using ANFIS and lexical features. Our proposed method can improve the text summarization by generating the precise summary of the input text. In future we can add more features for calculating the sentence score and rank them accordingly. Also in future we can combine ANFIS with other methods like genetic algorithm.

Fig:

REFERENCES

- [1] Aik, L.E.(2008). A study of neuro –fuzzy system in approximation based problems. Neuro – fuzzy system ANFIS: Adaptive Neuro – Fuzzy Inference System 24(2), 113-130.
- [2] Farshad Kyoomarsi, Hamid Khosravi, Esfandiar Eslami, Pooya Khosravyan Dehkordy, Asghar Tajoddin: Optimizing Text Summarization based on fuzzy logic, seventh IEEE/AACIS international conference on computer and information science.
- [3] Mr. Sarda A. T., Mrs. Kulkarni A. R. : Text Summarization using neural networks and Rhetorical Structure Theory, IJARCCE
- [4] Jose Vieira, Fernando Morgado Diasa, Alexandre Mota : Neuro fuzzy systems: A Survey
- [5] Rucha S. Dixit, Prof. S. S. Apte, Improvement of Text Summarization using Fuzzy logic based method, IOSR Journal of Computer Engineering
- [6] Taeho Jo, K Nearest neighbor for text summarization using feature similarity, ICCCEE, 2017.
- [7] Saif alZahir and Qandeel Fatima, Martin Cenek : New Graph-Based Text Summarization Method, IEEE, 2015, 978-1-4673-7788-1/15
- [8] Sebastian Suarez Benjumea, Elizabeth Leon Guzman : Genetic Clustering Algorithm for Extractive Text Summarization , 978-1-4799-7560-0/15 \$31.00 © 2015 IEEE , DOI 10.1109/SSCI.2015.139
- [9] Atif Khan, Naomie Salim, Haleem Farman, Clustered Genetic Semantic Graph Approach for Multi-document Abstractive Summarization, 978-1-4673-8753- 8/16/\$31.00 ©2016 IEEE
- [10] Arman Kiani –B, M. R. Akbarzadeh –T, Automatic Text Summarization Using: Hybrid Fuzzy GA-GP, 2006 IEEE International Conference on Fuzzy Systems, Canada
- [11] Mohamed ABDEL FATTAH, Fuji REN: Probabilistic Neural Network Based Text Summarization
- [12] Ladda Suanmali, Naomie Salim and Mohammed Salem Binwahlan: Fuzzy Genetic Semantic Based Text Summarization, 978-0-7695-4612-4/11 \$26.00 © 2011 IEEE, DOI 10.1109/DASC.2011.192

Use of Long Short-Term Memory for Enhancing Bi-Lingual Machine Translation Approach

Manish Rana, Mohammad Atique

Department of Computer Science, Sant Gadge Baba Amravati University, Amravati, India

Abstract— Long Short-Term Memory (LSTM) is a specific recurrent neural network (RNN) architecture that was designed to model temporal sequences and their long-range dependencies more accurately than conventional RNNs. In this paper, we explore LSTM RNN architectures and made some changes for its better performance. LSTM RNNs are more effective than DNNs. Here, we have changed the gates calculation and also have removed some unnecessary features of standard LSTM architecture. This architecture makes more effective use of model parameters than the others considered, converges quickly, and outperforms a deep feed forward neural network having an order of magnitude more parameters.

Keywords— Long Short-Term Memory, LSTM, recurrent neural network, RNN.

I. INTRODUCTION:

The Deep Neural Network (DNN) is an extremely expressive model that can learn highly complex vector-to-vector mappings. The Recurrent Neural Network (RNN) is a DNN that is adapted to sequence data, and as a result the

RNN is also extremely expressive. RNNs maintain a vector of activations for each time step, which makes the RNN extremely deep. Their depth, in turn, makes them difficult to train due to the exploding and the vanishing gradient problems [3] [13] [14].

There have been a number of attempts to address the difficulty of training RNNs. Vanishing gradients were successfully addressed by

Hochreiter & Schmidhuber (1997), who developed the Long Short-Term Memory (LSTM) architecture, which is resistant to the vanishing gradient problem. The LSTM turned

out to be easy to use, causing it to become the standard way of dealing with the vanishing gradient problem. Other attempts to overcome the vanishing gradient problem include the use of the powerful second order optimization algorithms [18] [19] and regularization of the RNN's weights that ensures that the gradient does not vanish [23], giving up on learning the recurrent weights altogether [15] [16] and a very careful initialization of RNN's parameters [25] [26]. Unlike the vanishing gradient problem, the exploding gradient problem turned out to be relatively easy to address by simply enforcing a hard constraint over the norm of the gradient [20] [23].

A criticism of the LSTM architecture is that it is ad-hoc and that it has a substantial number of components whose purpose is not immediately apparent. As a result, it is also not clear that the LSTM is an optimal architecture, and it is possible that better architectures exist.

Motivated by this criticism, we attempted to determine whether the LSTM architecture is optimal by means of an extensive evolutionary architecture search. We found specific architectures similar to the Gated Recurrent Unit

(GRU) [6] That outperformed the LSTM and the GRU by on most tasks, although an LSTM variant achieved the best results whenever dropout was used. In addition, by adding a bias of 1 to the LSTM's forgetting gate. We can close the gap between the LSTM and the better architectures.

II. LONG SHORT-TERM MEMORY:

In this section we will briefly explain LSTM architecture. The figure 1 is the traditional LSTM architecture. Standard RNNs suffer from both exploding and vanishing gradients [3] [13]. Both problems are caused by the RNN's iterative nature, whose gradient is essentially equal to the recurrent weight matrix raised to a high power. These iterated matrix powers cause the gradient to grow or to shrink at a rate that is exponential in the number of time steps. The exploding gradients problem is relatively easy to handle by simply shrinking gradients whose norms exceed a threshold, a technique known as gradient clipping [20] [23]. While learning would suffer if the gradient is reduced by a massive factor too frequently, gradient clipping is extremely effective whenever the gradient has a small norm the majority of the time.

The full LSTM's definition includes circuitry for computing S_t and circuitry for decoding information from S_t . Unfortunately; different practitioners use slightly different LSTM variants. In this work, we use the LSTM architecture that is precisely specified below. It is similar to the architecture of [10] but without peep-hole connections:

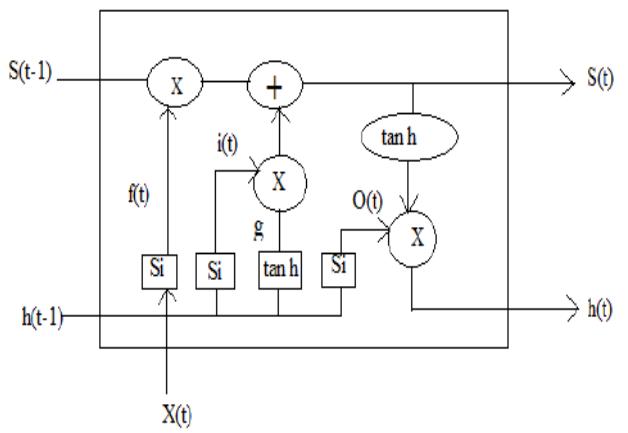


Fig 1: LSTM architecture

In the above LSTM architecture the symbols are defined as,

$S(t-1)$: Previous cell status

$h(t-1)$: Previous cell hidden state

$f(t)$: Forget gate

$i(t)$: Information gate

S_i : Sigmoid function

$X(t)$: Current input

X : Vector multiplication, in this paper it is represented by * notation.

$O(t)$: Output

$S(t)$: Current cell status

$h(t)$: Current cell hidden state

In this LSTM architecture the cell status store cell status. Based on current input LSTM takes decision that how much past information is to delete. This action is performed with the help of Forget gate. Once past information is deleted then new information is added to the cell using Information gate. The equations are,

$$i(t) = Si (Wxi Xt + Whi ht-1 + bi)$$

$$f(t) = Si (Wxf Xt + Whf ht-1 + bf)$$

$$g = \tanh (Wxg Xt + Whgt-1 + bg)$$

$$(t) = Si (Wxo Xt + Who ht-1 + bo)$$

$$S(t) = S(t-1) * f(t) + i(t) * g$$

$$h(t) = \tanh (S(t)) * O(t)$$

'W' is the weight vector initialized randomly. 'b' is the bias value also initialized randomly. All the weight vectors are updated after each iteration.

LSTM architecture learn more and more by training and work good for both long term as well as for short term memory.

III. METHODOLOGY:

To work on this architecture we are creating input data in the program. The created dataset will be in a range defined by the standard for LSTM architecture. For better performance of the architecture we made two changes in the standard architecture which helps the LSTM to work even more efficiently.

Firstly, In standard architecture amount of past information to delete and new information to add was decided separately hence was missed out some use full information. In the new architecture the amount of information to delete is calculated based on the amount of new information required to add. Hence in the new Architecture initially information gate will calculate new information to add and based on information gate output forget gate will calculate amount of information to forget.

Secondly, Due to the use of 'tanh' function while calculating $h(t)$ some use full information was lost hence we decided to remove this 'tanh' function. After removing 'tanh' from the $h(t)$ equation the architecture become more accurate and error rate has been reduced. The new modified architecture is,

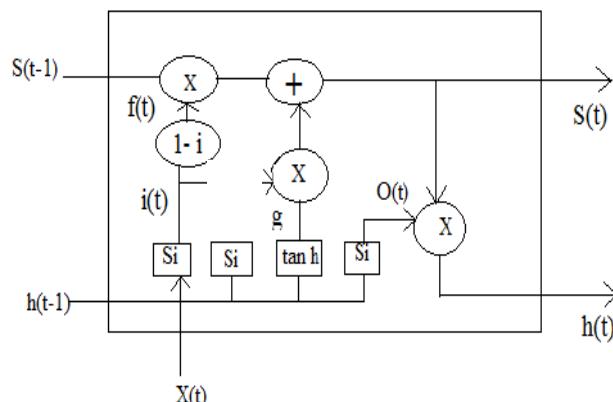


Fig 2: Modified LSTM architecture

After modification new equations are highlighted as,

$$i(t) = Si (Wxi Xt + Whi ht-1 + bi)$$

$$f(t) = (1 - i)$$

$$g = \tanh (Wxg Xt + Whgt-1 + bg)$$

$$(t) = Si (Wxo Xt + Who ht-1 + bo)$$

$$S(t) = S(t-1) * f(t) + i(t) * g$$

$$h(t) = S(t) * O(t)$$

Modified algorithm/architecture will not only work more accurately but also will reduce total execution time.

Result and Discussion:

We have implemented LSTM in python language and executed for multiple changes but this modified version of architecture gives us maximum accuracy. We also worked on the multiple modified architectures by changing normalization functions, changing equations and reconnecting different gates with different gates but all the architecture performed poorly compared to the standard architecture.

Standard LSTM is executed for a set of data for 0 to 99 iterations and the result is shown as the screen shot of the output. This screen short only was showing last part of the output with final loss at the end of the output.

```
cur_iter: 98
y_pred[0] : -0.500087106205
y_pred[1] : 0.200498204588
y_pred[2] : 0.0994746051746
y_pred[3] : -0.499579491524
loss: 7.08662403825e-07
cur_iter: 99
y_pred[0] : -0.500096072911
y_pred[1] : 0.200463300125
y_pred[2] : 0.0995059890975
y_pred[3] : -0.499595630139
loss: 6.31438766408e-07
```

Fig 3: Standard LSTM Output

In figure 3 after last iteration the final loss is 6.31438e-07. This loss is less than all the other loss of diffrent architecture except figure 2 modified LSTM architecture's loss.

Now, modified LSTM when execute for the same data input the result is shown as the screen shot of the output. This screen short only was showing last part of the output.

```
cur_iter: 98
y_pred[0] : -0.499712732281
y_pred[1] : 0.199925730777
y_pred[2] : 0.100122240059
y_pred[3] : -0.500299187737
loss: 1.92494594059e-07
cur_iter: 99
y_pred[0] : -0.499731288362
y_pred[1] : 0.19992779384
y_pred[2] : 0.100116124116
y_pred[3] : -0.500280730381
loss: 1.69714030648e-07
```

Fig 4: Modified LSTM Output

In figure 4 after last iteration the final loss is 1.6971403e-07. Modified LSTM's final loss is less than that of the standard LSTM architecture final loss.

Table I: Loss difference

	Standard LSTM	Modified LSTM
Loss	6.31438e-07	1.6971403e-07

IV. COMPARISON

Experiments with LSTM RNN architectures.

C P Depth N WER (%)

1024 512 3L 20M 10.7

1024 512 2L 15M 10.7

800 512 2L 13M 10.7

700 400 2L 10M 10.8

600 350 2L 8M 10.9

V. CONCLUSION:

Standard LSTM architecture works better than the RNN by handling the vanishing gradient problem. The LSTM architecture is not perfect. To make it more accurate we experimented on it by changing its architecture and hence come up with a new LSTM architecture which works better than that of the standard architecture. Thus, modifying standard LSTM architecture by changing the forget gate structure and removing unnecessary normalization function improved the LSTM performance.

REFERENCES:

- [1] Amit, Daniel J. Modeling brain function: The world of attractor neural networks. Cambridge University Press, 1992.
- [2] Bayer, Justin, Wierstra, Daan, Togelius, Julian, and Schmidhuber, J'urgen. Evolving memory cell structures for sequence learning. In Artificial Neural Networks— ICANN 2009, pp. 755–764. Springer, 2009.
- [3] Bengio, Yoshua, Simard, Patrice, and Frasconi, Paolo. Learning long-term dependencies with gradient descent is difficult. *Neural Networks, IEEE Transactions on*, 5 (2):157–166, 1994.
- [4] Bergstra, James, Breuleux, Olivier, Bastien, Fr'ed'eric, Lamblin, Pascal, Pascanu, Razvan, Desjardins, Guillaume, Turian, Joseph, Warde Farley, David, and Bengio, Yoshua. Theano: a CPU and GPU math expression compiler. In Proceedings of the Python for Scientific Computing Conference (SciPy), June 2010.
- [5] Boulanger-Lewandowski, Nicolas, Bengio, Yoshua, and Vincent, Pascal. Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and the transcription. The arXiv preprint arXiv:1206.6392, 2012.
- [6] Cho, Kyunghyun, van Merriënboer, Bart, Gulcehre, Caglar, Bougares, Fethi, Schwenk, Holger, and Bengio, Yoshua. Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078, 2014.
- [7] Chung, Junyoung, Gulcehre, Caglar, Cho, KyungHyun, and Bengio, Yoshua. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555, 2014.
- [8] Collobert, Ronan, Kavukcuoglu, Koray, and Farabet, Cl'ment. Torch7: A matlab-like environment for machine learning. In BigLearn, NIPS Workshop, number EPFL-CONF-192376, 2011.
- [9] Gers, Felix A, Schmidhuber, J'urgen, and Cummins, Fred. Learning to forget: Continual prediction with lstm. *Neural computation*, 12(10):2451–2471, 2000.
- [10] Graves, Alex. Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850, 2013.
- [11] Greff, Klaus, Srivastava, Rupesh Kumar, Koutn'ik, Jan, Steunebrink, Bas R, and Schmidhuber, J'urgen. Lstm: A search space odyssey. arXiv preprint arXiv:1503.04069, 2015.
- [12] Hinton, Geoffrey E and Shallice, Tim. Lesioning an attractor network: investigations of acquired dyslexia. *Psychological review*, 98(1):74, 1991.
- [13] Hochreiter, Sepp. Untersuchungen zu dynamischen neuronalen netzen. Master's thesis, Institut fur Informatik, Technische Universitat, Munchen, 1991.
- [14] Hochreiter, Sepp and Schmidhuber, J'urgen. Long shortterm memory. *Neural computation*, 9(8):1735–1780, 1997.
- [15] Jaeger, Herbert. The echo state approach to analysing and training recurrent neural networks-with an erratum note. Bonn, Germany: German National Research Center for Information Technology GMD Technical Report, 148:34, 2001.
- [16] Jaeger, Herbert and Haas, Harald. The Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304(5667):78–80, 2004.
- [17] Marcus, Mitchell P, Marcinkiewicz, Mary Ann, and Santorini, Beatrice. Building a large annotated corpus of english: The penn treebank. *Computational linguistics*, 19(2):313–330, 1993.
- [18] Martens, James. Deep learning via the hessian free optimization. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), pp. 735–742, 2010.
- [19] Martens, James Sutskever and Ilya. Learning recurrent neural networks with hessian-free optimization. In Proceedings of the 28th International Conference on Machine Learning (ICML-11), pp. 1033–1040, 2011.
- [20] Mikolov, Tom'a's. Statistical Language Models based on Neural Networks. PhD thesis, PhD thesis, Brno University of Technology, 2012. - , 2012.
- [21] Mikolov, Tomas, Karafiat, Martin, Burget, Lukas, Cernock'y, Jan, and Khudanpur, Sanjeev. Recurrent neural network based language model. In INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September 26-30, 2010, pp. 1045–1048, 2010.
- [22] Mikolov, Tomas, Joulin, Armand, Chopra, Sumit, Mathieu, Michael, and Ranzato, Marc'Aurelio. Learning longer memory in recurrent neural networks. arXiv preprint arXiv:1412.7753, 2014.
- [23] Pascanu, Razvan, Mikolov, Tomas, and Bengio, Yoshua. On the difficulty of training recurrent neural networks. arXiv preprint arXiv:1211.5063, 2012.
- [24] Plaut, David C. Semantic and associative priming in a distributed attractor network. In Proceedings of the 17th annual conference of the cognitive science society, volume 17, pp. 37–42. Pittsburgh, PA, 1995.
- [25] Sutskever, Ilya, Martens, James, Dahl, George, and Hinton, Geoffrey. On the importance of initialization and momentum in deep learning. In Proceedings of the 30th International Conference on Machine Learning (ICML-13), pp. 1139–1147, 2013.

- [26] Sutskever, Ilya, Vinyals, Oriol, and Le, Quoc VV. Sequence to sequence learning with neural networks. In Advances in Neural Information Processing Systems, pp. 3104–3112, 2014.
- [27] Zaremba, Wojciech and Sutskever, Ilya. Learning to execute. arXiv preprint arXiv:1410.4615, 2014.
- [28] Zaremba, Wojciech, Sutskever, Ilya, and Vinyals, Oriol. Recurrent neural network regularization. arXiv preprint arXiv:1409.2329, 2014.

Activity Recognition Based Machine Learning Models

Keval Chavda¹, Viral Haria², Bijal Patel³, Anupriya Babbar⁴

Computer Engineering, Thakur College of Engineering and Technology, Mumbai, India

Abstract- Activity Recognition systems are machine learning based models that uses classification. This AR system is developed specifically for the smart devices i.e. wearable or phone to recognize the human activities such as walking, running, swimming and cycling. This paper examines and compares the working and performance of several well recognized models like Logistic Regression, K-nearest neighbors, Support Vector Machine, Naïve Bayes, Decision Tree and Random Forest onto the dataset.

Keywords — *Machine Learning; Accuracy; Artificial Intelligence; Activity Recognition; Classification Models*

I. INTRODUCTION

Activity recognition or AR has been a key delinquent in studies related to multiple sensors. Activity recognition acts as a supporting stride in behavior analysis. Activity recognition converts low-level sensor data from sensors, such as gyroscope, pressure sensor, accelerometer, and GPS location, to a higher-level report of behavior primitives. There are sensors connected to the different parts of the bodies which in turn help to identify different movements related to our body. The multi-sensors differ from each other and rarely show similar results. This makes it difficult to obtain the final result whether what kind of the personation a body is maintaining. These problems are resolved largely using the machine learning algorithms. The sensors need to be correctly placed on the body parts thus making them very difficult to adjust to. At each different steps, the calculations are concatenated to one another and then the final result will be displayed based on the calculations. In this paper, we report the results of multiple experimentations using the several machine learning based approaches, on activity data collected in our simulated environment.

II. BACKGROUND

There has a great number of technological advancements that has shown an evident response to the betterment of the human life expectancy. The AR has caught the fascination of software engineering groups because of its capabilities on giving supporting customized data and its numerous applications in Human-Computer Interaction, medicine, sports, and human science. Machine learning is one such example that can be used for the above mentioned applications. Arthur Samuel described it as: "the field of study that gives computers the ability to learn without being explicitly programmed". We understand that we can apply

supervised classification as well as unsupervised classification for the testing of the datasets. For this paper, machine learning models like Logistic Regression, K-nearest neighbors, Naïve Bayes, Support Vector Machine, Decision Tree and Random Forest have been trained and tested. These algorithms require numerical support for the prediction of the results. The performance of these models have been examined according to both running time and prediction accuracy. In this paper, it is shown that the random forest model provides the best performance among the other classifier models.

III. PROBLEM STATEMENT

Research based studies in Activity Recognition have been centered around developing novel models for uncommon activities and training data, and for data collections. Here, an all-inclusive study has been performed over several well-known supervised learning models. Also, the effectiveness of dimensional reduction using principal component analysis on these models was studied.

IV. METHODOLOGY

A huge dataset composed of 150K instances, consistently distributed over 12 activities has been used in the study. Of this dataset, 80 percent of the data points is used for the training of the dataset and 20 percent of the data points is used for the testing of the dataset. The accuracy and the running time of these models was evaluated. Also, finally, the effect of component analysis on the performance of these models was evaluated.

V. CLASSIFICATION ANALYSIS

The details regarding the training and testing of the models with the datasets is shown in the table. Since the problem is a multi-class classification, the problem was evaluated using the sklearn libraries. Take a note that the Logistic regression is a linear classification model whereas the KNN, SVM, Naïve Bayes, Decision Tree and the Random Forest are nonlinear. Nonlinear classifiers are expected to display better results since the various activities exhibit non linearity while remarking the sensor data. The hyper parameter for every classifier ought to be set to an esteem that expands the precision of the model over the test data. The Principal Component Analysis is a statistical procedure used to convert the observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. Based on these results equivalent predictions is made by the Machine Learning model.

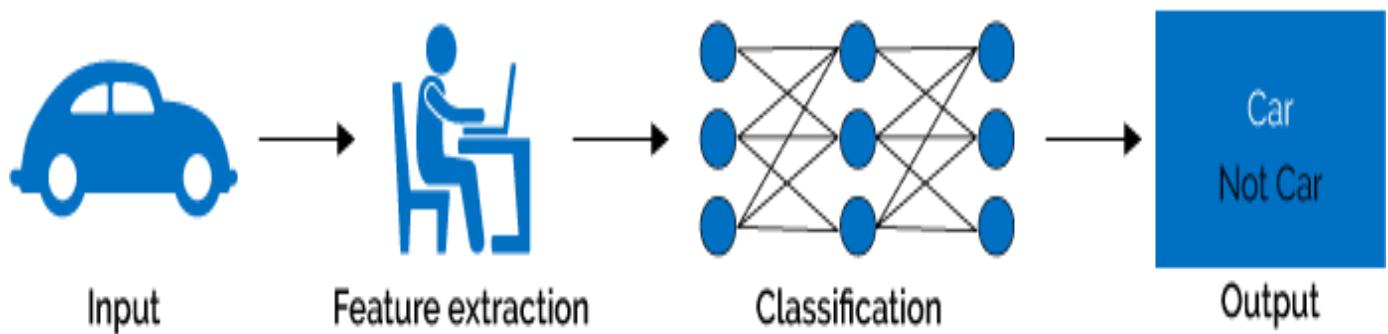
A. Logistic Regression

The logistic regression would probably be the one of the most poorly showing model for the accuracy. This is because it a linear classifier and thus cannot easily adapt to drastic changes.

C. SVM (Support Vector Machine)

An SVM model is a representation of the instances as points in space, plotted so that the instances of the distinct categories are divided by a clear slit that is as wide as possible. Newer data points are then mapped into the same space and prediction is made whether to which category the data points belongs to, based on

Machine Learning



Deep Learning

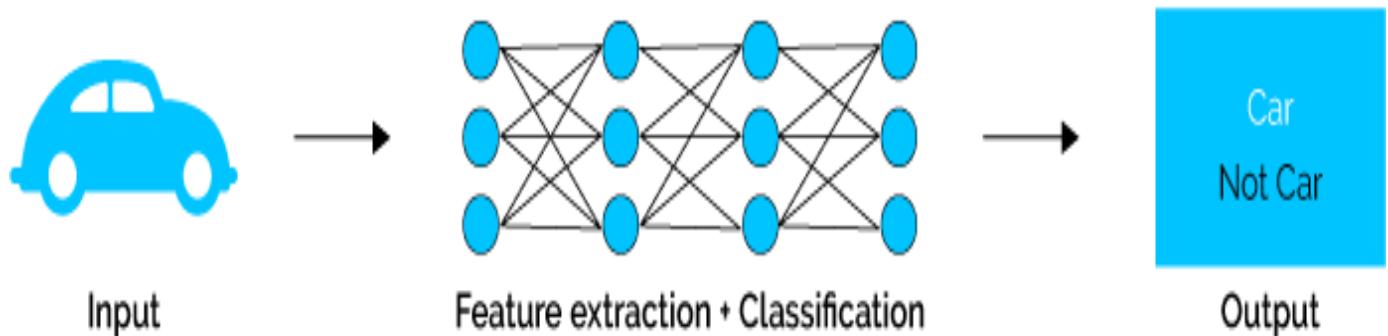


Figure 1. Machine Learning and Deep Learning

B. K-Nearest Neighbor

This classifier chooses the k nearest neighbors and predicts the newer data point using the Euclidean distance. The newer data point is assigned to the category with the most neighbors.

which side of the gap they fall.

D. Naïve Bayes:

Naïve Bayes is a family of probabilistic algorithms that take the use of probability theory and Bayes' Theorem to predict the category of a given subject. They are probabilistic, i.e. that they compute the probability of each category for a given subject, and then output the category with the highest one. It predicts the subject's datasets with the help of the Bayes theorem that uses the predefined values of the two sets. For example, it feeds whether an employee drives to work or walks to work and predict if a new employee is added whether it will walk or drive to work.

$$P(A|B) = (P(B|A) P(A))/P(B)$$

E. Decision Tree:

Decision trees are a kind of Supervised Machine Learning where the information data is ceaselessly fragmented according to a specific parameter. The decision tree can be clarified by two entities, decision nodes and decision leaves. The leaves can be stated as the final outcome and the nodes as where the data is fragmented.

F. Random Forest:

It is a type of an ensemble learning method. It operates by construction of multiple decision trees. These trees are then input to the training data. We need to be aware not to over fit the training dataset since it can cause indefinite problems into the predictions.

VI. RESULTS AND DISCUSSION

A. Description

For the testing purposes, the dataset used comprises of a collection of 4 sensors worn by 4 subjects while performing certain activity in total of about 8 hours. The dataset are recorded at every subject's belt, thigh, arm and ankle. Every activity was performed independently by the subjects.

B. Results

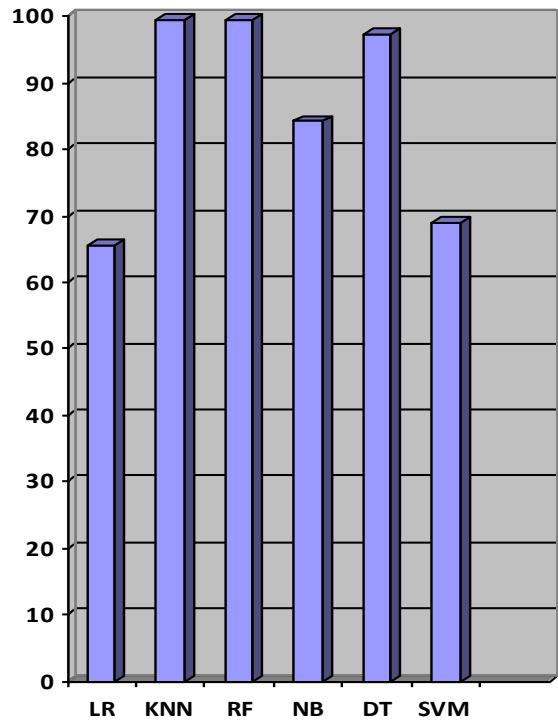


Figure 2. Prediction accuracy for all the models

This paper provides us with the various methods used to identify the activity. All the classifiers are used to test the AR dataset and then the readings were noted.

As displayed in the above chart, both K-nearest neighbors and Random Forest classifiers provided accuracy of about 99%. Whereas the Decision Tree, SVM provided with an accuracy of about 98%. However, Logical Regression and Naïve Bayes appeared to perform relatively poor to the others. As observed the linear recognition methods do not provide much accuracy. Like for example, if there is a nonlinear relationship between the predictors and outcome, in decision trees then accuracy will suffer. Random forest on the other provides better accuracy than decision trees as it combines multiple weak trees from bootstrapped predictors. Advanced ensemble techniques or techniques such as SVM also provides better accuracy than single decision tree. But however, the Random forest method gives the most accurate result with an accuracy of 99.8.

Apart from the accuracy, the running time is also a major factor for the performance of the classifier models. Random Forest and Decision Tree appeared to be the fastest of them all. On contrast, SVM and KNN have shown to utilize long time to execute.

VII. CONCLUSION

In this paper, the Activity Recognition (AR) was tested out and compared using the several classification models in the machine learning. Non-linear models such as KNN and random forest proved to work best mainly because the dataset contained non-linear points. However by considering the running time for comparison, random forest was the only one to show the fastest time when compared to the other models.

REFERENCES

- [1] <http://technav.ieee.org/tag/1074/machine-learning>
- [2] <http://www.jmlr.org/>
- [3] <https://www.omicsonline.org/open-access/a-comparative-study-on-machine-learning-classification-models-for-activity-recognition>
- [4] <https://www.superdatascience.com/machine-learning>
- [5] <http://www.statsoft.com/Textbook/Support-Vector-Machines>
- [6] <https://medium.com/fuzz/machine-learning-classification-models-3040f71e2529>
- [7] <https://www.geeksforgeeks.org/regression-classification-supervised-machine-learning/>
- [8] Tom M. Mitchell, “Decision Tree”, “Bayesian Learning”, Machine Learning, 1st edition
- [9] Yaser S. Abu-Mostafa, “Overfitting”, “Support Vector Machines”, Learning from Data, 1st edition
- [10] [14]Eunju Kim, Diane Cook and Sumi Helal: Human Activity Recognition and Pattern Discovery, IEEE International Conference on Pervasive Computing, 2010.

Incorporating Copying Mechanism in Sequence-to-Sequence Learning

Sansriti Mishra¹, Gautam Mishra², Anand Khandare³

¹*S.P. Jain School of Global Management, Mumbai, India*

²*Illinois Institute of Techology, Chicago, USA*

³*Thakur College of Engineering and Technology, Mumbai, India*

sansritimishra3122@gmail.com, mishra.a.gautam@gmail.com, anand.khandare@thakureducation.org

Abstract—We address an important problem in sequence-to-sequence (Seq2Seq) learning referred to as copying, in which certain segments in the input sequence are selectively replicated in the output sequence. A similar phenomenon is observable in human language communication. For example, humans tend to repeat entity names or even long phrases in conversation. The challenge regarding copying in Seq2Seq is that new machinery is needed to decide when to perform the operation. In this paper, we incorporate copying into neural network-based Seq2Seq learning and propose a new model called CopyNet with encoder-decoder structure. CopyNet can nicely integrate the regular way of word generation in the decoder with the new copying mechanism which can choose sub-sequences in the input sequence and put them at proper places in the output sequence. Our empirical study on both synthetic data sets and real-world data sets demonstrates the efficacy of CopyNet. For example, CopyNet can outperform regular RNN-based model with remarkable margins on text summarization tasks.

I. INTRODUCTION

Chatbot is an active research topic recently. The work can also be applied to other NLP tasks such as Machine Translation, Abstract Summarization. The general chatbot framework is all about a sequence-to-sequence model (seq2seq). This is straightforward, easier to implement. But of course, there are quite lots of space for an improvement. One of the them is about copying. In conversation, we normally have some repeated text as an example below

Hello, my name is chatbot

Nice to meet you, chatbot.

In general, seq2seq apparently does not handle repeated text well. Basically, the idea is to write simple code, and then ask seq2seq to execute the code. One interesting thing we learnt from the paper is that seq2seq is not good enough to execute a copy command, in general. For example:

command: print (123456789)

output: 123565756 (I made up the number, but you get the idea).

Also, as the input is longer, the seq2seq can produce less accurate. How to address that problem? In general, having an external memory is a very nice solution, I think. But it is just for learning to execute. Conversation is much harder problem, at least with current basic sample codes we want a seq2seq to execute.

Background on Neural Machine Translation

Back in the old days, traditional phrase-based translation systems performed their task by breaking up source sentences into multiple chunks and then translated them phrase-by-phrase. This led to disfluency in the translation outputs and was not quite like how we, humans, translate. We read the entire source sentence, understand its meaning, and then produce a translation. Neural Machine Translation (NMT) mimics that!

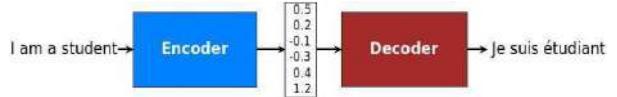


Fig 1. Encoder Decoder Architecture

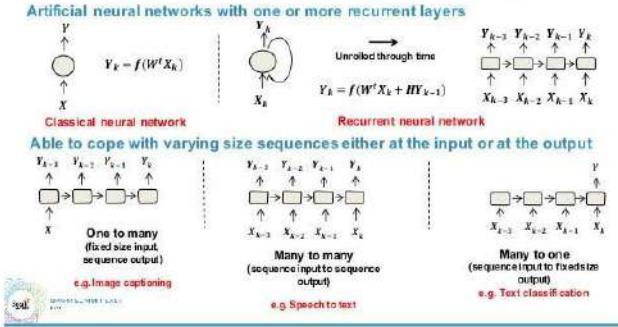
Fig 1. Encoder-decoder architecture – example of a general approach for NMT. An encoder converts a source sentence into a "meaning" vector which is passed through a decoder to produce a translation.

Specifically, an NMT system first reads the source sentence using an encoder to build a thought vector, a sequence of numbers that represents the sentence meaning; a decoder, then, processes the sentence vector to emit a translation, as illustrated in Figure 1. This is often referred to as the encoder-decoder architecture. In this manner, NMT addresses the local translation problem in the traditional phrase-based approach: it can capture long-range dependencies in languages, e.g., gender agreements; syntax structures; etc.

A goal of statistical language modeling is to learn the joint probability function of sequences of words in a language. This is intrinsically difficult because of the curse of dimensionality: a word sequence on which the model will be tested is likely to be different from all the word sequences seen during training. Traditional but very successful approaches based on n-grams obtain generalization by concatenating very short overlapping sequences seen in the training set. We propose to fight the curse of dimensionality by learning a distributed representation for words which allows each training sentence to inform the model about an exponential number of semantically neighboring sentences. The model learns simultaneously (1) a distributed representation for each word along with (2) the probability function for word sequences, expressed in terms of these representations. Generalization is obtained because a sequence of words that has never been seen before gets high probability if it is made of words that are similar (in the sense of having a nearby representation) to words forming an already seen sentence. Training such large models (with millions of parameters) within a reasonable time is itself a significant challenge. We report on experiments using neural networks for the probability function, showing on two text corpora that the proposed approach

significantly improves on state-of-the-art n-gram models, and that the proposed approach allows to take advantage of longer contexts.

Recurrent Neural Network basics



A. RNN Encoder Decoder

The encoder-decoder architecture for recurrent neural networks is the standard neural machine translation method that rivals and, in some cases, outperforms classical statistical machine translation methods. This architecture is very new, having only been pioneered in 2014, although, has been adopted as the core technology inside Google's translate service

Encoder-Decoder Architecture for NMT: -

The Encoder-Decoder architecture with recurrent neural networks has become an effective and standard approach for both neural machine translation (NMT) and sequence-to-sequence (seq2seq) prediction in general. The key benefits of the approach are the ability to train a single end-to-end model directly on source and target sentences and the ability to handle variable length input and output sequences of text.

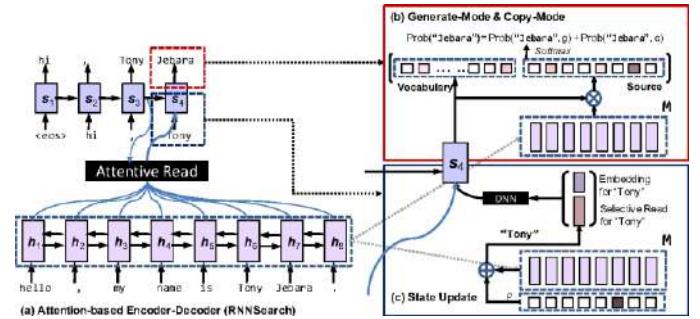
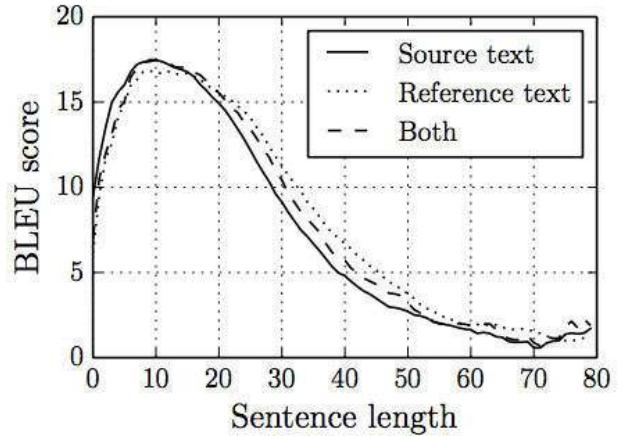
Model

An Encoder-Decoder architecture was developed where an input sequence was read in entirety and encoded to a fixed-length internal representation. A decoder network then used this internal representation to output words until the end of sequence token was reached. LSTM networks were used for both the encoder and decoder. The idea is to use one LSTM to read the input sequence, one timestep at a time, to obtain large fixed-dimensional vector representation, and then to use another LSTM to extract the output sequence from that vector. The final model was an ensemble of 5 deep learning models. A left-to-right beam search was used during the inference of the translations.

Loss in model skill with increased sentence length.

Copynet Model: - Seq2Seq learning heavily rely on the “meaning” for each word in source and target sequences, however, some words in sequences are “no-meaning” symbols and it is improper to encode them in encoding and decoding processes. For example, generating the response “Of course, read” for replying the message “Can you read the word ‘read?’” should not consider the meaning of the second “read”. By incorporating the copying mechanism, the decoder could directly copy the sub-sequences of source into the target. The basic approach is to jointly predict the indexes of the target word in the fixed vocabulary and/or matched positions in the source sequences

Identify applicable funding agency here. If none, delete this text box.



COPYNET MECHANISM

COPYNET follows the general encoder-decoder pattern and uses a bi-directional RNN to encode the source sequence. The encoded representation can be thought of as a short-term memory, M.

The decoder reads M and predicts the target sequence. It uses a similar attention mechanism to Bahdanau et al., with a few important differences:

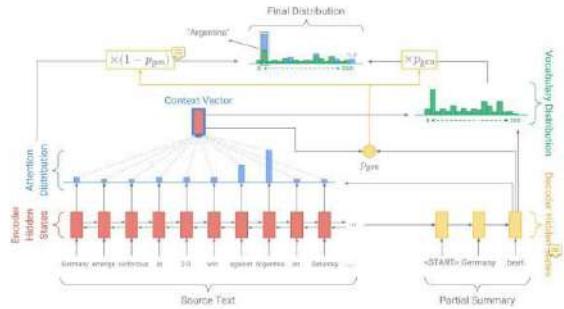
- Words are predicted based on a probabilistic model combining two modes: a generate mode and a copy mode. The generate mode uses the same scoring function as a generic encoder-decoder (see e.g. Bahdanau et al.). Copy mode picks words from the source sequence using the hidden states in M to represent each word, using a non-linear activation function (\tanh). The two modes are combined with a shared normalization term, and so are basically competing through a SoftMax function
- CopyNet uses both the previous state and a weighted sum of the hidden states in M in order to update each decoding state at every step. This selective read is designed for the copy mode and focuses attention on the source sequence encoded in the hidden state. A properly trained encoder will have captured both the semantics of a word and its location in the input in the hidden states in M. We hypothesize that COPYNET uses a hybrid strategy for fetching the content in M, which combines both content-based and location-based addressing. Both addressing strategies are coordinated by the decoder RNN in managing the attentive read and selective read, as well as determining when to enter/quit the copy-mode... Unlike the explicit design for hybrid addressing in the Neural Turing machine, COPYNET is subtler; it provides the architecture that can facilitate some particular location-based addressing and lets the model figure out the details from the training data for specific tasks.

COPYNET EXPERIMENTS

The authors use CopyNet in three different tasks: a synthetic dataset to show that it can learn rules requiring copying of symbols outside of its vocabulary (it can); a text summarization task; and simple single-turn dialogues. On text summarization, CopyNet ‘beats the competitor models by a big margin.’ But it’s the dialogue performance we’re most interested in this week.

- Dialogue data is collected from Baidu Tiega with real-life conversations covering greetings and sports etc.
- Patterns are mined from the set, e.g. “Hi, my name is Adrian” followed by “Hi, Adrian” can lead to the pattern “hi, my name is X -> hi, X” simply by looking for copied subsequences.
- The dataset is enlarged by filling the slots with suitable subsequences (e.g. name entities, dates etc.).

Using this slot filling, two datasets are created based on 173 collected patterns. CopyNet is able to accurately replicate critical segments from the input using copy mode and generates the rest of the answers smoothly using the generate mode.



Pointer-generator model. For each decoder timestep a generation probability $p_{gen} \in [0, 1]$ is calculated, which weights the probability of generating words from the vocabulary, versus copying words from the source text. The vocabulary distribution and the attention distribution are weighted and summed to obtain the final distribution, from which we make our prediction. Note that out-of-vocabulary article words such as 2-0 are included in the final distribution. Best viewed in color.

II. CONCLUSION OF THE EXPERIMENT

The copying attention mechanism is similar to the original sequence to sequence attentive mechanism. However, instead of simply using the encoder-generated hidden states as an additional input to the decoder, we create a probabilistic model to predict whether to use the hidden state to generate a word from the training vocabulary or to use the hidden state as positional information to copy a word from the article instance-specific source vocabulary. In the generate mode, the model uses each hidden state as in the original sequence to sequence attentive mechanism. In the copy mode, the model uses each hidden state as information about the word at the corresponding positional index. When determining the mode and the next word in the summary, we want to consider the probabilities of all possible words that appear in the union of the training vocabulary set and the source vocabulary set. The probabilities for each (word, mode) pair are calculated then combined to get the probability of each word. The probabilities generated by $P(g)$ is given by equations $\psi g = v^T c$ Where v and c are parameters that are being learned similar to word embeddings. The scoring function for the copy probabilities is given by $\psi c = \sigma(Wc h_j) s_t$ Where h_j is the hidden state corresponding to the j th word, s_t is the hidden decode state at time t , and w is a parameter. The idea of CopyNet is similar to

Pointer Networks in that both architectures allow the models to take advantage of the instance-specific source text (described in Gulcehre et al. and used in Nallapati et al.). However, Pointer Networks don’t combine the probabilities of words across the source and generating a word form the vocabulary, a target word is chosen from the entire probability distribution over both generate and copy modes as follows. $p(yt|st, ct, M, yt-1) = p(yt, g|st, ct, M, yt-1) + p(yt, c|st, ct, M, yt-1)$ Because we now have a chance to copy a word from the source, we need to embed that word, so we represent the previous word with a concatenation of the word embeddings and a weighted average of the hidden states from the source text, $yt-1 = [e(yt-1); \zeta(yt-1)]$. Here $e(yt-1)$ is given by the word embedding for word $yt-1$, which if the previous word was not in the vocab, is just 0s. Similar to the context vector, $\zeta = X J \beta t j h_j$ Image from [7]. 6 Figure 3: Copynet: architecture for copy and generation modes. 5 $\beta t j = e \psi c(h_j, st-1)$ $P k \psi c(h_k, st-1)$ Where h_j is the j th hidden state in the source text, and $p(c)|j$ is the probability for producing This allows us to also update the next hidden state with this new “embedding” and allows for the decode step to learn from the source text itself as well as use semantic meaning not only from our word embeddings, but directly from the source which makes it more powerful as we have the meaning of the word in the context in which it is defined. This is the main difference We generate a new state by using the LSTM by feeding it in the $yt - 1$, $ht - 1$, and ct . By training our new model end to end with this new copying attention mechanism, the network learned the appropriate parameters in order to use the encoder hidden state and chose whether to generate or copy the next word, and then chose the right word out of the source vocab or the training vocab. 4.5 Results Preprocessing Learning model ROUGE-1 ROUGE-2 ROUGE-L 1 None Textsum 0.08600 0.00771 0.07875 2 Tokenized Textsum 0.08154 0.00941 0.07623 3 Tokenized Textsum w/ Copy 0.04869 0.00465 0.04494 4 Entity-tagged Texsum 0.06434 0.00380 0.05730 5 Entity-tagged Textsum + GloVe 0.04035 0.01290 0.03820 6 Tokenized Texsum + GloVe 0.12123 0.03604 0.11342 Generally, tokenization did not produce noticeable gains, potentially because it increased the number of tokens in our source and summary texts, while we kept the number of encoding timesteps (i.e. input length) at 120 tokens. We truncated any articles and gold summaries that exceeded this length, which may have caused us to lose some information in certain examples. The model with the copying mechanism (experiment 3 in the table) was trained on a very small dataset of about 10,000 articles until the loss was below 0.01 as an initial experiment. Unfortunately, we were not able to train a full 10 epochs in time to produce meaningful decoding results. The training time is much greater because of the increased number of parameters we needed to train. However, on our small experimental subset of the dataset, the copying mechanism did successfully copy proper nouns and other tokens from the source texts. This gave the summaries an appropriate 7 subject (whereas our baseline attentive sequence to sequence summarizer often missed the correct main subject in the summary). For example: First 2 sentences: Police have uncovered a £1million ‘Downton Abbey’ mansion which had been transformed into a cannabis factory containing more than 1,000 plants. Wendreda House in March, Cambridge shire, was raided by officers yesterday after an anonymous tip-off about a pungent smell coming from the three-

story property. First highlight: wendreda house was raided by officers after a pungent smell was reported Copying mechanism summary: 1,000 1,000 plants wendreda house march march was by raided raided was , , march in wendreda wendreda plants Textsum summary manor ' mansion was raided by officers The copying mechanism summary shows many repeated words in sentences, something not seen in other models. We believe this to be in large part due to a lack of training time. After the few epochs that we were able to train on both the small experimental training set and the full training set, the language model was not fully trained, and the copy mechanism seemed to be dominating the output generation. Our best results came from our 6th experiment, as we combined a tokenized dataset and the pretrained embedding generated from Glove. We think this model did the best because it solved the problem of using such a small training set. Our model greatly benefitted from word embeddings that had been trained on a much larger corpus. This can easily be seen from our examples with this model as the words chosen are pertinent and salient to the article. However, we saw an increased amount of tokens in this model.

5 Conclusion We introduced a unique combination of copying attention and generation to an encoder-decoder model with a bidirectional LSTM-RNN encoder and RNN decoder. We completely integrated this new attention mechanism into the existing complex Textsum architecture. Adding Glove embeddings helped give semantic meaning to the input words, which helped improve the relevance of the output summaries to the corresponding input articles. Adding a copying attention mechanism to augment our attentive generation mechanism greatly reduced the number of tokens that were generated by our model. Finally we realize that the quality of our summaries often did not correlate strongly with their assigned Rouge scores. As

Rouge uses unigram and bigram counts to evaluate the summary, this makes it hard to mimic any human summary in news articles because what is relevant can differ greatly for people, especially in longer and longer articles such as the ones found in the CNN/Dailymail dataset. We could always see a significant difference between our unigram and bigram scores as its easier to pick up on the important words in the document, but because of the nature of abstractive summaries, what goes in between can be anything that makes sense for that language, ending with very low Rouge-2 scores even when the summary seems perfectly acceptable.

REFERENCES

- [1] Shiqi Shen, Yong Cheng, Zhongjun He, Wei He, Hua Wu, Maosong Sun, and Yang Liu. 2015. Minimum risk training for neural machine translation. CoRR, abs/1512.02433.
- [2] Alessandro Sordoni, Michel Galley, Michael Auli, Chris Brockett, Yangfeng Ji, Margaret Mitchell, Jian-Yun Nie, Jianfeng Gao, and Bill Dolan. 2015. A neural network approach to context-sensitive generation of conversational responses. arXiv preprint arXiv:1506.06714.
- [3] Rupesh Kumar Srivastava, Klaus Greff, and Jürgen Schmidhuber. 2015. Highway networks. arXiv preprint arXiv:1505.00387. Ilya Sutskever, Oriol Vinyals, and Quoc VV Le. 2014. Sequence to sequence learning with neural networks. In Advances in neural information processing systems, pages 3104–3112. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. (references)
- [4] Auli, M. Pointing the Unknown Words. 2016, 140–149. [2] Gu, J., Lu, Z., Li, H. [2]
- [5] Alexander M Rush, Sumit Chopra, and Jason Weston. 2015. A neural attention model for abstractive sentence summarization. arXiv preprint arXiv:1509.00685.

Auto Text Summarization

Dnyaneshwar Dhangar, Richa Tanna, Akanksha More, Ankita Thakur

*dnyaneshwar.dhangar@mctrgit.ac.in, rtanna65@gmail.com, akankshamore59@gmail.com,
ankitathakur1102@gmail.com*

Abstract: Research in text summarization is predominantly targets on measure of the worth of sentences for a summary. The proposed work has associated the Deep learning algorithm with fuzzy logic to improve the efficiency of the generated summary. The proposed work has two phases, they are training phase and testing phases. The training phase utilized to extract the benefits of fuzzy logic and deep learning algorithm for the efficient summary generation. Similar to every training phase, the proposed training phases is also possessed with well-known data and attributes. Latter to the training phase, the testing phases is implemented to check the efficiency of the proposed approach. The experimental evaluation of the proposed work provided the predictable results as, the average precision obtained is 0.37, the average recall is 0.86 and the average f-measure is obtained as 0.50%.

Keywords: Categorization, Feature matrix, Fuzzy Logic, Sentiment analysis.

I. INTRODUCTION

With the rapid growth in the quantity and complexity of documents sources on the internet, it has become increasingly important to provide improved mechanism to user to find exact information from available documents. Text summarization has become an important and timely tool for helping and interpreting the large volumes of text available in documents. Automatic document summarization is the summary of the source version of the original text while keeping its main content and helps the user to quickly understand large volumes of information. In this paper, a method for document summarization is proposed based on deep learning algorithm associated with fuzzy logic. The recent studies have showed that, the deep learning algorithms more impact on the text summarization process by pointing the most relevant objects from set of objects.

A. PURPOSE: Business leaders, analysts, paralegals, and academic researchers need to go through huge numbers of documents every day to keep updated, and a large amount of their time is spent just to figure out what documents are relevant and what are not. By extracting important sentences and creating comprehensive summaries, it's possible to quickly assess whether or not a document is worth reading. Automatic text summarization is also useful for students and authors. Imagine being able to automatically generate an abstract based for your research paper or chapter in a book in a clear and concise way that is faithful to the original source material. With the growing amount of data in the world, interest in the field of automatic summarization generation has been widely increasing so as to reducing the manual effort of a person working on it.

B. SCOPE

Automatic summarization involves reduction of a text file into a passage or paragraph that conveys the main meaning of the text. The searching of important information from a large text file is very difficult job for the users thus to automatic extract the important information or summary of the text file. This summary helps the users to reduce time instead of reading the whole text file and it provide quick information from the large document. In today's world to extract information from the World Wide Web is very easy. This extracted information is a huge text repository. With the rapid growth of the World Wide Web (internet), information overload is becoming a problem for an increasing large number of people. Automatic summarization can be an indispensable solution to reduce the information overload problem on the web.

II. SYSTEM ANALYSIS

A. EXISTING SYSTEM: Yan Liu et al have proposed a document summarization framework via deep learning model, which has demonstrated distinguished extraction ability in document summarization. The framework consists of concepts extraction, summary generation and reconstruction validation. A query-oriented extraction technique has been concentrated information distributed in multiple documents to hidden units layer by layer. Then, the whole deep architecture was fine-turned by minimizing the information loss in reconstruction validation part. According to the concepts extracted from deep architecture, dynamic programming was used to seek most informative set of sentences as the summary. Experiments on three benchmark dataset demonstrate the effectiveness of the framework and algorithm.

B. DISADVANTAGES OF EXISTING SYSTEM

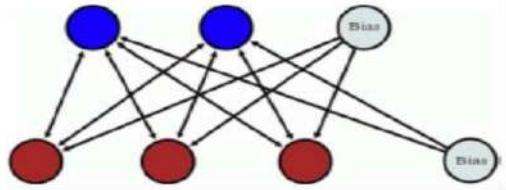
- The existing system is not user friendly and it is difficult to compute summaries using the existing methods and summaries suffer from inconsistencies and lack of balance results in lengthy summary.
- Detailed Information is not present.
- Existing methods does not give quick overview. It provides authors view not the user friendly view.
- Existing systems cannot summarise multiple documents of the same type.
- Existing systems cannot handle different languages.

III. PROPOSED SYSTEM

Proposed System consists of:

A. Restricted Boltzmann Machine

RBM is a stochastic neural network (that is a network of neurons where each neuron has some random behaviour when activated). It consists of one layer of visible units (neurons) and one layer of hidden units. Units in each layer have no connections between them and are connected to all other units in other layer as shown below in Figure.



Connections between neurons are bidirectional and symmetric. This means that information flows in both directions during the training and during the usage of the network and those weights are the same in both directions.

B. Preprocessing

Initially, the input to the proposed approach is a set of document from DUC 2002.Dataset that has to be summarized. The document utilized for text summarization is organized by a set of pre-processing steps like, sentence segmentation, stop words removal and stemming.

C. Segmentation

It is performed by identifying the delimiter commonly denoted by “.” called as full stop. This step is used to separate the sentences in the document. It is mainly useful for the user to understand each individual sentence which is there in the document.

D. Stop Words Removal

Stop words are removed mainly to reduce the insignificant and noisy words. These are predefined words such as a, an, in, by, etc., are called stop words which are filtered out before the pre-processing phase from the documents.

E. Stemming

Stemming is process of bringing the word to its base or root form for example using words singular form instead of using the plural. It basically removes the prefix and suffix of the concerned word to get the base form. There are many more number of algorithms, which are called as stemmers used to perform the stemming process.

F. Training Phase

On behalf of the training phase, the proposed approach defines five features sets. The feature sets are listed as follows,

G. Title Similarity Feature

The ratio of the number of words in the sentence that occur in title to the total number of words in the title helps to calculate the score of a sentence for this feature and it is calculated by the formula given below

$$\text{Title Feature } (f1) = \frac{|S \cap t|}{t}$$

H. Positional Feature

To calculate the positional score of sentence, the proposed approach considers the following conditions. If the sentence given is in the starting of the sentence or the last in the sentence of the

paragraph then the feature value f_2 is assigned as 1. Else if the sentence is in the middle of the paragraph then the feature value of f_2 is assigned as 0.

I. Term Weight Feature

The Term Frequency of a word will be given by $\text{TF}(f, d)$ where f is the frequency of the given word and d is text present the document. The Total Term Weight is calculated by Term Frequency and IDF for a document .Here IDF denotes the inverse document frequency which just implies that the term is common or rare across all documents.

$$\text{IDF}(t, D) = \log \left(\frac{D}{d \in D : t \in d} \right)$$

J. Concept Feature

The concept feature from the text document is retrieved using the mutual information and windowing process. In windowing process a virtual window of size „k“ is moved over document from left to right. Here we have to find out the cooccurrence of words in same window and it can be calculated by following formula,

$$f_4 \Rightarrow MI(w_i, w_j) = \log 2 \frac{P(w_i, w_j)}{P(w_i) \times P(w_j)}$$

K. POS Tagger Feature

Part of speech tagging is the process of categorizing the words of text on the basis of part of speech category such as noun, verbs, adverb, adjectives, they belong to. Algorithms such as hidden Markov

models, using dynamic programming are used to perform this task. The POS Tags on each document is feature five of the given documents.

L. Association of deep learning with fuzzy logic

The sentence matrix $S = (s_1, s_2, \dots, s_n)$ which is the feature vector set having element as s_i which is set contains the all the five features extracted for the sentence s_i . Here this set of feature vectors S will be given as input to deep architecture of RBM as visible layer. Some random values is selected as bias b_i where $i = 1, 2$ since a RBM can have at least two hidden layer. The whole process can be given by following equation: $S = (s_1, s_2, \dots, s_n)$.where $s_i = (f_1, f_2, \dots, f_n)$, $i \leq n$ where n is the number of sentences in the document.

M. Feature Matrix

Here sentence matrix where $S = (s_1, s_2, \dots, s_n)$ where

$s_i = (f_1, f_2, \dots, f_n)$, $i \leq n$ is the feature vector. The five features are the main attributes of the proposed text summarization algorithm. The whole documents under consideration are subjected for the feature extraction and a set of features are extracted accordingly. Now based on the collected features a feature matrix is formed by mapping the features values. The feature matrix is constructed according to the sentences extracted from the multiple documents. In addition to the five features, an additional attribute also associated with the feature matrix. The addition feature associated with the feature matrix is the class labels for each sentence.

IV. CHARACTERISTICS OF PROPOSED SYSTEM

- User Friendly
- Easily generates summary
- No problem with different languages
- Boost the summary generation time.
-

V. Architecture Diagram

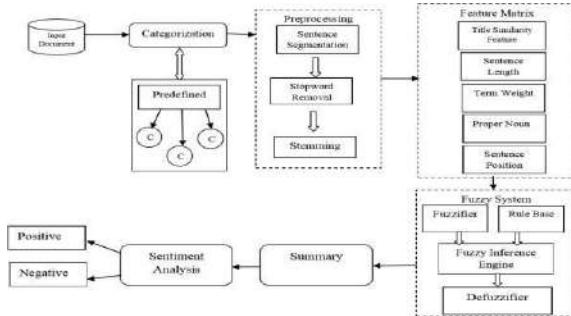


Figure 4.1 Architecture

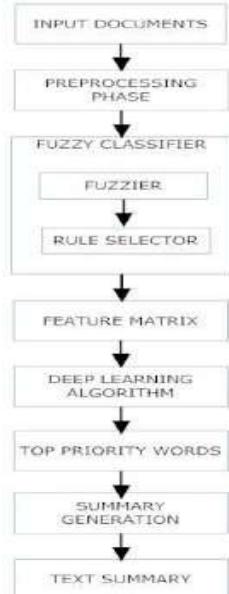


Figure 4.2 Overall Block Diagram of Text Summarization

VI. Methodology

This method considers each characteristic of a text such as sentence length, similarity to little, similarity to key word, etc. as the input of fuzzy system. Then, it enters all the rules needed for summarization, in the knowledge base of system. After that a value from zero to one is obtained for each sentence in the output based on sentence characteristics and the available rules in the knowledge base. The obtained value in the output determines the degree of the importance of the sentence in the final summary. The input membership function for each feature is divided into three membership functions which are composed of insignificant values. The important sentences are then extracted using IF-THEN rules according to the feature criteria.

VII. Software Environment

Front End

- Windows XP, Windows 7,8
- Visual Studio 2010
- Windows Operating System

Back End

- Windows XP, Windows 7,8
- Visual Studio 2010
- MS SQL Server 2008
- Windows Operating System
-

VIII. CONCLUSION

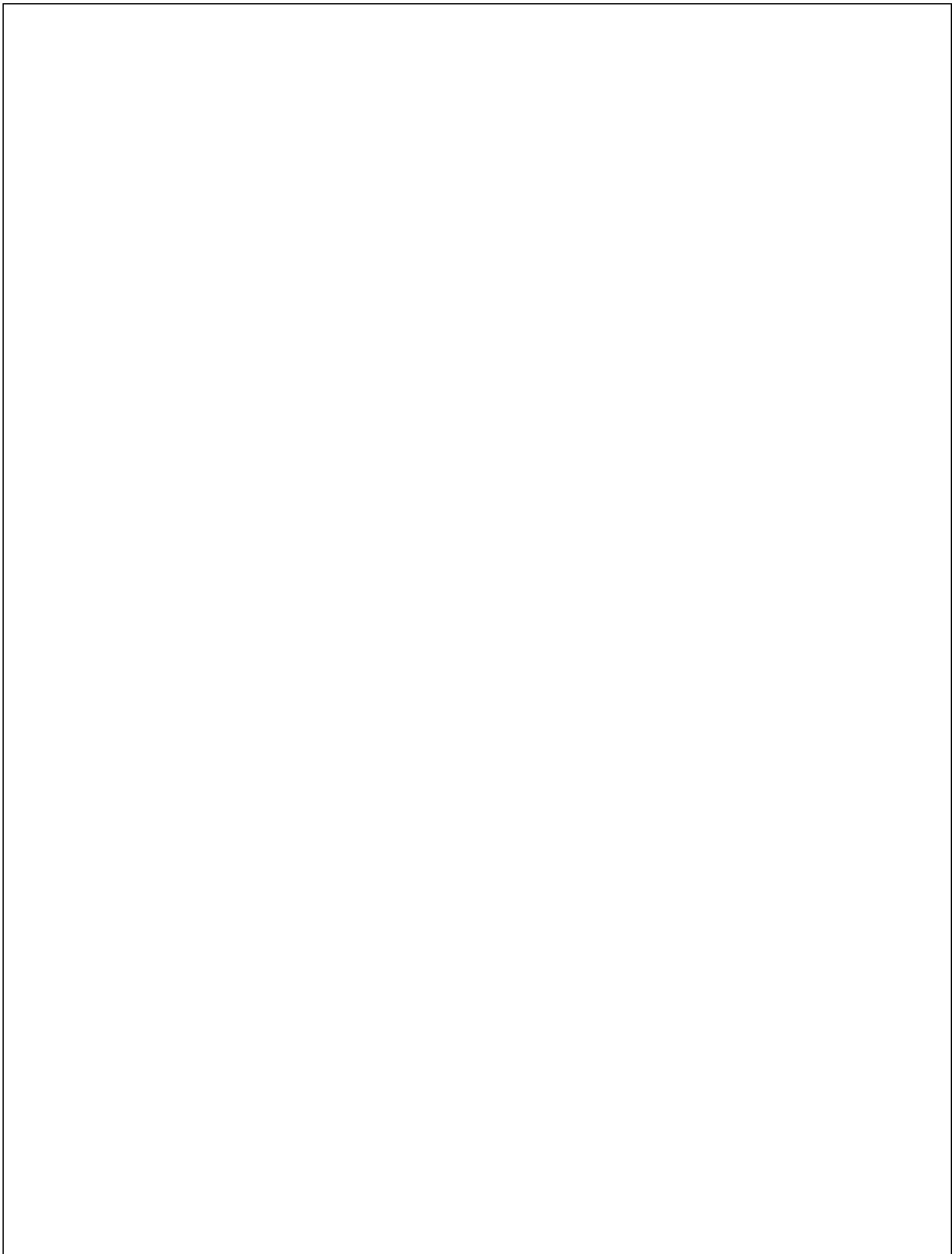
In this proposed work we have extracted five features for feature matrix from the set of sample dataset from DUC2002. The feature matrix is applied to our proposed work which associates the fuzzy logic with deep learning algorithm. The feature matrix is applied through the different levels of the RBM and finally the efficient text summary is generated. The result analysis shows that the proposed work produce the better performance than the existing work based on the evaluation metrics. The maximum Recall, Precision and F- Measure values for the current dataset of the proposed work is obtained as 0.37, 0.86 and 0.50 respectively for the proposed system.

REFERENCES

- [1] Allan Borra, Almira Mae Diola, Joan Tiffany T. Ong Lopez, Phoebus Ferdiel Torralba, Sherwin So, "Using Rhetorical Structure Theory in Automatic Text Summarization for Marcu-Authored Documents", In titaniaaddueduph, 2010.
- [2] B.Arman kiani, M.R. Akbarzadeh, "Automatic Text Summarization using: Hybrid Fuzzy GA-GP" IEEE International Conference on Fuzzy Systems,2006
- [3] C.Gordon and R.Debus, "Developing deep learning approaches and personal teaching efficacy within a preservice teacher education context," Web Intelligence

- and Intelligent Agent Technology, British Journal of Educational Psychology, Vol. 72, No. 4, 2002, PP. 483-511.
- [4] F. kyoomarsi, H. khosravi, E. eslami and M. davoudi, "Extraction-based text summarization using fuzzy analysis," Iranian Journal of Fuzzy Systems, Vol. 7, No. 3, 2010, PP. 15-32.
 - [5] G.Padmapriya, Dr.K.Duraiswamy, "An Approach For Text Summarization Using Deep Learning Algorithm", Journal of Computer Science, vol.6, no.11, 2013, PP.1-9.
 - [6] Inderjeet Mani, "Automatic summarization", Kluwer Academic Publishers, 2004.
 - [7] Jason Weston, Frederic Ratle and Ronan Collobert, "Deep Learning via Semi-Supervised Embedding, "International conference on Machine learning, 2008, PP.1168-1175.
 - [8] Jen-Yuan Yeh , K. Hao-Ren and Wei-Pang Yang , "iSpreadRank: Ranking sentences for extraction- based summarization using feature weight propagation in the sentence similarity network", Expert Systems with Applications, vol. 35, no. 3, October 2008, PP. 1451-1462.

Appendix



Accident Report & Road Surface Detection System

Usama Malim, Sarfaraz Khan, Junaid Ansari, Khan Mubashir

*Anjuman-I-Islam Kalsekar Technical Campus,
Department of Computer Engineering, University of Mumbai*

Abstract--India observe one of the highest road accidents in the world, as stated by highways ministry and Union road transport in 2016. Road Accident often occurs due to the bad road conditions, bad weather and rash driving etc. Accident occurring are often not noticed or they are reported late which often leads to severe injuries permanently or loss of the victim's life. One of the major problems of bad road is the potholes on the road.

Accident Report and Road Surface Detection System (ARARSDS) is a system which is designed for accident reporting and potholes on roads. This System has various sensors that are used for accident sensing and pothole's on roads. Using Smartphone's GSM&GPS the exact location of the accident is provided as well as the location of the potholes is also provided. Using the MAP in the ARARSDS we can navigate the location of the accident. Similarly by using the sensor the potholes are being sensed. This data which is being sensed by the various sensors and then processed by the processing module and the location of the potholes are displayed on the map in ARARSDS.

Further this kind of data can be sent to the authorities where such kind of data are stored and managed, which may be useful for maintaining the roads and providing the accident prone areas.

Keyword-- ARARSDS, GPS, GSM, Accident, Pothole.

I. INTRODUCTION

The number of road accident has been seen increasing in the last few years. In the last few there has been drastic change in the economy as well as population which often leads to more running automobiles. One of the major cause of road accident is the bad road condition and the weather condition. As soon as the accident occurs it is important to treat the sufferer as soon as possible. For the sufferer to be treated fast as possible now it is important to provide accurate information where the accident has taken place. The accurate information provided can save life of sufferer. This system uses the Smartphone as medium for providing the required location and alerting the authorities.

The potholes on the road are the problem related to the accident. By using the data provided by the sensor this issue can be solved. The data which is taken by the sensor is provided to the processing unit by the system.

Map is used for displaying the location of the potholes. The location of the potholes is displayed using a map. The data is displayed in the form of the pin on the map. On the road the area where the potholes are present are marked with the help of the pin.

II. WHAT IS ACCIDENT REPORT AND ROAD SURFACE DETECTION SYSTEM

Accident that occurs are usually unnoticed, in order to avoid this from happening we use such system that whenever the accident takes place it sends the alerting message with the location to the respective emergency authorities. With the help of this we will be able to recognize that the accident has taken place and we will not go unnoticed. As we know one of the major problems of the accident is bad road condition, whenever there is a pothole on the road and the vehicles enter in it may cause some problem. Thus this system is using the sensor for providing the location of the pothole.

III. NEED OF ACCIDENT REPORT AND ROAD SURFACE DETECTION SYSTEM

Accident usually occurring may lead to loss of the sufferer's life. Accident often occurring can result in loss of victim's life.

Thus the need turns up for a system that should detect and report the accident so as to provide help as soon as possible.

The condition of the roads are poor in India due to the bad weather condition and low maintenance of the road. Due to this there is an emergence of the potholes on the road. This system is designed to detect the potholes and provide location of the potholes with the help of the map.

IV. SYSTEM ARCHITECTURE

All the modules which are been used are shown below and are also explained below.

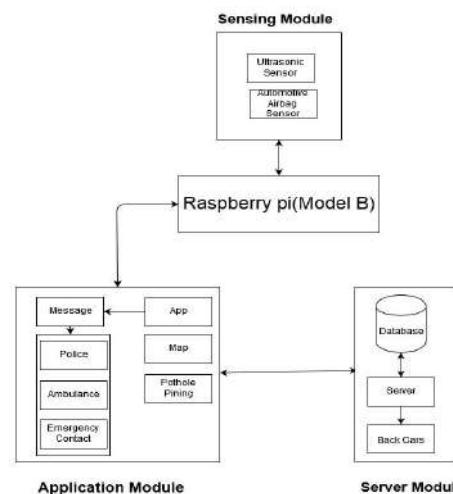


Fig . 1. System Architecture of Accident Alert and pothole Detection System

A. Sensing Module

The Sensing module is responsible for detecting the accident and potholes on the road. The Sensing module is the responsible for Detection of accident and pothole's on road. The sensing module incessant takes the data from the backdrop and provides data to the processing unit. The Sensor is responsible for sensing the trench incessant captures the data in its domain and provide that data to the processing unit. It interacts only with the processing module.

B. Processing Module

The main aim of the processing module is to process the incessant data received from the sensors. Raspberry pi is the processing module. Then this processed data is send to the app and then it is sending on the server and then the data is distributed.

C. Application Module

The application is android based, which provides simple user interface to user and displays the road condition and trench on the road on which the car is drown with the help of a map. The user interfaces in kept simple so that anyone can use it easily

D. Server

The server module is used for fetching the information of the users with the help of their apps. The data which is fetched is in the JSON format and then it is stored in MySql database format. This is provided to the user based on the location of the user or it can be pre-loaded accordingly.

V. FUTURE DEVELOPMENT

The system will be able to sense the trench using the real time data. The problem related to the seriousness of the trench can be solved. Where there is no network present the accident report cannot be used.

REFERENCES

- [1]. Vehicle Accident Alert and Locator IJECS-IJENS Vol: 11No: 2.
- [2]. An Image Processing Approach to Detect Lanes, Pot Holes and Recognize Road Signs in Indian Road International journal Modeling and Optimization vol . 2 No .6 Dec 2012.
- [3]. Pothole Detection System using Machine Learning on Android (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 7, July 2014)
- [4]. Automatic vehicle over speed, accident alert and locator system for public transport (Buses) International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 2 Issue 8 August, 2013 Page No. 2327-2331.
- [5]. A Web-Based Accident Reporting And Tracking System (Artsys) Using Sensor Technology International Journal of Advances in Engineering & Technology, Oct., 2015. ©IJAE ISSN: 22311963.
- [6]. RIPD: Route Information and Pothole Detection International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 12, December 2015
- [7]. Accident Alert And Trench (Pothole) Detection System International Journal of Innovative Science and Research Technology Volume 2, Issue 10, October– 2017 ISSN No:- 456 – 2165

College Bazaar

A Model for Smart Shopping

Vicky Prajapati

Computer Engineering Department
Thakur College of Engineering And Technology
Mumbai, India
vickyprajapati916@gmail.com

Deepak Singh

Computer Engineering Department
Thakur College of Engineering And Technology
Mumbai, India
sdeepak7190@gmail.com

Pawan Singh

Computer Engineering Department
Thakur College of Engineering And Technology
Mumbai, India
pawansinght20@gmail.com

Ms. Prachi Janrao

Computer Engineering Department
Thakur College of Engineering And Technology
Mumbai, India
prachi.janrao@gmail.com

Abstract—Development of the internet has increased the popularity of online shopping. Online shopping has become the third most popular internet activity, following email using/instant messaging and web browsing. The rapid diffusion of computer and information technologies throughout the business and consumer communities has resulted in dramatic changes.^[1] The application of the internet for purchasing behavior is a notable change in the way buyers and sellers interact. According to the data gathered by Interbank Card Center, the e-commerce volume increased.

College bazaar is an online web-application which lets you post images & a video of the study materials with contact information and short description about material of which you want to sell and the buyer who is interested to buy materials can directly contact to the seller. In this project there are a lot of categories ranging from books to aprons so whatever you are offering or want related to study materials are there. So, categories are well organized. It will help you avoid the hassles of shipping and stuff and you can easily buy or sell the materials in a meeting.

Keywords—E-commerce; Web-application;

I. INTRODUCTION (HEADING I)

E-commerce is fast gaining ground as an accepted and used business paradigm. More and more business houses are implementing web sites providing functionality for performing commercial transactions over the web. It is reasonable to say that the process of shopping on the web is becoming commonplace. The objective of this project is to develop a general purpose e-commerce store where any used stationery product in the college (such as resource books, files, printout sheets, double or single side sheets, notebooks, pens) can be

bought from the comfort of home through the Internet. An online store is a virtual store on the Internet where customers can browse the catalogue and select products of interest. The selected items may be collected in a shopping cart. At checkout time, the items in the shopping cart will be presented as an order. At that time, more information will be needed to complete the transaction. Usually, the customer will be asked to fill or select a billing address, a shipping address, a shipping option, and payment information such as credit card number. An e-mail notification is sent to the customer as soon as the order is placed.

II. EASE OF USE

A. Saving Time

This project is made to bridge the gap between the buyer and seller to meet very quickly by showing the result closest to the seller's location. This avoids the unwanted time of sending through delivery by making buyer and seller contact each other by themselves. One of the main features of the project is that any buyer can go the college-wise result filter and get the result of the search within the respective student's college. We ensure that only relevant information in such cases is shared between the respective person to avoid any misuse of the service.

B. Saving Money

As we have started building this project, so budget constraints will be challenging task. This project is designed by considering the reuse of useful resources. Every year new edition of same books came into the market with additional cost and only few changes. So, instead of using money in those types of book, anyone can buy the same book from us

added that they will save good money to buy something more useful.

III. MAIN TEXT

E-classified websites have taken the world very ahead with its advancement. Taking all the positives from this, we will find a better solution specifically in the field of education. As, the education is one of the necessary part of our life that needs to be handled perfectly. So, we are making an e-classified website specifically for the educational help. Here, all the stuffs related to education is put in a single platform. The existing platform is very complex to find the specified results. We need to add many search options with some of side filters. After doing all this, we are not sure to get the result as expected. The platform that we are building is capable of doing all these searches. We make sure, that the resource being searched is properly displayed, thereby avoiding any inconvenience. There are many technologies used in this project for development which is given below.

Technology Used

1. HTML5

HTML5 is the latest and most enhanced version of HTML. Technically, HTML is not a programming language, but rather a mark up language. With the updation of HTML to HTML5 it has many features included but one of the feature which is widely accepted is cross browser compatibility.

2. CSS3

Cascading Style Sheets (CSS) is a style sheet language used for describing the look and formatting of a document written in a markup language. When we build any product it needs to look very attractive. For this job cascading stylesheet is perfect.

3. JavaScript

JavaScript is a lightweight, interpreted programming language. It adds the dynamic nature to the website. All things happening in real time with any website uses javascript. The extent of its use depends on the requirements of the website. There are many types of framework developed in javascript just to provide better and faster access for development. It is complimentary to and integrated with Java. JavaScript is very easy to implement because it is integrated with HTML. It is open and cross-platform.

4. Bootstrap

Bootstrap is an open source toolkit for developing with HTML, CSS, and JS. Bootstrap is modular and consists of a series of Less stylesheets that implement the various components of the toolkit. These stylesheets are generally compiled into a bundle and included in web pages, but individual components can be included or removed. Bootstrap

provides a number of configuration variables that control things such as color and padding of various components

Bootstrap supports responsive web design. This means the layout of web pages adjusts dynamically, taking into account the characteristics of the device used (desktop, tablet, mobile phone). Starting with version 3.0, Bootstrap adopted a mobile-first design philosophy, emphasizing responsive design by default

Bootstrap provides a set of stylesheets that provide basic style definitions for all key HTML components. These provide a uniform, modern appearance for formatting text, tables and form elements. In addition to the regular HTML elements, Bootstrap contains other commonly used interface elements. The components are implemented as CSS classes, which must be applied to certain HTML elements in a page.

5. PHP

PHP is a popular and widely used programming language which is utilized to build dynamic web applications with MySQL database connections. For a programming language to be successful, it must be comfortable and widely accepted by a large number of web developers. PHP is outfitted with many open sources integrated development environments, moreover making it cost-effective.

PHP offers a plenty of benefits that will surely deliver your limits of developing something outstanding. Not only is it open-source but also feature-rich and has all the functionality that a proprietary or paid scripting language would offer. PHP is easy to install and set-up. It is the prominent reason of why PHP is the best language to learn. In the software industry, there are many IT companies were looking for PHP developers. There are various benefits of using PHP which attracts people towards it. In this way, let us discuss the most important reasons to utilize PHP in web development.

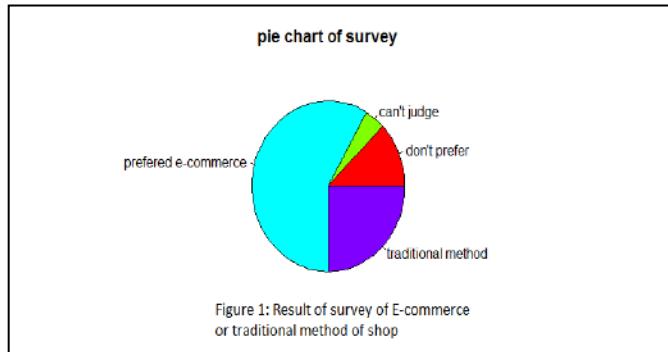
6. MySQL

MySQL is the most popular Open Source Relational SQL Database Management System. MySQL is one of the best RDBMS being used for developing various web-based software applications. It allows us to easily write the query and execute it with a simple go.

IV. STUDENT OPINION

We had discussions with our class members regarding e-commerce. We asked them whether they prefer e-commerce or traditional walk and shop method. 15% said they don't prefer e-commerce, 5% said they really can't judge, 70% prefer e-commerce and 30% prefer walk and shop method.^[4]

This experiment shows that students are very much interested about the E-commerce. They think it is very useful



for the unused materials to sell or buy. The web based environment will help them to put their unused items on our website and wastage of resources will be prevented.

Conclusion

The Internet has become a major resource in modern business, thus e-classifieds has gained significance not only from the

entrepreneur's but also from the customer's point of view. As per a survey, a lot of the stationery we buy in college goes unused after the end of a term. Our website helps curb the unwanted wastage of resources like reference books, notebooks, textbooks, files, writing sheets. We have designed the project to provide the user with easy navigation, retrieval of data and necessary feedback as much as possible. This project is aimed at helping students acquire study materials easily and at a low cost.

References

- [1] IEEE paper on E-Commerce Security Issues, Mohamad Ibrahim Ladan Computer Science Department, Haigazian University, Beirut, Lebanon.
- [2] IEEE paper on Autonomous Online Education System based on Intelligent Recommendation, QingsongTu Hubei University of Technology Wuhan, China.
- [3] Literature review of E-commerce , <http://www.scribd.com/document/178084863/Literature-Review-of-E-commerce>.
- [4] IJCSST paper on E- learning, MadhuBala, Neha Volume 10, No. 2, July December 2017.
- [5] Developing Web Applications 2nd Edition, Ralph Moseley & M. T. Savaliya.
- [6] Web technology and Design, C. Xavier.

Web Defender

¹ Mr. Shivam Singh
shivamsingh.tcet@gmail.com

¹ Mr. Dineshkumar Tanwar
dineshtanwar0307@gmail.com

¹ Mr. Aman Singh
amansingh.tcet@gmail.com

² Mrs. Vaishali Nirgude
vaishali.nirgude@thakureducation.org

Student ¹
(Department of Computer Engineering) Thakur College of Engineering and Technology

A.P. ²

Abstract— The goal of this project is to make your web application more difficult to hack. Web Application or any software for that matter will never be completely secure and free for the defect. It is only matter of time before a determined attacker will find some vulnerability or misconfiguration to exploit and compromise either your site or one of its user. You should take moment to come to terms with this truth before processing. Many people wrongly assumed that hiring a smart developer or deploying commercial security product will magically make their site hacker proof. So, we are making a virtual image of system implemented by top security features.

The outcome of the Project will be Self-configurable Virtual Machine Image which can be used by an enterprise to protect its own application by putting VM image in front of Application. Enterprise needs to develop an application which can demonstrate correct functionality. Non Functional Requirement of the application like Scalability and Security will be handled by Self configurable Virtual Machine. Here VM will protect Enterprise-wide application through the various Hierarchical mechanism. VMs will allow Access level Protection along with Predictability of Attack on Application.

Index Terms— Distributed Denial of service (DDOS), Domain Name System (DNS), Hypertext Transfer Protocol (HTTP), Kernel Virtual Machine (KVM), Open Web Application Security Project (OWASP), Reverse proxy, Virtual Machine (VM).

I. INTRODUCTION

The Project will be Self configurable Virtual Machine Image which can be used by an enterprise to protect its own application by putting VM image in front of Application. Enterprise need to develop application which can demonstrate correct functionality. Non Functional Requirement of the application like Scalability and Security will be handle by Self configurable Virtual Machine. Here VM will protect Enterprise wide application though various Hierarchical mechanism. VMs will allow

Access level Protection along with Predictability of Attack on Application. Now a day's internet is an insecure place and attacks keep occurring. One of the main reasons is the poor quality of the software used in systems and application software. So, we are came up with the innovative idea for making web apps secure from hackers and other attackers. We are making a product which will be act as intermediate between the client web app and users (legitimated users / hackers). This product will fully secured and configured by top 10 OWASP core rules.

Now a day's internet is an insecure place and attacks keep occurring. One of the main reasons is the poor quality of the software used in systems and application software. So, we are came up with the innovative idea for making web apps secure from hackers and other attackers. We are making a product which will be act as intermediate between the client web app and users (legitimated users / hackers). This product will fully secured and configured by top 10 OWASP core rules.

The end product will have the ability to secure any Web Application of client. Product will be flexible as per the requirement of virtual machine. Charges will be standard as of security services.

II. THEORY

The existing system include direct access of client and server, where in people are just relying on where they are hosting there website, or a platform where they are deploying there website which is completely dependent on the users concern.

The problems with the existing system are:

- a) The existing system involves many types of attack.
- b) Most of the attacks on enterprises networks are shifted towards cloud based application as more and more.
- c) In the current system there are less security feature.
- d) Very few existing systems provides a website security.

In today's world, where majority of people use website applications for their day to day needs, the application for advance marketing will be a success.

III. PROPOSED WORK

The application overcomes the problems in the current system in the following way:

- a) The idea is to create a fully equipped Open source based Virtual Machine will act as A Reverse proxy for any web applications with innovative modules.
 - b) Hide the actual location of Web application: Unauthorized user try to access the website directly from the web server (actual location). So we are making use of reverse proxy server .A reverse proxy taking request from internet and forwarding them to servers in internal network. Those who making request s to the proxy may not be aware of internal network.
 - c) Profiling Good access to Web application and create a Sets of permissible User inputs to web applications: Profiling (dynamic program analysis that measures the space (memory) or complexity of frequency and duration of function calls) of good access to web application and creating sets of permissible user inputs to web application .
 - d) Implementing standards OWASP Mod Security Core Rules Set. OWSAP mod security CRS projects goal is to provide an easily "pluggable" set of generic attack detection rules that provide a base level for any web application. The OWSAP mod security CRS provides protection if the following attack /threats categories:
- HTTP PROTOCOL PROTECTION
 - REAL TIME BLACK LIST LOOKUPS

Architecture of the system

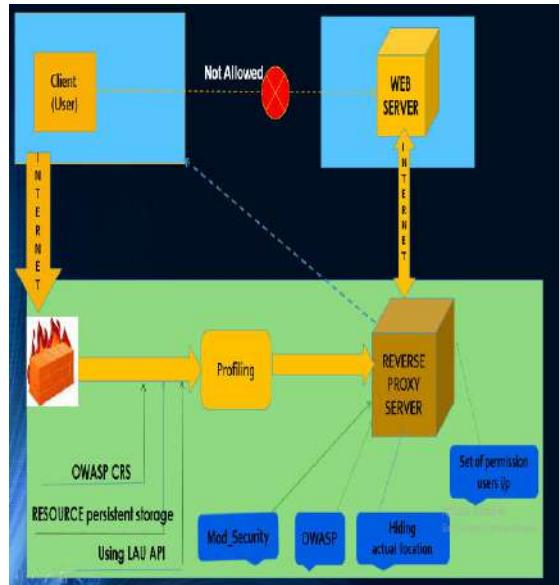


Fig.1 Architectural model

IV. METHODOLOGY AND DESIGN/ALGORITHM

- a) In this project, we as a team have decided to proceed with Spiral model. The reason to choose this specific model is due to its changing system requirements and measurable progress of system.
- b) Spiral model will follow a circular trail where once decision is made can change depending upon the requirement. The circular structure of spiral model is very favorable for our condition where client-interference will be at minimum and final assessment followed by presentation will be conducted at the end.

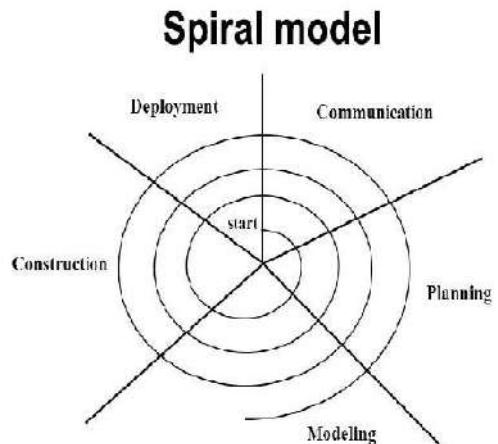


Fig.2 Spiral model

Pseudo Code Design

Web Defender implements four module:

- Detect denial service of attack
- Hide the actual location of web server
- OWASP modsecurity
- Set of permissible user inputs

In first phase of our project we have successfully implemented first module (Detect denial service of attack. Currently we are working on second module.)

Detect denial service of attack

The mod evasive apache module sometimes called as mod DoS evasive, It protects against Denial of service attack, Distributed Dos and brute force attack on the Apache web server. It can protect during attack and report abuses via mail and syslog facilities.

Our module works by creating an internal IP addresses and URL's table dynamically as well as denying any single IP address from any of the following:

- Requesting the same page more than a few times per second
- Making more than 50 concurrent requests on the same child per second
- Making any requests while temporarily blacklisted

If any of the above conditions are met, a 403 response is sent and the IP address is logged. Optionally, an email notification can be sent to the server owner or a system command can be run to block the IP address.

V. IMPLEMENTATION

Perl Script for DOS attack

```
#!/usr/bin/perl  
# test.pl: small script to test mod_dosevasive's  
# effectiveness  
use IO::Socket;  
use strict;  
for(0..100)
```

```
{  
my($response);  
  
my($SOCKET)= new IO::Socket::INET(Proto  
=> "tcp",  
PeerAddr=> "target.com:80");  
if (! defined $SOCKET)  
{  
    Die $!  
  
}  
  
}  
  
print $SOCKET "GET /?$_ HTTP/1.0  
\n\n";  
$response =<$SOCKET>;  
print $response;  
close($SOCKET);  
}
```

VI. RESULT AND DISCUSSION

Currently, We have successfully implemented first module of our project i.e. defended Dos attack using mod_evasive. Proof of our work has shown below figure. Further implementation of module will be done in next phase.

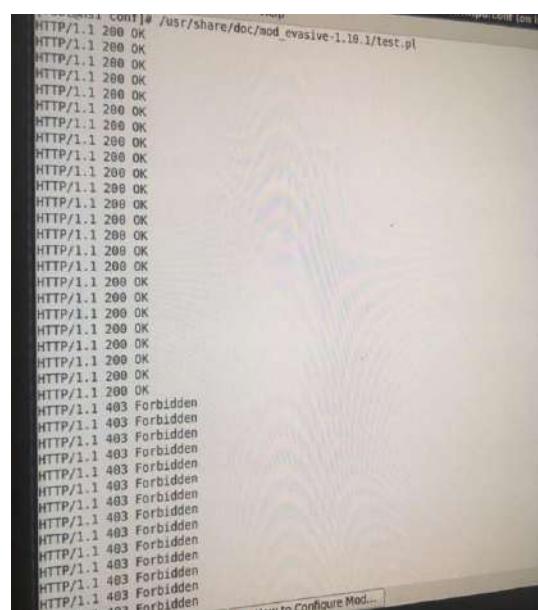


Fig. 3 Result of DoS attack

VII. CONCLUSION

Our product will help to provide VM based Security as a service model. Initially VMs will be hosted on Public cloud to make easily available to various enterprises. Any enterprise even an individual whose business is running online can take a product and acquire the best services needed for his/her web application.

VIII. FUTURE SCOPE

This project will majorly attract the people those running his /her business online and want to grow his business with full boost. In initial state it may not stand with the today's security service provider like GoDaddy and CloudFlare and many more. But it will be best product in future. In market there are many security service provider as competitor.

REFERENCES

- [1] RYAN C. BARNETT, "WEB APPLICATION DEFENDER'S COOKBOOK: BATTLING HACKERS AND PROTECTING USERS," PUBLISHED BY JOHN WILEY & SONS INC., 2013.
- [2] mod_evasive,
https://www.digitalocean.com/community/tutorials/how-to-protect-against-dos-and-ddos-with-mod_evasive-for-apache-on-centos-7
- [3] aws php documentation,
<https://aws.amazon.com/blogs/developer/provision-an-amazon-ec2-instance-with-php/>
- [4] Reverse Proxy with mod_proxy on CentOS 7,
https://www.digitalocean.com/community/tutorials/how-to-use-apache-as-a-reverse-proxy-with-mod_proxy-on-centos-7
- [5] BugTraq, <http://www.securityfocus.com/archive/1>
- [6] Auger, R., Barnett, R.: Web Application Security Consortium: Threat Classification Version 1.0. Web Application Security Consortium (2004), <http://www.webappsec.org>
- [7] The Open Web Application Security Project (OWASP) considers, in its 2016 top ten list. Available: <https://www.owasp.org/index.php/Top10-2016> – TOP 10.
- [8] Ivan Ristic : ModSecurity Handbook: The Complete Guide to the Popular Open Source Web Application Firewall, 2010 Feisty Duck Ltd Edition ISBN: 1907117024
- [9] Naxsi project (Nginx Anti XSS SQLInjection) May 2014.
Available:
https://www.owasp.org/index.php/OWASP_NAXSI_Project
- [10] The OWASP modSecurity Project Available:
https://www.owasp.org/index.php/Category:OWASP_ModSecurity_Core_Rule_Set_Project

FAKE PRODUCT REVIEW DETECTION AND REMOVAL

Aditya Darak¹, Akhil Chaudhary¹, Ayushi Mankar¹, Prajwal Mogaveera¹, Shiwani Gupta²

B.E (C.S.E.) Mumbai University¹

Assistant Professor, TCET, Mumbai²

Abstract— Online reviews are often the primary factor in a customer's decision to purchase a product or service, and are a valuable source of information that can be used to determine public opinion on these products or services. Reliance on online reviews gives rise to the potential concern that wrongdoers may create false reviews to artificially promote or devalue products and services. This practice is known as Opinion (Review) Spam, where spammers manipulate and poison reviews (i.e., making fake, untruthful, or deceptive reviews) for profit or gain. We propose to build a fraud risk management system based on data processing and intelligent risk-mitigation models. It captures fraudulent transactions based on user behaviors and network, analyses them in real-time using Data Mining, and accurately predicts the suspicious users and transactions. This system proposes behavioral approach using J48 CLASSIFIER to detect review spammers who try to manipulate the ratings on some target products.

Keywords— *Fake Review detection, Review Removal, J48, classifier*

INTRODUCTION

As the Internet continues to grow in both size and importance, the quantity and impact of online reviews continually increases. Reviews can influence people across a broad spectrum of industries, but are particularly important in the realm of e-commerce, where comments and reviews regarding products and services are often the most convenient, if not the only, way for a buyer to make a decision on whether or not to buy them. While online reviews can be helpful, blind trust of these reviews is dangerous for

both the seller and buyer. Many look at online reviews before placing any online order; however, the reviews may be poisoned or faked for profit or gain, thus any decision based on online reviews must be made cautiously. Furthermore, business owners might give incentives to whoever writes good reviews about their merchandise, or might pay someone to write bad reviews about their competitor's products or services. These fake reviews are considered review spam and can have a great impact in the online marketplace due to the importance of reviews.

LITERATURE SURVEY

Opinion Mining has attracted to a great deal of research earlier. However, not a great amount of work has been done in this field. Review Spam is very hard to detect unless read manually. Here are some of the work proposed and implemented. Paper [1] proposes different new features and gives the model and algorithm to construct each of these features. Although, it is not a good metric and the reduction is not substantial. Paper [2], have used linguistic features like unigram presence, unigram frequency, bigram presence, bigram frequency and review length to build a model and find fake reviews. Although, the main problem is data scarcity. Paper [3] proposes behavioral approach to detect review spammers who try to manipulate the ratings on some target products. Paper [4] proposes to employ categories of lexical semantic and linguistic features in the detection of online spam reviews. In Paper [5] we found that spotting the individual fake reviews was quite a difficult task but spotting the groups was comparatively easier one. Paper [6] first performed a comparison using real-life filtered (fake) and unfiltered (non-fake) reviews in Yelp. The results showed that the real-life data is much harder to classify, with an accuracy of only 67.8%.

Sr.No	Title	Publications	Work Done	Research Gap
1	Detecting Fake Reviews Utilizing Semantic and Emotion Model	IEEE'16	Computation of Semantic Similarity will result in 3 features including Review Density, semantics and emotion	It is not a good metric. The reduction is not substantial.
2	Fraud Detection in Online Reviews using Machine Learning Techniques.	IJCER'15	NAÏVE BAYES CLASSIFIER requires both linguistic & Behavioral features with 81% ACCURACY	It makes a very strong assumption on the shape of your data distribution. Another problem happens due to data scarcity
3	Fake Review and Brand Spam Detection using J48 Classifier.	IJCSIT'15	J48 CLASSIFIER resulted in behavioral approach using AMAZON dataset with 93% ACCURACY	Space complexity is very large.
4	Online Review Spam Detection by New Linguistic Features.	UNIVERSITY OF MARYLAND BALTIMORE COUNTY'15	Used Support Vector Machine (SVM) with 92% ACCURACY	The major downside of SVMs is that they can be painfully inefficient to train.
5	Fake Reviewer Groups' Detection System.	IOSR-JCE'14	Frequency Item set Mining(FIM) method finds clustered review which is comparatively easier to handle	Assumes database transaction is memory resident. Requires up to 'm' database scan.
6	Fake Review Detection:Classification and Analysis of Real and Pseudo Reviews.	UIC-CS'13	Used Amazon Mechanical Turk (AMT) With 90% ACCURACY	The appropriate 'wage' for each task is difficult to assess. Tasks that require rigorous programming cannot be done online.

DESIGN

Our research show that it is possible to detect spam comments with the proper selection of features which capture different characteristics of legitimate comments in order to differentiate them from spam comments. Initially input is selected for detecting the spammed reviews. After the selection the product is reviewed for spam detection. Spam detection technique is used to check for spams and spams are reviewed. After detecting the spam, the spam content is analyzed so as to ensure the nature of the spam. System controller is used to detect the fake spams which are to be deleted by the admin. After the process the users can view the final result.

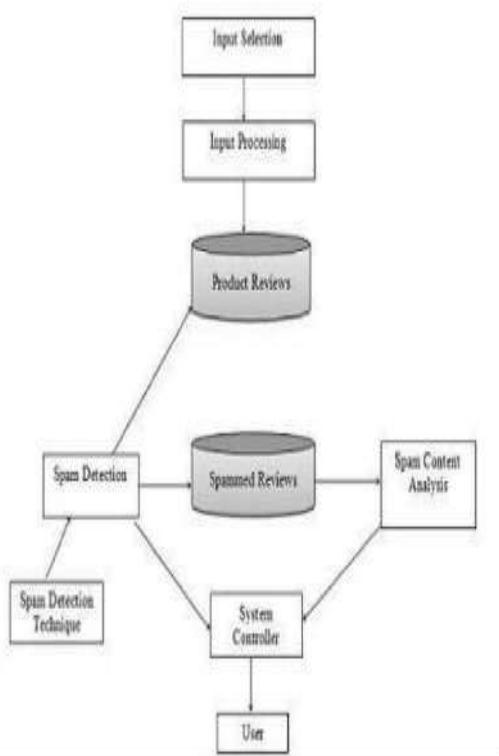


Fig.1 System Architecture

IMPLEMENTATION METHODOLOGY

The proposed system has been implemented earlier using various methodologies like Naïve Bayes, Sentimental analysis, J48, Natural language processing techniques and few others. All of them have resulted in successful implementation of the desired system. Efficient result has been taken from bank dataset using weka tool in the experiment. Naive Bayes classifier also showing good results. The experiments results shown in the study are about classification accuracy and cost

analysis. J48 gives more classification accuracy for class mortgage in bank dataset having two values Yes and No. The requirements are as follows:

Hardware Requirements

- Smart Phone
- Server Machine

Software Requirements

- Android 4.4+ OS
- JAVA
- MySQL
- Android Studio

EXPECTED RESULTS

Fake product review detection system will be useful to business organization as well as to customers. When used on e-commerce dataset the expected accuracy is 96%. Business organization can monitor their product selling by analyzing and understand what the customers are saying about products. Customers can make decision whether he/she should buy or not buy the products. This can be helpful to people to purchase valuable product and spend their money on quality products. Our application will do analysis and then post the genuine review on genuine products. And user can be sure about the products availability on that application and reviews too.

CONCLUSION

Determining and classifying a review into a fake or truthful one is an important and challenging problem. As part of future work, we can incorporate review spammer detection into the review detection and vice versa. Exploring ways to learn behavior patterns related to spamming so as to improve the accuracy of the current regression model. So as to evaluate our proposed methods, that conducts user evaluation on an Amazon dataset containing reviews of different manufactured products. We derive an aggregated behavior scoring methods for rank reviewers according to the degree that they demonstrate spamming behaviors. We found that here proposed methods generally outperform the baseline method based votes. We further learn a regression model from the user labeled ground truth spammers.

REFERENCES

- [1] Yuejun Li, Xiao Feng, Shuwu Zhang, "Detecting Fake Reviews Utilizing Semantic and Emotion Model" IEEE-2016.

- [2] Kolli Shivagangadhar, Sagar H, Sohan Sathyan, Vanipriya C.H ,“Fraud Detection in Online Reviews using Machine Learning Techniques.” IJCER-2015.
- [3] Sushant Kokate, Bharat Tidke, “Fake Review and Brand Spam Detection using J48 Classifier.” IJCSIT, 2015.
- [4] Amir Karam, Bin Zhou, “Online Review Spam Detection by New Linguistic Features.” UNIVERSITY OF MARYLAND BALTIMORE COUNTY’15
- [5] Kolhe N.M., Joshi M.M.Jadhav A.B., Abhang P.D., “Fake Reviewer Groups’ Detection System.” IOSR-JCE 2014.
- [6] Arjun Mukherjee, Vivek Venkataraman, Bing Liu, Natalie Glance, “Fake Review Detection:Classification and Analysis of Real and Pseudo Reviews.” UIC-CS’13

Online sports mart and Information system

Mrs.Ruta Pathak, Aakash Mishra, Gargee Kawale, Ramit Gupta, Aniket Kargaonkar

Thakur college of Engineering and Technology, Kandivali (E), Mumbai-400101

ABSTRACT-ONLINE SPORTS MART AND INFORMATION SYSTEM is a global system service about the news and the updates of mostly all the sports played in the world. It features articles, live coverage of all the sports (including videos and scores), highlights, player ranking and team ranking. The users will get complete information of the sports world. The website will have a database which will contain all the recognized users.

The system will contain a platform for the authorized users to buy stuffs related to different sports. (Online Shopping Concept)

The system will also include a concept where all the authorized users will be made available with only those sports feeds that they like and are subscribed for using Recommender Algorithm.

I.INTRODUCTION

THE PLATFORM PROVIDES SPORTS ENTERTAINMENT IN ITS PUREST AND MOST EXCITING FORM

This System creates an interactive platform between the sports and the people.

ONLINE SPORTS MART AND INFORMATION SYSTEM is a global system service about the news and the updates of mostly all the sports played in the world. It features articles, live coverage of all the sports (including videos and scores), highlights, player ranking and team ranking. The users will get complete information of the sports world. The website will have a database which will contain all the recognized users.

The system will contain a platform for the authorized users to buy stuffs related to different sports. (Online Shopping Concept)

The system will also include a concept where all the authorized users will be made available with only those sports feeds that they like and are subscribed for using Recommender Algorithm.

Event Calendar:

An event calendar is a system where users will get complete information about all the matches being played currently and all the anticipated matches that are yet to come.

An event calendar should be a place where everyone can go and instantly see all the activities that are taking place, i.e. matches are when, in/out of all activities. It should be live 100% of the time and also completely accurate.

It must also update itself regularly with all the current available information.

Recommender Algorithms:

What is recommender algorithm?

Algorithm that when a user logs in the system, he/she is provided with those information and notifications according to his last search and subscriptions.

Recommendation algorithms which suggest relevant information or suggestions to particular user based on its past search. It will also improve as per user's increasing experience with the system.

Recommender system has the ability to predict whether a particular user would prefer an item or not based on the user's profile and subscriptions.

Recommendation algorithms are part of machine learning and there are various methods to implement it. We first need to find parameters and its weight and according to that you require to generate one hypothesis function. This function is used to count or measure actual cost of any item. It basically counts the similarity with user's interest. Using this cost function, we can implement any algorithm like to suggest item. Recommender Algorithms for interest notifications includes study of machine learning, big data, data mining, meta data. Systems that recommend an item to a user based upon a description of the item and a profile of the user's interests. Content-based recommendation systems may be used in a variety of domains ranging from recommending web pages, news articles, restaurants, television programs, and items for sale. Although the details of various systems differ, content-based recommendation systems share in common a means for describing the items that may be recommended, a means for creating a profile of the user that describes the types of items the user likes, and a means of comparing items to the user profile to determine what to recommend. The profile is often created and updated automatically in response to feedback on the desirability of items that have been presented to the user.

Technology used: Recommender systems use algorithms to provide users with product or service recommendations. Recently, these systems have been using machine learning algorithms from the field of artificial intelligence. However, choosing a suitable machine learning algorithm for a recommender system is difficult because of the number of algorithms described in the literature. Researchers and practitioners developing recommender systems are left with little information about the current approaches in algorithm usage. Moreover, the development of a recommender system using a machine learning algorithm often has problems and open questions that must be evaluated, so software engineers know where to focus research efforts. This paper presents a systematic review of the literature that analyses the use of machine learning algorithms in recommender systems and identifies research opportunities for software engineering research. The study concludes that Bayesian and decision tree

algorithms are widely used in recommender systems because of their relative simplicity, and that requirement and design phases of recommender system development appear to offer opportunities for further research.

Recommender systems (RS) are used to help users find new items or services, such as books, music, transportation or even people, based on information about the user, or the recommended item. These systems also play an important role in decision-making, helping users to maximize profits or minimize risks. Today, RSs are used in many information-based companies such as Google, Twitter, LinkedIn, and Netflix. The field of RS has its origins in the mid-1990s with the introduction of Tapestry, the first RS. As the RS field evolved, researchers studied the use of algorithms from machine learning (ML), an area of artificial intelligence (AI). Machine learning has been studied since the late 1950s, with the emergence of the field of AI.

The system presents a summary list of items to a user, and the user selects among the items to receive more details on an item or to interact with the item in some way. For example, online news sites present web pages with headlines (and occasionally story summaries) and allow the user to select a headline to read a story. E-commerce sites often present a page with a list of individual products and then allow the user to see more details about a selected product and purchase the product. Although the web server transmits HTML and the user sees a web page, the web server typically has a database of items and dynamically constructs web pages with a list of items. Because there are often many more items available in a database than would easily fit on a web page, it is necessary to select a subset of items to display to the user or to determine an order in which to display the items. Content-based recommendation systems analyse item descriptions to identify items that are of particular interest to the user. Because the details of recommendation systems differ based on the representation of items.

Recommender systems are divided into three main categories, depending on the information used to drive the recommendations: collaborative, content-based, and hybrid filtering.

RSs using a collaborative approach consider the user data when processing information for recommendation. For instance, by accessing user profiles in an online music store, the RS has access to all the user data, such as the age, country, city, and songs purchased. With this information, the system can identify users that share the same music preference, and then suggest songs bought by similar users.

RSs with a content-based filtering approach base their recommendations on the item data they can access. As an example, consider a user who is looking for a new computer using an online store. When the user browses a particular computer (item), the RS gathers information about that computer and searches in a database for computers that have similar attributes, such as price, CPU speed, and memory capacity. The result of this search is then returned to the user as recommendations.

The third classification describes RSs that combine the two previous classifications into a hybrid filtering approach, recommending items based on the user and the item data. For example, on a social network, an RS may recommend profiles

that are similar to the user (collaborative filtering), by comparing their interests. In a second step, the system may consider the recommended profiles as items and thus access their data to search for new similar profiles (content-based filtering). In the end, both sets of profiles are returned as recommendations.

When using a collaborative or a hybrid filtering approach, RSs must gather information about the user in order to develop recommendations. This activity can be done explicitly or implicitly. Explicit user data gathering happens when users are aware they are providing their information. For instance, when registering for a new online service, users usually fill in a form that asks their name, age, and email. Other forms of explicit user data gathering are when users express their preferences by rating items using a numerical value or a preference such as a Facebook “like.” Implicit user data gathering accesses information about the user indirectly. For example, when visiting an online store, the server at the online store exchanges messages with the user’s computer, and based on that, the store’s RS may know the browser the user is using, as well as the user’s country. More advanced applications monitor user click and keystroke logs.

Besides the common recommendation process, in which users are presented with items that might be of interest, recommendations can be provided in other ways. Trust-based recommendations take into consideration the trust relationship that users have between them. A trust relationship is a link in a social network to a friend or a following connection. Recommendations based on trusted friends are worth more than those that do not have trust links. Context-aware recommendations are based on the context of the user. A context is a set of information about the current state of the user, such as the time at the user location (morning, afternoon, evening), or their activity (idle, running, sleeping). The amount of context information to be processed is high, making context-aware recommendations a challenging research field. Risk-aware recommendations are a subset of context-aware recommendations and take into consideration a context in which critical information is available, such as user vital signs. It is risk-aware because a wrong decision may threaten a user’s life or cause damage.

Phases of recommendation process:

1. Information collection phase

This collects relevant information of users to generate a user profile or model for the prediction tasks including user’s attribute, behaviours or content of the resources the user accesses. In E-learning platform, a user profile is a collection of personal information associated with a specific user. The success of any recommendation system depends largely on its ability to represent user’s current interests.

II.ITEM REPRESENTATION

Items that can be recommended to the user are often stored in a database table. Thetable shows a simple database with records (i.e., “rows”) that describe three sports. The column names (e.g., Type or Features) are properties of Sports. These properties are also called “attributes,” “characteristics,” “fields,” or “variables” in different publications. Each record

contains a value for each attribute. A unique identifier, ID in the table, allows items with the same name to be distinguished and serves as a key to retrieve the other attributes of the record.

ID	SPORT	TYPE	FEATURES	POPULARITY
1001	Cricket	Outdoor	Feature1	High
1002	Football	Outdoor	Feature2	Medium
1003	TableTennis	Indoor	Feature3	Low

Feature1: It includes users will be provided with news feeds, online purchase item, gallery images and live scores.

Feature2: It includes users will be provided with news feeds, online purchase item and gallery images.

The database depicted in the table could be used to drive a web site that lists and recommends different sports. This is an example of structured data in which there is a small number of attributes, each item is described by the same set of attributes, and there is a known set of values that the attributes may have. In this case, many machine learning algorithms may be used to learn a user profile, or a menu interface can easily be created to allow a user to create a profile. The next section of this chapter discusses several approaches to creating a user profile from structured data.

User Profiles:

A profile of the user's interests is used by most recommendation systems. This profile may consist of a number of different types of information.

Here, we concentrate on two types of information:

1. A model of the user's preferences, i.e., a description of the types of items that interest the user. There are many possible alternative representations of this description, but one common representation is a function that for any item predicts the likelihood that the user is interested in that item. For efficiency purposes, this function may be used to retrieve the n items most likely to be of interest to the user.

2. A history of the user's interactions with the recommendation system. This may include storing the items that a user has viewed together with other information about the user's interaction, (e.g., whether the user has purchased the item or a rating that the user has given the item). Other types of history include saving queries typed by the user.

There are several uses of the history of user interactions. First, the system can simply display recently visited items to facilitate the user returning to these items. Second, the system can filter out from a recommendation system an item that the user has already purchased or read. Another important use of the history in content-based recommendation systems is to serve as training data for a machine learning algorithm that creates a user model. The next section will discuss several different approaches to learning a user model. Here, we briefly describe approaches of manually providing the information used by recommendation systems: user customization and rule-based recommendation systems.

In user customization, a recommendation system provides an interface that allows users to construct a representation of their own interests. Often check boxes are used to allow a user to

select from the known values of attributes, e.g., the cuisine of restaurants, the names of favourite sports teams, the favourite sections of a news site, or the genre of favourite movies. In other cases, a form allows a user to type words that occur in the free text descriptions of items, e.g., the name of a musician or author that interests the user. Once the user has entered this information, a simple database matching process is used to find items that meet the specified criteria and display them to the user.

III. LITERATURE SURVEY

1. Author Name : Junpeng Gong

Year: May 2017

Title of the paper: An automatic generation method of sports news based on knowledge rules

Summary/Conclusion: Nowadays, with the massive demand for sports news, automatic generation systems based on the template technology has been deployed, which could generate massive sports news quickly and effectively. We propose an automatic generation method based on knowledge rules to select the template dynamically from a template set. The text generated by the system is flexible and the format is varied, which improves the quality of the generated news.

Gap Identification: We can further reduce the complexity of method without executing entire process whenever new data arrived as feature work.

2. Author Name : Ramesh Thakkar

Year: July 2017

Title of the paper: Research on the effect of the recommendation system on customer online shopping experience

Summary/Conclusion: In context of keen competition among the online shopping platforms, sellers and developers are finding critical factors of recommendation system that could impact customer experience. In consideration of this, we propose a research model that examines the effects of two characteristics of recommendation, i.e. recommended way and recommendation performance, on customer online shopping experience.

Gap Identification: The two algorithms used in the mentioned paper are combined together to perform the effective work. However, it needs a lot of data to effectively make recommendations and lots of storage is wasted.

3. Author Name : John Pagonis, Adrian F. Clark

Year: November 2005

Title of the paper: A genetic algorithm classifier for content-based recommender systems that does not require continuous user feedback

Summary/Conclusion: We present Engene, a genetic algorithm based classifier which is designed for use in content-based recommender systems. Once bootstrapped Engene does not need any human feedback. Although it is primarily used as an on-line classifier, in this paper we present its use as a one-class document batch classifier and compare its performance against that of a one-class k-NN classifier.

Gap Identification : In this paper, an algorithm called Engene is used as Content Based Recommender Algorithm. It does not

requires any human interaction but the algorithm is not feasible if the user keeps on changing the datas.

4 .Author Name : Ezgi Akar
Year: September 2004

Title of the paper: A review of literature on consumers' online purchase intentions

Summary/Conclusion: This paper is to depict the factors that have an impact on consumers' online purchase intentions through an in-depth analysis of the relevant literature.

Gap Identification: The review depicts different consumers' actions and problems but do not provide any solutions to them.

IV.METHODOLOGY



V.OBJECTIVES OF PROJECT

Help more people have a sporting habit for life.

To develop an understanding of the importance of sport in the pursuit of a healthy and active lifestyle.

To provide the opportunity to be inventive and creative in sporting activities.

To promote an awareness of and an ability to appreciate the aesthetic qualities of sporting performance and movement.

To use sport as a mean to develop social interaction and so learn about others as well as themselves.

To make people develop their interest in all the sports being played in the world.

VI.SCOPE OF PROJECT

The system will contain a platform for the authorized users to buy stuffs related to different sports. (Online Shopping Concept)

The system will also contain a platform for the authorized users to sell their own products that can be picked by other users. (Like OLX)

The users will also be made available with an event calendar of all the respective sports upcoming fixtures.

VII.APPLICATION OF PROJECT

The client and users for the respective system are those who are keen for sports and those who want to excel in sports world.

This System creates an interactive platform between the sports

and the people. Hence, it can be widely used by all the enthusiastic sports lovers.

The system contains Online Shopping Mart which can be used by all those people who are unable to manage time because of any reason to visit local stores for shopping.

Those people who are very busy and are unable to watch live matches can connect to our Live Event Calendar to get updated with upcoming matches and scoresheets.

Those people who are interested in any particular sports can connect to us and we will provide them with the respective subscribed sports feed only using Recommender Algorithms.

VIII.PROJECT IMPACT ANALYSIS

Social impact: The project developed will provide a user-friendly environment as the distribution of the data will be ordered.

Ethical or legal impact: The cluster randomized trial (CRT) is used increasingly in knowledge translation research, quality improvement research, community based intervention studies, public health research, and research in developing countries.

Environmental impact: No direct environmental impact.

Financial impact: The project developed will minimize the efforts to cluster the data manually using and also using predefined no of clusters. As a result, it will be benefited financially in various applications such as marketing.

IX.CONCLUSIONS

After the completion of the project we have made a conclusion that this type of projects are beneficial for the society for spreading sports enthusiasm among people.

- 1.Help more people have a sporting habit for life
- 2.Create more opportunities for young people to play sport
- 3.Nurture and develop talent
- 4.The System enhances the sports Broadcast in all parts of the world.
- 5.It is a boon to the sports industries as technology finds new ways to connect to the fans all over the world.
- 6.It supports the development of a global sport marketplace and it shapes how we consume the sport for generations to come.

X.REFERENCES

- 1.<http://www.bbc.com>
- 2.<http://www.skysports.com/>
- 3.<http://www.final-yearproject.com/2012/01/design-of-intranet-mail-client-system.html#.VwldRqR97IU>
- 4.<http://1000projects.org/design-of-intranet-mail-system-project.html>

Analysis and Comparative study of Mobile Malware Detection

Harshal Patil

Department of Information Technology
Theem COE,Boisar
patilharshal22@gmail.com

Azar Nabi

Department of Information Technology
Theem COE,Boisar
azhar.d.invincible@gmail.com

Abstract— In last few years, the number and variants of malicious applications for hand-held devices have grown vastly, particularly on platforms like android, which brings consistent challenges for detecting malicious apps. Using the analysis of network congestion and traffic, the traces of malicious apps have been endeavoured by the researchers. Our idea is to include machine learning with network traffic analysis for identifying malicious network performance. In many cases, the traffic network generated is gentle, and a very small portion of traffic is malicious, which when the traffic model skews towards modelling the benign traffic, leads to an imbalanced data problem.

Practical and dynamic techniques for malware detection are proposed in this literature that gives a fundamental survey on mobile malware detection dynamically. The introduction of security threats and different malware classifications with their definition and evolution is given. The evaluation measures and criteria of mobile malware detection is summarized of mobile malware.

We have analysed, commented and compared those methods of mobile malware detection that were proposed in last few years on the basis of evaluation criteria and measures. Finally, we extracted basic open issues in the research field of mobile malware detection and support future research mechanisms in this field.

1. INTRODUCTION

A mobile device becomes an open concurrent software platform that can run various mobile apps developed by not only mobile device manufacturers, but also many third parties. However, the third app development parties cannot ensure the security and integrity of their shipped apps. In the literature, existing malware detection methods encompass two different approaches

while collecting features: static and dynamic. The static method aims to find malicious characteristics or suspicious code segments without executing applications, while the dynamic approach focuses on collecting an app's behaviors information and proofs during its runtime. The static methods can be applied to detect known malware with high accuracy and speed. On the contrary, the dynamic methods can identify zero-day attacks accompanied with a relatively high false positive rate and heavy resource consumption. However, the static methods can do nothing facing with camouflage and

evasion techniques or zero days vulnerabilities. In this paper, we focus on dynamic detection of malware.

Specifically, the contributions of this paper are summarized below:

- Introduce the flourishing threats and behaviors of newest mobile malware in recent years and summarize most-frequently used features in mobile malware detection;
 - Summarize the criteria and measures that are used for evaluating the quality and performance of mobile malware detection and apply them as a research model
 - to study the pros and cons of the existing methods for dynamic mobile malware detection;
 - Review the recent advance in the literature of dynamic mobile malware detection.
- We compare and analyze existing methods according to the proposed criteria and measures for evaluating their performance and discussing their advantages and disadvantages;
- ¶We further come up with future mobile malware development and the requirements of mobile malware detection to figure out open issues and motivate promising future research directions.

2 .Overview of Mobile Malware

Mobile malware can be theoretically divided into several classes according to their malicious goals and behaviors. As a supplementary condition, distribution technique is another reference standard. Normally, there are two main distribution strategies: self-propagating and social engineering. The first approach uses different strategies to automatically install malware into mobile devices, like worms, while the second one takes advantage of user curiosity and unawareness of security to allure them to manually install apps (e.g., adwares). Herein, we summarize several basic types of mobile malware based on their malicious behaviors. Except for these mainstream categories, there are many non-mainstream malware, for example, piggybacked apps are originated from benign apps, and then become botnet or malicious apps by being injected into malicious payloads.

3.Detection Performance Evaluation Measures and Criteria

In order to protect mobile devices and resist threats , we summarize the requirements that a mobile malware detection method should satisfy and the measures for evaluating detection performance. Criteria are the requirements that a malware detection method should satisfy. Measures refer to the evaluation metrics that is used to evaluate the performance of a detection method. Both of them are needed for commenting and evaluating a mobile malware method. We can improve and optimize the method by using the criteria and measures to achieve high performance and reach design expectation.

4.Dynamic Mobile Malware Detection Methods

In this section, we review dynamic mobile malware detection methods that have been proposed since 2013. We mainly retrospect related papers from several databases:ACM, Springer, IEEE Explorer libraries, Elsevier Science and ScienceDirect, by searching key word: dynamic malware detection, malware classifier, mobile malware,smartphone malware, Android malware detection and so on. It is a challenge to detect

malware dynamically in mobile devices, especially, when malware designers use encryption algorithms and evading strategies . In general, the evolution of detecting mobile malware is stimulated by the techniques used by mobile malware. From static code segments analysis, to semantic analysis; from system calls to calling graph, from codes to behaviors, the detection techniques of mobile malware progress.

4.1 Classification Algorithms

Classification algorithms aim at classifying unknown samples with proper labels, such as malicious or benign. They serve as the most essential part of malware detection together with the features. The most popular technique used in classification is machine learning, accompanied with data mining methods. Data mining mainly uses statistical methods and programming methods to find patterns of features, which can be applied into machine learning to build classification models. Most classification algorithms fall into the scope of machine learning. we briefly introduce several popular machine-learning-based classification algorithms.

4.2 Dynamic Anomaly-based Detection

Anomaly-based detection aims at building a model that contains apps' normal behaviors,which are used to classify an app's maliciousness or benignancy. This technique has potential to detect previously unseen malware accompanied by a high omission rate, but it is a classical method.

To ensure correctly recognize normal behaviors and identify samples, the anomalybased detection approach contains two phases: the training phase where a profile model of normal behaviors is built according to the normal behaviors of a host and inspection apps; the testing phase

where the behavior information monitored during the execution of query samples is compared with the profile model. In the process, machine learning algorithms can be applied to detect anomalies, such as artificial intelligence algorithms, data mining methods and so on.

4.3 Dynamic Specification-based Detection

Specification-based detection is a special type of anomaly-based detection method. It originates from the inspiration of law. This method identifies authorized behavior rules by dynamically comparing observed behaviors with "pre-determined authorized behaviors (called specification) of generally accepted definitions of benign activities." Specification-based techniques use a rule set of valid and legal behaviors to decide the maliciousness or benignancy of apps. Programs or apps violating the specification are classified as malicious; otherwise, they are benign. Specification-based detection normally has two phases. A training phase is the attainment of a series of rule sets, which represent all the valid behaviors that any program can exhibit in the devices. A testing phase is to determine whether the behaviors monitored during app execution conform to the valid specification. The main limitation is it is impossible to comprehensively and correctly evaluate the entire set of valid program behaviors in order to find out all rules since their range is too wide.

4.4 Dynamic Signature-based Detection

Signature-based detection relies on signatures and monitored packets. This method uses pre-configured and pre-determined attack patterns that are given by experts to build a signature database or a pattern set to imply maliciousness of a program. That

is to say only malicious behavior patterns are stored. Newly defined malicious signatures can be further added into a detection knowledge database that is physically called repository. Generally, the signatures should identify any malware that exhibits malicious behaviors. Dynamic signature-based detection only uses information gathered during app execution to decide its maliciousness. This type of methods looks for behavior patterns to reveal true maliciousness.

The disadvantages of this method are listed below. First, although it is efficient to expose existing simple malware types, some oligomorphic, polymorphic and metamorphic malware can circumvent this detection. Second, timely updating repository with massive new signatures is essential, which is, however, difficult for resourceconstrained mobile devices to handle.

5 Open Issues and Future Research Trends

5.1 Open Issues

According to the above review and discussions, we figure out a number of open issues in this research field.

First, real-time detection is not well supported, almost all the detection methods are not real-time in terms of app runtime behavior monitoring, not mention real-time detection. It is realized real-time information-flow monitoring in smartphones, but it cannot realize real-time detection. GuardOL designed and implemented a hardware-based architecture for malware detection. It was said as a kind of real-time detection method, but actually not. It monitors apps' system calls and constructs a frequency-centralized model to identify malware on specialized FPGA. But it fails to reveal all system calls while extracting features. In practice, we need real-time malware detection during app execution, especially for finding out some intrusions or attacks happening at app runtime. Second, information privacy should be protected if private data are outsourced and malware detection is conducted at a distrusted third party. According to the aforementioned

literature, many cloud-based methods were proposed for mobile malware detection due to constrained resources of mobile devices. Many detecting schemes choose to identify malware in the cloud, such as Secloud and SMMDS. The collected data about mobile users need to be uploaded to the cloud to process. This could intrude user privacy since the cloud cannot be fully trusted. However, none of existing work provided privacy protection during the process of mobile malware detection. This is an important private data leakage issue that urges our efforts. In ,although the authors realized this problem, no proper solutions were proposed. The same expected privacy-preserving circumstance is met in that aims to calculate trust values at a third party to control unwanted traffic and this privacy preservation issue was solved, by applying homomorphic encryption algorithm to protect data and support encrypted data processing for unwanted traffic detection and control. In addition, protecting privacy may influence the performance of detection. Thus, investigating an efficient method to realize privacy-preserving mobile malware detection is practically significant.

5.2 Future Research Trends

Open issues motivate and direct future research. Considering the situation of current development and the constraint resources in the mobile devices, we suggest several promising future research directions in the field of dynamic mobile malware detection. Except for the future research trends listed below, high detection accuracy and efficient detection algorithms are always highly expected.

6 Conclusion

Mobile malware detection is significant to ensure the quality of mobile concurrent software. This paper gave a thorough review on dynamic mobile malware detection. We discussed the main techniques in dynamic malware detection proposed since 2013. We compared existing work with regard to their technique categories, applied classification algorithms, the features used for detection, target mobile operating systems,

the places of detection analysis, real-time detection support, privacy preservation support, the threats that can be revealed and overcome, the measures used for detection performance evaluation and performance test results. In addition, we also commented their pros and cons. Based on the review, we pointed out open research issues in order to direct future research trends.

References.

- 1.Guo, D., Sui, A., Shi, Y., Hu, J., Lin, G., Guo, T.: Behavior Classification Based Self-learning Mobile Malware Detection. *Journal of Computer (JCP)*. 9, 851-858 (2014)
- 2.Cui, B., Jin, H., Carullo, G., Liu, Z.: Service-Oriented Mobile Malware Detection System Based on Mining Strategies. *Pervasive and Mobile Computer*. 24, 101-116 (2015)
- 3.Almin, S.B., Chatterjee, M.: A Novel Approach to Detect Android Malware. *Procedia Computer Science*. 45, 407-417 (2015)
- 4.Sheen, S., Anitha, R., Natarajan, V.: Android Based Malware Detection Using a Multifea-ture Collaborative Decision Fusion Approach. *Neurocomputing*. 151, 905-912 (2015)
- 5.Khan, J., Abbas, H., Al-Muhtadi, J.: Survey on Mobile User's Data Privacy Threats and Defense Mechanisms. *Procedia Computer Science*. 56, 376-383 (2015)
6. Chen, P., Lin, S., Sun, C.: Simple and Effective Method for Detecting Abnormal Internet Behaviors of Mobile Devices. *Information Sciences*. 321, 193-204 (2015)
- 7.Enck, W., Gilbert, P., Han, S., Tendulkar, V., Chen, B., Cox, L.P., Jung, J., Mcdaniel, P., Sheth, A.N.: TaintDroid: An Information-Flow Tracking System for Real-time Privacy Monitoring on Smartphones. *Acm Transactions on Computer Systems*. 32, 393-407 (2014)
- 8.Suarez-Tangil, G., Tapiador, J.E., Peris-Lopez, P., Pastrana, S.Power-Aware Anomaly Detection in Smartphones: An Analysis of On-platform Versus Externalized Operation. *Pervasive & Mobile Computing*. 18, 137-151 (2015)
- 9.Memon, A.M., Anwar, A.: Colluding Apps: Tomorrow's Mobile Malware Threat. *IEEE Security & Privacy Magazine*. 13, 77-81 (2015)
- 10.Shen, T., Zhongyang, Y., Xin, Z., Mao, B., Huang, H.: Detect Android Malware Variants Using Component Based Topology Graph. In: 2014 IEEE 13th International Conference on Trust, Security and Privacy in computing and Communications (TrustCom). pp. 406-413. IEEE, Beijing (2014)
- 11.Ali, A., Rida, K., Marc, L.: A Gaussian mixture model for dynamic detection of abnormal behavior in smartphone applications. In: 2014 Global Information Infrastructure and Net-working Symposium (GIIS). pp.1-6. IEEE, Montreal (2014)
- 12.Caviglione, L., Gaggero, M., Lalande, J., Mazurczyk, W., Urbanski, M.: Seeing the Un-seen: Revealing Mobile Malware Hidden Communications Via Energy Consumption and

- Artificial Intelligence. IEEE Trans. Inform. Forensic Secur. 11, 799-810 (2016)
- 13.Das, S., Liu, Y., Zhang, W., Chandramohan, M.: Semantics-Based Online Malware Detection: Towards Efficient Real-Time Protection Against Malware. IEEE Trans.Inform.Forensic Secur. 11, 289-302 (2016)
- 14.Bai, X., Xing, L., Zhang, N., Wang, X., Liao, X., Li, T., Hu, S.: Staying Secure and Unprepared: Understanding and Mitigating the Security Risks of Apple ZeroConf. In: 2016 IEEE Symposium on Security and Privacy (S&P). pp. 655-674. IEEE, San Jose (2016)
- 15.Holland, B., Deering, T., Kothari, S.: Security Toolbox for Detecting Novel and Sophisticated Android Malware. In: 2015 IEEE/ACM 37th IEEE International Conference on Software Engineering (ICSE). pp. 733-736. IEEE, Florence (2015)
- 16.Rahman, M.: DroidMLN: A Markov Logic Network Approach to Detect Android Malware. In: 2013 12th International Conference on Machining Learning and Applications (ICMLA). pp. 166-169. IEEE, Miami (2013)
- 17.Fan, M., Liu, J., Wang, W., Li, H., Tian, Z., Liu,T.: DAPASA : Detecting Android Piggy-backed Apps through Sensitive Subgraph Analysis. IEEE Transactions on Information Fo-rensics & Security. In press (2017)
- 18.Suarez-Tangil, G., Tapiador, J.E., Peris-Lopez, P., Ribagorda, A.: Evolution, Detection and Analysis of Malware for Smart Devices. IEEE Communications Surveys & Tutorials. 16, 961-987 (2014)
19. Ping Yan, Zheng Yan, "A Survey on Dynamic Mobile Malware Detection".Software Quality Journal,pp.1-29,May 2017, Springer.

Automated Water Distribution and Effusion Detection by Adopting Embedded System

Mr.Kiran gurav

Information Technology Department
Theem College of Engineering
Boisar, India
Guravkiran800@gmail.com

Mr.Avinash Pardeshi

Information Technology Department
Theem College of Engineering
Boisar, India
avinashpardeshi260@gmail.com

Prof.Sneha Sankhe

Information Technology Department
Theem College of Engineering
Boisar, India
snehasankhe.88@gmail.com

Mr.krishnakumar Yadav

Information Technology Department
Theem College of Engineering
Boisar, India
Yadavkrishna30894@gmail.com

Mr.Ameya Gharat

Information Technology Department
Theem College of Engineering
Boisar, India
Ameya.gharat99@gmail.com

Abstract—as the increasing demand for water arising from global population growth and urbanization in recent years is stressing the water supply to its limits. On the other hand, water infrastructure such as pipes has been deteriorating due to aging. In most water-distribution systems; a large percentage of the water is lost in transit from treatment plants to consumers. Water loss can be attributed to several causes, including leakage, metering errors, public usage such as fire-fighting and pipe flushing, and theft. Leakage is usually the major cause.

Currently implemented water effusion system is Non-invasive Leakage distribution system, which detects leakage from outside the pipeline using sensors or visual inspection. Such system scans manually entire pipeline for damage. Also some of the new proposed system uses principle of ultrasonic transducer, which is not suitable for discontinuities in the fluid (bubbles), in such system consumer is unaware about effusion.

As our country moves towards Smart City, there is need to develop systems that provide efficiency to precious water infrastructure. We proposed the invasive Leakage distribution system that monitor flow rate of water in distribution pipe using Hall Effect sensor. The sensors are mounted at various node/points in water distribution pipe. We create a new algorithm called as an A2k for effusion detection. The Effusion in distribution pipe is detected by calculating difference in flow rate between two nearest sensor. At particular node, the system take current value and one previous value of flow sensor if both values are same, then it compare it with another flow sensor to find effusion between those two nodes. If the sensors sense any change in the rate of flow of water then the person at control station will close the automated valve at the pipeline by using GUI also system is able to close the valve itself. These will close the flow of water in that area of the pipeline.

Thus the water will not be wasted and its savage is done without wasting time. Graphical user interface is present at control station to monitor or control the water distribution, also registered users or consumers at particular location will get message regarding effusion in their area. The proposed system help government authority to control or monitor water distribution, future water consumption.

Keywords—Hall affect sensor, water effusion.

I. INTRODUCTION

Earlier the water infrastructure monitoring process was done by a human that caused an error. The presently available system leads to unnecessary loss of water and due to improper handling; water is not properly distributed to the end users. Also, issue of leakage is not handled properly that effects on our natural resources such as water. Also, consumers are not able to get alert related to some critical problems in water infrastructure.

II. OBJECTIVE

- Handle water Effusion issue by using flow sensors and actuators to save lots of water.
- All the manual work done by humans are reduced because of automated system apply on water infrastructure

- All the areas in the city have different need of water according to that we can change water pressure in different areas.
- Make government authorities work smart by adopting these system authorities are able to monitor or control entire water infrastructure using software interface.
- Alert or inform consumers about water effusion via SMS on mobile

III. SYSTEM DESIGN

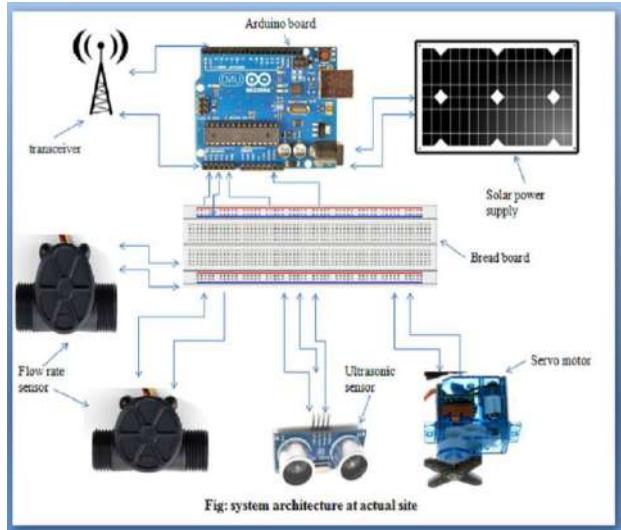


Fig a: system architecture at actual site

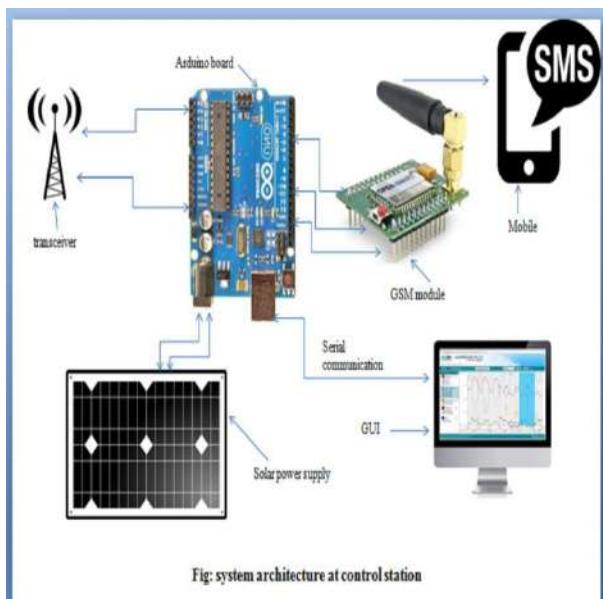


Fig b: system architecture at control station

IV. WORKING

One of our achievements is to successfully estimate if there is an effusion in the pipeline. These are done with the help of the flow rate sensors fixed into the pipelines at the distributed ends of the pipeline if the flow rate sensor senses any deflection between any two sensors. The alert is given at the control station computer GUI software, which means there is effusion in the particular area. This is handled by closing the automated valve in that particular area through software present at the control station. Different areas in the city have their different need as per their population and importance. This system can manage water pressure according to demand of water.

This system is based on embedded technology. The GSM Module is used for sending the messages to the registered person in the particular area. The quality sensors for checking the quality of the water flow in the region. The solar power supply is used to power Arduino board. The automated valve that is revolves 180 degrees so that the valve can be closed automatically.

This system helps government authorities to work smart. Inspection of the city's water infrastructure is a very complex task which is reduced by this system up to some extent. Distribution of water at different areas in the city can be handled properly. Because of real-time leak alerts, water wastage is reduced and handled efficiently. This will help authorities to work on many other problems in the city.

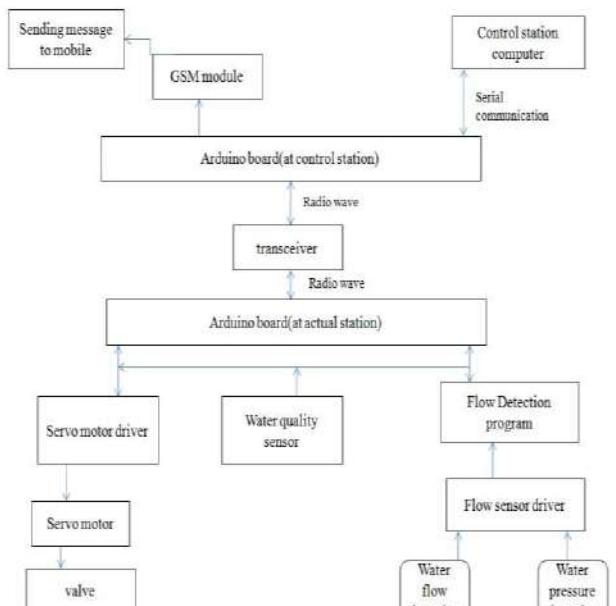


Fig: working.

V. ADVANTAGES

1. The idea (discusses to a social problem and suggests a solution): this system is based on social problem and provides complete water solution.
2. Effective utilization of natural resources: As whole water, infrastructure will be monitor by the system so there is effective utilization of resources.
3. Increase overall efficiency of water distribution: Authorities control water distribution using software interface so it can be more transparent.
4. Reduce operational expense: This system is able to detect effusion, so it reduces extra operational cost.
5. The power supply is providing by solar power panels as it requires being work on the whole day.
6. Alert consumers via SMS facility about water effusion.

Application

1. Business applications:

The system has many business applications as on the large scale it is very useful for city's water distribution and reduces the operational cost of government authorities.

2. Flow Detection system:

The flow detection system is used to detect the flow of oil in the pipe lines for oil industries.

3. Water pressure management system:

This system will apply in housing complexes for pressure management.

REFERENCES

- [1] Pratiksha Hattikatti, Savita Karwande, Poonam S. Rode, Mrunalini Bhandarkar "PLC Controlled Water Distribution System (International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering)" Vol. 4, Issue 5, May 2015.
- [2] Santosh Kashid, Sanjay pardeshi "A survey of water distribution system and new approach to intelligent water distribution system" 2014 IEEE.
- [3] Alexandre Santos, Mohamed Younis "A Sensor Network for Non Intrusive and Efficient Leak Detection in Long Pipelines" 2011 IEEE.
- [4] S.C.Hsia, S.W.Hsu, Y.J.Chang "Remote monitoring and smart sensing for water meter system and leakage detection" IET Wireless sensor systems.
- [5] Rakhel Kumar Parida, Vidhya Thyagarajan, Sreeja Menon "A Thermal Imaging Based Wireless Sensor Network for Automatic Water Leakage Detection in Distribution Pipes" IEEE CONECT2013 1569679473.
- [6] A.D.Harnawal, R.S.Pol, P.S.Done, A.D.Jambulkar, Prof.Dr.P.Patil "Wireless module for smart cities" (International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering) Vol-2, Issue-3 2016.
- [7] PHMSA, "Fact Sheet - Corrosion", <http://primis.phmsa.dot.gov/comm/FactSheets/FSCorrosion.htm?nocache=684>.

SECURE LOGGING AS A SERVICE—DELEGATING LOG MANAGEMENT TO THE CLOUD

Prof. Rosy Bhoi, Prof. Sanketi Raut, Prof. Ahmed shaikh

ABSTRACT

Securely maintaining log records over extended periods of time is very important to the proper functioning of any organization. Integrity of the log files and that of the logging process need to be ensured at all times. In addition, as log files often contain sensitive information, confidentiality and privacy of log records are equally important. However, deploying a secure logging infrastructure involves substantial capital expenses that many organizations may find overwhelming. Delegating log management to the cloud appears to be a viable cost saving measure. In this paper, we identify the challenges for a secure cloud-based log management service and propose a framework for doing the same.

EXISTING SYSTEM

Log files contain record of most system events including user activities; they become an important target for malicious attackers. An attacker, breaking into a system, typically would try not to leave traces of his or her activities behind. Consequently, the first thing an attacker often does is to damage log files or interrupt the logging services. It is very important to provide a logging in a secure manner and that the log records are adequately protected for a predetermined amount of time.

EXISTING TECHNIQUE

- Syslog protocol

DISADVANTAGE

- It does not ensure correctness of logs.
- High Cost required to maintain the logs

PROPOSED SYSTEM

In this paper, we propose a comprehensive solution for storing and maintaining log records in a server operating in a cloud-based environment. We address security and integrity issues not only just during the log generation phase, but also during other stages in the log management process, including log collection, transmission, storage, and retrieval. This successfully prevents the cloud provider or any other observer from correlating requests for log data with the requester or generator. Finally, we develop a proof-of-concept prototype to demonstrate the feasibility of our approach and discuss some early experiences with it. To the best of our knowledge, ours is the first work to provide a complete solution to the cloud based secure log management problem.

PROPOSED TECHNIQUE

Cryptographic protocols

ADVANTAGE

- More secure and confidential
- Low cost for store and manage log records in a proper manner.
- Very fast to transmit the log file.

SOFTWARE REQUIREMENT

- Operating system :- Windows7
- Front End :- Microsoft Visual Studio .Net 2010
- Coding Language :- C#
- Backend :- SQL Server 2005

HARDWARE REQUIREMENT

- Processor : Pentium Dual Core 2.00GHZ
- Hard disk : 40 GB

- Mouse : Logitech.
- RAM : 2GB(minimum)
- Keyboard : 110 keys enhanced.

FUTURE WORK

In the future, we plan to refine the log client implementation so that it is tightly integrated with the OS to replace current log process. In addition, to address privacy concerns current implementation allows access to log records that are indirectly identified by upload-tag values. We plan to investigate practical homomorphism encryption schemes that will allow encryption of log records in such a way that the logging cloud can execute some queries on the encrypted logs without breaching confidentiality or privacy. This will greatly reduce the communication overhead between a log monitor and the logging cloud needed to answer queries on logs.

PRIVACY PRESERVING DELEGATED ACCESS CONTROL IN PUBLIC CLOUDS

Prof. Rosy Bhoi, Prof. Sanketi Raut, Prof. Ahmed shaikh

ABSTRACT

Current approaches to enforce fine-grained access control on confidential data hosted in the cloud are based on fine-grained encryption of the data. Under such approaches, data owners are in charge of encrypting the data before uploading them on the cloud and re-encrypting the data whenever user credentials change. Data owners thus incur high communication and computation costs. A better approach should delegate the enforcement of fine-grained access control to the cloud, so to minimize the overhead at the data owners, while assuring data confidentiality from the cloud. We propose an approach, based on two layers of encryption that addresses such requirement. Under our approach, the data owner performs a coarse-grained encryption, whereas the cloud performs a fine-grained encryption on top of the owner encrypted data. A challenging issue is how to decompose access control policies (ACPs) such that the two layer encryption can be performed. We show that this problem is NP-complete and propose novel optimization algorithms. We utilize an efficient group key management scheme that supports expressive ACPs. Our system assures the confidentiality of the data and preserves the privacy of users from the cloud while delegating most of the access control enforcement to the cloud.

EXISTING SYSTEM

In the approaches subdocuments are encrypted with different keys, which are provided to users at the registration phase. The encrypted subdocuments are then broadcasted to all users. However, such approaches require that all or some keys be distributed in advance during user registration phase. This requirement makes it difficult to assure forward and backward key secrecy when user groups are dynamic. Further, the rekey process is not transparent, thus shifting the burden of acquiring new keys on users. It lays the foundation to make rekey transparent to users and protect the privacy of the users who access the content.

EXISTING TECHNIQUE

- Single Layer Encryption Approach

DISADVANTAGES

- In this approaches incur high communication and computation cost to manage keys and encryptions whenever user credentials change.
- In such approaches, the data owner has to enforce the ACPs and the privacy of the users from the content publisher is not protected.
- In some approaches, multiple encryptions of the same document are required which is inefficient.

PROPOSED SYSTEM

In this system, we propose a new approach to address this shortcoming. The approach is based on two layers of encryption applied to each data item uploaded to the cloud. Under this approach, referred to as two layer encryption (TLE), the data owner performs a coarse grained encryption over the data in order to assure the confidentiality of the data from the cloud. Then the cloud performs fine grained encryption over the encrypted data provided by the data owner based on the ACPs provided by the data owner. However, the way we perform coarse and fine grained encryption is novel and provides a better solution than existing solutions based on two layers of encryption.

PROPOSED TECHNIQUE

- Two Layer Encryption (TLE)

ADVANTAGES

- It provides a better way to handle data updates, and user dynamics changes.
- To support expressive access control policies.
- When user dynamics changes, only the outer layer of the encryption needs to be updated.

- Outer layer encryption is performed at the cloud, no data transmission is required between the data owner and the cloud.

SOFTWARE REQUIREMENT

- Operating system :- Windows7
- Front End :- Microsoft Visual Studio .Net 2010
- Coding Language :- C#
- Backend :- SQL Server 2008

HARDWARE REQUIREMENT

- Processor : Pentium Dual Core 2.00GHZ
- Hard disk : 40 GB
- Mouse : Logitech.
- RAM : 2GB(minimum)
- Keyboard : 110 keys enhanced.

FUTURE WORK

In future, we plan to investigate the alternative choices for the TLE approach further. We also plan to further reduce the computational cost by exploiting partial relationships among ACPs.

Ad Campaign Analytics

Ajinkya Naik, Ashutosh Dwivedi

B.TECH Students, Information Technology
Ramrao Adik Institute of Technology
Mumbai, India
ajinkyanai1997@gmail.com

Sahil Jadhav, Zaid Shaikh

B.TECH Students, Information Technology
Ramrao Adik Institute of Technology
Mumbai, India
zaid.r.shk@gmail.com

Abstract— Ad Campaign Analytics is the measurement, collection, analysis and reporting of data for purposes of understanding and optimizing business strategies taken to spread a central idea through ad campaigns. Nowadays it is used as a tool for business and market research and it also assesses and improves the effectiveness of ad campaigns from various organizations. Ad Campaign Analytics can also help companies measure the results of broadcast advertising campaigns.

In this project, we intend to develop a system which would help its users as well as clients with the generation of statistical and heuristic reports, provide multiple and lucrative basis for audience recognition and response, which can further help them to know their return on investment through logical notifiers. This will help the users as well as web admins to get a grasp on the total overall effectiveness of marketing and hence, predict the campaign performance as well as proper gist of current situations which can be taken into consideration while making future decisions regarding their respective business.

Keywords—ad campaign; analytics; business; market research;

I. INTRODUCTION

An advertising campaign is a coordinated series of linked advertisements with a single idea or theme. An advertising campaign is typically broadcast through several media channels. It focuses on a common theme and one or few brands or products, or is directed at a particular segment of the population. Successful advertising campaigns achieve far more than the sporadic advertising, and may last from a few weeks and months to years. Analytics is the discovery, interpretation, and communication of meaningful patterns in data. It is especially valuable in areas rich with recorded information. Analytics relies on the simultaneous application of statistics, computer programming and operations research to quantify performance.

II. LITERATURE SURVEY

A) Andy Bengel, Amin Shawki, who are software engineers at InfoTrust, LLC USA have mentioned in their paper that their approach offers a competitive advantage for an organization's marketing by providing a real-time insight into the tag behavior on your site. The main contributions of our approach and software are as follows:

- A web-based, real-time tag discovery and

analysis tool.

- Configurable interface for custom time range over the last seven days and website.
- Validate tags firing conditionally (tags that only fire when users come from a particular source or have a particular characteristic)
- Dashboard oriented visualization providing a complete picture of site's and tag performance
- A pleasing yet simple and intuitive look and feel with support for exploring the vast amount of information provided by the product in a drill-down manner thus making it easier to comprehend.
- Facilitates information capture on important site performance measures such as tag latency and page load time. Having all this information can be invaluable for marketers and analysts who can verify successful deployment of any tags that have been deployed recently, ensure that any tags that are taken off are not loading anymore and all this information is made available in real-time without having to schedule a website scan and studying the report later. Employing page-tagging can help an organization in ensuring compliance with privacy policies, source attribution and tracking visitor's browsing patterns to provide personalized and targeted advertising [2]. In the context of digital marketing, the concept of attribution is about understanding which channels or social media platforms are coming across as most effective in terms of attracting visitors to the site. The relevance about knowledge on source attribution can enable the marketers to decide where they should spend their marketing budget. By tracking the links that led the user to arrive at a certain page, tags can support source attribution.

B) Hana Anber, Akram Salah, A. A. Abd El-Aziz

Identify applicable sponsor/s here. If no sponsors, delete this text box (sponsors).

- Computer and Information Sciences Department, Institute of Statistical Studies and Research, Cairo University, Giza, Egypt.
- Computer Science Department, Faculty of Computers and Information, Cairo University, Giza, Egypt.
- Information System and Technology Department, Institute of Statistical Studies and Research, Cairo University, Giza, Egypt.

According to their research analyzing structured data have been widely used. In such case, the traditional Relational Database Management System (RDBMS) can deal with the data. With the increasing amounts of unstructured data on various sources (e.g. Web, Social media, and Blog data) that are considered as Big Data, a single computer processor cannot process such huge amount of data. Hence, the RDBMS cannot deal with the unstructured data; a nontraditional database is needed to process the data, which is called NoSQL database. Most studies focused on tools, such as R (the programming language and the software environment for data analysis). R has limitations when processing large data, and is not efficient in dealing with large volume of data. To solve this problem a hybrid big data framework is usually employed, such as Apache Hadoop (an open source Java framework for processing and querying vast amounts of data on large clusters of commodity hardware)

C) Nada Elgendi and Ahmed Elragal

In their research, they have examined the innovative topic of big data, which has recently gained lots of interest due to its perceived unprecedented opportunities and benefits. In an era where technologies are developed frequently and the information gets updated in every few seconds, different varieties of data are being produced daily. From this data we should be able to extract and utilize patterns as well as intrinsic details which might benefit business organizations. Big data analytics can be used for efficient decision making in business as well as to reap benefits for business organizations. For understanding decision making techniques and methods to analyze big data, the literature was reviewed. As a result this led to discussion of various concepts of big data, its characteristics and importance. Also, examination of some tools and methods used for analyzing big data were done. Hence, big data storage and management, big data analytics were detailed. Moreover, different data analytics were further discussed

D) Dippy Aggarwal

According to Dippy Aggarwal online marketers and analysts are not only interested in getting and studying a holistic view of their site's tag maintains a list of several hundreds of tags in the digital world. This allows us to categorize the tags identified on the website in terms of recognized names in the tag industry. Reporting shows what tags are firing on each page, what parameters are being passed

by each tags, and the values populating each parameter. The capabilities entail listening to the tags on pages with and without login, as well as the ones fired off on event occurrences such as button click etc. We present an analytics and visualization-driven solution to an organization's challenge of ensuring profitable return on investment (or ROI) on their marketing efforts. Our approach backed up by tool support empowers marketers and analysts by providing them with all the information related to tags that are deployed on their websites in real time. We believe that our paper contributes in highlighting the relevance of developing advanced solutions for digital marketing industry which is forecasted to be on consistent rise in the years to come.

III. CONTRIBUTION

The demand for data analysts is on a hike, the demand is rising and more organizations are hiring data analysts. As the need for jobs is growing, more people are gravitating towards this profession. Also, more and more businessmen are looking for world class analysts as this is how they will see a way to make a profit. A professional may thus need analytical skills as to understand the work patterns. Our system keeps a track on number of visitors visiting an ad and their location which will assist organizations in improving business strategies. It will enable marketers to measure ad campaign performance and find out which campaign drives the best visitors. Ad campaign analytics helps in identifying target audience and will increase ROI by keeping a track of performance of business (ads) related to visitor interests and demographic data.

The scope of the system is as follows:-

- To get an overview of user stats for traffic.
- To find out which online campaign brings the most traffic and conversions.
- To determine the region where the best visitors are located.
- To learn what people are searching for on your website.
- To visualize which ad people click on the most.
- To find out and analyze the worst performing ads.

IV. CONCLUSION

One of the most essential factors for progress of an organization is advertisement. The basic objective of advertisement is creating awareness among customers and building sales and profit. However, in this competitive era, advertisement alone is not enough. Due to which, further analysis is essential. Our project will provide all analytical data regarding the advertisement campaign and how it affects the sales while also notifying us about the return on investments. This project will provide organizations or businesses with aggregated data regarding the campaigns which are undertaken. This will be helpful for them to take proper future decisions and will help them boost their profits.

REFERENCES

- [1] Hana Ambar, Akram Salah, A. A. Abd El-Aziz "Literature Review on Twitter Data Analysis" Institute of Statistical Studies and Research, Cairo University, Giza, Egypt, Dec 2015.
- [2] Nada Elgendi and Ahmed Elragal "Big Data Analytics: A Literature Review Paper" Department of Business Informatics & Operations, German University in Cairo (GUC), Cairo, Egypt, August 2014.
- [3] Dippy Aggarwal "Simplifying web analytics for digital marketing" Ph.D. Student, Computer Science and Engineering University of Cincinnati Cincinnati, Ohio, USA, Jan 2015.
- [4] Andy Bengel, Amin Shawki "Simplifying web analytics for digital marketing" InfoTrust, LLC USA, Jan 2015
- [5] Marketing Analytics by Wayne Winston.
- [6] An Introduction to SVM: John Shawe-Taylor and Nello Cristianini, 2000.

WEBSITE FOR COLLEGE CANTEEN

GUIDE - MRS. VEENA KULKARNI

GROUP MEMBERS :

SIDDHARTH TIWARI(50)

MUKUL SUVARNA(45)

VIVEK SHARMA(26)

AKSHIT VAGHANI(51)

ABSTRACT

In today's world where people are running out of time almost everywhere, who would like to stand in a queue to order food in canteen and then again wait for the order to be served? The answer is no one. In order to deal with this problem in our college canteen where people spend more time on waiting for their food than eating it, our team has come up with the idea of preordering the food before going to the canteen. This will be done through website. Every student, teacher and other college members can order their food before going to the canteen and enjoy their food.

INTRODUCTION

We will provide each and every member of TCET with a unique user id so that the benefit is taken by our college members only. We will provide the whole menu on this website. The most important part of this project is the use online payment methods. We will connect our website through payment gateway so that it saves the papers used in our currencies as well as the bill. The user can have their food just by showing the message sent by the admin regarding the confirmation of their order.

ADVANTAGES: 1)Saves Time and Paper

2)Works Efficiently

Softwares Used: HTML, JAVASCRIPT AND DBMS

REQUIREMENT ANALYSIS

We have used majorly three programming method.

- HTML is the backbone of our website. The website is solely made on HTML. It is an programming language that's a bit lengthy but easy to code as well. We have used almost every property of HTML in our project.
- JAVASCRIPT used in the development of this website plays a hidden but a vital role. It's various functions are used in order to verify the user while registering for the website.
- DBMS is the used to get and store the information given by the user. The admin would get the data entered by the user through DBMS and will forward their order.

REFERENCE:

Various topics on internet.

Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factors

Parth Sharma¹, Aditya Samant², Adit Shrima³, Deepak Sharma⁴

^{1,2,3}Department of Computer Engineering

Thakur College of Engineering and Technology, Mumbai , India

sharmaparth33@gmail.com¹, adityasamant1996@gmail.com², aditshrima7@gmail.com³,
deepakmsharma51@gmail.com⁴

Abstract- Data mining techniques have been widely used in clinical decision support systems for prediction and diagnosis of various diseases with good accuracy. These techniques have been very effective in designing clinical support systems because of their ability to discover hidden patterns and relationships in medical data. One of the most important applications of such systems is in diagnosis of heart diseases because it is one of the leading causes of deaths all over the world. Almost all systems that predict heart diseases use clinical dataset having parameters and inputs from complex tests conducted in labs. None of the system predicts heart diseases based on risk factors such as age, family history, diabetes, hypertension, high cholesterol, tobacco smoking, alcohol intake, obesity or physical inactivity, etc. Heart disease patients have lot of these visible risk factors in common which can be used very effectively for diagnosis. System based on such risk factors would not only help medical professionals but it would give patients a warning about the probable presence of heart disease even before he visits a hospital or goes for costly medical checkups. Hence this paper presents a technique for prediction of heart disease using major risk factors. This technique involves two most successful data mining tools, neural networks and genetic algorithms. The hybrid system implemented uses the global optimization advantage of genetic algorithm for initialization of neural network weights. The learning is fast, more stable and accurate as compared to back propagation. The system was implemented in Matlab and predicts the risk of heart disease with an accuracy of 89%.

Keywords- data mining, heart disease risk factors, prediction and diagnosis systems.

I. INTRODUCTION

Heart diseases are the number one cause of death globally: more people die annually from Heart diseases than from any other cause. An estimated 17.3 million people died from Heart diseases in 2008, representing 30% of all global deaths. Of these deaths, an estimated 7.3 million were due to coronary heart disease and 6.2 million were due to stroke [1]. Recent research in the field of medicine has been able to identify risk factors that may contribute toward the development of heart disease but more research is needed to use this knowledge in reducing the occurrence of heart diseases. Diabetes, hypertension, and high blood cholesterol have been established as the major risk factors of heart diseases. Life style risk factors which include eating habits, physical inactivity, smoking, alcohol intake, obesity are also associated with the major heart disease risk factors and heart

disease [2,3].There are studies showing that reducing these risk factors for heart disease can actually help in preventing heart diseases [4]. There are many studies and researches on the prevention of heart disease risk. Data from studies of population has helped in prediction of heart diseases, based on blood pressure, smoking habit, cholesterol and blood pressure levels, diabetes. Researchers have used these prediction algorithms in adapted form of simplified score sheets that allow patients to calculate the risk of heart diseases [6]. The Framingham Risk Score (FRS) is a popular risk prediction criterion which is used in algorithms for heart disease prediction [7].

This study aimed at developing an intelligent data mining system based on genetic algorithm optimized neural networks for the prediction of heart disease based on risk factors' categories. The system was implemented using MATLAB R2012a.

II. DATA MINING TECHNIQUES

Data mining techniques are used to explore, analyze and extract medical data using complex algorithms in order to discover unknown patterns. Researchers are using data mining techniques for the diagnosis of many diseases such as heart disease [8], diabetes [9], stroke [10] and cancer [11] and many data mining techniques have been used in the diagnosis of heart disease with good accuracy. Researchers have been applying different data mining techniques such as naïve bayes, neural network, decision tree, bagging, kernel density, and support vector machine for prediction and diagnosis of heart diseases [13]-[15]. One of the systems [16] uses neural based learning classifier for classifying data mining tasks showed that neural based learning classifier system performs equivalently to supervise learning classifier. IEHPS [17] intelligent and effective heart attack prediction system was built using data mining and neural networks and it proposed extracting significant patterns for heart disease prediction using K-means clustering and used MAFIA algorithm to mine the frequent patterns. Polatet al., developed system using hybrid fuzzy and k-nearest neighbour approach for the prediction of heart disease, which had 87% accuracy in diagnosis [18]. In another system [19] neural network ensemble was used in the diagnosis of heart disease with an accuracy of 89.01%. Latha and Subramanian (2007), proposed an intelligent heart disease

prediction system using CANFIS and genetic algorithm which had a very low mean square error [20]. Analyzing the different techniques discussed, this paper proposes a novel system using genetic algorithm and neural network for predicting the risk of heart diseases. Genetic algorithm is used to optimize neural network weights. What is even more different in this paper is that it is the first time that such a hybrid technique is applied on risk factors for the accurate prediction of heart disease. Hence the main objective is not only to use this system in clinical decision support but to also use this system as risk indicator so that it helps people reduce the risks of having any heart disease in future.

III. MATERIALS AND METHODS

A. Data Analysis and Encoding

The problem with risk factors related to heart disease is that there are many risk factors involved like age, usage of cigarette, blood cholesterol, person's fitness, blood pressure, stress and etc. and understanding and categorizing each one according to its importance is a difficult task. Also a heart disease is often detected when a patient reaches advanced stage of the disease [21]. Hence the risk factors were analyzed from various sources [22]-[23]. The dataset was composed of 12 important risk factors which were sex, age, family history blood pressure, Smoking Habit, alcohol consumption, physical inactivity, diabetes, blood cholesterol, poor diet, obesity .The system indicated whether the patient had risk of heart disease or not. The data for 50 people was collected from surveys done by the American Heart Association [23]. Most of the heart disease patients had many similarities in the risk factors [24]. The TABLE I below shows the identified important risk factors and the corresponding values and their encoded values in brackets, which were used as input to the system.

TABLE I
RISK FACTORS VALUES AND THEIR ENCODINGS

	Risk Factors	Values
1	Sex	Male (1), Female (0)
2	Age (years)	20-34 (-2), 35-50 (-1), 51-60 (0), 61-79 (1) ,>79 (2)
3	Blood Cholesterol	Below 200 mg/dL - Low (-1) 200-239 mg/dL - Normal (0) 240 mg/dL and above - High (1)
4	Blood Pressure	Below 120 mm Hg- Low (-1) 120 to 139 mm Hg- Normal (0) Above 139 mm Hg- High (-1)
5	Heredity	Family Member diagnosed with HD -Yes (1) Otherwise -No (0)
6	Smoking	Yes (1) or No (0)
7	Alcohol Intake	Yes (1) or No (0)
8	Physical Activity	Low (-1) , Normal (0) or High (-1)
9	Diabetes	Yes (1) or No (0)
10	Diet	Poor (-1), Normal (0) or Good (1)
11	Obesity	Yes (1) or No (0)

12	Stress	Yes (1) or No (0)
Output	Heart Disease	Yes (1) or No (0)

Data analysis has been carried out in order to transform data into useful form, for this the values were encoded mostly between a range [-1, 1]. Data analysis also removed the inconsistency and anomalies in the data. This was needed. Data analysis was needed for correct data preprocessing. The removal of missing and incorrect inputs will help the neural network to generalize well.

B. Neural Network Weight Optimization by Genetic Algorithm

This system uses backpropagation algorithm for learning and training the neural network, but there are two major disadvantages with backpropagation algorithm. First is that the initialization of the NN weights is a blind process hence it is not possible to find out globally optimized initial weights and there is a danger that the network output would run towards local optima hence the overall tendency of the network to find out a global solution is greatly affected. The second problem is that backpropagation algorithm is very slow in convergence and there is a possibility that network never converges [25]. This problem of local optimum solution can be solved by optimizing the initial weights of neural network. For this we use a genetic algorithm which is specialized for global searching [26]. For this we first determine the number of inputs, layers and hidden neurons of the neural network and then we would use the backpropagation algorithm to train the networks using the weights optimized by GA.

C. Neural Network Architecture

A multilayered feed-forward network is used having 12 input nodes 10 hidden nodes and 2 output nodes. The number of inputs is based on the final set of risk factors for each patient which is given in TABLE I. number of hidden nodes must be decided for which the training is fast and the network gives the best output. The first step is to initialize the weights of neural network using the ‘configure’ function available in MATLAB. Then these configured weights are passed to the genetic algorithm for optimization according to the fitness function. Once the weights are optimized, the Levenberg-Marquardt backpropagation algorithm is used for training and learning and ‘trainlm’ is a network training function that updates weight and bias values according to Levenberg-Marquardt optimization. The ‘trainlm’ is often the fastest backpropagation algorithm in the toolbox, and is highly recommended as a first-choice supervised algorithm, although it does require more memory than other algorithms. Maximum number of epochs to train is set to a default value 100. The learning stops at a predefined minimum error after modifying network weights and adjusting them to an optimal quantity at which the classification is accurate. The predicted output would be presence or absence of a heart disease.

TABLE II
PATIENT'S CASE STUDY DATA IN ENCODED FORM

No	Sex	Age	Blood Cholesterol	Blood Pressure	Hereditary	Smoking	Alcohol Intake	Physical Activity	Diabetes	Diet	Obesity	Stress	Heart Disease
1	Female	35	High	Normal	No	No	Yes	Low	Yes	Poor	Yes	Yes	Yes
2	Male	70	Low	Low	No	No	Yes	High	Yes	Normal	No	No	No
3	Female	60	High	High	No	No	No	Normal	Yes	Poor	Yes	Yes	Yes
4	Female	36	Low	Normal	No	No	No	Normal	No	Good	No	No	No
5	Male	30	Low	Normal	No	No	Yes	High	No	Normal	No	No	No
6	Female	39	Low	Normal	Yes	No	Yes	High	Yes	Normal	No	Yes	No
7	Female	41	High	Normal	No	No	No	Low	No	Poor	Yes	No	No
8	Male	70	High	Normal	No	No	Yes	Low	No	Poor	Yes	No	Yes
9	Male	65	Normal	High	Yes	Yes	Yes	Normal	Yes	Poor	Yes	No	Yes
10	Male	30	Normal	High	No	Yes	No	Normal	No	Good	No	Yes	No
11	Female	31	Low	Normal	No	No	No	High	No	Normal	No	No	No
12	Female	29	Low	Normal	No	No	Yes	High	No	Good	No	No	No
13	Male	30	Low	Normal	No	No	Yes	Normal	No	Normal	No	No	No
14	Female	45	Normal	High	Yes	Yes	No	Normal	Yes	Normal	Yes	Yes	No
15	Male	25	High	Normal	Yes	Yes	Yes	Low	Yes	Normal	No	No	Yes
16	Female	37	Normal	Normal	No	No	No	Normal	Yes	Poor	No	Yes	No
17	Female	37	Normal	High	No	Yes	Yes	High	No	Poor	No	Yes	No
18	Male	53	High	Low	No	Yes	No	Normal	Yes	Normal	No	Yes	No
19	Male	57	High	Normal	No	Yes	No	Low	No	Poor	Yes	Yes	Yes
20	Male	52	High	Low	No	No	No	Normal	Yes	Poor	Yes	No	No
21	Male	48	Normal	Normal	Yes	Yes	Yes	Normal	No	Normal	No	No	Yes
22	Male	62	High	High	No	Yes	Yes	Normal	Yes	Normal	No	No	Yes
23	Male	56	Normal	High	No	Yes	Yes	Low	No	Poor	Yes	Yes	Yes
24	Female	27	Low	Normal	No	No	No	High	No	Good	No	No	No
25	Male	33	Normal	Normal	No	No	No	Normal	Yes	Good	No	No	No
26	Female	33	Normal	Normal	No	No	Yes	Low	Yes	Poor	No	Yes	No
27	Male	37	High	Normal	No	No	Yes	Normal	No	Normal	No	Yes	No
28	Male	43	Normal	High	No	No	No	Normal	Yes	Poor	Yes	Yes	Yes
29	Male	46	Low	Normal	No	No	No	Normal	Yes	Poor	Yes	Yes	No
30	Female	36	Low	Normal	No	No	No	Normal	No	Normal	No	No	No

31	Female	29	Low	Normal	No	No	No	Normal	No	Good	No	No	No
32	Female	47	Normal	Normal	No	No	Yes	High	Yes	Normal	No	Yes	No
33	Male	58	High	High	No	Yes	Yes	Normal	Yes	Normal	No	Yes	Yes
34	Male	44	High	Normal	Yes	Yes	Yes	Normal	No	Normal	Yes	Yes	Yes
35	Female	36	Normal	High	No	No	No	Normal	No	Good	Yes	No	Yes
36	Male	42	Low	Normal	Yes	No	Yes	Low	No	Poor	No	Yes	No
37	Female	25	Low	Normal	No	No	No	High	No	Poor	No	No	No
38	Female	28	Low	Normal	No	No	Yes	High	No	Normal	No	No	No
39	Female	26	Low	Normal	Yes	No	No	Normal	No	Normal	Yes	No	Yes

40	Male	28	Low	Normal	No	No	No	Normal	No	Poor	No	No	No
41	Female	45	High	Normal	No	No	Yes	Low	Yes	Poor	Yes	Yes	Yes
42	Male	63	Low	Low	No	No	Yes	High	Yes	Good	No	No	No
43	Female	55	High	High	No	No	No	Normal	Yes	Normal	Yes	Yes	Yes
44	Female	44	Low	Normal	No	No	No	Normal	No	Normal	No	No	No
45	Male	35	Low	Normal	No	No	Yes	High	No	Normal	No	No	No
46	Female	42	Normal	Normal	No	No	Yes	High	Yes	Good	No	No	No
47	Female	43	Normal	Normal	No	No	No	Low	No	Poor	Yes	No	No
48	Male	65	Normal	Normal	No	No	Yes	Low	No	Normal	Yes	Yes	Yes
49	Male	74	Normal	High	No	Yes	Yes	Normal	Yes	Normal	Yes	Yes	Yes
50	Male	36	Normal	High	No	Yes	No	Normal	No	Poor	No	No	No

IV. PARAMETER SETTINGS

The system was developed using MATLAB R2012a. Global Optimization Toolbox and the Neural Network Toolbox were used for implementing the algorithm [27]. The data for risk factors related to heart diseases collected from 50 people is provided in TABLE II. ANN is initialized with the ‘configure’ function, with each weigh being between -1.0 to +1.0. These weights are then passed to the genetic algorithm which uses the mean square error as the fitness function. The interconnecting weights and thresholds of the trained neural network are passed to the genetic algorithm. The number of neurons in the three layer neural networks is 12, 10, and 2 respectively in input, hidden and output layer. Hence there are $(12 \times 10 + 10) + (10 \times 2 + 2) = 152$ total weights and biases. The weights in the ANN are encoded in such a way each weight is being between -1.0 to +1.0. After that weights are assigned to each link. Weights adjustment using GA is done with ‘population size =20’. In this

application, each string or chromosome in the population represents the weight and bias values of the network. Fitness function is calculated for each chromosome based on mean square error. The fitness function

TABLE III
SOME PARAMETERS USED IN GA

Search Method	Genetic Algorithm
PopulationSize	20
Generations	100
CrossoverFraction	0.8000
MigrationInterval	20
MigrationFraction:	0.2000
EliteCount	2
TolFun	1.0000e-006

V. RESULTS AND DISCUSSION

The input data consisted of risk factors collected from 50 people through case studies provided at the website of the American Heart Association [23]. The data was encoded as shown in TABLE II. 70% of the data was used for training and 15% each

$$mse = \frac{\sum_{k=1}^n (O_k - T_k)^2}{n}$$

used is mean square error (mse) which is calculated as below:

After selection, crossover and mutation in GA, the chromosomes with lower adaptation are replaced with better ones, and the better and fitter chromosomes (optimized solutions) that correspond to the interconnecting weights and thresholds of neural network are generated. A small value, closes to zero, shows that the network has generalized well and is ready for the classification problem. In this method GA searches among several set of weight vectors simultaneously. The initial population is randomly generated. By selecting suitable parameters, like selection criteria, probability of crossover, probability of mutation, initial population, etc., to the GA, high efficiency and performance can be achieved.

for testing and validation. A confusion matrix is produced using Matlab and accuracy is determined (shown in TABLE IV) as Accuracy = $(TP + TN) / (TP + FP + TN + FN)$; where TP, TN, FP and FN denotes true positives, true negatives, false positives and false negatives, respectively. The accuracy of prediction of heart disease on the training data was calculated as 89% and accuracy on validation data

Training Set	34	96.2%
Test Set	8	92%
Validation Set	8	89%
Total instances	50	-

Gradient = 0.0061426, at epoch 18

was 96.2%. The least mean square error (MSE) achieved was

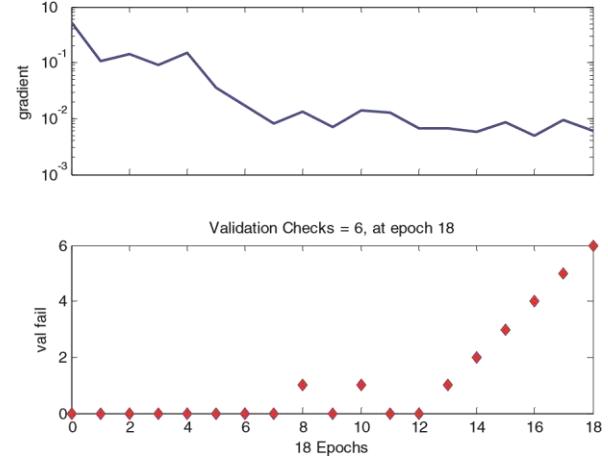


Figure 2: Training State Graph

0.034683 after 12 epochs, as shown in Figure 1. Results show genetic algorithm and neural network approach gives better average prediction accuracy than the traditional ANN.

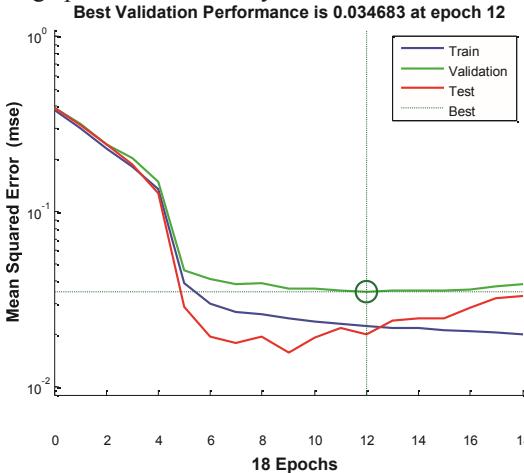


Figure 1: Performance Graph

TABLE IV
DATA SETS

Data Set	Number of Data	Accuracy (%)

VI. CONCLUSION

Data mining techniques and methods applied in patient medical dataset has resulted in innovations, standards and decision support system that have significant success in improving the health of patients and the overall quality of medical services. But we still need systems which could predict heart diseases in early stages. In this study, a new hybrid model of Neural Networks and Genetic Algorithm to optimize the connection weights of ANN so as to improve the performance of the Artificial Neural Network. The system uses identified important risk factors for the prediction of heart disease and it does not require costly medical tests. Risk factors data of 50 patients was collected and the results obtained showed training accuracy of 96.2% and a validation accuracy of 89% as specified in TABLE IV. With using hybrid data mining techniques we could design more accurate clinical decision support systems for diagnosis of diseases. We can build an intelligent system which could predict the disease using risk factors hence saving cost and time to undergo medical tests and checkups and ensuring that the patient can monitor his health on his own and plan preventive measures and treatment at the early stages of the diseases.

REFERENCES

- [1] "Global atlas on cardiovascular disease prevention and control", WHO, 2011.
- [2] Mozaffarian D, Wilson PW, Kannel WB, Beyond established and novel risk factors: lifestyle risk factors for cardiovascular disease. *Circulation* 117: 3031–3038, 2008.
- [3] Poirier P, Healthy lifestyle: even if you are doing everything right, extra weight carries an excess risk of acute coronary events. *Circulation* 117: 3057–3059, 2008.
- [4] Wood D, De Backer, Prevention of coronary heart disease in clinical practice: recommendations of the Second Joint Task Force of European and other Societies on Coronary Prevention. *Atherosclerosis* 140: 199– 270, 1998.
- [5] Anderson KM, Odell PM, Cardiovascular disease risk profiles. *Am Heart J* 121: 293–298, 1991.
- [6] Anderson KM, Odell PM, Wilson PWF, Kannel WB. Cardiovascular disease risk profiles. *Am Heart J.*, 121: 293–298, 1991.
- [7] Kannel WB, An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol* 110: 281–290, 1979.
- [8] R. Das, I. Turkoglu, and A. Sengur, Effective diagnosis of heart disease through neural networks ensembles, *Expert Systems with Applications*, Elsevier, pp. 7675–7680, 2009.
- [9] T. Porter and B. Green, "Identifying Diabetic Patients: A Data Mining Approach," *Americas Conference on Information Systems*, 2009.
- [10] S. Panzarasa et al, "Data mining techniques for analyzing stroke care processes," in Proc. of the 13th World Congress on Medical Informatics, 2010.
- [11] L. Li, T. H., Z. Wu, J. Gong, M. Gruidl, J. Zou, M. Tockman, and R. A. Clark, "Data mining techniques for cancer detection using serum proteomic profiling," *Artificial Intelligence in Medicine*, Elsevier, 2004.
- [12] V. A. Sitar-Taut et al., "Using machine learning algorithms in cardiovascular disease risk evaluation," *Journal of Applied Computer Science and Mathematics*, 2009.
- [13] K. Srinivas, B. K. Rani, and A. Govrdhan, "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks," *International Journal on Computer Science and Engineering (IJCSE)*, vol. 2, no. 2, pp. 250-255, 2010.
- [14] H. Yan, et al., "Development of a decision support system for heart disease diagnosis using multilayer perceptron," in Proc. of the 2003 International Symposium on, vol. 5, pp. V-709- V-712.
- [15] M. C. Tu, D. Shin, and D. Shin, "Effective Diagnosis of Heart Disease through Bagging Approach," *Biomedical Engineering and Informatics*, IEEE, 2009.
- [16] Hai H.Dam, Hussain A.Abbass and Xin Yao, "Neural – Based Learning Classifier Systems", *IEEE Transactions on Knowledge and Data Engineering*, Vol.20, No.1, pp.26-39, 2008.
- [17] Shantakumar B.Patil and Y.S.Kumaraswamy, "Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network", *European Journal of Scientific Research*, Vol.31, No.4, pp.642-656, 2009.
- [18] Polat , K., S. Sahan, and S. Gunes, Automatic detection of heart disease using an artificial immune recognition system (AIRS) with fuzzy resource allocation mechanism and k-nn (nearest neighbour based weighting preprocessing. *Expert Systems with Applications* 2007. 32 p. 625–631.
- [19] Das, R., I. Turkoglu, and A. Sengur, Effective diagnosis of heart disease through neural networks ensembles. *Expert Systems with Applications*, Elsevier, 2009. 36 (2009): p. 7675–7680.
- [20] Latha Parthiban and R.Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm", *International Journal of Biological and Life Sciences*, Vol.3, No.3, pp.157-160, 2007.
- [21] N. Elfadil and A. Hossen, "Identification of Patients With Congestive Heart Failure Using Different Neural Networks Approaches", *Journal Technology and Health Core*, vol. 17 Issue 4, December 2009.
- [22] Centre for Disease Control and Prevention, http://www.cdc.gov/heartdisease/risk_factors.htm.
- [23] American Heart Association, <http://www.heart.org/HEARTORG/Conditions>
- [24] D. Isern, D. Sanchez, and A. Moreno, "Agents Applied in Health Care: A Review", *International Journal of Medical Informatics*, 79(3), pp.146166, doi:10.1016/j.ijmedinf.2010.01.003, 2010.
- [25] J. G. Yang, S. Y. Weng,, *Applied Textbook of Artificial Neural Network*, Zhejiang University Press, Hangzhou, 2001.
- [26] Y. J. Lei, and X. W. Zhang, *Genetic Algorithm Toolbox of MatLab and its Application*, Xian University of Electronic Science and Technology Press, 2005.
- [27] J. Guo, and W. J. Sun, *Theory of Neural Network and its Implementation with MatLab*, Electronic Industry Press, Beijing, 2005.

Palm Vein Technology

Pooja Trivedi

trivedipooya429@gmail.com

Shikha Yadav

shikhayadav2910@gmail.com

Shivani Yadav

shivaniyadav19970114@gmail.com

Divya Dubey

diadubey14@gmail.com

Megharani Patil

megharani.patil@thakureducation.org

Abstract- Palm vein technology is one of the forthcoming technologies. It is the world's first contactless personal identification system that uses the vein patterns in human palms to confirm a person's identity. It is highly vein as it uses information contained within the body and is also extremely accurate because the pattern of veins in the palm is complex and unique to each individual. It works by capturing a person's vein pattern image while radiating it with near-infrared rays. It detects the structure of the pattern of veins on the palm of the human hand with the utmost precision. The sensor transmits a near-infrared beam towards the palm of the hand and the blood flowing through these back to the heart with reduced oxygen absorbs this radiation, causing the veins to appear as a black pattern. This pattern is recorded by the sensor and is saved in encrypted form in a database, on a token or on a smart card. This technology is designed in such a way that it can only detect the vein pattern of living person. Applications for the device include physical admission into secured areas, log-in to PCs or server systems, access to ATMs and other industry applications.

Keywords- Palm Vein Technology, Infrared beam, Sensor, Biometric, Authentication.

I. INTRODUCTION

In the present network society, where individuals can easily access their information anytime and anywhere, people are also experiencing the risk that others can easily access the same information anytime and

anywhere. Because of this risk, personal identification technology, which can differentiate between registered legitimate users and imposters, is now generating interest. Currently, passwords, Personal Identification Numbers i.e., PIN numbers or identification cards are used for personal ID. However, cards can be stolen, and passwords & numbers can be guessed or forgotten. To solve these problems, biometric authentication technology, which recognizes people by their unique biological information, is attracting attention. In biometric authentication, an account holder's body characteristics or behaviours are registered in a database and is then compared with others who may try to access that account to see if the attempt is legitimate. Fujitsu(Japanese multinational information technology equipment and Services Company) is researching and developing biometric authentication technology focusing on four techniques: faces, fingerprints, voiceprints, and palm veins. Among these four methods, because of its high accuracy and contactless feature palm vein authentication technology is being incorporated into various financial solution products for use in public places.

The Palm Secure sensor developed by Fujitsu is a biometric authentication solution offering maximum levels of security.

II. BACKGROUND

Automated measurement of Physiological and behavioural characteristics to determine or authenticate identity is known as biometrics. Increasing necessity of e-mail, Internet resources,

ATM's, public areas etc. lead to the birth of this feature which offers serious security concerns. The physiological characteristics would be physical human traits like fingerprints, hand shape, eyes, face and veins etc.

Why Biometrics is Important??

- The stress of always remembering the password gets lessen.
- Abuse of stolen id cards and passports will be reduced.
- Abuse of stolen credit cards will be prevented.
- Taking over foreign identity will be impossible.
- Access to any device/computers will not be possible for person without the right of admittance.
- Level of security will grow.

Different Biometric Technologies are:

Physical:

Iris, Retina, Vein Pattern, Face, Fingerprint

Behavioural:

Voice, Keystroke dynamics, Signature Dynamics, Walking Gait.

III. WORKING OF PALM VEIN TECHNOLOGY

Palm vein authentication works on a principle i.e. by comparing the pattern of veins in the palm (which appear as blue lines) of a person who is to be authenticated with a pattern stored in the database. Vascular patterns of each individual are unique and according to Fujitsu research even identical twins have different patterns, and since the vascular patterns exist inside the body, they cannot be ever stolen by any means may it be photography,

voice recording or fingerprints, thereby making this methodology of biometric authentication more secure and accurate than others.

Haemoglobin in the blood is oxygenated when it is in the lungs and carries oxygen to the tissues of the body through the arteries. After the oxygen is released to the tissues, the deoxidized haemoglobin is returned back to the heart through the veins. These two types of haemoglobin have different rates of absorbency. The deoxidized haemoglobin absorbs light at a wavelength of about 760 nm in the near-infrared region and that time when the palm is illuminated with near infrared light, unlike the image seen by the human eye, the deoxidized haemoglobin in the palm veins absorbs this light which reduces the reflection rate and causes the veins to appear as a black pattern. In vein authentication system based on this principle, the region used for authentication is photographed or captured with the help of near-infrared light and the vein pattern is finally extracted through image processing and then registered. The vein pattern of the person to be authenticated is then verified against the preregistered pattern.

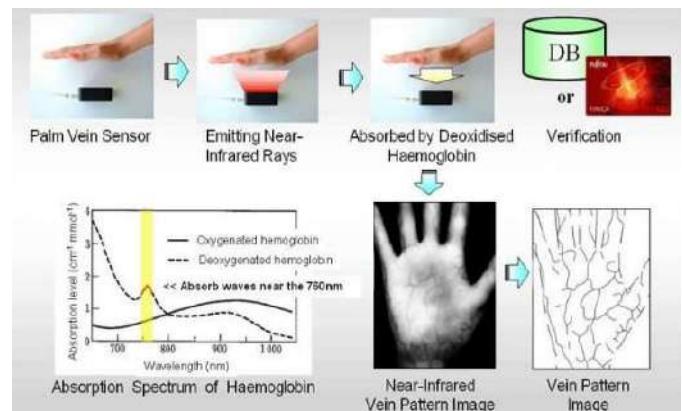


Fig: 1 Working of Palm Vein Technology

REGISTRATION THROUGH PALM VEIN TECHNOLOGY:

Step 1:

- Palm vein Authentication technology consists of a small Palm vein scanner that's easy and natural to use, fast and highly accurate.
- Simply hold your palm near scanner and within a second it reads your vein pattern. A vein picture is taken and palm pattern is registered.

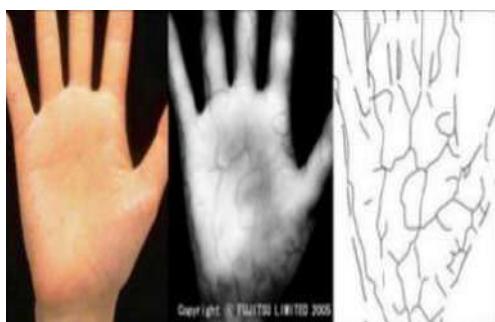


Fig2: Step 1

Step 2:

- The registered palm pattern is stored into the database along with the personal details of the client.



Fig3: Step 2

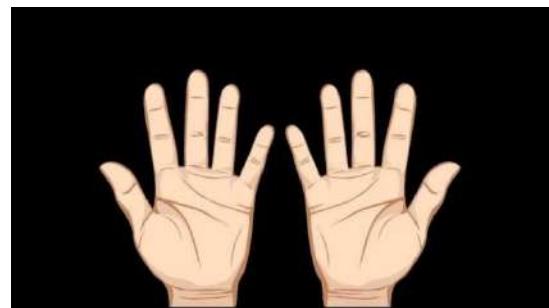


Fig: 4 Registering both Palm

IV. CONCLUSION

Palm vein pattern authentication technology developed by Fujitsu was being used in a wide range in Japan. The Fujitsu PalmSecure is a palm-vein based authentication system that utilizes the latest in Biometric Security Technology. Answering a worldwide need from governments to the private sector, this contactless device offers an easy-to-use, hygienic solution for verifying identity. Palm Veins are identified by comparing palm veins image stored in the database. The future scope of this project is, it can be used using a less expensive infrared camera so that it can be applicable everywhere in public sectors. If this technology is introduced in our country we can solve many problems such as password protection in ATM ,security in various fields and if we implement this technology in government we can make the employees to work accordingly.

With the new palm vein authentication device and with considerable experience in image recognition, Fujitsu expects to be a leader in providing solutions for the biometric security industry. Surely this technology will bring a revolution in the field of science and technology in the near future.

V. REFERENCES

[1] Mansi Manocha and Parminder Kaur, "Palm Vein Recognition for Human Identification Using NN", International Journal of Innovative Technology and Exploring Engineering(IJITEE)

[2] Fujitsu "Palm Vein Pattern Authentication Technology" White paper printed in the U.S.A. Qty. 3M WP10800306

[3] P. Ghosh and R. Dutta, "A new approach towards Biometric Authentication System in Palm Vein Domain"

[4] "Palm Vein Authentication Technology" white paper, Bioguard, Innovative Biometric Solutions,March, 2007.

[5] J. R. Vacca, "Biometric Technologies and Verification Systems", s.l.: Elsevier, 2007.

[6] S. M. E. Hossain, G. Chetty, "Human Identity Verification by Using Physiological and Behavioural Biometric Traits International Journal of Bioscience", *Biochemistry and Bioinformatics*, vol. 1, no. 3, pp. 81-113, September 2011.

[7]R. V. Yampolskiy, V. Govindaraju, "Behavioral Biometrics: a Survey and Classification", *International Journal of Biometrics*, vol. 1, no. 1, pp. 201-205, 2008.

Artificial Intelligence Based Chatbot For E-Commerce

Prakhar Agrawal

Thakur College of Engineering And Technology
Mumbai, India

Dhruvil Parekh

Thakur College of Engineering And Technology
Mumbai, India

Rohan Gupta

Thakur College of Engineering And Technology
Mumbai, India

Ms.Megharani Patil

Project Guide
Thakur College of Engineering And Technology
Mumbai, India

Abstract

Nowadays, businesses are slowly starting to deploy chatbots as a new method of communication with its customers. Due to the subject's infancy and lack of research on the subject, the purpose of this study is to explore the concept of mobile messenger chatbots and an attempt is made to determine the Dutch Millennials' intention to use chatbots as the next interface for mobile commerce. A research model is proposed based on the Technology Acceptance Model (TAM) and Innovation Diffusion Theory (IDT). Data is collected by means of an online survey among 195 participants. The proposed research model is tested by means of simple regression analysis and results are cross-validated using IBM Watson Analytics. All proposed hypotheses are supported. However, there is no unambiguous answer to whether Dutch Millennials have the intention to use chatbots as the next interface for commerce. Nonetheless, more than half of the respondents express a positive first impression towards chatbots. This study knows some limitations regarding external validity and the research model is limited to five independent constructs. Additional constructs or measurement tools could be used to obtain a deeper understanding regarding the subject. Moreover, using a real-life experiment may generate distinctive results. Organizations wanting to deploy messenger chatbots, marketers and chatbot developers should consider compatibility, the consumers' lifestyle and shopping preferences, for a successful implementation. Similarly, the consumers' privacy concerns and resistance to intrusive mobile advertisement are important topics to be considered.

1. INTRODUCTION

Chatbots are software agents that interact with the user in a conversation. The main goal of their creation was to resemble a human being in the way they perform said interaction, trying to make the user think he/she is writing to another human being. This has been implemented with varying degrees of success. One of the most popular languages for the definition of a chatbot knowledge base is artificial Intelligence Markup Language(AIML).This thesis focuses on the implementation of an AIML interpreter written in javascript to allow for a web-based client-side specific usage of AIML chatbots .The interpreter must guarantee the compliance of properly formed AIML documents, perform all the necessary pre-processing duties for the correct usage of the chatbot and ensure the correctness of both pattern matching of user input and chatbot response. The interpreter fully exploits the Document Object Model(DOM) tree manipulation functions of the jQuery library to achieve said goals ,treating AIML files as if they were normal XML files The result is a well performing ,fully functional AIML interpreter tailored around AIML 1.0 specification .A chatbot is software that is used to interact between a computer and a human in natural language. Naturally, it can extend daily life, such as help desk tools, automatic telephone answering systems, to aid in education, business and e-commerce.Have you ever had a problem with a service or a product and ended up contacting the company? Many of us have and chances are that the experience wasn't pleasant. The process can be very frustrating and in most cases, those customer support agents aren't trained very well and have limited knowledge and resources to actually help solve your issue .It's time to move on from those pesky agents! It's 2017 and technology is advancing. In this case, chatbots have started to take on real human support agents. ”.

2. LITERATURE SURVEY

Ecommerce web companies spend a lot of their investment maintaining their call centers. These call centers mainly handle queries of the users, regarding their products. This traditional method costs both time and money to the companies. One of the first and main goals of chatbots has always been to resemble an intelligent human person and make it hard or impossible for the other party of the conversation to understand their real nature (as in artificial). A Chat-bot is a conversational agent that interacts with users using natural language.

There are many applications that are incorporating a human appearance and intending to simulate human dialog, but in most of the cases the knowledge of the conversational bot is stored in a database created by a human experts. However, very few researches have investigated the idea of creating a chat-bot with an artificial character and personality starting from web pages or plain text about a certain person. This paper describes an approach to the idea of identifying the most important facts in texts describing the life (including the personality) of an historical figure for building a conversational agent that could be used in middle-school CSCL scenarios [1]

This paper describes a flexible method of teaching introductory artificial intelligence (AI) using a novel, Java-implemented, simple agent framework developed specifically for the purposes of this course. Although numerous agent frameworks have been proposed in the vast body of literature, none of these available frameworks proved to be simple enough to be used by first-year students of computer science. Hence, the authors set out to create a novel framework that would be suitable for the aims of the course, for the level of computing skills of the intended group of students, and for the size of this group of students. a synthesis of the traditional objectivist approach and a real-world-oriented, constructivist approach to teaching programming to novices. The main aim of implementing such a pedagogy was to engage the students in learning to which they personally relate while attaining intellectual rigor. Classroom experience indicates that students learn more effectively when the traditional objectivist approach is combined with a constructivist approach than when this orthodox approach to teaching programming to novices is used alone [2]

A chatbot aims to make a conversation between both human and machine. The machine has been embedded knowledge to identify the sentences and making a decision itself as response to answer a question. The response principle is matching the input sentence from user. From input sentence, it will be scored to get the similarity of sentences, the higher score obtained the more similar of reference sentences. The sentence similarity calculation in this paper using bigram which divides input sentence as two letters of input sentence. The knowledge of chatbot are stored in the database. The chatbot consists of core and interface that is accessing that core in relational database management systems (RDBMS). The database has been employed as knowledge storage and interpreter has been

employed as stored programs of function and procedure sets for pattern-matching requirement. The interface is standalone which has been built using programing language of Pascal and Java. [3]

Chatbots are software agents that interact with the user in a conversation. The main goal of their creation was to resemble a human being in the way they perform said interaction, trying to make the user think he/she is writing to another human being. This has been implemented with varying degrees of success. One of the most popular languages for the definition of a chatbot knowledge base is artificial Intelligence Markup Language(AIML).This thesis focuses on the implementation of an AIML interpreter written in javascript to allow for a web-based client-side specific usage of AIML chatbots.The interpreter must guarantee the compliance of properly formed AIML documents, perform all the necessary pre-processing duties for the correct usage of the chatbot and ensure the correctness of both pattern matching of user input and chatbot response. The interpreter fully exploits the Document Object Model(DOM) tree manipulation functions of the jQuery library to achieve said goals,treating AIML files as if they were normal XML files The result is a well performing,fully functional AIML interpreter tailored around AIML 1.0 specification.Achatbot is software that is used to interact between a computer and a human in natural language. Naturally, it can extend daily life, such as help desk tools, automatic telephone answering systems, to aid in education, business and e-commerce [4]

Chat bots typically provide a text-based user interface, allowing the user to type commands and receive text as well as text to speech response. Chat bots are usually stateful services, remembering previous commands in order to provide functionality. When chat bot technology is integrated with popular web services it can be utilized securely by an even larger audience. The college enquiry chat bot will be built using artificial algorithms that analyzes user's queries and understand user's message. This System will be a web application [3] which provides answer to the query of the student very effectively. Students just have to put their query to the bot which is used for chatting. The system will use the artificial intelligence algorithms to give appropriate answers to the user. If the answer is found invalid, then some system to declare the answer as invalid can be incorporated. These invalid answers can be deleted or modified by the admin of the system. The student will not have to go to the college for enquiring something. Student can use the chat bot to get the answers to their queries. Students can use this web based system for making enquiries at any point of time. This system may help students to stay updated with the college activities[5]

3. PROPOSED WORK

•AI has reached a stage in which chatbots can have increasingly engaging and human conversations, allowing businesses to leverage the inexpensive and wide-reaching technology to engage with more consumers.

•Chatbots are particularly well suited for mobile — perhaps more so than apps. Messaging is at the heart of the mobile experience, as the rapid adoption of chat apps demonstrates.

•The chatbot ecosystem is already robust, encompassing many different third-party chat bots, native bots, distribution channels, and enabling technology companies.

•We need a trained professionals to answer the queries on the answering side on the other side this system understands the queries and answer accordingly, which saves a lot of man-power

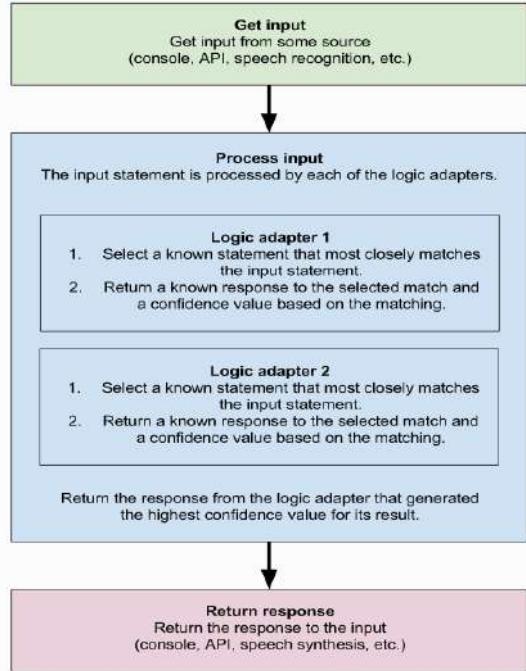
3.1 FEASIBILITY STUDY (FAD)

As businesses continue to automate their sales and customer support services in order to reduce costs, chatbots and other intuitive artificial intelligence programs are becoming more popular.

Chatbots are capable of helping businesses significantly cut labor costs. While complete automation of the customer service workforce is not feasible, automating a portion of the customer management and sales positions in the US is possible through chatbots and other automation technology, and this would result in considerable savings for businesses.

Many companies have invested more resources in better training for their customer service representatives, but the recent wave of artificial intelligence (AI) programming is changing the game, mainly through chatbots.

3.2 SOFTWARE ARCHITECTURE



A block diagram will show how the system of chatbot system will function. It shows the how the user asks a query then it is processed in 2 pass processor. In the first pass the keywords are found and in the second pass the sentence is analyzed.

3.3 Algorithm

Step 1: Start.

Step 2: Get the user query. (INPUT)

Step 3: Pre-processing of the query

Step 4: Fetch the remaining only keywords from the query

Step 5: Match the fetched keywords with the keywords in Knowledge base, and provide an appropriate response. The keywords will be matched with the help of keyword matching algorithm.

Step 6: Return the query response as an output to the user.

Step 7: Exit.

4 EXPECTED OUTCOME

For many online retailers business owners, who are not able to manage many customer conversations at a time generally, outsource their customer support to a call center. But if the basic customer support conversation is handled by their Bot won't that be more productive and cost-effective? Smaller online businesses are generally run by few people who are thankful for automation that can help them manage several conversations with their customers at a time.

5 CONCLUSIONS

The main objective of the project is to develop an algorithm that will be used to identify answers related to user submitted questions. The need is to develop a database where all the related data will be stored and to develop a web interface. The web interface developed will have two parts, one for simple users and one for the administrator.

A background research took place, which included an overview of the conversation procedure and any relevant chat bots available. A database will be developed, which will store information about questions, answers, keywords, logs and feedback messages. A usable system will be designed, developed and deployed to the web server.

6 REFERENCES

- [1] Emanuela Haller and Traian Rebedea, "Designing a Chat-bot that Simulates an Historical Figure", IEEE Conference Publications, July 2013
- [2] Maja Pantic, Reinier Zwitserloot, and Robbert Jan Grootjans, "Teaching Introductory Artificial Intelligence Using Asimple Agent Framework", IEEE Transactions On Education, Vol. 48, No. 3, August 2005.
- [3] S. J. du Preez, M. Lall and S. Sinha, "An intelligent web-based voice chat bot,"EUROCON 2009, EUROCON '09. IEEE, St.-Petersburg, 2009.
- [4] Y. Chen, W. Wang and Z. Liu, "Keyword-based search and exploration on databases," 2011 IEEE 27th International Conference on Data Engineering, Hannover, 2011.
- [5] Bayu Setiaji & Ferry Wahyu Wibowo "Chatbot Using A Knowledge in Database"2016 7th International Conference on Intelligent Systems, Modelling and Simulation.
- [6] Artificial Intelligence: A Modern Approach(By: Stuart Russell, Peter Norvig)
- [7] Introduction to Artificial Intelligence(By: Philip C. Jackson)
- [8] Designing Voice User Interfaces: Principles of Conversational Experiences By Cathy Pearl
- [9] Bot Business 101: How to start, run & grow your Bot / AI business By Ekim Kaya

ANALYTICS GOVERNANCE

...because just Data Governance is not enough

Kirti R Sharma
Tech. Consultant
Deloitte Consulting LLP
Austin, Texas, USA
kirti.asu@gmail.com

+1 315-944-7591Karan
R Sharma
Tech. Consultant
Deloitte Consulting LLP
Mumbai, India
karak92@gmail.com

+91 9867177490Suraj G
Singh
Data Scientist
Hill Country Holdings
Austin, Texas, USA
singhsurajg@gmail.com

+1 315-944-7625Rashmi
Thakur
Assistant Professor
Thakur College of Engnn. & Tech.
Mumbai, India
rashmi.thakur@thakureducation.org
+91 9869066194

Abstract

Analytics today is not just a buzz, trend or a fad; it is a revolution changing each & every business & taking the competition to the next level. Today, a financial service provider knows which customer is most likely to default this month, ride share apps know when & where you might need your next ride and you get a furniture discount coupon in your mail box because the retailers know what you have been looking for. Data has been the blood that has fueled the next Industrial revolution. In this Data Centric world, IT and business groups have to make decisions on what to do with the data used by, and created by, technology. This data is the driving force, which determines the success or failure of various Machine Learning, AI & business applications used for decision making by the organizations. All this had led to the need for a data governance framework – an overarching approach to how you collect, manage and archive data in your enterprise. This enormous governed data is processed and used to implement various analytical functions. Hence, a good data governance system is incomplete without a good analytics governance model. Analytics governance needs to become imperative to ensure that the data used, delivers exactly what is needed by creating value for the customers and complying with legal, confidential and quality issues. An Analytics team should be able to deliver valuable, sustainable and harmless analytical solutions.

In this paper, we discuss the need for a governance framework, explore the existing models, propose a model to integrate analytics and data governance

at an organization, with a focus on moving from project level governance to organization level governance and emphasize on the value it will create for consumers and organizations.

Keywords

Governance, Data, Analytics, Framework, Data Management, Organizations, Big Data, Sustainable Analytical Solutions, Chief Analytics Officer

I. Introduction

Governance in Analytics means laying down guidelines and procedures to support the development, deployment and maintenance of analytics projects. Organizations have started realizing the value of their data and are advancing towards implementing advanced analytical solutions. However, the processes followed to develop and deploy these solutions are still limited to just the project level. Hence, every time a new project comes into inception, a new process may be put into place. There is no standardization of execution and ownership, leading to highly volatile end products. For development of valuable and sustainable solutions, it is critical that an organization establishes an org-wide analytics governance framework at the earliest.

Many organizations today, are defining the way they govern their data. They are working towards laying a strong foundation to acquire relevant data, utilize it effectively, maintain the quality of data and most importantly, define its ownership and access

authorities. The need of the hour is to take this effort further and incorporate an analytics governance framework with the existing data governance model. This will not only help organizations to standardize the way they implement advanced analytics but also will optimize the time spent on developing solutions and create greater value by ensuring that the solutions are sustainable, measurable and repeatable.

II. Need for Analytics Governance

The need for analytics governance is strongly illustrated when evaluating the global prescriptive analytics market that is estimated to grow at a compounded annual growth rate of 31.7% from USD 1.16 billion in 2016 to USD 4.58 Billion in 2021 [5]. Enterprises, who are adhering to an analytics governance framework are going to be better positioned to assess present and future impacts such as this, and be better situated to take advantage of increasing analytics market, grow profits and mitigate risks.

The various departments in many organizations even today work in silos. Data governance surely resolves the challenges with data sourcing and management. However, it does not help with the isolated implementation of analytical projects, causing inconsistent, inflexible and non-repeatable solutions.

Governance is the critical last mile for analytics, explains Ryan Schmiedl, Vice President of Product Management at SAS. He adds, “It’s not enough to run analytics, get a decision and you’re done. There’s more to it. You put all the time into data preparation and model building, you need to make sure the resulting analysis is something you can complete and repeat [11].”

Hence, even with a data governance structure in place, there is still a need for placing regulatory policies and procedures for analytical processes. Various industries are all on the same journey, trying to overcome the challenges and obstacles in their way of emphasizing decision-making based on insights derived from data. With a structured analytics framework which resides centrally and spans across all departments in the organization, the way data is sourced from various departments, utilized to build efficient analytical solutions, delivers value to customers, distributed to third parties and secured

from unhealthy exploitation can all be controlled under one umbrella.

Data and analytics governance go hand in hand. The existence of one without the other is meaningless to ensure the intent of moving into the big data industry and exploiting the true power of the enormous data that organizations possess. No-one wants to wake up in the morning realizing that the reason they were depressed last week was because Facebook manipulated their emotions with an excuse of a research [12]. We don’t want to discover that our credit card has been blocked suspecting a fraud when we are shopping for our brother’s wedding, or worse, we don’t want to find out that we have been denied a credit card because the bank used other sources to determine our race and/or nationality and based its decision on that. Therefore, it is high time that analytics governance be deployed in this fast pacing digital industry where analytics is changing the face of businesses and consumer services.

III. Evolution of Governance

Traditionally, the governance process has been siloed at the department level of an organization with high autonomy, redundant designs and processes and no sharing of knowledge or skills. More recently, a hybrid approach has become popular which centralizes the design and support teams. This approach ensures communication between the business and implementation teams of various departments, thus standardizing the development processes. However, it has now become essential to adopt a centralized governance framework with one design and business team each, including representation from all departments and a shared development, maintenance and operations team. The implementation teams here, are spaced out at the department level but still maintain a cadence of communication. This framework highly supports the dynamic nature of analytics, allowing organizations to foster a governed agile delivery environment.

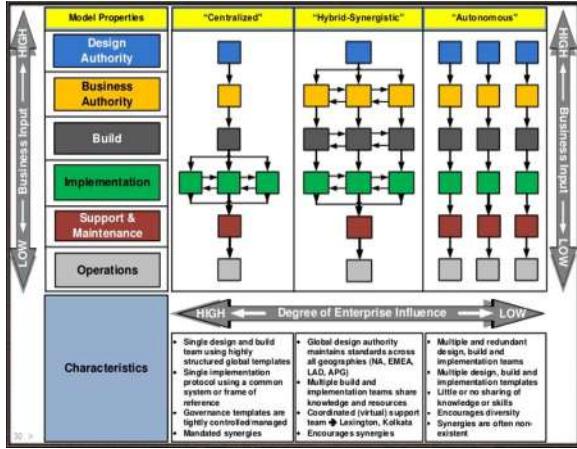


Figure 1: Evolution of Governance Frameworks

IV. Proposed Model

An effective analytics governance framework is contingent upon the coexistence of a data governance framework and its integration into the former. Our proposal of a good analytics governance framework is as follows:

1. Define the business problem or use case
2. Scope out the problem to various business units in the organization to evaluate the interaction of these units and how they might be affected
3. Define security, privacy, legal & other elements for the data used in the analytical solution that is planned to be built
4. Develop analytical solution with the 6 phases of data governance model
 - i. Cultural Tone of “Data Driven”
 - ii. Access to Data
 - iii. Stewardship of Data
 - iv. Quality of Data
 - v. Utilization of Data
 - vi. Acquisition of Data
 - vii. Overall Master Data Management

5. Track and Coordinate the progress of the implementation with a centralized delivery team
6. Implement & assess the effectiveness of the results from your analytical solution and ensure the delivery of a harmless, unoffending solution
7. Update the model timely, with new information and logic to keep it from becoming obsolete

Currently, there are many organizations that do not even have a good data governance model in place, let alone an analytics governance model. Hence, we believe that there is not just a need for an analytics governance but also a need to integrate the data and analytics governance into a single framework to expedite the transformation process and ensure a more robust organization wide model.

Our proposed integrated governance framework is shown below in figure 2.



Figure 2: Integrating Data and Analytics Governance

Adopting this integrated governance framework will increase operational efficiency, giving the Analysts an opportunity to utilize their time more effectively in deriving insights from well governed data, as well as Data Scientists, the bandwidth to build robust analytical models and improvise the same with a well governed analytics process.

The centralized governance team should ideally comprise of representatives from every business unit in the organization, along with representatives from the Data and the IT teams. Hierarchically, the team may reside just below the executive leadership.

V. Proposed Industry Applications

The proposed integrated model is general and hence can be applied to various industries and use-cases. Whether, retail or finance or healthcare, the right governance model will help to fully leverage the

power of data and analytics. In the absence of governance, it is not possible to completely harness the value of the enormous data that organizations are collecting. Also, it may create redundant solutions and it constraints the process of building re-usable, flexible and scalable analytical solutions.

When there is no data and analytics governance, every business unit in an organization may end up collecting data as per their requirements, formatting as per their needs and updating values a per their defined processes. There could also be non-uniformity in the way business units or business people store their data – for example a salesperson stores prospect data in an excel sheet which no one from the marketing department is aware of, thus rendering that data useless for any analysis or for analytical models. This creates highly inconsistent data and prevents from having a single version of truth. Moreover, if ownership is not defined and everyone and anyone can update the data as needed, the data loses its authenticity. Because of multiple copies of the same data being present across business units, analytical solutions will also get built in siloes, drastically increasing the time, cost and resources required for development. Hence, to derive the most value from data and analytics, it is imperative that various industries start focusing on deploying an integrated governance model.

VI. Use Cases

In this section, we discuss some use-cases across industries where our proposed model can be leveraged.

A. Retail Industry

The retail industry today is flooded with a sea of consumer data. Companies are investing millions of dollars to understand their customers from this rich data source & plan their products & its delivery accordingly. Big analytical models are driving businesses to make critical decisions on Sales, Marketing, and Merchandising etc. Most of these complex models involve parameters, which are widely used across all business fields & are highly co-related. Making two big data models for a problem, which may be solved by one, is not just resource consuming but also may hide few important relationships between these parameters. This is

where analytics governance can be highly resourceful.

For example, consider a retail department of a company developing a model for analyzing sales data to forecast inventory requirements of products & the merchandising department of the same company working to create a model capable to select the right product brand & vendor to stock up the inventory. When we see the problem together, we definitely see the possibility of a relationship; however not following an analytical governance approach may result in more time & resource investment.

If we follow the suggested approach, we can scope the sales prediction model to include parameters like product sub category, brand or vendor to understand the demand for products for each of them. This can enable the merchandising team to rank their collection of products based on customer ratings. Although this does not always mean that we will have a relationship between two models, but it is worth the time to invest. In most cases, the steps of data collection, data transformation & some preprocessing of data can be utilized in subsequent models to speed up the model building process even if they do not utilize the same or similar model.

B. Telecom Industry

If a telecom company has a good data governance but their analytics is still not governed, the various business departments could end up creating redundant analytical models. For example, the marketing department of the company may build a predictive model to determine a current customer as a potential customer for additional services. At the same time, the customer retention department may develop their model that determines the same customer to potentially churn.

Without a governed analytics model, the customer may actually give up the existing services due to constant offers from the marketing team, perceiving the same as harassment. Further, if both the teams present their findings to the executive leadership, with both reports having same customers identified as potential customers for additional services and potential churn customers, it is a great risk for the leadership to plan their business strategy.

If the company instead deploys our proposed integrated governance model, all the decisions, designs and results of their analytical models will be

in sync, delivering higher business value and conserving resources, time and cost. Our model ensures that business problems are scoped out centrally, data elements are considered in consensus, roles and responsibilities are well defined and most importantly, results are coordinated.

The model will help the business units to discover other important data elements that they might not have even thought of. For example, the marketing department may discover that the call center data for the customers should also be used to differentiate harassed customers.

C. Healthcare

Analytics & data science have recently brought about some rapid developments in the ever-change reluctant healthcare sector. In an industry where time & money have the biggest stakes, analytics has succeeded to reduce cost & increase efficiency of test results thereby minimizing time wastage. However, we cannot ignore the risk of data security that threatens to have big consequences. For example, misuse of customer health records for promotions by third party vendors. A constructive data & analytics governance structure provides the patient ownership & rights to secure their information and choose what information they would like to share.

VII. Value

The integration of analytics and data governance will most probably be a little slow paced process, with various consensual decisions to be made, architecture to be designed and most importantly, the new framework to be adopted. But, it will surely be worth every minute, penny and resource that was invested.

The value in implementing an analytics governance framework that is as fluid and dynamic as the data available for analysis, is that it provides for analytical agility relative to need and requirement. Enterprises will be able to deploy analytical strategies in real time, across the organization, all the while monitoring and improving their effectiveness, transitioning analytical resources as required and efficiently managing, controlling and mitigating regulatory issues and risks.

Moreover, the integration of the two governance frameworks will not just optimize the value of data, but also will make sure that the analytics processes are deployed and secured as efficiently as possible.

VIII. Conclusion

In Toto, we don't just need data management, we also need analytics management.

With data analytics in full swing, it is high time that organizations soon appoint a Chief Analytics Officer (CAO) to govern their analytical processes and to employ data as an actionable asset. The Chief Data Officer and the CAO can then work together to lay down a robust framework for governance. A good governance model will help manage your data to leverage it more effectively, support informed decision making, develop sustainable, harmless solutions and accelerate delivery and maintenance of the solutions.

It is important to understand that although it sounds simple, data & analytics governance is more complex than what we think. Data governance is a continual process that involves many detailed discussions & cross-functional information & data & process flows. Hence, even though it is not as interesting as recent analytics advancements like Deep learning, Natural language processing or IOT, data governance combined with analytics governance is bound to play a pivotal role as a solution to almost all organizational data problems of information retrieval, security & integrity.

Acknowledgement

We would like to extend our warm gratitude to our family for their patience and motivational words throughout the development of the paper. Last but not the least; we thank God for his spiritual bliss on us.

References

- [1] <https://erwin.com/blog/top-6-benefits-of-data-governance/>
- [2] <https://www protiviti com/US-en/insights/data-governance-combined-analytics-powerful-business-insights>
- [3] <https://www.cio.com/article/3235706/leadership-management/chief-analytics-officer.html>

- [4] <https://www.forbes.com/sites/adigaskell/2016/10/21/do-organizations-need-a-chief-analytics-officer/#3cd6be13ee8d>
- [5] <https://www.marketsandmarkets.com/PressReleases/prescriptive-analytics.asp>
- [6] Becoming an analytics driven organization to create value, EY publication in collaboration with Nimbus Ninety.
- [7] The science of winning in financial services, EY publication.
- [8] Exploring Next Generation Financial Services: The Big Data Revolution, May 2017.
- [9] Voices, David Pope on SAS. "The key to solving problems is to use ppt." SAS Voices. January 31, 2013. Accessed April 24, 2017. <http://blogs.sas.com/content/sascom/2013/01/31/the-key-to-solving-problems-is-to-use-ppt/>.
- [10] Occams Business Research & Consulting Pvt. Ltd. "Global Prescriptive Analytics Market is Anticipated to Grow at a CAGR of 31.7% by 2023 - Occams Business Research & Consulting." PR Newswire: news distribution, targeting and monitoring. April 04, 2017. Accessed April 24, 2017. <http://www.prnewswire.com/news-releases/global-prescriptive-analytics-market-is-anticipated-to-grow-at-a-cagr-of-317-by-2023---occams-business-research--consulting-618178763.html>.
- [11]"A framework for analytics governance." SAS. N.p., n.d. Web. 30 Apr. 2017. https://www.sas.com/en_us/insights/articles/analytics/a-framework-for-analytics-governance.html
- [12]Manager, Josh Parenteau Product. "The 5 Phases of Today's Analytic Workflow." CIO. N.p., 06 Mar. 2017. Web. 30 Apr. 2017. <http://www.cio.com/article/3177345/analytics/the-5-phases-of-today-s-analytic-workflow.html>
- [13]"6 Stages of an Effective Governance Model." Stratigent. N.p., 04 May 2015. Web. 30 Apr. 2017. <http://www.stratigent.com/community/analytics-insights-blog/6-stages-effective-governance-model>
- [14]Chandler, Neil, Bill Hostmann, Nigel Rayner, and Gareth Herschel. "Gartner's Business Analytics Framework." Gartner's Business Analytics Framework (n.d.): n. pag. 20 Sept. 2011. Web.
- [15]Simonson, Eric, Anupam Jain, and Rajesh Ranjan. "Analytics in Banking." Www.genpact.com. N.p., n.d. Web. <<http://www.genpact.com/docs/default-source/resource-/analytics-in-banking>>.
- [16]<https://www.healthcatalyst.com/demystifying-healthcare-data-governance>
- [17]<http://www.infinitive.com/2017/05/05/5-reasons-why-data-governance-matters-to-data-driven-marketers/>

FAULT TOLERANCE IN GRID COMPUTING

Ms.Tanushree
Bhattacharjee

tanu.b.4488@gmail.com

Harsheen Kaur
SE CMPN A
Mumbai, India
harsheenkaur23@mail.com

Anjali Thomas
SE CMPN A
Mumbai, India
thomasanjali07@gmail.com

Kinjal Bhatia
SE CMPN A
Mumbai, India
bhatiakinjal98@gmail.com

Mahima Choudhary
SE CMPN A
Mumbai, India.
mahima.choudhary10@gmail.com

I. ABSTRACT— THE BASIC IDEA OF GRID COMPUTING IS TO CREATE LARGE AND POWERFUL VIRTUAL COMPUTERS WHICH IS A COLLECTION OF HETEROGENEOUS, DISTRIBUTED ENVIRONMENT. GRID COMPUTING IS BECOMING A MAINSTREAM TECHNOLOGY FOR LARGE SCALE DISTRIBUTED RESOURCE SHARING AND SYSTEM INTEGRATION. GRID APPLICATIONS OFTEN INVOLVE LARGE AMOUNT OF DATA AND/OR COMPUTING RESOURCES THAT REQUIRE SECURE RESOURCE SHARING ACROSS ORGANIZATIONAL BOUNDARIES. TODAY, HIGHLY SECURE OR VIRTUAL GRID IS VERY DEMANDING IN WHICH YOU CAN SHARE ANY RESOURCE FROM ANY CLUSTER EVEN WITH EXISTENCE OF FAULT IN SYSTEM. THIS PAPER GIVES A METHOD TO IMPROVE THE RESOURCE UTILIZATION WITH MAXIMUM EFFICIENCY AND THROUGHPUT EVEN IN OCCURRENCE OF FAULT IN SYSTEM. IT ALSO INCREASES THE THROUGHPUT OF SYSTEM BY SIMULTANEOUS WORK OF LOG ENTRY BY CHECK POINTING APPROACH AND EXECUTION OF JOB OR BY REDUCING THE TIME.

Keywords- *Grid computing, fault tolerance, check pointing, throughput*

II. INTRODUCTION

The scientific communities were starting to look seriously at Grid computing as a solution to resource federation problems. For example, high energy physicists designing the Large Hadron Collider (LHC) realized that they needed to federate computing systems at hundreds of sites if they were to analyze the many petabytes of data to be produced by LHC experiments. The Grid: Blueprint for a New Computing Infrastructure

also had a catalyzing effect. Grid era in certain possibilities that transcend simply bigger, faster, and better. Grid computing was started long back ago, at a time when the application portability remained a major challenge, that is when many of the processor architectures competed for the dominance.

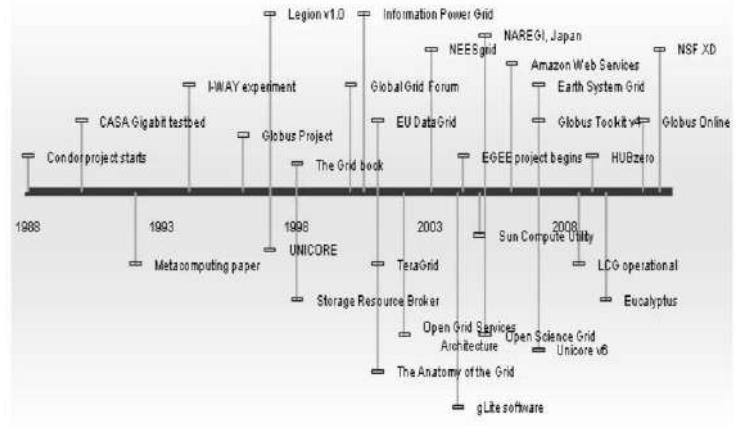


Figure 1: Abbreviated Grid timeline, showing 30 representative events during the period 1988–2011

GRID COMPUTING:

Grid computing is a collection of computers from multiple locations to reach a common goal. The grid can be thought of as a distributed system with non-interactive workloads that involve a large number of files. Grid computing is a special type of parallel computing that depends on complete computers connected to a network (private, public or the internet) by network producing hardware, compared to the lower efficiency of designing and constructing a small number of custom supercomputers.

In grid computing, the computers on the same network can work on a task together, thus function as a supercomputer. Typically, a grid works on various tasks within a network, but it is also capable of working on specialized applications. It is designed to solve problems that are too big for a supercomputer while maintaining the ability to process smaller problems.

A grid is connected by parallel nodes that form a computer cluster, which runs on an operating system. The cluster varies in size from a small work station to several networks. The technology is applied to a range of applications, like mathematical, scientific or educational tasks through several computing resources.

ARCHITECHTURE OF GRID COMPUTING:

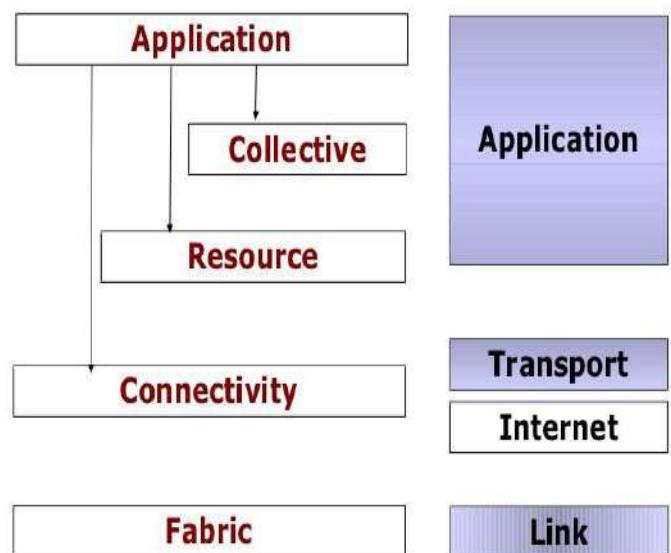
Protocols and services at five different layers as identified in the Grid protocol architecture are provided by Grids. In general, the higher layers are focussed on the user whereas the lower layers are more focussed on computers and networks.

At the Fabric layer, access to different resource types such as compute, storage and network resource, code repository, is provided by the grid. Grids usually rely on existing fabric components, for instance, local resource managers. As result of sharing operations at higher levels, fabric components implement the local, resource-specific operations that occur on specific resources (whether physical or logical). Richer Fabric functionality enables more sophisticated sharing operations.

Connectivity layer defines core communication and authentication protocols for easy and secure network transactions. The GSI (Grid Security Infrastructure) protocol underlies every Grid transaction. Communication protocols enable the exchange of data between Fabric layer resources. Authentication protocols build on communication services to provide cryptographically secure mechanisms for verifying the identity of users and resources.

The Resource layer defines functions for the publication, discovery, negotiation, monitoring, accounting and payment of sharing operations on individual resources. The Resource layer builds on Connectivity layer communication and authentication protocols to define protocols for the secure negotiation, initiation, monitoring, control, accounting, and payment of sharing operations on individual resources. When Resource layer implements these protocols, Fabric layer functions to access and control local resources is called.

Grid Architecture (Layered)



The Collective layer -While the Resource layer is focused on interactions with a single resource, the next layer in the architecture contains protocols and services (and APIs and SDKs) that are not associated with any one specific resource but rather are global in nature and capture interactions across collections of resources. For this reason, we refer to the next layer of the architecture as the *Collective* layer

The Application Layer is the highest layer of the structure. By calling upon services defined at any layer applications are constructed. At each layer, we have well-defined protocols that provide access to some useful service: resource management, data access, resource discovery, and so forth. At each layer, APIs may also be defined whose implementation exchange protocol messages with the appropriate service(s) to perform desired actions.

III. ISSUES

Since grid environments are extremely heterogeneous and dynamic, with components joining and leaving the system all the time, more faults are likely to occur in grid environment.

1. As and when the fault occurs at a grid resource which eventually results in failing to satisfy the user's deadline,

the job is rescheduled on another resource. As the job is re-executed, it consumes more time.

2. There are resources that fulfill the criteria of deadline constraint, but they have a tendency towards faults in computational based grid environments. The grid scheduler selects the same resource for mere reason that grid resource promises to meet user's requirements of grid jobs. This eventually results in compromising user's QOS parameters in order to complete the job.

3. Even if there is a fault in the system, a task running needs to be finished on its deadline. There is no meaning of a task which is not completed before its deadline. Hence, deadline is the major issue in real time.

4. It is about the ability to handle the growing amount of work, and the capability of a system to increase total throughput under an increased load when resources are added. Hence, fault tolerance in grid computing is important as the dependability of the grid resources may not be guaranteed. It is needed to enable the grid to continue its work when one or more resources fail. Hence, a fault tolerant system must be included to detect errors and recover them from them and thus avoiding the failure of the grid.

IV. REMEDIES

Job replication and job check pointing are the two often used techniques to accomplish fault tolerance in grid computing.

Job replication:

Job replication is based on the assumption that the probability of single resource failure is much higher than of a simultaneous failure of multiple resources. It copies the same job on different resources with redundant copies of a job, the grid can continue to provide a service in spite of failure of grid resource carrying out job copies without affecting the performance. Job replication is the method of replicating job on multiple servers such as in grid computing service is capable of receiving jobs, executing them, performing checksum operations on them, and sending the result back to the client.

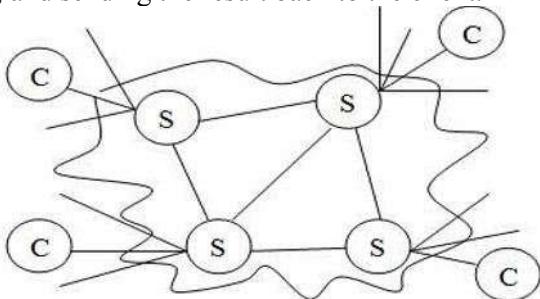


Fig 1.Distributed system with multiple clients and server

Data Replication is commonly used to enhance availability in Grid like environments where failures are more likely to occur. Components are replicated on different machines, and if any component or machine fail, then that application can be transferred and run on another machine having the required components. The main disadvantage of job replication technique is the additional resources used in executing the same job. This can cause grid over provisioning and can lead to great delays for other jobs waiting these resources to become free. Also, most of the existing replication based techniques are static. This means that the number of replicas of the original job is decided before execution and it is fixed number. Static job replication leads to excessive utilization of resources and also to excess load on the grid.

On the other hand, adaptive job replication can alleviate this extra load resulting from using fixed number of replica. Adaptive job replication techniques determine the number of replica according to the failure history of the primary resource allocated to execute the job. Thus, the number of replica will be different for each job according to the failure behavior of each resource in the past. Bad failure history means big number of replica and good failure history means small number of replica Job check-pointing:

Check-pointing is the ability to save the state of a running job to a stable storage. In case of any fault, this saved state can be used to resume execution of the application from the point in computation where the checkpoint was last registered instead of restarting the application from its very beginning. This can reduce the execution time to a large extent. Each interval starts when a checkpoint is established and ends when next checkpoint is established. A short check pointing interval leads to a large number of redundant checkpoints, which delay job processing by consuming computational and network resources. On the other hand, when a check pointing interval is too long, a substantial amount of work has to be redone in case of a resource failure. So, calculating the optimal length of a check pointing interval represents the main challenge when using this check pointing. Hence, the decision about the size of the check pointing interval and the check pointing technique is a complicated task and should be based upon the knowledge about the application as well as the system.

The efficiency of checkpoint is based on:

1. Checkpoint overhead in terms of time and resources consumed.
2. Checkpoint length plays a major role.
3. Compatibility and portability of checkpoints.

Various types of check pointing optimization have been considered by the researchers, e.g., Full check pointing, Incremental check pointing, Unconditional check pointing, dynamic check pointing, Synchronous and asynchronous check pointing etc. A check point may be system level, application level, or mixed level depends on its characteristics. Check-pointing is also categorized on the basis of In-transit or orphan message. These are Uncoordinated Check pointing, Coordinated Check-pointing, and Communication-induced Check-pointing. Check-pointing also can be classified is based on who instruments the application that do the actual capturing and re-establishing of the application execution state. These are Manual code insertion, Pre-compiler check pointing, Post-compiler check-pointing. A check point may be local or global on the basis of their scope. Check-point for separate process is local checkpoint and a check-point applied for set of processes is called global check point. Check-pointing have some demerits such as Check pointing causes execution time overhead even if there are no crashes.

V.RESULT AND DISCUSSION

The response time of a check pointing technique is not good compared with the job replication technique. This is due to the extra time needed to migrate the job to another resource when a resource fails. On the other hand, job replication technique does not need to migrate jobs between resources and the first returned response is employed. The required networking and computing resources of job replication techniques are much higher than those of check pointing techniques. Check pointing has another cost when writing checkpoint data to stable storage whenever a checkpoint is taken. This cost is

proportional to the size of the checkpoint data. Thus, we can use check pointing strategy for the resources constrained grids and job replication technique for real time applications. However, determination of the number of replica and the number and intervals of checkpoints are still big challenges.

VI. CONCLUSION

In all distributed environments fault tolerance is an important problem. Thus, by dynamically adapting the checkpoint frequency and the optimal number of replicas, based on history of information of failure and job execution time, which reduces checkpoint overhead and also, increases the throughput by which the proposed work achieves fault tolerance. Hence, following have been proposed new fault detection methods, client transparent fault tolerance architecture, on demand fault tolerant techniques, economic fault tolerant model, optimal failure prediction system, multiple faults tolerant model and self adaptive fault tolerance framework to make the grid environment is more dependable and trustworthy.

References

- <https://www.irjet.net/archives/V2/i9/IRJET-V2I9140.pdf>
- <https://pdfs.semanticscholar.org/12a9/e59170f0e67a62ff19e766a56f299d3d7378.pdf>
- <http://airccse.org/journal/ijcses/papers/0211cses07.pdf>
- <https://www.ijedr.org/papers/IJEDR1301011.pdf>

Research Paper on Miniature autonomous vehicle

Gautam Sharma, Monark Unadkat and Riya Sharma

**gautamsharma7234@gmail.com, monarkunadkat12@gmail.com,
riyasharmacs10@gmail.com**

anand.khandare@thakureducation.org

Abstract

We trained a convolutional neural network (CNN) to map raw pixels from a single front-facing camera directly to steering commands. This end-to-end approach proved surprisingly powerful. With minimum training data from humans the system learns to drive in traffic on local roads with or without lane markings and on highways. It also operates in areas with unclear visual guidance such as in parking lots and on unpaved roads.

The system automatically learns internal representations of the necessary processing steps such as detecting useful road features with only the human steering angle as the training signal. We never explicitly trained it to detect, for example, the outline of roads.

Compared to explicit decomposition of the problem, such as lane marking detection, path planning, and control, our end-to-end system optimizes all processing steps simultaneously. We argue that this will eventually lead to better performance and smaller systems. Better performance will result because the internal components self-optimize to maximize overall system performance, instead of optimizing human-selected intermediate criteria, e. g., lane detection. Such criteria

understandably are selected for ease of human interpretation which doesn't automatically guarantee maximum system performance. Smaller networks are possible because the system learns to solve the problem with the minimal number of processing steps.

1 Introduction

CNNs have revolutionized pattern recognition . Prior to the widespread adoption of CNNs, most pattern recognition tasks were performed using an initial stage of hand-crafted feature extraction followed by a classifier. The breakthrough of CNNs is that features are learned automatically from training examples. The CNN approach is especially powerful in image recognition tasks because the convolution operation captures the 2D nature of images. Also, by using the convolution kernels to scan an entire image, relatively few parameters need to be learned compared to the total number of operations.

While CNNs with learned features have been in commercial use for over twenty years , their adoption has exploded in the last few years because of two recent developments. First, large, labeled data sets such as the Large Scale Visual Recognition Challenge (ILSVRC) have become available for training and

validation. Second, CNN learning algorithms have been implemented on the massively parallel graphics processing units (GPUs) which tremendously accelerate learning and inference.

3 Data Collection

Training data was collected by driving on a wide variety of roads and in a diverse set of lighting and weather conditions. Most road data was collected in central New Jersey, although highway data was also collected from Illinois, Michigan, Pennsylvania, and New York. Other road types include two-lane roads (with and without lane markings), residential roads with parked cars, tunnels, and unpaved roads. Data was collected in clear, cloudy, foggy, snowy, and rainy weather, both day and night. In some instances, the sun was low in the sky, resulting in glare reflecting from the road surface and scattering from the windshield. Data was acquired using either our drive-by-wire test vehicle, which is a 2016 Lincoln MKZ, or using a 2013 Ford Focus with cameras placed in similar positions to those in the Lincoln. The

system has no dependencies on any particular vehicle make or model.

4 Network Architecture

We train the weights of our network to minimize the mean squared error between the steering command output by the network and the command of either the human driver, or the adjusted steering command for off-center and rotated images (see Section 5.2). Our network architecture is shown in Figure 4. The network consists of 9 layers, including a normalization layer, 5 convolutional layers and 3 fully connected layers. The input image is split into YUV planes and passed to the network. The first layer of the network performs image normalization. The normalizer is hard-coded and is not adjusted in the learning process. Performing normalization in the network allows the normalization scheme to be altered with the network architecture and to be accelerated via GPU processing. The convolutional layers were designed to perform feature extraction and were chosen empirically

through a series of experiments that varied layer configurations. We use strided convolutions in the

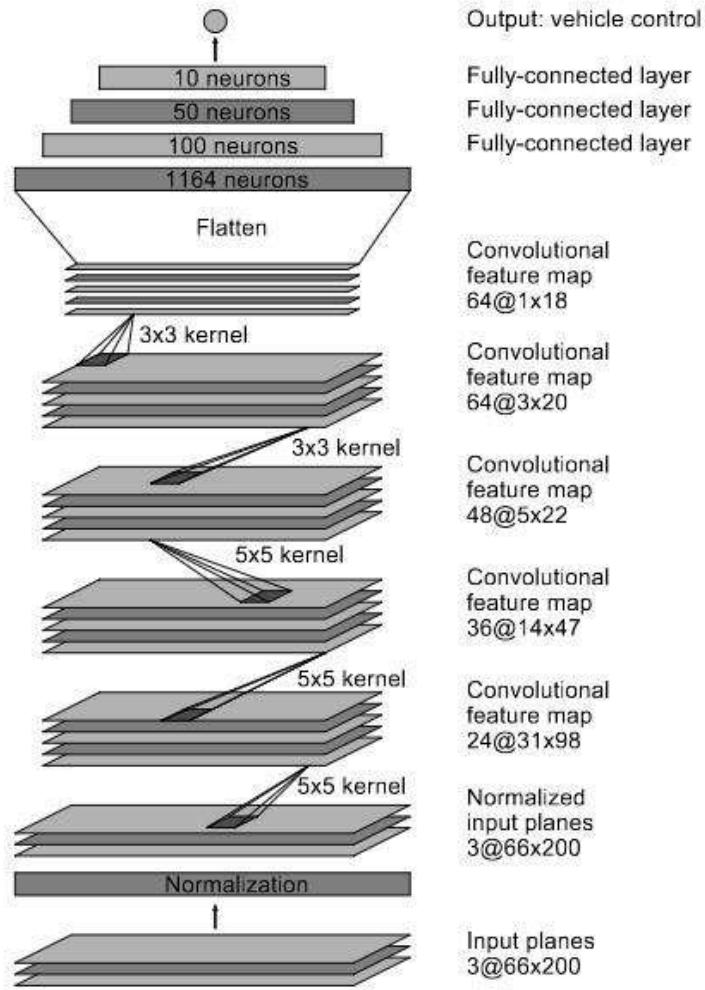
first three convolutional layers with a 2×2 stride and a 5×5 kernel and a non-strided convolution with a 3×3 kernel size in the last two convolutional layers. We follow the five convolutional layers with three fully connected layers leading to an output control

value which is the inverse turning radius. The fully connected layers are designed to function as a controller for steering, but we note that by training the system end-to-end, it is not possible to make a clean break between which parts of the network function primarily as feature extractor and which serve as controller.

5 Training Details

5.1 Data Selection

The first step to training a neural network is selecting the frames to use. Our collected data is labeled with road type, weather condition, and the driver's activity (staying in a lane, switching lanes, turning, and so forth). To train a CNN to do lane following we only select data where the driver was staying in a lane and discard the rest. We then sample that video at 10 FPS. A higher sampling rate would result in including images that are highly similar and thus not provide much useful information.



6. Conclusions

We have empirically demonstrated that CNNs are able to learn the entire task of lane and road following without manual decomposition into road or lane marking detection, semantic abstraction, path planning, and control. The CNN is able to learn meaningful road features from a very sparse training signal (steering alone). The system learns for example to detect the outline of a road without the need of explicit labels during training. More work is needed to improve the

robustness of the network, to find methods to verify the robustness, and to improve visualization of the network-internal processing steps

A Review on Cyber Security Issues and Methods For Resolving them

Vaishnavi Ganesh
Assistant Professor
Computer science and Engineering Department
P.I.G.C.E, Nagpur
vaishnavi.ganesh8@gmail.com

Abstract— The fast evolution of on-line and mobile channels has etched out new markets and brought large opportunities for aborting and established organizations alike. However, sadly the past decade has additionally witnessed important disruption to ecommerce payment processes and systems. The interconnected, anonymous and fast nature of those channels has inevitably diode to the event of malicious threats targeting ecommerce and retail services corporations, their individuals and their customers. These e-crime and digital fraud threats still evolve apace, with attackers utilizing progressively refined techniques to focus on vulnerabilities in individuals, processes and technologies. The e-crime threats, if with success completed, will undermine essential digital services, cause important injury to complete reputations, and end in wide money and operational pain for organizations and their customers. In the review of the proposed system a software can be extended in order to prevent the high-tech crime and cyber-terrorism and they spread horror by the rapid provision of information by the information security to collect internet users and the increase of the safety awareness of all banks and companies today their business online, there are millions of users who use the Internet to conduct online banking transactions. Cyber security is a critical issue now a days in various different domains in different disciplines. Another review consists of an analysis of cyber hacking attacks along with its experimental results and proposes a new methodology 3SEMCS named as three step encryption method for cyber security. By utilizing this new designed methodology, security at highest level will be easily provided especially on the time of request submission in the search engine as like google during client server communication. During its working a group of separate encryption algorithms are used. The benefit to utilize this three step encryption is to provide more tighten security by applying three separate encryption algorithms in each phase having different operations.

Keywords-Big Data, text extraction

INTRODUCTION

EVOLVING DEGREE OF THREATS

The threat landscape is ever evolving and increasingly challenging. Customer data with retailers and e-commerce firms has been increasing at a rapid pace. As per the incremental service provisioning in e-commerce, more data will be generated in the next two years than was generated ever before. Access to all this data has made the retail industry one of the primary targets for cyber-attacks. Some of the key threats today's organizations are vulnerable to include:

- User account takeover via robotic attacks, password guessing, HTML injection and Man-in-the-Middle or Man-in-

Dr. R.S.Kokhale
HOD
Computer science and Engineering Department
P.I.G.C.E, Nagpur

the-rower. Account peeking is a very common behavior by fraudsters as it allows them to validate the login credentials, identify higher value accounts and understand the controls which must be defeated to complete future unauthorized transactions.

- Business Logic Abuse or the use of portal's functionality for malicious or exploitative purposes (e.g., abuse of loyalty point programs or shopping cart functionality, fraudulent account set up, Scripted attacks to find valid coupon codes.). Impact of such abuse may include effect on the genuine customer due to unauthorized use of coupon offers, overall decrease in revenue due to offer abuse, incremental portal overhead due to scripted attacks and site scraping by resellers or coupon aggregator sites.
- Distributed-Denial-of-Service or DDOS attack on the application layer where a deluge of page requests coordinated by a bad actor overwhelms the server and brings the site down.
- Site or Architecture Probing to gather as much information about site structure and security vulnerabilities as possible to prepare for an attack on that site.
- Site & Inventory Scraping or data theft perpetrated by copying large amounts of data from a website, typically via automated scripts.

ISSUES

Cyber Security issues lead to brand degradation and change in consumer behavior. Attacks are exploiting weaknesses in traditional controls, some very destructive. Traditional controls around Point of Sale and other IT systems are necessary but not adequate – greater emphasis must be placed on preventative controls, rapid detection, and rapid response.

Retail innovations that drive growth (e.g. Digital, Omni-channel retailing, social etc.) also create cyber risk. Cyber risk management strategy must be a component of business strategy, and can't simply be delegated to IT.

1. Lack of appropriate control and transparency add to cyber security risk. Despite growing frequency and sophistication of cyber-attacks on the ecommerce industry, payment settlement agreements between credit card networks, the banks and the merchants have remained a closely guarded secret. Neither the government nor any database shares the list of defaulters with the public. Banks and credit card companies determine fault on a case-by-case basis through private contracts with individual merchants. Fines and the reasons for them remain sealed. Due to the lack of transparency, the majority of customers is not aware of any cyber security breaches and remains vulnerable to cyber attackers.

2. E-commerce firms and retailers face heat to increase efforts to ensure greater cyber security. In the wake of recent data-security breaches at large retail corporations, retailers have been pushed to spend more to ensure tighter customer data security. While the traditional retailers have been investing millions of dollars to compete with online retailers the cyber-security threats have multiplied their operational expenditures.

3. Third-party cyber risk As firms look to exploit the competitive edge they gain from the data they capture about their customers, they are increasingly leveraging the expertise of third parties Such as analytics specialists and social marketers. Couple this with increasingly lengthy and complex supply chains; retail organizations are increasingly becoming enmeshed in very complex, interconnected value chains where sensitive data is shared and dependencies are introduced between business critical systems. Firms are rapidly waking up to the realization that they often have very little visibility in these areas, and that they do not have a good understanding of where their customers data is travelling, and what their risks are. We should focus on to map these interconnections, develop robust risk management frameworks, and provide firms with assurance that they have understood and actively managed the risk of each partner relationship.

4. Inadequate joint efforts by banks and retailers to counter cyber security threats While collaborated efforts are expected to ensure tighter cyber-security, banks and retailers differ in terms of responsibility sharing. Banks want retailers to bear more of the costs of replacing cards after breaches occur whereas retailers say banks have been slow to adopt new, more secure debit card technology.

IDENTITY THEFT BREEDERS AND DAMAGES

Breeder identification can be gained by any means; its significance is in its use for obtaining additional, separate, false, or fraudulent means of identification controlled exclusively by the perpetrator without the victim's knowledge, or ability to know. An identity thief can fraudulently use obtained personal information to generate other means of identification, ranging from open new accounts, apply for loans and credit cards, to receive governmental benefits (General Accounting Office, 1998). Breeder ID means occur most often in the course of committing credit card fraud for the purpose of establishing the —authenticity required to obtain a new account, although their incidence is fairly frequent in conjunction with check fraud, document fraud/counterfeiting, signature forgery, and bank/loan fraud as well. Thus, as long as the identity thieves have knowledge of or keep a record of the stolen identities, deeper and long-term damage to the victims can explode or surprise the victims at any time after the initial damage. For that reason, in addition to financial and credit damages, some victims of identity theft may suffer from varied psychological, social, and/or legal disturbances. These hidden costs are considerable but usually are not addressed. The recent supplement of the National Crime Victimization Survey shed some light in this regard

– the emotional distress experienced by some types of identity theft victims (e.g., open new account, stolen personal information) were comparable to an average violent crime victim.

Dumpster Diving/Trashing

Identity thieves can rummage through trash of residences or businesses looking for bills, paper documents, storage devices, and even discarded credit cards containing personal information. This way of stealing identifying information is fairly labor-intensive and is restricted to limited geographic areas. Consequently, suspects are relatively easy to locate by law enforcement agencies.

Old-Fashioned Stealing

Via traditional stealing methods, identity thieves either target goods that include personal information or obtain victims personal identification as a byproduct of pickpockets. The targets are those usually containing personal identifying information, such as wallets and purses, mail, especially bank and credit card statements, pre-approved credit offers, new checks, and tax information. Old-fashioned stealing can also occur when offenders steal personnel records from institutions or bribe/coerce/deceive employees who have the access.

Changing Address

Identity thieves divert victim's mail, particularly billing statements, to another physical location by completing a change of address form. This type of identity theft is usually conducted by filing the change-of-address form with the U.S. Post Office. Thus, the U.S. Postal Inspection Service is intuitively the corresponding law enforcement agency accountable for preventive/deterring actions.

Skimming

Skimming occurs when legitimate transactions are processed by swiping credit/debit cards in retail stores or any other type of institutions where swiping cards is required. Generally, the credit/debit card numbers are stolen by a special storage device built in or attached to the swipe machines. The card information is stolen simultaneously when a legitimate business transaction occurs. The thief can be anyone who has access to the swipe machine, including, but not limited to, technicians of swipe machine vendors, and retail stores' staffs/owners. Skimming sometimes can be completed by perpetrators who attach a slim seem-like-real cover on a given ATM machine.

Pretexting

Pretexting involves a series of deceptive actions that obtain victim's personal information from the owner of the information, institutions that hold the information, and/or other individuals who may have knowledge of the information. Pretexters may pretend to have different roles (e.g., customer service representatives, survey researchers, the victims or the victim's authorized representatives) in order to collect pieces of victim's personal information. In sum, as a technique of social engineering, pretexting is a cluster of pretenses with the ultimate intention of taking financial advantage of the victims.

Hacking

Hacking was perceived as a creative activity that helped overcome the limitations of computers about a half century ago when such machines were not common, but the image of hacking changed, largely influenced by the media, to a threatening force in 1980s (Britz, 2009). The developed categories of hackers (e.g., white hat, black hat, and gray hat) are usually not mutually exclusive (McQuade, 2006; Parker, 1998) because whether their intention is malicious is uncertain from discovered evidence. Even though contemporary hacking is usually associated with stealing valuable information other than personal information (e.g., business secrets, confidential documents) and properties (e.g., copyrighted artifacts, billing) in cyberspace, it can be used as a means to obtain identifying information. Stolen identity information sometimes can be a —by-product|| of hacking for other purposes. Hacking is attractive for the reason that offenders do not have to physically appear at the —crime scene|| to —rob|| or —steal|| from institutions. Instead, exploiting online financial and billing systems is enough to illegitimately gain privileged information.

Especially after database technology is widely utilized by varied institutions to store and manage huge amounts of data, a copy of the database itself is very valuable in the black market. As more money, transactions, and even resources are moved to and managed in the virtual space for the sake of efficiency and convenience, it is likely hacking will remain a seductive means of identity stealing.

Phishing

Phishing is the pursuit of personal financial information with the intent to commit fraud by relying upon the recipient's inability to distinguish bogus emails, messages, web sites, and other online content, from legitimate ones – they all designed to appear with legitimacy.

Phishers can use a combination of tricks involving web sites, emails, and malicious software to deceive potential victims for the purpose of stealing their personal identity information and financial account credentials. The significance of phishing is that it enables remote identity theft. Precisely, phishing significantly reduces the risk and the costs to identity thieves because no physical contact, such as dumpster diving or old-fashioned stealing, is needed to complete the crime. Consequently, the chance of being caught at the crime scene is virtually eliminated. Another significance of phishing is its popularity in the U.S. where the largest proportion (25%) of phishing sites are hosted, compared to other countries in the world. A typical phishing attack begins when phishers (offenders) send out massive amounts of email (spam) or messages with bait, which is intended to trigger the targeted victim's intuitive interests. Usually, the unsolicited emails ask recipients, with a sense of urgency often exaggerated by an alleged security breach, to log onto the provided URL and confirm their personal information details, particularly their password of access. Typically these fraudulent emails are designed to look like they are from large and well-known financial institutions, such as Bank of America, Citigroup, or PayPal. In the past several years, however, observers have witnessed that phisher's Spyware (Malicious Software).

Overview of Project Cyber Security on Online Service Avoidance:

This phishing attack is done in different way and also much type of attacks is available but the common procedure is same for all type of attacks.

Procedure of Phishing Attacks.

Phishing attack procedure is depicted in Figure.

Following steps are involved in phishing attack[1] .

- Phishers set up a faked Web site which looks accurately like the legitimate Web site, including setting up the web server, applying the DNS server name and making the web pages similar to the destination Website, etc. victim's individuality to gain the illegal financial benefits. The e-mail directs the user to visit a Web site where they are asked to bring up-to-date personal information, such as passwords, credit card details and bank account numbers.
- Send large quantity of spoofed e-mails to target users in the name of those legitimate companies and groups, trying to convince the prospective victims to visit their Web sites.
- Receivers obtain the e-mail, open it, click the spoofed hyperlink in the e-mail and input the required information.
- The personal information is transmitted from a phishing server to the phisher.
- Phishers steal the personal data and perform their fraud such as transferring money from the wounded.

The phisher uses the personality information of the victim to the goal website and impersonates the victim's individuality to gain the illegal financial benefits.

The e-mail directs the user to visit a Web site where they are asked to bring up-to-date personal information, such as passwords, credit card details and bank account numbers.

The proposed work of the above mentioned project is as follows:-

In the proposed work to reduce the risk and the improvement of the Usability user's online service by creating the Phish tank. The phish tank is the web service database for Phishing websites. It gives some regular service to the API Developer. Use heuristic methodology for the automatic classification of phishing URLs as potentially in nature. This method can be used to prevent a phishing attack either by masking the potential phishing URLs or by notification to the user about the potential threat. Since the focus is on the URL itself, this approach can be applied anywhere that a URL can be embedded, e.g. in e-mails, Web sites, chat, just to name afew[2,3].

To improve the safety of the user Auto Responder email is implemented to send the message to the person who had sent threaten message. The keyboard typing words will be monitored to detect threaten words. As soon as threat words and any other worm deducted, an massage will be send to admin with ip address, time date, threat words with related sentences and the same will be sent by e mail to the person who had sent that threaten message. Here we are using check reader for sending automated emails. By using this we cancheck how many emails are bounced and track the warning message and check whether email is viewed by the sender or not.

The Phishing URL Detection Framework:

In this the user gives the URL as an input this may be phishingURL or legitimate URL [4].

The database contains the both phishing and legitimate URLs the feature extraction extract the user input by the use of database and page or Alexa ranking. After this extraction the Machine learning algorithm is applied to the URL, then the output URL is displayed where the URL is legitimate or phishing.

The Phishing Email Detection Framework:Here procedure for detection of Email phishing is described:

- Sender creating the messages, its send to the check reader.
- Check reader checks the messages, if there is no worm is detected then the message will be send to the receiver.
- If there is any worm is detected then it is blocked then the message and the sender information will send to the Admin.
- Admin will send the warning message to the sender.

The admin also monitor the warning message is read by the sender or not.

Methods and Algorithms: Suggested here is a heuristic-based approach for classifying phishing URLs with the URLs only to information. Then the problem of detecting phishing URLs as a binary taxonomy problem is treated with phishing URLs. First, phishing and legitimate sites will be collected to build the dataset. Then apply different machine learning algorithms to build models of training data. Thereafter, the two algorithms that are used here are as follows:

Random Forest algorithm, one of the most efficient machine learning algorithm in order to build prototypes of training data from the pairs of values and functions class labels. The prototypes are then separate set of test data and the data instance of the predicted class is compared with the actual data class.

Content-based algorithm, (work on the publicly available data on the URLs), which focuses on the essentials, to distinguish phishing sites, legitimate.

Overview of the 3 step Encryption Method for Cyber Security in Modern Cryptography

In this paper, authors designed a new private browser named “RIM ROCKS” whose function is to provide security on the time of client server communication. If any user wants to use this new designed private browser then he or she must be registered. After the confirmation of the registration, user will start browsing from the internet. This new designed methodology “3SEMCS” is termed as Three Step Encryption Method for Cyber Security. The complete working of this new designed methodology is based on several encryption algorithms. In addition, the major significance to utilize this new designed methodology is it may provide security from phishing websites through passing URL'S from phish tank [5] during client server communication. Presently, security professionals added extensions in the form of options in Google Chrome and Firefox [6], [7] for the detection of phishing sites. This new designed may help to provide online

security from the phishers especially on the time of client server communication.

PROPOSED PROCEDURE-3SEMCS

Table.1: Nomenclature for 3SEMCS:

3SEMCS	3-Step Ecryption method For Cyber security
DES	Data Encryption Standard
SHA-1	Secure hash Address
BFA	Brute Force Algorithm
URL	Uniform Resource Locator
T_L	Time_Limit

3SEMCS (Browser,Status,Time_Limit, Hash Address, Key Size, Encrypion_Algo (BFA/DES/SHA), Index _Pointer, Google Web Page, URL Address)

Step-1) Design a Personnel Browser.

Step-2) On_Mouse_Click:= Browser_Open and STATUS:= READY TO USE.

Step-3) Confirm Registration. [SET: = User_ID and Pwd: = STRING].

Step-4) When USER SEND REQUEST ON SEARCH ENGINE: = ACCESS FILE FROM WEB THEN Software automatically Generate Encrypted_Hash_Address.

In Addition Check the Status of The Website.

IF (CHK_URL_WEBSITE:= TRUE)

{

Not included In Phish Tank.

This website is legitimate or Original.

}

ELSE

{

Website is Fake or affected by Phisher.

}

//

Rim Rocks will correspondingly check either the website either it is effected by the phisher or not. Phish Tank help for checking the addresses of different websites.

Step-5) AS REQUEST PROCEED:=Movement_of_Encrypted_Hash_Address_Start.

Step-6) APPLY ENCRYPTION ALGORITHM: = URL_Of_Web_Page. (DES/SHA).// 2-step encryption is provided.

Step-7) AFTER THAT APPLY CAST-128 bit = On_Already_Encrypted_Hash_Address* in Step 6. //3-step encryption is provided.

Step-

8) Set:=Session_Key_On_Already_Encrypted_Hash_Address ** of Step7THEN CHECK WHETHER THESTATUS_OF_WEB_PAGE.

IF (T_L = 1 Mintue)

{

Index_Pointer:=

MOVE NEXT TO CURRENT_STATUS_OF_GOOGLE_PAGE.

}

ELSE

{

```

Index_Pointer:= 1 st Page_of_Google
OTHERWISE Repeat step 2 to step 4.
}
Step-9)
IF
(Attacker_Send_Request:=Copying_Path_From_URLAddress
)
{
THEN
Web_Page:= Expire
and Generates a Warning_Message.

}

Else
{
URL_Address:= COPIED.
}
Step-10) END.

```

Conclusion

This paper did a review of Cyber Security Issues and overcoming them. The first method offered a solution for the problem of phishing sites and URLs with Page Ranking and Phish Tank based function for random forest algorithm. It has been demonstrated that by applying -based Web Mining heuristic methodology on Random Forest and Content based algorithm. This system also offers a solution to the problem of the threat to the mail and bad words. Because the Internet is the unique situation in relation to the geography and identity, E-mail alert is required for the Internet to govern itself. So take advantage of the developments in technology and the increased efficiency of the operation in the report handling.

Next method reviewd is 3SEMCS that istermed as a Three-Step Encryption Method for CyberSecurity. The complete working of this proposed procedure is based on auto-generated encrypted hash address where the movement of encrypted hash address towards next google page shows two-step encryption* on the path by utilizing strong encryption algorithms like DES and SHA-1. On the next move on google page this will further provide encryption up to third highest level on path** for more tighten the security by utilizing brute force algorithm. In addition, this designed methodology may help to provide security from the phishers. The study of three step encryption method is actually enhance the potential of upcoming encryption technologies and its implications to defense and government users. In this way, authors say the use of new designed private browser provides a more secure channel of communication during information exchange on the time of client server communication.

REFERENCES

- [1]. Joby James, L. Sandhya and Ciza Thomas, 2013. "Detection Of Phishing URLs Using Machine Learning Techniques", International Conference onControl Communication and Computing (ICCC).
- [2].Paul A. Pavlou, 2003. "Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model". "International Journal of Electronic Commerce", 7(3): 69-103.
- [3]. Li, Y.H. and J.W. Huang, 2009. "Applying theory of perceived risk and technology acceptance model in the online shopping channel," World Acad. Sci. Eng.Technol., pp: 53-76.
- [4] Weifeng, Zhang and Hua Lu, 2013. "Web Phishing Detection Based on Page Spatial Layout Similarity". "Informatics", 37: 231-244.
- [5]Shukla Kumar and Piyush,2015.System Design- investigation and Countermeasure of Phishing Attacks using Data Mining Classification Methods and Its Analysis. International Journal of Advanced Science and Technology.
- [6]Wang HAINING AND Yue Chuan, Anti-Phishing is offence & defense. The college of William and Marry.Xin Yang and Ouyang Xi, April-2012.An empirical analysis of the effectiveness of browser based anti- phishing solutions, International Journal of digital content technology and its applications.

Auto Text Summarization

Prof. Dnyaneshwar Dhangar
dnyaneshwar.dhangar@mctrgit.ac.in

Richa Tanna
rtanna65@gmail.com

Akanksha More
akankshamore59@gmail.com

Ankita Thakur
ankitathakur1102@gmail.com

ABSTRACT

Research in text summarization is predominantly targets on measure of the worth of sentences for a summary. The proposed work has associated the Deep learning algorithm with fuzzy logic to improve the efficiency of the generated summary. The proposed work has two phases, they are training phase and testing phases. The training phase utilized to extract the benefits of fuzzy logic and deep learning algorithm for the efficient summary generation. Similar to every training phase, the proposed training phases is also possessed with well-known data and attributes. Latter to the training phase, the testing phases is implemented to check the efficiency of the proposed approach. The experimental evaluation of the proposed work provided the predictable results as, the average precision obtained is 0.37, the average recall is 0.86 and the average f-measure is obtained as 0.50%.

Keywords

Categorization, Feature matrix, Fuzzy Logic, Sentiment analysis.

1. INTRODUCTION

With the rapid growth in the quantity and complexity of documents sources on the internet, it has become increasingly important to provide improved mechanism to user to find exact information from available documents. Text summarization has become an important and timely tool for helping and interpreting the large volumes of text available in documents. Automatic document summarization is the summary of the source version of the original text while keeping its main content and helps the user to quickly understand large volumes of information. In this paper, a method for document summarization is proposed based on deep learning algorithm associated with fuzzy logic. The recent studies have showed that, the deep learning algorithms more impact on the text summarization process by pointing the most relevant objects from set of objects.

1.2 PURPOSE

Business leaders, analysts, paralegals, and academic researchers need to go through huge numbers of documents every day to keep updated, and a large amount of their time is spent just to figure out what documents are relevant and what are not. By extracting important sentences and creating comprehensive summaries, it's possible to quickly assess whether or not a document is worth reading. Automatic text summarization is also useful for students and authors. Imagine being able to automatically generate an abstract based for your research paper or chapter in a book in a clear and concise way that is faithful to the original source material. With the growing amount of data in the world, interest in the field of automatic summarization generation has been widely

increasing so as to reducing the manual effort of a person working on it.

1.3 SCOPE

Automatic summarization involves reduction of a text file into a passage or paragraph that conveys the main meaning of the text. The searching of important information from a large text file is very difficult job for the users thus to automatic extract the important information or summary of the text file. This summary helps the users to reduce time instead of reading the whole text file and it provide quick information from the large document. In today's world to extract information from the World Wide Web is very easy. This extracted information is a huge text repository. With the rapid growth of the World Wide Web (internet), information overload is becoming a problem for an increasing large number of people. Automatic summarization can be an indispensable solution to reduce the information overload problem on the web.

2. SYSTEM ANALYSIS

2.1 EXISTING SYSTEM

Yan Liu et al have proposed a document summarization framework via deep learning model, which has demonstrated distinguished extraction ability in document summarization. The framework consists of concepts extraction, summary generation and reconstruction validation. A query-oriented extraction technique has been concentrated information distributed in multiple documents to hidden units layer by layer. Then, the whole deep architecture was fine-turned by minimizing the information loss in reconstruction validation part. According to the concepts extracted from deep architecture, dynamic programming was used to seek most informative set of sentences as the summary. Experiments on three benchmark dataset demonstrate the effectiveness of the framework and algorithm.

2.2 DISADVANTAGES OF EXISTING SYSTEM

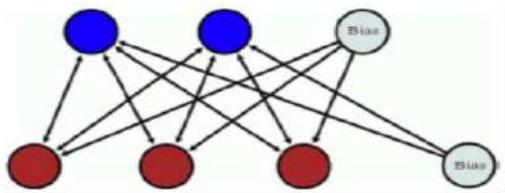
- The existing system is not user friendly and it is difficult to compute summaries using the existing methods and summaries suffer from inconsistencies and lack of balance results in lengthy summary.
- Detailed Information is not present.
- Existing methods does not give quick overview. It provides authors view not the user friendly view.
- Existing systems cannot summarise multiple documents of the same type.
- Existing systems cannot handle different languages.

3. PROPOSED SYSTEM

Proposed System consists of:

1. Restricted Boltzmann Machine

RBM is a stochastic neural network (that is a network of neurons where each neuron has some random behaviour when activated). It consists of one layer of visible units (neurons) and one layer of hidden units. Units in each layer have no connections between them and are connected to all other units in other layer as shown below in Figure.



Connections between neurons are bidirectional and symmetric. This means that information flows in both directions during the training and during the usage of the network and those weights are the same in both directions.

2. Preprocessing

Initially, the input to the proposed approach is a set of document from DUC 2002.Dataset that has to be summarized. The document utilized for text summarization is organized by a set of pre-processing steps like, sentence segmentation, stop words removal and stemming.

3. Segmentation

It is performed by identifying the delimiter commonly denoted by “.” called as full stop. This step is used to separate the sentences in the document. It is mainly useful for the user to understand each individual sentence which is there in the document.

4. Stop Words Removal

Stop words are removed mainly to reduce the insignificant and noisy words. These are predefined words such as a, an, in, by, etc., are called stop words which are filtered out before the pre-processing phase from the documents.

5. Stemming

Stemming is process of bringing the word to its base or root form for example using words singular form instead of using the plural. It basically removes the prefix and suffix of the concerned word to get the base form. There are many more number of algorithms, which are called as stemmers used to perform the stemming process.

6. Training Phase

On behalf of the training phase, the proposed approach defines five features sets. The feature sets are listed as follows,

- Title Similarity Feature

The ratio of the number of words in the sentence that occur in title to the total number of words in the title helps to calculate the score of a sentence for this feature and it is calculated by the formula given below

$$\text{Title Feature } (f1) = \frac{S \cap t}{t}$$

- Positional Feature

To calculate the positional score of sentence, the proposed approach considers the following conditions. If the sentence given is in the starting of the sentence or the last in the sentence of the paragraph then the feature value f_2 is assigned as 1. Else if the sentence is in the middle of the paragraph then the feature value of f_2 is assigned as 0.

- Term Weight Feature

The Term Frequency of a word will be given by TF (f, d) where f is the frequency of the given word and d is text present the document. The Total Term Weight is calculated by Term Frequency and IDF for a document .Here IDF denotes the inverse document frequency which just implies that the term is common or rare across all documents.

$$IDF(t, D) = \log \left(\frac{D}{d \in D : t \in d} \right)$$

- Concept Feature

The concept feature from the text document is retrieved using the mutual information and windowing process. In windowing process a virtual window of size ‘ k ’ is moved over document from left to right. Here we have to find out the cooccurrence of words in same window and it can be calculated by following formula,

$$f_4 \Rightarrow MI(w_i, w_j) = \log 2 \frac{P(w_i, w_j)}{P(w_i) \times P(w_j)}$$

- POS Tagger Feature

Part of speech tagging is the process of categorizing the words of text on the basis of part of speech category such as noun, verbs, adverb, adjectives, they belong to. Algorithms such as hidden Markov

models, using dynamic programming are used to perform this task. The POS Tags on each document is feature five of the given documents.

7. Association of deep learning with fuzzy logic

The sentence matrix $S = (s_1, s_2, \dots, s_n)$ which is the feature vector set having element as s_i which is set contains the all the five features extracted for the sentence s_i . Here this set of feature vectors S will be given as input to deep architecture of RBM as visible layer. Some random values is selected as bias 3_i where $i = 1, 2$ since a RBM can have at least two hidden layer. The whole process can be given by following equation: $S = (s_1, s_2, \dots, s_n)$. where $s_i = (f_1, f_2, \dots, f_n), i \leq n$ where n is the number of sentences in the document.

8. Feature Matrix

Here sentence matrix where $S = (s_1, s_2, \dots, s_n)$ where $s_i = (f_1, f_2, \dots, f_n), i \leq n$ is the feature vector. The five features are the main attributes of the proposed text summarization algorithm. The whole documents under consideration are subjected for the feature extraction and a set of features are extracted accordingly. Now based on the collected features a feature matrix is formed by mapping the features values. The feature matrix is constructed according to the sentences extracted from the multiple documents. In addition to the five features, an additional attribute also associated with the feature matrix. The addition feature associated with the feature matrix is the class labels for each sentence. The Figure below represents the feature matrix of the set of documents under consideration.

3. Characteristics of Proposed System

- User Friendly
- Easily generates summary
- No problem with different languages
- Boost the summary generation time.

4. Architecture Diagram

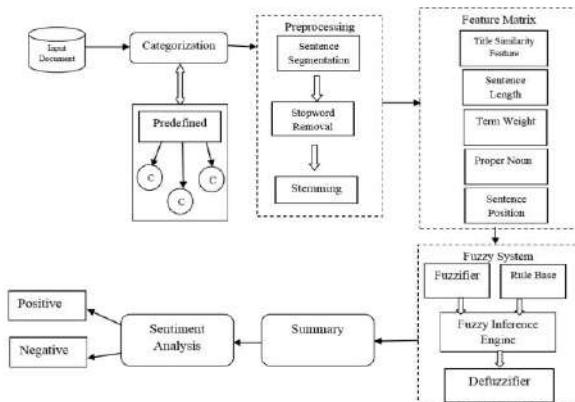


Figure 4.1 Architecture

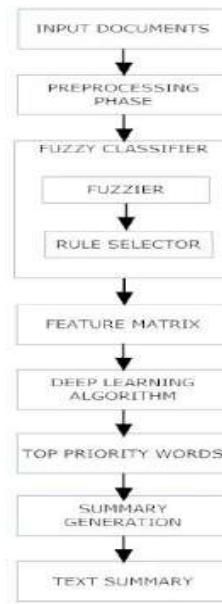


Figure 4.2 Overall Block Diagram of Text Summarization

5. Methodology

This method considers each characteristic of a text such as sentence length, similarity to little, similarity to key word, etc. as the input of fuzzy system. Then, it enters all the rules needed for summarization, in the knowledge base of system. After that a value from zero to one is obtained for each sentence in the output based on sentence characteristics and the available rules in the knowledge base. The obtained value in the output determines the degree of the importance of the sentence in the final summary. The input membership function for each feature is divided into three membership functions which are composed of insignificant values. The important sentences are then extracted using IF-THEN rules according to the feature criteria.

6. Software Environment

6.1 Front End

- Windows XP, Windows 7,8
- Visual Studio 2010
- Windows Operating System

6.2 Back End

- Windows XP, Windows 7,8
- Visual Studio 2010
- MS SQL Server 2008
- Windows Operating System

7. CONCLUSION

In this proposed work we have extracted five features for feature matrix from the set of sample dataset from DUC2002. The feature matrix is applied to our proposed work which associates the fuzzy logic with deep learning algorithm. The feature matrix is applied through the different levels of the RBM and finally the efficient text summary is generated. The result analysis shows that the proposed work produce the better performance than the existing work based on the evaluation metrics. The maximum Recall, Precision and F-Measure values for the current dataset of the proposed work is obtained as 0.37, 0.86 and 0.50 respectively for the proposed system.

8. REFERENCES

- [1] Allan Borra, Almira Mae Diola, Joan Tiffany T. Ong Lopez, Phoebus Ferdiel Torralba, Sherwin So, "Using Rhetorical Structure Theory in Automatic Text Summarization for Marcu-Authored Documents", In titaniaaddueduph, 2010.
- [2] B.Arman kiani, M.R. Akbarzadeh, "Autmatic Text Summarization using: Hybrid Fuzzy GA-GP" IEEE International Conference on Fuzzy Systems,2006
- [3] C.Gordon and R.Debus, "Developing deep learning approaches and personal teaching efficacy within a preservice teacher education context," Web Intelligence and Intelligent Agent Technology, British Journal of Educational Psychology, Vol. 72, No. 4, 2002, PP. 483-511.
- [4] F. kyoomarsi, H. khosravi, E. eslami and M. davoudi, "Extraction-based text summarization using fuzzy analysis," Iranian Journal of Fuzzy Systems, Vol. 7, No. 3, 2010, PP. 15-32.
- [5] G.Padmapriya, Dr.K.Duraiswamy, "An Approach For Text Summarization Using Deep Learning Algorithm", Journal of Computer Science, vol.6, no.11, 2013, PP.1-9.
- [6] Inderjeet Mani, "Automatic summarization", Kluwer Academic Publishers, 2004.
- [7] Jason Weston, Frederic Ratle and Ronan Collobert, "Deep Learning via Semi-Supervised Embedding, "International conference on Machine learning, 2008, PP.1168-1175.
- [8] Jen-Yuan Yeh , K. Hao-Ren and Wei-Pang Yang , "iSpreadRank: Ranking sentences for extraction-based summarization using feature weight propagation in the sentence similarity network", Expert Systems with Applications, vol. 35, no. 3, October 2008, PP. 1451-1462.

Interoperability of Electronic Health Record

Vinay Dodeja, Akshay Gunani, Dinesh Nandi

Department of Computer Engineering,VESIT

Mumbai, India

vinay.dodeja@ves.ac.in

akshay.gunani@ves.ac.in

dinesh.nandi@ves.ac.in

Mr.Prashant Kanade

Assistant Professor

Department of Computer Engineering,VESIT

Mumbai, India

prashant.kanade@ves.ac.in

Abstract— Health record of an patient to be clinically significant it needs to be from birth, not less than. As one progresses through one's life, every record of every clinical encounter represents a health associated event in one's life. Each of these records may be important or not at all required depending on the current problems that the person is suffering from. Thus, it becomes necessary that these records be available, arranged as a when person visit doctor, and be clinically relevant to provide a summary of the various healthcare events in the life of a person. An Electronic Health Record (EHR) is a digital version of patient's medical records that get generated during any clinical encounter and make information available instantly . In this paper we present interoperability of EHR without affecting privacy of individual.

Keywords— Health Record, Interoperability, Medical

I. INTRODUCTION

From the perspective of Indian Medical care system, patients visit several doctors, throughout their life time right from visiting a primary health center to community health. Health records get generated with every clinical meet during the inpatient or emergency visits. However, as it is paper based most of the health records are either lost by the patients or remain in the supervision of health care providers and gets destroyed. As per the maintenance period of medical records generally followed by hospitals is 5 years for out-patient records and 10 years for in patient records. Medical records are however retained permanently. We do have the concepts of EMR/EHR in India. But there are certain barriers to it. The idea behind any technology or a invention is to make things simple and easy for everyone.

To Store the health record of patients to digital system and accessing the record whenever required.So simple that even peoples in rural area can run it and to achieve interoperability of that record without affecting the security and privacy of the user.

II. METHODOLOGY

A. IBM Bluemix

It is a cloud platform as a service developed by IBM.The IBM bluemix cloud platform help you solve real problems and drive business value with applications, infrastructure and services.It enables developers to create, deploy, and manage applications on the cloud fast and simply.

The 3 principles of IBM bluemix:

Security:

Our platform keeps you secure with enforced standards, tested processes, dedicated tools, and the most respected provider partners in the industry.

Privacy:

Access to your data is strictly controlled and monitored in accordance with IBM's internal privileged user monitoring and auditing programs.

Compliance:

We adhere to IBM security standards across the IBM Cloud portfolio. We work with independent auditors and third-party organizations to meet the industry's most stringent guidelines.

B. Blockchain Technology

A blockchain is a decentralized and distributed digital ledger that is used to record transactions across many computers so that the record cannot be altered retroactively without the alteration of all subsequent blocks and the

collusion of the network. A blockchain originally blockchain is a continuously growing list of records called *blocks*, which are linked and secured using cryptography.

A blockchain can serve as "an open, distributed ledger that can record transactions between two parties efficiently and in a verifiable and fixed way.

C. *BigchainDB*

The BigchainDB is a blockchain database that combines the benefits of distributed databases (scale, query ability) and blockchains (decentralized, immutable / audit trails, assets / exchanges).

It looks, acts and feels like a database with added blockchain characteristics and it is complementary to decentralized storage, processing and communication building blocks. Rather than trying to expand blockchain technology, BigchainDB starts with a big data distributed database and then adds blockchain characteristics - decentralized control, immutability and the transfer of digital assets.

BigChainDB takes the best parts from distributed, "big data" databases and adds blockchain parts. But why decentralize control of data? The impetus for firms to move to a decentralized environment is to gain operational efficiency, among other benefits.

BigchainDB and MongoDB have teamed up to offer the best of two worlds – blockchain technology and high performance distributed databases. BigchainDB and MongoDB are building the core characteristics of blockchains on top of a mature distributed database that has a rich feature set, monitoring and backup tools that enterprises trust. This new blockchain database technology has the scalability needed in big data environments, by building on top of best-in-class distributed databases like MongoDB. This unlocks the potential for highly interesting applications in big data: shared control of infrastructure, and the possibility for a universal data exchange.

Being a big data database, it has the scale to actually hold the data itself, unlike traditional blockchains. As that database fills up, one can add more databases, and connect them with Interledger protocol for interoperability.

Blockchains are secure by design and are an example of a distributed computing system with high Byzantine fault tolerance. Decentralized consensus has therefore been achieved with a blockchain. This makes blockchains potentially suitable for the recording of events, medical records.

III. LITERATURE SURVEY

Indian Journal of Science and Technology, Vol 9(3), DOI: 10.17485/ijst/2016/v9i3/86391, January 2016: Implementation of Cloud based Electronic Health Record (EHR) for Indian Healthcare Needs

EHR means the digital version of the patients medical report, in store the data in real time, it contains medication and treatment history which includes the broader view of patients care and it also contains patients medical history, diagnosis, medications, treatment plans, immunization data, allergies, radiology images, laboratory and test results.[1] Methods/Statistical Analysis: The main intention of EHR is to have access to evidence based tools that health providers can make use to make decision and disease diagnosis about the patients care delivery.[1]

Ariel Ekblaw , Asaph Azaria, John D. Halamka, MD , Andrew Lippman, MIT Media Lab, Beth Israel Deaconess Medical Center August 2016: A Case Study for Blockchain in Healthcare: "MedRec" prototype for electronic health records and medical research data

A long-standing focus on compliance has traditionally constrained development of fundamental design changes for Electronic Health Records (EHRs)[2][5]. We now face a critical need for such innovation, as personalization and data science prompt patients to engage in the details of their healthcare and restore agency over their medical data. In this paper, we propose MedRec: a novel, decentralized record management system to handle EHRs, using blockchain technology. Our system gives patients a comprehensive, immutable log and easy access to their medical information across providers and treatment sites. Leveraging unique blockchain properties, MedRec manages authentication, confidentiality, accountability and data sharing—crucial

considerations when handling sensitive information. A modular design integrates with providers' existing, local data storage solutions, facilitating interoperability and making our system convenient and adaptable. We incentivize medical stakeholders (researchers, public health authorities, etc.) to participate in the network as blockchain "miners". This provides them with access to aggregate, anonymized data as mining rewards, in return for sustaining and securing the network via Proof of Work. MedRec thus enables the emergence of data economics, supplying big data to empower researchers while engaging patients and providers in the choice to release metadata. The purpose of this paper is to expose, in preparation for field tests, a working prototype through which we analyze and discuss our approach and the potential for blockchain in health IT and research[3][5]

Kevin Peterson, Rammohan Deeduwanu, Pradip Kanjamala, and Kelly Boles Mayo Clinic: A Blockchain-Based Approach to Health Information Exchange Networks

Sharing healthcare data between institutions is challenging. Heterogeneous data structures may preclude compatibility, while disparate use of healthcare terminology limits data comprehension. Even if structure and semantics could be agreed upon, both security and data consistency concerns abound[2][3][5]

Centralized data stores and authority providers are attractive targets for cyber attack, and establishing a consistent view of the patient record across a data sharing network is problematic. In this work we present a Blockchain-based approach to sharing patient data. This approach trades a single centralized source of trust in favor of network consensus, and predicates consensus on proof of structural and semantic interoperability.[3]

IV. PROPOSED SYSTEM

The proposed system diagram for the storing and accessing data represented by Fig. 1

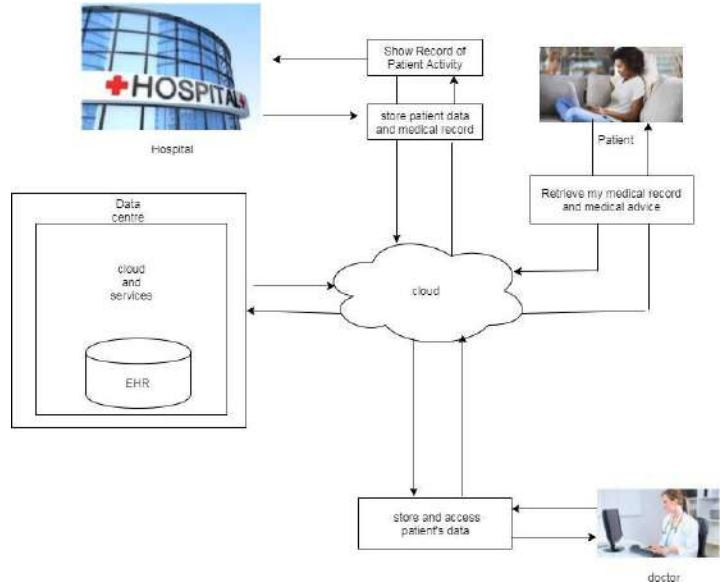


Fig. 1 EHR system architecture diagram.

The system has three modules patient , doctor and hospital namely. The patient module is used to collect personal information and store it in the data storage. The Doctor module takes data and store and upload reports. After uploading reports patient can access . The data will be on your fingertips to save time.

V. RESULTS

So after studying about various methods of gaining interoperability of EHR. A new cloud based technique can be adapted in future i.e IBM bluemix which an implementation of IBM's Open Cloud Architecture, that enables you to rapidly create, deploy, and manage your cloud applications and also maintains security, privacy and compliance of health records. Now to store thousands of patient's records a new technique known as BigchainDB is introduced which is a blockchain database that combines the benefits of distributed databases (scale, query ability) and blockchains (decentralized, immutable / audit trails, assets / exchanges). And it can store upto petabytes of data resulting in zero storage issues.

VI. CONCLUSION

The proposed system is to make things simple and easy for everyone. So sharing of patient's health record considering the drawbacks of existing systems(PHR, EHR,Paper based) like centralized, storage, availability and security so we are coming up with interoperability of electronic health record to overcome the issues of existing system. Starting with the conversion of centralized system to decentralized system and the entire patient records will be stored in a cloud based system that provides better security. And the entire data will be at your fingertips resulting in saving your time and money.

Blockchain-Based Approach to Health Information Exchange Networks.

VII. REFERENCES

- [1] Indian Journal of Science and Technology, Vol 9(3), DOI: 10.17485/ijst/2016/v9i3/86391, January 2016: Implementation of Cloud based Electronic Health Record (EHR) for Indian Healthcare Needs.
- [2] Ariel Ekblaw , Asaph Azaria , John D. Halamka, MD , Andrew Lippman ,MIT Media Lab, Beth Israel Deaconess Medical Center August 2016:A Case Study for Blockchain in Healthcare: "MedRec" prototype for electronic health records and medical research data.
- [3] Kevin Peterson, Rammohan Deeduvanu, Pradip Kanjamala, and Kelly Boles Mayo Clinic: A

[4] R. Weitzman, L. Kaci, K. D. Mandl, "Sharing medical data for health research: The early personal health record experience", J. Med. Internet Res., vol. 12, pp. 1-10, 2010.

[5] Qi Xia, Center for Cyber Security, University of Electronic Science and Technology of China, Chengdu, China- MeDShare: Trust-Less Medical Data Sharing Among Cloud Service Providers via Blockchain

Augmented Situation Awareness and Intelligence Using Data Mining

Rohit Shashi Dubey

Department of Information Technology
University of Mumbai
Mumbai, Maharashtra, India
rohititboyy@gmail.com

Akash Mishra

Department of Information Technology
University of Mumbai
Mumbai, Maharashtra, India
aaloknath.mishra@gmail.com

Sunil Yadav

Department of Information Technology
University of Mumbai
Mumbai, Maharashtra, India
sunil.yadav@slrtce.in

Swapnil Chowkekar

Department of Information Technology
University of Mumbai
Mumbai, Maharashtra, India
chowkek@gmail.com

Devvyashwar Raj

Department of Information Technology
University of Mumbai
Mumbai, Maharashtra, India
divyashwarraj@gmail.com

Abstract—Geospatial Data has an important role for advancing sustainable and development goals. Geospatial technology can highly benefit the traditional data in the field of administration, statistics and business analytics. In this report, we have present geospatial approach towards criminology. Geospatial provides an alternative to existing method of approach to analyze the crime pattern geographically

Keywords-Kriging Model; Geospatial Data; Crime; Prediction;

I. INTRODUCTION

Data mining is a popular technological method that converts chunks of data into useful knowledge and information that can help the data users make intelligent choices and take efficient steps for their own benefit. Specifically, data mining used for hidden patterns amongst huge sets of data that can help to understand, predict, and guide future behavior. Data Mining is the set of methodologies used in analyzing data from various dimensions and perspectives, finding previously unknown hidden patterns, classifying and grouping the data and summarizing the identified relationships.

There are enormous datasets that allow academics and practitioners to identify and map different types of land use (e.g. residential, commercial, industrial, leisure and Public services) different types of neighborhood (e.g. Disadvantaged inner cities, affluent suburbs, Student areas), together with street networks and major transport routes. Information such as this, referred to as “Geospatial data”, can potentially be very useful in understanding crime patterns and in explaining why crime hotspots occur in particular places. However, the extent to which beneficiaries are aware that such data exists is largely unknown and the added value that such data might bring is

largely unexplored. An important component of this project is to investigate geospatial data for crime mapping and crime analysis.

II. SPATIAL PREDICTION – KRIGING ALGORITHM

Spatial prediction involves some component of randomness. This is crucial with geostatistics when you’re making inferences on a data set kriging weights are estimated from the variogram. The quality of the estimate surface is reflected in the quality of the weights. You want weights that give an unbiased prediction and the smallest variance.

Kriging is an advanced geostatistical procedure that generates an estimated surface from a scattered set of points with z-values. Unlike other interpolation methods in the Interpolation toolset, to use the Kriging tool effectively involves an interactive investigation of the spatial behavior of the phenomenon represented by the z-values before you select the best estimation method for generating the output surface. Ordinary kriging based on centroids of administrative units to produce a surface of homicide rates and to identify clusters. However, recent advances in geostatistical methodology, such as area-to-area (ATA) and area-to-point (ATP) kriging and Poisson kriging, have opened up new opportunities.:

Variography

Fitting a model, or spatial modeling, is also known as structural analysis, or variography. In spatial modeling of the structure of the measured points, you begin with a graph of the empirical semi variogram, computed with the following equation for all pairs of locations separated by distance h:

Semi variogram(distance) = 0.5 * average ((value – value)²)

Each pair of locations has a unique distance, and there are often many pairs of points. To plot all pairs quickly becomes unmanageable. Instead of plotting each pair, the pairs are grouped into lag bins. For example, compute the average semi-variance for all pairs of points that are greater than 40 meters apart but less than 50 meters. The empirical semi-variogram is a graph of the averaged semi-variogram values on the y-axis and the distance (or lag) on the x-axis.

Fig (b): Gaussian Model

A. Mathematical Model

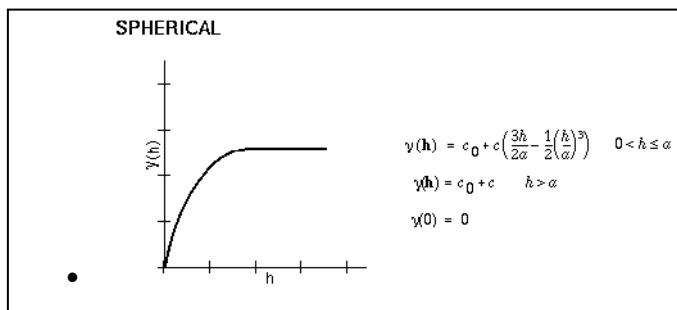


Fig (a): Spherical Model

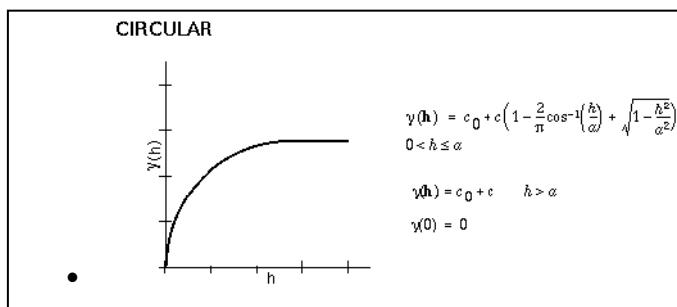
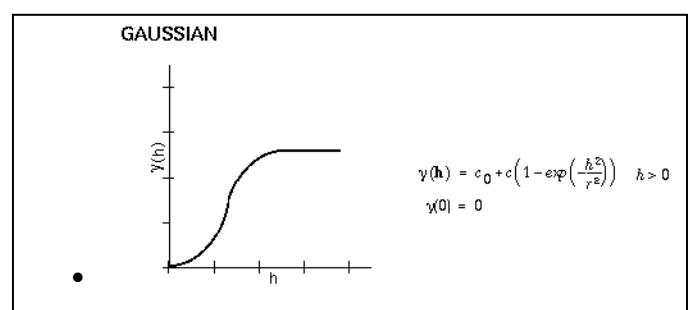


Fig (c): Circular Model

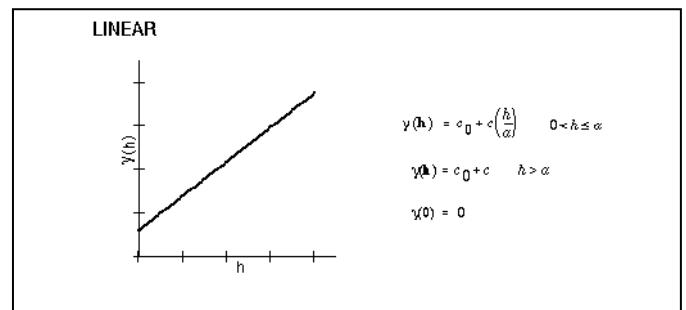


Fig (d): Linear Model

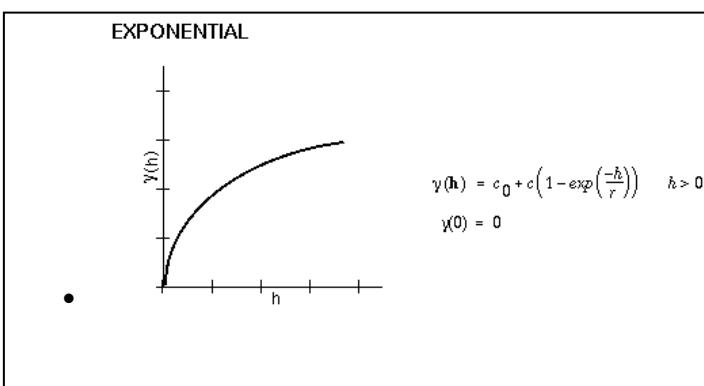


Fig (e): Exponential Model

B. Syntax

```
Kriging_3d (in_point_features, z_field, out_surface_raster,
semiVariogramProps, {cell_size}, {search_radius},
{out_variance_prediction_raster})
```

Kriging example 1 (Python window)

```
import arcpy
from arcpy import env
env.workspace = "C:/data"
arcpy.Kriging_3d("ca_ozone_pts.shp", "OZONE",
"c:/output/krigout",
```

$$\text{"Spherical", 2000, "Variable 12"})\alpha + \beta = \gamma. \quad (1)$$

C. Kriging example 2 (stand-alone script)

This example inputs a point shapefile and interpolates the output surface as a Grid raster.

```
# Name: Kriging_3d_Ex_02.py
# Description: Interpolates a surface from points using kriging.
# Requirements: 3D Analyst Extension
# Import system modules
```

```
import arcpy
from arcpy import env

# Set environment settings
env.workspace = "C:/data"

# Set local variables
inFeatures = "ca_ozone_pts.shp"
field = "OZONE"
outRaster = "C:/output/krigoutput02"
cellSize = 2000
outVarRaster = "C:/output/outvariance"
kModel = "CIRCULAR"
kRadius = 20000

# Check out the ArcGIS 3D Analyst extension license
arcpy.CheckOutExtension("3D")

# Execute Kriging
arcpy.Kriging_3d(inFeatures, field, outRaster, kModel, cellSize,
kRadius, outVarRaster)
```

III. REVIEW OF METHODS

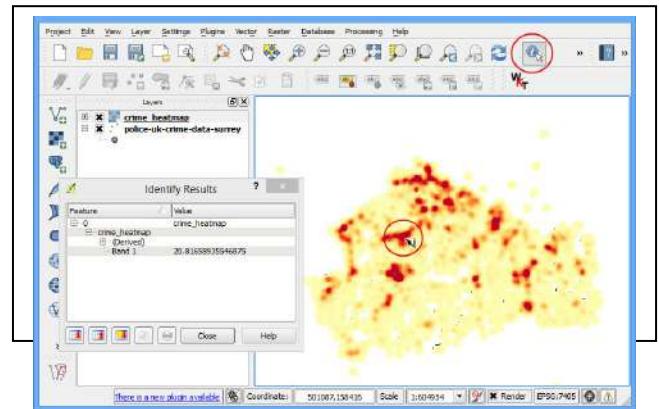
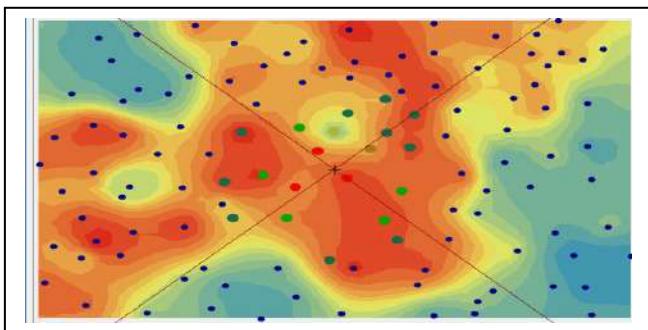
VARIOGRAM ANALYSIS

Relies on the α and β = semivariogram. In simple terms, semivariograms quantify autocorrelation because it graphs out the variance of all pairs of data according to distance. Chances are that closer things are more related and have small semivariance, while far things are less related and have a high semivariance. But at a certain distance (range), autocorrelation becomes independent. Where that variation levels off, it's called (sill). This means there is no longer any spatial autocorrelation or relationship between the closeness of your data points.

HOTSPOT-ANALYSIS

We have been able to successfully apply some spatial techniques which include the use of location, the development of kernel surface estimation algorithms, and Local Indicators of Spatial Association such as Getis and Ord Gi. Some researchers have also recognized that a simple spatial concentration of crime can be valuable. A number of similar programs are available which will not give a specific shape but would produce a hotspot surface map called heatmap, rather than similar to the temperature that we see in the weather report. These maps actually show kernel density surface map of the crime intensity also, this process involves estimating the density of crime across an entire two-dimensional study area, based on the known locations of KDE begins with overlaying grid on top of study area and calculating a density estimate based on the center points of each grid cell. Each distance between an incident and the center of a grid cell is then weighted based on a specific method of interpolation and bandwidth. The heat map process using QGIS and shows a number of parameters that must be considered before a density estimate can be produced. These parameters include the grid cell size, the method of interpolation, and the bandwidth.

QGIS provides an ease of implementation and interpretation. It is directly automating an algorithm to draw the data's cluster and process it.



Application advantages:

The system will be designed to predict the different situations leading criminal activities and will create awareness by using data mining prediction techniques. We will use the vast amount of data which is available locally and if available in social media such as twitter and by using these data we will try to predict the outcome of the different criminal situation. Moreover, with this project we could also compare different algorithms and use other methods in social data mining to predict the outcome.

1) GIS aids crime analysis by:-

- a) Identification and mapping crime incidents and events that may require further investigation;
- b) Supporting pattern and analysis across multiple jurisdictions;
- c) Improving the implementation of various policing methodologies to reduce overall crime and disorder.
- d) Integrating traditional and nontraditional law enforcement data to improve overall analysis.
- e) Educating the public with visual information to clarify crime concerns and enlist community action.
- f) Providing tools and techniques to capture crime series and forecast future crime occurrences.

g) Tactical, strategic, or administrative problem: Any data containing location information can be displayed and analysed using geographic information system (GIS) technology. GIS is an essential part of a crime analyst's (Police, Law Enforcement officers etc.) toolkit – means of creating valuable information for the concerned officers (on-duty) in the field. By incorporating traditional law enforcement data with data such as demographics, infrastructure, and offender tracking, user can use GIS to transform information into actionable intelligence. GIS will also help to improve critical decision making in a rapidly changing environment and have a direct impact on the safety of the on-duty officers and the citizens they are serving for. Every crime problem is related to some location, whether it's an address, street, ZIP Code, or district. GIS can help user to leverage the locational aspect of the data to analyse, understand, and build solutions to the problems user foreseen.

h) Violence against women in India:

Crimes which are directed specifically against women and in which only women are victims are characterised as crime against women. It is equally important to clarify the concept of 'Violence against Women'. Violence is also known as abuse and includes any sort of physical aggression or, misbehaves. When violence is committed at home it becomes domestic violence and involves family members such as children, spouse, parents or servants. Domestic violence may involve different means such as hitting, kicking, biting, shoving, restraining, and throwing objects. In broad terms, it includes threats, sexual abuse, emotional abuse, controlling or domineering, intimidation,

stalking, passive/covert abuse and economic deprivation, rape, abduction, kidnapping, murder (all cases of criminal violence, dowry death, wife battering, sexual abuse, maltreatment of a widow and for an elderly women (all cases of domestic violence) and eve-teasing, forcing wife/daughter-in-law to go for feticide, forcing a young widow to commit sati, etc (all cases of social violence), are issues which affect a large section of society. The United Nations (UN) defined 'Violence against Women' in 1993 in declaration on the 'Elimination of Violence against Women'. It defines it as any act of gender-based violence that results in, or is likely to result in, physical, sexual or psychological harm or suffering to women, including threats of such acts, coercion or arbitrary deprivation of liberty, whether occurring in public or private life..

B. Conclusion

In this paper, spatial prediction of crime location is analyzed and predicted. It is used to make short-term forecasting of property crime for one city of Maharashtra. The result shows that geospatial model fits the data well and makes higher accurate forecasting. This work is proved to be very helpful to the local police stations and municipal governments in improving the efficiency of decision-making and emergency management.

ACKNOWLEDGMENT

This work is based on data collected as part of the project "Augmented Situation awareness and Intelligence using Data Mining," by Bachelor of Engineering students of Shree L.R Tiwari College of Engineering, under the supervision of MR. Sunil Yadav, Department of Information Technology, Shree L.R Tiwari College of Engineering, Thane, Maharashtra.

REFERENCES

- [1] ↗ ShyamVaranNath, "Crime Pattern Detection Using Data Mining", IEEE, 2006.
- [2] ↗ Messner, S. F., Anselin, L., Baller, R. D., Hawkins, D. F., Deane, G. and Tolnay, S. E. (1999) The spatial
- [3] Patterning of county homicide rates: An application of Exploratory Spatial Data Analysis, Journal of
- [4] Quantitative Criminology, 15(4), 423–450.
- [5] Kyriakidis P. A Geostatistical Framework For Area-To-Point Spatial Interpolation. Geographical Analysis. 2004;36:259–89.
- [6] ↗ Gupta R., Rajitha K., Basu S., Mittal S.K., 2012. Application of GIS in Crime Analysis: Gateway to Safe City.
- [7] Patterns: Frontiers of practice (pp. 77–85). Thousand Oaks, CA: Sage.
- [8] ↗ Johnson, C.P., 2000. Crime Mapping and Analysis Using GIS. Geomatics 2000:
- [9] ↗ Leigh, A., Read, T. and Tilley, N. (1996) Problem-oriented policing, Police Research Group: Crime
- [10] ↗ Detection and Prevention Series, 75.

Speaker Dependent Speech Recognition

Shreya S. Hegiste
Thakur College of
Engineering and
Technology,
Mumbai University.

Mrinal A. Bageshwari
Thakur College of
Engineering and
Technology,
Mumbai University.

Vedashree J. Dalvi
Thakur College of
Engineering and
Technology,
Mumbai University.

Prachi P. Nagane
Thakur College of
Engineering and
Technology,
Mumbai University.

Abstract— The objective of this paper is to provide a summary of the state-of-the-art of speaker dependent speech recognition and its applications. Speaker Dependent Speech is Voice - The next disruption. For this Alexa- Amazon's voice service will be used which is the brain behind tens of millions of Amazon's devices, it creates a more personalized experience for the end users. Voice experiences are significant when they offer a quicker, uncomplicated, or more amusing way of doing everyday chores. A voice user interface (VUI) can be used by giving a voice input to control computers and devices. Today's real-world VUIs are growing smarter at an ever-increasing pace, learning and understanding the end user's speech patterns over the course of time to an extent that they are capable of even building their own vocabulary. VUIs are fast-developing and which is why voice is poised to be the next big disruption in the computing world.

Keywords—VUI (Voice User Interface), Amazon Alexa, Voice recognition, HCI (Human Computer Interaction)

With growing years, it is seen that a new form of Human Computer Interaction is emerging and quickly becoming conventional. The potential of VUI is now ever-spreading. They are partaking in voice experiences beyond customer service and search. They're talking to Alexa to control the lights, change the volume, play games, etc. [6]

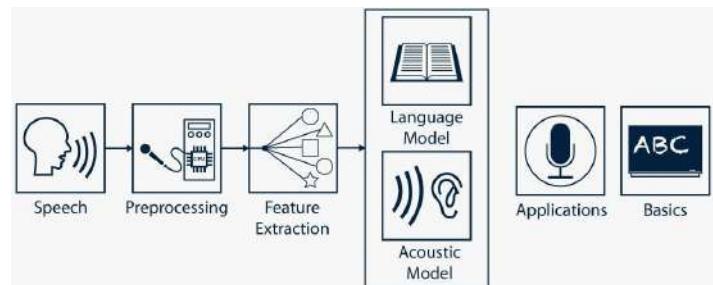


Figure 1: Block diagram of a typical Speech Recognition System

I. INTRODUCTION

Google desktop world is synonymous to search engine but the next search engine is voice. One in four searches on Google using mobile devices is done by voice. Apart from the existing voice recognition technologies there is a need for development of very strong voice device algorithms.[7] So, it is only logical that the next technological development to be natural language speaker dependent speech recognition/ Voice recognition for HCI. Speech Recognition can be defined as the process of converting speech signal to a sequence of words by means Algorithm implemented as a computer program. [1] Machine recognition of speech involves generating a sequence of words that best matches the given speech signal. Some of the known applications include virtual reality, Multimedia searches, auto-attendants, travel Information and reservation, translators, natural language understanding and many more Applications.

Speaker dependent speech recognition is a technique in computing technology by which specialized software and systems are created to identify, distinguish and authenticate the voice of an individual speaker. Speaker dependent speech recognition evaluates the voice biometrics of an individual, such as the frequency and flow of their voice and their natural accent. Speaker dependent speech recognition is also known as Voice recognition.

II. TYPE OF UTTERANCES

Speech recognition system can be separated in different classes by describing what type of utterances they can recognize.

1. *Isolated Word:* Isolated word recognizers usually require each utterance to have quiet (lack of an audio signal) on both sides of the sample window. It accepts single words or single utterance at a time. These systems have "Listen/Not-Listen" states, where they require the speaker to wait between utterances (usually doing processing during the pauses)[9]. Isolated Utterance might be a better name for this class.
2. *Connected Word:* Connected word system are similar to isolated words but allow separate utterance to be "run together minimum pause between them.
3. *Continuous speech:* Continuous speech recognizers allows user to speak almost naturally, while the computer determine the content. Recognizer with continues speech capabilities are some of the most difficult to create because they utilize special method to determine utterance boundaries.
4. *Spontaneous speech:* At a basic level, it can be thought of as speech that is natural sounding and not rehearsed .an ASR System with spontaneous speech ability should be

able to handle a variety of natural speech feature such as words being run together.[8]

III. TYPES OF VOCABULARY

The vocabulary size affects the complexity, processing essentials, performance and precision of the speaker recognition system. Some of the applications require only few words and others require very large and innovative dictionaries. The diverse types of vocabularies can be classified as follows:

- A. Small Vocabulary – tens of words
- B. Medium vocabulary – hundreds of words
- C. Large Vocabulary – thousands of words
- D. Very-large Vocabulary – millions of words
- E. Out-of-Vocabulary – mapping a word from the vocabulary the unknown word.

Various other characteristics like environment variability, speaker accent, tone of voice, pitch, age, speed of speech and the discrepancy in the signal also makes the speech recognition system more complex. Thus, efficient speaker recognition systems are developed based on the accuracy needs of the consolidating applications. Hence, this paper particularly focuses only on speaker dependent speech recognition that is voice recognition.

IV. SPEECH RECOGNITION TECHNIQUES

The goal of speech recognition is for a machine to be able to "hear," "understand," and "act upon" spoken information. The earliest speech recognition systems were first attempted in the early 1950s at Bell Laboratories, Davis, Biddulph and Balashek developed an isolated digit Recognition system for a single speaker [1]. The goal of automatic speaker reorganization is to analyze, extract characterize and recognize information about the speaker identity. The speaker reorganization system may be viewed as working in four stages:

1. Analysis
2. Feature extraction
3. Modeling
4. Testing

V. SPEAKER RECOGNITION

The speaker recognition process based on a speech signal is treated as one of the most exciting technologies of human recognition. [5] Speaker recognition is basically divided into two parts: recognition and identification. This is a way to automatically identify who is the speaker on the basis of individual information included in speech. The algorithm has to compare a voice signal from a speaker and give access to the functions only if the speaker is recognized as known. The system has been previously trained for working with only the known speaker and denies access to any other speaker for

security purpose. Basically, voice is used to confirm the identity of a speaker. [4]

VI. SPEAKER VERIFICATION /IDENTIFICATION

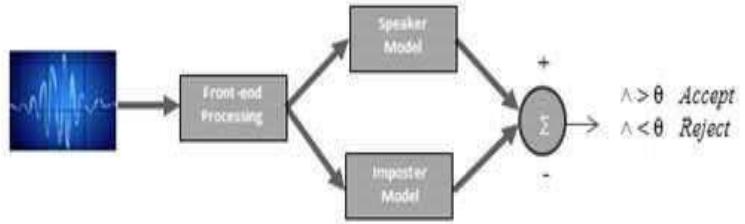


Figure 2: Speaker Verification Process

The theory that needs to be approved is: whether the test speech comes from an asserted speaker or from an imposter. In this process, attributes obtained from the speech signal are compared to a model representing the asserted speaker to a previous registration and to some models of the imposter. Likelihood ratio statistic decides whether to accept or reject the speaker where the ratio depicts the difference in log field of speakers. [5]

VII. PERFORMANCE OF THE SYSTEM

Performance of each system depends on its accuracy and speed. Accuracy is estimated with word error rate (WER) whereas speed is estimated by real time factor. Word Error Rate(WER) is calculated using the following formula:

$$WER = \frac{S + D + I}{N}$$

Where

- S is the number of substitutions,
- D is the number of the deletions,
- I is the number of the insertions,

N is the number of words in the reference. [1]

VIII. FUTURE SCOPE

With the ever-rising value of speed, we are starting to see consumers replace touch-typing with voice typing. The proof is in the behavior. The reason that the people are using more content than ever is because the consumption tools are now convenient, and we can save time. But there is still a lot of content which is to be absorbed which gives us a massive opportunity to produce on another level.

Providing the best to the speakers and making their voice the most priceless commodity, the motive of this paper is to apply voice recognition to the lifeline of Mumbai- The Local Train.

This paper intends at giving a brief but a deep understanding of how voice could take you to places, literally. The Mumbai Local Train schedule will used as the database. All users could leverage, not with the touch of their fingers but just a few words using Alexa right whenever they want needless of any knowledge regarding the schedules.

IX. REFERENCES

- [1] Santosh K. Gaikwad, Bharti W. Gawali, Pravin Yannawar, “A Review on Speech Recognition Technique”.
- [2] “A Study on Speech Recognition”, Mridula Shanbhogue1, Shreya Kulkarni2, R Suprith3, Tejas K I4, Nagarathna N5, Dept. Of Computer Science & Engineering, B.M.S. College of Engineering, Bangalore, India.
- [3] “Speech and Speaker Recognition: A Tutorial” by Samudravijaya K, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005.
- [4] “Recognition Techniques and System Classification” Arvind Gupta *Govt. M.V.M College Bhopal, Praveen Shrivastava **Sadhu vaswani College Bhopal.
- [5] Chapter 5: “Speaker Recognition - Verification and Identification”.
- [6] “Why Voice is on the Rise”, Alexa Skills Kit, <https://developer.amazon.com/alexa-skills-kit/vui> .
- [7] “VOICE IS THE NEW SEARCH [Gary Vaynerchuk]”, <https://www.youtube.com/watch?v=FYoRjPpdigM>
- [8] “Speech Processing: A Review”, Sunita Dixit- Research scholar, Pacific university, Udaipur. Dr. MD Yusuf Mulge- Principal, PDM College of Engineering for Women, Bahadurgarh.
- [9] H.Hermansky, “Perceptual Linear Predictive Analysis of Speech”, The Journal of the Acoustical Society of America, vol. 87, no. 4, pp. 1738–1752, 1990.

Intelligent Mannequin for Human-Computer Interaction

Suyash Awasthi

Computer Engineering Department

Thakur College of Engineering and Technology

Mumbai, India

suyash.ak47@gmail.com

Bavel Singh

Computer Engineering Department

Thakur College of Engineering and Technology

Mumbai, India

bavelrsingh@gmail.com

Marmik Shah

Computer Engineering Department

Thakur College of Engineering and Technology

Mumbai, India

marmik.shah5@gmail.com

Jinsith Maroli

Computer Engineering Department

Thakur College of Engineering and Technology

Mumbai, India

maroli.jinsith@gmail.com@gmail.com

Prof. Vikas Singh

Computer Engineering Department

Thakur College of Engineering and Technology

Mumbai, India

Vikas.singh@thakureducation.org

Abstract- Human Computer Interaction (HCI) as the name suggests, is related to humans and computers and the way, both interact with each other. Interaction between users and computers occurs at the user interface (or simply interface), which includes both software and hardware; for example, characters or objects displayed by software on a personal computer's monitor, input received from users via hardware peripherals such as keyboards and mice, and other user interactions with large-scale computerized systems such as aircraft and power plants.

I. INTRODUCTION

Human computer interaction (HCI) as the name suggests, is related to humans and computers and the way, both interact with each other. Interaction between users and computers occurs at the user interface (or simply interface), which includes both software and hardware; for example, characters or objects displayed by software on a personal computer's monitor, input received from users via hardware peripherals such as keyboards and mice, and other user interactions with large-scale computerized systems such as aircraft and power plants.

On the machine side, techniques in computer graphics, operating systems, programming languages, and development

environments are relevant. Intelligent Mannequin proposes the idea of having a mannequin interact with Real Users and have a complete conversation based on guidelines provided

II. OBJECTIVE OF PROJECT

A. Saving Time

Let's take a look at the typical profile of a small business owner. They are busy with day to day tasks and supporting their clients. These business owners feel quite overwhelmed and lack time to do everything they need to do to run their business successfully.

This is where a useful Embodied agents comes in; the owner can delegate certain menial tasks to the agent, such as customer demands, customer service, or even personal support like organizing meetings.

B. Saving Money

As mentioned above, small business owners operate within time constraints. They also have budget constraints, which can be challenging if they need to seek out service providers or special experts. Intelligent agents can now do tasks a human would normally do, but with more precision, on a larger scale, and at affordable costs. A few years ago, you would typically ask agencies or freelancers to do certain activities for you, which proved to be costly.

III.PROJECT / TECHNICAL OBJECTIVES:

- Achieving Dynamic conversation understanding.
- Reducing error rates
- Increasing data dictionary and NLP success rates
- Integrating Different systems to form a complete platform.
- Developing complete test analog and digital understanding database for analysis.

IV.MAIN TEXT

Hci in the large is an interdisciplinary area. it is emerging as a specialty concern within several disciplines, each with different emphases: computer science (application design and engineering of human interfaces), psychology (the application of theories of cognitive processes and the empirical analysis of user behavior), sociology and anthropology (interactions between technology, work, and organization), and industrial design (interactive products).

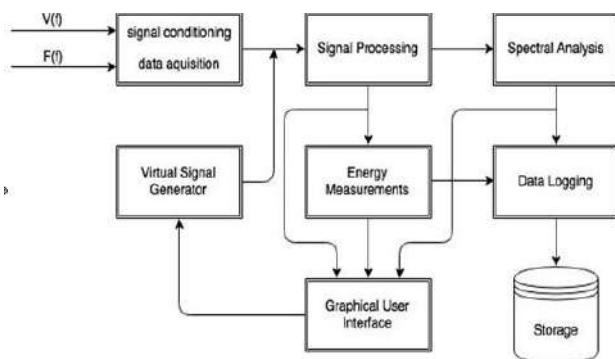
A. Luis:

- Digital technology and A.I. is growing at an exponential rate. As "Moore's law" explains, over the history of computing hardware, the number of transistors in a dense integrated circuit has doubled approximately every two years. This means that in the future, a computer can design a better one that can design a better one and so on. Today, A.I. is all around us. If you own a smartphone, tablet, pc or even a car, it's very likely that it runs some form of A.I. (Siri, Cortana, Google Now, etc.).
- There are also projects by IBM and Google that produce supercomputers capable of beating humans basically at any game. Language Understanding Intelligent Service (LUIS) offers a fast and effective way of adding language understanding to applications. With LUIS, you can use pre-existing, world-class, pre-built models from Bing and Cortana whenever they suit your purposes -- and when you need specialized models, LUIS guides you through the process of quickly building them.
- LUIS draws on technology for interactive machine learning and language understanding from Microsoft Research and Bing, including Microsoft Research's Platform for Interactive Concept Learning (PICL). LUIS is a part of a project of Microsoft Cognitive Services. Multimodal techniques can be used to construct a variety of interfaces. Of particular interest

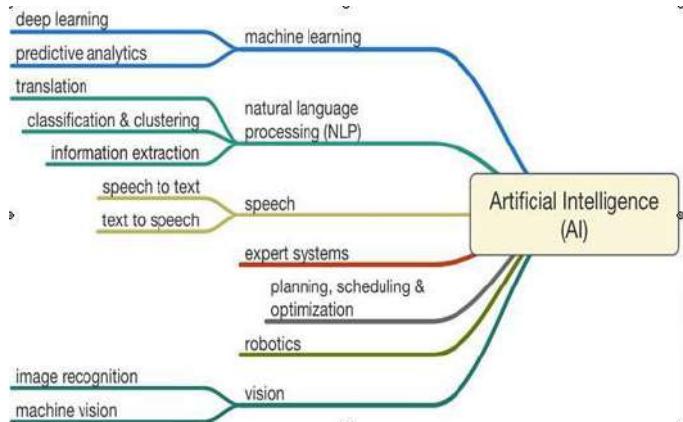
for our goals are perceptual and attentive interfaces. Perceptual interfaces as defined in, are highly interactive, multimodal interfaces that enable rich, natural, and efficient interaction with computers.

1. LARGE-SCALE BODY MOVEMENTS
2. GESTURE RECOGNITION
3. GAZE DETECTION

B. Architecture of Project



C. Architecture response of AI



V. Problem Statement: Problem Statement:

How do we decide how to represent an intelligent system in its interface, and how do we decide how the interface represents information about the world and about its own workings to a user? The rubric representation covers at least three topics in this context:

- (1) How a computational system is represented in its user interface,

- (2) How the interface conveys its representations of information and the world to human users,
And
- (3) How the system's internal representation affects the human user's interaction with the system.

Objective:

- **Resolve customer inquiries**

Recognize a customer query and correctly resolve and referents so that a grammatically valid query is sent to the answer service.

- **Correctly identify intent**

Recognize when the user has asked an elliptical question. This is necessary in order to ensure the query is not sent to the answering service. Elliptic queries are identified but they are not resolved in INTELMAN.

- **Interface with existing question-answering service**

Use existing question-answering services in order to maintain the systems status as a chatbot-like interface. This will allow the system to be used with several different QA platforms.

- **Correctly guess the most likely gender of a name**

Gender agreement is important for being able to bind the referent with a correct anaphor. i.e. binding "he" with "William".

VI .Market Survey

- Most of the organizations skip this fundamental assessment and directly jump into the next phase of identifying the platform for implementation. This is common across every organization, and it ends up either failing, or requiring more unexpected resources(time, effort and money).
- Some Organizations may not have the talent, best practices and resources to do this exercise. We recommend hiring external consultants and experts to prevent having to spend more time and money later, or failing completely.

VII. THE ORGANIZATION OF THE PROJECT REPORT IS AS FOLLOWS:

- It consists of the introduction to the project. The problem definition of the project is explained. Also the motivation for the project, scope, objectives, outcomes, impacts of the project has been discussed.
- IT mainly focuses on the literature review and the proposed work of the project
- IT deals with the initial stages of the project such as requirement gathering, planning and analysis.
- IT focuses on the design part of the project. It consists of the Data Flow Diagrams (DFD), UML Diagrams such Use case diagrams, Flowcharts etc.

- IT consists of the conclusion of the project.
- a) will be centered on the page; all previous will be in two columns.

VIII.Conclusion

- The subject of Human Computer Interaction is very rich both in terms of the disciplines it draws from as well as opportunities for research. Discussed here was just a small subset of the topics contained within HCI. Emotions can significantly change the message: often it is not what was said that is most important, but how it was said. How we define and think about our relationships with computers is radically changing.
- How we use them and rely on them is also being transformed. At the same time, we are becoming hyper connected and our interactions are being increasingly etched into our digital landscapes. There is more scope than ever before to solve hard problems and allow new forms and creativity.
- One important HCI factor is that different users form different conceptions or mental models about their interactions and have different ways of learning and keeping knowledge and this helps in improvements in Technology.

IX. Reference

- Akamatsu et al.: Multimodal Mouse: A Mouse-Type Device with Tactile and Force Display. Presence, Vol. 3,
- R. L. Anderson: A Real Experiment in Virtual Environments: A Virtual Batting Cage. Presence, Vol. 2, No. 1, pp. 16-33 (1993) [Asce95a] Ascension: Ascension trackers technical data. <http://www.world.std.com/~ascen> (1995)
- [Asce95b] Ascension: Ascension trackers technical data. <ftp://ftp.std.com/ftp/vendors/Ascension/tecpaper.ps> (1995) [Asthe95] P. Astheimer: Acoustic Simulation for Visualization and Virtual Reality. EUROGRAPHICS'95 State Of The Art Reports, pp. 1 - 23(1995)
- [Atla95] Atlantis: Atlantis VR Systems. http://vr-atlantis.com/vr_systems_guide/vr_systems_list2.html (1995) [Azum94] R. Azuma, G. Bishop: Improving Static and Dynamic Registration in an Optical See-through HMD. Proceedings of SIGGRAPH'94, pp. 197-204 (1994) [Badl93a] N. Badler, M. Hollick, J. Granieri: Real-Time Control of a Virtual Human Using Minimal Sensors. Presence, Vol. 2, H. Fuchs, R. rix: The Effect of Update Rate on the Sense of Presence within Virtual Environments. Virtual Reality: Research, Development and Application, Vol. 1, No. 1, pp. 3-16 (1995)

‘Decision Support System for Medical Diagnosis Using Data Mining ‘

Mr. Nitin Harane

Department: Computer Engineering
Thakur college of engineering and technology
Mumbai, India
Email:nitinharane13@gmail.com

Mr. Rahul Iyer

Department: Computer Engineering
Thakur college of engineering and technology
Mumbai, India
Email: kvrahul28126211@gmail.com

Mr. Siddharth Singh

Department: Computer Engineering
Thakur college of engineering and technology
Mumbai, India
Email: siddharth162009@gmail.com

Mr. Ian Noronha

Department: Computer Engineering
Thakur college of engineering and technology
Mumbai, India
Email: ian.noronha7@gmail.com

Abstract- The healthcare industry collects huge amounts of healthcare data which unfortunately are not “mined” to discover hidden information for effective decision making. The medical environment is still information rich but knowledge weak. There is a wealth of data possible within the medical systems. However there is a lack of powerful analysis tools to identify hidden relationships in data for detecting a disease number of tests should be required from the patient. These conditions describe the unexpected health conditions that directly control body parts. But using data mining technique the number of test should be reduced. This reduced test plays important role in saving time and improving performance. Data mining techniques in medical field like association rule mining with classification, clustering is implemented to analyze the different kinds of Medical problems. One of the important problems in data mining is classification. Advanced data mining techniques can help in reducing this situation. This project describes about a prototype using techniques from data mining, in which we will developing a system where user will give input as symptoms, the system will analyze the symptoms and predict the disease as per the symptoms visible. Along with the result the System will also give precautions, prescribed medicine for the predicted Result.

I. INTRODUCTION

Data mining is process of obtaining hidden knowledge from huge volumes of raw data. Data mining is used to discover knowledge out of data and presenting it in a form that can easily be understood by humans.

Disease Prediction plays a important factor in data mining. Data Mining is used extensively in the field of medicine to predict

different diseases such as heart disease, lung cancer, breast cancer, diabetes etc. This project analyzes the diseases using different algorithms. Medicinal data mining is highly important for exploring the unknown patterns in the data sets of medical domain .These patterns can be used for medical diagnosis in raw medical data available.

Heart disease, diabetes and cancer were the major cause of casualties in the world. Half of the medical causalities that occur in the countries like India, United States are due to these diseases. Medical data mining techniques like Association Rule Mining, Clustering, Classification Algorithms such as Decision tree, and Clustering Algorithm like K-Means are the data mining techniques used in medical field. With the help of this technique, the accuracy of such disease can be analyzed.

II. OBJECTIVE OF PROJECT

Decision making in healthcare is primarily done in two areas. The first area (lower level) involves patient management, diagnosis and treatment, record keeping, finance and inventory management. The other area involves higher level decision to laboratory applications or medical imaging applications. DSS is built based on the data that are derived by data mining techniques and is effective in reducing cost incurred by the hospital by preventing adverse medical events and improving quality of care. Another system that improves quality and delivery of health care services is Strategic Healthcare Decision Support Services (SHDS) which is a synergy between knowledge management and data mining techniques.

III.PROJECT / TECHNICAL OBJECTIVES:

- Achieving Dynamic conversation understanding.
- Reducing error rates
- Increasing data dictionary
- Integrating Different systems to form a complete platform.
- Developing complete test analog and digital understanding database for analysis.

IV.MAIN TEXT

1st we will train our system by uploading dataset onto our system. The dataset will be containing data like disease, age, gender, factors, symptoms, cure, food, habits, etc. Based on these Factors the Systems will be trained. The System will be trained using Classification algorithm like NAÏVE BAYES or ID3, etc. The Classification algorithm will generate a rule set onto the System. The User will provide input to the System like his age, gender, and Symptoms. Based on the Symptoms the System Will predict the disease and notify the User. Along With the Predicted Result i.e., Predicted Disease, the System Will also suggest the Precautions, Suggested medicine,etc.

A. DATA MINING

Data mining is the process of combining the different data source and derives the new pattern from that data collection. The following diagram represents different stages of data mining process: Data mining uses information from past data to analyze the outcome of a particular problem or situation that may arise. Data mining works to analyze data stored in data warehouses that are used to store that data that is being analyzed. That particular data may come from all parts of business, from the production to the management. Managers also use data mining to decide upon marketing strategies for their product. They can use data to compare and contrast among competitors. Data mining interprets its data into real time analysis that can be used to increase sales, promote new product, or delete product that is not value-added to the company.

Data mining involves six common classes of tasks:

Anomaly detection (Outlier/change/deviation detection) – The identification of unusual data records, that might be interesting or data errors that require further investigation.

Association rule learning (Dependency modeling) – Searches for relationships between variables. For example a supermarket might gather data on customer purchasing habits. Using association rule learning, the supermarket can determine which products are frequently bought together and use this information for marketing purposes. This is sometimes referred to as market basket analysis.

Clustering – is the task of discovering groups and structures in the data that are in some way or another "similar", without using known structures in the data.

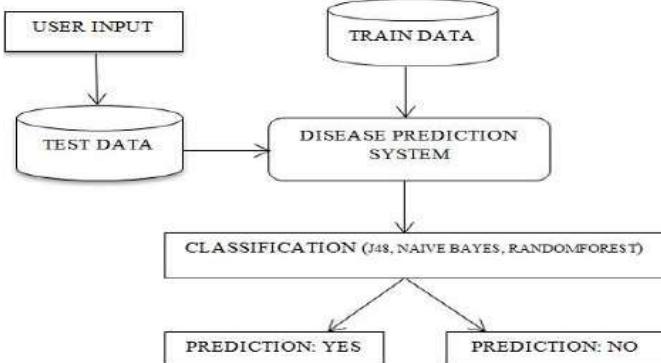
Classification – is the task of generalizing known structure to apply to new data. For example, an e-mail program might attempt to classify an e-mail as "legitimate" or as "spam".

Regression – Attempts to find a function which models the data with the least error.

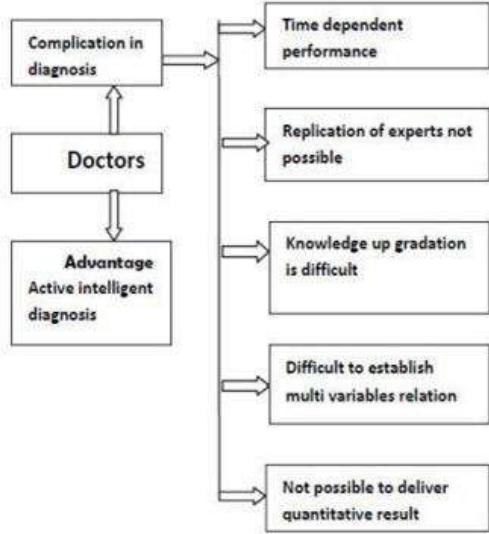
Summarization – providing a more compact representation of the data set, including visualization and report generation.

Sequential Pattern Mining – Sequential pattern mining finds sets of data items that occur together frequently in some sequences. Sequential pattern mining, which extracts frequent subsequences from a sequence database, has attracted a great deal of interest during the recent data mining research because it is the basis of many applications, such as: web user analysis, stock trend prediction, DNA sequence analysis, finding language or linguistic patterns from natural language texts, and using the history of symptoms to predict certain kind of disease.

B. Architecture of Project



C. Block diagram



V. Problem Statement:

Problem Statement:

Many hospital information systems are designed to support patient billing, inventory management and generation of simple statistics. Some hospitals use decision support systems, but they are largely limited. They can answer simple queries like "What is the average age of patients who have heart disease?", "How many surgeries had resulted in hospital stays longer than 10 days?", "Identify the female patients who are single, above 30 years old, and who have been treated for cancer." However, they cannot answer complex queries like "Identify the important Preoperative predictors that increase the length of hospital stay", "Given patient records on cancer, should treatment include chemotherapy alone, radiation alone, or both chemotherapy and radiation?", and "Given patient records, predict the probability of patients getting a heart disease."

Clinical decisions are often made based on doctors' intuition and experience rather than on the knowledge-rich data hidden

in the database. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. This suggestion is promising as data modeling and analysis tools, e.g., data mining, have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions.

Objective:

The Aim of the present study is the development and evaluation of a Medical Decision Support System for the treatment of patients with Heart Disease, diabetes and hepatitis. According to one survey, heart disease is the leading cause of death in the world every year. Just in the United States, almost 930,000 people die and its cost is about 393.5 billion dollars.

The mission of decision support systems is to improve effectiveness, rather than the efficiency of decisions. The use of data mining helps institutions make critical decisions faster and with a greater degree of confidence.

VI .Market Survey

- Most of the organizations skip this fundamental assessment and directly jump into the next phase of identifying the platform for implementation. This is common across every organization, and it ends up either failing, or requiring more unexpected resources(time, effort and money).
- Some Organizations may not have the talent, best practices and resources to do this exercise. We recommend hiring external consultants and experts to prevent having to spend more time and money later, or failing completely.

VII. THE ORGANIZATION OF THE PROJECT REPORT IS AS FOLLOWS:

- It consists of the introduction to the project. The problem definition of the project is explained. Also the motivation for the project, scope, objectives, outcomes, impacts of the project has been discussed.
- IT mainly focuses on the literature review and the proposed work of the project
- IT deals with the initial stages of the project such as requirement gathering, planning and analysis.
- IT focuses on the design part of the project. It consists of the Data Flow Diagrams (DFD), UML Diagrams such Use case diagrams, Flowcharts etc.
- IT consists of the conclusion of the project.

VIII.Conclusion

Healthcare is the most important factor affecting human life. Due to heavy work load, personal healthcare is not a possible thing. This System Will be useful for common people, where the user can give input as Symptoms to the System and the System will predict the disease along with the precautions, prescribed medicine for that disease.

IX. Reference

[1] Keerthana T K , “Heart Disease Prediction System using Data Mining Method” *International Journal of Engineering Trends and Technology (IJETT) – Volume 47 Number 6 May 2017.*

[2] Geetika Sunil Agrawal, Prof. Dinesh D. Patil, “Study of Heart Disease Prediction Using Data Mining”, International Journal of Innovative Research in Computer and Communication Engineering Vol. 5, Issue 6, June 2017.

[3] Sonali S. Jagtap, “Prediction and Analysis of Heart Disease”, International Journal of Innovative Research in Computer and Communication Engineering Vol. 5, Issue 2, February 2017.

[4] Vincy Cherian, Bindu M.S, “*Heart Disease Prediction using Naïve Bayes and Laplace Smoothing Technique*”, *International Journal of Computer Science Trends and Technology (IJCST) – Volume 5 Issue 2, Mar – Apr 2017*

[5] Nidhi Bhatla Kiran Jyoti, *An Analysis of Heart Disease Prediction using Different Data Mining Techniques*, International Journal of Engineering Research & Technology (IJERT), 2012.

Dynamic Blood Bank

Nilay Darji, Pulkit Kedia, Archit Kogta, Azad Nautiyal, Rajesh Singh

(Computer Engineering , TCET/Mumbai University, India)

Abstract:

Blood is a saver of all existing lives in case of emergency needs. Blood is one of the most essential necessities of all human beings. In emergency situations, such as accidents, the common problem is need of blood in the hospitals. Despite of the increasing requirement of the blood in emergency cases, about only 5% Indian population donates blood. The issue is not insufficient number of donors, but finding a willing donor at the right time. With our project, in this paper, we will propose an effective way to overcome such emergency requirement of blood. We are going to build a network of people who can donate blood during an emergency. The main aim of the project is saving life of human beings. This project allows user to see the information of the donors such as name, contact number, address, blood group as well as other medical details of the donors. This App provides the list of Hospitals or Blood bank near you using GPS. Since almost everyone has a Smartphone phone with them, it gives instant GPS tracking as well as communication. The services will be access only by the person who is willing to donate blood.

Keywords:

Android, Blood, Blood bank, Donor, Acceptor, Database System, GPS System.

I. INRODUCTION

Blood is a body fluid in humans and other animals that delivers necessary substances such as nutrients and oxygen to the cells and transports metabolic waste products away from those same cells^[3]. For the treatment in medical field, the need of blood is very important. Every minute someone requires blood for their urgent need to save his/her life. The Blood bank acts as mediator between the donor and the acceptor. It

receives the blood from the donor, stores it and preserves it for future blood transfusion

In emergency case. It also monitors the blood group database and sends it to the hospitals that needs particular blood group in urgent need. It is possible in some scenario that the patient is not able to get the required amount of blood due to poor

network database among the blood banks which leads to the lack of knowledge of updated record of all blood donors. Now a day's almost everyone have mobile phones with them, which helps the patient to get required blood at any time using this App.

II. LITERATURE SURVEY

(I)Literature Review: The Optimization of Blood Donor Information and Management System by Technopedia.

P. Priya, V. Saranya, S. Shabana, Kavitha Subramani [1] has proposed an efficient web application that timely updates the information related to the donors, the acceptor and the patients. In this system admin has access to whole information about blood bank system. The service provided by the system is needed and valuable to health sector where a quality of the blood is considered for the safety of the patient through a systematic process by the blood management system. The proposed work is also having the security regarding the important details of donor which can be misused by other persons. It maintains the data of available blood groups, and if it finds the lower stock than the required stock of any particular blood group, then it notifies donors having that same blood group to donate it as soon as possible. Along with the web application, an Android mobile application is proposed to find the donors in emergency cases in nearby areas.

A Survey Paper on E-Blood Bank and an Idea to use on Smartphone

Blood is an essential component for all animals. In case of emergency requirement, it is considered as lifesaving element. Online blood bank never gives the direct contact

between blood donor and blood bank. The main drawback of the existing systems is time consuming, also it is very costly.

Tushar Pandit, Satish Niloor, A.S. Shinde [2] has introduced comparison between existing system and improved system. The new idea

Is to move from desktop system to much faster and efficient mobile system. E-blood bank is automation system. The major function of E-blood bank is to get a single network that comprises of all the blood banks in that state. It also validates stores and circulates various information using computation technology. This data will help the people in the need of blood on urgent bases by quick access through the App and save the life.

Blood Bank Management Information System in India.

Vikas Kulshreshtha Research Scholar, Dr. Sharad Maheshwari [3] has introduced the review of the main features, merits and demerits provided by the existing Web-Based Information System for Blood Banks. Blood is one of the most precious elements of the life. ‘Blood Bank’ refers to the particular section of hospital where the storage of blood and blood product occurs. It is also responsible for proper testing that is performed to minimize the problem of transfusion. Big sized cold container holds blood and blood products at constant temperature. This system gives quick access to the donor records collected from all over the country. It provides effective search that needs blood in their city as soon as possible.

(ii) Existing System:

Blood Connect:

Blood cannot be manufactured in the factory; it can only collect from the donor. To complete to this demand, Blood Connect was launched on 1st April, 2010 (as a project under NSS IIT Delhi) with an aim of solving the problem of blood shortage in India.

Shortage of about 3 million blood units is faced by India, According to WHO data. Even if additional 2% of blood is donated by young generation of India, this problem can be easily solved. For that purpose, Blood Connect acts as a mediator between the willing donor and patient.

E-Blood Bank:

This App helps to find people donating blood regularly in nearby area. You can contact them via phone number or address.

If you need blood in emergency case, then can get the location of the user as well if he/she has registered with the App. If your blood group is same as the blood group that is needed on urgent bases, you will get push notification. You can find nearby hospitals and access them.

Features:

- Find donor with your specific blood group and with your respective states and cities.
- Send notification: This will help you know who all are having the same blood group in your local area.
- Find nearby hospitals in maps.
- Provided helpline numbers in case of emergency.

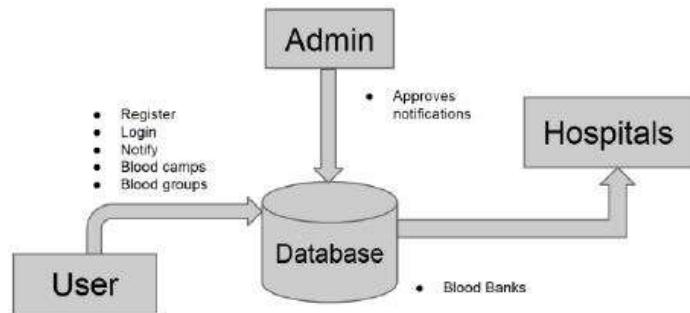
III. PROPOSED SYSTEM

Person has to first download the App. After downloading the App, registration form will appear. For registration basic information like name, address, DOB, Contact number, blood group, emergency contact number etc. are supposed to be filled. If he/she has already registered, then he/she has to login.

The user gets different options:

- Blood banks
- Search donors
- Search nearby places
- Requirement for blood
- Request for blood
- Speed dial
- Medical first aid
- Profile update

Block diagram



IV. METHODOLOGIES

Android Studio:

Android Studio is the official integrated development environment (IDE) for Android app development, based on IntelliJ IDEA. Android Studio is designed specifically for Android development. It offers flexible Gradle-based build system along with code templates to build common app

features. It offers good layout editor with support for drag and drop theme editing and built-in support for Google Cloud Platform, which makes it easy to integrate Google Cloud Messaging and App Engine and much more.

It features a new and better interface design views where you can see the interface you are working on and its related components.

Blood Bank and an Idea to use on Smartphone”. Dept. of I.T Sinhgad Academy of Engineering, Pune, India. Year 2015.

[3].<https://en.wikipedia.org/wiki/Blood>

V. CONCLUSION

We have proposed an effective android blood bank application. The system provides better communication between Blood banks and blood donors to save someone's life in emergency cases. It also interconnects the blood information available only to the local areas to improve the blood donation service quality. It takes minimum efforts and time to complete the whole process. This App allows people to request for blood in urgent case, find blood bank for willing blood donors, getting information of specific donor in emergency cases and many more.

REFERENCES

[1]. P. Priya1, V. Saranya2, S. Shabana3, Kavitha Subramani4, “The Optimization of Blood Donor Information and Management System by Technopedia”. Department of Computer Science and Engineering, Panimalar Engineering College, Chennai, India, Volume 3, Special Issue 1, February 2014

[2]. Tushar Pandit, Satish Niloor and A.S. Shinde, “A Survey Paper on E-

Travel and Overtime Allowance Manager for Western Railways

Vaiditya Chauhan¹, Ankit Gupta², Yash Nair³, Dr. Sheetal Rathi⁴

B.E (C.S.E) Mumbai University ^{1, 2, 3}

Associate Professor, TCET, Mumbai ⁴

Abstract

The management of allowances in Western Railways is a common and critical operation, as it has an effect on the payment of salaries of employees. Currently, the manual calculation of allowances is too complex and takes a lot of time efforts. If these processes are automated, it would be of great benefit as it would require less amount of time to calculate the salary of the employees. So, a mobile application that has a system to facilitate the above cause is considered as a solution here. We are going to digitize an existing system of periodic calculation of Travel Allowance and Overtime Allowance, which currently involves considerable amount of paperwork. This existing methodology consumes an incredibly huge amount of time, considering the verification procedures and transportation of the voucher forms from the various railway stations to the Western Railway Headquarters and vice versa. The proposed application will be capable to automatically calculate the allowances according to the inputs fed in by the employees, thereby simplifying the existing system and reducing the time consumed.

Keywords

Travel Allowance, Overtime Allowance, Android Application

I. INTRODUCTION

The Mumbai Suburban Railway is a public transit system serving Mumbai Metropolitan Region of Western Railways, Maharashtra. It consists of thirty seven stations from Dahanu Road to Churchgate Station. It is operated by Western Railways (WR) and serves to approximately 3.52 million commuters daily.

The employees of the Western Railways are entitled to a number of allowances, most commonly being Travel Allowance (TA) and Over-time Allowance (OTA). In the existing scenario, these allowances are paid accordingly upon the filling and verification of voucher forms, which is a long and tedious process. The entire process is dependent on paper work and manual calculation which makes it error prone. Thus, we have taken an initiative to develop an android application in order to digitize the existing system and thereby contribute towards the Digital India mission.

The proposed application will have separate modules for filing travel and overtime allowances. The app will facilitate addition of new users through registration, which will only be permitted if approved by a higher

authority (in our case, the station manager). The technologies that we will be using for app development are Android, PHP and MySQL. The user interface will be designed with the help of Adobe Photoshop CC.

II. BACKGROUND

With a view contribute towards the Digital India initiative, we wished to make software which can help our government to execute their tasks efficiently. Hence, with this aim we approached the Western Railways Headquarters to know more about their existing issues. The organization proposed this problem statement of allowance management and thus, we have built an android application for the same.

III. LITERATURE REVIEW

The system developed by Poonamdeep Kaur, Dr. Dinesh Grover is a web based software developed using PHP, HTML, CSS, JavaScript, AJAX, MySQL. The main features of this software are calculation of allowances, taxes and other deductions, printing individual pay slips and deduction vouchers [1]. The software system developed by Arjun V. Singh, Siddesh V. Chaphekar and Yogesh S. Sawant is developed using VB.net as frontend and Microsoft Access 2007 SQL server 2008 as backend. The planned system has a database that stores all information related to personnel allowances, deductions, taxes and net pay of the employees. The features of the system include importing attendance from Biometric machine, sending details regarding salary and attendance before finalizing salary, Faculty Management, Overtime Calculation, through mail sending salary slips, HRD programs like offer letter, appointment letter, promotion letter etc[2]. The software implemented by Kritika Mahajan, Shilpa Shukla, and Nitasha Soni is also a desktop based system, which is developed in HTML, CSS and JQuery as frontend, C#, ASP.net is used for backend. For data parsing, JSON and Ajax is used. The system tracks the number of hours worked and a record of employee data including their pay, allowances, deductions and taxes on monthly bases. The fresh definitions are reflected from the current month onwards [3]. The application developed by Manish Singh, Prachi Singh, Rohil Singh, Shubham Singh, and Shiwani Gupta is a web based portal. The features of the application includes applying for leave, viewing leave history, viewing

leave stats and granting/rejecting leave applications. The Payroll Management module deals with viewing payroll & tax deductions and automatic tax calculation [4].

The allowances in the Western Railways are calculated using the grade pay. The general formula is:

$$\text{OTA} = \text{Grade Pay} * \text{Principal Amount} * \text{Work hours}$$

where,
 $\text{Work Hours} = \text{Actual Duty Hours} - \text{Roster Duty Hours}$

Similarly, the travel allowance can also be calculated as follows:

$$\text{TA} = \text{Grade Pay} * \text{Principal Amount} * \text{M.F}$$

where,

$$\text{M.F (Multiplying factor)} = \begin{cases} 30\% & \text{time} < 06 \text{ Hrs.} \\ 70\% & \text{time} < 12 \text{ Hrs.} \\ 100\% & \text{time} > 12 \text{ Hrs.} \end{cases}$$

IV. PROBLEM IDENTIFICATION

The current scenario is that each western railway employee who is entitled to and wants to claim OT has to fill up the vouchers on daily basis against their Roster duty timings and manually calculate the extra duty hours done by him/her, either extra duty was done in day or night shift and their grades (on which the amount to be multiplied varies). This process is carried out for 14 days(cycle of 14 days is fixed throughout the year).Now this filled voucher of 14 days of all the employees are then submitted to Station in charge who will then verify all the entries and send it to Headquarters for final authentication. There these vouchers are authenticated and brought back to the station and final entry is made in salary database.

Here, the major problems are:

- Duration of the whole procedure is immense as it involves manual transportation.
- Transportation cost increases only for the sake of carrying vouchers.
- Fully paper based procedure where employee can only fill up the vouchers when they are present at their workplace.
- No proper verification can be done as Station in charge gets the information at the end of 14 days not on the basis of what is happening on a daily basis.
- If due to any reasons, the paper voucher get destroyed or misplaced, there is no way the employee can claim the allowance and may have to

go through a long and complex process which is time consuming.

Similarly with TA, only the data types are getting changed and rest of the things is the same and it is also having similar problems as OT voucher.

V. PROPOSED WORK

This android application aims to radically transform the existing manual system. The basic steps that will be performed by the system are as follows:

1. Accept the OT and TA data from the user and store them over a central remote database.
2. Create a real time database to help the station manager keep track on each and every employee's progress.
3. Make calculations using the prescribed formulae.
4. Reflect the results of the calculations to the user and to the higher authorities for verification and validation.
5. Generate cycle-wise reports (in PDF/Word Document format) periodically and forward the same for verification to the verifying authority.
6. Prompt the user to make changes as and when required upon the notification triggered by the station manager during validation.

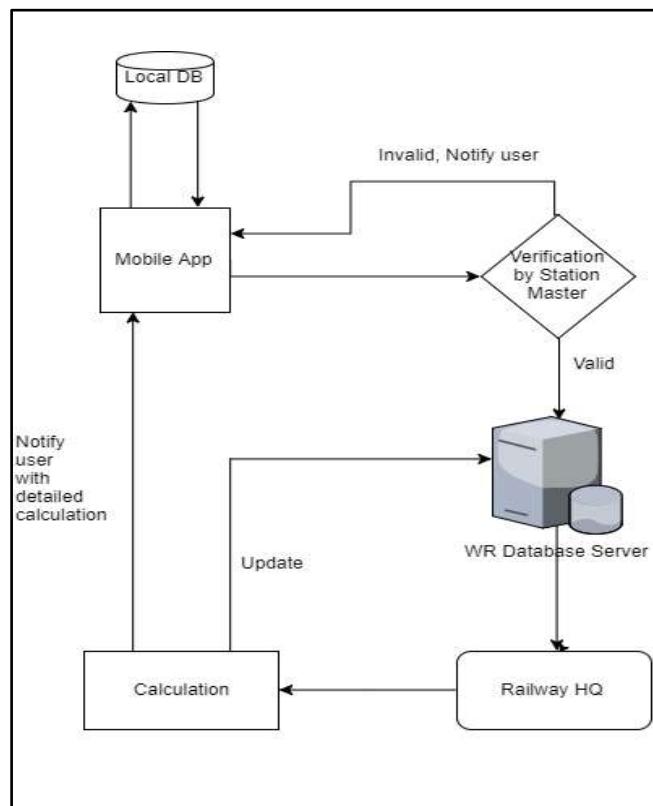


Fig.3 (a)

The android application is based on Overtime (OT) and Traveling Allowance (TA) for Western railway employees. So following are the categories of employees present at each station who can access the application for their procedure of applying and calculation of OT and TA.

- **Operating:**
This category includes number of station masters, point's men and station porters. The station managers are not eligible for OT allowances.
- **Commercial:**
This category includes booking clerks.
- **Signaling and Telecommunication:**
This category includes signal inspector, electronic signal maintainer (Khalsi).
- **Overhead equipment:**
It includes all of the technicians and workers.

VI. SYSTEM OVERVIEW

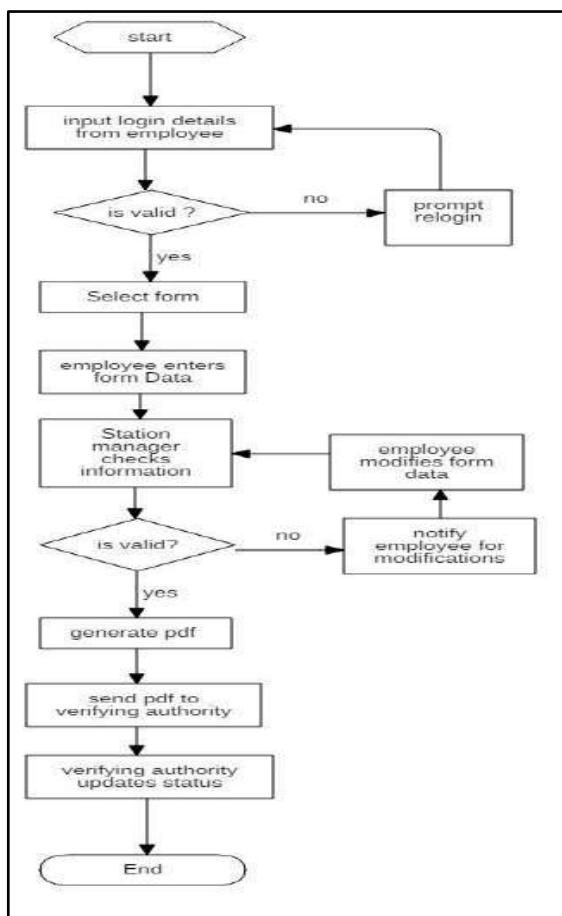


Fig.3 (b)

The application will provide an interface to the employees to login their OT and TA Voucher details, which will essentially be an e-Voucher form. These two forms will be available as two different modules.

Module 1: Overtime Allowance

The overtime allowance module allows the user to select the form, fill in the form and review them later. This module includes the various employee working categories which are namely Intensive working hours and Normal working hours. It displays a calendar of current cycle which displays a summary of previously filled vouchers. It also gives the user option to edit past vouchers if they contain any errors. The employees in Intensive working hours category include a total work time of 6 hours, above which the work hours spent by the workers fall under allowances for which they are paid extra based on the formula specified earlier.

Module 2: Travel Allowance

If the employee leaves his/her designated work station or headquarters for official duty then this travel is sponsored by the Western Railways, meaning he/she is given extra monetary benefit which is calculated and included in their salary. The allowance awarded to travelling employees depend upon various factors such as time spent in travelling, total distance covered by the employee, etc. The TA section allows automation of this calculation and makes the process faster. This software module also displays the calendar to choose the day on which he/she travelled to claim the allowances from their organization. Unlike Overtime allowance, here all the employees are entitled for the Travel allowances.

On the completion of the periodic cycle all the forms of a particular station will be converted into a report and will be forwarded to the validating authority, i.e. station master. The station master may accept or reject the report based on the correctness and authenticity of the report. If correct then the report will be forwarded to the verifying authority. If not the user will be notified to identify and correct the errors. The Railway headquarters upon receiving the report will verify and release the allowances and reflect in the database.

VII. EXPECTED OUTCOME

The application will present the user the choice to select which type of allowance form he wishes to fill. The travel allowance module and overtime allowance modules will get presented once the user makes the decision. For example, suppose the user selects overtime allowance module. The application will present the list of cycles comprising of both the previously filled ones and the current cycle. The user will be permitted to make changes for a few previous cycles and the current cycle only.

Previous Cycles		
01/01/2017	TO	14/01/2017
15/01/2017	TO	26/01/2017
29/01/2017	TO	11/02/2017
Current Active Cycle		
12/02/2017	TO	26/02/2017

Fig.4 (a)

After the selection of cycle, a calendar will be displayed as shown in fig 4(b). The dates will be color coded as per the status of the forms.

SPECIMEN

Fig.4 (d)

TOTAL NOS. EMPLOYEE	TOTAL WORKING HOURS	TOTAL AMOUNT PAID

VIII. RESULTS

Fig.4 (b)

The user will select the date on which he wishes to fill in the allowance form. The form will look like in fig 4(c)

Fill OverTime Form

Shift:

ROSTERED DUTY:
14:00 - 22:00

ACTUAL DUTY HOURS:	START TIME	END TIME
	15:30	22:30

DESCRIPTION:
emergency

SUBMIT

Fig 4(c)

Similarly, the Travel Allowance module will be displayed and filled in. At the end of the cycle, a report as shown in fig 4(d) will be generated and sent for further verification.

Existing System	Proposed System
The existing system is primitive and outdated	The proposed system is modern and advanced
Error prone	Chances of error are minimized
The entire process is time consuming	Time consumption is very less
Not economical since the expenditure for the entire process is recurring and quite considerable	This digital alternative is economically sustainable as it is a onetime investment
Less efficient	More efficient

IX. FUTURE SCOPE

The software is basically built according to the requirements of the western railways, thus following their norms and regulations. Thus, any other organization having similar procedures may use this application by making some minor modifications. Also, the system can be linked to the existing payment modules in future; thereby enabling the automatic payment of allowances to the bank accounts of the employees.

X. CONCLUSION

The travelling and overtime allowances are the most common allowances offered to the Western Railway employees by the organization. The current procedure for filing and verification of the same is a manual procedure involving considerable amount of paperwork and human effort. With a view to provide a digitized and error free alternative, we propose to develop a user friendly android application that will replicate the entire procedure digitally , thereby minimizing the calculation errors and the transportation issues incurred while filling the paper voucher forms. Since this is a first of its kind initiative, this application will also help in setting ground for radical improvements in the railway administration at large.

10. Definition of Travel Allowance
<http://indianrailwayemployee.com/content/daily-allowance-popularly-known-ta>
11. Android Documentation
www.developer.android.com

ACKNOWLEDGEMENTS

We are extremely thankful to the officials of the Western Railways who provided expertise that greatly assisted the project. We also express our gratitude to the concerned authorities for sharing their pearls of wisdom with us during the course of this project.

REFERENCES

1. Poonamdeep Kaur, Dr. Dinesh Grover (2012) “Computer Based Payroll System Implementation for E-Governance at Punjab” Agricultural University, Department of Computer Science.
2. Arjun V. Singh, Siddesh V. Chaphekar, Yogesh S. Sawant (2016) “Automated Payroll System” University of Mumbai, Department of Computer Science.
3. Kritika Mahajan, Shilpa Shukla, Nitasha Soni (2015) “A Review of Computerized Payroll System” University of Lingaya, Department of Computer Science.
4. Manish Singh, Prachi Singh, Rohil Singh, Shubham Singh, Shiwani Gupta (2017), “Leave and Payroll Management System”, University of Mumbai, Department of Computer Science.
5. Railway Establishment Rules.
<http://www.indianrailways.gov.in/railwayboard/uploads/directorate/establishment/estmanualmain.jsp>
6. OT Voucher Form (Hard Copy).
7. TA Voucher Form (Hard Copy).
8. Station Handbook.
9. Definition of Over Time Allowance
<http://indianrailwayemployee.com/content/overtime-allowance>

Opinion Extraction and Sentiment Knowledge Discovery of Twitter data

Ms. Vani Singala

Thakur College of Engg. and Tech.
Student

Ms. Nidhi Mishra

Thakur College of Engg. and Tech.
Student

Ms. Aneri Moradia

Thakur College of Engg. and Tech.
Student

Ms. Sonu Mehta

Thakur College of Engg. and Tech.
Student

Mrs. Megharani Patil

Thakur College of Engg. and Tech.
Guide

Abstract—Ongoing increase in wide-area network connectivity promise vastly augmented opportunities for collaboration and resource sharing. Now-a-days, various social networking sites like Twitter, Facebook, MySpace, and YouTube have gained so much popularity and we cannot ignore them. They have become one of the most important applications of Web 2.0. With the rise of social networking epoch, there has been a surge of user generated content. Micro-blogging sites have millions of people sharing their thoughts daily because of its characteristic short and simple manner of expression. We propose and investigate a paradigm to mine the sentiment from a popular real-time micro-blogging service, Twitter, where users post real time reactions to and opinions about “everything”. We expound a hybrid approach using both corpus based and dictionary based methods to determine the semantic orientation of the opinion words in tweets. A case study is presented to illustrate the use and effectiveness of the proposed system

Keywords—Sentiment analysis; opinion mining; Twitter data analysis

I.INTRODUCTION

Sentiment Analysis (SA) and summarization has recently become the focus in the field of data science, this is mainly because analysis of online text is beneficial and demanded in many different applications. One such application that has already been developed is the pragmatic analysis of different user environment and preferences specified on Facebook, which helps other companies to narrow down its target users and which specifies the probability success for an advertising company's product.

Microblogging sites have millions of people sharing their thoughts daily because of its characteristic short and simple manner of expression. It has become increasingly important to analyze these informal expressions of speech and identify the inherent meaning of a tweet and thus bifurcate the opinions of a majority consensus and take necessary actions based on it. In this project we have tried a novel solution to target-oriented (i.e. aspect-based) sentiment summarization and SA of short informal texts with a main focus on Twitter posts known as “tweets”. We have compared different algorithms and methods for SA polarity detection and sentiment summarization.

This project is implemented using hybrid polarity detection and learning method essentially uses both corpus based and dictionary based methods to determine the semantic orientation of the opinion words in tweets. [2] Hybrid aspect -based learning that we have worked on in this project is inherently better than unsupervised or supervised learning based aspect detection system, as a supervised algorithm suffers from a large time and memory complexity due to its inability to automatically train itself on a large set of data, and unsupervised learning algorithms requires lot of efforts to acquire and maintain a lexicon, and our system has tried to reduce inconsistencies and ambiguous outputs, For instance, for the following tweet - "I absolutely detest Samsung galaxy note but like it's camera ", here a non-aspect -based system will straight away categorize this tweet as extreme negative due to the presence of 'absolutely', but our aspect-based system will identify this tweet as negative for the product as a whole , but also categorize it's aspect(here camera) separately as positive, this gives the manufactures focused results to dwell on.

I. BACKGROUND

PD of subjective texts has been extensively studied in the literature. It is commonly known as classifying subjective opinions into positive and negative. Existing PD methods may be categorized into three main categories: supervised, unsupervised, and hybrid.

Supervised, Unsupervised, and Hybrid Polarity Detection

Supervised methods are Machine Learning (ML) approaches in which a classifier is trained based on a feature set, using labeled training data. Pang et.al. were one of the very first to perform SA on online movie reviews. They tested ML approaches, namely, SVM, MaxEnt and NB classifiers, and trained them on different feature sets including unigrams. Their findings showed that an SVM trained by making use of a unigram bag-of-words feature set, outperforms all other approaches presented in their work. Draw backs of standard ML approaches are that for opinion polarity detection they are both domain (dataset topic) and temporally (datasets collected in different time periods on an annual basis) dependent. Additionally, for classifiers to be trained on huge feature sets (e.g. unigrams) long training sessions are required and due to the huge number of features they have a large time and memory complexity.

Unsupervised Methods are mostly based on a Sentiment Lexicon (SL), in which each sentiment-bearing word is associated with either a sentiment score or a set of sentiment bearing seed words as explained bellow. These methods use different algorithms to compute a sentiment score for a given document. A SL could be automatically generated using a set of positive and negative seed-words in the literature mostly referred to as Semantic Orientation from Association; or it could be manually built. There exist a number of such manually built SLs. The SentiStrength lexicon which is specially designed for informal texts (i.e. its words' stemming rules were adopted from the social networking website

MySpace), is one example. Drawbacks of lexicon-based methods are effort for acquiring a lexicon as well as maintaining them due to language change overtime. Additionally, lexicons could also have domain dependency and the domain-independent ones might show a lower accuracy.

Hybrid methods adopt a combination of both of the above-mentioned categories to perform opinion mining. In most of such systems, a SL is used to generate features for training an ML classifier. We refer to such features as sentiment features. For instance, [1] adopts an SL to count the number of occurrences of positive and negative words in tweets and use these two measures along with other features for training their classifier. Our proposed hybrid approach is another example for such hybrid methods, which could significantly reduce the drawbacks of the two former methods. Hybrid approaches could solve most of the major problems of supervised methods, namely, the problem of huge feature sets and time and memory complexity. On the contrary, a need for domain dependent lexicons still exists. However, it appears to us that creating many domain dependent lexicons and utilizing topic detection algorithms to detect different domains and using the associated lexicon for SA of each domain may resolve this problem.

II. A SENTIMENT SUMMARIZATION SYSTEM

An aspect-based summarization system is a multi-document summarization system that summarizes documents based on different aspects of a target item. For example, an aspect-based sentiment summarizer might summarize reviews about hotels' and restaurants' reviews, based on different aspects, namely, décor, service, food, value, etc. Analogously, an aspect-based sentiment summarizer, summarizing reviews about a Smartphone might present summaries regarding the Smartphone's battery life, camera, display, etc.

Our approach to sentiment summarization consists of four main steps:

1. Finding the target item's features (e.g. battery in the context of Smartphone's) that most people have commented on in the corpus of tweets.
2. Grouping the documents containing the same item feature in the same cluster.
3. Classifying those documents using a polarity detection system.
4. Output a textual, as well as, a graphical summary, based on the classified documents.

In the following subsections we will describe each of the above mentioned steps:

A. The Aspect Detector Module

Our system can summarize documents based on the most important item features that are mentioned in multi-documents. Nevertheless, these target items' features could be provided to the system manually by plugging in an ontology containing domain vocabulary. Conversely, they could be detected automatically, however, with lower precision.

B. Documents Clustering Module

The next step is to cluster all the documents that have the occurrence of the same aspect word within them in the same cluster. The clustering module is responsible for performing this task.

C. A Hybrid Polarity Detection System

This section describes our new target-oriented hybrid system. This system is similar to that presented in [3]. However, here we propose some improvements to it that increases the classification accuracy over the previous accuracy baseline. Nevertheless, for the sake of completeness we explain the system [3]. It consists of three major modules: a preprocessing module, a lexicon-based sentiment feature generator module and finally a Machine Learning (ML) Classifier module.

The Preprocessing Module

This module performs a number of preprocessing steps as described in the following:

- @username is replaced with “ATUSER”.
- URLs are removed.
- “#word” is replaced with “word”.
- Slangs (abbreviations) are replaced with their actual phrase equivalences. A manually built slang dictionary is used for this purpose.
- Each document (or tweet) is split into smaller text snippets based on ‘.’, ‘!’ and ‘?’.
- We therefore, define a text snippet as a number of words that occur in between two of the above mentioned punctuations.
- The target (of sentiment)word is replaced by “TARGET”. This target word could be given to the system as an input query word.

Block Diagram/Architecture

The structure of the twitter analysis system is as follows:

- User: They are registered user of twitter who shares different views. The statement passed in these views needs to be analyzed. After User shares his opinion, the sentence goes to three modules for analyzing process.

- Retrieval Module: Here the sentences breaks into words for the analysis process. It acts as the first module.
- Preprocessing Module: This is the main module in analyzing process. Its main function is to filter the data present in the Retrieval module. Firstly, it removes URL such as @tags, #tags, etc. After which it removes errors. Then the filtered data enters to emoticon tagger from which the words related to opinions are extracted and forwarded to POS tagger. At the end of this filtering process a transaction file is produced where data are divided in the form of emoticons, verbs, adverbs and adjectives.
- Scoring Module: It is the third module of the process where the results of the analyzing process are produced. It is basically divided into two subparts-
 - Word Seed List: Here the scores of verbs and adverbs are produced.
 - Log Linear Regression Classifier: Here the scores of adjectives are produced
- Twitter Sentiment Scoring Module: It is the final stage of the process and the scores for opinions are generated

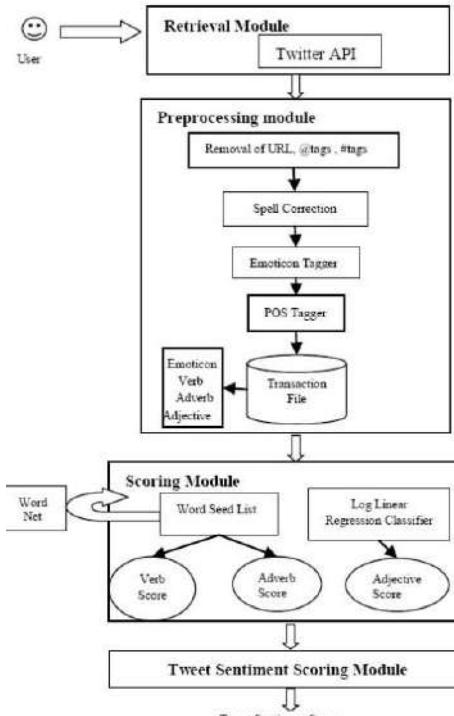


Figure: Block Diagram

Sentiment Feature Generator Module

This module starts with replacing slangs with their equivalences using a slang dictionary. To build this slang dictionary, we manually collected a slang dictionary by using as many online resources that we could find. Then, in the second step this module uses the SentiStrength lexicon to tag all sentiment-bearing words in each document with their corresponding sentiment scores. Likewise, according to a list of emoticons, it tags happy emoticons with a sentiment score of “+1” and sad ones with a score of “-1”. It further tags all intensifiers (e.g. absolutely) and diminishers (e.g. might) with their corresponding scores. Additionally, it tags negation words with “NEGATE”. Finally, if a word did not belong to any of the mentioned categories, it would be tagged with the score “0”.

After having all words in a document tagged either by their Score or type, now we should handle occurrence of intensifiers, diminishers, and negations. First, we intensify the strength of a sentiment-bearing word that appears after an intensifier, by the score of that intensifier word. Analogously, in the case of diminishers, we weaken the strength of a sentiment-bearing word that appears after a diminisher

word by the strength of that diminisher. Finally, for handling negations, we flip the polarity of the score of a sentiment-bearing word that appears after a negation. Then we weaken the flipped sentiment score by 1. That is if the flipped score is positive, we subtract it by 1 and if it's negative we sum it by 1. Note that in all cases we ignore the words with "0" tags that appear in between one of the above mentioned valence shifters and a sentiment-bearing word in a single text snippet while performing the above mentioned computations.

Table 1 presents this feature set.

f1	Document (or tweet) overall sentiment score using the unsupervised polarity detection algorithm
f2	Number of positive words
f3	Number of negative words
f4	Number of negation words
f5	Number of negation words followed by a positive word
f6	Number of negation words followed by a negative word
f7	Inverse sentiment
f8	Number of positive words followed by target
f9	Number of negative words followed by target
f10	Number of negation words followed by target
f11	Number of positive words followed by a negative word
f12	Number of negative words followed by a positive word
f13	Number of target words followed by a positive word
f14	Number of target words followed by a negative word
f15	Number of negation words followed by a positive word which is followed by target
f16	Number of negation words followed by a negative word which is followed by target

Table 1. Features used in our system

Feature f1 is an overall sentiment score for an entire document. In order to compute this feature, we aggregate the words' scores according to the tagging process explained in the beginning of the current subsection. We define the decision threshold of '0' for classifying words. That is, if the sentiment score of a word is less than '0' that word is tagged as negative, and otherwise if the score is greater than '0' it is tagged as positive. Then we compute an overall sentiment score for each document by aggregating them. If the score is equal to or greater than '0' we classify it as positive and otherwise negative.

The rest of the features mentioned in the table are mostly self-explanatory. The feature f7 named "inverse sentiment", uses some heuristics to search for patterns that most of the words in a document have the opposite polarity of the actual overall document polarity. These heuristics search for occurrence of words that often convey happening a situation or condition. For instance, the words "if", "after", "before" and "until" are among this category of words. If any of these words were present in a document, the algorithm that f7 uses, further analyze that document based on some heuristics to see if the sentiment words used in that document are meant to convey their opposite polarity. We explain the heuristic regarding "if statements" in detail. In the following, one example sentence is given:

"If I don't get an iphone for Christmas, I would be sad."

In the above example, it could be seen that the actual polarity of the sentence regarding "iPhone" is positive, whereas, the sentence only contains words and patterns that usually occur in a negative context. "don't get an iphone" is a negative statement regarding iPhone and its sets f10 to "1". Similarly, the word "sad" is a negative word and its occurrence sets f3 and

negation words are tagged with "Negation".

f14 to "1". These features are often more likely to be present in negative sentences. However, because of the existing patterns, the inverse sentiment feature (f7) is also set to "1". The heuristic searches to find out if both the "if clause" and the "main clause" in the example sentence, include a negative sentiment pattern and if so it detects this pattern as an inverse sentiment. Thus, our system can easily detect and correctly classify most such cases, while for many existing SA systems handling such cases is a challenge.

Machine Learning Classifier

The machine learning module is a linear SVM that takes as input the feature set described in Table 1 and accordingly classifies the tweets (or documents) to separate classes.

D. A Textual and Graphical Summary Generator Module

In this subsection we present our methods for generating graphical as well as textual summarizations.

Non-Textual summarization

After having the classified documents, we can easily tell what percentage of documents are positive or negative regarding each attribute of our target item. Therefore, in such scenario, we have already come up with a numerical summary of the entire corpus regarding most important attributes of our target item.

III. EVALUATION

Our dataset consists of 940 tweets annotated by a group of 22 human annotators from which 470 have a positive polarity and 470 have a negative polarity. We gathered our dataset by making use of word spotting based on occurrence of the word "iPhone". To comply with the gold standard, each tweet went through a 3 persons voting process. That is, out of a collection of tweets collected from Twitter that all contained the word "iPhone", each tweet was read and labeled (positive or negative) by three people. Then for each tweet, if at least two out of the three people had agreed on the same class label, that tweet was accepted to have that class label, and otherwise it was rejected from our data set. As explained, challenges of such dataset are the presence of a target word toward which SA must be done, presence of comparison between two different Smartphone's, and informality of text statements.

A. Evaluation Metrics

We evaluate all the methods presented in this paper using accuracy on single class and overall accuracy as presented in the following:

$$\text{SingleClassAcc.} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{OverallAcc.} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (3)$$

Where TP, FP, TN, and FN are the number of true positives, false positives, true negatives and false negatives.

Furthermore, for testing any of the supervised classifiers as well as our hybrid method we use 10-fold cross validation.

B. Experimental Results

Polarity Detection Algorithms

In this subsection we present the results of our experiments. Reference [3] conducted an experiment, testing the effect of presence or absence of emoticons and stop words in the unigram feature set to train and test SVM, Naïve Bayes (NB), as well as, Maximum Entropy (MaxEnt) classifiers on the same dataset described in the beginning of current section. By this experiment they showed that an SVM trained on a unigram feature set with emoticons and stop words included, outperformed all other combinations with an overall accuracy of 86.70% and thus, it is a base line for our experiments.

IV. CONCLUSION

The rise of micro-blogging has changed the means of communication nowadays. The large amount of information about many different consumer products on Twitter has made this website an attractive source of data for opinion mining and SA of public opinion about many different entities including various products as well as technology related products. In this project, we have built a sentiment summarization system by combining various methods and algorithms, in order to summarize opinionated texts in the domain of consumer-products. We showed that our proposed target-oriented hybrid method outperforms the unigram (supervised/ unsupervised) algorithm baseline. Additionally, we have illustrated that our unsupervised PD and ranking algorithm could be utilized to generate extractive summaries of a corpus of multi-documents for an efficient aspect based summary of data. According to our experiments we believe that in the context of SA, moving towards sentiment features that require cognitive thinking rather than conventional text processing features would be a promising solution to this problem. Detection of sarcasm, as well as, finding more features that could semantically interpret a document are also implemented with the most accuracy provided till date. Additionally, building a multi-domain context dependent lexicon would be a future direction of this project

V. REFERENCES

- [1] Hogenboom, A.; van Iterson, P.; Heerschap, B.; Frasincar, F.; Kaymak, U., "Determining negation scope and strength in sentiment analysis," Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on , vol., no., pp.2589,2594, 9-12 Oct. 2011
- [2] Lai, P., "Extracting Strong Sentiment Trends from Twitter", Jan 2011.
- [3] O'connor, B.; Balasubramanyan, R.; Routledge, B.; Smith, N. From "Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series." International AAAI Conference on Weblogs and Social Media, North America, May 2010.
- [4] Pak, A., and Paroubek, P., "Twitter as a corpus for sentiment analysis and opinion mining" in Proceedings of LREC, May 2010.
- [5] "Sentiment analysis and summarization of twitter data" [online]. Available: <http://ieeexplore.ieee.org/document/6755222/>
- [6] "Sentiment Analysis on twitter" [online]. Available: <https://www.ijcsi.org/papers/IJCSI-9-4-3-372-378>
- [7] "Semantic pattern for Sentiment analysis of twitter" [online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-11915-1_21
- [8] "Sentiment analysis of twitter data using Machine learning approaches and semantic analysis" [online]. Available: <http://ieeexplore.ieee.org/document/6897213>
- [9] "Semantics pattern for sentiment analysis on twitter" [online]. Available: https://www.researchgate.net/268133308_Semantic_Patterns_for_Sentiment_Analysis

FESCCO: Fuzzy Expert System for Career Counselling

Prof. Megha V Gupta Computer Engineering NHITM Thane, India meghavgupta@outlook. com	Prof. Pranali Patil Computer Engineering NHITM Thane, India pspatil289@gmail .com	Sanika Deshpande B.E. Student Computer Engineering NHITM Thane, India ndsanika@gmail.com	Sravani Arisetty B.E. Student Computer Engineering NHITM Thane, India sravs.dec6@gmail. com	Shraddha Asthana B.E. Student Computer Engineering NHITM Thane, India shraddha596@gmail. com
--	--	---	--	---

Abstract: Artificial intelligence involves two basic ideas. First, it involves studying the thought processes of human beings. Second, it deals with representing those processes via machines (like computers, robots, etc.). Artificial intelligence (AI) technologies and techniques have useful purposes in every domain of mental health care including clinical decision-making, treatments, assessment, self-care, mental health care management and more. This application involves an AI based fuzzy expert system which helps the students to give a basic idea or insight of possible career opportunities, to enable them to move forward towards the path most suitable for them in all respects. This project will give a personal aid to the students taking into consideration, the student's interest and aptitude test result. The fuzzy expert in our project will choose accurate careers for the user accordingly.

Keywords: *Fuzzy System, career, expert system, aptitude, interest*

1.1 Introduction

In today's growing world of opportunities, it is important to have acute knowledge of the various careers available across the globe, to be able to select careers that are either high in demand or suit a person's personality.

A staggering 9 out of 10 people aged between 21 and 65 say they regret rushing their career choices, with many picking a university course at random because they simply don't know what they want to do and feel pushed to decide because they are 'running out of time'. Over half of the 3,000 students that responded to a GTI Media survey said their parents tried to exert influence over their choice of career or course. It, therefore, comes as no surprise that 20% of students currently enrolled at university say they would have chosen a different course if given the chance, with 18% saying they regret their choice of degree. The same study found that 18% of the 1,805

respondents cite a lack of initial research as the main cause of their disappointment. As a result of wrong career choices, China and India reported their unemployment rates in July and September this year at 4.7% and 4.3% respectively, as work-satisfaction is not achieved by them in a field that they didn't happily choose for themselves. These rates are approximately equal to 5.8 million, each, of Indians and Chinese (citizens). Beyond the financial impact there is to make a poor career choice, many scientific studies show that the wrong career can make you seriously ill - the total number of cases of work-related stress and depression in 2015/2016 was 488,000 with a prevalence rate of 1,500 for every 100,000 workers. Career guidance for students, particularly in rural areas is a challenging issue in India. In the present era of digitalization, there is a need of an automated system that can analyze a student for his/her capabilities, suggest a career and provide related information.

For selecting the right career, we must bring down our list of appropriate careers to a selected few, which can be done by a juxtaposition of one's abilities and areas of interest. To overcome this, Artificial intelligence (AI) can be highly useful. Artificial Intelligence is the field within computer science that seeks to explain and to emulate, through mechanical or computational processes, some or all aspects of human intelligence. Included among these aspects of intelligence are the ability to interact with the environment through sensory means and the ability to make decisions in unforeseen circumstances without human intervention.

Expert systems are a well-known area of Artificial Intelligence which use human knowledge to solve problems that normally would require human intelligence. Education system will be revolutionized with the introduction of expert systems in this field because of the following:

- (1) Educational planning and decision-making ability, manage student records, student counseling, and special education programs.
- (2) Teacher training and education, specifically, identifying training needs and using computer assisted instruction to teach information and skills.
- (3) Intelligent tutoring systems that guide students through instruction according to their individual strengths and weaknesses.

Artificial intelligence has a form called fuzzy expert system that uses a collection of membership functions (fuzzy logic) and rules (instead of Boolean logic) to reason about data. A fuzzy expert system is an expert system that utilizes fuzzy logic as the paradigm to express rules and thus uses a fuzzy inference engine to reason about this type of rules. Fuzzy logic uses a scale for degrees of truth that range between 0 and 1 rather than the typical Boolean logic which uses either the 0 or 1 value to describe false and truth.

2.1 Study of existing system.

This paper suggests a fuzzy based approach towards this issue. It has two parts; in the first part a student will be analysed for his/her capabilities and in the second part the available courses, job aspects related to their capabilities will be suggested. To analyse a student, marks in various subject in 10+2 standards and vocational interest in different fields have been considered and fuzzy sets have been formed. On example basis, fuzzy inference rules have been framed for analysing the abilities in engineering, medical and hospitality fields only. In second part, concept of composition of relations has been used to suggest the related courses and jobs. [1]

The above paper is limited, and analyses abilities based on aptitude only in the fields of engineering, medical and hospitality. It also does not consider the interest of the individual towards a subject/field while suggesting courses.

In [2], an Intelligent Algorithm is implemented to design an expert application on smart phone. In this research, the expert system that will aid the student in his registration decision is designed as a fuzzy expert system and implemented as a mobile application that runs under the Android operating system.

Here the proposed system uses six inputs and produces a single output; these inputs represent fuzzy linguistic variables that can be set to certain fuzzy values. Table 1 lists the inputs and output and their corresponding values. Input variables and values were chosen based on an actual survey conducted on college students. The system uses Mamdani-style

inference to reason from several fuzzy rules that represent the knowledge base.

Table 1: Variables and values [2]

	Variables	Symbol	Values
1	Perceived teaching efficiency of lecturer	A	inefficient, average, efficient
2	Past performance	B	bad, good, very good
3	Perceived difficulty of course	C	hard, average, easy
4	Appeal of course topic	D	unlike, average, like
5	Friend in course	E	none, some, all
6	Cost of course	F	high, moderate, low
7	Recommendation of registering the course	R	high, average, low

The above system suggests various courses based the ratings given to it by the students. It does not take into consideration the individual while suggesting the courses. The implementation of the fuzzy system was understood from this paper.

In [3], Career Advisor Expert System Based on Myers Briggs Personality Assessment is discussed. It advises the user based on his/her personality. This is achieved through the method of creating facts from the Myers-Briggs Type Indicator (MBTI) thereby mapping them to common Careers using a rule-based system based on the sixteen Personality Types (according to Myers-Briggs). It uses prolog for implementation.

Table 2: Key domains of Myer Briggs Personality indicator [3]

I. HOW WE PREFER TO DIRECT OUR ATTENTION AND ENERGY?	
1. Extraversion (E)	2. Introversion (I)
II. HOW WE PREFER TO OBSERVE THE WORLD?	
3. Sensing (S)	4. Intuition (N)
III. HOW WE PREFER TO MAKE DECISIONS?	
5. Thinking (T)	6. Feeling (F)
IV. HOW WE PREFER TO ORIENT OURSELVES ON LIFE?	
7. Judging (J)	8. Perceiving (P)

The table 2 shows four key domains of Myer Briggs Personality indicator. Careers are suggested according to the 8 personality types shown above. It only focuses on the personality types and not the abilities of an individual.

The paper [4], presents the design of a multi-expert system for educational and career guidance based on a multi-agent paradigm and the semantic web. It uses ontology for the decision making of careers. The use of ontology allows to identify the different guidance concepts and semantic links between them, to establish a better representation of the existing while staying within the context of sharing and reusing of knowledge. The proposed system uses the following experts: The psychologist, sociologist, pedagogue, economist, Coach, supervisor. It is a multi-expert system supervised and organized around a multi-agent system, which is based on the n-tier model of the web application. The presence of many agents makes the system very complex. And since it was a web application it was not easily accessible like a mobile app.

This section talks about Dr. John Holland's theory that has been extensively tested in the United States and many other countries, summarized the process of career choice in the following statements:

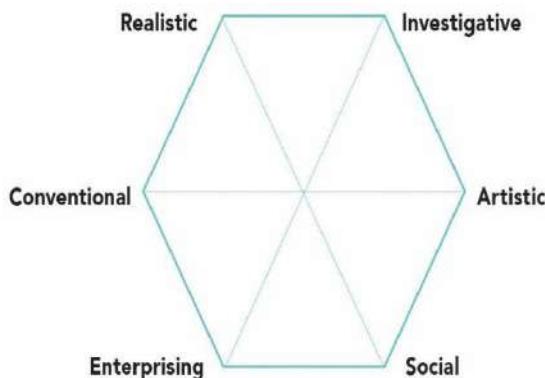


Figure 1: Holland Codes [6]

People can be described as a combination of two or more of six personality types, which he titled Realistic, Investigative, Artistic, Social, Enterprising, and Conventional. One's personality type can be measured by taking a reliable, valid assessment whose items are based on this theory.

Occupations, work environments, and school courses can similarly be described as a combination of two or more of the same six "personality" types.

Individuals of a given type seek environments of the same or highly similar type.

To the extent that individuals can enter educational or work environments of the same type as their own, it is likely that they will be satisfied and persist in them. [5]

Career guidance appeared alongside advances made in the social sciences in the early 1900s. Frank Parsons, author of the 1909 work "Choosing a vocation" and who was also part of the development of "vocational bureau" in Boston, USA, was one of the founders. Parsons states that occupational decision making occurs when people have achieved:

- An accurate understanding of their individual traits (aptitudes, interests, personal abilities)
- A knowledge of jobs and the labour market
- Rational and objective judgement about the relationship between their individual traits, and the labour market. [6]

3.1 FESSCO

To build a fuzzy expert system whose role should be to recommend various successful career possibilities to an individual based on various aspects of the individual given as input. The system should use aptitude and achievement assessments, to help clients evaluate their interests, skills, and abilities. The system should be capable of deducing realistic career possibilities based on aptitude and interest analysis. The application must enable the user to explore various existing career options and their scope. It should also enable the user to take tests and give a detailed report about the suitable careers in order of preferences. The system should display the list of renowned institutes for the suggested careers.

3.2 System analysis

System analysis report for FESCCO contains input given by user and output given by system. It consists of following subsystems:

Expert System:

An expert system is a computer system that emulates the decision-making ability of a human expert. Knowledge base and Inference engine are two integral parts of an expert system.

• Fuzzy Logic Controller:

Fuzzy logic is an approach to computing based on "degrees of truth" rather than the usual "true or false" (1 or 0) Boolean logic on which the modern computer is based. Fuzzy Logic Controller mathematical system that analyses crisp input values in terms of logical variables

that take on continuous values between 0 and 1.

- **Inference Engine:**

Inference engine is a component of the expert system that applies logical rules to the knowledge base to deduce new information.

- **Knowledge base:**

A knowledge base is a technology used to store complex structured and unstructured information used by the expert system

3.3 FESCCO Overview

The FESCCO system will be fuzzy expert system. The system will have properties of a fuzzy logic as well expert system. A fuzzy expert system is an expert system that utilizes fuzzy logic as the paradigm to express rules and thus uses a fuzzy inference engine to reason about this type of rules. Fuzzy logic uses a scale for degrees of truth that range between 0 and 1 rather than the typical Boolean logic which uses either the 0 or 1 value to describe false and truth.

The first step for the user will be to register. The user will have to enter detailed information about himself. This will include Name, email ID, password and details about education required for further analysis. Once registration is done user can login using his email ID and password. This will take the user to his home page. The home page will contain the profile of the user along with various options. The user can explore various career options available. He can also take tests after which a personalized report for the user will be generated.

The tests will be into sub parts:

- 1) **Interest Analysis:** This will have questions which will analyses the fields of interest of the user based on the answers. These questions will have options based on levels of whether he agrees/disagrees to the given statement or has a strong to weak interest in a field. The answers will be analysed by the fuzzy system.
- 2) **Aptitude Analysis:** This will contain questions based on aptitude. It may further be divided in several sections like analytical, verbal, logical, etc. The results of these sections will show abilities of the user.

The results of both the tests will be combined and possible successful careers for the user will be displayed. There will be more than one careers displayed in order of preferences. The institutions for the given careers will also be displayed.

3.4 Proposed Design/Architecture

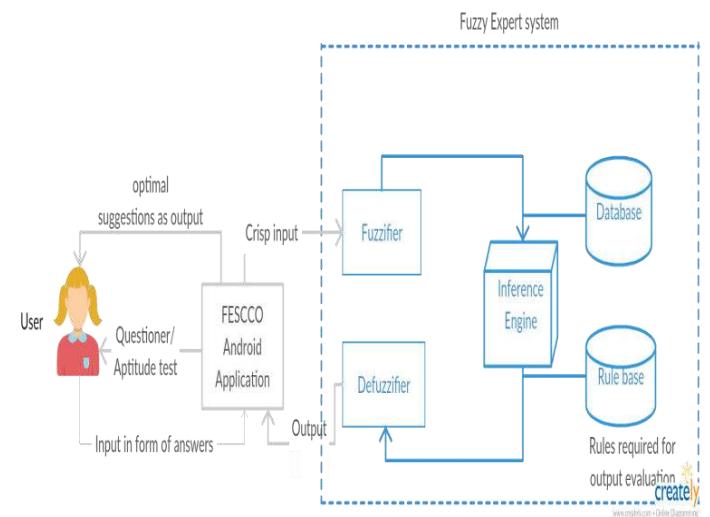


Figure 2: Proposed System Architecture

The system will consist of following components:

- 1) User: The user for the can be anyone who wants to seek advice on his/her career. The user must register after which he can login and start using the app. User can explore various options and take test to generate appropriate career options.
- 2) FESCCO Android App: It will act as an interface for the user. It will allow user to take tests and display results.
- 3) Fuzzy Inference System:
 - a) Fuzzifier: Converts the crisp input to a linguistic variable using the membership functions stored in the fuzzy knowledge base.
 - b) Inference Engine: Using If-Then type fuzzy rules converts the fuzzy input to the fuzzy output.
 - c) Defuzzifier: Converts the fuzzy output of the inference engine to crisp using membership functions analogous to the ones used by the fuzzifier.
- 4) Databases:
 - a) Rule base: It contains rules for the decision-making process of analysing interests and career choices.
 - b) User Database: Contains user details and test status of the users.
 - c) Test Database: Contains test questions and test details.

4. Applications

Assessing to deduce an **optimal** career option: There are uncountable career choices in the world. But, there are a finite number of career options that are suitable for everyone based on his own personality, IQ, grades, interests, preferences, etc. The processing of all answers given by the user in these aspects is done in the most precise form to deduce an optimal career path that is suitable or rather best for the user, based on his skills and capabilities that are assessed at the moment.

Available and accessible by students at any time, any day and at their convenience: Students have doubts or questions about which career to choose, what passion to pursue, and how beneficial can any career choice be for them and for the surrounding environment. These doubts can arise as random times of the day like 8pm, 2am, or at any day including a public holiday when relatives come over to meet them, and question them about their future goals. Meeting a user's requirement anytime anywhere is the best service that can be offered by an application, and FESCCO aims at achieving it.

This information system can serve as a complementary tool for real life counsellors: Not all counsellors have knowledge of each existing field in the country, let alone in the world. The application of FESCCO's Database is to reach out as an assistant to real-life counsellors while they assess a student based on their desired set of tests, and suggest him a career that he has the potential to excel in.

Provides guidance tips to all kinds of students: Not only is this app useful for students who are confused whether to choose Science, Commerce or Arts, and not only for the students who are confused in which specialization to take further in the above branches, but also for the people who have taken a year drop, or who wish to switch their field of studies, or who just simply wish to understand their own selves better in terms of their IQ, interests, potential, scope in a certain field, etc.

Explore all available career options: For curious and confused people alike, enlightening them with all career options available is a good way to explain which fields require what education, and thereby enable a user to view a career option not only suitable for him but also other options he could opt for. Also, exploring all career options is a major advantage to the real-life counsellors, as it gives them a platform to view all careers systematically, without feeding in details of each student

5. Conclusion

Our FESCCO system will establish an automated process like a one-to-one meeting with a career counsellor and will aid to 'plan' a career true to the student's grade, IQ, hobbies, interests, and other predominant specifications entered by the user at the time of registration. We aim at achieving optimum results in the time we have, as we believe that this app will be a boon to all those people out there who are at their major crossroads in life, striving to know in which stream and domain their strengths lie, and what all career options are awaiting them.

References

- [1] Raj Kishor Bisht "A FUZZY BASED CONCEPTUAL FRAMEWORK FOR CAREER COUNSELLING", Advanced Computational Intelligence: An International Journal (ACII), Vol.2, No.4, October 2015
- [2] Walid Mohamed Aly, Khaled Ahmad Eskaf, Amir Serry Selim "Fuzzy Mobile Expert System for Academic Advising", 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE).
- [3] A. Iwayemi, B. F. Oladejo, D. S. Adeleke "Career Advisor Expert System Based on Myers Briggs Personality Assessment" Department of Computer Science, University of Ibadan, Ibadan, Nigeria.
- [4] Essaid EL HAJI, Abdellah AZMANI, Mohamed EL HARZLI "Multi-expert system design for educational and career guidance: an approach based on a multi-agent system and ontology", IJCSI International Journal of Computer Science Issues, Vol. 11, Issue 5, No 2, September 2014
- [5] JoAnn Harris-Bowlsbey, Ed.D. Kuder Research Faculty "Overview of Career Guidance: Its Foundations, Objectives, and Methodology", White Paper, Kuder.
- [6] Samuel T. Gladding Counselling- A Comprehensive Profession; Pearson Publication, Seventh Edition;

Application of Big Data in Disease Diagnostics- Prospects and Challenges.

Prachi Dalvi¹
Assistant Professor
TCET-Mumbai

Anupriya Babbar²
Assistant Professor
TCET-Mumbai

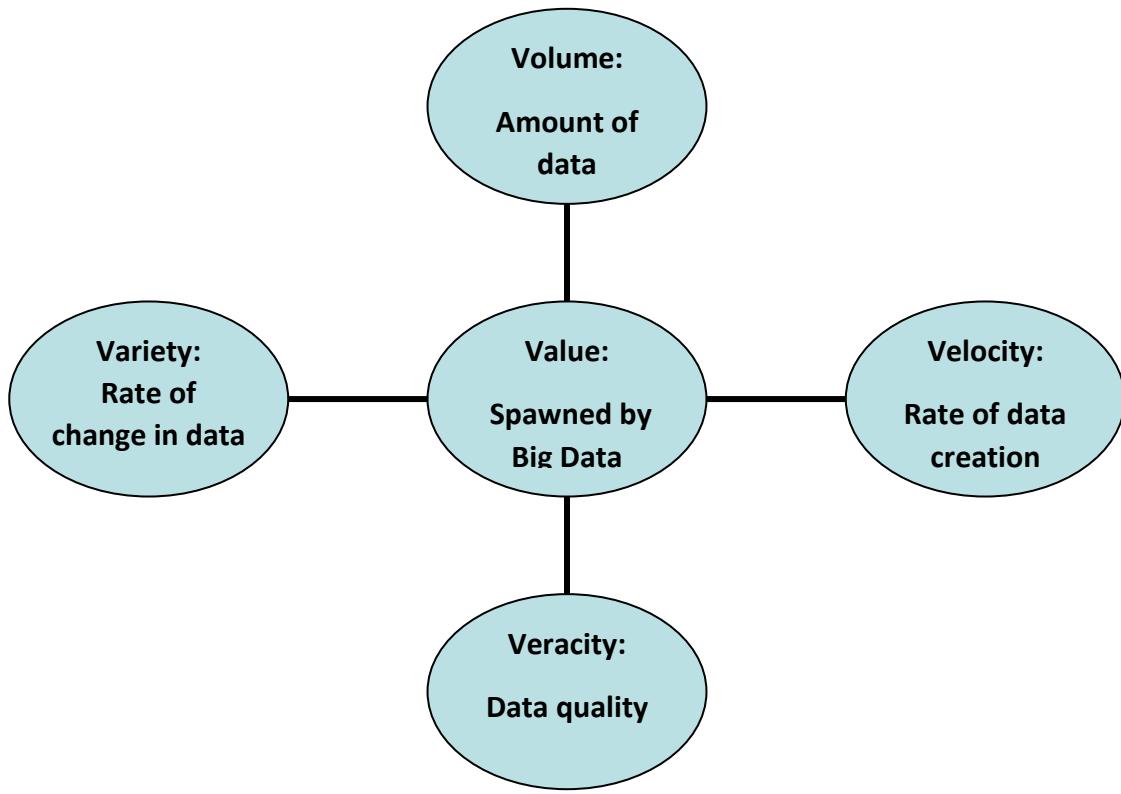
Abstract- In the recent developments of advanced database technologies, in which vast amount of data is created every second and is placed in database for analysis. The capability to analyze the data on various diseases will be extremely helpful in escalating the boundaries of what can be gained from current diagnostics models to analyze particular pattern of disease growth. In this paper, we focus on main areas where big data can be applied for disease diagnostics and better results can be achieved by applying this technology.

Keyword- Big Data, Map-Reduce

Introduction- Medical science has advanced a lot since a decade, lots of new machineries and technologies were introduced to examine. Large amount of data is collected, then what? Not only has the volume of data increases exponentially, but also the type and nature of this data has vividly evolved. This data can be used as an affirmation to examine the disease patterns and irregularities so that treatment quality can be improved on the basis of evidential proofs in the form of large volume of up to date data. Big data allows to analyze structured, semi- structured and un-structured data and predict the patterns and irregularities of disease which can play a crucial role in the field of disease diagnostics.

Why Big Data?

Big Data is known for five V's... Volume, Velocity, Veracity, Variety and finally the most important is the Value. It can improves the outcomes and reduce the cost. The collecting, storing, analyzing and retrieving of data for decision making are rapidly reduces the paper work towards the automation. Data generated in mass volume are complex to examine by outmoded approach. Backers engenders data from various assorted sources such as clinical observations, lab tests, therapies, surveys, insurance and bills, all these data is in diverse forms. Collation and investigation of such data is a tedious task which can be overtaken by Big Data to produce the prolific results.



Implementation of Big Data in disease diagnostics-

1. Source of big data in disease diagnostics are as follows:

A. Machine engendered data:

The data are generated from various machines that are used in the diagnosis of disease, example comprises of **Android Sleep App**(this useful app can pair with your sleep tracker and wake you up at the best possible time so you're rested and don't feel groggy at all.), **Pebble Time**(measuring your sleep is only one of the built-in health tracking applications), Fitbit Surge(This touch-screen wristwatch not only tracks your steps and sleep, but also alerts you to incoming phone calls and text messages, keeps tabs on your heart rate with a built-in optical heart rate monitor and uses GPS to track outdoor activity.), **PIP** (a tiny device designed to give immediate feedback about your stress levels.)

B. Human engendered data:

This type of data comprise of data which is generated by human either in the form of clinical observation notes, summaries of hospital admission, discharge or in the form of electronic health record.

C. Transaction engendered data:

This data comprise of records related with insurance, claims or billing.

D. Observation engendered data:

This data comprise of notes or records which are generated from various interpretations.

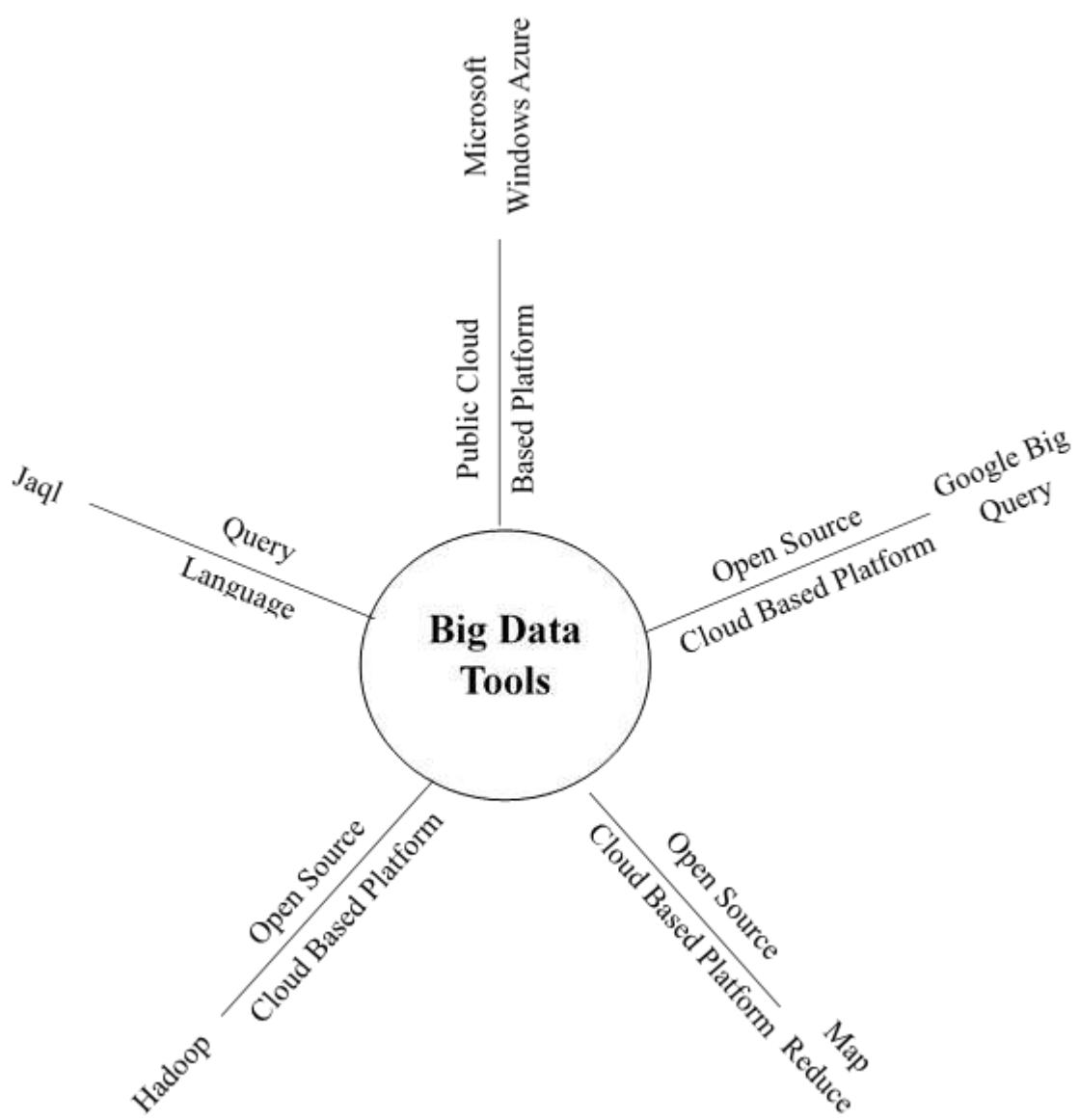
E. Miscellaneous data:

This data can be derived from various sources such as statistical data, countless surveys, researches and reference materials.

2. Process:

This huge amount of data which is doubled or tripled in every minute can be in any format. Big data tools such as Map-Reduce, JAQL, Hadoop and Google Big query are beneficial to drill out the value from gigantic data.

2.1 Tools and Technologies for processing Big Data in disease diagnostics:



3. Analyze and Extract:

Big data is an analytical tool to drive out the results from massive data but it must possess finely developed user interfaces. Analyzed data gives the information in the form of values, these extracted values can further be utilized for many applications such as disease diagnostics, health care, disease control and many more.

Benefits of Big Data in disease diagnostics:

In today's world, disease diagnostics system has perceived rapid change which comprise of increase in the use of electronic apparatus to improve the cure methods. Though, increase in volume of data has made it difficult for traditional applications to meritoriously retrieve the useful data. It is the obligatory to cognize the importance of big data in disease diagnostics. Use of Big Data make the complete conversation interactive between the patient and the doctor, patient is also completely involved. This system is completely evidence based system as treatment of patient depends largely on available scientific evidences on the contrary, it reduces the cost of system up to an extent as reports can be directly sent to the smartphones of the stake holders. It also improves, public health surveillance as it can predict disease patterns, track disease out breaks and transmission. Big Data guarantees the premature diagnosis of **disease**, **it ensures the correct methods of treatments for the particular disease based on value**.

Conclusion: In today's scenario, with the increased of stress level, peoples are highly cognizant towards their health and wishfully wants to pay towards the health care. In this paper, we made an efforts to express the various benefits that medical system can make in disease diagnostics by using Big Data Technology.

Capabilities of Big Data analytics allow health system for the early diagnosis of disease, better cure methods and also provides surveillance and privacy for tailored treatments.

This papers provides an insight for the forthcoming researchers to know the influence of Big Data in disease diagnostics. Since it is a growing area, researchers has an opportunities to come up with new ideas.

REFERENCES

- [1] Hilbert & López 2011 "IBM What is big data? — Bringing big data to the enterprise". www.ibm.com. Retrieved 2013-08-26.
- [2] <http://whatsthebigdata.com/2012/06/06/a-very-short-history-of-bigdata/>.
- [3] How President Obama's campaign used big data to rally individual voters. By Sasha Issenberg on December 19, 2012, MIT Technology Review. <http://www.technologyreview.com/featuredstory/509026/howobamas-team-used-big-data-to-rally-voters/>

[4] <http://www.forbes.com/sites/kashmirhill/2012/02/16/how-targetfigured-out-a-teen-girl-was-pregnant-before-her-father-did/>

Digital Marksheets generator by using QR code

Abrar Mungi, Mehfooz Ali Sayyed, Mustafa Khan, Mohd Saad Shaikh

*Anjuman-I-Islam's Kalsekar Technical Campus,
Department of Computer Engineering, University of Mumbai*

Abstract— In today's digital world we have often hear the news about fake marksheets and unauthenticated certificate, so it is a big challenge to provide security and authenticity of digital data. Our aim is to provide a digital marksheet which can't be modified from third party and also user can access their marksheets in any smart device. For that we are using encryption and decryption technique with the help of QR code based system.

Keywords: QR code, Encryption, Decryption

1. Introduction

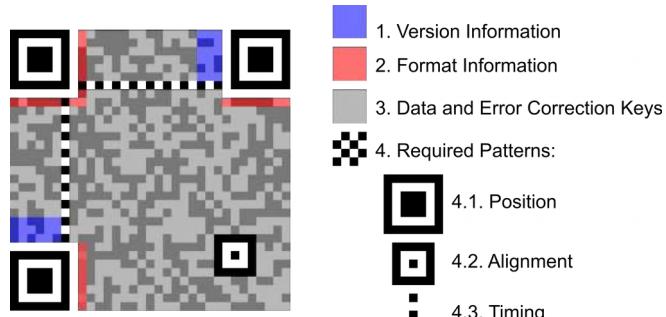
As we know that in this era we are moving towards digital world, so many of the people are using internet for storing the data on the cloud so that data can be accessed efficiently from any corner of the world. Also many of the people avoid carrying the hard copies of their documents wherever they are going, as taking care of these documents is a tedious task for them. And due to this, there may be a chance of getting lost of their documents which will be probably more difficult for them to get these lost documents back again, as they have to follow some procedures to get their lost documents back. So to avoid all these stuffs, we mostly used to put it all on the internet like google drive or cloud etc, we just scan our certificates and useful documents and kept it on any of the cloud as these documents can be used in emergency work. And nowadays we often hear about fake marksheets as many of the unauthorized users are creating the certificate with their own name and if someone wants to verify that certificate then long procedure she/he have to follow like first write an application to the University or the Institution in which certificate number is mentioned and then the University or Institution will search for that certificate by its

certificate number and if they found then cross verification will be done and if they didn't find then that particular certificate is said to be fake as its details are not available with University or Institute in their databases. So now the question is how to get rid from all these things. For that reason we are proposing a system to overcome all these things, the system will provide digital marksheets with the help of QR code i.e. user can view marksheets by just scanning the QR code of that particular marksheets. QR code used to store very less amount of data in it around 4KB, so to store whole marksheets in QR code we will use some compression techniques. Generally QR code consists a machine readable code consisting of an array of black and white squares, typically used for storing URL or other information for reading by the camera on a smartphone.

1.1 What is QR code?

A QR code (abbreviated from Quick Response code) is a type of matrix - barcode (or two-dimensional bar code) [5]. QR Codes were introduced by the Japanese corporation Denso Wave Incorporated[10]. Originally it was designed to be used to track parts in the vehicle manufacturing industry, but its use has since grown tremendously[8]. QR codes are increasing their popularity as they appear more place in the Urban environment. Apart from marketing, QR Codes have been also adopted in different areas such as the on-line payments. The use of QR codes is license-free. A bar code is 2 dimensional code that contains data in only one direction but QR Code contains information in both the vertical and horizontal directions therefore QR code can hold considerably greater information than a bar code[11]. These Quick Response (QR) codes are

versatile. In this new mark-sheet system, we embed a QR code in the mark sheet which encodes the mark detail of the student, grades obtained, student's name, roll number, year, semester etc. So any HR or any employer who needs to verify the marksheet can just take his mobile phone and use the QR code reader application which is available for free of cost and scan the code. If the student or alumni wants to see his marks digitally or wants to send his own academic data to any other educational institute in digital format, then he has to just scan the QRCode, decrypt the data and send the information.[9] This system can easily be implemented and it is virtually impossible to alter the QR code that is embedded within a marksheet. So it increases the complexity of the forget process and hence it is computationally secure as per the laws of security. The system ensured several layers



In the above figure, the three prominent boxes in the Corners of each node indicates the location of the image(so it can be centered by the image parser). Another smaller box in the lower right aligns the image. The rest of the blocks indicate the information specific to the international standards and finally the identifying data contained within the code[7][6].

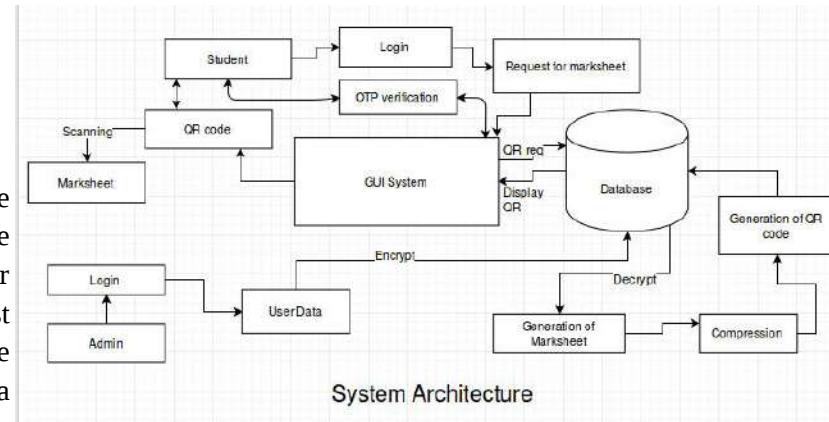
1.2 What is Digital Marksheets?

Digital Marksheets are softcopies of the marksheets which are generated by scanning the QR code. Digital Marksheets consist of all the information of the student like student's Name, Roll number, Marks, Percentage, etc.

1.3 Need of Digital Marksheets:

In today's world we are often hear the news about fake marksheets and unauthenticated certificates so that we should have such system in which the modification can't be possible. And also if someone wants to verify whether the marksheet is genuine or fake then instead of writing applications to Universities and Institutes, we just have to scan the QR code which is there on the marksheet. If it displays the same marksheet after scanning the QR code then it means that the marksheet is genuine and if it doesn't display the marksheet or not displaying the same marksheet then it means that this marksheet is not genuine. So this saves our time for verification process. And also students did not have to stand and waste their time in queue for getting the marksheets from the exam cell instead they can just go to this system's website and from there they can get the QR code through which they get their marksheet's softcopy after scanning it.

2. System Architecture:



There are two Ends of the system:

1. Admin End.
 2. User End.

2.1 Admin End:

All the back-end work is done by the Admin. Admin will login with its credential. After login Admin will enter the

information of students which will be get stored into the database in an encrypted form and by using these information marksheets will be created by the system itself. System will generate the QR code for particular marksheets of the student and keep the QR code stored in the database.

2.2 User End:

Students will enter their information such as Name, Seat Number and semester then OTP verification will be done. After successful verification, system will display the QR code. As the student scan QR code, marksheets will be displayed and student has two options, either to save the marksheets or to print it.

3. Encryption:

The data entered by the admin get encrypted and saved into the database using PHP's encryption function `mcrypt_encrypt()` followed by `mcrypt_rjindael_256` method (AES encryption) and result will be encoded into base64 by using `base64_encode()` function for providing more security to the data.

4. Decryption:

The data which is encrypted in the database is get fetched first from the database and get decrypted using PHP's decryption function `mcrypt_decrypt()` followed by `mcrypt_rjindael_256` method (AES decryption) and result will be decoded into base64 by using `base64_decode()` function after that this decrypted data is used for generating the softcopy of the marksheets.

5. Compression:

The marksheets softcopy will be compress using JPEG image compression technique as the image of the marksheets will be in large size.

7. Literature Review:

Sr No	Paper Publication	Author	Title	Published Year	Content	Demerits
1	IEEE	Somdip Dey Steanne Solution Ltd, UK	New Generation of Digital Academic-Transcripts using encrypted QR Code TM Use of encrypted QR Code TM in Marksheets (Academic Transcripts)	26 December 2016	Digitize the mark-sheets in the form of encrypted QR Code TM using TTJSAs encryption technique.	It is only encrypting the tabular content of the marksheets which is in black & white
2	DRCSIT	Madha v V Dahigondne, Vinod Kadam Dr. Babasaheb Ambedkar Technological University	QR Code Based Digitized Marksheets System	26 February 2014	Encrypting the marksheets in the form of QR code and also decrypting it using DES algorithm (56 bits).	It is not storing the data in an encrypted format into the database just encrypting the data in the QR code format.
3	International Journal of Applied Engineering Research	Delphin Raj K. M and Nancy Victor	Secure QR Coding of Images Using the Techniques of Encoding and Encryption	12 November 2014	Converting marksheets image into QR code using conversion, compression and encryption technique.	It is difficult to implement.
4	International Journal of Cyber-Security and Digital Forensics (IJCSDF)	Somdip Dey Steanne Solution Ltd, UK	An Image Encryption Method: SD-Advanced Image Encryption Standard: SD-AIES	2012	Image encryption using SD-AIES	It is just encrypting the image and making the image in an unreadable form.

Conclusion:

We have selected this topic because Marksheets is a sensitive document and to preventing it from getting misused by the unauthorized user. As distribution of the Marksheets takes too much time and maintaining these mark sheets is also a big deal so to overcome all these problems this system is proposed.

Acknowledgement:

We would like to thank our guide Prof. Mubasshir Khan and Prof. Apeksha Gopale for their great support for making this paper and guiding us in the right direction.

References:

- 1) New Generation of Digital Academic-Transcripts using encrypted QR Code TM Use of encrypted QR Code TM in Mark-sheets (Academic Transcripts). IEEE.org (26 December 2016).
- 2) QR Code Based Digitized Marksheets System IJRCST.com (26 February 2014).
- 3) Secure QR Coding of Images Using the Techniques of Encoding and Encryption. IJAER. (12 November 2014).
- 4) An Image Encryption Method: SD-Advanced Image Encryption Standard: SD-AIES. IJCSDF (2012).
- 5) Vinod J. Kadam, S.P.Vaidya, "Bar Code Technology for Library - SWOT analysis, Strategies, Survey, Cross-case analysis and Application: A Case study of Dr.Babasaheb Ambedkar Technological University", International Conference on Science and Technology (ICST-2K14),2014,Pune.[accepted]
- 6) [http://en.wikipedia.org/wiki/Reed
%E2%80%93Solomon_error_
correction](http://en.wikipedia.org/wiki/Reed%E2%80%93Solomon_error_correction) (Retrieved 2014-01-02)
- 7) Jimmy Rogers,What Are QR Codes? Digitize your World & Back Again (online) <http://www.makeuseof.com> 27th October, 2008
- 8) <http://www.whatisaqrco.de.co.uk/> (Retrieved 2014-01-02)
- 9) Somdip Dey,"New Generation of Digital Academic-Transcripts using encrypted QR CodeTM "IEEE 2013 International Multi Conference on Automation, Computing, Control, Communication and Compressed Sensing (iMac4s 2013), Kerala, India, March 2013
- 10) Cryptography and Network Security, William Stallings, Prentice Hall of India.
- 11) A. Sankara Narayanan,QR Codes and Security Solutions,International Journal of Computer Science and Telecommunications [Volume 3, Issue 7, July 2012]

Biometric Keyless Car Unlocking System using Smartphone and Raspberry Pi 3

Prashant H Rathod ,prashantrathod14@gmail.com

Komal R. Shetty, komalshetty95@gmail.com

,Prashant K. Dalal,prashantdalal9559@gmail.com,

Abstract—Security is one of the major concerns of people and this project helps you deal with the security of your car. It places emphasis on central locking system of car which is replaced by biometric locking system (Fingerprint Scanner). The current remote key fob is replaced by an android application installed in the smartphone. The android application has biometric interface for fingerprint authentication which is compared with already stored fingerprint. Once successfully authenticated, the Bluetooth of smartphone turns on and user can select the option to lock or unlock according to the preference. Raspberry pi is used for the purpose of pairing between smartphone and car.

Keywords—Android Smartphone, Biometrics,Bluetooth, Fingerprint scanner, Raspberry pi.

I. INTRODUCTION

This Paper includes smart phone and raspberry pi for locking and unlocking door and trunk of the car .Firstly, the android application will authenticate the user by scanning the fingerprint of the user. If the user proves to be genuine, it will trigger the Bluetooth of smart phone and provide consoles to lock/unlock or access the trunk of car. The smart phone will connect to the raspberry pi through Bluetooth. The signals that are received by the raspberry pi from the smart phone will be transmitted through the GPIO pins to the car door's actuators which will trigger the locking and unlocking of the door and likewise for the trunk. Here Raspberry pi will be fitted in the same manner as RF receiver is installed in the car. The raspberry pi will receive the power supply from the car's battery so that it always remains active even when the car is not in ignition.

Thus this system allows us to lock and unlock the car by sending a Bluetooth signal from biometric based android application via a smartphone to the raspberry pi fittedinside the car.

II. RELATED WORK

In this Paper implementation of car door unlocking using Bluetooth technology and anti-theft system using GSM technology has been referred[1]. The paper shows an application which was implemented and developed for mobile phone to access car door locking system. This system was developed by integrating both hardware and software by using EmbeddedBlue 506 Bluetooth technology .EmbeddedBlue and Smart Phone were used as communication devices. The software was designed using the Dynamic C which was compiled and loaded into Rabbit Core Module [2].This paper shows

the closing and opening of car door and trunk with the mobile application. The function was executed and implemented by using 89c2051 and 89s52 microcontroller. The function which was built can be operated within the range of 10 meters from the vehicle through Mobile Bluetooth[3]. In this paper we are using SIFT pores matching technique for the fingerprint verification. [4]. The raspberry pi increases the usage of mobile technology to provide essential security to our homes and for other control applications[5]

Proposed System:

The system architecture is the step by step representation of working of the system. It shows the proper flow of signals. This architecture bears two modules, first module explains about the libraries and fingerprint packages used and second module explains the actual flow of signals.

Module 1:

The android application consists of the package java.lang.Object which inherits the fingerprint manager class extended to Object. It also consists of the class KeyGenerator extended to Object. These classes use different encryption algorithms like AES, ARC4, Blowfish, DES etc.Firstly a request for the fingerprint authentication permission within the project manifest file. Before requesting, verify that the smartphone is protected by a PIN, password or pattern because fingerprints can only be registered on devices which are secured by PIN or password. A FingerprintManager class instance is created. The android Keystore container stores the fingerprint of the owner. Access for the android keystore container is gained and encryption key is generated and stored in the Keystore container. An instance of cipher class is initialized using the key generated. A CryptoObject is created and assigned to the fingerprint manager instance using the cipher instance. Finally call the Authenticate method of *FingerprintManager* instance.

Module 2:

The module 2 shows the communication between the smart phone and raspberry pi. Once the owner is successfully authenticated, the Bluetooth of the smart phone is turned on. The smart phone then sends the Bluetooth signal to the raspberry pi and requests for connection. The Bluetooth receiver of the raspberry pi at the other end accepts the request only from the binded bluetooth MAC address and UUID (unique ID) of the android application. The smart phone's Bluetooth MAC address is binded to the raspberry pi and the android application is binded to a python script. Once done with the pairing, the user can then press the lock/unlockor trunk button. The instruction which is received by the python script running inside the raspberry pi is always in byte format. The instruction received in byte format is converted into string format and after string comparison the respected GPIO pin is activated. Finally a response message is send to the application displaying the status of the actions performed.

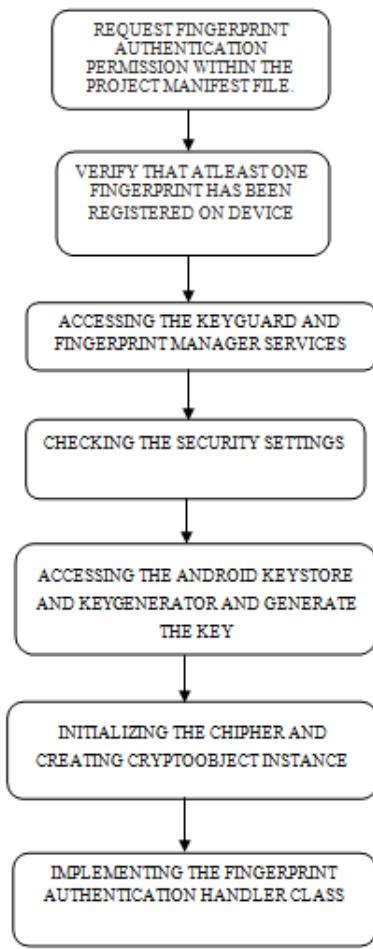


Fig 1 Users and Smartphone Interaction

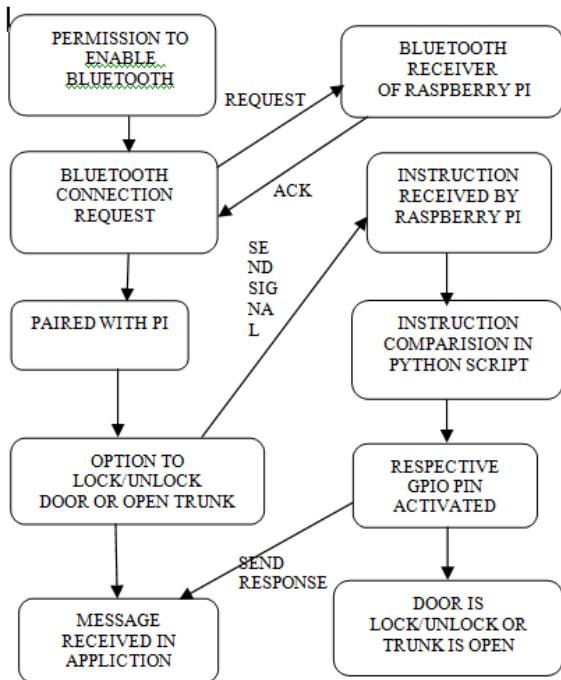


Fig 2 Flow Chart of Smartphone and Raspberry pi Interaction

III. TESTING

The following test cases were conducted to check the vulnerabilities of this project.

The following table shows the difference between the expected and actual time taken by raspberry pi to boot up. As soon as the Raspberry pi boots up, its GUI is displayed through which we can manually execute the python script via LX Terminal. To eliminate the overhead of manually executing the python script, the script was set to autorun which took about 13.80 sec including the bootup time which is considerable.

Table 1 Python Script Autorun Testing

Parameter	Expected outcome	Actual outcome	Status
Script autorun on Boot	13-15 sec	13.80 sec	Ok

The following table shows the handshaking time taken between the android application and the raspberry pi. As soon as the instruction is sent for the first time from the android application, handshaking takes place. The results obtained during the handshaking process are as follows

Table 2 Handshaking

Parameter	Expected outcome	Actual outcome	Status
Handshaking	1.5-3 sec	Beyond 3 sec	working

Once done with handshaking, the communication takes place without much of a delay. The results are as follows

Table 3 After Handshaking

Parameter	Expected outcome	Actual outcome	Status
After Handshaking	1.5-3 sec	1.8 sec	working

IV. RESULT AND DISCUSSION

Whenever raspberry pi receives the Bluetooth signals from android application, it transmits this signal to the relay because of which the coil inside the relay induces. These signals are then further transmitted from GPIO to the actuator which changes the polarity of the motor which is present inside the actuator and thus further locking and unlocking of the car door is possible.

The below circuit easily gets saturated and all the components get heated and the actuator operates for a lesser amount of time. So in order to overcome this, we have placed a switch between the power supply and the above circuit so this switch will provide the power supply to the circuit for a time period of 1 sec which will avoid heating of components. So this switch will give supply to the circuit for 1 sec after the raspberry pi receives signal from the application and then it cuts off the power supply to the circuit.

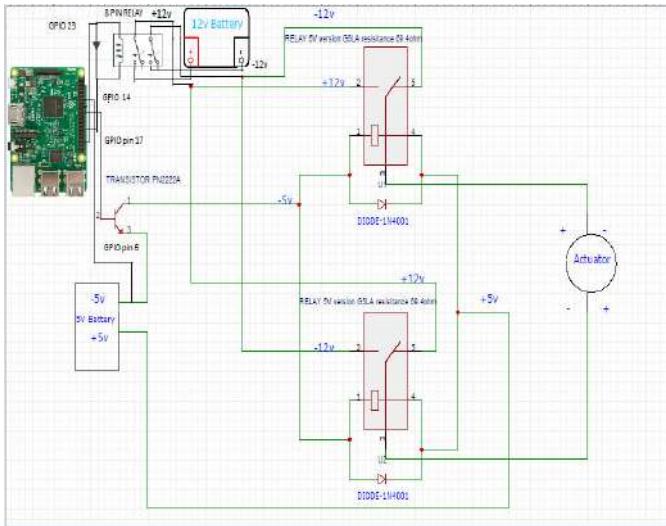


Fig 3 lock/unlock circuit

This is the first activity where the owner places the finger on the fingerprint scanner for authentication. Based on the fingerprint already stored in the smartphone, it will check if the provided fingerprint matches the fingerprint already stored in the smartphone. If matched, it will go to the next activity. Thus the first activity simply shows a symbol for placing the finger on the fingerprint scanner.

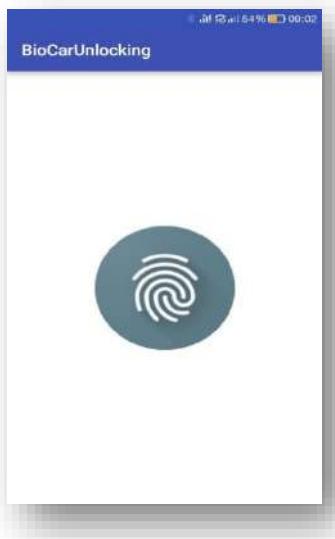


Fig 4 First Activity

This is the second activity after successful authentication. This enables the user to access all three buttons namely door open/unlock, door close/lock and open trunk. As soon as the user clicks any of the following buttons, a signal is sent to the raspberry pi and the following actions are successfully performed and the status is displayed in the textbox.

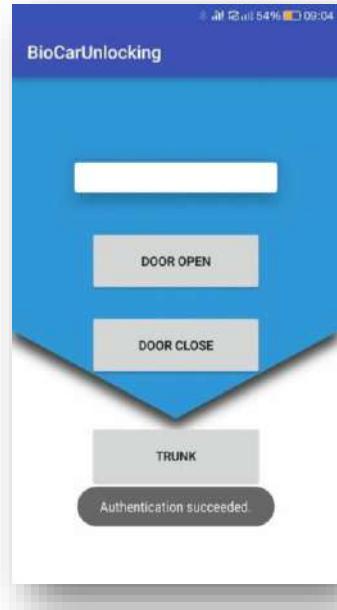


Fig 5 Second Activity

V. CONCLUSION

This system was implemented with the help of an android application which uses biometric for authentication and a raspberry pi. The performance of this system is measured by comparing with existing systems in terms of security, efficiency, feasibility and cost effective. The scope of this system can be extended by adding GPS module to the system which will help to locate the car. This module can be implemented for all the mid-range cars as well as Garage doors and similar concerns.

REFERENCES

- [1] B V D S sekhar, Dr.G.PSaradhiVarma, S. Venkatramana, Ch. Arjun, Ch. Nikhil, "Secure Automotive Locking Control And Anti Theft Using Gps And Bluetooth", International Journal for Innovative Research in Multidisciplinary Field, 8th August 2016.
- [2] HarizHazli Bin Aziz, Noor HafizahAbdul Aziz and Kama Azura Othman, Mobile Phone Car Ignition System Using Embedded Blue 506 Bluetooth Technology, IEEE (2011 IEEE Control and System Graduate Research Colloquium)
- [3] Gurjashan Singh Pannu, Mohammad Dawud Ansari, Pritha Gupta, "Design And Implementation of Autonomous Car Using Raspberry Pi", International Journal Of Computer Applications, March 2015.
- [4] N Kumar, A Singh, "The Study and Optimization of Fingerprint Verification using SIFT Approach on Pores and Ridges of Fingerprints", International Journal of Advanced Research in Computer Engineering and Technology November 2012.
- [5] Ms. Renuka Chuimurkar, Prof. Vijay Bagdi, "Smart Surveillance Security & Monitoring System Using Raspberry Pi and PIR Sensor", International Journal for Innovative Research in Multidisciplinary Field, Jan 2016.

Travel offline virtual support through locational reminder and suggestions using data analytics

Shaikh Farhan Firoz ahmed
Department of Computer Engineering
Anjuman-I-Islam's Kalsekar Technical Campus
New Panvel, India
mistershaikh397@gmail.com

Khan Mahir Ali
Department of Computer Engineering
Anjuman-I_Islam's Kalsekar Technical Campus
New Panvel, India
mahirkhan1995@gmail.com

Shaikh Arfat Izharul haq
Department of Computer Engineering
Anjuman-I-Islam's Kalsekar Technical Campus
New Panvel, India
arfats729@gmail.com

Prof. Kalpana Bodke
Department of Computer Engineering
Anjuman-I-Islam's Kalsekar Technical Campus
New Panvel, India
kalpana.n.meher@gmail.com

The combination of smart phone and web service is that the trend of the longer term data development and software applications. As the tourists are provided with printed guide booklet to find out interesting places during the visit. As the tourists are not aware of the information and proper navigation, tourists are unable to visit all the destinations. Mobile phones were normally used for communication purpose. Travel offline virtual support through locational reminder and suggestions using data analytics. . The Android application design to show how the system works and has a lot of functionality for tourists and travellers such as booking, maps , events, weather, travel partner and many others which are discussed in this paper Nowadays mobile phones are used for various technologies like GPS and browsing over Internet. Mobile phones are equipped with various functionality. We can use these functionalities in our system. The proposed system provides many services to the users like displaying the shortest route between the sources and destinations the tourist specify. We are using data and web scrapping to get the data from the websites. we will be using various API available in market for maps and navigation purposes. The application gets the current location of the user through GPS in the form of longitude and latitude and this information is send to the server. Goal of this project is to guide the tourists to travel on their own.

Global Positioning System, Responsive Web Service, Smart phone.

I. INTRODUCTION

While travelling tourists expect to get personalized access to tourism information at anytime, from anywhere through any internet enabled device. Mobile applications can provide the user with such a general access. With the evolution of technology, internet enabled devices have made it effortless to access information anywhere, anytime.

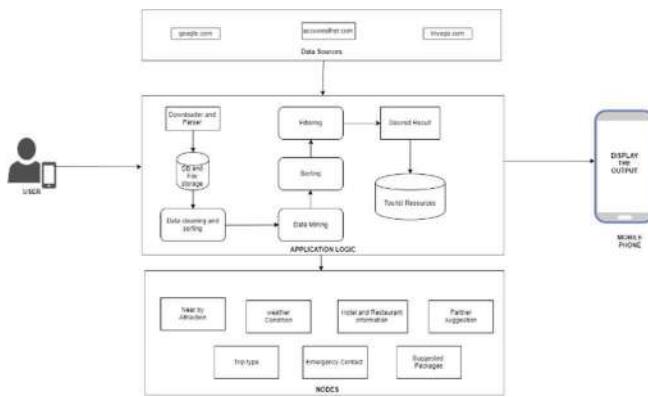
The task is to replace the hard copy tour guides books with mobile applications. We will also provide E-books to our customers. Travelers will use mobile applications and websites due to the benefit they present over hard copy books. With the fast developments in technology, internet enabled devices offer many advanced features rather than making calls, so the number of smart phone users increases day by day, not only in urban areas but also in rural areas. Tourists at a new place always faces problem finding the location of different facilities such as shop, hospitals, hotels, restaurants and bus stops. Our focus for this project is on software support for location based applications; we are not just interested in the location but also other elements of the user's context, such as memorables locations in view, attractions and equipment nearby, such as public telephones and toilets. In this paper we describe the processes involved in designing the Tourist Guide application and in particular we focus on the context sensitive features of the system as well as the reliability of the navigation system. When analyzing the tourism systems of other countries it was found that there have been many experimentation carried out regarding internet enabled application based tourism systems. The Location based services for mobile devices is an application that provides information and services to users based on their

KEYWORDS:

Electronic tourist guide, Mobile tourist guide, Tourist Information System, Tourist Application, Android, GPS-

location. The tourism sector is viewed as one of the most important engines of growth and development in the Indian economy mostly from the northern part and as such, is a key focus in the Government's economical strategy. The Android application design to show how the system works and has a lot of functionality for tourists and travellers such as booking, maps , events, weather, travel partner and many others which are discussed in this paper. An offline version of android application for the system which won't need full time Internet connection get services and as soon as the application gets connected to Internet then the application should be automatically get synchronized with the real time service data. The system provide information query of the hotel, scenery, restaurant, traffic and so on. The system is a combination of devices and Internet services and will facilitate traveler the with all the information requires during the trip. In this paper we have proposed a system or mobile application which will track the current position of the user and send it to the server which will send the requested information for the nearby attractions.

II. PROPOSED SYSTEM



The system focuses on travel off line virtual support through locational reminder and suggestions using Data analytics. The proposed system has the following features:

1. Social Login: Social login is a sign in technology through which the users is authenticated on various applications and sites by connecting through a social site rather than entering again the ID and password on each website. Many people are now using social authentication which are publicly exposed authenticated sites or social media such as Google and Facebook.

2. Nearby Attractions: This feature of the proposed system helps the tourist to find the nearby tourist spots like shopping, adventures activities, hiking location, resorts and amusement park and many more.

3. Weather updates: Time to time updates regarding the climate will notified to the tourists. For regular updates of weather the tourist should have connected to internet.

4. Nearby hotels: The system provides general information of hotel, restaurant, shops, hospitals, and companies. As well as the newest events of the plaza and shops. The system provides service of hotel, restaurant and cinema-ticket reservations. The system provides service of displaying the shortest path between the sources and destinations the visitors specify.

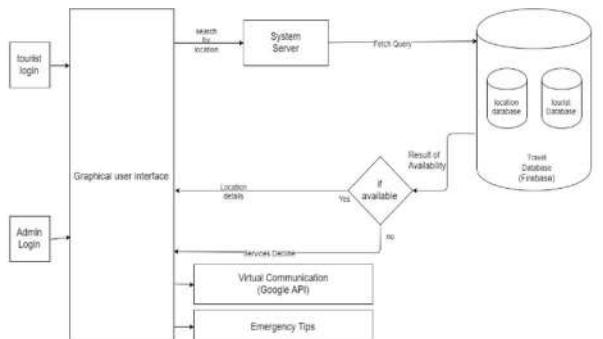
5. Trips types: The system will ask the user which type of trip is he/she is making like business trip, family trip or solo trip which is in trend and increasing day by day. According to the type of trip the tourist is making the system guides the researchers and guests that came for conferences to spend an easy tour in the country.

6. Partner Suggestions: As the tourist selects the type of trip, if the user select the solo trip this feature of our system will as for the user if he/she wants any partner for trekking.

7. Suggested packages: The system will provide the tourist packages based on the search location according to the type of trip the user is going to make.

8. Emergency Contacts: The system will provide the emergency contacts like Hospitals, Police station, Ambulance, Railway helpline, Woman helpline.

Flow of System:



III. LITERATURE SURVEY

Data can be referred as representing some existing information in a more better form for processing and analyzing. Data extraction is a process of retrieving data out of data sources that are unstructured or poorly structured. Generally, unstructured data sources include emails, web pages, spool files, etc. Hence, this process of extracting data from the web is referred to as Web Scraping. Web scraping, a type of data scraping, is a process of extracting data from the web world through various methods. With the help of web scraping services, the unstructured data is converted into structured data which is stored centrally. The aim of the web scraping is to obtain, store and analyze data.

A web scraper composes of two parts:

Data Extractor (to extract data from crawled links): After crawling all the web pages of a website, the links are filtered out from it. As there is a lot of unrelated data present on the web page, data extractor is used to extract the required data and convert it into a usable format.

Web Crawler (to crawl links) : A web crawler generally crawls a web page using recursive algorithms in which it scans the page first, finds the links present on the page which later is stored in a type of data structure. Then fetches the first page of that link, stores them into the same data structure and recursively repeats the process till all the links get crawled.

Some web scraping techniques:

1. Tree-based: The nature of the web pages is semi-structured which is one of the most exploited features in the extraction of web data and can be represented as labeled ordered rooted trees. In this technique, labels represent the tags of HTML and the tree represents different levels of nesting elements which construct the web pages. This representation is generally referred as Document Object Model or DOM.
2. Human Copy-paste: This is one of the most common techniques used by people who do not belong to a technical background. This human manual examination technique is quite helpful when websites scraping sets constraints or barriers to prevent human automation.
3. HTTP Programming: A socket programming is used to retrieve web pages by posting HTTP requests to the web servers.

Web Mining

Web mining is an application of data mining techniques which aims at extracting useful data patterns and adopts most of the techniques of data mining to discover potentially useful information from the web. Web mining helps to evaluate the performance of a business by recognizing customer behavior which indirectly helps to boost business. It examines the content

of the web and also the result of the search. Web mining is a part of Information Retrieval and Information Extraction systems and is a tedious task as the web source lack of structure which leads to problems such as information overloading.

There are two approaches to web content mining :

- a. Database Approach : Multilevel databases are used to extract meta data from data at a lower level and is organized in a structured format.
- b. Agent Based approach : Agent based approach directly mines the content of web documents by using three types of agents.

- Intelligent search agents : They automatically search according to the given query by using domain characteristics and profiles of the user.
- Information categorizing agents : They use different functions and techniques to retrieve data.
- Adapted web agents : They fetch data from welldefined databases containing schema and attributes.

Algorithm for Data compression:

Lossy data compression algorithm

Lossy compression is the data compression technique that uses approximations and limited data discarding to represent the original content. These techniques are used to reduce the size of the data for storing, transmitting and handling data. This technique is opposite to lossless data compression which does not degrade the data. The amount of data lost possible using lossy compression is much higher than lossless techniques.

Lossy compression technique reduces file sizes before degradation is noticed by the user. Even the degradation of the data is notice by the user, further data reduction may be desirable (e.g to reduce transmission time to reduce storage needs).

Lossy compression is most commonly used to compress data like audio, video, and images especially in applications such as streaming media and internet calls. By opposition, lossless compression is basically use for text and data files, such as financial records and text files. It is better to make a main lossless file which is to be used to produce new compressed files; for example, a multi-megabyte file can be used at full size to produce a full-page advertisement in a glossy magazine, and a 100 kilobyte lossy copy can be made for a small image on a web page.

IV. RELATED WORK

Trip Advisor:

TripAdvisor, is a travel website company providing hotels and travel related booking . It also includes interactive travel forums. The website services are free to users, who provide most of the content, and the website is supported by a hotel booking facility and an advertising business model.

Rome2rio:

Rome2rio is a multimodal transport search engine. Rome2rio's platform is capable of long-distance trip planning as well as intra-city trip planning. Users can input any address as the origin and destination and Rome2rio searches a database of flight, train, bus and driving routes and price options for travelling to that destination.

Triposo:

Triposo is a social travel site and application . Triposo was released which included location-based software which allowed the application to tell the user recommendations depending on the weather, time, and other variables. The application has reportedly been downloaded 10 million times.

The application, available for iOS and Android, will show the user recommendations on where to go depending on information they've given to the application. This includes Facebook details. The application works without internet connection. In order for the application to work without connectivity, it downloads information about the users destination before departure.

CONCLUSION AND FUTURE SCOPE:

The main aim of the research project was to develop a tourist guide system for tourist travelling. An important assumption made while developing the system was that the users have the basic idea about using an android mobile device and they are familiar with the English language. In order to use the location based services user needs to be in a place where the mobile device receives GPS data accurately. The system needs a better network connection for communication between the mobile device and the server. Otherwise it takes a long time to receive the data from database. The web and Android application developed to show how the architecture works and has a lot of important features for tourists such as booking, maps and places, events, weather etc.

An offline version of android application for the system which won't need full time Internet connection get services and as soon as the application gets connected to Internet then the application should be automatically get synchronized with the real time service data. The system provide information query of the hotel, scenery, restaurant, traffic and so on. The overall system combines a smart phone and Internet services and will allow trip related tour and life for user. In this paper we have proposed a system or mobile application which will track the current position of the

user and send it to the server which will send the requested information for the nearby attractions. In future this system can use for hotel booking and travel booking services.

ACKNOWLEDGMENT

We, the authors, are grateful to Prof. Kalpana R. Bodke for her great support and guidance throughout this study.

REFERENCES

- [1] Emilio Ferrara, Pasquale De Meo, Giacomo Fiumara, Robert Baumgartner, "Web Data Extraction, Applications and Techniques: A Survey", June 5, 2014
- [2] Surbhi Sharma, Dinesh Soni, Dr. Arvind K Sharma, "Explorative Study of Web Data Mining Techniques and Tools: A Review", International Journal of Computer Science And Technology, Vol . 8, Issue 1, Jan March 2017
- [3] Google Maps. (n.d.). Retrieved May 21, 2013, from <https://maps.google.com>
- [4] Google Places API. (n.d.). Retrieved May 20, 2013, from Supported Place Types: https://developers.google.com/places/documentation/supported_types
- [5] M. H. Goadrich and M. P. Rogers, "Smart Smartphone Development: iOS versus Android," Proceedings of the 42nd ACM technical symposium on Computer science education, 2011.
- [6] Google APIs Console. (n.d.). Retrieved May 21, 2013, from <https://code.google.com/apis/console/>
- [7] Wikipedia for information collection https://en.wikipedia.org/wiki/Main_Page
- . H. Nwana D. Ndumu "A perspective on software agents research" The Knowledge Engineering Review vol. 14 no. 2 pp. 1-18 1999.
- [8]. D. Ndumu J. Collins H. Nwana "Towards Desktop Personal Travel Agents" BT Technological Journal vol. 16 no. 3 pp. 69-78 1998.
- [9]. J. N. Suarez D. O'Sullivan H. Brouchoud P. Cros "Personal Travel Market: Real-Life Application of the FIPA Standards" Technical Report BT Project AC317 1999.
- [10]. M. Umlauft G. Pospisil G. Niklfeld E. Michlmayr H. Werthner E. Veit "LoL@ a Mobile Tourist Guide for UMTS" in Journal on Information Technology & Tourism Congnizant vol. 5 no. 3 pp. 151-164 March 2003.
- [11]. W. Swinger C. Grün B. Pröll W. Retschitzegger A. Schauerhuber "Context-awareness in mobile tourism guides-A comprehensive survey" Rapport Technique. Johannes Kepler University Linz 2005.
- [12]. H. K. Eden U. Gretzel "A taxonomy of mobile application in tourism" E-review of Tourism Research vol. 10 no. 2 pp. 47-50 2012.
- [13]. M. Kenteris D. Gavalas D. Economou "Evaluation of Mobile Tourist Guides".
- [14]. Jian Meng Neng Xu "A Mobile Tourist Guide System Based on Mashup Technology" ISBN 978-1-4244-7618-3/10©2010IEEE.
- [15]. H. K. Eden U. Gretzel "A taxonomy of mobile application in tourism"

E-review of Tourism Research vol. 10 no. 2 pp. 47-50 2012.

[16]. M. Andress "The road to secure web services" InfoWorld
vol. 24 no. 2 pp. 52 2002.

Image Auto-tagging

Javed Gouri

Department of Computer Engineering
Anjuman-I-Islam's Kalsekar Technical Campus
University of Mumbai

Abstract—Due to the advancement in the field of multimedia technologies, there is an increase in the computerized and digital images. An image may contain a tree, house, mountain, etc due to which a real life object can be categorized into multiple categories. There have been several studies on automatic image annotation where they utilize machine learning techniques to annotate digital images due to its need. Face detection and recognition is already being used in many real world applications. The traditional methods of retrieving an image such as annotating images manually is time-consuming and expensive, especially for an continuously increasing image database. The problem in the existing applications is that it does not tag the other objects present in the pictures, and sometimes they also have a problem with tagging people. In this paper, we propose a system of automatic image annotation using convolution neural networks that takes into account the textual queries or keywords and searches for the related in the database. Image auto-tagging is a classification task that aims to tag or label an image by identifying the objects present within the same image. This new system also has an advantage of automatically determine the image on the basis of the keyword entered by the user. It can also be used to improve information content for the description of the image.

Keywords—Image Annotation; Deep Learning; Django; Images; Labels

I. INTRODUCTION

Storing an image and later finding it is a very tedious task especially in a big image dataset. It is, thus, very difficult to remember each and every image by its name especially when there are similar images and names. Meta data tags are used to tag objects in images. Generation of these tags automatically is desired by the user. Automatic image annotation creates a problem where the task is to predict multiple textual label for describing unseen image in image and its visuals. It's is very subjective and time consuming to tag and annotate image manually. It also creates problems due to difference in opinions, languages, vocabulary education, culture, etc. Nowadays, many mobile phones and digital cameras have built-in GPS which is attached to the flash of the camera and phones which gives a geo-code that assigns a location tag to the image. In some mobile phones this tag is generated automatically with the GPS co-ordinates available.

In this paper, we propose a system which aims to predict the objects present in an image and give them a certain tags which can be later used to search the image by voice commands or by exploring the server.

II. WHY THERE IS A NEED TO ANNOTATE AN OBJECT IN IMAGE

Searching galleries of endless mislabeled photos, trying to search a particular image is both frustrating and time consuming. Hence, there is a need to annotate image. Annotating objects in images has following advantages :

A. Ease of use

By tagging the images it will become easy to access the image files and there won't be any need to remember the name of each and every image. It is also difficult to tag the image manually. This will need more human efforts along with a lot of time and expense. Hence, there is a need of a system that will automatically annotate the objects present in the image so that the image can be easily available to the user every time he needs it without manually remembering the name of each and every image.

B. Efficient related search

The annotation of object along with the image will help us access images related to the search. It will not only display the desired image, but also the related searches from which can pick more images. This will help in increasing the efficiency for searching

III. AUTOMATIC IMAGE ANNOTATION

Automatic image annotation is an automated process in which the computer system assigns a meta data automatically to a image in the form of captions or keywords so as to organize, retrieve or locate images from database. Thus, we can say that automatic image annotation allows user to have an efficient access to a large database of images and gives the ability to organize and search multimedia information from it. It is an application of computer vision and is also known as Linguist Indexing or Automatic image tagging.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar :

IV. PROPOSED SYSTEM AND MODULES

A. System Architecture

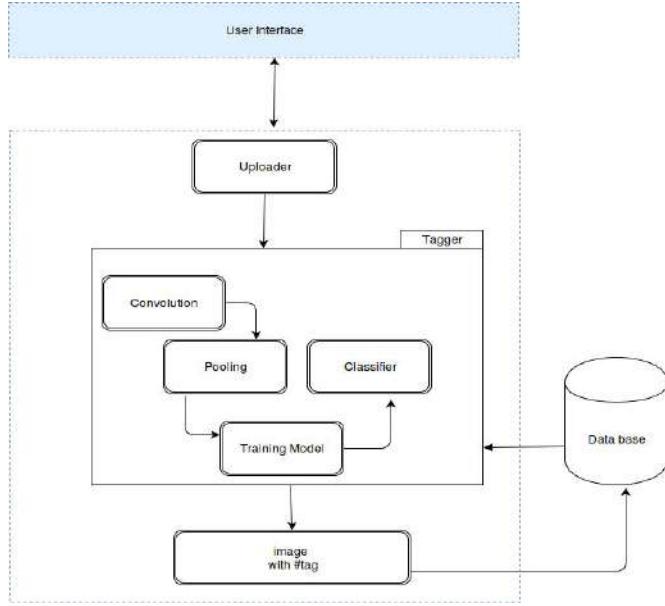


Fig. 1. System Architecture

B. Modules

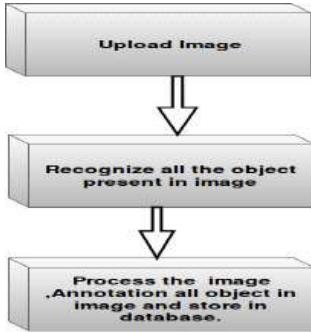


Fig. 2. Tag object in an image

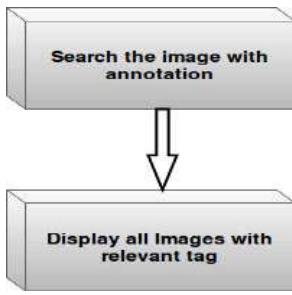


Fig. 3. Fetch data from database

C. User Interface

The User Interface will be a web or android application which will have a search bar where the user can fetch their

images from the database by merely entering the relevant tag name. The output displayed will be the images related to the user's tag input.

D. Uploader

This will basically be an API which will maintain the synchronization between an user's local and global images. Local images refers to the user's personal images whereas global images refers to the images which are already stored on cloud. The uploader will upload the recent images of the user on cloud which will be later sent for image-processing and tagging.

E. Auto-tagger

Automatic assignation of captions to a digital image by a computer system is called as Automatic image annotation. This module is important for tagging object present in image. Basically this module consists of convolution, pooling, training model, and classifier.

1) Convolution

Each layer of convolution is of size $H \times W \times D$ which learns D features of $H \times D$, where H refers to the height and W refers to the width of learned kernels. The size of the kernel determines a component's maximum size which it can precisely capture. If the size of kernel is too small, the layer would not be able to learn a meaningful data shape representation or distribution. Hence, large kernels (17×5) are proposed but they do not allow invariance within its range. The convolution axes are crucial aspects of the layers of convolution.^[6]

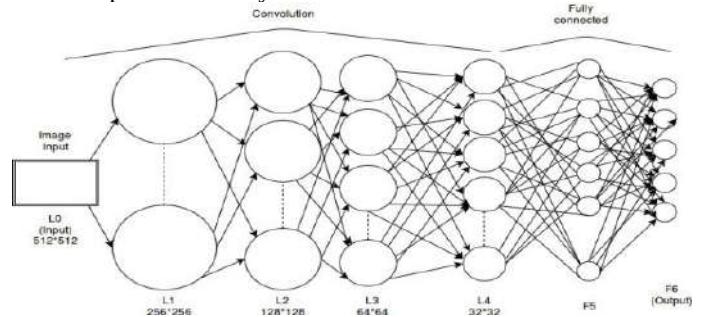


Fig. 4. Convolutional Neural Networks

A convolution operation is applied to the input by the convolution layers to pass the result to the next layer. A response of an individual neuron is emulated by convolution to visual stimuli. The data is processed by each convolution for its particular field. CNNs tolerate translation of the input image such as rotation, etc. because of Tiling. The number of parameters which are free are reduced using convolution operation which improves generalization. This implies that the problem in training traditional neural-network layers is resolved by using back propagation. There are one or more convolutional layers in a convolutional neural network. It is then followed by one or more fully connected layers which are identical to the layers in a standard multilayer neural network. The CNN's architecture is designed to take an advantage of the input image's 2D structure.

2) Pooling

Convolutional networks may include local or global pooling layers, which combine the outputs of neuron clusters at one layer into a single neuron in the next layer. For example, max pooling uses the maximum value from each of a cluster of neurons at the prior layer. Another example is average pooling, which uses the average value from each of a cluster of neurons at the prior layer. It partitions the input image into a set of non-overlapping rectangles and, for each such sub-region, outputs the maximum. The intuition is that the exact location of a feature is less important than its rough location. The spatial size of the representation, the number of parameters and the amount of computation in network can be controlled by the pooling layer to have a control over fitting. An another form of translation invariance is provided by pooling operations.

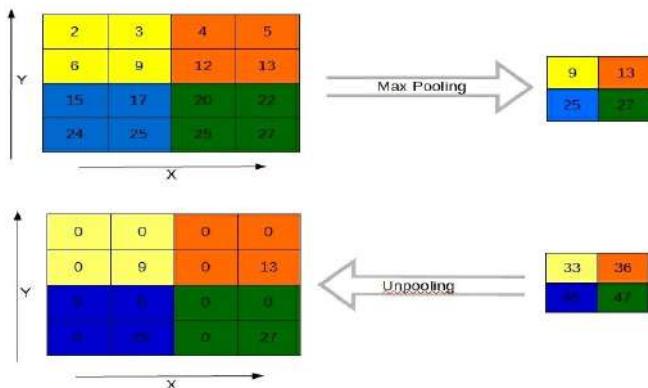


Fig. 5. Pooling and Unpooling

Generally, a max function of Pooling is used as an operation to reduce the size of feature map. Majority of works relying on CNN structures has adopted it. To reduce the size of feature map, Pooling employs sub sampling to preserve the information of the activation in the region. This helps pooling size in determining the tolerance within each layer's location variance and provides a trade-off between two different

aspects that can affect a network performance. The network does not have enough invariance of distortion if the pooling size is very small. If the image size is very large, the features' location may be missing when needed.

3) Training Model And Database

In the training model, the data is trained for getting identified and a specific keyword or tag is assigned to the object. After assignment of tags, the image is then stored in the database for further access.

CONCLUSION

Personal photos highly differ in appearance but are increasingly used and also contain a great deal of information about the user. Our system will combine all the image data with specific keyword which will increase the efficiency of annotation performance. We believe our proposed system could provide a useful service to the millions of users who keep labeling the photos manually.

ACKNOWLEDGMENT

We, the authors would like to thank our guide, Prof. Tabrez Khan for his support and thorough guidance.

REFERENCES

- [1] G. Eason, B. Noble, and I.N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955. (references)
- [2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.
- [3] I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.
- [4] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.
- [5] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
- [6] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

WEB PORTAL ON CONFERENCE ALERT

Authors Name/s per: Mohammad Hamid(13CO46)
COMPUTER, AIKTC
PANVEL, INDIA
Email-id: shaikhhamid039@gmail.com

Authors Name/s: Shaikh Uzair Ahd(13CO59)
COMPUTER, AIKTC
PANVEL, INDIA
Email-id: uzairshaikh869@gmail.com

Author Name/s: Siddiqui Mohd. Sharique(11CO45)
COMPUTER, AIKTC
PANVEL,INDIA
Email-id: shariquesiddiqui100@gmail.com

Authors Name/s per Bind Rahul(13CO20)
COMPUTER, AIKTC
PANVEL, INDIA
Email-id: rbind7172@gmail.com

Abstract—We are highlighting on the topic of Conference Alert, which is very fundamental approach for the people. Those who are interested in conference alert. For such types of people at the beginning has to visit several sites to collect the information about the conference, it requires a lots of time wasting to achieve the proper information. We are developing a web-portal on conference alert through the scraping from different web sites. Where all the users will get proper information about the conference without the wasting time and it will also notify to the user on which date and time the conference will be going too held on. The library that we are going to use to extract the data from different websites by using Beautiful Soup. It is an incredible tool for pulling out the information from a web pages. You can use it to extract tables, lists, paragraph and you can also put filters to extract information from web pages. Python idioms and few simple methods are provided by Beautiful Soup for navigating, searching, and modifying a parse tree. It is a toolkit for dissecting and extracting a document or information what you need. It required less code to develop an application.

Keywords—Web scraping, Beautiful soup, Data analyzer

I. INTRODUCTION

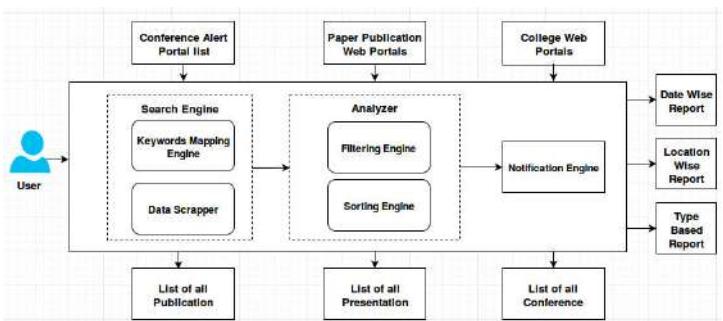
A conference is generally understood as a meeting of several people to discuss a particular topic. A conference differs from the others in terms of knowledge and purpose, the term can be used to cover the general concept. It is a gathering of delegates representing several groups. At a conference, innovative ideas are thrown about and new information is exchanged among experts. This technology can help people, organizations who wants to organize any conference or want to join the conference it can also help to the college who wants to published the paper. Scraping is the act of extracting data or

information from websites, with or without the consent of the website owner. Web scraping is a computer software technique of extracting information from web sites. This technique mostly focuses on the transformation of unstructured data (HTML format) on the web into structured

data (database or spreadsheet). Scraping can be done manually, or automated. In most cases, it's the latter because of its efficiency. Scraping of content or prices is mostly done with malicious intents, and there are several techniques by which this is done.

The purpose of this project is for the user who not get exact information about the conference on which date and time. The different platform and technologies are for developing this web portal. We recommended different point on the basis of the conference alert to be held on the given date and time. We will also scrap the data from another websites and will display it on our portal which will be beneficial for the users who search the information on different sites.

II. SYSTEM ARCHITECTURE



A. Scraping

Scraping is very fundamental approach in the web system .it is the act of extracting data or information from websites, with or without the consent of the website owner In our project we will scrap the data or information from different websites which is related to conference or paper publications and data will be scraped by Keywords.

Libraries required for Scraping:

Urllib2:

It is a Python module which can be used for fetching URL s. It defines functions and classes to help with URL actions (basic and digest authentication, redirections, cookies, etc.)In our project we will use Urllib2 to fetch the URL from different website after it will be scraped by Beautiful Soup to the particular information from that URL which has been scraped by the Urllib2.

Beautiful Soup:

It is an incredible tool for pulling out information from a web pages. You can use it to extract tables, lists, paragraph and you can also put filters to extract information from web pages .It does not fetch the URL it only scrap the information therefore we will use both Beautiful Soup as well as Urllib2.

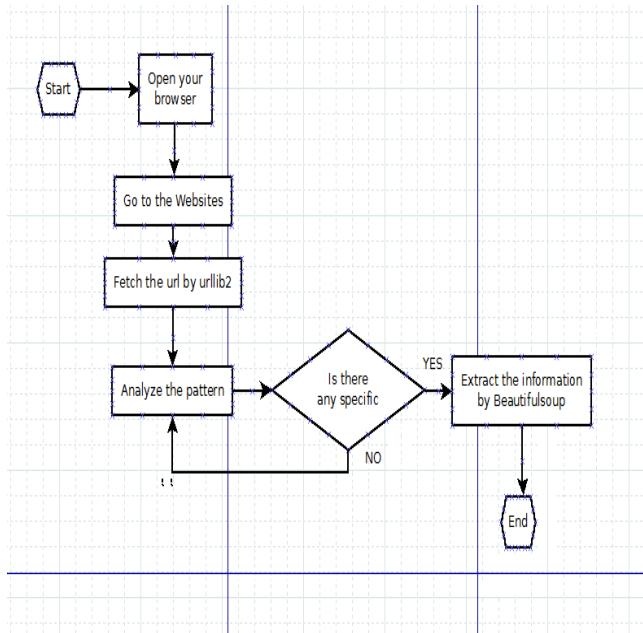


Fig: Flowchart of scraping

B. Data Analysis

Data Analysis is a process of inspecting, cleaning, transforming and modeling data with the goal of discovering useful information, suggesting conclusions and supporting decision-making .In our project data analysis will use to clean the data. Filter will remove the unwanted data then it will perform sorting on the basis of data wise and location wise.

III. LITERATURE SURVEY

Websites	Notification by	Alert by
www.conferencesindia.org	E-mail, SMS	Facebook,twitter,google+
www.allconference.com	E-mail 1	Facebook,twitter
www.cacuss.ca	E-mail	Facebook,twitter
www.conferencealert.com	E-mail	Facebook,twitter,google+, instagram
www.conferencealert.in	E-mail	Facebook,twitter,google+, linkedin
www.ieee.org	E-mail	Facebook,twitter,google+, linkedin,youtube,instagram

We have visited lots of websites which is related to the conference, paper publication some websites notify to the user's via email, message notification but some websites allow only email notification to the users but there is no websites are available which allow the users via desktop notification sometimes user doesn't have time to check the email or mobile to check schedule of conference which has been sent so therefore after based on the result we will also provide desktop notification as well as email, message notification to our users so therefore it will be totally beneficial for that users who doesn't have time to check the email.

ACKNOWLEDGEMENT

We are extremely thankful to our guide Prof. **ANSARI MUKHTAR AMIR** for their valuable guidance and for providing all the necessary facilities, which were indispensable in the completion of this project report. We are also thankful to Department of Computers of Anjuman-i-islam Kalsekar Campus, New Panvel for their valuable time, support, comments, suggestions and persuasion. Required facilities, Internet access and important books.

CONCLUSION

Conference Alerts helps in promoting conferences for Academic and scientific studies. The absolute need to attain an international conference at India could be met by using conference alerts to search for the meeting on appropriate areas of interest to present once innovative research articles.

REFERENCES

- [1] <http://scraping.pro/basic-web-scraping-principles/>
- [2] <http://www.analyticsvidya.com/>
- [3] M. Turland, PHP Architect Guide Wen scraping with PHP Muskeeters.me, LLC, 2010.
- [4] D. T. Larose dan C. D. Larose, iscovering Knowledge in Data An Introduction to Data Mining, New Jersey: John Wiley Sons Inc, 2014.
- [5] M. Bakaev dan T. Avdeenko, “Data Extraction for Decision-Support Systems: Applicationin Labour Market Monitoring and Analysis,” vol. 4, pp. 23-26, 2014.)
- [6] Bonifacio, T. E. Barchyn, C. H. Hugenholtz dan S. W. Juebzke, “CDST:Free Canadianclimate data scraping tool,” pp. 13-16, 2014.
- [7] <https://www.scrapehero.com/5-tips-for-scraping-big-websites/>
- [8] <https://www.promptcloud.com/data-scraping-vs-data-crawling/>

Author Index Multicon-W:-CRTCE-2018

A

Aakash Kanojiya -118	29-32
Aakash Katiyar -42	Appendix
Aakash Mishra -15	Appendix
Aakash Patel - 101	133-136
Aakashkumar Yadav -32	233-236
Abhijeet Prasad - 112	27-28
Abhijeet Saha -46	241-244
Abhimanyu Singh -49	151-154
Abhishek Yadav -32	233-236
Adit Shrimal - 89	Appendix
Aditya Darak -14	Appendix
Aditya Samant - 89	Appendix
Aditya Sarvankar -17	225-228
Ahamad Husen Shekh -68	Appendix
Ahamad Husen Shekh -69	Appendix
Aishwarya Manjalkar -77	197-200
Aishwarya Sontakke - 81	159-162
Aishwarya Verma -73	17-22
Ajaykumar Verma -20	141-144
Ajinkya Naik - 80	Appendix
Akanksha Joshi -62	179-184
Akanksha More - 122	267-270

Akash Mishra - 126	175-178
Akash Singh -114	Appendix
Akhil Choudhary -14	Appendix
Akib Maredia - 128	97-100
Akshay Gunani - 124	Appendix
Akshit Vaghani - 88	Appendix
Ali Akbar Nathani -55	245-246
Ameya Gharat -65	Appendix
Amit Dubey -10	41-46
Amit Mehra -7	111-116
Amit Tiwari - 82	69-72
Anand Khandare - 121	263-266
Anand Khandare - 99	75-80
Anand Khandare -105	171-174
Aneri Moradia -31	Appendix
Angela Roy -114	Appendix
Aniket Korgaonkar -15	Appendix
Anita Lahane - 111	85-88
Anjali Thomas - 109	Appendix
Anjali Yeole -66	59-60
Ankit Gupta -28	Appendix
Ankita Thakur - 122	267-270
Ankur Sinha - 99	75-80
Anupriya Babbar -37	Appendix
Anupriya Babbar -39	129-132
Anuradha Maurya - 77	197-200

Anurag Shetty -29	145-146
Anushri Arora -62	179-184
Archana Gulati - 95	251-254
Archana Patankar -45	147-150
Archit Kogta -23	Appendix
Arfat Shaikh -51	Appendix
Arul Shalom -49	151-154
Arzoo Jivani -16	Appendix
Ashish Kothari -13	121-124
Ashish Yadav -114	Appendix
Ashlesh Patel - 98	23-26
Ashuta Thakur - 75	191-196
Ashutosh Dwivedi - 80	Appendix
Ashutosh Mishra - 128	97-100
Ashwin Yadav -114	Appendix
Ashwini Patil - 98	23-26
Atif Thakur -53	55-58
Avinash Pardeshi -65	Appendix
Ayushi Mankar -14	Appendix
Azad Nautiyal -23	Appendix
Azhar Nabi -64	Appendix

B

Bavel Singh -21	Appendix
Bharat Joshi -6	105-110

Bharat Rathore -32 233-236

Bijal Patel -106 259-262

C

Chaitanya Zarekar -25 47-52

Chandrashekhar Bora -7 111-116

Chinmay Verlekar -118 29-32

D

Deep Suchak -12 137-140

Deepak Sharma - 89 Appendix

Deepak Singh -16 Appendix

Deepali Joshi - 97 73-74

Devyashwar Raj - 126 175-178

Dhananjay Theckedath - 129 Appendix

Dhruvi Bhagat -13 121-124

Dhruvil Parekh -102 Appendix

Dinesh Nandi - 124 Appendix

Dineshkumar Tanwar -8 Appendix

Divya Dubey - 92 Appendix

Dnyaneshwar Dhangar - 122 267-270

Drashti Shrimal -71 185-190

F

Farhan Shaikh -51 Appendix

G

Gargee Kawale -15 Appendix

Garima Choudhary -13 121-124

Gautam Kumar Jha - 78 247-250

Gautam Mishra - 121 263-266

Gautam Sharma - 117 Appendix

Gautamkumar Ojha - 130 Appendix

H

Harsh Shah - 98 23-26

Harshal Patil -64 Appendix

Harshali Patil -11 117-120

Harshali Patil -7 111-116

Harsheen Kaur - 109 Appendix

Himanshu Agrawal -38 5-10

I

Ian Noronha -22 Appendix

Ishita Sharma - 90 205-210

J

Jainee Parmar -46 241-244

Jainil Jain -10 41-46

Javed Gouri -56	Appendix
Jenil Shah -104	217-220
Jhanvi Shah - 81	159-162
Jinsith Maroli -21	Appendix
Jonathan Kumar -17	225-228

K

Kalindi Awasthi -9	1-4
Kalpana Bodke -51	Appendix
Kalpana Gangwar -86	167-170
Kanishk Verma - 133	33-36
Karan Sharma - 108	Appendix
Kaustubh Shah -26	229-232
Keval Chavda -106	259-262
Kinjal Bhatia - 109	Appendix
Kiran Gurav -65	Appendix
Kirti Sharma - 108	Appendix
Komal Shetty -48	Appendix
Krishna Kumar Yadav -65	Appendix
Kunal Singhal -12	137-140

L

Loukik Salvi -67	11-16
Lydia Suganya -25	47-52

M

Madhuri Rao -38	5-10
Madhuri Rao -84	163-166
Mahima Choudhary - 109	Appendix
Mahima Jain -120	89-92
Mahir Khan -51	Appendix
Manish Rana - 112	27-28
Manish Rana - 89	Appendix
Manish Rana -103	255-258
Manisha Sharma - 90	205-210
Manjiri Bhat -66	59-60
Manoj Dixit -118	29-32
Marmik Shah -21	Appendix
Mayannk Kumaar -25	47-52
Meenaz Gigani -53	55-58
Megha Gupta -34	Appendix
Megharani Patil - 92	Appendix
Megharani Patil -102	Appendix
Megharani Patil -31	Appendix
Meher Dodhia -120	89-92
Mihir Rane - 130	Appendix
Mohammad Atique -103	255-258
Monal Daswani -66	59-60
Monark Unadkat - 117	Appendix
Mrinal Bageshwari - 132	Appendix

Mukesh Patel -62 179-184

Mukul Suvarna - 88 Appendix

Mungi Abrar Nizamuddin - 47 Appendix

N

Namrata Bodas - 81 159-162

Narayan Mundhara -66 59-60

Nazneen Pendhari - 87 201-204

Neha Kamble -74 61-64

Nidhi Mishra -31 Appendix

Nihir Shah - 98 23-26

Nilay Darji -23 Appendix

Nilofer Peerzade -53 Appendix

Ninad Virkar -29 145-146

Nishant Shishodia -12 137-140

Nitin Harane -22 Appendix

Nitin Harane -116 Appendix

O

Omar Dalvi -17 225-228

P

Parth Sharma - 89 Appendix

Pathan Naziya - 134 37-40

Pawan Singh -16	Appendix
Pawankumar Dubey -18	Appendix
Piyush Akoliya -33	233-236
Pooja Trivedi - 92	Appendix
Prachi Dalvi -37	Appendix
Prachi Janrao -12	137-140
Prachi Janrao -16	Appendix
Prachi Nagane - 132	Appendix
Pragnesh Patil - 98	23-26
Prajakta Samant - 75	191-196
Prajwal Mogaveera -14	Appendix
Prakash Shukla -29	145-146
Prakhar Agrawal -102	Appendix
Pranali Patil -34	Appendix
Pranit Patil -46	241-244
Pranjali Sankhe - 111	85-88
Prashant Dalal -48	Appendix
Prashant Kapri -49	Appendix
Prashant Rathod -48	Appendix
Pratik Patel -118	29-32
Pratik Singh - 112	27-28
Pratik Singh -19	125-128
Pratiksha Singh -105	171-174
Pratiksha Warule - 95	251-254
Priyankar Tiwari -67	11-16
Pulkit Kedia -23	Appendix

R

R.R Sedamkar - 129	Appendix
R.R. Sedamkar -82	69-72
R.S.Kokhale-119	Appendix
Raghavendra Sedamkar - 87	201-204
Raghvendra Sedamkar -115	Appendix
Rahul Gupta -10	41-46
Rahul Iyer -22	Appendix
Rahul Singh -20	141-144
Rahul Yadav -26	229-232
Raj Shah - 101	133-136
Rajesh Singh -23	Appendix
Ramit Gupta -15	Appendix
Rashmi Thakur -13	121-124
Rashmi Thakur - 108	Appendix
Rashmi Thakur -6	105-110
Ravish Deshmukh -50	53-54
Ravish Deshmukh -53	Appendix
Ravishankar Singh -19	125-128
Rekha Sharma - 130	Appendix
Rekha Sharma -18	Appendix
Rekha Sharma -104	217-220
Richa Tanna - 122	267-270
Rishabh Mishra - 128	97-100

Rishabh Rathod -32	233-236
Rishi Shah -7	111-116
Ritesh Mishra -6	105-110
Riya Sharma - 117	Appendix
Rohan Gupta -102	Appendix
Rohit Dubey - 126	175-178
Romil Badami -18	Appendix
Romit Bonkar -66	59-60
Rosy Bhoi -68	Appendix
Rosy Bhoi -69	Appendix
Ruchit Tripathi -104	217-220
Ruhi Bajaj -38	5-10
Rushikesh Shinde -33	237-240
Ruta Pathak -15	Appendix
Ruta Pathak -33	237-240

S

Sahil Jadhav - 80	Appendix
Sakshi Singh - 90	205-210
Sakshi Singh -33	237-240
Salonee America -120	89-92
Sandeep Pardeshi -25	47-52
Sanika Deshpande -34	Appendix
Sanketi Raut -68	Appendix
Sanketi Raut -69	Appendix

Sansriti Mishra - 121	263-266
Saqib Khan -42	Appendix
Saurabh Bagde - 79	65-68
Saurabh Indoria -9	1-4
Saurabh Rai -4	221-224
Saurabh Singh -20	141-144
Shaheen Sayyed -105	171-174
Shailesh Sangle -17	225-228
Shailesh Sangle -32	233-236
Shashank Salian - 130	Appendix
Sheetal Rathi -73	17-22
Sheetal Rathi -116	Appendix
Sheetal Rathi -20	141-144
Shikha Yadav - 92	Appendix
Shivam Mishra -6	105-110
Shivam Pandey -11	117-120
Shivani Yadav - 92	Appendix
Shiwani Gupta -14	Appendix
Shiwani Gupta -4	221-224
Shraddha Asthana -34	Appendix
Shradha Mhatre -11	117-120
Shreya Hegiste - 132	Appendix
Shreyash Sharma -74	61-64
Shruti Pimple - 111	85-88
Shubham Chhaparia -9	1-4
Shubham Patane -49	151-154

Shubham Singal -26	229-232
Shubham Vishwakarma -20	141-144
Siddharth Singh -22	Appendix
Siddharth Tiwari - 88	Appendix
Simran Malawat -11	117-120
Sneha Sankhe -65	Appendix
Snehal Patil -73	17-22
Sonu Mehta -31	Appendix
Sravani Arisetty -34	Appendix
Sruthi Harish - 90	Appendix
Sudarshan Dasarathy -104	217-220
Sudha Shankar -84	163-166
Sudhir Sawarkar - 95	251-254
Sufiyan Pawaskar -52	155-158
Sujata Alegavi - 115	Appendix
Sujay Salvi -33	237-240
Sunil Yadav - 126	175-178
Surabhi Singh - 111	85-88
Suraj Maurya -49	151-154
Suraj Singh - 108	Appendix
Suyash Awasthi -21	Appendix
Suyash Singh - 112	27-28
Swapnil Chowkekar - 126	175-178
Swaraj Kasliwal -19	125-128
Swastik Chavan -120	89-92

T

Tanmay Patil - 112	27-28
Tanushree Bhattacharjee -46	241-244
Tanvi Sanzgiri -105	171-174
Tanya Talwar -105	171-174
Tasneem Mirza -45	147-150

U

Uday Menon -26	229-232
Uzair Shaikh -58	Appendix

V

Vaidhya Chauhan -28	Appendix
Vaishali Nirgude - 96	211-216
Vaishali Nirgude -118	29-32
Vaishali Nirgude -8	Appendix
Vaishnavi Ganesh -119	Appendix
Vamsi Akhumukhi -10	41-46
Vani Singala -31	Appendix
Varun Shah -4	221-224
Vedashree Dalvi - 132	Appendix
Veena Kulkarni - 131	101-104
Veena Kulkarni -29	145-146

Veena Kulkarni - 88	Appendix
Veena Kulkarni -19	125-128
Vicky Prajapati -16	Appendix
Vidyadhari Singh - 110	81-84
Vidyadhari Singh -9	1-4
Vidyadhari Singh -10	41-46
Vikas Nirgude - 96	211-216
Vikas Singh - 133	33-36
Vikas Singh -21	Appendix
Vinay Dodeja - 124	Appendix
Vinitta Sunish -71	185-190
Vipul adwait - 125	93-96
Vipul adwait - 125	93-96
Vipul Babbar -39	129-132
Viral Haria -106	259-262
Vishal Rajak -29	145-146
Vishal Singh -104	217-220
Vishwa Solanki - 101	133-136
Vivek Giri -18	Appendix
Vivek Sharma - 88	Appendix

Y

Yash Kaul -7	111-116
Yash Nair -28	Appendix
Yash Shah - 110	81-84

Yash Shah - 81

Appendix

Yashraj Kotian -18

Appendix

Z

Zaid Shaikh - 80

Appendix

Zankar Murudkar -42

Appendix

