Lecture Notes in Electrical Engineering 643

Vinit Kumar Gunjan · Sabrina Senatore · Amit Kumar · Xiao-Zhi Gao · Suresh Merugu *Editors*

Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies

Springer

Lecture Notes in Electrical Engineering

Volume 643

The book series *Lecture Notes in Electrical Engineering* (LNEE) publishes the latest developments in Electrical Engineering—quickly, informally and in high quality. While original research reported in proceedings and monographs has traditionally formed the core of LNEE, we also encourage authors to submit books devoted to supporting student education and professional training in the various fields and applications areas of electrical engineering. The series cover classical and emerging topics concerning:

- Communication Engineering, Information Theory and Networks
- Electronics Engineering and Microelectronics
- Signal, Image and Speech Processing
- Wireless and Mobile Communication
- Circuits and Systems
- Energy Systems, Power Electronics and Electrical Machines
- Electro-optical Engineering
- Instrumentation Engineering
- Avionics Engineering
- Control Systems
- Internet-of-Things and Cybersecurity
- Biomedical Devices, MEMS and NEMS

For general information about this book series, comments or suggestions, please contact leontina.dicecco@springer.com.

To submit a proposal or request further information, please contact the Publishing Editor in your country:

China

Jasmine Dou, Associate Editor (jasmine.dou@springer.com)

India, Japan, Rest of Asia

Swati Meherishi, Executive Editor (Swati.Meherishi@springer.com)

Southeast Asia, Australia, New Zealand

Ramesh Nath Premnath, Editor (ramesh.premnath@springernature.com)

USA, Canada:

Michael Luby, Senior Editor (michael.luby@springer.com)

All other Countries:

Leontina Di Cecco, Senior Editor (leontina.dicecco@springer.com)

**** Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, SCOPUS, MetaPress, Web of Science and Springerlink ****

More information about this series at http://www.springer.com/series/7818

Vinit Kumar Gunjan · Sabrina Senatore ·
Amit Kumar · Xiao-Zhi Gao ·
Suresh Merugu
Editors

Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies

Springer

Editors
Vinit Kumar Gunjan
Department of Computer Science
and Engineering
CMR Institute of Technology
Hyderabad, India

Amit Kumar
BioAxis DNA Research Centre
Private Ltd
Hyderabad, India

Suresh Merugu
CMR College of Engineering
and Technology
Hyderabad, India

Sabrina Senatore
University of Salerno
Fisciano, Italy

Xiao-Zhi Gao
School of Computing
University of Eastern Finland
Kuopio, Finland

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-981-15-3124-8 ISBN 978-981-15-3125-5 (eBook)
https://doi.org/10.1007/978-981-15-3125-5

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Contents

Artificial Neural Network and Partial Pattern Recognition to Detect Malware

Aparna Gautam, Rajesh Gopakumar and G. Deepa

Abstract Cyber Security is a challenging research area nowadays because of the continuously increasing dependency of people on the Internet for day-to-day activities. Secure data transfer is a critical matter of concern for users while using the Internet. Secure data transfer demands that the data should be from a reliable and authenticated source. Secure data transfer is essential to avoid malicious code variants or software aka hidden malware. Detecting malware is a critical task as the current generation malware uses rapidly evolving techniques, and gets hidden in a genuine-looking file. However, malware detection can get a lot more progress if we deploy a machine learning technique, such as Deep learning for this purpose.

Keywords Malware detection · Security PE files · Opcode n-grams · Obfuscated malware · Artificial neural network · Partial pattern recognition

1 Introduction

Malware is a software that is created to attack a computer system to gain access and manipulate the critical data of a user. It can be used to breach the security and privacy of a user. Malware is of many types such as virus, Trojan, worm, botnet, rootkits, backdoors, spyware and adware etc. [1]. These malware programs harm the system security and data of the concerned user. So, detection of malware is vital before it damages the system or data in any way. Obfuscation is a technique used by malware creators to bypass malware detection mechanisms [2]. This evolving property makes the malware detection more difficult than that of a normal malware.

A. Gautam (✉) · R. Gopakumar · G. Deepa
Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India
e-mail: eraparnagautam@gmail.com

R. Gopakumar
e-mail: rajesh.g@manipal.edu

G. Deepa
e-mail: deepa.ganesan@manipal.edu

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_1

2 Literature Review

Previously a lot of work has been done to analyze and detect malware from a given data. Static and dynamic analysis is used to check the presence of malware in a program. To detect malware.

2.1 Signature-Based Malware Detection

Each malware-family has a unique bit string or byte-sequence known as malware signature [3] that is used to detect specific virus. Malware-signature checks for a specific bit string in the given executable file to detect malware.

2.2 Behavior-Based Malware Detection

Behavior-based malware detection system [4] detects the malicious code by judging its behavior. This analyses the potential mal-behavior by checking the services used and requests made by the program to access specific types of files. This technique takes too much time in scanning the potential malicious behavior of a file. This method uses two techniques to detect malware:

- **Family-Behavior Graph**

Family-behavior graph method [5] uses the concept of dependency graph to detect malware. It works only on known malware variants.

- **Machine-Learning Technique**

Machine Learning provides the facility of learning and improve for the existing system [6]. Here, we need not to program the system explicitly and the system does not remain as a static system and works dynamically on the given data [7]. Some of them are described below.

2.3 Heuristic-Based Malware Detection

Heuristic approach to detect malware is same as Behavior-based malware detection except for the fact that heuristic-approach works on detecting previously unknown malware as well. Evasion techniques are used to hiding the similarity of a new malware to the previously-known malware, so it is difficult to detect these malware through signature technique. Obfuscated malware also comes in the category of malware that use evasion techniques [8].

So, a lot of work has been done in the field of malware-detection. But there is still a long way to go as the malware detection schemes have some problems listed as follows:

- Malware Detection System is still not fully automated yet.
- Present malware-systems lack when it comes to detecting malware which use evasion techniques. So the objective of this work is to deploy deep learning and pattern recognition technique to detect obfuscated malware in the given data.

3 Methodology

The method built in this work to detect malware works on two main systems that are artificial neural network and partial pattern recognition. Artificial neural networks is used for constructing feature vector and classification purpose. Partial pattern recognition is used in this method for malware recognition and detection as it can recognize the malware even if it is partially similar to previous malware.

The proposed methodology works in six phases (Fig. 1).

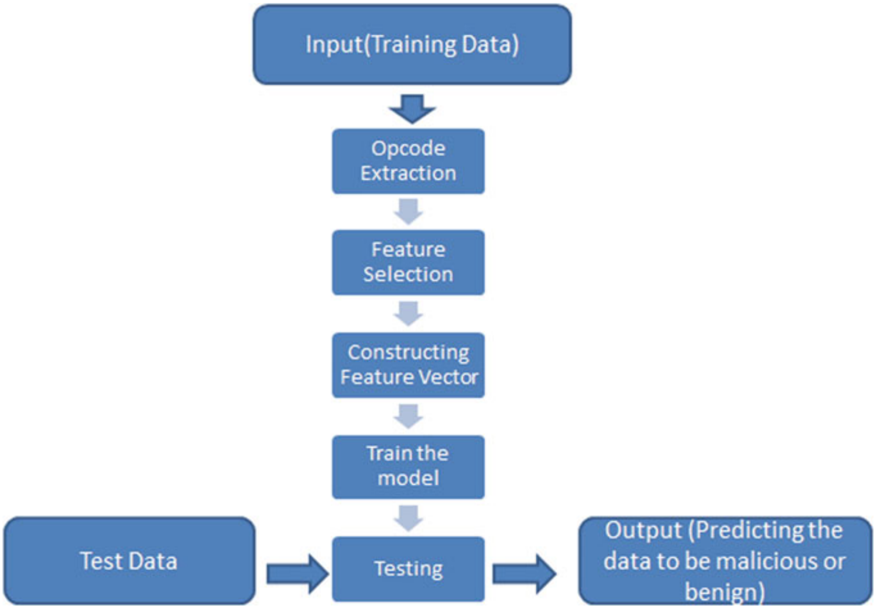


Fig. 1 Methodology

3.1 Opcode Extraction

Opcode extraction step comes under the pre-processing of data. As the malware files are packed, it is necessary to unpack them. For this purpose, unpacking of files and extraction of data is necessary. The malware file is scanned and all executables are unpacked using PEid [9]. These unpacked executables are portable executables files. After unpakcing executables, all PE files are decompiled using IDAPro [10] and extracted opcodes are extracted during this phase. In this work, 3 n-grams opcode extraction is conducted.

3.2 Feature Selection

In this work, initially seven features were selected that included pixel intensity count as well but in the later phase it was shown that pixel intensity feature was adding more complexity in creating the feature vector as it was not much beneficial in classifying the malware samples into their respective families so this feature was removed from the selected feature list to create more discrete feature vectors that can differentiate between the malware samples and classify them in malware families as specified in the kaggle dataset. Rest six features that were included in creating the feature vector are mentioned below:

- **ResourceSize**

Denotes the size of the resource section. Some virus files may have no resources. Clean files may have larger resources.

- **IatRVA**

Describes the relative-virtual address of the import-address table. The value IatRVA is 4096 for benign files and 0 for virus files.

- **VirtualSize2**

Describes the size of the virtual section. Most of the viruses usually have only one section and second section value is 0 for them.

- **ImageVersion**

Describes the version of the file. It is defined by user. Benign programs have more than one version and a large set of image-version. Malware have an ImageVersion value of 0.

- **Segment Count**

The segments in a memory model are used to group data into specific segments or blocks that are referenced by pointers located in the segment registers. Every segment describes only one specific kind of data.

- **DebugSize**

Defines the size of the debug-directory table. Most of the executable files consists of a debug directory. Hence benign have a non-zero value for DebugSize.

3.3 Constructing Feature Vector

After extracting opcodes from the PE files, opcode n-gram algorithm is applied on opcodes to get executable in the form of n-grams. In this work, n-gram of length three have been used to get sequential opcodes each of length three. The selected features are used to construct feature vector. On the basis of Document frequency, N/2 benign and N/2 malicious n-grams are selected and feature vector with N elements are produced as output. The output of this phase would be fine-tuned training dataset consisting of feature vector. After constructing feature-vector, classification of feature is needed to differentiate the specifications of each feature. For this purpose, artificial neural network are used in this work.

3.4 Feature Classifier

The artificial neural networks contain neurons and are used as a method to extract information from the raw data. Neural network [11] gives good performance with noisy data as well. It contains three components i.e. input, hidden layer and output. Hidden layer can be more than one and each layer contains some feature vectors that help in classifying data more precisely. Input layer depends upon the number of input that are being given to neural network. The number of internal layers is adjusted according to the feature vector chosen. The output of one hidden layer is given as the input of second hidden layer and so on. In this work, number of neurons in each hidden layer are the mean of neurons of input layer and output layer.

Let us consider the number of neurons in each hidden layer as HN; Number of neurons in input layer as IN and number of neurons in output layer as ON; then number of neurons in each hidden layer i.e. HN would be:

$$HN = \frac{(\mathrm{IN} + \mathrm{ON})}{2}.$$

In this paper, feed-forward back propagation artificial neural network is used. To construct artificial neural network, the tool used is Neural Designer [12] and required conditional coding part is done in Python language. Artificial Neural Network uses activation function as an input to the neurons of the hidden layer. The activation function used in this work are Rectifier Linear Unit (ReLU) and Softmax. Here, ReLU is used in the hidden layers of the artificial neural network because it helps the ANN in back-propagating the error and improving the learning rate. The range of ReLU is 0 to x. Softmax is used in the output layer of the ANN as this activation function is specifically designed for the inputs that are needed to be classified in multiple output categories. The range of this Softmax is 0 to 1 for each input class.

3.5 Training the Model

In this step, the partial pattern recognition algorithm is trained using the classified samples to create final part of the proposed model. Pattern Recognition is used to recognize specific pre-classified pattern in the input data. Partial Pattern Recognition is a pattern recognition technique that does not need complete pattern matching to determine the similarity between input data and pre-classified patterns. So, by using partial pattern recognition technique, we can determine if an input is benign or malicious without scanning the entire input data that results in faster detection of malware.

3.6 Testing

Test datasets used in this paper, are given as an input to model and output is produced (Fig. 1) in the form of predicting whether the given sample is malicious or benign. The combination of feature classifier and partial pattern recognition algorithm is used to detect malicious input data. The ANN extracts and classifies the feature of input while detection framework (partial pattern recognition model) matches the pattern similarities of input data to the malicious patterns, and if found any; input data is considered as malware.

4 Dataset Used

In this paper, two datasets have been used. These datasets are by Microsoft Kaggle and contains obfuscated malware. The huge training and testing data provided by the datasets helps in building a reliable and effective malware detection method. Second dataset mentioned in this section provides a good amount of benign portabel executable files that helps in training the artificial neural network. These datasets are specified as follows:

1. Microsoft Kaggle 2015 Dataset [13]
2. Benign and Malicious PE Files [14].

Table 1 Malware family classification

Malware family	Number of samples
Ramnit	1541
Lollipop	2470
Vundo	470
Simda	45
Tracur	754
Kelihos_ver1	400
Kelihos_ver3	2944
Gatak	1227
Obfuscator.ACY	1012

Table 2 Comparison

Detection framework	Dataset used	Accuracy achieved (%)
Generating malware images	Hauri, Kaist cyber security research center	96
DeepAM	Comodo cloud security center	97.8
DFEL	NSL-KDD	99.29
CNN	Vision research lab	94.5
Our approach	Kaggle datasets	99.85

5 Results

A malware detection system that is automated and can detect unknown malware in real life without human-intervention. In this work, opcode extraction work is done with 3-gram opcode as 3 n-gram opcode gives more explicit view of the file code. Artificial Neural Network classifies the given dataset into malware family classes. The classification results obtained by ANN are shown in Table 1.

Accuracy achieved while classifying features using ANN is 100%. The proposed framework detects 10,857 inputs accurately out of 10,873 test data inputs while combining whole detection framework (ANN and Partial Pattern Recognition) gives 99.85% accuracy. A comparison study of the accuracy achieved by other malware detection frameworks based on deep learning and our approach is shown in Table 2.

References

1. Bettany A, Hasley M (2017) Windows virus and malware troubleshooting
2. You I, Yim K (2010) Malware obfuscation techniques: a brief survey. In: International conference on broadband, wireless computing, communication and applications

3. Zhou Y, Han M, Liu L, He J, Wang Y (2018) Deep learning approach for cyberattack detection. In: IEEE conference on computer communications workshops (INFOCOM WKSHPS), IEEE Infocom MiseNet Workshop
4. Dornhackl H, Kadletz K, Luh R, Tavolato P (2018) Defining malicious behavior. In: 9th International conference on cyber conflict
5. Ding Y, Xia X, Chen S, Li Y (2018) A malware detection method based on family behavior graph. Comput Secur 73:73–86
6. Xin Y, Kong L, Liu Z, Chen Y, Li Y, Zhu H, Gao M, Hou H (2018) Wang C (2018) Machine learning and deep learning methods for cybersecurity. IEEE Access 6:35365–35381
7. Apruzzese G, Ferreti L, Marchetti M, Colajanni M, Guido A (2018) On the effectiveness of machine and deep learning for cyber security. In: 10th International conference on cyber conflict
8. Bazrafshan Z, Hashemi H, Mehdi S, Fard H, Hamzeh A (2013) A survey on heuristic malware detection techniques. In: 5th Conference on Information and Knowledge Technology (IKT)
9. PEid Tool. https://www.softpedia.com/get/Programming/Packers-Cryptors-Protectors/PEiD-updated.shtml
10. IDA Pro. https://www.hex-rays.com/products/ida/support/download.shtml
11. Neural Designer. https://www.neuraldesigner.com
12. Raman K, Selecting features to classify malware
13. Kaggle Dataset. https://www.kaggle.com/c/malware-classification/data
14. Benign and Malicious PE Files. https://www.kaggle.com/amauricio/pe-files-malware/home

Epidemiology and Forecasting of Cholera Incidence in North India

Kumar Shashvat, Rikmantra Basu, Amol P. Bhondekar and Arshpreet Kaur

Abstract Cholera is a gastrointestinal infection caused by vibrio cholera. The cholera becomes dangerous if not treated on time; it may leads to dehydration and death. In this paper the objective is to develop and compare the predictive models to forecast the number of cholera incidence. The second objective of the paper is to see correlation with number of cholera incidence with ecological variable. The Monthly cholera incidences from 2014 to 2017 were taken from integrated diseases surveillance programme, Government of India. The data was analysed by four statistical models Autoregressive integrated moving average (ARIMA), exponential smoothing (decomposition method), linear regression and XG Boost. Based on model comparison indices such MAE and RMSE it was found that exponential smoothing model and XG Boost model performed equally better in comparison to all the models used for analysis in terms of mean square error and root mean square error.

Keywords ARIMA · Exponential smoothing · Linear regression · Cholera

1 Introduction

Cholera is one of the most epidemic diseases especially in the developing countries like India, Pakistan and Bangladesh. The cholera is a gastrointestinal infection caused by vibrio cholera symptoms of cholera are diarrhoea and vomiting, after an incubation period of 2 h to 5 days [1–4]. Cholera has continued to be global threat to a public health and is also epidemic in parts of Asia, the Middle East. The features that may lead to endemic areas include flooding earthquakes poor sanitation conditions in the particular area. Worldwide, about 1.4–4.3 million cases and 28,000–142,000 deaths are due to the occurrence of cholera in a region.

K. Shashvat (✉) · R. Basu · A. Kaur
National Institute of Technology, Delhi, India
e-mail: shashvat.sharma13@gmail.com

A. P. Bhondekar
Central Scientific Instruments Organization, Chandigarh, India

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_2

A microbiological and ecological variable plays an important role for the occurrence of disease. Cholera is one of the most rampant water related infectious diseases in many regions of the world especially a country like India where the slum areas are there where poor sanitation is there. The explicit alternations in the climatic conditions influence the spread of an infectious disease. The V.cholerae is highly encouraged by the contamination of food and water.

The cholera has been regularly resurfacing in and around Chandigarh, concurs the onset of the monsoon season from June–September from the past few years. The importance of the collection of epidemiological data on cholera outbreaks lies in formulation the plans and preparedness to suit a particular affected area of the region. The quick actions can be taken in the field of epidemiology before it will proceed to an endemic and larger outbreaks will lead to the higher morbidity and mortality rate of diseases [5–8].

The spread of the infectious diseases can be predicted by applying various prediction models. In this paper we develop and compare the predictive models for the occurrence of cholera in a Chandigarh region. The incidence of this disease is mainly depends upon the contamination of water and food, but the ecological variables such as the rainfall and humidity also plays an important role in order to have rise in the number of cholera incidence. To investigate this we have used four predictive models.

2 Materials and Methods

2.1 Data Collection

To achieve the objective of this paper database of cholera incidence has been provided by integrated diseases surveillance programme, Government of India. The respective data for [2014–2017] of the patient has been collected. To complete the study regarding the relationships of the ecological variables such as rainfall and humidity was collected from www.indiastat.com.

2.2 Application of Software Used for Data Analysis

The cholera database is analyzed using four different predictive models which are exponential smoothing model Autoregressive integrated moving average model. The complete analysis was performed in R studio (R 3.5). Table 1 shows the scenario of cholera incidence from year **2014** to **2017**.

Table 1 Scenario of cholera incidence 2014–2017

Months	2014	2015	2016	2017
Jan	0	1	0	0
Feb	0	0	0	0
Mar	0	0	0	0
Apr	0	0	0	0
May	0	0	0	1
Jun	0	0	36	5
Jul	1	3	18	1
Aug	0	56	1	3
Sept	5	6	2	0
Oct	0	1	0	0
Nov	0	0	0	0
Dec	6	0	0	2
Total	12	67	57	12

2.3 Methods

2.3.1 Descriptive Analysis

Table 2 shows the seasonal incidence of each month between years **2014** and **2017**. It can be clearly depict that from July to September there is an evaluation trend in number of cholera incidence.

Table 2 Seasonal incidence 2014–2017

Months	2014	2015	2016	2017
Jan	0	0.18	0	0
Feb	0	0	0	0
Mar	0	0	0	0
Apr	0	0	0	0
May	0	0	0	0.16
Jun	0	0	7.57	0.83
Jul	0.16	0.54	3.78	0.16
Aug	0	10.1	0.21	0.5
Sept	0.83	1.09	0.42	0
Oct	0	0.18	0	0
Nov	0	0	0	0
Dec	1	0	0	0.33

2.3.2 Autoregressive Integrated Moving Average Model (ARIMA)

Autoregressive Model

An AR model is one of the time series models in which y_t is depends only its past values y_{t-1}, y_{t-2}, y_{t-3} etc. Here y_t is the function of the past values. The common representation of an AR model where it is depend on p of its past values called as AR (p) model is represented as

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + \cdots + \beta_p y_{t-p} + \varepsilon_t \tag{1}$$

Moving Average Model

The Moving average models are the models where the forecast of the value is done by taking the error terms [5]. In this model y_t depends upon the random error terms which follow a white noise process i.e. $y_t = f(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2} + \cdots)$. A common representation of the model where it is depend on q of its past values called as MA (q) model is represented as

$$y_t = \beta_0 + \varepsilon_t + \varnothing_1 \varepsilon_{t-1} + \varnothing_2 \varepsilon_{t-2} + \cdots + \varnothing_q \varepsilon_{t-q} \tag{2}$$

Autoregressive Moving Average Model

The ARMA model uses the past values as well as the error terms to forecast the value. The general form of this model which depends upon the p of its own past values and q past values of the white noise.

$$\begin{aligned} y_t = {} & \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + \cdots + \beta_p y_{t-p} \\ & + \varepsilon_t \beta_0 + \varepsilon_t + \cdots + \varnothing_1 \varepsilon_{t-1} + \varnothing_2 \varepsilon_{t-2} + \cdots + \varnothing_q \varepsilon_{t-q} \end{aligned} \tag{3}$$

Autoregressive Integrated Moving Average Model

ARIMA model has been widely used in forecasting of time series data. It is instigated by AR model, MA model and the amalgamation of AR and MA model [6, 7]. In ARIMA model a non-stationary time series is made stationary by pertaining differencing of the data points.

The common mathematical representation of this model is ARIMA (p, d, q).

$$\gamma(L)(1 - L)^d x_t = \theta(L)\varepsilon_t \text{ i.e.}$$

$$\left(1-\sum_{i=1}^{p}\gamma_i L^i\right)(1-L)^d x_t=\left(1+\sum_{k=1}^{q}\theta_k L^k\right)\varepsilon_t \quad (4)$$

here, p, d, q are integers greater than or equal to zero and consign to the order of the autoregressive, integrated and moving average parts of the model. The d reins the level of difference. Generally, d = 1 is enough in most of the cases.

2.3.3 Linear Regression

A linear regression is a linear approximation of casual relationship between two or more variables. There is dependent variable i.e. is predicted y and independent variable $x_1 + x_2 + x_3 + x_k$ predictor. The dependent variable y is a function of independent variable x_1 to x_k $y = F(x_1 + x_2 + x_3 + x_k)$. The regression models are highly valuable as they are the most common ways to make predictions [8].

The mathematical representation of simple linear regression model is as follows

$$y = \beta_0 + \beta_1 X_1 + \varepsilon \quad (5)$$

Here y is the dependent variable and X_1 is the independent variable.

2.3.4 Exponential Smoothing

The exponential smoothing model is a kind of forecasting technique which uses the linear combination of the past values of the series Forecast = F(y_{t-1}, y_{t-2}, y_{t-3} ...). Here the future forecast is the function of different lag period $y_t, y_{t-1}, y_{t-2}, y_{t-3} \ldots$

$$s_t = \alpha_0 + \alpha_1 y_t + \alpha_2 y_{t-1} + \alpha_3 y_{t-2} \quad \alpha_1 > \alpha_2 > \alpha_3 \quad (6)$$

$s_t =$ Forecast value of the time series.

2.3.5 XG Boost (Extreme Gradient Boosting)

The XG boost is an ensemble model; it builds a decision tree iteratively. It assigns more weight to the wrong prediction and the basic tuning parameters for this model are the trees, depths. The XG boost builds a more accurate model the gradient boost as objective function is a sum of specific loss function [9].

$$\mathrm{obj}(\theta) = \sum_{i=1}^{n} l\left(y_i - \widehat{y_i}\right) + \sum_{k=1}^{k} \omega(f_k) \quad (7)$$

3 Epidemiological Features

The Cholera is mutually epidemic as well as the endemic disease. The epidemicity and endemicity of the diseases will depend upon the characteristics of the agent and those of system called environment. Characteristics of the agent will influence its ability to survive in the environment, its influence the average number of organisms is required to cause infection.

3.1 Agent

The organism that causes cholera is labelled as V. cholerae O group 1 or vibrio cholera O1 and O139. The term epidemic strain is also used for these vibrios. The vibrios is biochemically similar to the epidemic strains (vibrio cholera O1 and O139), but do not agglutinate in vibrio cholera O1 and O139 antiserum have been referred to in past as non agglinating (NAG) or as non cholera vibrous [10].

3.2 Mode of Transmission (Vibrio)

The transmission of vibrio can be occurred through man to man via uncontrolled water resource such as waste water from lakes, ponds it poses a great threat in transmission of the bacteria. Ingestion of the contaminated drinks is also associated with the outbreak of cholera incidence [11, 12]. Fruits and vegetables washed with the contaminated water is also a major source for the spread of the infection. The incubation period of cholera is from few hours up to five days, but commonly 1–2 days.

Faecally Contaminated Water → Contaminated Food and Drinks → Direct Contact

3.3 Environmental Factors

The vibrio transmission is readily possible in a community with poor environmental sanitation. The environmental factors of importance include contaminated food and water [13]. There are various social factors which are responsible for the transmission and occurrence of the cholera in India.

4 Results and Discussions

There were 148 cholera incidences in total from 2014–2017 in Chandigarh region as shown in Table 1. There was cholera outbreak every year of the study period. The years of 2015, 2016 had relatively higher number of incidence as compared to the year 2014 and 2017. The seasonal incidences calculated as shown in Table 2. It is clearly depicted from the seasonal incidences table that the outbreaks concentrated in **June to September** every year, which are comparatively warmer months. as well as the incidences are more in the monsoon season as compared to coldest months from **December to March**.

This paper extracted the yearly correlation of the ecological variables such as the rainfall and humidity with the number of cholera incidence for the study period. To find out the relationship between the number of cholera incidence and between the ecological a p value staticall test has been applied and coefficient of correlation has been calculated as shown in Table 3. The coefficient of correlation r clearly shows that the rainfall has a positive correlation with the number of cholera incidence but in case of humidity the humidity plays a major role for the outbreak of cholera but not significant as compared to the rainfall.

To validate the performance of our models we have used two performance measures mean absolute error and root mean square error. Table 4 shows the error comparison of various predictive models that has been applied to forecast the number of cholera incidence for the next year. Table 4 validates that the exponential smoothing performed relatively better as compared to other predictive models. Table 5 depicts the forecast value from the models that has been applied. The novelty of the work is

Table 3 Correlation of rainfall and humidity with cholera Incidence

Year	p value (Humidity)	p value(Rainfall)	Coefficient of correlation (rainfall)	Coefficient of correlation (humidity)
2014	0.087	0.28	0.34	−0.055
2015	0.2	0.16	0.44	0.4
2016	0.087	0.26	0.35	−0.055
2017	0.97	0.013	0.69	0.013

Table 4 Comparison of models

Models	Mean absolute error	Root mean square error
ARIMA	6.118924	9.632753
Exponential smoothing	4.513889	9.632753
Linear regression	6.093678	10.83353
XG Boost	5.268956	8.387952

Table 5 Forecast value till October 2018

Models	Forecast value
ARIMA	40
Exponential Smoothing	40
Linear Regression	26
XG Boost	33

that we have developed a new model to forecast the number of incidence as well as analyzed the application of predictive modelling such as XG boost which is a new application on the medical databases.

5 Conclusion

This study has analyzed the ecological variables which are responsible for the outbreak of cholera in a Chandigarh region. The study is based upon the four years monthly data of cholera outbreak. Our results show that the ecological variables are an important indicator for the spread of cholera in a particular region. The staticall analysis shows that the rainfall plays a significant role in the development of cholera incidence. The rainfall has strong effect on the cholera. Our staticall analysis compares the four predictive models. Our result shows that the geographical condition also plays a role in the occurrence of a probability of diseases. The forecast value for the transmission of the diseases will be an early warning for a spread of diseases in a particular region. The results will be more significant by combing the environmental and geographical factors will enhance the prediction of infectious diseases.

Acknowledgements The author thankfully recognizes the support from Integrated Diseases Surveillance Programme, Chandigarh for sharing the database. We express gratitude to Director Health and Services, Chandigarh Dr. G. Dewan for providing sanction for the collection of information of Chandigarh district.

Ethics and Consent The data used in this manuscript was sourced from the Integrated Diseases Surveillance Programme, Govt of India with permission no. E-151913.

Competing Interest No competing interest was reported by the authors.

References

1. LeBrun DG, Talwar D, Pham TA, Banskota B, Spiegel DA (2017) Predictors of healthcare seeking delays among children with chronic musculoskeletal disorders in Nepal. J Epidemiol Glob Health 7(4):299–304. https://doi.org/10.1016/j.jegh.2017.10.002
2. Rajendran K, Sumi A, Bhattachariya MK, Manna B, Sur D, Kobayashi N, Ramamurthy T (2011) Influence of relative humidity in Vibrio cholera infection: a time series model. Indian J Med Res 133(2):138–145

3. Adagbada AO, Adesida SA, Nwaokorie FO, Niemogha MT, Coker AO (2012) Cholera epidemiology in nigeria: an overview. Pan Afr Med J 12(1):1–12. https://doi.org/10.11604/pamj.2012.12.59.1627
4. Of L, Month THE (2002) Health services and the public health, pp 835–840
5. Van den Bergh F, Holloway J, Pienaar M, Koen R, Elphinstone C, Woodborne S (2011) A comparison of various modelling approaches applied to Cholera case data. ORiON 24(1):17–36. https://doi.org/10.5784/24-1-57
6. Zeeshan Zafar M, Gulzar H (2016) A case study: cholera. Occup Med Health Aff 04(06):4–6. https://doi.org/10.4172/2329-6879.1000252
7. Pezeshki Z, Tafazzoli-Shadpour M, Nejadgholi I, Mansourian A, Rahbar M (2016) Model of cholera forecasting using artificial neural network in Chabahar City. Iran. Int J Enteric Pathog 4(1):1–8. https://doi.org/10.17795/ijep31445
8. Xu M, Kan B, Wang D (2015) Identifying environmental risk factors of cholera in a coastal area with geospatial technologies. Int J Environ Res Pub Health 12(1):354–370. https://doi.org/10.3390/ijerph120100354
9. Puri S, Sarpal SS, Kumar A, Goel NK (2014) An outbreak of Vibrio cholerae in Vikas Nagar, Chandigarh, India. Med J Dr. DY Patil Univ 7(2):147–151. https://doi.org/10.4103/0975-2870.126320
10. Kumar M, Sharma VL (2008) Outbreak investigation of cholera in a slum of Northern India. Open J Epidemiol 4(4):243–247. https://doi.org/10.4236/ojepi.2014.44031
11. Sule IB, Yahaya M, Aisha AA, Zainab AD, Ummulkhulthum B, Nguku P (2017) Descriptive epidemiology of a cholera outbreak in Kaduna state, northwest Nigeria, 2014. Pan Afr Med J 27:1–8. https://doi.org/10.11604/pamj.2017.27.172.11925
12. Siziya S (2017) A review of the epidemic-prone enteric diseases in Zambia: cholera, typhoid fever and bacterial dysentery. Health Press 2(1):6–11
13. Jutla A, Whitcombe E, Hasan N, Haley B, Akanda A, Huq A, Alam M, Sack RB, Colwell R (2013) Environmental factors influencing epidemic cholera. Am J Trop Med Hyg 89(3):597–607. https://doi.org/10.4269/ajtmh.12-0721

A Novel Approach for Improving Routing Efficiency and Minimizing Propagation Mean Delay for DADCQ Protocol in VANET

Dharmaveer P. Choudhari and S. S. Dorle

Abstract VANET provides the information in the dynamically changing environment in the rapidly changing topological environment so as to provide the seamless flow of data between the vehicles in the in ad hoc environment. Basically VANET is the subset of MANET. For effective routing from the source vehicle to destination vehicle in the dynamic environment the broadcast protocol must withstand this dynamically changing topology of the network. The mobility of network in urban scenarios is highly variable so the protocol must be able to adapt changes as per the topology of VANET. This paper focuses on the detection of the attacks such DDoS and Eavesdropping attacks and its effect to the protocol parameter such routing efficiency and mean delay. After detection of these attacks a novel solution is adopted from improving routing efficiency and mean delay of the protocol so as to make network reliable and more secure.

Keywords DADCQ · Efficiency · D_c · Routing

1 Introduction

The main purpose of the VANETs is to provide the real-time information sharing of the road and safety messages those are sent by the various vehicles present in the VANET network. So by this way we can very safely avoid the accidents on the road not also this but by providing the accurate information about the traffic we can also saves the energy consumption of the vehicles present in the VANET network. Talking more about VANET networks these are also capable of providing the services like news, FM and entertainment which provides the fun to the driver along with the safety messages along with the information of the network.

D. P. Choudhari (✉) · S. S. Dorle
Department of Electronics Engineering, G.H. Raisoni College of Engineering, Nagpur, India
e-mail: dharmaveerc@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_3

1.1 Literature Review

Broadcast protocol is usually designed to broadcast the message in one or the two directions with the high reliability and optimal latency the examples of such type is as explained in [1–3]. Several protocols for VANET have been made and discussed. Whereas Tonguz et al. have suggested the forwarding node selection methods those are based on the fundamental statistical methods as mention above in. After analysis we found that the p-persistence protocol gives the best results. And more detail study of these methods for the highway scenarios which has been explained and rolled in the scheme called as DV-CAST as explained in. But the DV-CAST has shown the poor performance in the urban scenarios so in this case the authors have proposed the variant which is called as UV-CAST that is designed for the urban scenarios.

2 Application Distance Method for 1D and 2D Networks

Distance method uses the finest distance from transmitting node to receiving node prejudice between re-broadcasters node and the non re-broadcasters node in the network. So initialize $D = 1$ and check if a message is received set ‘d’ to the distance to sender $D = \min \{D, d/r\}$ so set a random back off timer [4].

For this we need to find the minimum value of D_C at the many values of the global density which is termed as ‘λ’. Then we must plot the optimal values and we need to find the suitable estimate function D_C (λ). Then substitute the local node density for the global node density ‘λ’ to get the value of D_C (N). The values of the variables D_{max}, β and α depends upon the external factors such node distribution pattern and channel quality.

$$D_C(N) = D_{max} - \beta e^{\alpha N} \quad (1)$$

2.1 Analysis of Spatial Distribution in VANET

Spatial Distribution is nothing but the science of analyzing how the nodes are in the network and how they form the pattern in the network. We can choose any sample size and any radius ‘r’ for the transmission. Here we have chosen sample size as 40 and the radius as $r = 8$ for our simulation results. Here the sample size is chosen as 30 and a circular cell and radius $r = 5$ where r is the radius of transmission. The process of calculating the frequency values for nm is as such initialize $nm = 0$ for $m = [0, N]$ and generate random point , y within the node’s transmission area, set

= 0. For neighbor & node itself, if it is within r/5 of x, y increment m & increment nm. The quadrant function is indicated as follows:

$$Q = \frac{V[n]}{E[n]} \tag{2}$$

2.2 Adaption of Node Distribution Pattern for 1D and 2D Uniform Distributions

As discussed earlier the DADCQ protocol uses the distance method for the forwarding nodes for 1D and 2D distribution patterns. For the method the adaption to the node density means that rebroadcasting threshold Dc varies when the distribution pattern of the nodes in the network changes [5]. Dc depends upon the three factors α, Dmax and β.

As discussed earlier Dc is a function of α, D_{max} and β so at several node densities which provides at least 99% reach ability in many of simulations runs we need to these values by hand selecting the values for Dmax and then performing a least mean squares to calculate the values of α and β.

Table 1 shows the calculated values of the function parameters which are found by using this method as discussed above. As shown in Fig. 1 compares the threshold curves foe 1D and 2D distributions. Which clearly shows that threshold value must be lower for the 1D network. The parameter which differs between the two curves is the exponential multiplier of the rate which is 'α'.

Table 1 Node density and threshold values 1D and 2D networks

Distribution pattern	Dmax	α	β
1-D uniform	0.91	−0.040	1.2
2-D unifrom	0.91	−0.098	1.4

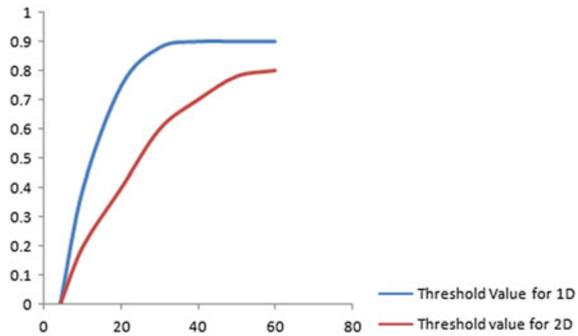


Fig. 1 Node density versus threshold value for 1D and 2D networks

Table 2 Values for K, D_{max}, α and β for 1D and 2D network

Type of network	K	D_{max}	α	β
1D	16	0.610	−0.060	1.00
1D	20	0.620	−0.060	0.80
2D	8	0.688	−0.060	5.00
2D	12	0.700	−0.060	4.88

3 Reduction of Fading and It's Analysis in the Network

Fading in the VANET can be modeled ad the Rician Fading. Fading signal is nothing but the line-of-sight which is scattered among the multiple paths our need is to find the function that so that the fading can be quantified and it should adapt to the threshold value [6]. For our network simulations we have used Rician Fading function and here in this k indicates the relative strength of the line-of-sight signal in Rician fading (Table 2).

The standardization for the VANET network is around the standard IEEE 802.11 Dedicated Short Range Communication (DSRC) and which uses the frequency range of 5.85–5.92 GHz band.

4 Design of Quadrant-Based Threshold

In case of VANETs we have considered the both distributions 1D distributions and 2D distributions so we need a mathematical function that will directly become accustomed to these 1D and 2D scenarios of the network and such as function which is capable of doing is Quadrant function which will accomplish the task in the network. Here the sample size is chosen as 30 and a circular cell and radius $r = 5$ where r is the radius of transmission. The process of calculating the frequency values for nm is initialize $nm = 0$ for $m = [0, N]$ and generate random point, y within the node's transmission area, set = 0. And for neighbor & node itself, if it is within r/5 of x, y increment m and nm.

5 Analysis of Routing Efficiency and It's Improvement for DADCQ Protocol in VANET

In topological routing each node must be aware of the layout of the network and should also be able to forward the packets using the available information about the links and the nodes in the network [7]. Whereas position based routing should be aware of the nodes location in the packet forwarding for the network. Another

Table 3 Showing values of routing efficiency for eavesdropping and DDoS attacks under different scenarios

Number of communications	Routing efficiency when eavesdropping and DDoS attacks are considered	Routing efficiency when only eavesdropping attacks are considered	Routing efficiency when only DDoS attacks are considered	Routing efficiency after removal of eavesdropping and DDoS attacks
5	2.77	3.47	2.51	5.01
10	1.84	2.89	2.4	3.69
15	2.24	1.74	2.23	2.74
20	1.01	2.04	2.58	2.54
25	1.04	1.64	2.02	2.74
30	1.07	1.71	1.73	2.35
35	1.42	1.29	1.45	2.42
40	1.12	1.29	2.73	2.06
45	0.82	1.21	1.52	1.95
50	0.73	0.89	1.77	1.77

important type of routing protocol is the proactive protocol. Each of the entry in the table contains the information of the next hop node which is used in the path to the destination node.

Table 3 shows the data for the routing efficiency of the DADCQ protocol under the different situations. The column 1 indicates the number of communication being performed between the nodes, column 2 indicates the routing efficiency when no attacks are removed in the network.

The column 3 and 4 indicates the routing efficiency considering eavesdropping and DDoS attack single at a time while masking the other attack if first attack is taken. And finally the last column indicates the improved efficiency as compared to the column 2, 3 and 4.

6 Analysis and Minimizing Mean Delay for DADCQ Under Eavesdropping and DDoS Attacks

We have considered the various situations of mean delay. Now in this section our analysis is for mean delay due to Eavesdropping nodes and DDoS nodes in the VANET. Eavesdropping is nothing but the unauthorized reading or accessing the data of the users. These Eavesdropping attacks are known as the sniffing attack in which someone is trying to steal the data that the vehicles or computers have transmitted over the ad hoc network.

Table 4 Showing values of eavesdropping and DDoS attacks under different scenarios of attacks

Number of communications	Mean delay when eavesdropping and DDoS attacks are considered	Mean delay when only eavesdropping attacks are considered	Mean delay when only DDoS attacks are considered	Mean delay after removal of eavesdropping and DDoS attacks
5	1009	918	277	128
10	1062	1008	262	143
15	1122	898	206	179
20	1136	920	163	106
25	1048	893	133	85
30	1016	922	88	92
35	1131	929	115	65
40	1112	868	82	87
45	1081	868	81	47
50	1068	927	89	47

Here we considering the DDOS attacks and what are their effects on the communication mean delay. The following Table 4 shows that the as DDOS attacks are present in the network so the mean delay is initially high and as the number of communication increases mean delay decreases.

Table 4 above shows the values of number of communication and the values of the mean delays due to the DDOS attacks and Eavesdropping attacks in the network.

7 Conclusion

From Table 3 of columns 2, 3 and 4 we may observe that due to DDOS and Eavesdropping we got much lower values of the routing efficiency for the protocol which clearly suggests that some other nodes those are attackers nodes which are utilizing the network resources and so the routing efficiency of the network is reduced and making protocol less efficient. The column 5 of Table 3 and shown in Fig. 2 shows the values of the routing efficiency after removal of these attacks and so because of which we got the higher values for routing efficiency which is been also indicated in Fig. 2. Similarly for the Table 4 of columns 2, 3 and 4 we may observe that due to DDOS and Eavesdropping we got much higher values of the mean delay which clearly indicates that the attacker nodes are consuming the resources of the network. Also the column 5 of Table 4 and also shown in Fig. 3 shows the values of the mean delay after removal of these attacks and so because of which we got the lower values for mean delay. This is also indicated in Fig. 3. So finally we conclude that we have

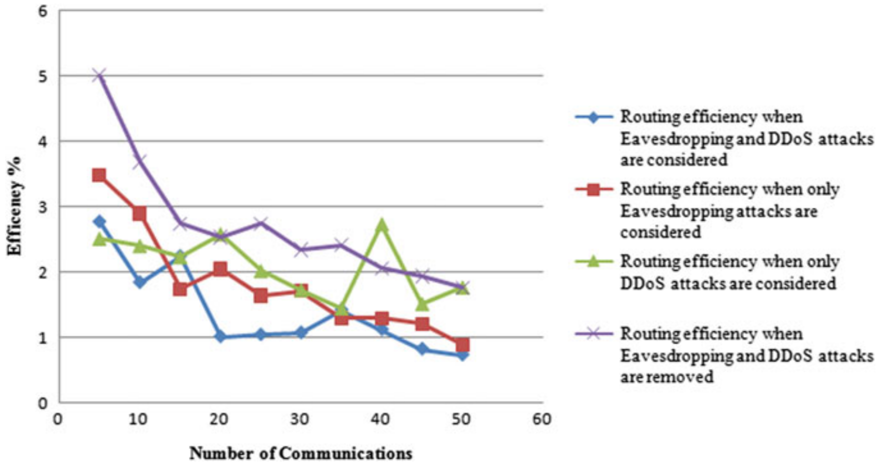


Fig. 2 Routing efficiency under eavesdropping and DDoS attacks and routing efficiency after the removal of both attacks

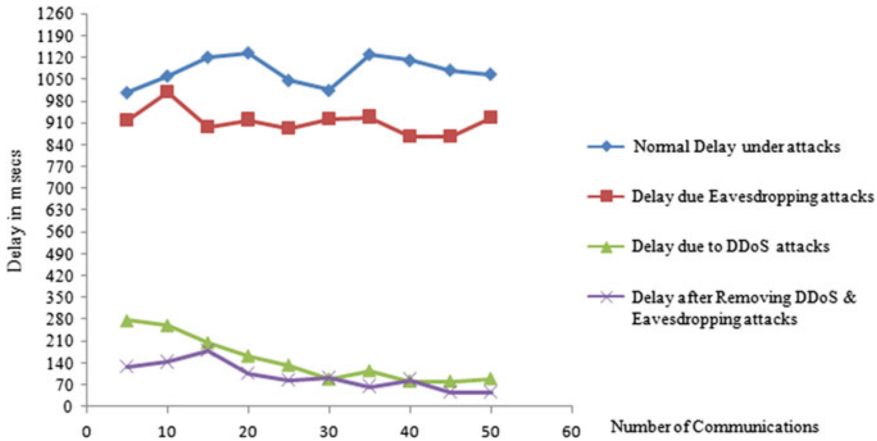


Fig. 3 Mean delay under eavesdropping and DDoS attacks and mean delay after the removal of both attacks

efficiently detected the eavesdropping and DDoS attacks and we have improved the routing efficiency and mean delay of the protocol to large extent.

References

1. Janech J, Lieskovsky A, Krsak E (2012) Comparation of strategies for data replication in VANET environment. In: 26th International conference on advanced information networking and applications workshops
2. RoselinMary S, Maheshwari M, Thamaraiselvan M, Early detection of DOS attacks in VANET using attacked packet detection algorithm (APDA)
3. Leccese F (2013) Remote-control system of high efficiency and intelligent street lighting using a ZigBee Network o f devices and sensors. IEEE Trans. Power Deliv 28(1):21–28
4. Slavik M, Mahgoub I (2013) Spatial distribution and channel quality adaptive protocol for multihop wireless broadcast routing in VANET. IEEE Trans Mob Comput 12(4):722–734
5. Vidgren N, Haataja K, Patino-Andres JL, Ramirez-Sanchis JJ, Toivanen P (2013) Security threats in ZigBee-enabled systems: vulnerability evaluation, practical experiments, countermeasures, and lessons learned. In: 46th Hawaii international conference on system sciences
6. Tayal S, Tripathy MR, VANET-challenges in selection of vehicular mobility model
7. Slavik M, Mahgoub I (2010) Statistical broadcast protocol design for unreliable channels in wireless ad-hoc networks. In: IEEE globecom 2010 proceedings
8. Slavik M, Mahgoub I (2012) Analysis of beaconing message rate in VANET multi-hop broadcast protocols. IEEE
9. Park YH, Park YH, Moon SJ (2010) ID-based private key update protocol with anonymity for mobile ad-hoc networks In: International conference of computational science and its applications (2010)

Human Computer Interaction as a Mode of Improvement in Human Lifestyle

Sani Hasan

Abstract The following paper discusses about some of the current commonly available new technology in the society along with their current uses and the other aspects of their uses, while focusing on the reasons behind their need and place in the modern society and their duty in the betterment of the human life in the modern times by the use of the technology.

Keywords HAL, Hybrid Assistive Limb · CASPER, Centralized Automatic Sleeping Pod Enhanced Revitalization

What is to be assumed by the human computer interaction in today's society? Is it simply a person sitting with smart phone in his/her hand or laptop or perhaps a person posting a video on Youtube from his/her home half way across the world and still reaching millions of people?

Yes, in a broad sense this can be called human computer interaction but this is the very tip of the iceberg that we shall examine. The human computer interaction these days is not limited to a small box fitted with apps to make life easier and entertaining but we have reached a point where it has integrated into our life and has become an essential factor for survival in the modern society among food, clothing and a roof over our heads.

Now the situation has become food, clothing, roof over our heads and a Wi-Fi under the roof. The limit is far from our reach and it is still fascinating that we have innovation that helps not only to lessen the physical and mental load of the human beings but also the psychological aspects of it.

For e.g. People living in an apartment in a city know the trouble of wanting to keep a pet. Pets are always there for you, they are huge stress relievers and many a times help to take care of the building stress of owner. But living in closed spaces most options like a cat or dog go out of the question. So what could be the solution?

'Vector' by Anki [1] which boasts exceptional abilities of reaction to stimulus using a Qualcomm Snapdragon quad core chip.

S. Hasan (✉)
Institute of Technology and Management, Aligarh, UP 202001, India
e-mail: sanihasan513@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_4

- It reacts to surrounding noises.
- It has an exploration mode built in which allows it to interact and explore with its surroundings.
- It not only learns things but also forms an ‘opinion’ about them.
- You can pet it and watch it get happy while making little noises.
- It is able to return to its charger once it detects low battery charge.

Vector is a great example of human stress relief robotics. A little pet robot for those who just can’t afford to give a cat or a dog the time and care it needs. Vector is self sufficient in these aspects.

Consider the recent hype about the delivery drones. The Google delivery drones [2] gained government approval in Australia. This brings up another great aspect of the use of these drones. Their use in the delivery of rations and supplies in disaster stricken zones where accessibility is low. Their use could save many lives.

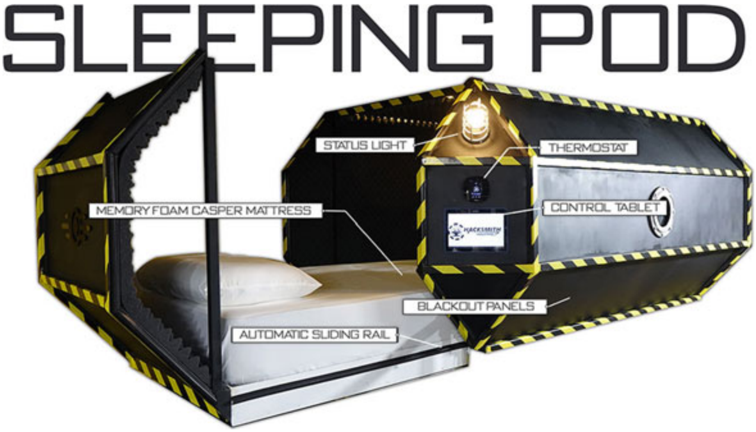
Not only these but there are many examples such as burglar alarms used in home security systems that alert the authorities and the owner in case of a mishappening or a break-in.

Along with the terms of security and times of hardships. Computers are also embedded in the smaller parts of our lives. Such as the Inamo restaurant [3] with its customer interactive tables, in the Inamo restaurant you can order your food with the help of an interactive table and watch as the chefs prepare your meal.

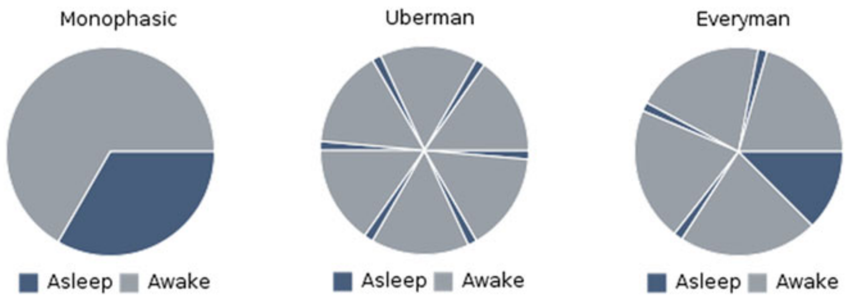
Virtual Reality for one's entertainment and socializing is another such example, where people are now able to for the first time live in "A Virtual Moment". People can interact with a virtual environment using their avatars. A virtual reality that one can feel and smell doesn't seem like a farfetched idea now.

Another way to approach for an even more technologically embedded human life style is to intelligently optimize a certain principle with a machine that compliments it.

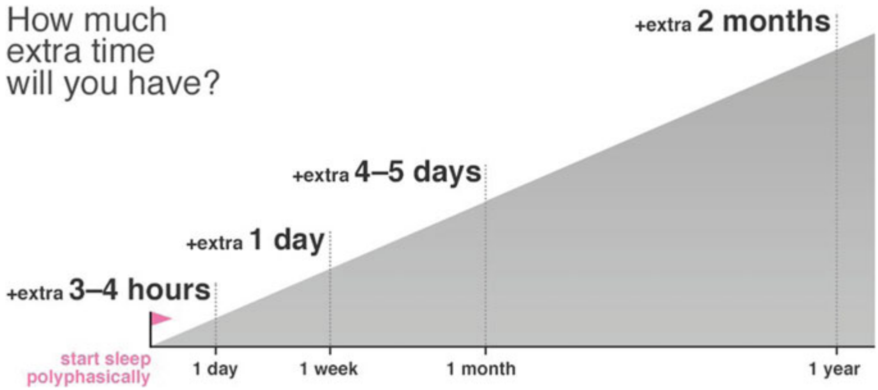
Let us take a situation of a deadline crunch time. Which is common place in businesses and their projects. The things often seen are a sleepless employees and fuming bosses over incomplete tasks. A similar situation occurred in front of James Hobson, a former engineer/product developer in Canada who quit his job to pursue his passion of making fictional sci-fi technologies a reality and founded the Hacksmith Industries. He faced the problem "that my need for sleep was getting in the way of my ability to work on my projects". The solution he found was a sleeping pod which he named Centralized Automatic Sleeping Pod Enhanced Revitalization or CASPER [4] where he used the 'Everman/Polyphasic sleep cycle' which he implemented into his schedule.



The beauty of this is that instead of sleeping for the average of 7–8 h a day a person instead has to sleep only for 4 h a day with three 20 min naps with equal intervals and a 3 h nap in the night.



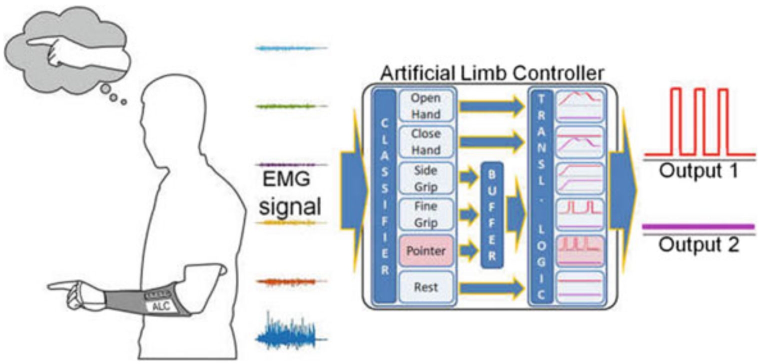
Although this is not recommended for long period of time but this will easily change the idea of a lethargic and sleep deprived employee with an unfinished quota of work to a healthy employee with properly finished work. The diagram below shows the hours the polyphasic cycle add to an otherwise jam packed time schedule up to 1 extra day in a week and 4–5 extra days in a month.



A different aspect is the medical application of technology which is quite commonplace in today's society. From a simple pacemaker in the heart to a complex prosthetic. How much complexity have we achieved in the field is the deterrent here. Just as you can say that a pirate with a wooden leg and a hook for a hand can be called to posses 'prosthetics', but that will not hold true in today's society.

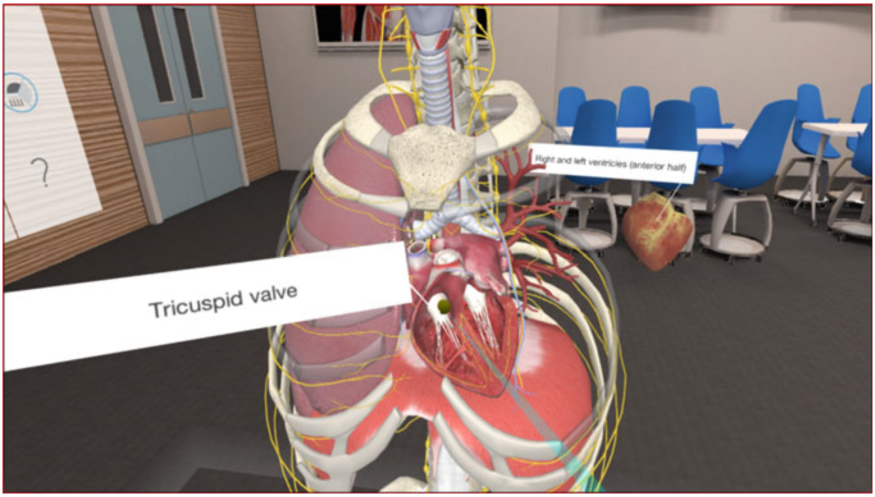
We have reached a point where most appendage prosthetics are able to imitate their biological counterparts in their complexity and their ability to deliver. The world's fastest runner with prosthetics 'Oscar Pistorius', also referred to as 'The fastest man with no legs', 'Blade Runner' with both legs amputated and supported by prosthetics made with the design with the help of Human Computer Interaction.

What is left now is precision and complexity but that boat has already sailed with prosthetics equipped with the Myoelectric Pattern Recognition (MPR) which are able to pick up neural receptions at the stump levels and allows Real time inference of human motor volitions and has a great potential to be used in intuitive control of robotic devices.



Another example are HAL [5] suits developed by Japan's Tsukuba University and robotics company Cyberdyne which have proved tremendously helpful in rehabilitation of paralysis patients (Full body paralysis, half body paralysis). The basic design is an exoskeleton that is strapped to the body of the patients, the suit to be strapped is decided by the condition of the patients. The effects have been way better than just simple rehabilitation.

Not only in the field of prosthetics but also in the training of medical students, cut-sections of the body, simulating every nerve, vein, artery and organs perfectly in a virtual reality. Allowing the students to experience what is to be expected without the loss of cadavers which can be used more due to fewer losses in the institution stock.



Many more such aspects are steadily emerging in the modern society.

References

1. Vector by Anki, https://anki.com/en-us/vector.html
2. Google Drones
3. Inamo Restaurant, https://www.inamo-restaurant.com
4. Sleeping Pod C.A.S.P.E.R by Hacksmith Industries, https://www.thehacksmith.ca/
5. HAL suits by Tsukuba University. www.tsukuba.ac.jp and Cyberdyne https://www.cyberdyne.jp

New Fast and Efficient Techniques for Image Denoising Using VQ, PCA, SVD, Soft Thresholding and Wiener Filter

Rachana Avinash Dhannawat and Archana Bhupendra Patankar

Abstract Image denoising is process of removal or reduction of noise from degraded images. During formation of images itself, images get degraded, so it is unavoidable. In this paper two fast, easy and efficient techniques are developed for image denoising. These techniques use VQ, PCA/SVD, Soft thresholding, Wiener filtering for denoising on non-overlapping blocks. These techniques are compared with existing VQ techniques and results are found better in terms of PSNR, SSIM objective criteria. The proposed techniques improved PSNR by 12% and 9% for color and gray images respectively as compared to noisy images while existing VQ technique has improvement of 9% and 0.47% respectively. SSIM is improved by 67% and 45% respectively for color and gray images by proposed methods while existing VQ technique has improvement of 27% and 3% respectively. These techniques are also compared with state-of-art LPG-PCA technique, and it is found that the proposed techniques have given 5% better PSNR than LPG-PCA, but SSIM is not improved. The key advantage of these techniques is time required to execute is 97% less than LPG-PCA. So, these techniques can be considered as promising improvements in terms of computational complexity.

Keywords VQ · LBG · PCA · SVD · Soft thresholding · Wiener filter · Image denoising · LBG-PCA-ST-W · LBG-SVD-ST-W · LPG-PCA

1 Introduction

Image denoising is process of removal or reduction of noise from degraded images [1, 2]. During formation of images itself, images get degraded, so it is unavoidable. If atmospheric conditions are not favorable then noise is increased much more. It is an ill posed problem and many solutions are possible to it. Several techniques are

R. A. Dhannawat (✉)
Vidyalankar Institute of Technology, Mumbai, India
e-mail: rachanadhannawat82@gmail.com

A. B. Patankar
Thadomal Shahani Engineering College, Mumbai, India
e-mail: patankararchu@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_5

developed for image denoising [3] such as LPG-PCA [4], BM3D [1], Sparsity based such as KSVD [5], KLLD [6], NCSR [2], etc. All of these techniques are complex and many iterations are needed to get good quality images. Here in this paper we propose new techniques which are very fast and direct. These are easy and require very less time to execute.

The basic difference between above mentioned existing techniques and our method is that these techniques use overlapping blocks for calculations, so many times a single pixel is processed. While our techniques use VQ [7–10] as first step in which image is divided into non-overlapping blocks. Later we modify the results by using PCA/SVD [11–14], soft thresholding [15, 16], Wiener filtering [17]. In literature LBG [7–10] is used for image restoration but there was scope to improve the quality of denoised image. We have attempted to improve the results. We observed that results in terms of PSNR and time are better for our methods when compared with existing VQ techniques [7–10] as well as LPG-PCA [5]. As compared to LPG-PCA the time required is 97% less for our proposed method and PSNR of denoised image is also improved.

Next section of paper discusses in detail proposed techniques. Following section presents results and analysis and finally the findings are concluded.

2 Proposed Techniques

The proposed method consists of 4 steps. First step is vector quantization which maps low frequency components to high frequency components using codebook. This mapping helps to recover high frequency components degraded by low frequency degradation. Further denoising is done by applying PCA or SVD dimensions reduction techniques which will select significant components only, omitting noise. After this soft thresholding is applied for better results and finally Wiener filter is applied as post processing to obtained results. Figure 1 shows complete flow of proposed techniques. In detail importance and working of all steps is given in following section.

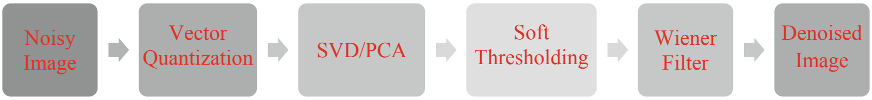


Fig. 1 Basic flowchart of proposed techniques

2.1 Vector Quantization

Noise is high frequency component and it affects high frequency components in a noisy image, as it mixes with high frequency components. The idea to denoise the noisy image is to separate out the low frequency and high frequency components and find mapping between low frequency and high frequency components. The denoised image is obtained as addition of low pass filtered image and the corresponding high frequency information. LBG [7–10] algorithm is used here for vector quantization [7–10].

Image restoration is least square problem and two basic techniques to solve least square problems are PCA and SVD.

2.2 PCA

Statistically, PCA is a de-correlation technique. PCA groups components with maximum variance in one direction. These components correspond to image data and generally noise has small variance. By transforming the original dataset into PCA [11, 14] domain and preserving only the several most significant principal components, the noise and trivial information can be removed.

2.3 PCA for Denoising

It is assumed that noise N in the corrupted image A is white additive Gaussian with zero mean and image A and noise N are uncorrelated. Noisy image A_N is given as $A_N = A + N$. Procedure for image denoising using PCA [11, 14] is as follows.

1. Mean value of A_{Ni} is calculated as $m_{ni} = 1/n \sum_{j=1}^{n} A_{Ni}(j)$.
2. Noisy image A_N and noiseless image A both are centralized as A_{CNi}, $A_{CNi} = A_{Ni} - m_{ni}$, as noise is of zero mean.
3. For noisy image covariance matrix is calculated as

$$CVAN = 1/n\ A_{CN}A_{CN}^T = 1/n\left(A_cA_c^T + A_cN^T + N\ A_c^T + NN^T\right).$$

4. As image and noise are uncorrelated, $A_c\ N^T$, NA_c^T are approximately zero. So CVAN is written as $1/n\left(A_cA_c^T + NN^T\right) = CVA_C + CVN$.
5. CVA_C can be decomposed as $CVA_C = O_cSO_c^T$.CVA_C and CVN have same eigen vector matrix O_c^T.
6. Applying orthogonal PCA transformation matrix O_c^T to noisy image A_N gives

$$Y_N = O_c^TA + O_c^TN.$$

7. O_c^T A is decorrelated dataset of A. Most energy of noiseless O_c^T A concentrate in several most important components, while energy of noisy O_c^TN distributes evenly all over the image.
8. Noise in Y_N can be suppressed by using any shrinkage algorithms such as linear minimum mean square error estimation technique [4]. In this technique covariance of Y_N is used as weight coefficient w. In flat areas this weight coefficient w is close to zero and most of the noise is suppressed when Y_N and weight coefficient w are multiplied. i.e. $Yd = Y_N * w$.
9. By transforming Yd back to time domain, the denoised result for A_{CN} can be obtained as

$$Ad = \left(O_c^T\right)^T * Yd.$$

10. Mean values m_i are added back to Ad to get final denoised image.

2.4 SVD

SVD [12–14] is matrix decomposition technique. Using singular value decomposition, any real matrix A can be decomposed into a product of three matrices U, S and V as $A = USV^T$, where U and V are orthogonal matrices and S is diagonal matrix. If A is $m \times n$ matrix, U is $m \times m$ orthonormal matrix; V is $n \times n$ orthonormal matrix. S is of size $m \times n$. SVD has property to divide image into two distinct subspaces data and noise subspace. This property makes SVD useful for noise reduction. Largest singular values correspond to image data and noise corresponds to eigen images associated with smallest singular values. SVD can be used even if matrix is not square [12–14]. The basic steps for denoising using SVD and PCA are same as mentioned above. PCA can be calculated using SVD and vice versa. In denoising both methods can be used and have little difference in working.

2.5 Soft Thresholding [15, 16]

Larger coefficients correspond to image signals while lower coefficients correspond to noise components. Applying thresholding to lower components can reduce noise significantly. Therefore, soft thresholding is applied as next step for denoising. Soft threshold [15, 16] is calculated as

$$\begin{aligned} D(U, \lambda) &= \text{sign}\,(U)\max(0, |U| - \lambda) \text{ for all } |U| > \lambda \\ &= x \text{ otherwise.} \end{aligned}$$

2.6 Wiener Filter [17]

Wiener Filter [17] gives minimum mean square error between original image and noisy image and it works well with noise corrupted image as compared to inverse filter. It is used in the proposed technique as post processing operation.

3 Results and Analysis

Table 1 compares basic LH-LBG method with our developed methods LBG_PCA_ST_W and LBG_SVD_ST_W for 7 color images. These techniques are also compared with averaging and median filter for denoising. The noise considered is Gaussian noise with standard deviation 20. The stages wise output is also displayed in Table 1. We observe that these two proposed techniques have best results as compared to all other methods. It is seen from Table 1 that LBG_SVD_ST_W technique gave best PSNR and SSIM [18] for parrot and butterfly images and for all other images LBG_PCA_ST_W technique gave best PSNR and SSIM [18]. Average PSNR and SSIM are better for LBG_PCA_ST_W technique. The percent increase in PSNR as compared to noisy image by LBG_PCA_ST_W technique is 12% and by LBG_SVD_ST_W technique it is 11%. By the LH-LBG technique it is 4% and averaging and median filter the increase is 9%. The percent increase in SSIM as compared to noisy image by LBG_PCA_ST_W technique is 67% and by LBG_SVD_ST_W technique it is 65%. By the LH-LBG technique it is 27% and averaging and median filter the increase is 51% and 50% respectively. Figures 2 and 3 display the comparison of 5 techniques along with noisy image using graphs for PSNR and SSIM respectively. The resultant reconstructed butterfly images as an example, by various techniques are displayed in Fig. 4.

We compared LBG_PCA_ST_W and LBG_SVD_ST_W these two techniques for 9 more color images and found that average results of both techniques for average of 16 color images are almost same. The average PSNR and SSIM of LBG_PCA_ST_W are 33.640 and 0.784. The average PSNR and SSIM of LBG_SVD_ST_W are 33.639 and 0.784. In average it is observed that LBG_PCA_ST_W gave better results for a greater number of images (11) while LBG_SVD_ST_W gave better results for a smaller number of images (5).

We implemented same methods for 9 gray images and Fig. 5 displays denoised cameraman images as an example. Tables 2 and 3 compares PSNR and SSIM of above discussed 5 techniques for 9 gray images respectively. From Table 2 it is clear that averaging filter has given best PSNR for house and butterfly images and LBG_PCA_ST_W technique has given best PSNR for baboon and boat images. For all other images LBG_SVD_ST_W gives best PSNR and hence average PSNR is also best by LBG_SVD_ST_W technique. For gray images it is observed that LH-LBG has given very less 0.47% improvement in PSNR while the proposed techniques have improved PSNR 9% as compared to noisy image. Figure 6 displays a graph

Table 1 PSNR comparison of 6 techniques for 7 colour images

Images	Parrot		Butterfly		House		Barbara		Starfish		Leaves		Tower		Average	
Technique\metric	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Noisy	29.38	0.38	29.31	0.59	29.29	0.36	29.31	0.47	29.29	0.54	29.31	0.72	29.30	0.31	29.31	0.48
LH_LBG (256)	30.79	0.53	30.45	0.70	30.71	0.50	30.53	0.61	30.48	0.67	30.54	0.81	30.71	0.46	30.60	0.61
LBG_PCA	31.33	0.59	30.71	0.71	31.36	0.54	30.82	0.64	31.14	0.70	30.92	0.81	31.20	0.50	31.07	0.64
LBG_PCA_ST	31.33	0.59	30.80	0.71	31.32	0.54	30.84	0.64	31.04	0.70	30.94	0.81	31.16	0.50	31.06	0.64
LBG_PCA_ST_W	33.24	0.79	31.88	0.82	33.55	0.73	32.29	0.77	32.28	0.80	31.40	0.89	34.21	0.76	32.69	0.80
LBG_SVD_ST	31.39	0.59	30.76	0.71	31.31	0.55	30.86	0.65	31.02	0.70	30.94	0.81	31.23	0.51	31.07	0.65
LBG_SVD_ST_W	33.28	0.79	32.01	0.82	33.46	0.73	32.24	0.77	32.19	0.80	31.27	0.89	34.17	0.75	32.66	0.79
Averaging filter	32.667	0.726	31.513	0.789	32.443	0.665	32.076	0.744	31.840	0.772	30.610	0.8442	32.662	0.668	31.858	0.727
Median filter	32.182	0.665	31.354	0.771	31.986	0.608	31.622	0.694	31.518	0.733	31.060	0.840	32.157	0.6021	31.860	0.719

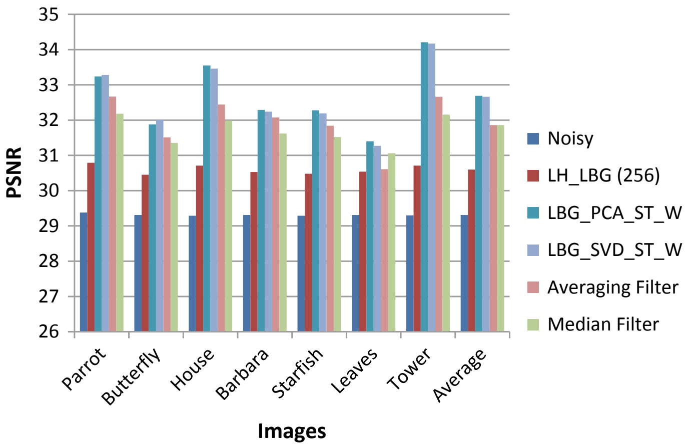


Fig. 2 Graph comparing PSNR of different techniques for 7 color images

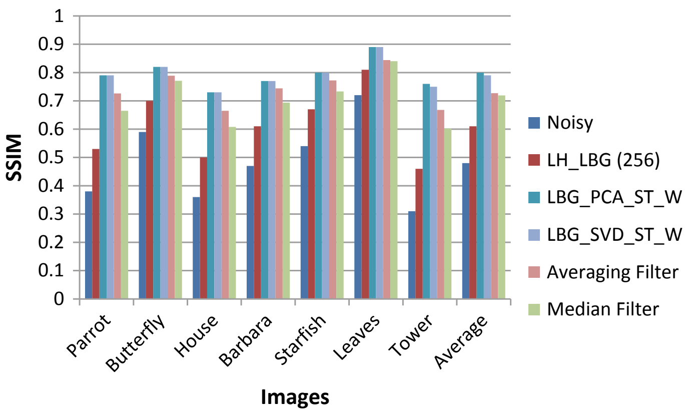


Fig. 3 Graph comparing SSIM of different techniques for 7 color images

comparing PSNR of 5 techniques for 9 gray images.

Table 3 indicates that LBG_SVD_ST_W has given best SSIM for all 9 images and also the best average SSIM. LBG_PCA_ST_W technique also has given the same best SSIM for 5 images. LH-LBG method has improved SSIM by only 3% while the proposed method has improved SSIM by 45% as compared to noisy image. Figure 7 displays a graph comparing SSIM of 5 techniques for 9 gray images.

Table 4 shows average PSNR, SSIM [19] and time for LPG-PCA and proposed LBG-PCA-ST-W technique for 4 gray and 2 color images only for comparison purpose. Gray images considered are house, Lena, butterfly, cameraman. Color images considered are parrot and Barbara. From Table 4 it is clear that PSNR of our proposed technique is better than LPG-PCA [5] by 5% and time required is very much 97%

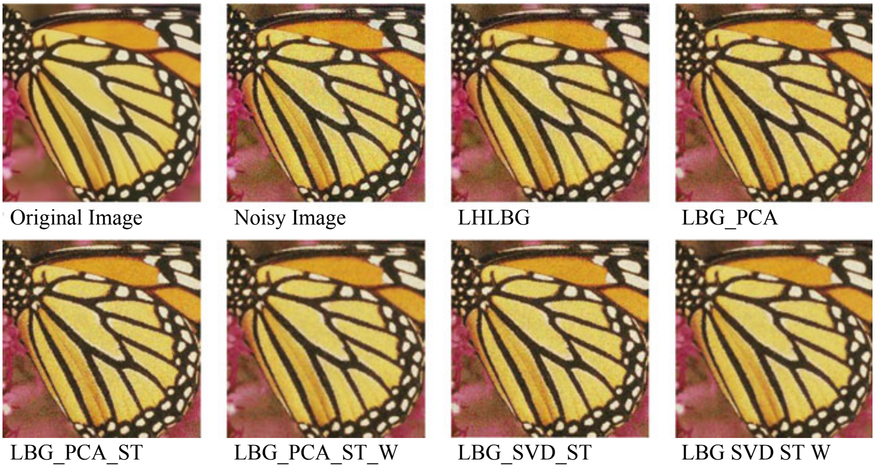


Fig. 4 Reconstructed butterfly images

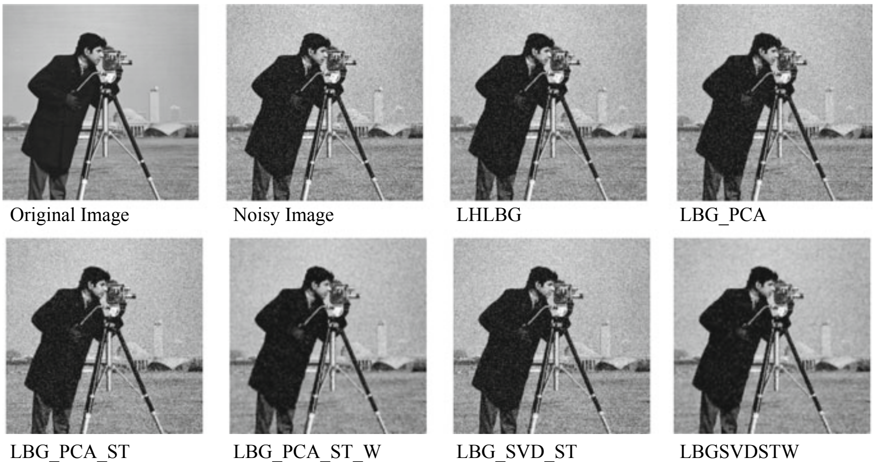


Fig. 5 Reconstructed cameraman images

less than LPG-PCA. SSIM is better by 15% for LPG-PCA [5] technique. Almost same results we got for LBG-SVD-ST-W technique.

Table 2 PSNR of all techniques for 9 grey images

Technique\images	Noisy	LH_LBG	LBG_PCA	LBG_PCA_ST	LBG_PCA_ST_W	LBG_SVD_ST	LBG_SVD_ST_W	Averaging filter	Median filter
House	29.30	29.46	29.50	29.58	33.03	29.47	33.06	33.066	32.276
Butterfly	29.27	29.43	29.44	29.54	32.27	29.41	32.34	32.351	31.675
Lena	29.31	29.47	29.52	29.59	32.56	29.52	32.67	32.736	32.043
Barbara	29.30	29.47	29.49	29.48	31.58	29.54	31.58	31.565	31.028
Baboon	29.29	29.21	29.27	29.21	29.73	29.23	29.72	29.624	29.501
Cameraman	29.51	29.66	29.77	29.67	32.26	29.70	32.34	31.861	31.603
Boat	29.27	29.47	29.59	29.60	32.31	29.55	32.13	32.301	31.654
Peppers	29.36	29.54	29.58	29.55	32.31	29.44	32.35	32.396	31.785
Pentagoan	29.31	29.43	29.41	29.29	31.08	29.36	31.14	31.190	30.943
Average	29.324	29.46	29.507	29.501	31.903	29.468	31.925	31.899	31.390

Table 3 SSIM of all techniques for grey images

Technique\images	Noisy	LH_LBG	LBG_PCA	LBG_PCA_ST	LBG_PCA_ST_W	LBG_SVD_ST	LBG_SVD_ST_W	Averaging filter	Median filter
House	0.35	0.37	0.39	0.38	0.68	0.39	0.69	0.681	0.618
Butterfly	0.52	0.54	0.54	0.54	0.80	0.54	0.80	0.796	0.759
Lena	0.43	0.46	0.47	0.48	0.75	0.47	0.76	0.752	0.702
Barbara	0.52	0.54	0.55	0.55	0.73	0.55	0.73	0.713	0.676
Baboon	0.68	0.65	0.64	0.64	0.57	0.64	0.57	0.475	0.454
Cameraman	0.41	0.43	0.44	0.43	0.69	0.43	0.70	0.657	0.609
Boat	0.48	0.50	0.52	0.51	0.75	0.51	0.75	0.747	0.697
Peppers	0.46	0.48	0.48	0.48	0.74	0.47	0.74	0.737	0.693
Pentagoan	0.62	0.63	0.62	0.62	0.74	0.62	0.74	0.739	0.718
Average	0.496	0.511	0.516	0.514	0.716	0.513	0.72	0.700	0.658

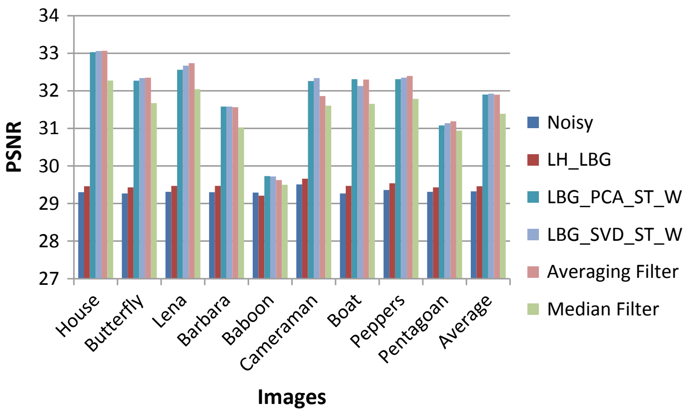


Fig. 6 Graph comparing PSNR of different techniques for 9 gray images

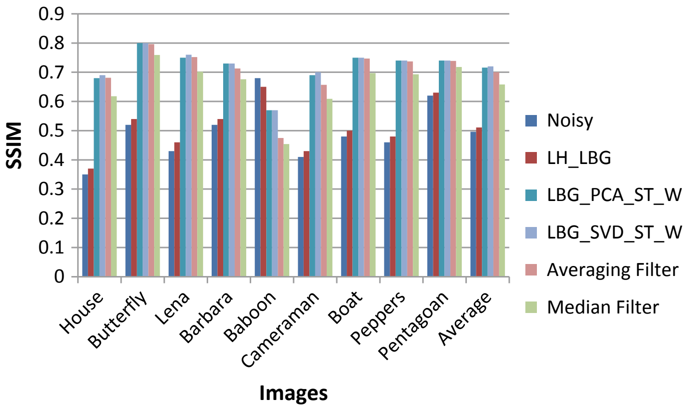


Fig. 7 Graph comparing SSIM of different techniques for 9 gray images

Table 4 Average PSNR, SSIM and TIME of LPG-PCA and VQ-PCA-ST-W techniques

Technique\metric	LPG-PCA	VQ-PCA-ST-W
PSNR	31.20487	32.60883
SSIM	0.87863	0.745867
TIME	361.483	10.47728

4 Complexity Analysis of Proposed LBG_PCA_ST_W and LBG_SVD_ST_W with LPG_PCA

Complexity of LBG algorithm [19] is given as: nmtnLBG + n (m − 1) t, where nLBG consist of (2k − 1) addition and k multiplications.

Where t: number of vectors in training sequence, n: number of iterations, m: codebook size, k: code vector size.

PCA/SVD calculations are done m times. Soft thresholding and Wiener filtering are just post processing steps done only once. Soft thresholding requires one multiplication and one addition with an image. Wiener filter requires 2 divisions, 1 addition and 1 multiplication.

Soft thresholding and Wiener calculations are very less as compared to LBG and PCA/SVD, so we will ignore them. LBG-PCA [4] consist of two main steps codebook generation and PCA calculations. Both steps are done separately. First codebook of desired size generated and then PCA/SVD calculated for generated codebooks. So, no overlap in procedure and procedure is very fast. The maximum complexity of proposed method is by PCA/SVD which comes as O (k^3).

If we compare LPG-PCA then LPG requires ($2k^2 - 1$). $(L - k + 1)^2$ additions, k^2. $(L - k + 1)^2$ multiplications [5], where L and K are sizes of blocks chosen for grouping. In LPG K times single pixel processed in horizontal direction and k times in vertical direction. So, it is very much complex and requires many calculations. PCA calculations are called n*m times, where n, m is size of an image. The complexity comes as O (K^5). There are two passes for algorithm so all above calculations are done twice. Hence, overall complexity is high.

5 Conclusions

This paper proposed two techniques for image denoising using VQ, PCA, SVD, Soft thresholding and Wiener filtering. The techniques use non-overlapping blocks so calculations and complexity is reduced much in these techniques as compared to state of art methods. It is observed that proposed LBG-PCA-ST-W and LBG-SVD-ST-W techniques gave best PSNR and SSIM for all images when compared with existing VQ technique. Results are compared for 16 color images and 9 gray images. The proposed techniques improved PSNR by 12% and 9% for color and gray images respectively as compared to noisy images while existing VQ technique has improvement of 9% and 0.47% respectively. SSIM is improved by 67% and 45% respectively for color and gray images by proposed methods while existing VQ technique has improvement of 27% and 3% respectively. When compared with state-of-art algorithm LPG-PCA, these techniques have improved PSNR by 5% but SSIM is not improved. The key advantage of these techniques over LPG-PCA is that PSNR is improved and time is reduced near about by 97%. So, these are very fast and efficient techniques for image denoising. These can be considered as promising improvements in terms of computational complexity. Further improvements can be done for better SSIM.

References

1. Dabov K, Foi A, Katkovnik V, Egiazarian K (2007) Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Trans Image Process 16(8):2080–2095
2. Dong W, Zhang L, Shi G, Li X (2013) Nonlocally centralized sparse representation for image restoration. IEEE Trans Image Process 22(4):1620–1630
3. Dhannawat Rachana, Patankar Archana (2018) A short review on image restoration of noisy and blurred images and comparison of state of art algorithms. Int J Adv Res Comput Sci 9(5):181–185
4. Zhang L, Dong W, Zhang D, Shi G (2010) Two-stage image denoising by principal component analysis with local pixel grouping. Elsevier J Pattern Recogn 43:1531–1549
5. Aharon M, Elad M, Bruckstein A (2006) K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans Signal Process 54(11):4311–4322
6. Chatterjee P, Milanfar P (2009) Clustering-based denoising with locally learned dictionaries. IEEE Trans Image Process 18(7):1438–1451
7. Nakagaki R, Katsaggelos AK (2003) A VQ-based blind image restoration algorithm. IEEE Trans Image Process 12(9):1044–1053
8. Nakagaki R, Katsaggelos AK (2003) A VQ-based blur identification algorithm. In: IEEE international conference on acoustics, speech and signal processing, pp 725–728
9. Nakagaki R, Katsaggelos AK (2002) A VQ-based image restoration algorithm. In: IEEE international conference on image processing, pp 305–308
10. Lu T-C, Chang C-Y (2010) A survey of VQ codebook generation. J Inf Hiding Multimedia Signal Process 1(3):190–203
11. Watkins DS (2010) Fundamentals of matrix computation. New Jersey, USA, pp 259–526
12. Sadek RA (2012) SVD based image processing applications: state of the art, contributions and research challenges. Int J Adv Comput Sci Appl 2(1):265–270
13. Guo Q, Zhang C, Zhang Y, Liu H (2015) An efficient SVD-based method for image denoising. IEEE Trans Circ Syst Video Technol (99):1–13
14. Dhannawat R, Patankar AB (2016) Improvement to blind image denoising by using local pixel grouping with SVD. Elsevier Procedia Comput Sci 79:314–320
15. Vohra R, Tayal A (2011) Image restoration using thresholding techniques on wavelet coefficients. Int J Comput Sci (IJCSI) 8(5):400–404
16. Naz A, Talukdar A, Sarma K (2012) Digital image restoration using discrete wavelet transform based approach. IRNet Trans Electr Electron Eng (ITEEE) 1(2):53–57
17. Yoo J-C, Ahn CW (2014) Image restoration by blind-Wiener filter. IET Image Process 8(12):815–823
18. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–611
19. Shanbehzadeh J, Ogunbona P (1997) On the computational complexity of the LBG and PNN algorithms. IEEE Trans Image Process 6(4):614–616

Elicitation and Analysis of Security Requirements and Patterns for IoT Based Health Monitor

E. R. Aruna, A. Rama Mohana Reddy and K. V. N. Sunitha

Abstract Security must be incorporated in the software development process at each phase such as requirements, design, implementation, testing and deployment of the applications to deliver the secure software to end user community. We are using security patterns at each phase to achieve secure Software Development life cycle (Secure SDLC). In this paper, we have identified security requirements, suggested and analyzed the security patterns for security goals during requirement phase for IoT based Health Monitor. Nowadays usage of life care device applications are increasing due to feasible connectivity provided by Internet of things (IoT) which requires more security.

Keywords IoT · Secure SDLC · Security patterns · Health monitor

1 Introduction

In recent years, security is important to consider for developing any software because hackers, developers, end users are disclosing the large number of vulnerabilities due to malicious code. Hence Software security is becoming an important concern in software development process. As there can be large number of vulnerabilities identified in software due to poor analysis, poor design and poor development, hence a better process needs to develop software to enhance security concerns at each phase of SDLC [1].

Previously in software development process security was a concern while testing phase, but now, it has been proven that security must be integrated in all phases of SDLC [2]. Eric Bodden says "The U.S. Department of Homeland study says 90%

E. R. Aruna (✉)
Department of IT, Vardhaman College of Engineering, Hyderabad, India
e-mail: ayam.ac2003@gmail.com

A. R. M. Reddy
Department of CSE, SVU College of Engineering, SV University, Tirupati, India

K. V. N. Sunitha
Department of CSE, BVRIT College of Engineering, Hyderabad, India

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_6

of cyber-attacks are neither by faulty crypto, nor by networks or hardware but by application-level implementation vulnerabilities" [3].

People involved in Requirements Gathering, Designing, Developing, Testing need not be security experts, because they can select and use the security pattern as per the context from the catalog. Security issues of unsecure applications are addressed by the security Patterns. Security patterns encapsulated abstract solutions to recurrent security problems to achieve security goals in software development.

Security vulnerabilities may occur in various stages of the life cycle and protecting the assets from these vulnerabilities are more increasingly important, in this context security mechanisms or Guidelines using security patterns makes much easier for the requirement engineers, designers, developers in the development process.

The Internet of Things (IoT), involves data communication among billions of devices and users over the network. Secure communications among devices and users seldom considered which cause threat to whole system. For security critical applications this can be dangerous which causes sensitive data exposure, Data Manipulation, Denial of service etc. So that IoT based security critical applications such as Health Monitoring Devices, Wearable's, Smart Homes, Connected Cars etc. cannot be ignored.

2 Related Work

Ali et al. in [4] describes the right stuff for secure applications, security goals for unsecure IoT based applications and eight security patterns documentation. Lee et al. [5] applied five security patterns for the development of IoT software based System. Avelet et al. [6] proposed, security is incorporated during design and implementation of IoT Health Monitor. Kolias et al. [7] identified 3 major security problems in IoT applications. Meher et al. [8] summarize the problem space of security patterns representation (extended use case diagram) while designing security requirements. This provides functional requirements along with security threats. This extension graphical way gives security expert Knowledge for mitigating threats.

Moosavi et al. [9] proposed a secure end-user authentication and authorization architecture based DTLS hand shake, session resumption, smart gateways which acts as communication protocols between device and Fog layers, Fog layer and Cloud layer.

This work is organized into 4 major sections, in Sect. 3 we discussed about design and functioning of IOT Health Monitor, in Sect. 4 we have identified the security requirements of IoT Health Monitor during requirement Phase, in Sect. 5 suggested the security requirement patterns for the security goals, Sect. 6, analyzed the security requirement patterns for the security goals and discussed these patterns are support for the designers without much security exercise in next level.

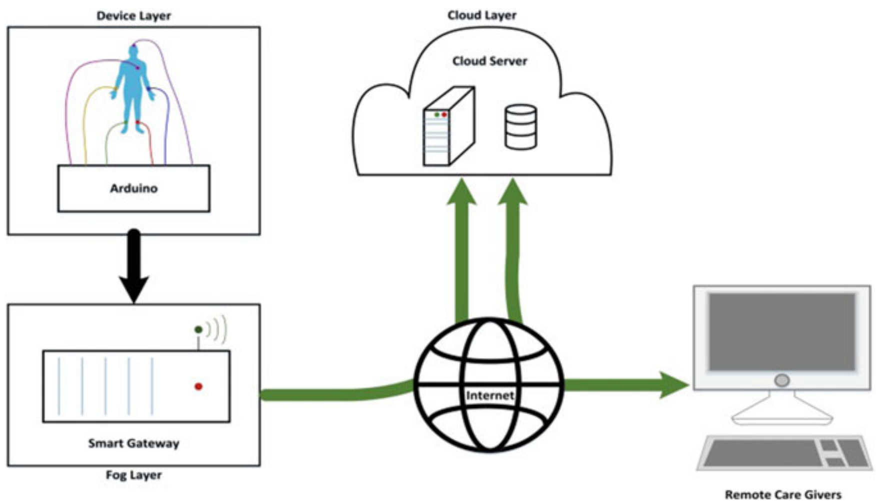


Fig. 1 Health monitor architecture using IoT

3 Design and Functioning of Health Monitor Using IOT

IoT health monitor machine is having three layers which enables remote tracking of patient information. Figure 1 shows the architecture of IoT Health Monitor devised as per Moosavi et al. [9]. This architecture has Device layer, Fog Layer and Cloud Layer. The device layer, biomedical signals are captured through sensors from human body in the home/Hospital. This biomedical information is transferred to the fog layer through WI-FI/Internet. The fog layer additionally introduced for efficient mobility and extends the cloud paradigm to the network edge. Fog layer consists of smart gateways and behaves like secure communication protocols between device layer and cloud layer. The cloud layer contains health data server and maintains the patient's information.

Our aim is to make sensitive data communication through Wi-Fi/Internet is 100% secure from device layer to fog layer, from fog layer to cloud layer, from cloud layer to the person may be a Doctor or Guardian to monitor the patient condition from remote location using security Patterns.

4 Security Requirements Elicitation

As a best practice, security requirements must be identified during the initial planning stages [10]. Due to the lack of serious check in functional and nonfunctional requirements at the requirements phase, leads to more than 50% of the defects, else if we try to catch, this will reduce the cost for fixing the issues and also reduce the defects impact [11]. Developers are concerned more on core functionality, if security concerns are not defined in the early stages, the developer implements security mechanisms based on his own interpretations, in worst case, missing security also in the software development [4]. Security Requirements must focus on high level security objectives and also possible threats and attacks.

Several techniques such as Misuse cases Attack Trees, Issue Based Information Systems, Misuse Activities etc. are available to elicit the security requirements. In our application, the following security requirements are derived based on Misuse case Technique.

Misuse case Techniques are closer with regular use cases, misuse cases facilitate better analysis how threats and security requirements are relating to functional requirements. We have described the misuse cases in textual representation with certain fields such as Name, iteration, basic, extensive, alternate paths, Preconditions, Post conditions, rules and threats.

- Req1: Data must be continuously or intermittently sensed from the patient and analyzed. Authorization for sensed data and authentication is required to access the sensed data.
- Req2: Data Communication between all the system components such as Arduino, Sensors, Health Information Server(cloud), Gateways while exchange of data must ensure authenticity, confidentiality, data integrity.
- Req3: Access to the heath information in cloud server must be focused, to ensure users access their own specific records.
- Req4: Input fields such as data from arduino to database and database to users (Doctor/Guardian) must be validated before processing, storage and granting privileges to access.
- Req5: There must be secure process, which protects Denial of Service attacks since IoT health monitor is real time web application.
  Amir et al. presented security risks, threats and new classification of cyber attacks in three layers (Application, Network and perception layers) of IOT based Heath Infrastructure [12].
- Req6: Session Management mechanism must be implemented against remote unauthorized access or accidental disclosure. Session must be ended either when doctor/Guardian logged out or Inactive for a period of Time.
- Req7: In Addition to the sensitive data storage, application logon credentials must also be secured against spoofing/Tampering.
- Req8: The web interface between IoT device to database server must be protected because security vulnerabilities such as SQL Injection (SQLi) causes privileges escalations and pirating the users session respectively.

Throughout the next stages security requirements must be evaluated. However, it is recommended to define security objectives in next stages based on evolving levels of security concerns.

5 Identification of Security Requirement Patterns for Security in Requirements Phase

Several techniques are available to identify the security patterns for a specific goal. Few of them are Ordered Matrix Matching [13], Text Classification Technique [14], Goal Oriented Requirements Language Approach and Prolong Rules, Pattern Search Engine [15], ontological approach [16].

We have used here the general approach to select the pattern based on the security properties, application domain and Constraints.

We have selected the security patterns from the following sources, Sahumacher et al. [17], Withall [18], Hatebur [19], Beckers et al. [20], Motii et al. [21], Yoder et al. [22], Yoshizawa et al. [23], Fernandez [24] and "cloudpatterns.org".

The pattern selection to analyze Security requirements of IoT based Health Monitor are:

- To access the patient data by the doctor/Guardian must be authenticated (Identified User), for secure access mechanisms we have chosen 'User Registration Requirement Pattern' and 'User Authentication requirement Pattern' (Req1).
- To access the patient data, verification of the access rights to care takers must be defined using 'Access control requirements Pattern' and 'Authorization pattern' (Req1).
- Data communication between device layer to fog layer, for secure data transmission use 'Confidential Data Transmission using Symmetric Encryption Pattern' (Req2).
- Data communication between fog layer to cloud layer must be secured using 'Virtual Private Network (VPN) abstract pattern' (Req2).
- Ensure care takers access their specific records from cloud patient Information server database whenever they need using 'Check Point pattern' combined with 'Singe access point pattern' protects from DoS attacks and also Validation of all the Input data before storage, process and granting privileges to access(Req3, Req4, Req5).
- To maintain secure session management use 'Session-Based Role-Based Access Control and Session-Based Attribute-Based Access Control patterns' (Req6).
- Secure Storage of Patient Data and Logon credentials data use 'Cloud Storage Data at Rest Encryption pattern' (Req7).
- To mitigate SQL Injection vulnerability use 'Intercepting validator' Pattern (Req8).

6 Security Requirements Patterns Analysis and Its Impact on Security Objectives of IoT Health Monitor

- 'User Registration Requirement Pattern', in our application doctor/patient Guardian must make their identity known to the system by providing their details to register. Server acts as a repository of available client details those who registered.

- 'User Authentication Requirement Pattern' performs validating the user before grant or deny to access the patient sensitive information. To grant or deny access to individual requests login-and-authentication is required.
- 'Access control requirements Pattern' and 'Authorization pattern' provides right to allow or deny doctor/Guardian to perform read, modify or delete actions on patient information in the cloud database. These patterns provide generic access control requirements are unique to each situation while access and also help us to determine the importance of access right.
- 'Confidential Data Transmission using Symmetric Encryption Pattern' considers the symmetric encryption, the content of the secret data is preserved using preconditions (Secret Keys). The respective preconditions are only known by the sender and receiver machines. If the preconditions match both sides then only secure data transmission takes place, trusted paths also required between sender and receiver. Post conditions also required similar to preconditions for confidential data communication. Preconditions and Post conditions matching is necessary between device layer to fog layer.
- 'Virtual Private Network (VPN) pattern' provides security mechanisms to provide confidentiality and integrity of data. This pattern works in virtual sense uses cryptographic tunnel with private and public ends.
- 'Check Point patterns' and 'Single access point patterns' define clear entry point, assessed by implementing security policy. These patterns provide access to the health monitor application for caretakers and ensure the application is not damaged by these remote users.
- Check point Pattern takes certain actions such as denying or ignoring messages. If the pattern context does not match with user actions, then denial of service attacks can be prevented.
- 'Session-Based Role-Based Access Control pattern' defines access session by allowing access to patient information based on the role and limit of the right at a given time.
- 'Session-Based Attribute-Based Access Control pattern' defines access session by allowing access to patient information based on the attributes and properties of the Health Monitor.
- 'Cloud Storage Data at Rest Encryption pattern' applying cryptographic key management system policies, procedures to encrypt patient information which is in cloud server database.
- 'Intercepting validator' mitigates sql injection attacks by performing validation at cloud server and controls user requests with respect to its specifications.

The identified patterns provide the true properties to achieve security goals. Security requirement Patterns define security mechanisms to stop few threats, but these requirement patterns will not treat how to mitigate completely at this stage.

The precise text-based specification of security requirements into a set of patterns provide a basis for better communication between the requirements engineering to design. However, evolving security risks during design can be alleviating with security design patterns.

7 Conclusion and Future Work

In this paper, we have identified the security requirements of IoT based Health Monitor application in Requirements Phase. We have suggested security Patterns at early stage of SDLC. These Patterns can claim by the next level people, they are able to work with support of these patterns without much security exercise.

In our future research work, applying security design patterns for IoT applications, validates and analyzes these patterns impact on secure SDLC.

References

1. Tung Y-H, Lo S-C, Shih J-F, Lin H-F (2016) An integrated security testing framework for secure software development life cycle. In: 2016 The 18th Asia-Pacific network operations and management symposium (APNOMS), IEICE
2. Karim NSA, Albuolayan A, Saba T, Rehman A (2016) The practice of secure software development in SDLC: an investigation through existing model and a case study. In: Security and communication networks (2016), wileyonlinelibrary.com, Wiley, Ltd
3. Bodden E (2018) State of the systems security. In: 2018 ACM/IEEE 40th international conference on software engineering: companion proceedings
4. Ali I, Asif M (2018) Applying security patterns for authorization of users in IoT based applications. In: International conference on engineering and emerging technologies (ICEET), IEEE
5. Lee W-T, Law P-J (2017) A case study in applying security design patterns for IoT software system. In: IEEE international conference on applied system innovation
6. Fernandes AM, Pai A, Colaco LMM (2018) Secure SDLC for IoT based health monitor. In: ICECA 2018, IEEE
7. Kolias C, Stavrou A, Voas J, Kuhn R (2016) Learning of internet of things security "Handa-on". IEEE Computer and Reliability Societies
8. Maher ZA, Shah A, Shaikh H, Rahu GA (2018) A methodology for modeling and analysis of secure systems using security patterns and mitigation use cases. In: ICCCE, IEEE
9. Moosavi SR, Gia TN, Nigussie E, Rahmani AM, Virtanen S, Tenhunen H, Isoaho J (2016) End-to-end security scheme for mobility enabled healthcare internet of things. J Future Gener Comput Syst 64:108–124
10. Assal H, Chiasson S (2018) Security in the software development lifecycle. In: Fourteenth symposium on usable privacy and security, USENIX Association, 12–14 Aug 2018
11. Geogy M, Dharani A (2015) Prominence of each phase in software development life cycle contributes to the overall quality of a product. IEEE
12. Djenna A, Saidouni DE (2018) Cyber attacks classification in IoT-based-healthcare infrastructure. In: 2018 2nd cyber security in networking conference (CSNet), IEEE
13. Alvi AK, Zulkernine M (2017) Security pattern detection using ordered matrix matching. In: 2017 International conference on software security and assurance, IEEE
14. Hasheminejad SMH, Jalili S (2009) Selecting proper security patterns using text classification. In: 2009 International conference on computational intelligence and software engineering, IEEE
15. Weiss M, Mouratidis H (2008) Selecting security patterns that fulfill security requirements In: 2008 16th IEEE international requirements engineering conference
16. El Khoury P, Mokhtari A, Coquery E, Hacid M-S (2008) An ontological interface for software developers to select security patterns. In: 2008 19th International conference on database and expert systems application, IEEE

17. Sahumacher M, Fernandez-Buglioni E (2006) Security patterns, integrating security and systems engineering
18. Withall S (2007) Software requirement patterns
19. Hatebur D, Heisel M, Schmidt H (2007) A pattern system for security requirements engineering. In: The second international conference on availability, reliability and security (ARES'07), IEEE
20. Beckers K, Côté I, Goeke L (2014) A catalog of security requirements patterns for the domain of cloud computing systems. In: Proceedings of the 29th annual ACM symposium on applied computing—SAC '14
21. Motii A, Hamid B, Lanusse A, Bruel J-M (2015) Guiding the selection of security patterns based on security requirements and pattern classification. In: Proceedings of the 20th European conference on pattern languages of programs—EuroPLoP '15, ACM
22. Yoder J, Barcalow J (1997) Architectural patterns for enabling application security. In: PLoP '97
23. Yoshizawa M, Washizaki H, Fukazawa Y, Okubo T, Kaiya H, Yoshioka N (2016) Implementation support of security design patterns using test templates. www.mdpi.com
24. Fernandez EB, Pernul G (2006) Patterns for session-based access control. ACM

Classification of Remote Sensing Images Based on K-Means Clustering and Artificial Bee Colony Optimization

M. Venkata Dasu, P. V. N. Reddy and S. Chandra Mohan Reddy

Abstract The Satellite image classification is a complex process. It gives the information about the land cover and thematic land use in remotely sensed data. Image classification is the one of the important application of remote sensing. It is the process of obtaining details about the objects without physical contact. The main objective of image classification is to classify various objects or classes in satellite images. This paper gives an automatic method for the hyper spectral image classification and it classifies the images into multiple land cover objects. The proposed method assigns each pixel in the image to a group of pixels based on reflectance. In this work the acquired input satellite image is preprocessed and then applied classification methods. The proposed classification method includes both spectral and spatial information. In this paper the proposed method uses K-Means clustering algorithm for segmentation and Artificial Bee Colony optimization algorithm is used for classification. The existing method is Particle Swarm Optimization. Both the methods are compared in terms of sensitivity, specificity, overall accuracy and kappa coefficient. The Proposed method has better statistical values compared to the proposed method. This work is simulated in MATLAB R2017b version with system configuration of i3 processor and 8 GB.

Keywords Segmentation · Image classification · Artificial bee colony optimization · Particle swarm optimization

M. Venkata Dasu (✉)
JNTUA, Anantapuramu, AP, India
e-mail: dassmarri@gmail.com

P. V. N. Reddy
SVCE, Kadapa, AP, India
e-mail: principal@svck.edu.in

S. Chandra Mohan Reddy
JNTUACEP, Pulivendula, AP, India
e-mail: email2cmr@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_7

1 Introduction

This Processing and analysis of satellite image is an attractive growth in various applications. These applications include forestry, military, urban planning and agriculture. In achieving these applications classification plays a vital role [1]. A multispectral image contains definite frequencies in electromagnetic spectrum. Classification methods are classified into two methods; they are supervised and unsupervised classification. Supervised classification requires training data i.e. it is required to choose some gray levels (pixels) from each object or class, these are called training pixels [2–4]. The training field is known as ground truth data or the training set. In unsupervised classification is based on pixel classification. Here the user specifies the number of classes. Unsupervised classification is quick and easy to run.

A satellite image contains spatial and spectral components. In this paper we proposed optimization based classification approach. It is a supervised classification approach. In this work we address the solution for the challenges in classification of Hyper-Spectral images with optimization algorithms [5, 6]. The proposed method is Artificial Bee Colony Optimization (ABC) algorithm and existing method is PSO (Particle swarm Optimization algorithm). The performance of the proposed method is compared with the performance metrics like sensitivity, specificity, overall accuracy and kappa coefficient.

The paper is organized as follows. Section 2 describes methodology of proposed work. In Sect. 3 explained about the existing (PSO) and proposed (ABC) optimization algorithms. Section 4 presents the experimental results and discussion on land cover classification and the performances of existing and proposed algorithms are shown and finally Sect. 5 summarizes the total work presented in this paper as conclusion.

2 Methodology

The proposed methodology contains preprocessing and classification. These are explained clearly in following subsections (Fig. 1).

A. **Preprocessing**: The objective of image preprocessing is to improve the image data that eliminates the unwanted distortion and makes the images suitable for further processing. In the proposed methodology the first step is image reshaping and second step is conversion of high resolution input satellite image into gray scale. Gray scale conversion makes the segmentation and classification easier.

B. **Segmentation**: In computer vision image segmentation plays a most significant role for processing of an image. Segmentation is the process of dividing an image into smaller objects. The goal of the segmentation is to simplify the representation of an image. In this paper k-means segmentation technique is used. The detailed explanation of K-means clustering algorithm is explained below.

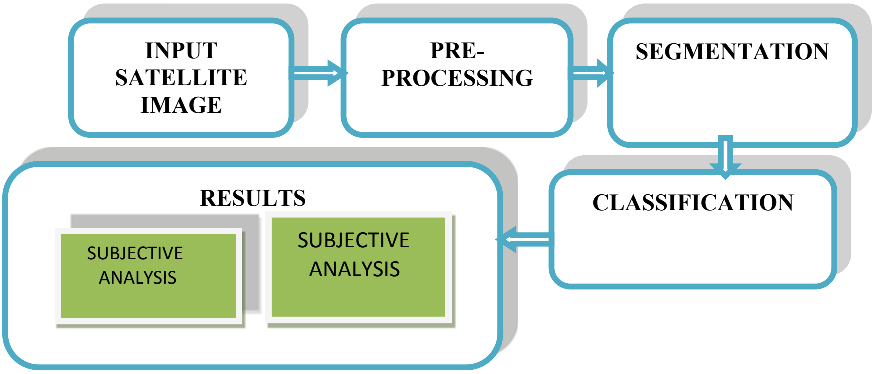


Fig. 1 Block diagram of proposed model

K-Means Clustering Algorithm: Clustering methods are used to identify the group of objects with similar characteristics in the image. K-means clustering algorithm is an unsupervised approach which is used on unlabeled data. The objective of the algorithm is to determine the similar groups in the data and numbers of groups are represented with the variable K. This algorithm has a behavior that it works iteratively to assign each data point to any one of K groups based on features. The main steps of algorithm shown as follows.

- Set the K number of clusters.
- Set a first partition of data.
- Relocation: Move objects from one group to another group to obtain a better partition.
- Optimize some objective function for evaluating the partition.
- Finally the algorithm minimizes the objective function. The objective function is

$$\mathrm{M} = \sum_{m=1}^{k} \sum_{n=1}^{l} \left\| y_n^{(m)} - c_m \right\|^2 \quad (1)$$

3 Classification Methods

The optimization techniques which are depends on the swarm intelligence are called as meta-heuristic techniques or algorithms. They have gained attractiveness in solving complex optimization problems. Swarm intelligence algorithms are inspiring from natural phenomena. In this paper land cover types in satellite images are classified using optimization techniques. The optimization methods used are ABC (Artificial Bee Colony) and PSO (Particle Swarm Optimization. The two algorithms are described clearly in this section.

(A) **Particle Swarm Optimization Algorithm (PSO)**

It is a metaheuristic algorithm proposed by Kennedy and Eberhart. This algorithm is based on the bird flocking. The birds are flying together in a space for food by adjusting their velocity and distances for optimum search [7, 8]. It has two approaches first approach is cognitive and second approach is social. It imitates a bird flying in search space and going towards to Gbest (Global optimum value). In PSO a particle can be defined as $R_j \in (x, y)$ where $j = 1, 2, 3, \ldots, D$, where D represents dimension and J represents real number. A Particle in group has its own position and velocity. Theses are initialized randomly and each particle have to maintain local best known as Pbest and global best Gbest to know the particle velocity [9, 10]. The velocity of each particle calculated as

$$S_j(q+1) = A \times S_j(q) + p_1 t_1 \left[\text{Pbest} - y_i(t)\right] + p_2 t_2 \left[\text{Gbest} - y_i(t)\right] \quad (2)$$

Position of each particle is calculated as

$$\text{Position} = y_i(t+1) + S_j(q+1) \quad (3)$$

(B) **Artificial Bee Colony Optimization (ABC)**

ABC motivated from the foraging nature of honey bees, they introduced to solve the unconstrained functions. It was invented by Dervis Karaboga in 2005 [11, 12]. ABC contains three groups among first group is employed bees second group is onlooker bees and third group is scout bees. The employed bees discovers the food source randomly [13]. The Employed bees shares the details about food source with the onlooker bees in colony by waggle dance and the information includes the direction and distance of the source [14]. The period of waggle dance is equivalent to the nectar content [15]. Therefore onlooker bees attracts for good source position. After finding new food source by onlooker bees and scout bees, they become employed bees, in (Artificial Bee colony) ABC employed bees and onlooker bees performs exploration in search space and scouts control the exploration process [16, 17].

The ABC contains four steps.

- **Initialization process**: The initial food sources are randomly produced via the equation

$$Y_n = p_j + \text{rand}(0, 1) * \left(v_j - p_j\right) \quad (4)$$

Where v_j and p_j are upper and lower bound of solution space.

- **Employed bee phase**: The neighbor food source v_{mi} is determined by

$$v_{mi} = x_{mi} + \phi_{mi}(x_{mi} - x_{ki}) \tag{5}$$

- **Onlooker bees**: The source food quantity is determined by profitability and it is given as profitability is

$$p_m = \frac{f_m(x_m)}{\sum f_m(x_m)} \tag{6}$$

where $f_m(x_m)$ is the fitness of x_m.
- **Scout Phase**: The scout bees are randomly search for the new solution. This new solution x_m will be represented by the expression

$$x_m = l_i + \text{rand}(0, 1) * (u_i l_i) \tag{7}$$

4 Results and Discussions

In this work we considered multispectral satellite image which is acquired from public database such as digital globe. The input image is a high resolution mage. It is shown in Fig. 2. The input image is converted into gray scale image for further processing. The converted gray scale image is shown in Fig. 3.

Fig. 2 Input image. *Courtesy* Digital globe

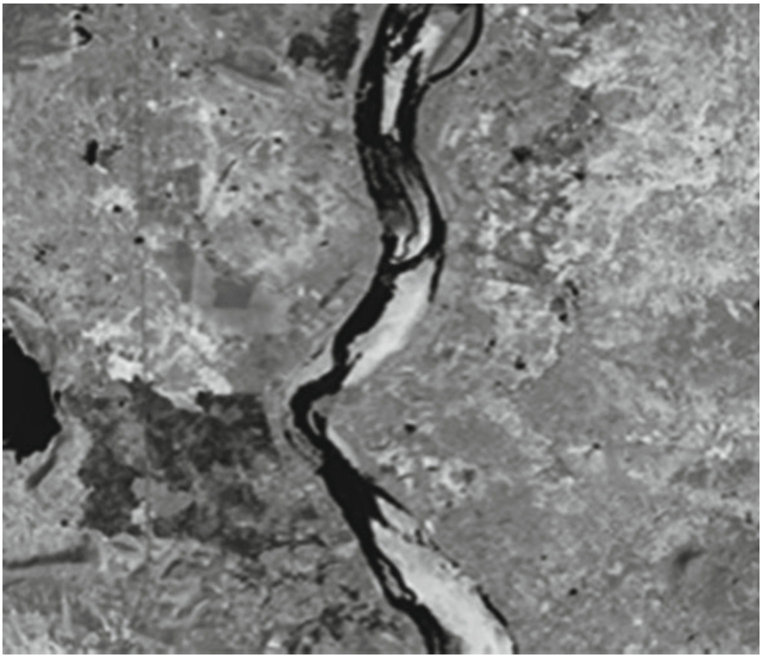


Fig. 3 Gray scale image

Figure 4 shows the Segmented image. It is the important application in computer vision. In this work K-means clustering algorithm is used for segmentation. It makes the edges sharper in the image. Here the number of clusters are taken as k = 4.

Figure 5 shows the classified image using Artificial Bee colony Optimization (ABC). The classified image contains four classes and each class is represented with different color.

In classified image Blue color indicates water content, yellow gives trees, green shows grass and red describes the dry land in multispectral input image. The performance of proposed method is compared with the existing method in terms of sensitivity, specificity, overall accuracy and kappa coefficient. Table 1 shows the comparison of performance metrics.

The obtained values of proposed method (ABC) are better than the Existing Method (PSO). Figure 6 shows the graphical representation of obtained parameters.

5 Conclusion

In this paper, a swarm intelligence based optimization algorithms are used to classify the objects or classes in the high resolution multispectral images. In this work k-means segmentation technique is employed for segmentation. ABC works on total image by

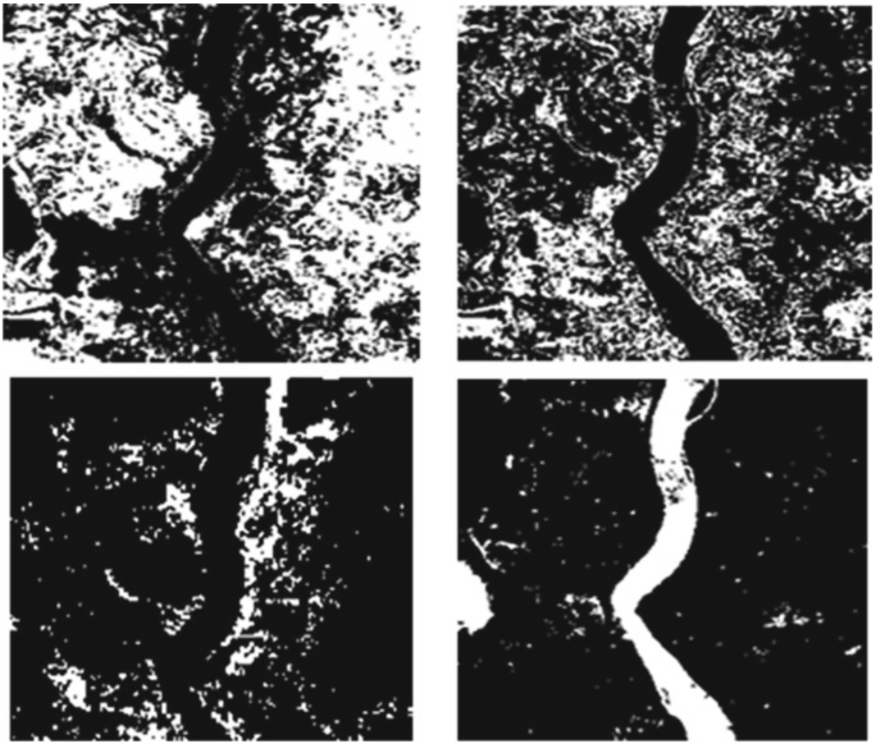


Fig. 4 Clustered image

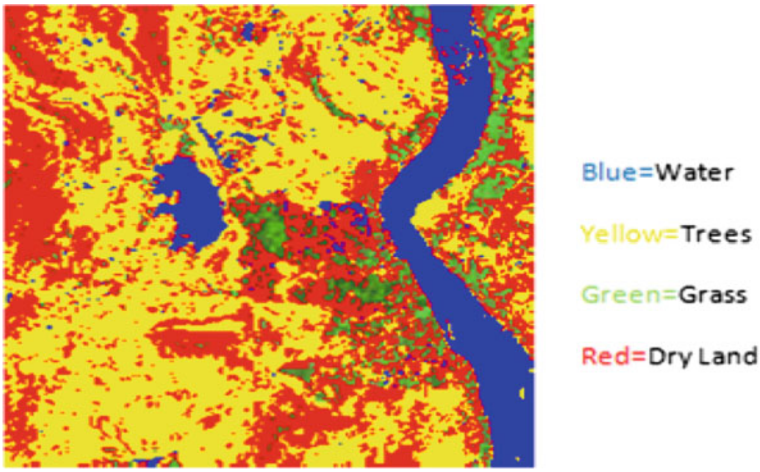


Fig. 5 Classified image

Table 1 Comparison of obtained parameters

S. No.	PSO	ABC
Sensitivity	0.81	0.92
Specificity	0.85	0.97
FPR	0.011	0.021
Overall accuracy	0.89	0.94
Kappa coefficient	0.87	0.92

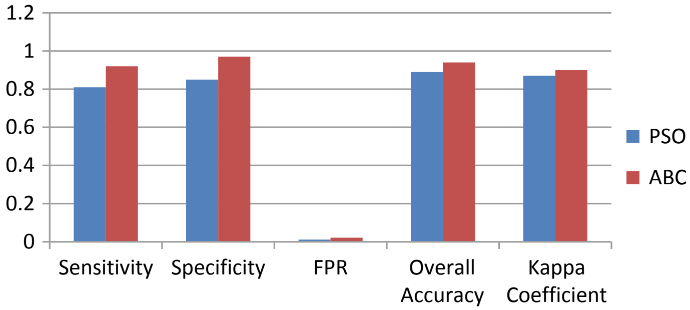


Fig. 6 Graphical representation of parameters

considering each pixel in the image and hence it can recognize heterogeneous parts in the image so that it classifies the each object perfectly. The objects in satellite image are classified using PSO and ABC. From Table 1 we can conclude that the performance of proposed method is better than the existing method.

Acknowledgements The First author express sincere thanks to JNTUA, Anantapuramu, Andhra Pradesh, where he is a research scholar and AITS, Rajampet, A. P. for providing good research facilities.

References

1. Thakur N et al (2017) A review on image classification techniques. Int Res J Eng Technol (IRJET) 1588–1591
2. Menaka D et al (2014) Land cover classification of multispectral satellite images using QDA classifier. In: International conference on control, instrumentation communication and computational technologies (ICCICCT), pp 1383–1387
3. Senthilnath J, Omkar SN, Mani V, Karnwal N, Shreyas PB (2013) Crop stage classificat ion of hyperspectral data using unsupervised techniques. IEEE J Sel Top Appl Earth Observ Rem Sens 6(2) (2013); Smith TF, Waterman MS (1981) Identification of common molecular subsequences. J Mol Biol 147:195–197. doi:10.1016/0022-2836(81)90087-5
4. Omkar SN, Sivaranjini V, Senthilnath J, Suman M (2010) Dimensionality reduction and classification of hyperspectral data. Int J Aerosp Innov 2(3):157–163

5. David L (2002) Hyper spectral image data analysis as high dimensional signal processing problem. IEEE
6. Richards JA, Cluster-space representation for hyper spectral data classification. IEEE J Sel Top Geosci Rem Sens Lett 16
7. Imran M, Jabeen H, Ahmad M, Ababs Q, Bangyal W, Ababs Q, Opposition based PSO and mutation operators(OPSO with power mutation) In: 2nd International conference on education technology and computer, Shanghai, p V4-506–508
8. Pant M, Thangaraj R, Singh VP, Abraham A (2008) Particle swarm optimization using sobol mutation. In: First international conference on emerging trends in engineering and technology, Nagpur, Maharashtra, pp 367–372
9. Imran M et al (2012) An overview of particle swarm optimization variants. In: Malayasian technical universities conference on engineering & technology
10. Alam S et al (2014) Research on particle swarm optimization based clustering: a systematic review of literature and techniques. Elsevier
11. Holland JH (1975) Adaptation in natural and artificial systems. The University of Michigan Press
12. Dorigo M (1992) Optimization, learning and natural algorithms. Ph. D. Thesis
13. Kennedy J, Eberhart R (1995) Particle swarm optimization. In Proceedings IEEE International Conference on Neural Networks, vol 4, pp 1942–1948
14. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings IEEE International Conference on Neural Networks, vol 4, pp 1942–1948
15. Abu-Mouti FS et al (2012) Overview of artificial bee colony (ABC) algorithm and its applications. IEEE
16. Karaboga D (2005) An idea based on honey bee swarm for numerical optimization. Erciyes University, Engineering Faculty, Computer Engineering Department, Tech Rep TR06, pp 1–10
17. Yang XS (2008) Nature-inspired metaheuristic algorithms. Luniver Press, London

Internet of Things and Its Applications: An Overview

Sanjeeva Polepaka, M. Swami Das and R. P. Ram Kumar

Abstract Today the use of Internet of Things (IoT) applications increasing rapidly to use automation in the areas including home, hospital, office, and others. In the future, the people expected to use IoT applications in the year 2025 by 11 trillion dollars business worth with use. The problem is to develop an Intelligent IoT system application based on user specifications, and the demand for application with the use of elements, models, and providing services. The objective of the paper is to explore IoT elements, framework-architecture, analysis, and monitoring, and case study applications, such as, Heath care System, Pilotless plane, Smart lighting, Home IDS, Smart Parking, Smart Irrigation, and Intelligent Transport System applications. Finally, effective utilization of elements, protocols, components, and sensors networks, AI applications, and Machine learning applications predict the future patterns based on present and past data of applications. The conclusion of the proposed model improves the quality of the IoT system.

Keywords IoT · Web service · Cloud computing · Pervasive computing

1 Introduction

IoT is the internetworking of physical devices (i.e. connected devices, electronics systems, software, sensors, actuators) and other things using and network connectivity. These connected things by using sensors together to communicate using sensors and analyze and operations of various applications, for example, Healthcare, Smart cities, etc. Nowadays the demand to use of IoT applications in the world increasing rapidly for example, in 2017–18 the people usage is 2.4 billion dollars, and by

S. Polepaka (✉) · M. Swami Das · R. P. Ram Kumar
Malla Reddy Engineering College (A), Hyderabad, Telangana, India
e-mail: psanjesus@gmail.com

M. Swami Das
e-mail: msdas.520@gmail.com

R. P. Ram Kumar
e-mail: rprkvishnu@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_8

expected use of 2020 and 2025 is around 1.7 trillion dollars and 11 trillion dollars, respectively [1, 2].

The users use IoT applications around the world effectively by using automated tools and applications. The use of IoT based applications are accessed and stored information, the analysis and monitoring the applications according to the on the real-time scenario of operations and functions. Some of IoT based automated applications are Smart home, Intelligent Transportation, etc. The objective of the proposed method deals with building an Intelligent IoT application with I/O interfaces, sensors (to collect the data) and developing various applications based on the specifications and demands, analyze the data and report information to stakeholders. The rest of the paper covers with Sect. 2 Related work, Sect. 3 Elements of IoT, Sect. 4 Frame work-architecture, Sect. 5 Analysis, Monitoring, and reporting; Sect. 6 Case study applications, and Sect. 7 Conclusions with Future scope.

2 Related Work

Park [3] proposed an M2M to IoT ecosystem; this consists of architecture M2M service, service, and interconnecting to other devices. Mehmood [4] use of IoT based applications are growing, sensors, world digital, machine learning, to provide better security, mobile apps. Akram [5] Emergency medical service, medical treatment by the hospitals. The event, that the patient health conditions immediately informed to the hospital and guardians, with the help a system called Emergency Medical IoT system.

Mehtaa [6], used IoT systems for Services into things, internet, and semantic oriented communication, the advantages are Application use machine to machine interactions and disadvantages not specified a Generalized model for building intelligent IoT application. Lin [7] suggested a building IoT model by social network model; Model is social IoT, dynamic groups, a goal from node to node interactions by considering trusty relationships. The advantage is service factors by the structure of the social network, based on the address of social IoT, the trustworthiness of objects.

3 Elements of IoT

The elements of IoT as, first, namely I/O interfaces are used for Identification-sensing, second Communication (i.e., IoT communication to devices. Stakeholders (citizens, government and public service providers Communication parties using Transport between devices to backend and users. Wi-Fi, WinMax, BlueTooth, GSM, GPS protocols), third, is Application and computations (i.e., Offer services; Smart transportation, healthcare, safety and security, smart cities, E-security) and fourth is Analysis and reporting (i.e. Analyze information and predictions the report information send

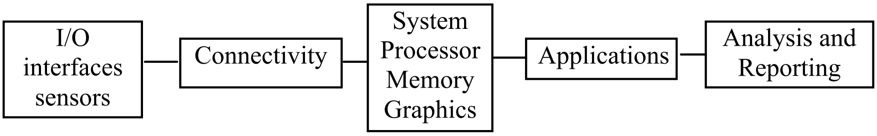


Fig. 1 IoT building model

back to the stakeholders). Figure 1 shows the elements of the IoT Model for building various applications. IoT devices protect, interoperable, and communicate to together devices with a unique identity.

I/O interfaces are sensors actuators connected devices to access system with applications using networked communications. Communication, connectivity and internet communication to access the applications using sensor controlled devices, and finally the real-time data collected from input devices, stored and analyzed with recommendations and actions [8]. The elements of IoT are Sensors are used to collect the data. Network components, protocols are used for communication to the web, and data storage processing, applications. Sensors use of devices to remote and continuous monitoring of applications. A network of sensors, Actuators, Mobile devices, the IoT is connected technologies gained and strength the need of services Monitoring system [9]. Products used in IoT is radio connectivity chip with radio protocols sensors is chip contains a measurement of environmental/electrical variables, Microcontrollers contains a processor, storage with the low-cost chip. The system with application software is integrated into modules services IoT solutions deploy, manage, and maintain. IOT applications like Intelligent Home appliances [10].

Figure 2 shows the remote and local user of IoT in Cloud-based applications consists of users, Wireless Sensor Networks (WSN) is centralized digital wireless solutions to distributed pervasive wireless systems. The networks with IoT devices connecting the World Wide Web. These connected devices, network systems require low cost, low power consumption wireless sensors, and WSN [11]. For example, it has users, local and remote, and WSN, has a front end, data sensor, with communication protocols to IoT applications resources. These are the local resources with remote resources, IoT software systems to prove human interactions to cloud and IoT applications [12]. Cloud computing with centralized cloud servers, distributed computing, direct connection to the internet, and Hybrid computing using IoT. Cloud computing, services on demand, Platform as Service (PaaS), Infrastructure as Service (IaaS), Software as Service (SaaS), Ubiquitous information, many devices operating and interacting the devices.

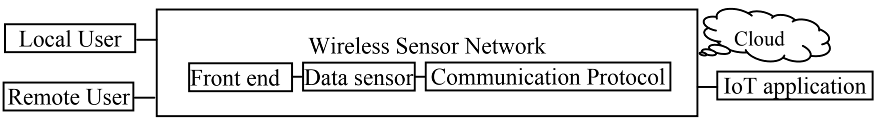


Fig. 2 Communication for the remote and local user of IoT in cloud base applications

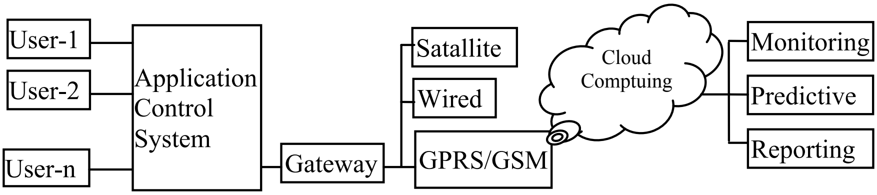


Fig. 3 IoT based generalized application-monitoring and reporting system

IoT applications usability of the product and services, Smart wearable, smart homes, smart cities, smart enterprise, smart environment [9]. To improve resource utilization by the relationship between the human and nature, dynamically control the operations, Intellectual entity by human society, physical system, Transport, internetworking, accessibility, the usability of systems [1]. IoT is a physical, virtual object to access applications by configurations, networking, and dynamic systems to process the information.

4 Framework-Architecture

The Generalized a model—architecture shown in Fig. 3 consists of Users are the end users service consumers. Application system provides services (i.e., various services to the end user, for example. the healthcare system, smart cities, etc.). The framework application main components are users, application control system, With communications services, using satellite, Wired, wireless and GPRS systems to process information to cloud-based applications for providing services like, (1) IaaS (2) SaaS and (3) PaaS, cloud models, and Analysis by Monitoring, and reporting information. The data collected from form resources and services, The system provides portability, high availability, and application components, use interface, components, backend database, and data representations, evaluation at servers to analyze the information and predictions, results to stakeholders based on analysis reports (i.e., Intelligent Information Systems (AI systems)) using Data Mining and machine learning algorithms [9].

5 Analysis, Monitoring, and Reporting

To analyze the data collected from sensors of various locations, which is coming from WSN and stored on Cloud servers. The system use of Nonfunctional things like nodes, gateway, nodes, connections, communications, gateways, sensors, etc. functional requirements, IaaS, cloud providers, optimal query, fetch relevant data, access, visualization of clients, using internet standards, secured access, the deployment has configured, deploy the service and run service. The collected data predicted models based on present and past data predict the future patterns, for example, forecast damage, recommend and suggestions, and best algorithms used, the Wireless node is designed to process the data, gateways, wireless antennas used, routers are used [10]. IoT smart objects communicate together to make an easy life. IoT smart devices, IoT framework, model for the development of IoT application, IoT software architecture, physical objects, networks, sensors, traditional systems, and cloud resources keeping human users.

The IoT application a small sender and gateway connection with records and keep track of the data and transferring using gateway and communication protocols and send to IoT Analyzer and monitor. This will record the events, and according to the events, continuously record using sensors and recorded in the cloud, and preventive measure and safety, monitoring climate situation like Smart Heritage application will use wireless sensors, and hidden nodes in industrial, scientific technologies. Considering the various important parameters (features) used for identification of scenario, patterns, predicted reports to visualizations. [13]. For example, in healthcare application based on clinical data, the system will predict the precautions-treatment to the patient in time into the hospital and immediate actions. The various kinds of case study applications are discussed in Sect. 6.

6 Case Studies—Applications

(1) **Health Care System**: Healthcare devices are used to record the patient's data. The Continuous monitoring of health conditions and correlate the physiological parameters and health data for prediction and analysis. (1) Record the clinical data. (2) Provide treatment by a doctor. (3) Reduce the healthcare cost by accuracy diagnoses using IoT devices. (4) Management—storage, visualization, remedies, and functions based on analysis, and predictions [3].
(2) **Pilotless Plane**: Artificial Intelligence application pilotless plane using sensors, data and processing, and operations, functions flight will move the specified source to destination based on IoT controlled system.
(3) **Smart Lighting**: Home automation, the control the lights remotely using Web applications The system is designed with the use of Model is REST based architecture which provides the services. First, the sensor device receives the request and (mode and sate, Mode(auto or manual using PUT request), receives the GET request, the mode updated from the database, set the light state ON/OFF based on put request State is updated based on the status of the database [8].
(4) **Home Intrusion Detection System**: Door sensors, and raise alerts are used in Home IDS if necessary to record the data. Read the Motion of Sensor data at each door, detect door sensor and motion of opening, each sensor at regular intervals detected the motion, and events stored in the and alert sent to the database. Based on the analysis of stored data with new data find the intrusion and report accordingly virtual entities, room with attribute states, using REST

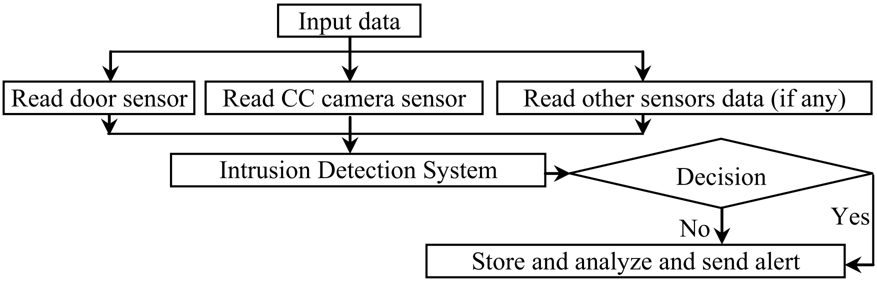


Fig. 4 IoT based home intrusion detection system

Table 1 IDS dashboard alarms status

Room No.	Doors	Alarm status	Doors
Room2	Door1	ON/OFF	Door2
Room1	Door2	ON/OFF	Door3

services. If an intrusion is detected, means send the information to the Owner of the house, or Police, etc. Figure 4 shows Home IDS and Table 1 shows the alarm status.

(5) **Smart Parking**: This can be used for Metro railway Stations by Virtual Parking state; attribute: Parking slot: ID, Parking lot, Attribute state: attribute name: slot location, state: State: is Empty or Occupied.

(6) **Smart irrigation**: Soil moisture sensors, determine the amount of moister in the soil, and flow of water using irrigation pipes, is moister goes down means pre-defined threshold, the data moister levels collected from the cloud and analyze a plan for watering and scheduling plants based on the report [14].

Pseudo Code for Agriculture—Smart Irrigation

Step 1. Read the moisture data stream from various channels
Step 2. Set mode, trigger, find the threshold
Step 3. Sensors connected to moisture, Read the level of moisture
Step 4. If (level < threshold), trigger pin TRUE
else trigger Pin FALSE
Step 5. Store moisture data stream with the current value and controller set up
Step 6. If (moister location and update) true do according to the instructions/watering, and plant the trees based on Recommendation and suggestions.
Step 7. Record the results and information to the formers and user.

(7) **Intelligent Transport Systems**: Use of vehicles, trucks, cars, too have sensors. The sensors collect the data from GPRS/satellite route and traffic, secured IoT enabling technologies. RFID is to identify and track the data of things. Sensors are data collection and detecting any changed in the physical status of things; Nano Technology to make smaller and smaller things ability to connect and

interact. M2M used for internetworking functions. The footprints of M2M lies in internetworking, industries domains, discovery and security applications [3].

7 Conclusion and Future Scope

In future, by 2020, ability the devices located indoors to receive geological signals, locating the people using every object, and efficient teleoperation and monitor control, software agents and sensor controlled devices using web applications. IoT objects are sensor controlled network infrastructure.

IoT with sensor and actuator technologies use as smart home, smart grids, intelligent transportation, and smart cities. The devices to detect changes in the physical status of things, collect details, and send to cloud data storage systems. The main three core sectors used IoT are Enterprise, Home, and Government applications [10, 15].

The reliability is one of the most important parameters, collecting, devices, the operation time of companies. IoT in real time monitoring the advantage of sensor-controlled reliability monitoring products and service in the organization used embedded meters, automatically detect the failures. In the future, the data collection factors tracking the objects, the things using processors, software, and sensors connection with communication infrastructures, using wired or wireless communications [16].

IoT in the future, when objects are interacting with message communications, by using the internet, it is easy interaction of devices in other areas like home applications, surveillance cameras, smart city, and another application. The main objective of IoT applications with the use of communication facilities to control the operations, activities. The other services of Smart city including (health, management of waste, quality air monitoring, control traffic, energy consumption, smart parks, noise monitoring and automation of buildings smart lighting that will optimize the street lights intensity of time, and weather conditions that exploit to increase number of communications connections using street light controllers.

Automation public buildings, schools, admin offices, museums with temperature, humidity, and enhance web service applications, Service-Oriented Architecture (SOA) and Representational State Transfer (REST). Use of IoT for both user and developers able to for data transfer HTML, XML, for constrained EXI, for HTTP, TCP for constrained in application transports used SOAP and UDP, for network IPv4, IPv6, and constrained IPV6 and 6LoWPAN are used in IoT nodes. For the protocols of functional data collection from sensors, various applications, smart city, smart health, etc., and transportation protocols that use network connectivity with entities. For data format, and application transports, HTTP and TCP used, SOAP is binary format transported used for UDP transport, retransmission for reliable services, using REST protocols get, post, delete and put methods, for network protocols, efficient communication, to access interoperability of IoT nodes on the internet [17]. The effective utilization of sensors, protocols, AI applications to achieve the quality of

IoT applications. In the future, the IoT service provider will use the suggested model-architecture with elements and concepts in various real-time sectors will improve the overall performance of systems.

References

1. https://www.tcs.com/internet-of-things
2. IoT 2020: Smart and secure IoT platform, pp 1–181
3. Park H, Kim H, Joo H, Song JS (2016) Recent advancements in the Internet-of-Things related standards: a one M2M perspective. In: ICT Express. 126–129. https://doi.org/10.1016/j.icte.2016.08.009
4. Mehmood Y, Ahmad F, Yaqoob I, Adnane A, Imran M, Guizani S (2017) Internet-of-things-based smart cities: recent advances and challenges. IEEE Commun Mag 16–24. https://doi.org/10.1109/mcom.2017.1600514
5. Akram A, Anjum M, Rehman M, Bukhary H, Amir H, Qaisar R (2017) Life savior: an integrated emergency response system. In: International conference on information technology (ICIT), pp 1002–1006. https://doi.org/10.1109/icitech.2017.8079981
6. Mehta R, Sahni J, Khanna K (2018) Internet of things: vision, applications and challenges. Procedia Comput Sci 132:1263–1269
7. Lin Zhiting, Dong Liang (2018) Clarifying trust in social internet of things. IEEE Trans Knowl Data Eng 30(2):234–248. https://doi.org/10.1109/TKDE.2017.2762678
8. Bahga A, Madisetti V (2015) Internet of things a hands-on approach, Universities Press, pp 20–326
9. Gupta R, Gupta R (2016) ABC of Internet of Things: advancements, benefits, challenges, enablers, and facilities of IOT, CDAN. IEEE, pp 1–5. https://doi.org/10.1109/cdan.2016.7570875
10. Hassanalieragh M, Page A, Soyata T, Sharma G, Aktas M, Mateos G, Kantarci B, Andreescu S (2015) Health monitoring and management using internet-of-things (IoT) sensing with cloud-based processing: opportunities and challenges. IEEE, pp 285–292. https://doi.org/10.1109/scc.2015.47
11. https://en.wikipedia.org/wiki/Internet_of_things
12. Hernández MEP, Reiff-Marganiec S (2016) Towards a software framework for the autonomous internet of things. IEEE, pp 220–227. https://doi.org/10.1109/ficloud.2016.39
13. Perlesa A, Pérez-Marín E, Mercadoa R, Damian Segrellese J, Blanquere I, Zarzoc M, Garcia-Diegob FJ (2017) An energy-efficient internet of things (IoT) architecture for preventive conservation of cultural. Futur Gener CS 1–64. https://doi.org/10.1016/j.future.2017.06.030
14. https://www.cbc.ca/radio/thecurrent/the-current-for-july-31-2018-1.4767971/would-you-fly-in-a-pilotless-plane-ai-aircrafts-are-on-the-horizon-1.4768015
15. Sucheta PSV, Vijaya Lakshmi D, Swami Das M (2018) Building an IoT web services for intelligent systems. Int J Pure Appl Math 120(6):3101–3117
16. Bowman K, Cline B (2015) How the internet of things will improve reliability tracking. IEEE, pp 1–5. https://doi.org/10.1109/rams.2015.7105179
17. Zanella A, Bui N, Castellani A, Vangelista L, Zorzi M (2014) Internet of things for smart cities. IEEE Internet Things J 1(1):22–32. https://doi.org/10.1109/jiot.2014.2306328
18. May P, Ehrlich H-C, Steinke T (2006) ZIB structure prediction pipeline: composing a complex biological workflow through web services. In: Nagel WE, Walter WV, Lehner W (eds) Euro-Par 2006. LNCS, vol 4128. Springer, Heidelberg, pp 1148–1158. https://doi.org/10.1007/11823285_121
19. Foster I, Kesselman C (1999) The grid: blueprint for a new computing infrastructure. Morgan Kaufmann, San Francisco, pp 13–50

20. Czajkowski K, Fitzgerald S, Foster I, Kesselman C (2001) Grid information services for distributed resource sharing. In: 10th IEEE international symposium on high-performance distributed computing. IEEE Press, New York, pp 181–184. https://doi.org/10.1109/hpdc.2001.945188
21. Foster I, Kesselman C, Nick J, Tuecke S (2002) The physiology of the grid: an open grid services architecture for distributed systems integration. Technical report, Global Grid Forum
22. National Center for Biotechnology Information. http://www.ncbi.nlm.nih.gov

Smart KYC Using Blockchain and IPFS

Nikita Singhal, Mohit Kumar Sharma, Sandeep Singh Samant, Prajwal Goswami and Yammanuru Abhilash Reddy

Abstract Know your customer, also known as know your client or simply KYC, is the process that businesses and financial institutes must employ to identify their clients and assessing any kind potential risk due to illegal intentions and foul play for the business relationship in compliance with a national regulating body. The term KYC is often used to refer the bank regulations and the anti-money laundering regulations which are in place to govern such activities. Also due to bribery and foul play, companies of all sizes are compelled to employ KYC for the purpose of ensuring their consultants, agents, or distributors follow rules set by anti-bribery compliant. With population of India around 1.3 billion, a secure and faster system for sharing sensitive information like KYC document which may contain personal document, capable of handling this vast amount of this data is of high demand. While the implementation of such a system isn't new, the present systems have drawbacks. The proposed system will replicate the functionality of the legacy KYC system. By using the immutable property of Distributed Ledger Technology (DLT) and Inter Planetary File System (IPFS), a tamper-proof system can be formed. This paper aims to address some of the shortcomings of the current system and propose implementation of innovative features to develop a more secure and comprehensive system. The proposed system will allow customer and business institute to verify and record the customer KYC document into the DLT. The proposed system will use IPFS which will greatly improve the storage efficiency of DLT.

N. Singhal · M. K. Sharma · S. S. Samant (✉) · P. Goswami · Y. A. Reddy
Army Institute of Technology, Pune, India
e-mail: sandeepsamant1702@gmail.com

N. Singhal
e-mail: ngupta@aitpune.edu.in

M. K. Sharma
e-mail: mmks987654@gmail.com

P. Goswami
e-mail: prajwalgoswami1996@gmail.com

Y. A. Reddy
e-mail: abhilashreddyy1@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_9

Keywords Blockchain · IPFS · KYC · DLT

1 Introduction

With growing rate of money transactions each day, Banks have to deal with a myriad of sham transfers and the culprit gets away easily with it [1]. This has led to a condition where we need to shift gears and abandon the old methodology of doing KYC [2] and document verification. People in last few years have proposed ideas to adopt Blockchain [3] as the solution to the given problem. But not much change was observed either because of too much complexity or flaws in the system. The motivation behind the paper is to present a simple and efficient solution to the current KYC system with the help of modern day technology which will greatly reduce complexity in the present system and improve the customer experience. Our proposed system not only solves the problem of doing KYC but can also provide assistance to verify attested documents as required by other non-banking organizations. This paper aims to address the various issues that bank and customer face and also revamp the former KYC system by adding a new dimension IPFS [4].

1.1 IPFS

IPFS viz. Interplanetary File System. It is a p2p (peer to peer) distributed hypermedia protocol that is designed to act as a ubiquitous file system for all computer systems. It is also open-source. It is a complex and highly ambitious project with many major and important implications on the structure of internet in the future. In p2p (peer to peer) protocol every node possesses a collection of hashed files. Any client who wants to retrieve any of these file is presented with a simple abstraction layer where only hash of the file needs to be called to get the file. IPFS then digs across the nodes and supplies the client with the called file.

1.2 DLT

Distributed ledger technology or DLT is a recording of transaction of assets in detailed format stored in digital system simultaneously at multiple places. Unlike traditional databases, DLT don't have any central data store or any kind of administration functionality. In DLT, every node process and checks the integrity of every item and create its record by doing a consensus on each item's veracity. A distributed ledger can be used to store both static and dynamic data such as registry and transactions respectively.

2 Present Versus Proposed System

With population of India crossing the 1.3 billion mark and over 300 million banks account to be administered, banks are finding it difficult to manage the tedious paperwork involved. The old KYC process comprised of the user sharing necessary identity documents with the banks or financial institutions such that the bank carries out background check and verification for the same which involves replication of the provided documents multiple times thus increasing paperwork and risk factor. Owing to the risk and cost involved the old system is not at all efficient. With the advent of new technologies each day many researchers have proposed solutions to the above problem quite recently. In 2015 the research carried out by Arasa and Ottichilo [5] suggested that there were four variable factors that explain 78.3% of the level of compliance of KYC requirements by commercial banks in Kenya. Soni and Duggal [6] in 2014 used Big Data Analytics to put forth a solution for problem of large data involved in KYC. They used techniques like Fuzzy matching and Map Reduce for the same. Moyano and Ross in 2017 [7] suggested the use of Distributed Ledger Technology (DLT) for doing KYC such that KYC is carried out only once for each user thus improving the cost involved. But still there is a problem of huge storage cost involved. Though the present system is secure and tamper proof but it is not cost effective. Storage on Blockchain is way too expensive. Estimates state it cost $100 per GB of storage. One can easily buy a 500 GB hard drive for $100 in today's world [8]. The proposed system makes use of IPFS to tackle this problem. IPFS abbreviates to *Interplanetary File System* and is a peer to peer distributed file system with advantages like content addressing, no duplication etc. For every file we get a unique IPFS hash which is of a constant length. Rather than storing complete file, IPFS stores the hash of the file on the blockchain. This hash can then be used to locate the actual file, thus avoiding duplication of file on the blockchain. It is just like calling someone to your place and rather than physically moving your accommodation to them, you provide them a pointer to where you actually stay such that it is easy for them to reach you. We will also be employing a two ways authenticity check of documents by linking the user's files with the verifier id.

3 Proposed Architecture

The system proposed in this paper is a Distributed Leger Technology (DLT) based document verification or KYC system which uses a decentralized File storage (IPFS) to store user details and documents for verification [9]. The system proposed in this paper incorporates all the functionality provided by a standard KYC system. The user will have to provide a username for the creation of the wallet which will be used for storing all data relevant to the user under his username. A key pair containing private and public key [10] will be generated which will support adding or updating in the blockchain against the username and identifying the user respectively.

The file to be verified will first be uploaded to the IPFS node which will return an IPFS hash pointing to the document in the IPFS node. This hash is then stored in the blockchain which is then verified by an authorized party which on successful verification adds the hash and the user public key in the verified document list with its public key and verifier's address in the document details. The user can then provide the digital document from which the hash can be calculated or directly provide the IPFS hash along with verifier address through which the concerned party can confirm the authenticity of the document by cross examining the presence of the documents IPFS hash both in the list of verified documents in the blockchain pointed by the verifier's address and user's address.

The proposed system has multiple components which are

(a) Document Submission
(b) Notary Verification and
(c) Third Party Verification.

Each of the components is independent of each other which makes the system more flexible and easier to incorporate any changes.

3.1 Document Submission

Document submission is the stage when the user who wants to submit documents or provide KYC, uploads the documents to the decentralized file storage IPFS for the documents to be accessible by the notary and to ensure that the data that is notarized cannot be tampered with as the documents can only be accessed based on the hash of the document. This ensures that each distinct document that is uploaded to the decentralized file system has a different hash and the data once approved by notary always be available as makes it is extremely difficult for anyone who is trying to delete the document as it is stored with enough redundancy and at various nodes in the network which might be geographically very far apart. The documents related to a specific user are stored in a smart contract on the blockchain. As the documents are stored on IPFS which is tamper proof and available without depending on a specific organization or service like Amazon S3, Azure Blob storage and as the IPFS hashes of this data is stored on blockchain which again is always available independently due to which the documents which are stored in this stage are independently available with-out depending on a specific service or organization to maintain the data. The contract which stores the IPFS hashes has access control mechanisms so that only the user can add or change data in the contract.

3.2 Notary Verification

Notary verification stage is the stage after the user uploaded the documents to IPFS, the user sends a request to a notary who decide on the authenticity of the data. If the data in the documents is authenticated, then the IPFS hash of the document is added to the list of verified documents by the notary. Notary can be any one based on who is authorized to verify the document. In case the document is a driving license, then the notary can be Transport Department of Government. If the document is a certificate specifying the salary of an individual, then the notary can be employer of that individual. Similarly, this data can be authorized by the appropriate party. This verification can be done independently and can be reused anywhere.

3.3 Third Party Verification

Third party verification is the last stage of the proposed KYC verification process. Before this stage user must upload all the documents and get them authorized by respective authorities by adding them to the list of IPFS hashes of approved documents. In this stage the user submits the contract address of the user, IPFS hash of the specified document and address of approving authority. The address of the approving authority should be publicly accessible for anyone to verify. The third party who wants to verify the document can access the document based on the IPFS hash and then verify if the document is approved by checking if the IPFS hash of the document is present with the contract (Figs. 1 and 2).

4 Benefits of the System

The benefits of the proposed system include Better governance of data as Data alterations can be tracked and monitored, direct access to the KYC data saving huge amount of time and lastly storage space is effectively utilized as inspite of saving the complete file on Blockchain, only hash is stored over it.

5 Future Aspects

Today cryptography can help us to achieve secrecy and integrity of our data and as a result authentication and anonymity to our communications. We look forward to implementing cryptography for document uploading process with end to end encryption. Incorporating cryptography in our proposed system will increase its security and improve the integrity aspect as well.

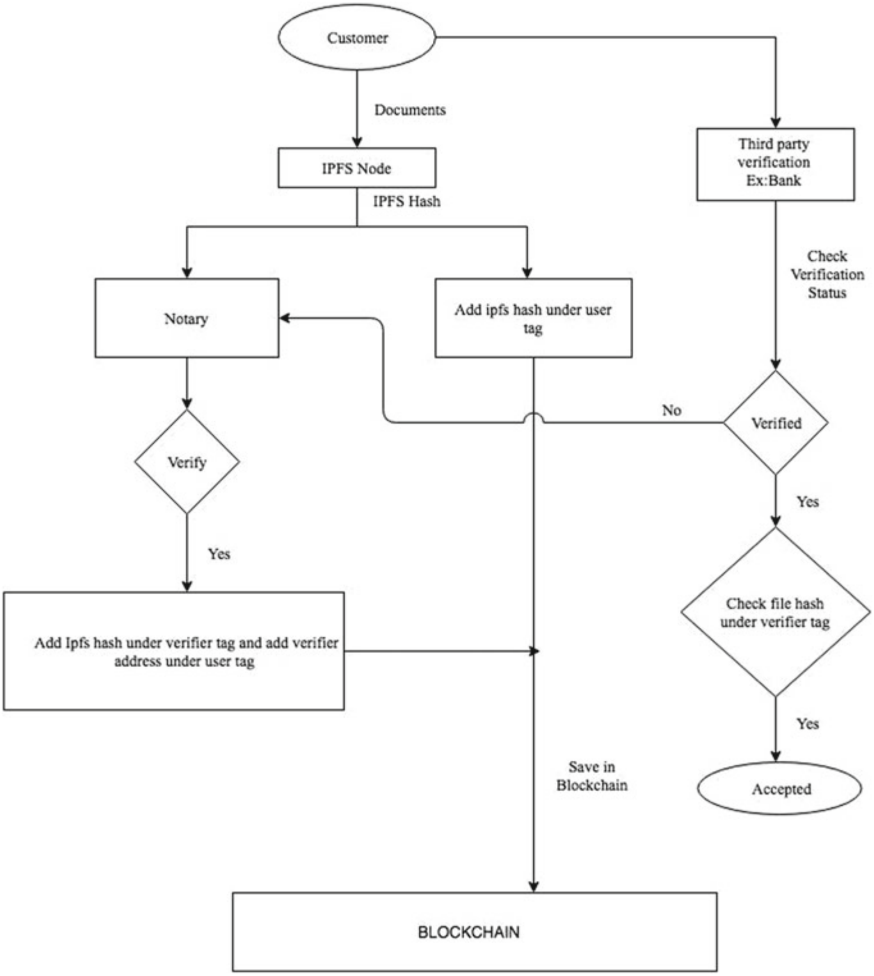
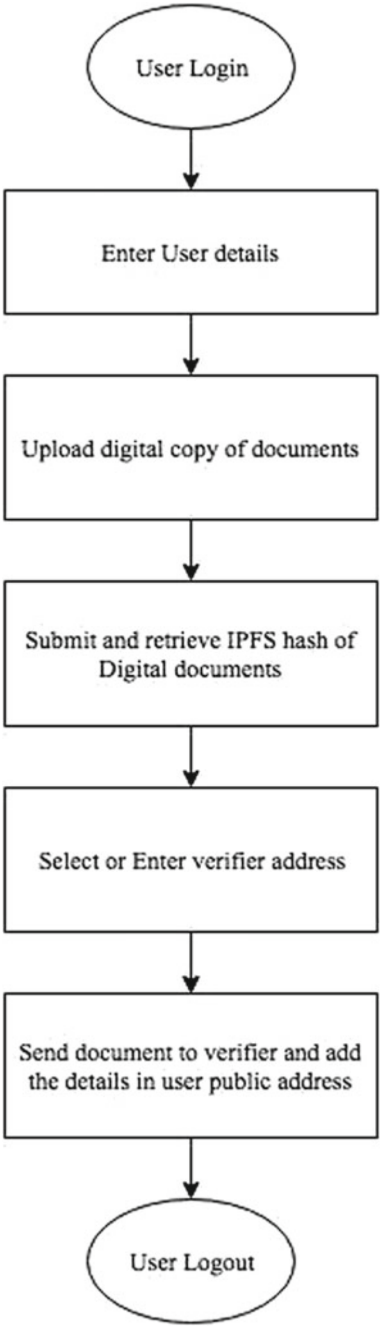


Fig. 1 Design of decentralized KYC system

6 Conclusion

Blockchain is a disruptive technology [11] and its use is being enhanced each day. The proposed system is a needed replacement of the one in existence although such systems' discussion is not unprecedented, but we've tried to build it with improvisation on the storage part, all thanks to IPFS. IPFS is cost effective and it is used throughout, making our system space efficient. The user will also be benefitted from

Fig. 2 UI data flow



the security aspect of the architecture as third-party involvement is completely eliminated. Lastly, we have to tried to compare our system with IPFS against the previous systems. The proposed system will bring a change in the lives of both user and the banks. In future we propose to implement this system on a larger scale to seek more results.

References

1. Canada: Immigration and Refugee Board of Canada. India: Availability and prevalence of fraudulent identity documents, including membership cards of political parties (2011-April 2014), 5 May 2014, IND104839.E, available at: https://www.refworld.org/docid/538c369f4.html
2. Rajput VU (2013) Research on know your customer (KYC). ISSN 2250-3153
3. Zheng Z, Xie S, Dai H, Chen X, Wang H (2017) An overview of blockchain technology: architecture, consensus, and future trends. In: 2017 IEEE 6th international congress on big data
4. Benet J (2014) IPFS-content addressed, versioned, P2P file system. arXiv preprint. arXiv:1407.3561
5. Arasa R, Ottichilo L (2015) Determinants of know your customer (KYC) compliance among commercial banks in Kenya. J Econ Behav Stud 7(2):162
6. Soni A, Duggal R (2014) Reducing risk in KYC (know your customer) for large Indian banks using big data analytics. Int J Comput Appl 97(9)
7. Moyano JP, Ross O (2017) KYC optimization using distributed ledger technology. Bus Inf Sys Eng 59:411–123 https://doi.org/10.1007/s12599-017-0504-2
8. Kohli K (2018) Ipfs + blockchain = decentralised file storage, Coinmonks, available at: https://medium.com/coinmonks/ipfs-blockchain-decentralised-file-storage-9ef3a1fa307b
9. Pachaiyappan V, Kasturi R block chain technology (DLT technique) for KYC in FinTech domain: a survey. Int J Pure Appl Math
10. Merkle RC (1980) Protocols for public key cryptosystems. In: Proceedings 1980 symposium on security and privacy, IEEE Computer Society
11. Donaldson J (2017) What is blockchain and why is it disrupting every industry? available at: https://mojix.com/blockchain-technology

Lexical Analysis and Mathematical Modelling for Analysing Depression Detection of Social Media Reviews

V. Kakulapati and S. Mahender Reddy

Abstract Majority of people are suffered with depression though out the world. This is the major reason for disability of humans. When depressed person is diagnosed by physicians physically via referring to clinical methodology. Though, the major portion of the diseased are not approach physicians at beginning the stage of depression, this direct to promote worsening of patient conditions. In the meantime, these persons are taking support of social media to reveal feelings and sharing their actions as a result, social media is the main source of leverage for detecting physical and mental illness of such people. We propose a multidimensional depressive detection model to investigate how different modes of social interaction influence the user's interactions on social networks and evolution of lexicons. This is complete by comparing their emotional intelligence in social interactions. In this work, we describe three different use cases; those are mathematical model, lexical data of the psychological phenomenon and lexico-psychological factor. As opposed to the psychometric methods used to capture such factors, our model uses lexical features to compute the social media interactions. Our experiment results exhibit highly accurate detection of depression so that physicians are easily diagnosis the patients.

Keywords Depression · Lexical · Reviews · Social media · Rain forest · Model

1 Introduction

Many people are suffering by depression or over anxiety is a universal problem throughout the world. The characterization of mood disorder is mainly depending on unhappy, confusing, guilty feeling, inferiority complex and life threaten ideas like suicidal feelings and lack of concentration [1]. This disease can be enduring or recurrent, resulting considerable effect on people who are suffering by depression

V. Kakulapati (✉)
Sreenidhi Institute of Science and Technology, Hyderabad, India
e-mail: vldms@yahoo.com

S. Mahender Reddy
Otto Friedrich University of Bamberg, Bamberg, Germany
e-mail: mahendersheri@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_10

and also affected by their relations, society etc. [2]. Though, the efficient diagnosis of depression is limited by presently available investigative techniques, these methods are mainly based on physician expertise and patient cooperation with the physician, hazard a variety of objective preconceptions [3]. Contrasting variations in mood, medical depression is a general psychological muddle so as to long lasting which results the ailment and condensed performance. In addition, when a person suffered with the severe level of depression directed to suicide. According to WHO [4] reports more than 350 million public globally are influenced by depression. Depression is the main causes over two-thirds of suicides per year [5]. The percentage of suicidal cases of depressed patients is 30 times more than people without these disorders [6]. No two depressed people have similar types of symptoms or behaviours. Psychological physicians' diagnoses the depressed people by counselling frequently on referring to the common utilized investigative and Statistical approaches of psychological muddle criterion.

Today, people are habituated to pose their feelings and psychological disorders as well as their private status in social media like what's up, Twitter and Facebook. The social media sites are main resource of the user generated the contents instantaneously replicate not only their routine life, but also the psychological conditions of addicts. Researchers developed several approaches by utilizing the social media for objective and psychological healthiness investigations [7–9].

In this work we implement a lexical analysis of social media analysis by developing mathematical model. This can implement by applying Rain Forest algorithm for relatively fanciful reasons. There are many classification trees to select from, and they all grow fast in Rain-Forest algorithm, which removes the gap among the drawbacks to the main memory datasets of algorithms in the machine learning, statistics approaches and the scalability requirements.

2 Related Study

The analytical and statistical instruction booklet of psychological Disorders [10] establishes nine types of depressive signs, for example, discouraging feel and moderate significance. Without considering the attitude feature, reviews are taking from different social media is analysing the classification of depressed group reorganization [11]. By analysing depression concerned issues from social media users [12] perception of the online websites and natural language models. Through the prospective use [13] of online social networks for identify and analyse the most important psychological disorders in the human being.

Developments in identifying psychological conditions are acknowledged in the CLPsych (Computational Linguistics and Clinical Psychology) in 2016 distribute assignment [14] wherever post insertions [15] are utilized to verify the kind of severity of posts circulated by online users. For diagnosis of psychological illness, n-grams character, n-grams words, n-grams lemma of lexical analysis and POS n-grams, addictions of syntactic features, public attitude examples such as frequently pose comments in social networks and the rate of retweet and the personal details like age, gender, and behaviour [16] were also reflect as strong indicators. Researcher developed psychological disorder identification from the utilization of lexicon-based methods to natural language processing methods and latent Dirichlet topic methods. The recent studies have tried to improve models' performance with the use of vector space illustrations and RNN (recurrent neural network) layers with consideration [17] to identify and clarify comments describing predicament. In their study, execute a model which generates aggressive consequences for discovering psychological disorder of social media users.

Psychological emotions and societal features generally develop associations among offline societal communications or occurrences and gloominess [18] and happens in online communities, as self-respect reconciles directing associate communications and psychological indications [19], which is also describing the consumer utilize and affected by the social media. Meditation, an approach of response where a user preserves a reactive and repetitive centre of attention on their agony [20], which exhibits person anxiety existence occurrences and the growth of psychological disorder [21], and the online communities offers the prospect for a user to both mentally reflect on awful incidents and comprise the whole social media more draw attention to limitations. Moreover, the societal impact has been proven to sensible connections among anxiety and psychological disorder, this impact performing like a buffer to psychological disorder aspects [22]. From all these, is applicable to the social media as a probable prospective for increasing and accommodating online networks for susceptible consumers.

3 Our Proposed Model

Depression analysis: In this proposed work first use case is expanding mathematical model by utilizing rain forest algorithm to formalise the recognition of Psychological disorder based on the collection of features accumulated from the social networks especially twitter reviews. In the second phase is to model the emotional occurrence of apprehension, its origin, and persistence, etc. using lexical information accessible through online networks. Finally, the lexicon-emotional feature is confining the maximum of every user with certain psychological state.

Rain forest algorithm: this algorithm is an ensemble of including classification trees. This algorithm gives significant results on a number of realistic complications, as it is neither responsive to noise nor subject to overfitting in the data set. This algorithm in general shows, a considerable performance enhancement more than many other classifications of decision tree algorithms. This can construct by merging the predictions of numerous trees, each tree is prepared in segregation.

4 Result Analysis

The data set collecting from various comments and tweets from users in social media on various topics. Pre-processing dataset for removing unnecessary symbols, commas, # symbols, stop words etc. Processing the textual data in terms of keywords, i.e., the word counts of all the words and then extracting only depression words which are predefined by us. Calculating the word count of depression words in 5, 10, 25…. Then calculate the Poisson posterior distribution for depression words. Random variations of the word count to sentences Bayesian plots.

We found words in tweets like 1 sad 2 bore 3 depressed 4 lonely 5 discouraged 6 guilty 7 bad 8 ashamed 9 dissatisfied 10 miserable 11 terrible 12 lost 13 sorrow 14 unhappy 15 dull (Table 1; Figs. 1, 2 and 3).

Table 1 Tweeter comments are classified as words

S. no.	Sentence	Words
1	10	2
2	15	5
3	25	7
4	35	8
5	50	12
6	75	16

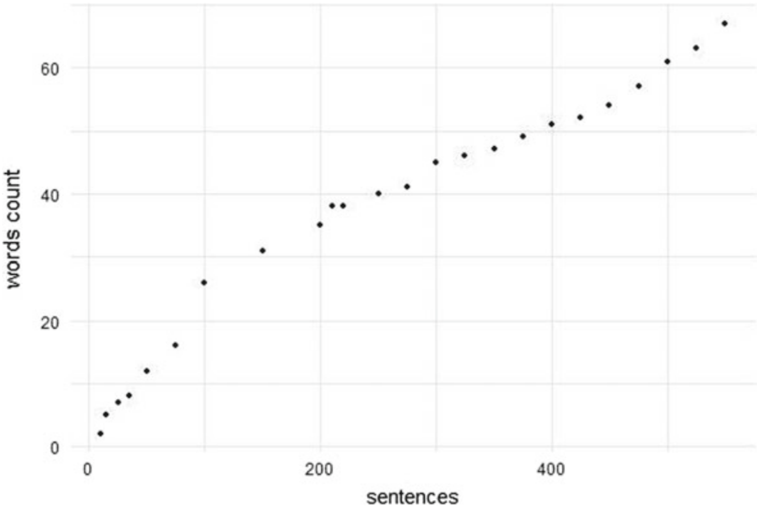


Fig. 1 Each sentence of tweeter comments and their word count

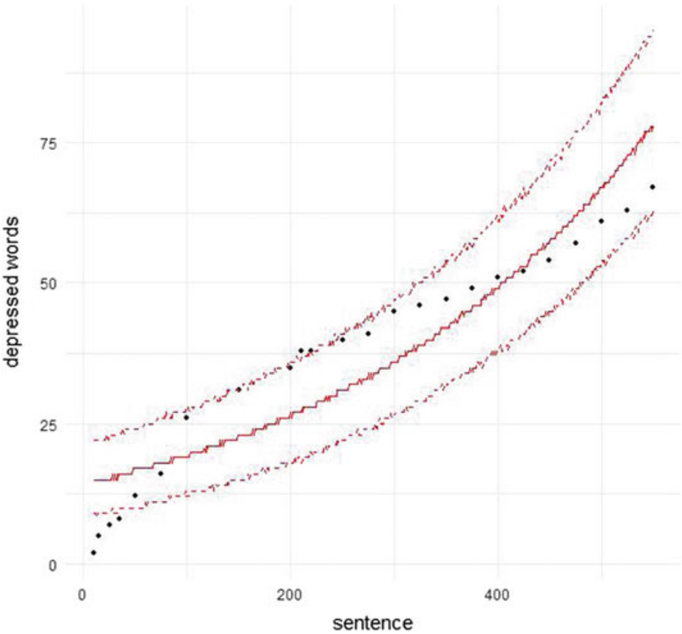


Fig. 2 Depressed words in sentences

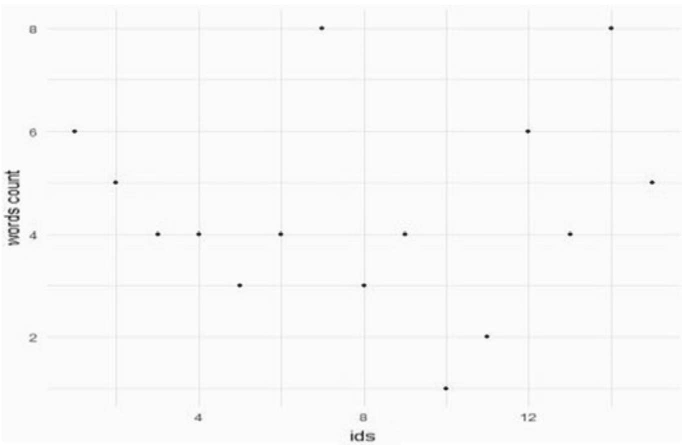


Fig. 3 The depression word count in review comments

Estimates					
Mean sd		10%	50%	90%	
(Intercept)	2.7	0.1	2.6	2.7	2.8
Sentence	0.0	0.0	0.0	0.0	0.0

(continued)

(continued)

mean_PPD	37.1	1.8	34.9	37.1	39.4
Log-posterior	−100.1	1.0	−101.4	−99.8	−99.1

Diagnostics:
mcse Rhat n_eff

(Intercept)	0.0	1.0	860
Sentence	0.0	1.0	1039
mean_PPD	0.0	1.0	1506
Log-posterior	0.0	1.0	806

id	Count
11	6
22	5
33	4
44	4
55	3
66	4

Model Info:
function: stan_glm

Formula	count ~ id				
Algorithm	Sampling				
Priors	see help('prior_summary')				
Sample	2000 (posterior sample size)				
Observations	15				
Predictors	2				
Estimates					
Mean sd			10%	50%	90%
(Intercept)	1.4	0.3	1.1	1.4	1.7
Id	0	0	0	0	0
mean_PPD	4.5	0.8	3.5	4.5	5.5
Log-posterior	−35.9	1.0	−37.2	−35.6	−35.1

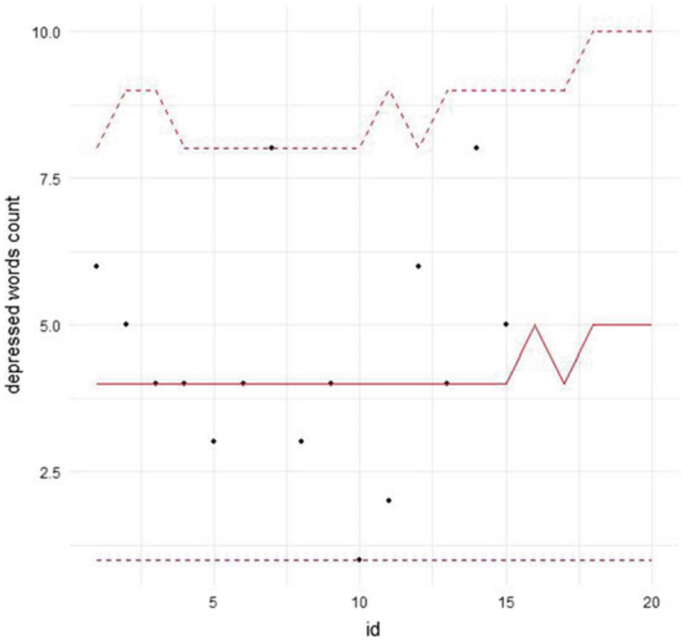


Fig. 4 Random forest analysis of depression words

Diagnostics:

	mcse	Rhat	n_eff
(Intercept)	0.0	1.0	873
id	0.0	1.0	917
mean_PPD	0.0	1.0	1470
Log-posterior	0.0	1.0	875

For every parameter, mcse is Monte Carlo standard error, n_eff is a crude measure of effective sample size, and Rhat is the potential scale reduction factor on split chains (at the convergence Rhat = 1) (Fig. 4).

5 Conclusion

In this work, analysed the involvement of the feature procedure and recognized psychological disorder users on an extensive social network dataset. In this work, implement the mathematical model of lexical analysis and applying rainforest algorithm which is an ensemble knowledge method that joins forecasts of diverse uncorrelated classification trees for retrieving depression words from the social media.

6 Future Enhancement

In the future, expanding research by applying similarity measures between user mental disorders and provide the corresponding depression aspects as recommendation explanations. Further depression might be pertinent to calculate whether or not the prediction performances along with twitter reviews are analyzing by various datasets.

References

1. Aaron TBMD, Brad AAPD (2014) Depression: causes and treatment. ClinGeriatr Med 14:765–786
2. Willner P (2016) The chronic mild stress (CMS) model of depression: history evaluation and usage. Neurobiol Stress 6:78
3. Mundt JC, Snyder PJ, Cannizzaro MS, Chappie K, Geralts DS (2007) Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. J Neurolinguist 20:50–64
4. Mathers C, Boerma J, Fat D (2008) The global burden of disease: 2004 update. WHO, Geneva, Switzerland
5. US Department of Health and Human Services (2000) Healthy people 2010: understanding and improving health, vol 2. US Government Printing Office, Washington, DC
6. Guze SB, Robins E (1970) Suicide and primary affective disorders. Br J Psychiatry 117(539):437–438
7. Coppersmith G, Dredze M, Harman C (2014) Quantifying mental health signals in Twitter. In: the workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality, pp 51–60
8. Lin H, Jia J, Nie L, Shen G, Chua TS (2016) What does social media say about your stress? In: Proceedings of the twenty-fifth international joint conference on artificial intelligence, pp 3775–3781
9. Akbari M, Hu X, Liqiang N, Chua TS (2016) From tweets to wellness: wellness event detection from twitter streams. In: Proceedings of the thirtieth AAAI conference on artificial intelligence, pp 87–93
10. Whooley O (2014) Diagnostic and statistical manual of mental disorders (DSM), Wiley
11. Nadeem M, Horn M, Coppersmith G, Sen S (2016) Identifying depression on Twitter. arXiv preprint 2016. arXiv:1607.07384
12. Park M, McDonald DW, Cha M (2013) Perception differences between the depressed and non-depressed users in Twitter. In: Proceedings of the international conference on weblogs and social media, pp 476–485
13. Choudhury et al. (2013) Predicting depression via social media. In: Proceedings of the international conference on weblogs and social media, pp 128–137
14. Milne DN, Pink G, Hachey B, Calvo RA (2016) CLPsych 2016 shared task: triaging content in online peer-support forums. In: Proceedings of the third workshop on computational linguistics and clinical psychology, Association for Computational Linguistics, San Diego, CA, USA, pp 118–127

15. Mac Kim S, Wang Y, Wan S (2016) Data61-CSIRO systems at the CLPsych 2016 shared task. In: Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology, vol 1. Association for Computational Linguistics, San Diego, CA, USA, pp 128–132
16. Preotiuc-Pietro D, Eichstaedt J, Park G, Sap M, Smith L, Tobolsky V, Schwartz HA, Ungar L (2015a) The role of personality, age and gender in tweeting about mental illnesses. In: Proceedings of the workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality, pp 21–30
17. Kshirsagar R, Morris R, Bowman S (2017) Detecting and explaining crisis. In: Proceedings of the fourth workshop on computational linguistics and clinical psychology—from linguistic signal to clinical reality, Association for Computational Linguistics, Vancouver, pp 66–73
18. Cole DA, Turner JE Jr (1993) Models of cognitive mediation and moderation in child depression. J Abnorm Psychol 102(2):271–281
19. Bosacki S, Dane A, Marini Z, YLC-CURA (2007) Peer relationships and internalizing problems in adolescents: mediating role of self-esteem. Emotional Behav Difficulties 12(4):261–282. https://doi.org/10.1080/13632750701664293
20. Nolen-Hoeksema S, Wisco BE, Lyubomirsky S (2008) Rethinking rumination. Perspect Psychol Sci 3(5):400–424. https://doi.org/10.1111/j.1745-6924.2008.00088.x
21. Michl LC, McLaughlin KA, Shepherd K, Nolen-Hoeksema S (2013) Rumination as a mechanism linking stressful life events to symptoms of depression and anxiety: longitudinal evidence in early adolescents and adults. J Abnorm Psychol 122(2):339–352. https://doi.org/10.1037/a0031994
22. Wang X, Cai L, Qian J, Peng J (2014) Social support moderates stress effects on depression. Int J Ment Health Syst. 8(1):41. https://doi.org/10.1186/1752-4458-8-41

Application of Neural Networks and Lifting Wavelet Transform for Long Term Solar Radiation Power Prediction

Manju Khanna, N. K. Srinath and J. K. Mendiratta

Abstract In the past research scholars, have tried to predict solar radiation for a short duration for power prediction. However, with increase in applications requiring predicted data for solar radiation need for long duration prediction has increased. Present work carried out by application of concurrent Neural Networks and lifting wavelet transform for long-term power prediction. A brief review of Neural Networks and second-generation wavelets carried out for their application in long-term solar radiation power generation. Algorithm developed and experiments conducted with results presented for long-term prediction of solar radiation. Results produced present quite accurate prediction for a duration up to three months.

Keywords Lifting Wavelet Transform · Multi-resolution analysis · Riesz basis · NARX neural network

1 Introduction

Solar Radiation data is required in varied uses, i.e. atmospheric energy balance studies, analysis of solar load of buildings, planning for operation of renewable power plants, metrology and environmental impact analysis [1–4]. The requirement for solar radiation data has become essential due to increase in number of solar energy utilizations.

M. Khanna (✉) · N. K. Srinath
RVCE, Bangalore, India
e-mail: manjukhanna1977@gmail.com

N. K. Srinath
e-mail: srinathnk@rvce.edu.in

M. Khanna
MVJCE, Bangalore, India

J. K. Mendiratta
Life Time Senior Member IEEE, Bangalore, India
e-mail: mendiratta.jk@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_11

In the past, various computer models were developed involving very large computational analysis with empirical relations. Selection among these models two features are to be considered

(1) Meteorological and other kind of data used in the model.
(2) Accuracy of model.

With exponential growth in the renewable energies, need of precise modeling, prediction and forecasting of solar radiation has increased. Prediction of solar radiation only for a few hours, say from 12–24 h is not sufficient. In practice, solar radiation is available as a time series and so difficult to predict from underlying models for generating the required data. Artificial intelligence, involving neural network, fuzzy logic and hybrid systems have been applied for modeling and forecasting from time series [5]. Many scholars have applied recurrent neural networks (RNNs) to utilize the statistical properties of time series data [6–8]. Capizzi et al. have applied second generation wavelets along with recurrent neural network to predict solar radiation by using observed meteorological data, for few days' prediction [9].

As has been mentioned above, most of the researches have carried out solar radiation prediction limited to a day or to the maximum of a few days, resulting in short time prediction, which is not sufficient to size the grid and energy storage devices for load scheduling for a long durations. Present work utilizes Recurrent Neural Networks and Lifting Wavelet Transform for long-time prediction.

2 Overview of Basics

Brief overview of basics required for long term solar radiation prediction is as below:

2.1 Recurrent Neural Networks

A neural network is a machine to perform like a human brain. It can perform in a linear or non-linear mode and learn by supervisor with training samples. Further, it is adaptive in nature thereby adapting to the environment changes. Mathematically a neuron 'k' is expressed by [10]. The neural networks can be in single or multi-layer configurations. Single layer network comprises of an input and an output layer. The multi-layer feed Forward network comprises of input, output and hidden layers.

The recurrent networks consist of an additional feedback loop, which may be either a hidden layer or an output layer. For temporal processing in neural networks, Time Delay Neural Networks (TDNN), with multi-layer feed forward perceptrons are most commonly used. With back propagation, the network comprises the output with desired or target response at each delay instant. The output $y_j(n)$ and desired response $d_j(n)$ at time 'n', produce instantaneous sum of square error:

$$E(n) = \frac{1}{2}\sum_{j} e_j^2(n) \quad (1)$$

Index ‘j’ being a neuron in the output layer and $e_j(n)$, the error signal defined by:

$$e_j(n) = (d_j(n) - y_j(n)) \quad (2)$$

With objective of minimizing, a cost function computed over total time:

$$E_{total} = \sum_{n} E(n) \quad (3)$$

A recurrent neural network maps the input to an output space. Such networks respond temporally to an externally applied input signal. These networks with single input, when applied to a tapped delay line with a memory of ‘q’ units, produce an output ahead of input by one time unit. Therefore, a model with present value of input u(n), produce an output y(n + 1) with unit delay, and called a Non-linear Auto Regressive with exogenous inputs (NARX) with mathematical representation:

$$\mathrm{y(n+1) = F(y(n)\ldots y(n-q+1), u(n)\ldots y(n-q+1))} \quad (4)$$

where F(.) being a non-linear function of its arguments. A dynamic network with a tapped delay at its input is called Focused Time Delay Neural Network (FTDNN) without back propagation and is well suited for time series prediction. Such a network trains faster than other dynamic networks. For multi-step ahead prediction, the network is configured in the form of a closed loop. However, if each time a neural network is trained, results a new solution due to different initial weights and bias values and with different training, testing and validation data.

2.2 Wavelet Transforms

They can be thought as data building blocks, allowing representation efficiently with fast computing, through dyadic translation and dilation of a function: $\psi_{j,m(x)} = \psi\left(2^j(x-m)\right)$. named as **first generation wavelets** [7], with properties:

(i) Create a Riesz basis in $L_2(R)$ with a variety of function spaces $\mathcal{F}$.
(ii) The coordinate functional $\bar{\Psi}_{j,m}$, denoted by: $\gamma_{j,m} = \bar{\Psi}_{j,m}(f)$, generates orthogonal or duals (bi-orthogonal) wavelets.
(iii) They and their duals generate vanishing polynomial moments in local space and generate frequency, which decays towards slow frequencies.
(iv) Create multi-resolution analysis, leading to fast wavelet transform, allowing passing the function ‘f’ along with its coefficients $\gamma_{j,m}$ in a linear time.

Above features have been exploited in varied applications [9].

By generalization, these wavelets not only translate and dilate, but also provide:

(i) Bases solution to differential and integral equations not only on R^n, but also on arbitrary, non-smooth domains of R^n also.
(ii) First generation wavelets provide bases for spaces with translation invariants, whereas second generation wavelets make them suitable for differential functions and analysis on curved surfaces with weighted approximation.
(iii) They also provide solution to many algorithms adapted to irregular sampled data.

So, the generalization not only preserves the properties of first generation wavelets, but also above properties, and therefore is named **second generation wavelets**.

2.3 Construction of Second Generation Wavelets (Lifting Scheme) [7]

Several schemes for construction of wavelets have been developed: for functions on interval [11–14], bounded domains [15], spline for irregular samples [16, 17], and weighted wavelets. Lifting scheme provides a simple construction technique for second generation wavelets. A multi-resolution analysis with working upwards is applied to achieve desired properties of custom designed filters. However, a check is required on these filters for generating functions with a stable basis for smoothness of the wavelets.

2.4 Multi-resolution Analysis (MRA) [18]

For closed subspaces, a wavelet generates direct sum of decomposition in $L_2(R)$:

$$V_j = \ldots \dot{+} W_{j-2} \dot{+} W_{j-1}, \quad j \in \mathrm{Z} \text{ in } L_2(\mathrm{R}) \tag{5}$$

$\dot{+}$ Indicates "direct sums" of the components. For an orthogonal function 'f' its decomposition is "mutually orthogonal" with the subspaces:

(1) $\ldots \subset V_{-1} \subset V_0 \subset V_1 \ldots$

(2) $clos L_2(\underset{j \in Z}{\cup} V_j) = L_2(R)$

(3) $\underset{j \in Z}{\cap} V_j = \{0\};$

(4) For V_j $j \in Z$, comprising of a Riesz basis, provides a scaling functions $\{\varphi_{j,k} \mid k \in K(j)\}$ with K(j) as a general set of index.

With the above properties, if a single function subspace V_0, generated by $\phi \in L_2(R)$ with:

$$V_0 = clos_{L_2(R)}\langle \phi_{0,k} : k \in Z \rangle \tag{6}$$

For a function $\phi_{j,k}(x) \equiv 2^{j/2}\,\phi(2^j x - k)$, subspaces V_j are generated by ϕ itself, just as W_j are generated by a function ψ $W_j \equiv clos_{L_2(R)}\langle \psi_{j,k} : k \in Z \rangle$ with duals of $\phi\ as\ \bar{\phi}\ and\ \ of\ \psi\ as\ \bar{\psi}$ satisfying:

$$\langle \psi_{j,k}\ \bar{\psi}_{l,m} \rangle = \delta_{j,l}\,.\,\delta_{k,m} \quad j, k, l, m \in Z$$

From above analysis, we define:

Definition 1 A function $\phi \in L_2(R)$ generates a MRA, producing a nested sequence of closed subspaces V_j, satisfying properties (1), (2), (3) and (4), such that $\{\phi_{0,k}\}$ forms a Riesz basis of $V_{0.}$

To meet the above definition, two variables A and B are assumed such that $0 < A \le B < \infty$, with

$$A\|\{c_k\}\|_{l^2}^2 \le \left\| \sum_{k=-\infty}^{\infty} c_k \phi_{0,k} \right\|_2^2 \le B\|\{c_k\}\|_{l^2}^2 \tag{7}$$

For all bi-infinite square sum, able sequences

$$\{c_k\},\ \text{i.e.}\ \|\{c_k\}\|_{l^2}^2 = \sum_{k=-\infty}^{\infty} |c_k|^2 < \infty \tag{8}$$

If a MRA is generated by a function ‘φ’ it is defined as a “scaling function”.

3 Construction of Lifting Wavelet Transform

For construction of these wavelets, two things are required: (i) a set of partitioning and (ii) a filter

A finite filter in context of MRA is defined [17, 19]:

$$\{h_{j,k,l} \mid j \in Z, k \in K(j), l \in K(j+1)\} \tag{9}$$

If:

(i) for each j and k, coefficients $h_{j,k}$, there are a non zero finite number thereby making set

$$L(j,k) = \{l \in K(j+1) \, | h_{j,k,l} \neq 0\} \text{ a finite}$$

(ii) Similarly for j and l, coefficients $h_{j,k,l}$ are a non-zero finite number making a finite set

$$\mathrm{K}(j,l) = \{k \in \mathrm{K}(j) \, | h_{j,k,l} \neq 0\},$$

(iii) With $L(j,l)$ and K(j, l) uniformly bounded for all j, k, l.

Taking a signal's of a finite sequence L, $a_0 = (a_{0,n})$ (n = 0, 1, …, L−1), characterizing the signal.

The sequences $a_1 = (a_{1,n})$ and $d_1 = (d_{1,n})$, are defined:

$$a_{(m+1),n} = \sum_{k=2n}^{2n=N} h_{k-2n} a_{mk} \tag{10}$$

Also for

$$d_{(m+1)n} = \sum_{k=2n+1-N}^{2n+1} (-1)^k h_{2n+1-k} a_{mk} \tag{11}$$

From above equations, with m = 0, $a_{(m+1),n}$ will consist of half the length of a_0, with decomposition:

$$a_{0n} = \sum_{k=-\infty}^{\infty} \left[h_{n-2k} a_{1k} + (-1)^n h_{2k+1-n} d_{1k} \right] \tag{12}$$

a_1 and d_1 can be reconstructed precisely for every $n \in \mathrm{Z}$ [19].

For $p = (p_n)$, being a finite sequence, energy $\|p\|^2$ is defined by:

$$\|p\|^2 = \sum_{n=-\infty}^{\infty} p_n^2$$

Considering 'p' as a vector, with a square root $\|p\|$ and energy $\|p\|^2$, $\|p\|_2$ is called a Euclidean norm.

Thereby, the sequence a_0, a_1, d_1 will satisfy the equation:

$$\|a_0\|^2 = \|a_1\|^2 + \|d_1\|^2 \tag{13}$$

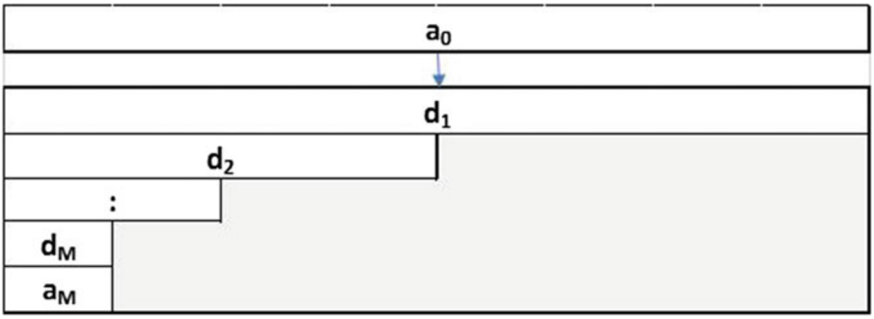


Fig. 1 LWT of a sequence a_0

Recurring application of above equations shall generates a decomposition of the a0 into a matrix of sequences d_1, d_2, …, d_M, a_M as shown in Fig. 1.

These sequences possess the same amount of information as a_0:

$$\|a_0\|^2 = \|a_M\|^2 + \sum_{m=1}^{M} \|d_m\|^2 \tag{14}$$

In the above process, wavelet coefficients of the signal 's', in the above matrix are reduced by a power of '2' at each level. Thus, depending upon the size of the signal samples, feature extraction of the signal is carried out. The 'a_M' coefficients representing approximate of signal features, while various 'd_M' components representing details of the signal features.

$$u_k = \sum_{j=1}^{m} w_{kj} x_j \tag{15}$$

$$y_k = \phi(u_k + b_k) \tag{16}$$

4 Experiment

4.1 Solar Radiation Prediction with Neural Networks

Data required for long duration prediction for solar radiation has been taken from NREL-USA site, which has been recorded at an interval of 5 min with a reset time of 24 h and has been recorded for complete year.

4.2 Modeling of NARX Neural Network

For problem in hand, different NARX models with various hidden layers were tried and the one with 10 hidden layer was considered suitable for the problem. Input data taken for the experiment, raw data as given in above site alternative, featured data by use of Lifting Wavelet Transform on raw data was taken and output with corresponding predicted data is considered for NARX neural network. To deal with small to medium data set, Levenberg-Marquardt (LM) algorithm is used for training, testing and validation with 50, 45 and 5% of data required for the experiment.

4.3 Results

Figures 2 and 3 show prediction results using raw data for the month of February, with output of predicted results for the month of March. Similarly featured data found by application of Lifting Wavelet Transform for the month of February and

Fig. 2 Actual data plot

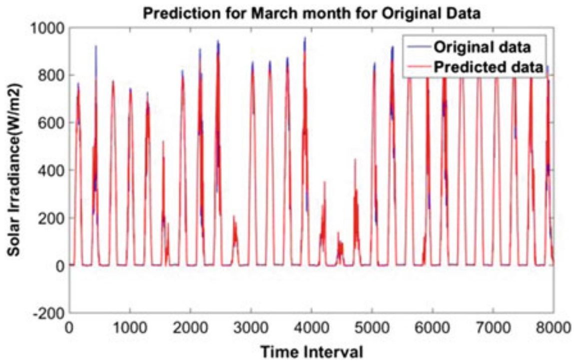
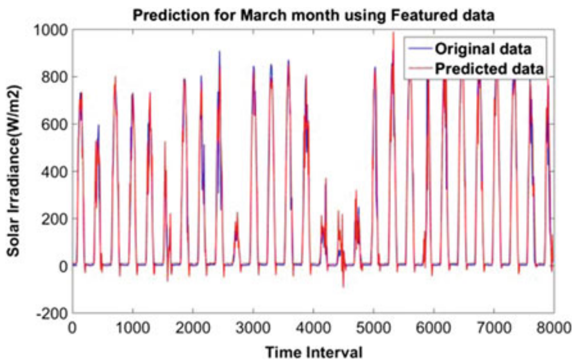


Fig. 3 Featured data plot



predicted output for the month of March. Both the output results for both data type plotted in the figures mentioned.

Above plots shows similarity in plots for prediction with original data as well as with featured data. Advantages of feature extraction using Lifting Wavelet Transform requires less resources along with reduction in computation time for prediction. Following Table 1 show comparative analysis of the error performance with both the methods.

Above Table 1 shows the performance of one- month prediction. Though with featured extracted data performance in root means square error is more compared to actual data, but with reduction in data to be handled from 4000 to 500 makes a lot of difference in computer resources required and computation time. Similarly, prediction has been carried out for three months, with plots shown in Figs. 4 and 5.

From Figs. 4 and 5, we find prediction for three months also has not much difference, except one spike in radiation, which may be due some error in recording. This spike is not seen in the featured data, as the data get filtered during feature extraction. Table 2 provides quantitative performance of the process of solar radiation prediction for a duration of three months, with featured and without featured data.

Table 1 Comparison of performance with actual and featured data-one month

S. No.	Operations	RMSE error	% samples
1	Training	42.041	50
2	Testing	41.238	45
3	Validation		5
4	Overall	42.078	
5	Training	62.316	50
6	Testing	74.515	45
7	Validation		5
8	Overall	67.775	

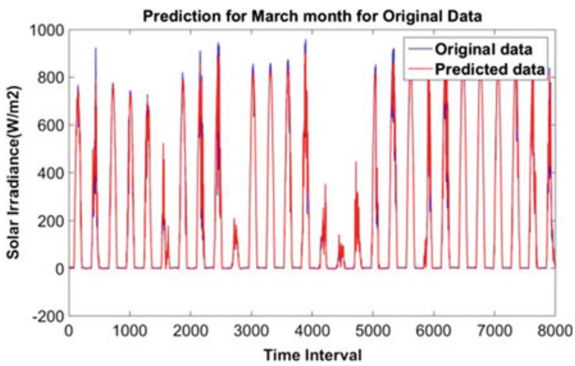


Fig. 4 Actual data plot-3 months

Fig. 5 Predicted data plot-3 months

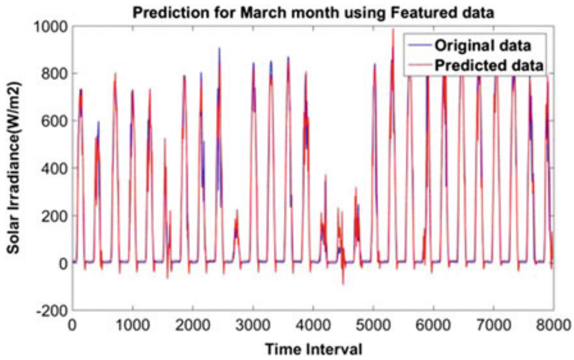


Table 2 Comparison of performance with actual and featured data-three months

S. No.	Operations	RMSE error	% samples
1	Training	67.753	50
2	Testing	231.060	45
3	Validation		5
4	Overall	163.085	
5	Training	116.514	50
6	Testing	138.166	45
7	Validation		5
8	Overall	127.799	

In this case also, the performance with featured data is poor compared to non-featured data, but with drastic reduction of memory requirement and so the reduction in computational time, featured technique is preferred.

5 Conclusion

From above analysis of lifting wavelet Transform and neural networks, along with results obtained from the experiment, it is concluded, that with featured data the system resource requirement along with computational effort reduces drastically. Specially, if the prediction is carried out for longer durations, performance degradation is negligible compared to reduction in resource requirement and computational effort.

References

1. Wong LT, Chow WK (2001) Solar radiation model. Appl Energy 69:191–224
2. Muneer T, Younes S, Munawwar S (2005) Discourses on solar radiation modeling. Renew Sustain Energy Rev 11:551–602
3. Mellit A, Kalogirou SA (2008) Artificial intelligence techniques for photovoltaic applications: a review. Progr Energy Combust Sci 34:547–632
4. Sen Z (2008) Solar energy fundamentals and modeling techniques: atmosphere, environment, climate change and renewable energy. Springer, Berlin
5. Palit AK, Popovic D (2005) Computational intelligence in time series forecasting: theory and engineering applications. Springer
6. Yan W (2012) Toward automatic time series forecasting using neural networks. IEEE Trans Neural Netw Learn Syst 23(7):1028–1039
7. Qi M, Zhang GP (2008) Trend time series modeling and forecasting—with neural networks. IEEE Trans Neural Netw 19(5):808–816
8. Borges RV, d'Avila Garez A, Lamb LC (2011) Learning and representing temporal knowledge in recurrent networks. IEEE Trans Neural Netw 22(12):2409–2421
9. Capizzi G, Napoli C, Bonanno F (2012) Innovative second generation wavelets construction with recurrent neural networks for solar radiation forecasting. IEEE Trans Neural Netw Learn Syst 23(11):1805–1815
10. Haykin S (2005) Neural networks-a comprehensive foundation, 2nd edn, Pearson Education
11. Anderson L, Hall N, Jawerth B, Peters G (1993) Wavelets on closed subsets of the real line. In: Recent advances in wavelet analysis, Academic press, New York, pp 1–61
12. Auscher P (1992) Wavelets with boundary conditions on the interval. In: Chui CK (ed) Wavelets: a tutorial in theory and applications, Academic Press, San Diego, pp 217–236
13. Chui C, Quak E (1992) Wavelets on a bounded interval. In: Bracess D, Schumaker LI (eds) Numerical methods of approximation theory, Birkhauser-Verlag, Basel, pp 1–24
14. Cohen A, Daubechies I, Jawerth B, Vial P (1993) Multi-resolution analysis, wavelets and fast algorithms on an interval. C. R. Academic Science. Paris Ser, I. Math, I(316): 417–421
15. Cohen A, Daubechies I, Vial P (1993) Multi-resolution Analysis, wavelets and fast algorithms on an interval. J Appl Comput Harmonic Anal I(1):54–81
16. Cohen A, Dahmen W, DeVore R Multi scale decompositions on bounded domains, Preprint University of South Carolina
17. Buhmann MD, Micchelli CA (1992) Spline pre-wavelets for non-uniform knots. Numer Math 61:455–474
18. Dehmen W, Micchelli CA (1993) Banded matrices with banded inverses II: locally finite decompositions of the spline spaces. Constr Approx 9(2–3):263–281
19. Daubechies I (1988) Ortho-normal bases of compactly supported wavelets. Comm Pure Appl Math 41:909–996

Modeling Sustainable Mobility Using Fuzzy Logic: An Application to Selected Indian States

Ubaid Illahi and Mohammad Shafi Mir

Abstract Modeling sustainable mobility is a challenging task among the researchers, planners and evaluators. Indicator-based approaches are, however, meritorious, owing to the communicative benefits among the diverse interdisciplinary group of people such as researchers, planners, evaluators, and stakeholders. The main aim of the paper is to develop a sustainable mobility model using a combination of statistical and machine learning tools. A set of sustainable mobility indicators, categorized into three dimensions of sustainability i.e. Environmental, Social, and Economic are developed. Then, the indicators are normalized, weighted and aggregated by Factor Analysis. The sub-indices corresponding to the three dimensions are then aggregated using Fuzzy Logic. This model was applied to 16 Indian states and 1 UT, which were shortlisted based on the data availability. Based on the computed Fuzzy Sustainable Mobility Index (I_{FSM}), these states and UT were ranked. Tamil Nadu, Telangana, and Andhra Pradesh were the best performing states with computed I_{FSM} of 0.727, 0.726, and 0.726 respectively. Whereas, Odisha, Uttarakhand, and Haryana with computed I_{FSM} of 0.578, 0.555, and 0.513 respectively were the least performing states.

Keywords Sustainable mobility · Indicators · Factor analysis · Fuzzy logic · Composite indicator/index

1 Introduction

Mobility, on one hand, plays a major role in achieving sustainable development goals. On the other hand, it can hinder in orientating towards sustainability, if not planned, designed, operated, and maintained carefully and efficiently. To do so, it is necessary to quantify and model what we call as sustainable mobility. This, however, is not an

U. Illahi (✉) · M. S. Mir
Civil Engineering Department, National Institute of Technology, Srinagar, Jammu & Kashmir 190006, India
e-mail: ubaidillahi@nitsri.net

M. S. Mir
e-mail: shafi@nitsri.net

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_12

easy task and still remains a challenge among the researchers, planners and evaluators [1, 2]. Sustainable mobility is becoming more and more vast, inter-disciplinary and complex. So there is a dire need to adopt a multi-dimensional approach to solve problems related to it.

The main aim of this paper is to develop a sustainable mobility model which would generate a composite indicator i.e. Fuzzy Sustainable Mobility Index (I_{FSM}) using various statistical and machine learning tools. In this paper, factor analysis is used as a statistical tool and fuzzy logic as a machine learning tool. The model so developed is applied to the selected Indian states which are then ranked on the basis of the computed I_{FSM}. To validate the fuzzy sustainable mobility model, the ranking was again done by means of an index developed by Factor Analysis.

The paper is structured in the following manner; Firstly, the model development is shown by the hierarchical diagram. This is followed by introducing sustainable mobility indicators (SMIs) and their development. Next, the methods of normalization, weighting, and aggregation by means of statistical tools have been presented. This is followed by the integration of sub-indices by Fuzzy Logic in MATLAB. Then, some light is thrown on how the study areas were shortlisted and how the data was collected. Lastly, a brief and succinct discussion on the application of the developed model followed by some takeaways is presented.

2 Methodology

The methodological approach for modeling sustainable mobility has been presented by means of a flow chart (see Fig. 1).

2.1 *Development of Sustainable Mobility Indicators*

Sustainable Mobility Indicator (SMI) is a variable which can either be a simple determinant or a complex metric developed with the help of various mathematical/statistical models [3]. The main aim to develop each SMI is to represent an important characteristic of sustainable mobility. The framework to design and develop SMIs should be given utmost importance as it forms the building blocks for achieving the sustainable mobility goal [4]. The selection of suitable SMIs, however, remains a challenging task owing to the diversity and vast area of concern [5]. Researchers have debated this topic in depth and have suggested some robust frameworks [6]. However, it is important to understand that no framework can be accepted wholesale everywhere [3]. It needs to be customized keeping the strengths and weaknesses of the concerned organization in mind. Moreover, the goal-oriented approach should be given due consideration while customizing the frameworks for sustainable mobility.

In this study, firstly the determinants were obtained from the databases [7–11]. Secondly, the SMIs were developed. A total number of 12 SMIs, each categorized

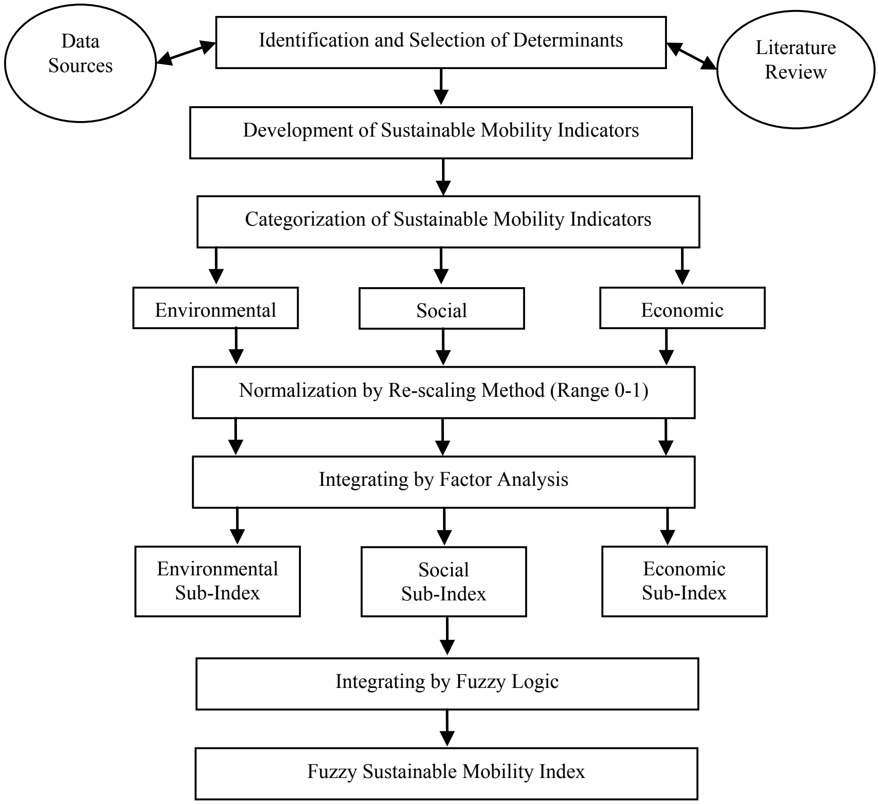


Fig. 1 Methodology adopted to develop sustainable mobility model

into three different dimensions i.e. environmental, social and economic, were finally used in the analysis. The developed SMIs used in the analysis, corresponding to the respective dimension and effect on sustainable mobility have been shown in Table 1.

2.2 Normalization, Weighting and Mathematical Integration

The selected SMIs do not have the same units of measurement. Moreover, the range of the values obtained is large. To analyze the data, thus, it is important to normalize it. Different normalization techniques are available in the literature. Comparative analysis of these techniques is done by the researchers [12]. Taking these guidelines into consideration, the re-scaling method of normalization (range 0, 1) has been used in this research. Depending upon whether the SMI has a positive or negative effect on sustainable mobility, normalized indicators (NI) have been calculated using Eq. (1).

Table 1 Developed sustainable mobility indicators

Dimension	Indicator	Effect
Environmental	Annual average air pollution[a] by transportation	–
	Annual average GHG[b] emissions by transportation per capita	–
	Annual average GHG emissions by transportation per area	–
	Per capita sales of petroleum products in transportation	–
	SRTU's fleet fuel efficiency	+
	Total road length per annual million VKT	–
Social	An road traffic fatalities per million VKT	–
	Total number of registered vehicles per capita	+
	Annual NMT fatalities per total number of licensed drivers	–
	SRTUs passengers carried per bus per day	+
Economic	Annual transportation revenues per GSDP	+
	Annual transportation expenditures per GSDP	–

Note GHG = greenhouse gas; *SRTUs* = state road transport undertakings; *VKT* = vehicle kilometres travelled; *GSDP* = gross state domestic product
[a]includes SO_2, PM_{10} and NO_2
[b]includes CO_2, CH_4 and N_2O

$$\boldsymbol{NI} = \begin{cases} \frac{x_i - x_{min}}{x_{max} - x_{min}} \textbf{ if the effect is positive} \\ \frac{x_{max} - x_i}{x_{max} - x_{min}} \textbf{ if the effect is negative} \end{cases} \tag{1}$$

where, $\boldsymbol{x_{min}}$ and $\boldsymbol{x_{max}}$ are the minimum and maximum values of the indicator $\boldsymbol{i}$.

In the index development, the next step is to assign a weight to each SMI followed by aggregation into a composite indicator. A number of methods of weighting are available in the literature, each with pros and cons [13]. Statistical/mathematical weighting approaches are often advantageous over methods like equal weighting and Analytical Hierarchy Process (AHP) methods. Therefore, factor analysis has been used in this research (refer to Eqs. 2–5).

$$\boldsymbol{w_i} = \frac{\left(\boldsymbol{FactorLoding_i}\right)^2}{\boldsymbol{Eigenvalue_i}} \tag{2}$$

$$\boldsymbol{FI_c} = \sum_i \boldsymbol{w_i NI_i} \tag{3}$$

$$\boldsymbol{\alpha_i} = \frac{\boldsymbol{Eigenvalue_i}}{\sum \boldsymbol{Eigenvalue_i}} \tag{4}$$

$$\boldsymbol{I_{cx}} = \sum_i \boldsymbol{\alpha_i FI_c} \tag{5}$$

where $\boldsymbol{w_i}$ is the weight assigned to the indicator $\boldsymbol{i}$; $\boldsymbol{\alpha_i}$ is the weight assigned to the composite factor index $\boldsymbol{FI_c}$; $\boldsymbol{I_{cx}}$ is composite sub-index corresponding to Environment/Social/Economic dimension.

2.3 Integrating the Sub-indices with Fuzzy Logic

Three sub-indices designated as Environmental sub-index (I_{ES}), Social sub-index (I_{SS}) and Economic sub-index (I_{CS}) developed in the previous step are used as input whereas an aggregated index is the output in the fuzzy logic designer as shown in Fig. 2. Next, five-member functions (VL—Very Low, L—Low, M—Moderate, H—High and VH—Very High) are defined for input as well as the output variables (refer to Figs. 3 and 4). This is followed by adding rules to the model. A total of 10 rules were added in the Rule Editor. Then, the data was fed in the form of [I_{ES}; I_{SS}; I_{CS}] as input to the model which generated the output. Finally, three 3D surface plots were obtained showing the relationship of two sub-indices with the output.

Fig. 2 Fuzzy logic designer for sustainable mobility index

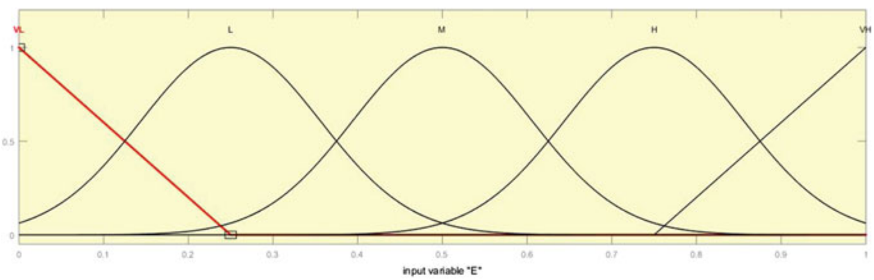


Fig. 3 Input membership functions

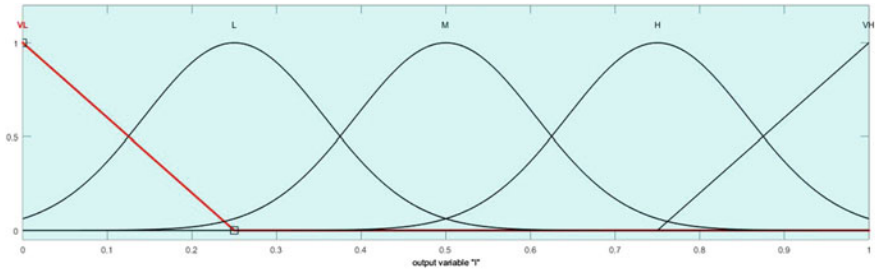


Fig. 4 Output membership functions

3 Study Areas and Data Collection

To apply the developed fuzzy sustainable mobility model, initially all the 29 states and 7 Union Territories (UTs) of India were taken into consideration. However, due to data constraints, only 16 states which include Andhra Pradesh, Goa, Gujarat, Haryana, Jammu and Kashmir, Karnataka, Kerala, Maharashtra, Meghalaya, Mizoram, Odisha, Rajasthan, Tamil Nadu, Telangana, Uttar Pradesh, Uttarakhand, and 1 UT i.e. Delhi were finally shortlisted.

The latest data for all the selected indicators was available for the year 2015–2016 which was obtained from the following sources: MORTH, NHAI, MOP&NG, CPCB and OGDP [7–11].

4 Results and Discussions

Developed fuzzy sustainable mobility model is validated with a statistical tool (Factor Analysis) using SPSS. The results showed that the value(s) of the Fuzzy Sustainable Mobility Index (I_{FSM}) so developed varied but uniformly across all the states and as a result the ranking of the states was not affected. To rank the states with respect to each other, the computed I_{FSM} for each state was plotted in a horizontal bar diagram as shown in Fig. 5. As far as assessing and evaluating sustainable mobility is concerned, the higher value of I_{FSM} indicates better performance of the state and vice versa (Table 2).

5 Conclusion

A sustainable mobility model was developed with a combination of statistical tool called Factor Analysis (in SPSS) and Machine Learning tool called Fuzzy Logic (in MATLAB). The model was applied to the SMIs data corresponding to 16 Indian states and 1 UT, which were shortlisted based on the data availability. It was found that

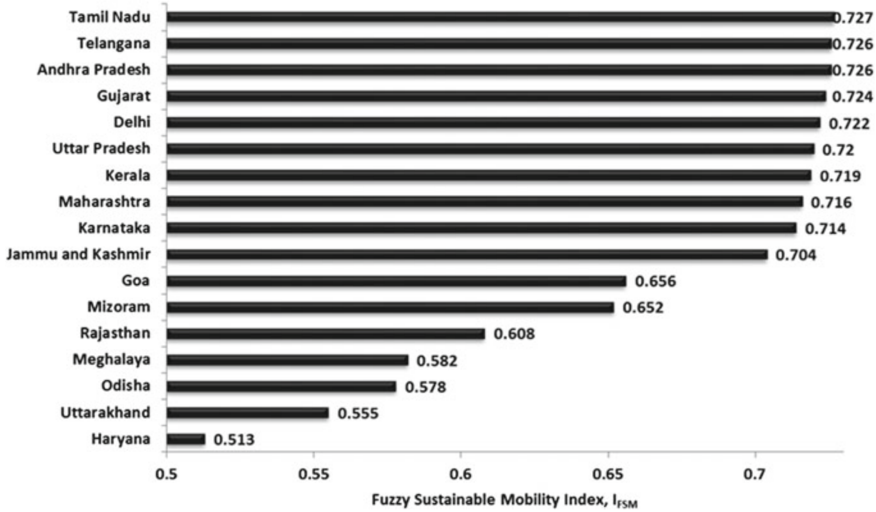


Fig. 5 Ranking the study areas based on the computed fuzzy sustainable mobility index

Table 2 Ranking the states and UT based on fuzzy sustainable mobility index

State/UT	I_{ES}	I_{SS}	I_{CS}	I_{FSM}	Rank
Andhra Pradesh	0.8676	0.5059	0.4872	0.726	2
Delhi	0.5227	0.8962	0.4906	0.722	5
Goa	0.2722	0.5518	0.4832	0.656	11
Gujarat	0.7431	0.6564	0.4895	0.724	4
Haryana	0.6053	0.4504	0.4452	0.513	17
Jammu and Kashmir	0.8727	0.4108	0.4701	0.704	10
Karnataka	0.8279	0.6104	0.4903	0.714	9
Kerala	0.8045	0.6390	0.4853	0.719	7
Maharashtra	0.7386	0.6166	0.4901	0.716	8
Meghalaya	0.7493	0.4041	0.4868	0.582	14
Mizoram	0.7909	0.4734	0.4900	0.652	12
Odisha	0.7408	0.3166	0.4903	0.578	15
Rajasthan	0.7634	0.3915	0.4897	0.608	13
Tamil Nadu	0.8483	0.7153	0.4895	0.727	1
Telangana	0.8859	0.7077	0.4878	0.726	2
Uttar Pradesh	0.8850	0.3858	0.4894	0.72	6
Uttarakhand	0.7271	0.4075	0.4842	0.555	16

Note I_{ES} = environmental sub-index; I_{SS} = social sub-index; I_{CS} = economic sub-index; I_{FSM} = fuzzy sustainable mobility index

Tamil Nadu topped the ranking with a computed Fuzzy Sustainable Mobility Index (I_{FSM}) score of 0.727 followed by Telangana and Andhra Pradesh with computed I_{FSM} of 0.726 each. The least scores were obtained for Odisha, Uttarakhand, and Haryana with computed I_{FSM} of 0.578, 0.555, and 0.513 respectively.

It is suggested that the model be used with a bigger SMI set covering the dimensions of sustainable mobility in a broader way.

References

1. Ramani TL, Zietsman J (2016) Sustainable transportation—alternative perspectives and enduring challenges. Int J Urban Sci 20:318–333. https://doi.org/10.1080/12265934.2016.1217784
2. Clevenger CM, Ozbek ME, Simpson SP, Atadero R (2016) Challenges in developing a transportation sustainability rating system that meets the preferences of a department of transportation. J Transp Eng 142:1–7. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000830
3. Gudmundsson H, Hall RP, Marsden G, Zietsman J (2016) Sustainable transportation: indicators, frameworks and performance management, Springer Texts in Business and Economics Henrik
4. Pei Y, Barrella E, Meyer M, Ross C, Amekudzi A (2010) Performance measurement frameworks and development of effective sustainable transport strategies and indicators. Transp Res Rec J Transp Res Board 2163:73–80. https://doi.org/10.3141/2163-08
5. Castillo H, Pitfield DE (2010) ELASTIC—a methodological framework for identifying and selecting sustainable transport indicators. Transp Res Part D 15:179–188. https://doi.org/10.1016/j.trd.2009.09.002
6. Rassafi AA, Vaziri M (2005) Sustainable transport indicators: definition and integration. Int J Environ Sci Tech 2:83–96. https://doi.org/10.1007/BF03325861
7. Ministry of Petroleum and Natural Gas, Government of India. http://petroleum.nic.in/
8. Ministry of Road Transport & Highways, Government of India. http://morth.nic.in/
9. Open Government Data Platform India. https://data.gov.in/
10. National Highways Authority of India, Ministry of Road Transport & amp; Highways, Government of India. http://www.nhai.org/
11. Central Pollution Control Board, Ministry of Environment, Forest and Climate Change, Government of India. http://cpcb.nic.in/
12. Reisi Varzaneh M (2014) Quantifying transport sustainability: development of an index. http://hdl.handle.net/11343/40789
13. Guo J, Zhao Y, Wilson M, Fernandez IC, Zhou B, Gan X, Wu J (2017) When to use what: methods for weighting and aggregating sustainability indicators. Ecol Indic 81:491–502. https://doi.org/10.1016/j.ecolind.2017.05.068

Shravya—A Framework for Enunciation of Sanskrit Shlokas and Phrases

Apoorva Maheshwari, Eravatee Raje, Mihir Karkare and Shubham Joshi

Abstract Through this paper we are proposing a framework for enunciation of Sanskrit shlokas and phrases. Sanskrit shlokas have been known to have a cathartic effect on the mind and soul. But due to the richness of the language and the lack of a large number of teachers, few are able to tap its benefits. We propose a new method of sandhi splitting based on the shloka characteristics and the utilization of the seq2seq model with attention for generating shlokas. On testing, the model generates an MOS of 3.25 out of 4.

Keywords Sanskrit · Sandhi Splitting · seq2seq · Tacotron · TTS

1 Introduction

Sanskrit is one of the oldest languages in the world with elaborate oral and written specifications. This makes it rather difficult for a person to recite shlokas. In order to understand where to pause while enunciation of the shlokas, a number of rules have been defined, which have been passed on as shruti (oral tradition). A great deal of work has been done to split the words through various implementations of Sandhi Vicched—a set of rules enumerated in Panini's Astadhayayi.

Hartzell et al. [1] undertook a study on Professional Vedic Sanskrit Pandits in India training from childhood in an ancient, formalized tradition of oral Sanskrit text memorization and recitation, mastering the exact pronunciation and invariant content of multiple 40,000–100,000 word oral texts to analyse the effect of Sanskrit shlokas in the brain. The MRI scans showed massive gray matter density and cortical thickness increases in Pandit brains in language, memory and visual systems.

Sanskrit being a fairly rich language has a large number of rules which make *Sandhi Vicched* or sandhi splitting more difficult and time consuming. Both rule-based and more recently neural networks based approaches have been developed but the accuracy is not a hundred per cent. Utilizing any of these algorithms into the framework would be futile as they take a large amount of time and data to train

A. Maheshwari (✉) · E. Raje · M. Karkare · S. Joshi
Pune Institute of Computer Technology, Pune, India
e-mail: apoorva19.am@gmail.com; sankrutikshravya@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_13

and provide the desired outcome. Further, we suspect that the transliteration of the words leads to further ambiguity and hence inaccuracy. Moreover, as shlokas follow a specific tune, *Sandhi Vicched* for them become simpler and a simple rule-based approach proved sufficient.

The task of enunciation aligns with the task of text-to-speech synthesis given certain rules about where to take pauses in speech. In order to produce a cathartic effect, it's important for the enunciation to have a natural sound, exactly similar to the human voice. Although, there have been a lot of recent advancements in TTS, it is very challenging to produce a sound that is able to capture and retain human qualities. We seek to use a TTS model that is most suitable for providing rich conditioning on language attributes.

The framework we propose strives to eliminate the challenges faced in *sandhi vicched* and provide near human pronunciation of Sanskrit shlokas.

The paper has been organized as follows. In Sect. 2, we discuss the literature survey for Shravya and also identify the usability and the drawbacks. In Sect. 3 we introduce Shravya. Section 4 concludes the paper.

2 Literature Survey

As there is no consolidated research done in developing such a framework, a number of avenues had to be explored such as Natural Language Processing on Sanskrit, Sandhi splitting, melody (template) generation and text-to-speech deep learning models. Finally, we have divided our research into two sections—the *Sandhi Vicched* module and the enunciation module.

2.1 Sandhi Vicched

Sandhi Vicched can be considered as creating a parser but that would be an oversimplification as there are a number of rules to consider. Due to a large number of permutations, many new methods have been developed. Some are discussed below.

Krishna [2] pass the sentence through Sanskrit Heritage Reader which provides all the possible segmentations for the input sentence along with morphological information for each of the segments. A graph is formed by taking the morphological segments as nodes. Each node has three attributes—the word-form, the lemma and the POS tag. Path-constrained random walks are carried out. The weights are defined as the co-occurrence probability of the attribute value at the target node given the attribute value at the source node. It utilizes the grammar rules to identify the paths that can be drawn.

Another interesting approach was developed by Natarajan [3] in their paper Statistical Samdhi Splitting. The paper tries to upgrade the current algorithm for Samdhi splitting by proposing a principled modification of the posterior probability function.

The paper also proposes a Bayesian Word-Segmentation method. The authors have upgraded the current algorithm by expressing Maximum a posteriori (MAP) through the noisy channel model framework. As the probability of obtaining the valid morphological constituents given that a particular rule is applied is 1, MAP is expressed in the terms of the probability of generating the morphs.

The flexibility of neural networks has been utilised for Sandhi Vicched as well. Hellwig [4] in his paper Using Recurrent Neural Networks for joint compound splitting and Sandhi resolution in Sanskrit interprets Sandhi and compound resolution as a sequence labelling task. The paper defines Paninian prescriptions for inter-word Sandhi in 5 rules. The paper further utilizes a recurrent neural network. The size of the input layer is equal to the number of unique phonemes and the number of output layer nodes is equal to the target classes. LSTM cells are used in the hidden layer because they are less susceptible to vanishing gradient.

Aralikatte et al. [5] in their paper Sanskrit Sandhi Splitting using seq2(seq)^2 have discussed that the problem with the rule-based system of Astadhayayi is that is brute force and hence even with complete knowledge of all the rules it is difficult to identify the position of splitting in a given compound word. The authors propose a seq2(seq)^2 model. A simple RNN proved to be inefficient as it encoded on the context of the preceding characters before the split. The most important part of learning to split compound words is to correctly identify the locations of the split. Hence the authors propose to use two decoders in the bi-directional encoder-decoder model. The first decoder is the location decoder which identifies the place of inserting the split. The second decoder performs the character decoding and generates the Sandhi output. The model is estimated to be better than a simple RNN, a bidirectional RNN and a bidirectional RNN with attention. It gives an accuracy of 95 and 79.5% for location prediction and split prediction.

2.2 Enunciation—Shloka Generation

Most recently, there has been a lot of work in tackling the problem of text-to-speech synthesis (TTS). After the release of Wavenet [6], which introduced a new generative model operating directly on the raw audio waveform to produce human-like voice, there have been many advancements in synthesizing speech with neural networks.

Wang et al. [7] introduced Tacotron, a neural text-to-spectrogram conversion model, that synthesizes speech directly from characters and uses Griffin-Lim Algorithm for spectrogram-to-waveform synthesis. Shen et al. [8] further made improvements by implementing Tacotron 2, a system composed of a recurrent sequence-to-sequence feature prediction network that maps character embeddings to mel-scale spectrograms, followed by a modified WaveNet model acting as a vocoder to synthesize time-domain waveforms from those spectrograms. This system can be trained directly from data without relying on complex feature engineering and achieves state-of-the-art sound quality close to that of natural human speech.

DEEP VOICE 3 [9] highlights a fully-convolutional character-to-spectrogram architecture, which enables fully parallel computation and trains an order of magnitude faster than analogous architectures using recurrent cells. The architecture is capable of multi-speaker speech synthesis by using trainable speaker embeddings, a technique described in DEEP VOICE 2 [10].

3 Shravya

See Fig. 1.

3.1 Sandhi Vicched Module

Sanskrit belongs to the inflectional category. Sandhi rules dictate the morphophonemic changes at the morpheme or word boundary in terms of alteration to the sounds due to the neighbouring sounds or due to the morphological behaviour of adjacent words.

The shlokas in Anushthup format follow strict melody rules. Hence this leads to the concatenation of words, which makes reading the shlokas very difficult. Consider the following lines from shlokas:

पश्यञ्शृण्वन्स्पृशञ्जिघ्रन्नश्नन्गच्छन्स्वपञ्श्वसन् ।
प्रलपन्विसृजन्गृह्णन्नुन्मिषन्निमिषन्नपि ।

These are the lines from *Bhagvad Gita, Pancham Adhyaya*. As seen above, without *Sandhi Vicched*, the above lines can't be read. There are a number of ways for performing Sandhi Vicched such as Statistical Sandhi Splitting by A Natarajan et al.,

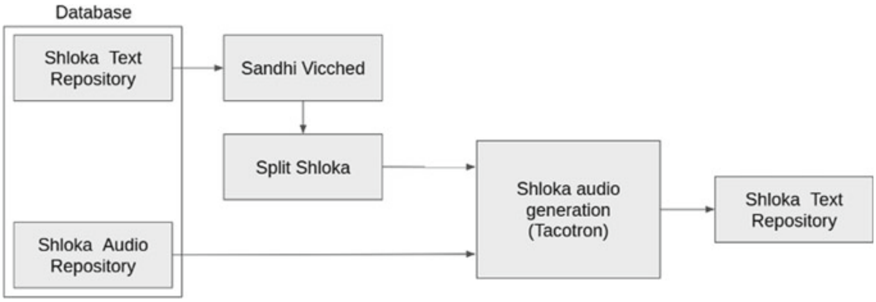


Fig. 1 Architecture of Shravya

Using Recurrent Neural Networks for joint compound splitting and Sandhi resolution in Sanskrit by O Hellwig, Sanskrit Sandhi Splitting using seq2seq by A Sankaran et al., and Word Segmentation in Sanskrit using Path Constrained Random Walks by A Krishna et al. Due to the constant nature of the shlokas (16 diphones in each line of the shloka), certain rules can be drawn out, which help in Sandhi Vicched.

Hence after passing the shloka through the Sandhi Vicched module, the shloka would look like as follows:

पश्यञ्शृण्वन्स्पृशञ्जिघ्रन् - नश्नन्गच्छन्स्वपञ्श्वसन् ।
प्रलपन्विसृजन्गृह्णन् - नुन्मिषन्निमिषन्नपि ।

Sandhi can take place between vowel and vowel, vowel and semivowel, semivowel and semivowel, consonant and consonant and between visarga and other sounds. Sanskrit characters can be directly manipulated with the help of the open source Classical Language Toolkit (cltk).

The algorithm is as follows:

```
program Sandhi Vicched (split_shloka)
   Data: input_shloka, sandhi_length
   Result: split_shloka
   Tokenize shloka;
   var        position_marker := 0;
              overflow_marker := 0;
   begin
      while word in shloka do
        if word_length and overflow_marker > sandhi_length then
        calculate position of split;
                 set poisiton_marker and overflow_marker;
                 while overflow_marker do
                          split word further;
                 end
        end
      end
    end.
```

3.2 Shloka Audio Generation

Our model is based on Tacotron (Wang et al.), a recently proposed state-of-the-art end-to-end speech synthesis model that predicts mel spectrograms directly from grapheme or phoneme sequences. The predicted mel spectrograms can either be synthesized directly to the time-domain via a WaveNet vocoder [8], or by first learning a linear spectrogram prediction network, and then applying Griffin-Lim spectrogram inversion [11].

3.3 Training

Generating Shloka Text Repository: To train the model, we tried multiple methods. Initially, we tried to train the model to work with the Devanagari Script (Sanskrit's script). This did not give optimum results as Devanagari Script has a large character set as compared to the Roman Script.

For the next test, we tried to work with the International Phonetic Alphabet (IPA). This also did not produce satisfactory results due to a large character set.

Finally, we tried to transliterate the Devanagari Script into the Roman Script using the itrans library. We received maximum accuracy with transliterated shloka text as input.

Generating Shloka Audio Repository: As there was no data available in Sanskrit, we had to create and augment our own data set to train the model to fit our requirements. We recorded eight hours of shlokas from various scriptures namely, The Bhagvad Geeta, The Durga Saptashati, etc.

3.4 Esting

The output generated after the model is trained for 62,000 steps has clear pronunciation and produces a human-like sound indistinguishable to the human ear (Fig. 2).

4 Conclusion

Please A number of methods can be used for *Sandhi Vicched* but the rule-based method for *Anushthup*, gives a high accuracy of approximately 95%. For the *shloka* generation system, one may experiment with giving input in transliterated format, *Devanagari* script or International Phoneme Alphabet. Though transliterated format may give higher efficiency due to few numbers of unique characters.

5 Future Scope

We intend to extend this framework for other *shloka* types such as *Chatushpad* and *Vedas*. Further research and analysis needs to carried out in what would be better for catering to a large number of *shloka* types: rule based or deep learning based.

Acknowledgements We would like to thank our mentor Dr. S. C. Dharmadhikari for her guidance and encouragement. We would also like to thank our mentors from C-DAC ACTS, Pune for their

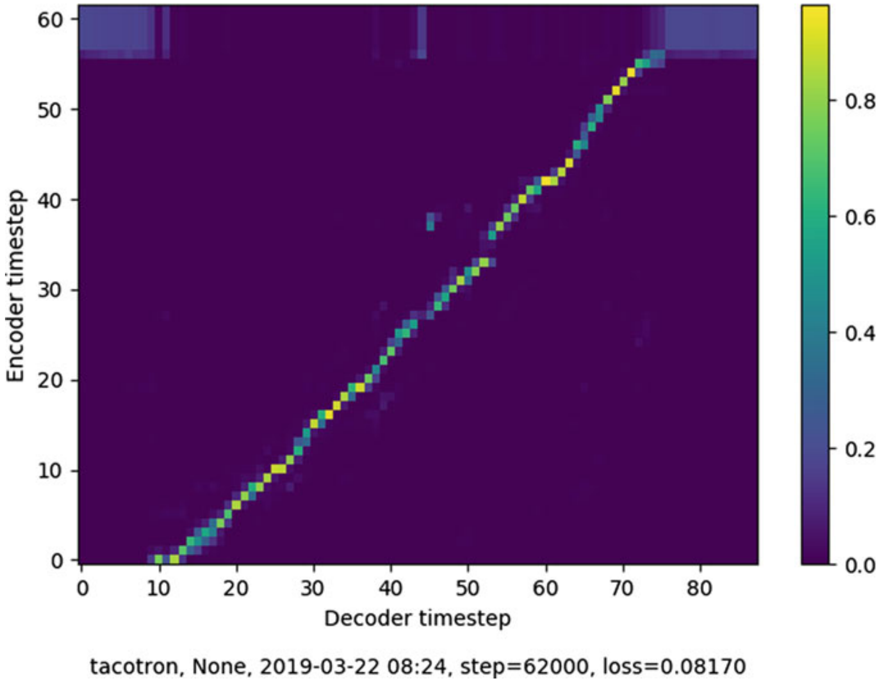


Fig. 2 Graph after 62,000 steps

mentorship. We would like to extend our gratitude to Dr. Ujjwala Jha, Center of Advanced Learning for Sanskrit and Prakrit, Pune for her support in evaluating the results.

References

1. Hartzell JF, Davis B, Melcher D, Miceli G, Jovicich J, Nath T, Hasson U (2016) Brains of verbal memory specialists show anatomical differences in language, memory and visual systems. NeuroImage 131:181–192
2. Krishna A. Word segmentation in sanskrit using path constrained random walks
3. Natarajan A. Statistical Sandhi splitting
4. Hellwig O. Using recurrent neural networks for joint compound splitting and Sandhi resolution in Sanskrit
5. Aralikatte R, Gantayat N, Panwar N, Sankaran A, Mani S (2018) Sanskrit Sandhi splitting using seq2(seq)^2. CoRR
6. Oord AVD, Dieleman S, Zen H, Simonyan K, Vinyals O, Graves A, Kalchbrenner N, Senior A, Kavukcuoglu K (2016) Wavenet: a generative model for raw audio. arXiv:1609.03499
7. Wang Y, Skerry-Ryan R, Stanton D, Wu Y, Weiss RJ, Jaitly N, Yang Z, Xiao Y, Chen Z, Bengio S, Le Q, Agiomyrgiannakis Y, Clark R, Saurous RA (2017) Tacotron: towards end-to-end speech synthesis. In: Proceedings Interspeech, pp 4006–4010

8. Shen J, Pang R, Weiss RJ, Schuster M, Jaitly N, Yang Z, Chen Z, Zhang Y, Wang Y, Skerry-Ryan RJ, Saurous RA, Agiomyrgiannakis Y, Wu Y (2017) Natural TTS synthesis by conditioning wavenet on mel spectrogram predictions. arXiv preprint. arXiv:1712.05884
9. Ping W, Peng K, Gibiansky A, Arik SÖ, Kannan A, Narang S, Raiman J, Miller J (2017) Deep voice 3: 2000-speaker neural text-to-speech. CoRR, vol. abs/1710.07654
10. Arık S, Diamos G, Gibiansky A, Miller J, Peng K, Ping W, Raiman J, Zhou Y (2017) Deep voice 2: multi-speaker neural text-to-speech. NIPS
11. Griffin D, Lim J (1984) Signal estimation from modified short-time Fourier transform. IEEE Trans Acoust, Speech, Signal Process 32(2):236–243

Energy-Efficiency in WSN Based on SCMC Protocol

Rajkumar Krishnan, Ganeshkumar Perumal, M. Anandaraj, K. Selvaraj and A. Sangeetha

Abstract In Wireless Sensor Networks (WSNs) the QoS is considered to be an important factor which is based on the performance of end-to-end delay, energy-efficiency and packet delivery ratio. In WSN, the single path routing mechanism faces a lot of problems in providing a better communication channel, as there is no option for alternative path in case of path failure. This problem can be overcome by implementing multipath routing mechanism with well designed protocol. In this paper, it is intended to propose Single cluster with Multi-hop Communication (SCMC) protocol to achieve better QoS by means of improved bandwidth, packet delivery ratio as well as end-to-end delay. It is a Multi-hop Communication Protocol with a single cluster head that balances the energy level and achieves high energy conservation with increased lifetime. The single cluster forms dynamic multi-paths with multi-members for multi-hop communications. This work is implemented in Ns-2 and compared with ECMP and SPEED protocols. The result shows that the proposed protocol is more effective in terms of minimum delay, low energy consumption and increased delivery ratio than the other two protocols.

Keywords Sensor networks · Single cluster with multi-hop communication protocol · Quality of service · Single path routing · Alternative path

1 Introduction

A WSN comprises of an arrangement of scaled down minimal effort sensors [1]. The sensors fill in as remote information obtaining gadgets for all the effective performing artist hubs that procedure the sensor readings and set forward a proper action [2]. All sensor systems are portrayed by the necessity for vitality productivity,

R. Krishnan (✉) · G. Perumal · M. Anandaraj · K. Selvaraj · A. Sangeetha
PSNA College of Engineering and Technology, Dindigul, Tamilnadu, India
e-mail: rajkumardgl@psnacet.edu.in

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_14

adaptability and adaptation to non-critical failure [3]. The fundamental idea inside topology administration is the capacity to keep up a completely associated network constantly in the middle of the lifetime of WSN [4].

This paper is organized as follows: Sect. 2 discusses the literature survey with various related protocols. Section 3 is about the proposed mechanism and the reason for designing multipath routing protocols for cluster-based as well as event-based WSN and also depicts the new multipath routing algorithm. Section 4 presents and evaluates the performance of the proposed work by comparing it with the existing protocol by means of graphical result. Finally, Sect. 5 provides the conclusion of the paper.

2 Related Work

The hardware based protocol design was implemented to solve energy problem. It was not suitable for large scale networks [1]. The hybrid protocol was used for energy efficiency. But it was very less dynamic [2]. Some of the QoS based routing systems were discussed in [3]. The earlier routing protocol enabling QoS on those days was Sequential Assignment Routing (SAR) protocol [5].

Advantage of cluster based QoS aware routing protocol based on queuing model is its capacity of handling both real time as well as non real time traffics [4]. But, it only concentrated on end-to-end delay which connects links as per the cost functions.

SPEED4 gives guaranteed solution to real-time end-to-end delay issues. The packets must be delivered within the time limit and compute end-to-end delay by allocating the total distance as per the packet delivering speed of the nodes [6]. The problem is it causes jamming when the network is overcrowded.

By varying delivery speeds, Multi-path and Multi-Speed Routing Protocol (MMSPEED) for achieving QoS was developed. Optimum reliability cannot be achieved in this protocol [7]. Multi constrained QoS multi-path routing (MCMP) protocols apply best routes from the available path for delivering the packets to the destination nodes [8].

ECMP protocol dealt with limited hops and minimum energy consumption by achieving implementation of path which satisfies QoS requirement with low energy consumptions [9, 10].

LEACH overcomes energy hole problem but suffers delay issues as the cluster heads were far from the base stations [11]. Protocol HEED8, in which the cluster head was selected by combining the relationship between residual energy and reference energy in HEED [12, 13].

The environment based CACH protocol concentrated only on network traffic and load balancing among the cluster headers [14]. NSN routing algorithm focuses only on energy consumption and not QoS issues [15]. EEHC protocol was implemented only with 2 level hierarchies [16]. QEMPAR protocol increased the network life time by considering end-to-end delay and energy consumption only [17].

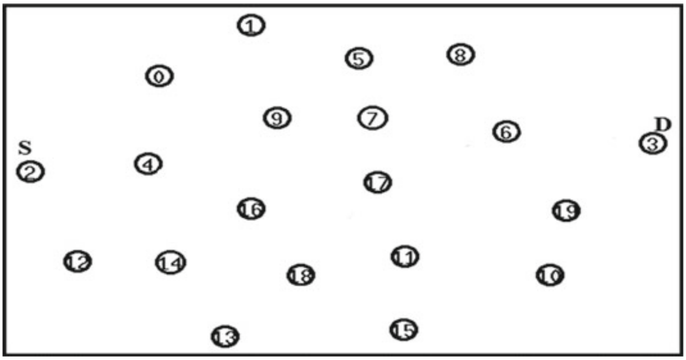


Fig. 1 Sensor node deployment dynamic structure

3 Proposed Work

3.1 Proposed System Architecture

Figure 1 shows the architecture of proposed work. Sensors are smaller in size as per the applications [3]. Mostly it uses gateway node or sink node for gathering data. It is provided with a huge number of densely deployed sensor nodes [5].

Improper node deployment maximizes the complexities of WSN. Another important issue in WSN is energy consumption as energy optimization increases the lifetime of WSN [4].

3.2 Proposed Design Goal

In this paper, a novel Single cluster with Multi-hop Communication (SCMC) protocol is proposed to achieve QoS by means of improved bandwidth, PDR as well as end-to-end delay. Our proposed mechanism states that the cluster head remains the same for all nodes which are connected randomly. This single dynamic cluster head will perform for all nodes according to their energy levels. The source node will be the cluster head which allocates data to other nodes as per the energy level. The cluster head balances the energy level with all individuals within its transmission range. Once the single cluster head is identified, cluster transmission of data takes place dynamically from all the other nodes to the cluster head. The proposed mechanism achieves high energy conservation with increased lifetime. The best thing is that at regular intervals, the energy level of cluster nodes are reassessed and compared with the remaining nodes forming a dynamic cluster. In order to attain better performance and successful transmission, the nodes with higher energy levels are employed for communication which can conserve energy for future use also. The following section discusses the formation of the proposed mechanism in a detailed manner.

3.3 Dynamic Cluster Formation

It is a single cluster formation with dynamic multi-paths. At any cause, any path failure means an automatic multipath will be created dynamically. In the proposed work, the source node is always the cluster head and at anytime, dynamically the other nodes form a new cluster path and accept the source node as CH (cluster head).This is the formation of single cluster head with multimembers for multi-hop communications. In this formation, if any path fails, it automatically chooses an alternative path to achieve minimum end-to-end delay. It will avoid congestion.

Figure 2 shows that by means of the proposed Multipath Cluster with Dynamic Mode Routing Protocol, a cluster is formed among 2, 12, 16, 7, 6 and 3 in which node 2 is the source node and 3 is the destination node. The data gets forwarded from one node to another as shown in the order.

Figure 3 shows the selection of alternative source path. When a cluster fails, automatically a dynamic cluster is formed among the nodes 4, 16, 11, 6 and 19 which have the same source node 2 and destination node 3. As shown in Fig. 3, when the cluster time elapses, it dynamically forms a new cluster

Multipath Cluster with Dynamic Mode Routing Protocol (MCDMRP)

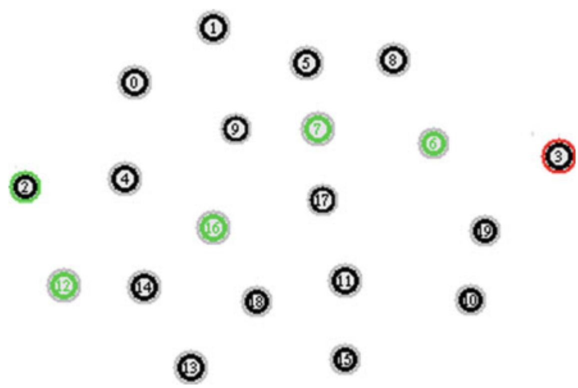
`Set of nodes initialize` N_{th}

$$for(N_{th} = 0; N_{th} < n; N_{th} + +)$$

`Set threshold of cluster` $N_{th} \rightarrow CH[i]//$ high energy nodes act as cluster head

`If` $CH[i] = 1; //$ group members busy, choose alternative cluster

Fig. 2 Dynamic multi cluster 12, 16, 7, 6



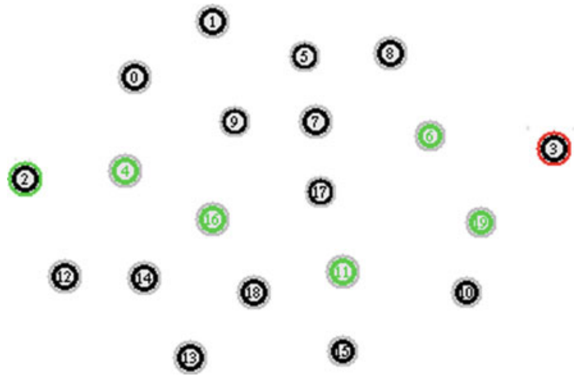


Fig. 3 Dynamic multi cluster 4, 16, 11, 6, 19

`else if` $CH[j] = 0;//$ group is available with adequate energy level for all nodes to communicate
`If` $CH[j] = 1;//$ group members busy, choose alternative cluster Path Selection process initiated
`for` $(P_{th} = CH_{i,j})$
`If` $CH[i] = 1;//$ all paths are busy, choose alternative path.
`else if` $CH[j] = 0;//$ this path is free from the cluster
`If` $CH[j] = 1;//$ this path is busy, choose alternative path
`end if`
`end`

4 Experimental Results

The graphical result of Fig. 4 shows that the proposed protocol SCMC is compared with ECMP and SPEED in the aspect of data rate with packets sent per second. Here, the total coverage area is 1500 * 1500 m^2 with 1000 packets. The total number of nodes involved in transmission is 24. The simulation time is 15.0 ms. Based on these parameters, our proposed algorithm is far better in performance. It is denoted in red color and the other protocols performance are shown in green and blue accordingly with their data rates.

Figure 5 shows the energy consumption among the number of nodes after completing the packets transmission. These 24 nodes carry out packet transmission using each protocol until the packets reach their destination successfully. Among these three protocols, the proposed SCMC protocol provides better performance in terms of minimum energy.

Figure 6 is the comparison of end-to-end delay to number of packets for MCDMRP, ECMP and SPEED protocols in which dynamic cluster mechanism shows

Table 1 Parameters and values used for simulation

Parameter	Value
Channel Type	Channel/WirelessChannel
Propagation Model	Propagation/TwoRayGround
Network Interface Type	Phy/WirelessPhy
MAC Type	Mac/802_11
Interface Queue Type	Queue/DropTail/PriQueue
Link Layer Type	LL
Antenna	Antenna/OmniAntenna
Area (M * M)	1500 * 1500
Maximum Packet Size	1000
Number of Mobile node	24
Routing Protocol	SCMC
Simulation time	15.0 ms

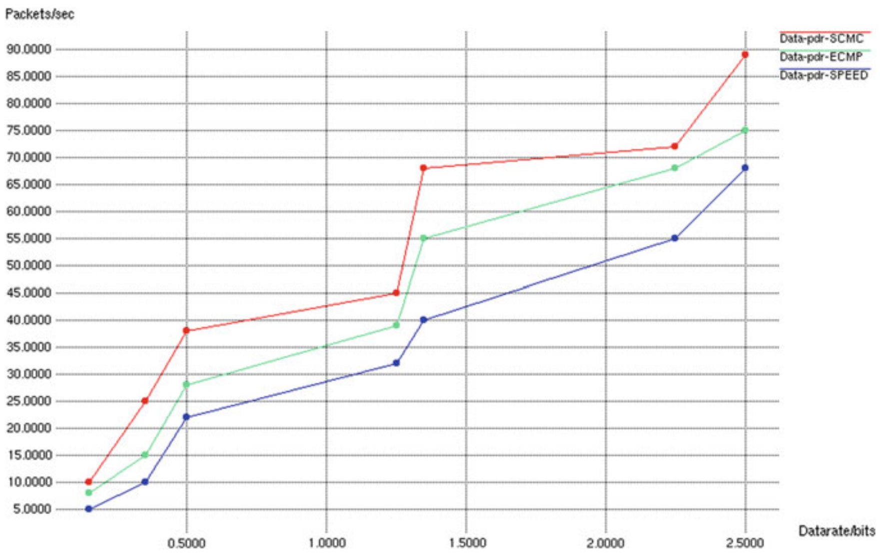


Fig. 4 Data rate/bits versus packets/sec

minimum delay. The other two protocols increase delay as the number of packets increase. From the above graph, it is clearly known that the proposed protocol is more efficient than ECMP and SPEED.

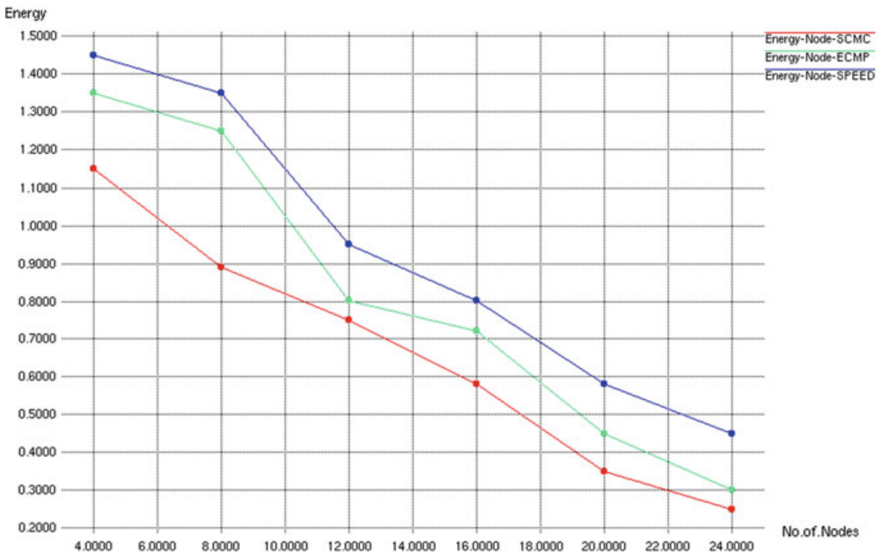


Fig. 5 No. of nodes versus energy

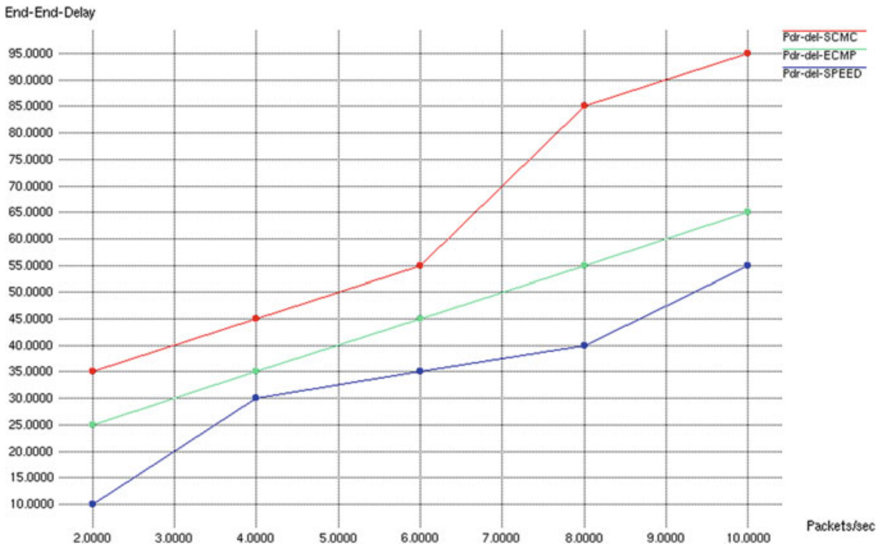


Fig. 6 Packets/sec versus end-to-end delay

5 Conclusion

This paper depicts the process of developing a unique protocol which processes source link as cluster head and establishes Multi-hop Communication Protocol by forming multiple paths dynamically. If any path or link fails, automatically dynamic paths are developed till communication reaches its end. In this path arrangement, the source node will be the cluster head which allocates the data to the other nodes as per the energy level. The proposed mechanism ensures the dynamic creation of an automatic multipath in case of path failure. In order to prove the effectiveness, the performance of the proposed scheme is compared with ECMP and SPEED in several aspects such as end-to- end delay, throughput and packet delivery ratio. In all aspects, the attained results are graphically expressed which prove that SCMC is unique and far better than ECMP and SPEED in overall performance. Thus, the proposed work provides comparatively better outcomes in QoS for the Wireless sensor networks.

References

1. Wang T, Zhang G, Yang X, Vajdi A (2016) A trusted and energy efficient approach for cluster-based wireless sensor networks. Int J Distrib Sens N 1–13. https://doi.org/10.1155/2016/3815834. Hindawi Publishing Corporation
2. Priya B, Manohar SS (2016) Life time enhancement of cluster based wireless sensor network through energy efficient MAC protocol. Circ Syst 7:2296–2308 Scientific Research Publishing
3. Chen D, Varshney PK (2004) QoS support in wireless sensor networks: a survey. In: International conference on wireless networks (ICWN), vol 1. CSREA Press, USA, pp 1–7
4. Akkaya K, Younis M (2003) An energy aware QoS routing protocol for wireless sensor networks. In: IEEE workshop on mobile and wireless networks (MWN), Providence, pp 1–20
5. Sohrabi K, Pottie J (2000) Protocols for self-organization of a wireless sensor network. IEEE Pers Commun 7(5):16–27
6. He T, Stankovic JA, Lu C, Abdelzaher T (2003) SPEED: a stateless protocol for real-time communication in sensor networks. In: International conference on distributed computing systems, Providence, RI
7. Felemban E, Lee C-G, Ekici E (2006) MMSPEED: multipath multispeed protocol for QoS guarantee of reliability and timeliness in wireless sensor networks. IEEE T Mobile Comput 5(6):738–754
8. Huang X, Fang Y (2008) Multiconstrained QoS mutlipath routing in wireless sensor networks. Wirel Netw 14:465–478
9. Bagula AB, Mazandu KG (2008) Energy constrained multipath routing in wireless sensor networks. In: 5th international conference on ubiquitous intelligence and computing (UIC), vol. 5061, 23–25, pp 453–467
10. Ganichev I, Dai B, Godfrey PB (2010) YAMR: yet another multipath routing protocol. ACM Sigcomm Comp Com 40(5):14–19
11. Heinzelman WB, Chandrakasan AP, Balakrishnan H (2000) Energy-efficient communication protocol for wireless microsensor networks. In: 33rd Hawaii international conference on system sciences, pp 1–10
12. Younis O, Fahmy S (2004) HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. IEEE T Mobile Comput 3(4):366–379
13. Lobiyal DK, Mohapatra DP, Nagar A, Sahoo MN (2016) Distributed clustering in ad-hoc sensor networks: a hybrid, energy-efficient approach. In: International conference on signal, networks, computing and systems

14. Haque ME, Matsumoto N, Yoshida N (2009) Context-aware cluster-based hierarchical protocol for wireless sensor networks. Int J Ad Hoc Ubiq Co 4(6):379–386
15. Chaaran KN, Younus M, Javed MY (2010) NSN based multisink minimum delay energy efficient routing in wireless sensor networks. Eur J Sci Res 41(3):399–411
16. Kumar D, Aseri TC, Patel RB (2009) EEHC: Energy efficient heterogeneous clustered scheme for wireless sensor networks. Comput Comm 32(4):662–667
17. Heikalabad SR, Rasouli H, Nematy F, Rahmani N (2011) QEMPAR: QoS and energy aware multi-path routing algorithm for real-time applications in wireless sensor networks. IJCSI Int J Comput Sci Issues 8(1):466–471

Extending Bidirectional Language Model for Enhancing the Performance of Sentiment Analysis

Eashan Arora, Sakshi Mishra, K. Vimal Kumar and Pawan Upadhyay

Abstract Sentiment analysis provides an analysis about the writer's emotion conveyed in the text. It uses natural language analysis to predict the sentiment of the writer. The proposed work describes an effective and efficient natural language approach to realize multi-classification for textual data according to their respective sentiments. Proposed system employs a RNN language model based on Long-Short term memory (LSTM) over pre-trained word vectors that was generated from different language model for sentiment classification. In order to improve the performance of sentiment analyser, we have used a bidirectional language model—embeddings from language model layer (ELMo). We have merged the Natural language processing and Deep learning techniques to analyse and achieve an important emotion from the long sentences fed to the system. The proposed system is compared with state-of-art approaches such as—Simple Multilayer Perceptron, recurrent neural network (RNN) and LSTM. These techniques are thus being applied to classify sentiments on the dataset of imdb-movie review (Mass et al. in Learning word vectors for sentiment analysis. Association for Computational Linguistics, pp 142–150, 2011 [1]). These techniques have improved the system to capture syntactic and semantic relationship that can help in identifying the sentiments.

Keywords Bidirectional language model · Sentiment analysis · ELMo · Multilayer perceptron · RNN · LSTM · Stacked LSTM

E. Arora (✉) · S. Mishra · K. V. Kumar · P. Upadhyay
Jaypee Institute of Information Technology, Noida, India
e-mail: eashanarora2798@gmail.com

S. Mishra
e-mail: mishrasakshi1208@gmail.com

K. V. Kumar
e-mail: vimalkumar.k@gmail.com

P. Upadhyay
e-mail: pawanupadhyay@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_15

1 Introduction

In various areas, the feedback provided by the customers regarding some product/service has an incomparable relevance according to the context. Opinions which are expressed in the form of feedbacks provide an opportunity to get an idea about the likes and dislikes of community on the whole. Thus, for any data centric company and their data scientists who look into extracting some meaning out of an unstructured textual data, sentiment analysis will be one of the initial steps towards such a task with a very little investment. Using natural language processing (NLP), extracting textual features that can be used for sentiment analysis preserves its popularity according to their usage. The two main focuses in NLP are language processing and language generation. The ability to recognize the text and understand the meaning is performed in language processing whereas language generation involves generating natural language from linguistic data. The emphasis of this paper will be on the processing of textual data and mapping it features to perform efficient sentiment analysis. This paper also addresses the core problem of sentiment analysis on a particular context.

The main contribution of this paper is the improvement of sentiment analysis by making use of new word embeddings from a language model that can capture the language specific features. This language specific feature contributes in learning the parameters of network based their corresponding labels. The main idea is to automatically train model without pre-processing of data. The literature study shows [2] that a deep contextualized language model such as ELMo will be a very powerful language embedding model for sentiment analysis because it makes use of a bidirectional analysis using the network. Also, it generates the word vectors based on the word usage in terms of syntactic and semantics in that sentence. The network should hold a memory in order to deal with sequential data such as a text. The meaning of a word has dependency not only on the context of the previous word but on the whole sentence. Thus, the proposed system uses a bidirectional language model which has been trained on 1 Billion words from benchmark dataset and accepts strings in the form of raw or tokenized form. The model output is mean pooled vector representation of the input fed to it.

The paper is organized as follows: in Sect. 2, we present a brief description on some related works on sentiment analysis and our approach towards it. The dataset used and its description presents in Sect. 3. Section 4 describes the software packages and models we used in our proposed study. Experimental results are presented and analysed in Sect. 5. We also present upcoming work and approach towards it in last section of the paper after concluding section that summarizes the findings of this proposed work.

2 Related Work

Review given by people gives the user's thought which plays an essential role in gathering up any important information. Resources such as online reviews and personal blogs provide an opportunity to understand and seek other's opinion. This makes sentiment analysis an important task.

Sentiment analysis has grown up remarkably during the early 2000s. However, first approach towards such analysis has been done by Hatzivas-Silogou and Mckneown [3] in 1997. In this research, they used wall street journal data to perform research on identifying the semantic orientation of adjectives. Since dictionaries do not show sentiments, they made a basic assumption that sentence itself would serve that information. They demonstrated that the conjunction between the adjectives provide an indirect information regarding the sentiments of the sentences.

Consider the example sentence—*movie was good but bit slow before interval*, in which good and slow are the two dissimilar words that are connected by the conjunction but their main focus were on identifying the sentiment orientation of adjectives and not the whole sentence. When analysing a word's semantics it is being found that each word have an isolated meaning of itself and also having some meaning in context of a sentence and this concept is called as the "Principle of Compositionality". According to this principle [4], the meaning of word depends on the meaning of its predecessor in the whole sentence. Thus describes its meaning in respect of whole text.

In recent years, artificial neural network gains more importance in the field of sentiment analysis. Some techniques that are designed for the representation of sentences through semantic composition are convolution neural network (CNN), recurrent neural network (RNN), Long-short-term memory (LSTM), etc.

Recurrent neural network came in across during 1980s but gained an importance since last few years because of increase in computational power and also helps in dealing up with large amounts of data. Because of its internal memory, RNN's are able to store important information regarding the input and then predicting an output through it. But RNNs has problem in dealing with the long-term dependencies [5]. This problem leads to the invention of LSTM during 1997 (discovered by Hochreiter and Schmidhuber). It is an extension for recurrent neural network that makes the usage of larger internal memory thus to deal with long-term dependencies and also enables RNNs to remember their inputs over a long period of time [6].

In recent years, sentiment analysis was explored a lot. In a research on sentiment classification using product reviews [7], the researchers identified the class of sentiment polarities both at the sentence-level and review-level. Later on, the prediction of sentiments using texts on social media was explored and to improve its accuracy new features were introduced [8].

To deal with the problem of polysemy i.e. the concept that word has different meanings according to usage; proposed system uses ELMO since other embedding models focuses on one vector output representation, combining of different meanings of single word all together. ELMo [9] is the embedding technique proposed by

Allen NLP. Unlike other embedding models it is dynamic as it changes every time depending on the context even if the word is same, totally depending on the nature of the sentence.

Dataset

To illustrate and evaluate the need for bidirectional language model for sentiment classification, data for training and testing was collected from an online source platform-Kaggle. The IMDBmaster dataset is a set of 7998 reviews from different people having their particular views on a particular movie. Each of these movie reviews is classified as either being "positive" or "negative". The data is divided into a train and validation set, trained on 5358 sentences and validation on 2640 reviews. Among these set, 50% are positive reviews and rest 50% are negative. The training database contains the imdb reviews as input which accepts it in the form of raw strings or tokenized strings.

3 Proposed Work

The proposed system aims to provide a deep learning model that can capture the language specific features in an appropriate manner such that it can improve the accuracy of sentiment analysis. In order to capture the language specific features, a bidirectional language model is being used. The features captured are represented in the form of vectors which are further used by sentiment analysis models to map the relation with their respective sentiments. The sentiment analysis model is developed using different models such as multi layer perceptron model, long short term memory (LSTM) network and stacked LSTM network.

3.1 Word Embedding Using Bidirectional Language Model

Recent developments in improving computational power of computers have taken the development in NLP research to a new dimension. The language model takes a new shift in how words are vectorized or encoded. Word-embeddings have been a major reason how different NLP model deal with language. Previously models like continuous bag-of-words (CBOW), Skip gram model and glove have been used for embeddings task. But, each of them has certain disadvantages. Due to these disadvantages a bidirectional sentence level analysis is being employed in this system instead of textual feature extraction using a directional (left-to-right) sentence level analysis. One such model is embeddings from language model (ELMO) that uses bidirectional analysis (biLM-bidirectional language model) to learn both word (e.g., syntax and semantics) and linguistic context. Currently the bidirectional language model shown in Fig. 1 has been used for 6 Downstream NLP tasks namely: question answering, textual entailment, semantic role labeling and named entity extraction.

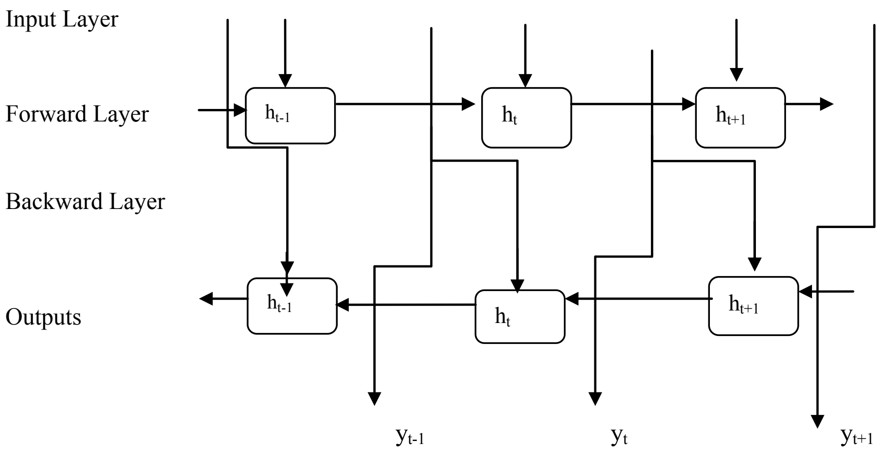


Fig. 1 Bidirectional language model

ELMO gives an embedding according to the context of the sentence it's used in, so that it can capture both word-level meaning and contextual information. The language model gains the understanding by getting trained to predict the next word based on the current word in the current sentence. Since ELMO employs bidirectional approach, it actually goes a step further and trains the model such that it can extract the sense of the next word and also the previous word.

This language model is basically concatenations of the activations on several layers of the biLMs (bi-directional Language Models). Different layers of a language model encode different kind of information on a word (e.g. Part-Of-Speech tagging is well predicted by the lower level layers of a biLSTM while word-sense disambiguation is better encoded in higher-levels).

3.2 Sentiment Analysis Approach

Sentiment analysis is being used in many other NLP applications, so that, it can improve the overall accuracy of those applications. In general, sentiment analysis can be used for predicting customer satisfaction based on their reviews. Based on these review, it can also be used for finding the popularity of a particular product. Since it has wide scope for integration, the need for improving its accuracy is predominant. In this proposed system, the deep learning models are used for mapping the textual features with its corresponding sentiment labels. The efficiency of bidirectional model is evaluated using different sentiment analysis models such as multi-layer perceptron, long short term memory (LSTM) and a stacked LSTM.

Multi-layer perceptron based sentiment analyzer

The multilayer perceptron network used for sentiment analysis has one input layer and an output layer. But, there is flexibility in the number of hidden layers to be used. In this proposed system, the number of hidden layers was kept in varying manner and the accuracy of network is found to be good when there were three hidden layers being used in it. This is due to improvement in the learning of network parameters.

LSTM based sentiment analyzer

The recurrent neural network (RNN) was introduced in this proposed system to capture the features in an appropriate manner. Since the recurrent neural network has a vanishing gradient issue, the long short term memory network (as shown in Fig. 2) was later introduced in place of simple RNN. As in recurrent neural network, when the error is propagated it goes through the layers of neurons which are connected to themselves. The hidden layers which are connected to other hidden layer have network weight parameter called the recurrent weight. The weight is applied many times on top of itself which causes the gradient error to decline rapidly. That is, the weight of the layers on very far left are a bit updated much slower than the weight on the other layers on the far right and this creates a domino effect.

$$
\begin{aligned}
&\text{IF:}\\
&\quad \text{WREC} = \text{SMALL} \rightarrow \text{Vanishing gradient}\\
&\quad \text{WREC} = \text{Large} \rightarrow \text{Exploding gradient}
\end{aligned}
$$

As LSTM are capable of handling long term dependencies. For LSTMS WREC = 1, there is only 2 point-wise operations as well understand for the down and there's

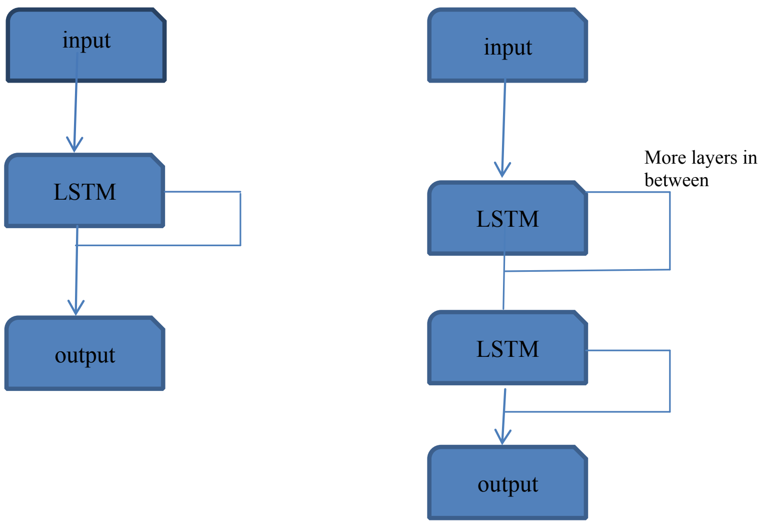


Fig. 2 LSTM model versus stacked LSTM model [10]

no complex neural network layered operations. LSTM'S have a memory cell which just goes through time, sometimes it might be removed or sometimes things might be added to the memory cell. Otherwise it flows through times freely and therefore when you back propagate through these LSTMS there is not much of a problem of vanishing gradient.

Stacked LSTM

Since LSTM model consist of single hidden layer followed by feedforward output layer, the textual features are learnt in a coarse manner. In order to fine tune the learning of features there is need for multiple LSTM layers. So, the proposed system uses stacked LSTM layers which is the extension to this model that has multiple hidden LSTM layers as shown in Fig. 2. Stacked LSTMs were first introduced by Graves et al. in their application of LSTMs to speech recognition. Stacked LSTM helps to increase the depth of network which in turn helps to increase the accuracy and efficiency of the model. A single large hidden layer can be used to approximate functions but increasing the depth of the network is another solution that only requires few neuron and trains faster. In stacked LSTM, one LSTM layer provides a sequential output which then serves as an input for the next LSTM layer. It has been argued that stacked LSTM layer allows network to learn at a different time scales over an input to make a better use of parameters [10].

4 Result Analysis

Firstly, the proposed bidirectional model performance is evaluated over the previous built models such as, multi-layer perceptron and LSTM. So that importance of bidirectionality in sentiment analysis is detected properly. Initially for the multilayer perceptron model, a dense network layer with default 256 neurons was used and for input vector the training data samples are fed to the ELMO embedding layer which is of a 2D shape vector with 1024 as its dimension.

For the LSTM based model, the analysis on this layer was performed and we found that LSTM does not support masking and lengths of different variable size. To overcome this, a masking function in the ELMO model is employed. Further to train the ELMO embedding layer, input training sentences are used. Afterwards embeddings from the Language Model are fed to the LSTM layer with a 3D shape of the same dimensions as the dense network.

The further validation is done on the testing data samples; the model is trained for 10 epochs with each of network layers and each network yields a particular accuracy as shown in Table 1 (Fig. 3).

Table 1 Accuracy of bidirectional model over different sentiment classification models

ELMo + Baseline model	RUN1	RUN2	RUN3	RUN4	RUN5
MLP	86.62	87.7	87.4	87.35	86.51
LSTM	87.25	86.43	87.23	87.23	87.42
Stacked LSTM	87.12	85.98	86.82	87.16	86.79

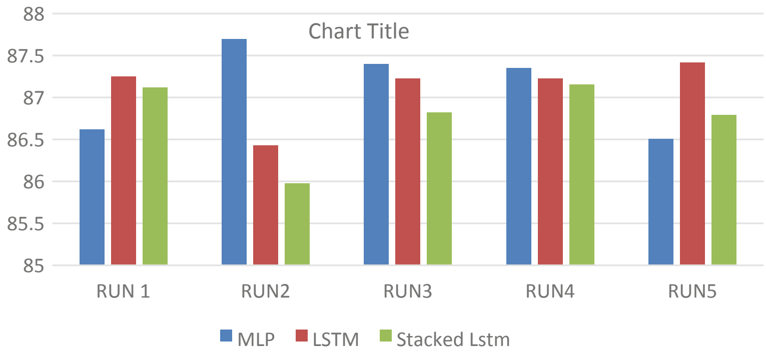


Fig. 3 Results of proposed sentiment analyzer in different runs

5 Conclusion and Future Work

The proposed sentiment analyzer was developed and was found to improve the accuracy of sentiment analyzer with a bidirectional language model-ELMO. While analyzing a particular text, the semantic information of a specific word cannot be extracted by considering it as an individual unit. But, the semantic information can be extracted using the context usage in the whole sentence. That is why we used a model that not only focuses on the context of the word but also on the context of the whole sentence—both forward and backward. We tried implementing the Bidirectional model in different layers of the network such as multilayer perceptron network. LSTM and Stacked LSTM. A network is built to classify the sentiment of the IMDB movie review dataset. It is found that the performance of proposed sentiment analyzer using stacked LSTM network and bidirectional language model was so good. It is found to be 87% accurate compared to other variants developed. The use of bidirectional language model has significant improvement in sentiment analyzer. But, still there is scope for further improvement which can be developed using bidirectional model to predict the sentiment analysis as well.

References

1. Maas AL, Daly RE, Pham PT, Huang D, Ng AY, Potts C (2011) Learning word vectors for sentiment analysis. In: Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies, vol 1. Association for Computational Linguistics, pp. 142–150
2. Peters, ME Neumann M, Iyyer M, Gardner M, Clark C, Lee K, Zettlemoyer L (2018) Deep contextualized word representations, NAACL
3. Hatzivassiloglou V, McKeown KR (1997) Predicting the semantic orientation of adjectives. In Proceedings of the eighth conference on European chapter of the association for computational linguistics (EACL '97), Association for Computational Linguistics, Stroudsburg, PA, USA, pp 174–181. https://doi.org/10.3115/979617.979640
4. Mehler A (2007) Compositionality in quantitative semantics. A theoretical perspective on text mining. In: Aspects of automatic text analysis. Studies in fuzziness and soft computing, vol 209. Springer, Berlin, Heidelberg
5. Hochreiter S (1991) Untersuchungen zu dynamischen neu-ronalen netzen.Master's thesis, Institut fur Informatik, Technische Universitat, Munchen
6. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
7. Fang X, Zhan J (2015) Sentiment analysis using product review data. J Big Data 2(1). https://doi.org/10.1186/s40537-015-0015-2
8. Alaoui IEI, Gahi Y, Messoussi R, Chaabi Y, Todoskoff A, Kobi A (2018) A novel adaptable approach for sentiment analysis on big social data. J Big Data 5(1). https://doi.org/10.1186/s40537-018-0120-0
9. Peters M, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, Zettlemoyer L (2018) Deep contextualized word representations. In: Proceedings of the 2018 conference of the North American chapter of the association for computational linguistics: human language technologies, vol 1 (Long Papers), vol 1, pp 2227–2237
10. Sak H, Senior A, Beaufays F (2014) Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In: Proceedings of the annual conference of the international speech communication association, (INTERSPEECH), pp 338–342

Energy-Efficiency Driven Strategy for Resource Aggregation Based on Future Resource-Usage Inference in a Cloud Computing Environment

N. Asha and G. Raghavendra Rao

Abstract Aiming at the huge energy consumption issues, this paper proposes a novel strategy for aggregation of Virtual Machine (VM)s using future resource usage inference. With the rapid and enormous development in the area of cloud computing, datacenters, a major part of cloud infrastructure swallows large amount of electrical energy and leads to increased cost and CO_2 emissions. In this direction, consolidation of VMs also referred as aggregation in our work anticipates resource usage in the near future within each host. Based on this work, a model is proposed which identifies a source host machine that is either Excess-Provisioned (EP) or Less-Provisioned (LP), a target host machine for accommodating the VMs selected from the above machines. Our experiments show through a rigorous analysis with the peer works in existence, minimization in the number of active hosts.

Keywords Cloud computing · Resource usage prediction · Energy efficiency · Linear regression

1 Introduction

Cloud Computing is becoming popular in recent times. Many companies such as Amazon, IBM, and Rackspace are renting out services for the customers through their own infrastructure (IaaS). Due to the high demand for resources, datacenters have been deployed in large number at sparse geographical locations. These datacenters consume huge amount of power not only with respect to the infrastructure housed within it, but because of improper, unplanned usage of these resources. Therefore, there is a necessity of building a methodology that makes best use of resources in an efficient manner. Virtualization [1, 2] is a promising technology that enables the efficient utilization of resources in a sharing basis of Physical Machine (PM) among multiple VMs which are isolated from each other.

N. Asha (✉) · G. Raghavendra Rao
NIE, Mysuru, India
e-mail: asha.n.naveen@gmail.com

G. Raghavendra Rao
e-mail: grrao56@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_16

The main issue in the cloud computing infrastructure is the increased energy consumption of datacenters reflecting on the increased operational expenses too. A program designed and developed without proper care to energy consumption reduction may lead to generation of large amount of heat from the components involved and hence demands cooling equipments for them and affects environment. Therefore, to enhance the proper usage of the resources in the datacenter, certain challenges in the virtual environment to be addressed are (i) determination of EP and LP hosts based on certain ceiling values (ii) selection of VMs for migration from the above cases (iii) to search for a better placement in the host for the VMs migrated. In our work, we have organized into different sections. To begin with, examination of background work is carried out in accordance with the energy conscious resource provisioning; next section addresses the predicting of future resource usage achieved through Linear Regression [3] techniques, and finally addressing the issues of EP and LP host problems through efficient algorithms. Several experiments are carried out in the simulation environment as it is feasible to make changes in a realistic way and the results are analyzed with that of the models developed in [4, 5]. There is of course an improvement found in our work with respect to the energy consumption reduction by 25–30%.

2 Related Work

There exists an extensive amount of work in the techniques which are part of virtualization technology. Beloglazov and Buyya [6] have proposed a method for allocation of VMs aiming at minimizing power consumption. This work suggests a heuristic based method about decision of VM allocation without violating Service Level Agreement (SLA), which is a contractual engagement between an end-user and cloud provider. This method also suggest for decreased migration count so that unnecessary complexities can be avoided. During the time of VM migration, hosts are turned on to sleep state. About 83% of the power savings with a minimum tolerable SLA violation is observed. Duy et al. [7] have proposed a method to optimize the server consumption in coordination with a scheduling algorithm. Optimization is carried out with a workload predictor using neural network based on past history of utilization. Based on this algorithm, many of the nodes were put to switch-off mode to reduce the count of active servers that contributed to power consumption. Results have proved that about 46% of power savings.

Dong et al. [8] considered CPU utilization rate and bandwidth of hosts in a homogeneous cloud environment to propose reduced energy consumption and total count of alive hosts. This helped them to come up with an algorithm for VM allocation to conserve power and hence to save energy. The authors of Beloglazov and Buyya [9] have presented a heuristics based algorithm for an energy efficient datacenter with a strong SLA maintenance. Steinder et al. [10] emphasized on management of heterogeneous workloads through VM migration among the host nodes as the fresh tasks started arriving. But the cost incurred during migration was more and power saving was not considered seriously.

3 Problem Formulation

Let us consider a scenario like the one shown in Fig. 1. VMs are spread across host nodes where there are three host nodes and eight VMs. In this setup, host-node1 is launching five VMs indicating EP, host3 is with a single VM by signaling LP and host-node2 with two VMs launched appears to be in a balanced state. The case of EP is leading to strong SLA violation and LP is demanding the host-machine to be active thereby increasing the energy consumption. The states of each of the host-nodes are determined through a fixed ceiling strategy. Both the situations are

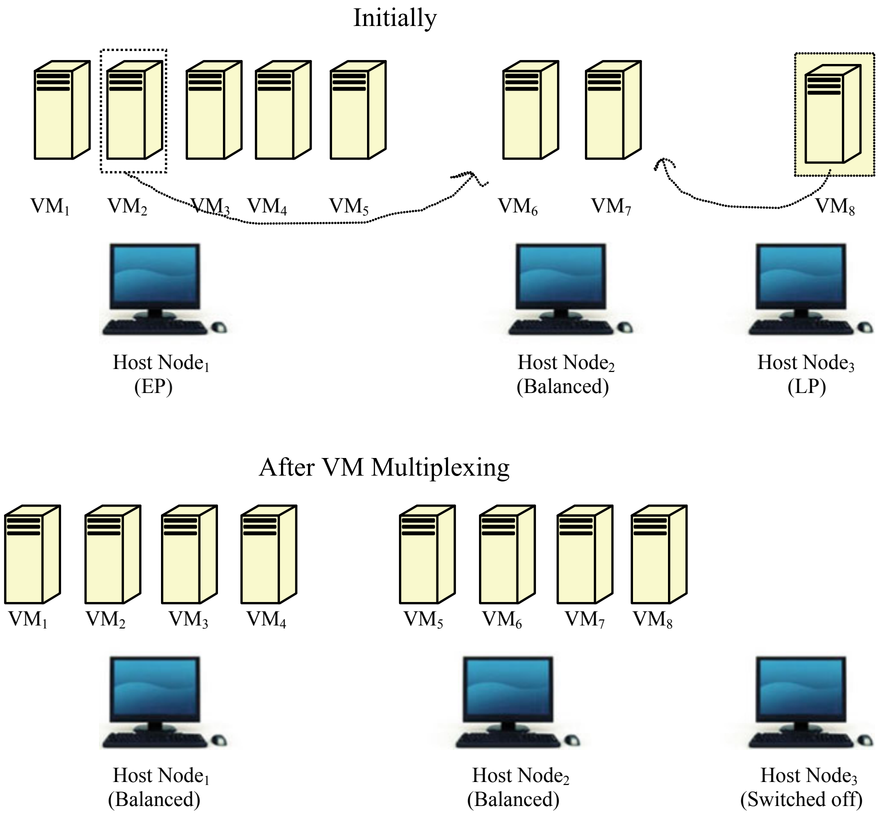


Fig. 1 VM multiplexing

depicting improper usage of resources by wasting the energy and degradation in quality of service (QOS) and performance of the system is observed.

If we think of applying the idea of VM multiplexing, above condition can be better improved and the process can be divided into three stages as

- Source host node selection based on EP and LP state
- Choosing of one or more VMs from the from the previous stage
- Destination host node selection to launch the chosen VMs from the previous stage or if any fresh demands for VMs arrived.

After applying this strategy, we could observe only two host nodes to be active in the entire system and also in a balanced state. The third node discovered in LP state is turned off to further bring down the energy consumption rather than maintaining in idle state.

4 Proposed Model

It is a normal scenario in cloud environment that VMs are created with diversified operating systems and executing users' requests within them by monitoring the access rights and serve the requests arriving on-demand. Cloud providers usually will take care of a balanced distribution of load and the monitoring the activity of hosts on a regular basis in spite of unpredictable demands. Suppose if there are 'm' VMs and 'n' requests, each of the VM can be hosted on different physical machines as long as the host can accommodate enough resources. At the same time, we also have to balance the demand against the allocation with minimized energy consumption. It can be achieved through inferring the demands or future usages of the resources. Our proposed algorithm emphasizes on the same before mapping VMs onto Physical machines.

Figure 2 shows the model proposed, which concentrates on IaaS (Infrastructure as a Service) among other various cloud services. Task Queue Scheduler holds demands in queue that follows predictor entity which infers the future resource demand based on the past history and EP of the overall infrastructure. Followed by the EP discovery, migration as the remedial practice is considered by shifting VMs from source host node to other. At this point of time, VM allocation is to be done by finding an appropriate placement for them. Actually, resource allocation would have allocated resources for the entire session of the task execution. Depending on the load at execution time, system can either include or delete resources. Isolation, coordination between VMs is maintained through reservation of an entire session for the job execution until its completion.

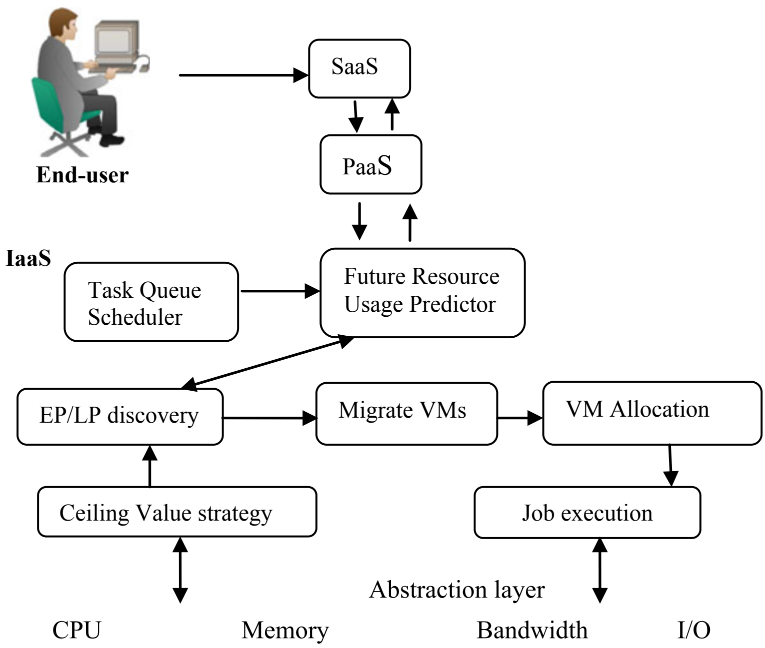


Fig. 2 System model

5 Implementation Details

Our implementation idea is mainly divided into two parts: at first, discovering LP hosts, EP hosts and forecasted to-be EP host. Next, shifting of all VMs will take place from LP hosts and few from the rest of two host types. This follows with the identification of a target host for accommodating these migrating VMs from the LP hosts so that the hunt starts with the maximum loaded host. Computation of both source host and target host's CPU usage is carried out. This process is continued until all VMs in source host are shifted. If any one VM is left, entire migration process is reverted and the migration list continues to hold null value. If not, source host is changed to sleep mode. Later, it is shifting of VMs from EP hosts. During this process, shifting happens with the VM with highest request that is if current CPU usage crosses beyond the maximum CPU capacity of host.

In both of these cases, host identification is done based on following conditions. Firstly, we have to check whether the target host has enough resources for accommodating incoming VM. If it can, it will continue to host it. At the same time, if it is found that some resource's CPU usage is nearing cent percent, a fixed ceiling value of 80% is set to control the quantum of CPU requested by VM. Next is anticipation of the target host for EP in the near future by applying prediction rule based on past dataset samples.

6 Evaluation with Assessment

6.1 Simulation Details

As it is difficult to test our work on a large scale cloud infrastructure, we have conducted the experiments on the CloudSim Simulator [11], a popular toolkit due to its efficient and extensible provisioning of resources. To evaluate our algorithm performance, 500 heterogeneous hosts are considered. A comparison analysis of the usage performance is carried out by varying the ceiling value with those of the benchmark algorithms. Energy consumption and the number of migrations are computed based on the real workload collected for the three days (9th, 11th and 12th April 2011) as a part of PlanetLab [12], an infrastructure-based project called CoMon within Cloudsim. For each simulation run, one of the VMs from a particular day of the workload is allocated to our working VM. All of the VMs in this environment of real workload are single cored.

Our HRUF algorithm uses the ceiling value to control its usage which means that the forecasted average utilization of all VMs executing on the host-node and demanded usage by the newly arrived VMs must be less than the UT of current CPU usage which can be represented as

$$UT * TCPU_{util}\left(D_p\right) > VCPU_{util}(VM) + FCPU_{util}\left(D_p\right) \tag{1}$$

VM-FRU algorithm collects future demanded VM statistics using least square regression method. The constraint used here is that VM can be accommodated in the target host only if the forecasted demand of the VM and average usage of the VMs executing on the host machine must be less than that of the UT of the total usage. This can be represented as

$$UT * TCPU_{util}\left(D_p\right) > VCPU_{util}\left(D_p\right) + FCPU_{util}(VM) \tag{2}$$

6.2 Result Analysis

Performance analysis of the algorithms is carried out to find a suitable target host for the migrated VMs and the count of migration statistics is collected across different ceiling values. Both of these results are represented in Fig. 3 and Fig. 4. With a comparative analysis of all the three algorithms, it is shown that FRUC algorithm has performed well. This is because target host has controlled VM migration-count due to based on future utilization inference. Much of the LP hosts are turned off by shifting the VMs running within them to other hosts. Also FRUC has emerged with energy savings in comparison with other approaches. However, it is observed that

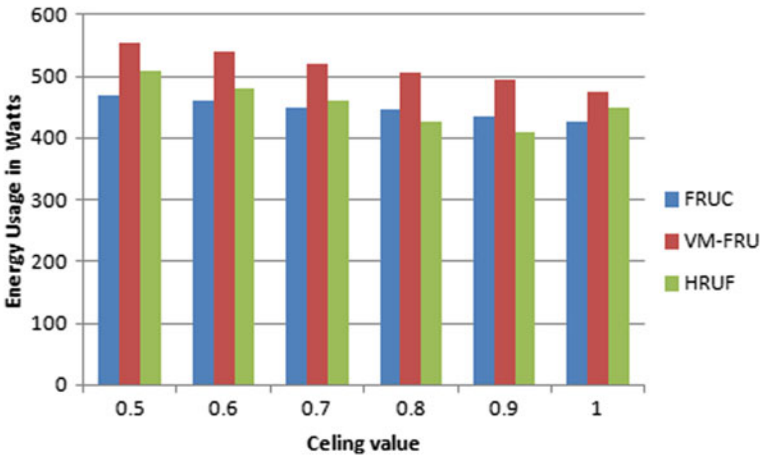


Fig. 3 Energy consumption of the algorithms with varied ceiling values

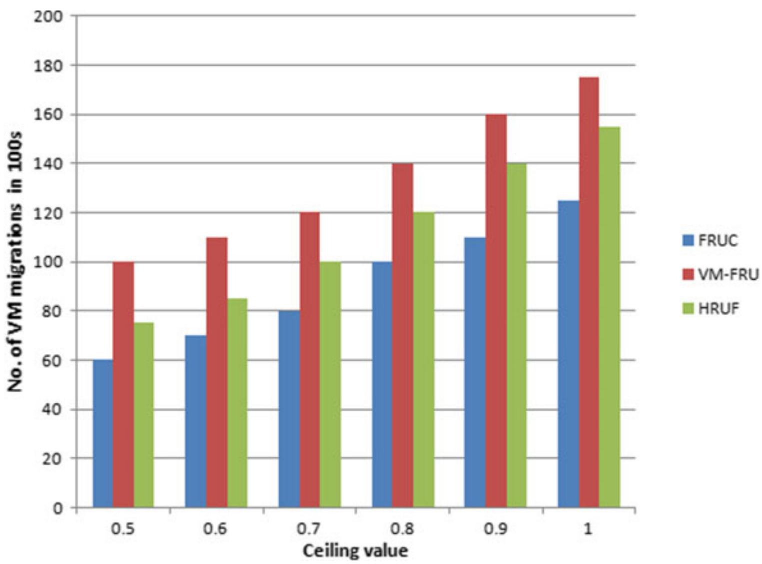


Fig. 4 Analysis of the algorithms in terms of number of migrations

closer to the hundred percent of the ceiling value, performance is better with respect to the energy consumption, but vice versa with the number of migrations.

Table 1 List of symbols used with their meaning

D_P	Target host machine
UT	Fixed ceiling value
$TCPU_{util}(D_P)$	Overall CPU usage of VMs on target host
$VCPU_{util}(D_P)$	Requested CPU usage by VMs on target host
$VCPU_{util}(VM)$	CPU usage consumed by VM
$FCPU_{util}(VM)$	Forecasted CPU usage consumed by VM
$FCPU_{util}(D_P)$	Forecasted CPU usage by VMs on target host

7 Conclusion

The increased number of datacenters as a part of cloud computing infrastructure expansion is causing a major hike in the operational expenses since much of the energy is consumed by most of its equipments. Therefore, it is required to minimize the power and hence the energy consumed as a part of green computing. This paper proposed an energy-efficiency driven inference algorithm that checks for Host EP and LP during VM allocation. Thus, an inference of the future resource usage based on past history is computed using linear regression concept. Also, CPU is the resource considered for prediction as it consumes majority of the energy in the entire infrastructure. As observed from the results presented, our proposed algorithm is outperforming with respect to number of migrations and energy consumption after comparing with the benchmark results. A fixed ceiling strategy is followed to detect the extreme conditions of allocation. That is, if the forecasted usage exceeds this value, host is considered to be EP and VM migration is recommended for few of the hosts. Similarly, by discovering LP hosts, all VMs within them are suggested for migration. For the future scope of this work, it is good to consider the other resource types such as bandwidth and disk usage to promote a QOS based virtual environment that helps to achieve energy efficiency.

Meanings of Symbols used

See Table 1.

References

1. Barham P, Dragovic B, Fraser K et al (2003) XEN and the art of virtualization. In: ACM symposium on operating systems principles (SOSP'03), ACM, New York, USA, pp 164–177
2. Sotomayor B, Montero RS, Llorente IM, Foster I (2009) Virtual infrastructure management in private and hybrid clouds. IEEE Internet Comput 13:14–22
3. Seber GAF (2003) Linear regression analysis, Wiley-Interscience
4. Beloglazov A, Buyya R (2012) Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers. Concurr Comput Pract Exp 24(12):1397–1420

5. Ferreto TC, Netto MAS, Calheiros RN, De Rose CAF (2011) Server consolidation with migration control for virtualized data centers. Future Gener Comput Syst 27(8):1027–1034
6. Beloglazov A, Buyya R (2010) Energy efficient resource management in virtualized cloud data centers. In: Proceedings of the 2010 10th IEEE/ACM international conference on cluster, cloud and grid computing, IEEE Computer Society, pp 826–831
7. Duy TVT, Sato Y, Inoguchi Y (2010) Performance evaluation of a green scheduling algorithm for energy savings in cloud computing. In: IEEE international symposium on parallel distributed processing, workshops and Phd Forum (IPDPSW), pp 1–8
8. Dong Y, Zhou L, Jin Y, Wen Y (2015) Improving energy efficiency for mobile media cloud via virtual machine consolidation. Mob Netw Appl 20(3):370–379
9. Beloglazov A, Buyya R (2010) Energy efficient resource management in virtualized cloud data centers. In: Proceedings of 10th IEEE/ACM International CCGrid, pp 826–831
10. Steinder M, Whalley I, Carrera D, Gaweda I, Chess D (2007) Server virtualization in autonomic management of heterogeneous workloads. In: 10th IFIP/IEEE international symposium on integrated network management, New York, pp 139–148
11. Calheiros RN, Ranjan R, Beloglazov A, Rose CAFD, Buyya R (2011) CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. J Softw Pract Exp 41(1):23–50. https://doi.org/10.1002/spe.995
12. Park K, Pai V (2006) CoMon: a mostly-scalable monitoring system for PlanetLab. ACM SIGOPS Oper Syst Rev 40:65–74

A Comparison of Deep Learning Algorithms for Plant Disease Classification

V. Vimal Adit, C. V. Rubesh, S. Sanjay Bharathi, G. Santhiya and R. Anuradha

Abstract The Plant Disease is a major threat to Farmers, but faster identification remains a difficult task due to lack of incorporating latest technologies. The classification of plant disease is very challenging as it has only minimal features and hard to differentiate. The proposed technique uses a public dataset of 76,000 images of diseased plant leaves. We train the model using LeNet, AlexNet and Inception V3 for categorizing 38 varieties of plant diseases. The proposed model gives a comparison study of above three deep learning models and from the results we can predict Inception V3 performs better with an accuracy of 98% when compared to other models.

Keywords Plant disease classification · LeNet · AlexNet · Inception V3

1 Introduction

One of the important sectors of the Indian economy is agriculture. Farmer's economic growth depends on the quality of the products. In field of agriculture, detection of disease in plants plays an instrumental role. Plants are highly prone to diseases that in turn affect the growth of the plant. Finally, it affects the ecosystem. In order to detect a plant disease at a very initial stage, use of automated disease detection technique is advantageous. Manual detection of plant disease using leaf image is a tedious job. Several deep learning architectures are proposed for plant disease classification. Each has its own strength and weakness. This makes the right choice of model for leaf disease identification difficult.

LeNet is a classical architecture which has seven layers among which three are convolutional layers, two are pooling layers and one fully connected layer followed by output layer. LeNet is a simple architecture. AlexNet is a popular CNN architecture. It has five convolutional layer and three fully connected layers. It also has dropouts which is not used in LeNet architecture. Inception V3 is a modern CNN

V. Vimal Adit (✉) · C. V. Rubesh · S. Sanjay Bharathi · G. Santhiya · R. Anuradha
Department of Computer Science and Engineering, Sri Ramakrishna Engineering College, Coimbatore, India
e-mail: vimaladit.1601252@srec.ac.in

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_17

architecture. Model is made up of symmetric and asymmetric layers including convolution, average pooling, max pooling, concat, dropout and fully connected layers. Architecture of Inception V3 is complex than that of AlexNet and LeNet. Models are made efficient by using factorized convolution and aggressive regularization. In this paper, the comparison in performance of plant disease classification among LeNet, AlexNet and Inception V3 is shown

2 Literature Survey

Recent techniques in plant disease identification and prediction use Deep Learning Architecture. Pooja et al. [1], has attempted to identify diseases on leaves by using segmented images. Lee et al. [2] proved to obtain an accuracy of 65.4% on the Caltech-101 object classification task. Mehrotra et al. [3] implemented a classifier for Devanagari character recognition which obtained 98.19% accuracy. Simard et al. [4] trained a classifier for the MNIST handwriting recognition training set that performed with 99.6% accuracy. Ferentinos [5], Plant disease detection and diagnosis through deep learning models using healthy and diseased leaf images. Lee [6], Combined Deep convolutional neural network of multiple processing layer for extracting the features of the leaf images and AdaBoost algorithm to assemble recognizable images from SoftMax classifiers. Albawi [7], work focuses on important issues and parameters that effects the efficiency and working of convolutional neural network. Guo [8] performed Image classification using Convolutional neural network to analyze learning method and optimization algorithms. Jassmann modeled Convolutional neural network to classify the type of a tree using one of its leaves. CNN yields remarkable results in computer vision. Szegedy, attempted to improvise the model efficiency using factorized convolution and aggressive regularization.

2.1 Dataset

The dataset consists of images of size 256×256 pixels. Since all the images used are of the same size, filtering and pre-processing of data is not required for this dataset. The dataset is obtained from Kaggle plant disease dataset. It consists of totally 38 classes of diseases like Tomato Mosaic virus, Potato late blight etc., and each class consist of about 2000 images. The amount of data for training occurs is increased by using data augmentation. Figure 1 shows a sample images of Plant Disease dataset.

Fig. 1 A sample view of the plant disease dataset

3 Model Architecture

3.1 LeNet

The architecture of LeNet is small and primitive. It was first trained on MNIST dataset and provided great results. It consists of 7 layers among which 3 are convolutional, 2 are pooling, 1 fully connected and 1 output layer. Hence this architecture use less resources for training the model. Figure 2 (source: https://medium.com/@sidereal/cnns-architectures-lenet-alexnet-vgg-googlenet-resnet-and-more-666091488df5) displays a sample model of LeNet Architecture.

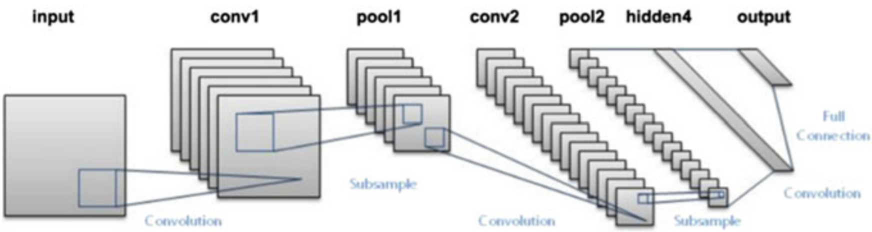


Fig. 2 A sample model of LeNet architecture

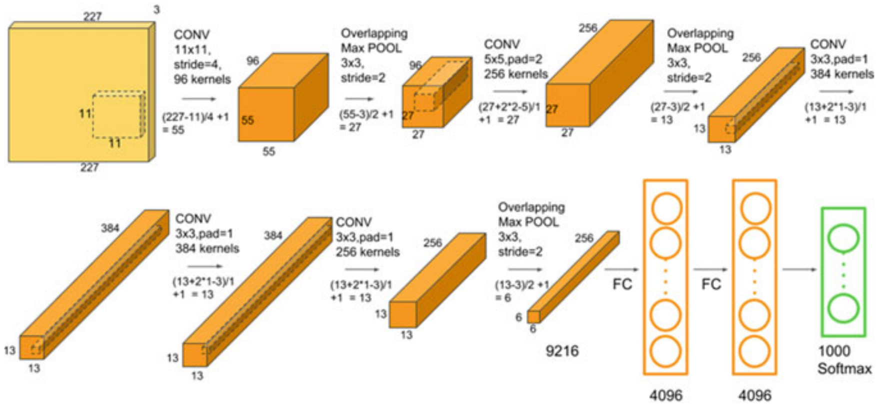


Fig. 3 A sample model of AlexNet architecture

3.2 AlexNet

AlexNet is a popular CNN architecture. It was first used on ImageNet dataset at the 2012 ImageNet competition and won the competition by a large margin. Its size was much larger when compared to other CNN architectures at that time. It has 5 convolutional layer and 3 fully connected layers. It also has dropouts before the first and second fully connected layer. Figure 3 (source: https://www.learnopencv.com/understanding-alexnet/) displays a sample model of AlexNet Architecture.

3.3 Inception V3

Inception v3 is the third iteration of GoogleNet. This is a modern architecture when compared to Alexnet and LeNet. Inception v3 is a 42 layered architecture. It uses factorization for reducing the number of connections without decreasing network efficiency. It also uses one auxillary classifier as a regularizer. Figure 4 (source: https://towardsdatascience.com/how-to-train-your-model-dramatically-faster-9ad063f0f718) shows a sample Inception v3 architecture.

4 Experiments and Results

In order to make the comparisons fair, the above mentioned models are trained for the given values mentioned in Table 1 Though LeNet is a popular architecture it does not hold well for 38 classes of images. Its performance is poor when compared to the other 2 models. The result is due to its reduced architecture framework. Figures 5

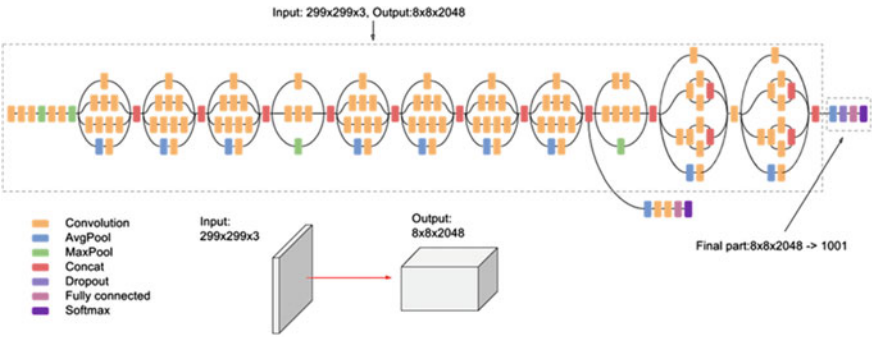
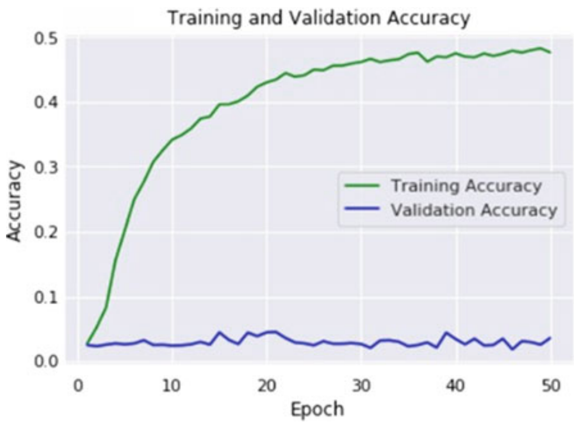


Fig. 4 A sample model of inception V3 architecture

Table 1 Parameters used for training the models

Parameter	Value
EPOCHS	50
BATCH_SIZE	128
STEPS_PER_EPOCH	550
VALIDATION_STEPS	64

Fig. 5 Performance of LeNet model



and 6 depicts the performance of LeNet Model. AlexNet performs well compared to LeNet model due to the inclusion of many convolution layers. Figures 7 and 8 depicts the performance of AlexNet Model. Figures 9 and 10 shows the performance of Inception V3 Model.

Inception V3 gives a good accuracy in terms of training data accuracy but when it comes to new data its accuracy drops. This was due to the amount of data used for training, which resulted in overfitting for the training dataset.

Table 2 illustrates the performance of the three models for various epochs. From

Fig. 6 Training and validation loss of LeNet model

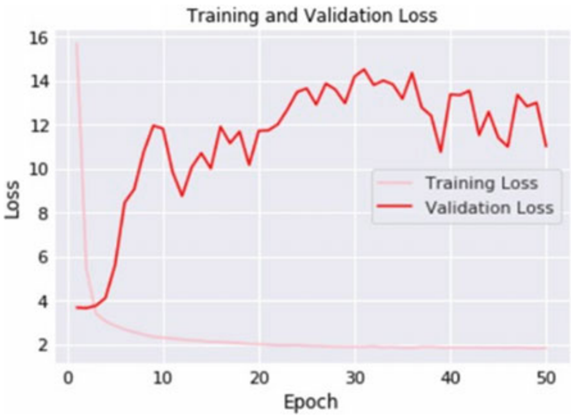


Fig. 7 Performance of AlexNet model

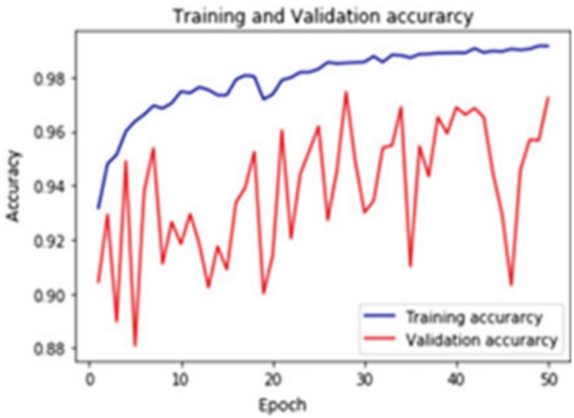


Fig. 8 Training and validation loss of AlexNet model

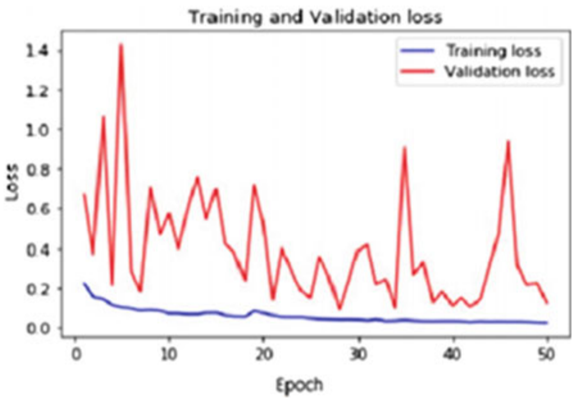


Fig. 9 Performance of inception V3 model

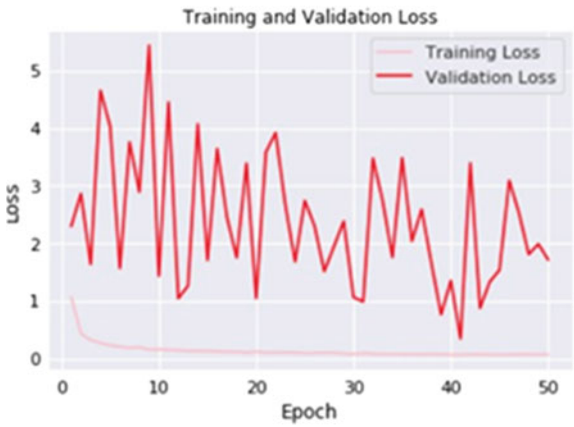


Fig. 10 Training and validation loss of inception V3 model

Table 2 Performance comparison with increase in epochs

Number of epochs	10	20	30	40	50
LeNet	34.14	42.99	46.15	47.48	47.63
AlexNet	96.78	96.82	97.25	96.83	96.97
Inception V3	95.32	96.83	97.55	97.78	98.32

the Table it is clear that Inception v3 attains a larger training accuracy compared to AlexNet Model. As LeNet is primitive with less number of convolution layers, the accuracy in training is very less compared to the other two models. Figures 11 and 12 demonstrates the comparison of three models with respect to training accuracy and Loss factor.

By comparing these performances we can infer that for this dataset, AlexNet yields the best result by providing high validation accuracy and low loss (Fig. 13).

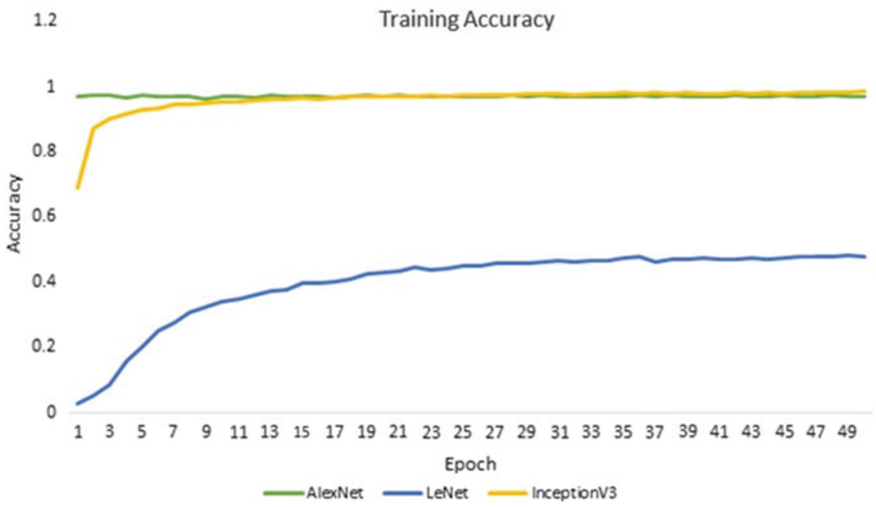


Fig. 11 Performance comparison of three models

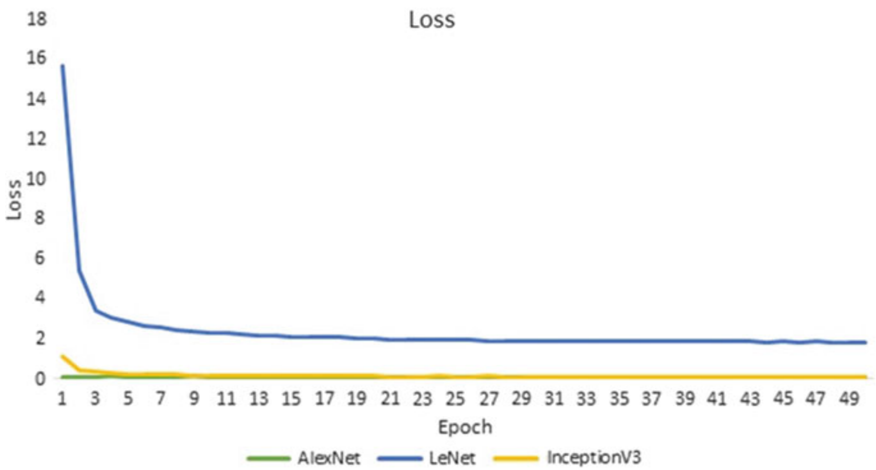


Fig. 12 Error loss comparison of three models

5 Conclusion

In this paper we have shown the efficacy of using plant disease image data for Categorizing using LeNet, AlexNet and Inception V3 models. Experiments show a significant gain in prediction accuracy for disease using Inception V3 comparative to other two models. This work has also demonstrated that consistency of data collection from different parts of the leaf is important as validation accuracy becomes inconsistent in Inception V3 model comparative to AlexNet which shows a better

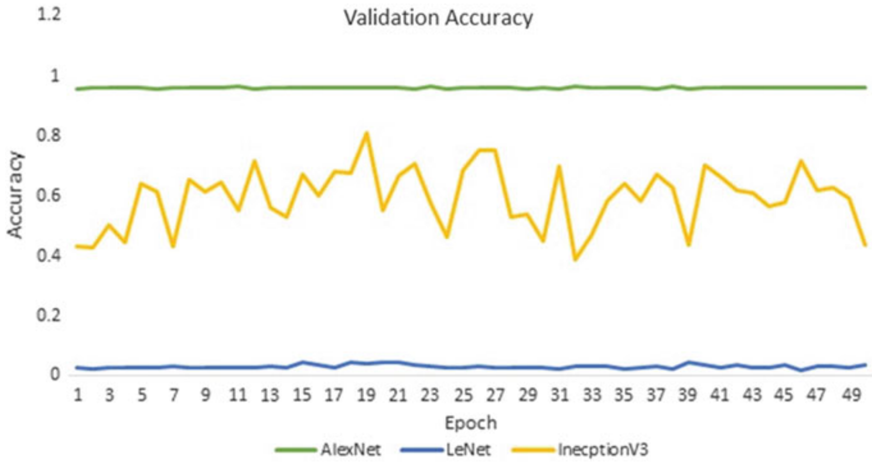


Fig. 13 Validation accuracy comparison of three models

validation Accuracy. This work can be extended for other deep learning models in prediction of plant disease.

Acknowledgements The authors would like to thank Sri Ramakrishna Engineering College for providing Resources for their work.

References

1. Pooja V, Das R, Kanchana V (2017) Identification of plant leaf disease using image processing techniques. In: IEEE international conferences on TIAR 2017, pp 130–133
2. Lee H, Grosse R, Ranganath R, Ng AY (2011) Unsupervised learning of hierarchical representations with convolutional deep belief networks. Commun ACM 54:95–103
3. Mehrotra K, Jetley S, Deshmukh A, Belhe S (2013) Unconstrained handwritten Devangari character recognition using convolutional neural networks. In: Proceedings of the 4th international workshop on multilingual OCR, New York, NY, USA
4. Simard PY, Steinkraus D, Platt JC (2003) Best practices for convolutional neural networks applied to visual document analysis. In: Proceedings of the seventh international conference on document analysis and recognition, Edinburgh, UK
5. Ferentinos KP (2018) Deep learning models for plant disease detection and diagnosis. Hellenic Agricultural Organization "Dementer", Institute of Soil & Water Resources, Department of Agricultural Engineering, 61 Dimokratias Av., 13561 Athens, Greece
6. Lee S-J, Chen T, Yu L, Lai C-H, Image classification based on the boost convolutional neural network
7. Albawi S, Mohammed TA, Al-Zawi S (2017) Understanding of a convolutional neural network. In: ICET2017, Antalya, Turkey
8. Guo T, Dong J, Li H, Gao Y (2017) Simple convolutional neural network on image classification. In: IEEE 2nd international conference on big data analysis

Active Learning Using Margin Sampling Strategy for Entity Recognition

Ankit Agrawal and Sarsij Tripathi

Abstract Different machine learning techniques require large amount of labeled training data to work well. It is very costly and difficult to label the unlabelled data. The active learning approach can be used in this case as it selects the most appropriate data from the unlabeled data pool. In this work, we have applied active learning using CRF classifier over biomedical disease names and adverse effect corpus using margin sampling strategy. Experiment have been carried for selecting the most uncertain samples for labeling by defining different threshold values for the margin used to measure the uncertainty. The active learning approach performs well with much less data requirement and its performance increases with the increase in threshold values of the margin.

Keywords Named entity recognition · Conditional random field (CRF) · Margin sampling

1 Introduction

Named entity recognition is one of the information extraction subtask in which the various mentions of different entities are identified and correctly classified into predefined categories. It is used in different information extraction tasks such as machine translation, text summarization, question answering system, etc. [1]. The supervised machine learning algorithms performs well for different information extraction problems for which sufficient labeled data is readily available. However, labeling of the unlabeled data is much difficult, costly and time taking. The active learning based approaches are widely adopted as one of a solution to this problem. The active learning approaches can select the most informed instances of data for labeling and training [2, 3]. Hence, the amount of data used by the active learner is quite less in

A. Agrawal (✉) · S. Tripathi
National Institute of Technology Raipur, Raipur, India
e-mail: aagrawal.phd2017.cse@nitrr.ac.in

S. Tripathi
e-mail: stripathi.cs@nitrr.ac.in

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_18

comparison to the supervised machine learning approaches. In this paper, the active learning algorithm is used for entity recognition of disease names and adverse effect corpus using the margin sampling strategy.

2 Related Work

Recently researchers are putting their effort towards named entity recognition task using semi-supervised or unsupervised approaches. Out of these work, some are discussed in this section. For named entity recognition from the clinical documents, new active learning approach is proposed by [4]. The authors have experimented the existing and proposed algorithm with different query strategies which includes diversity-based and uncertainty-based sampling strategies. A semi-supervised model is proposed by [5] which uses distant supervision to generate the weakly labeled training corpus which is further refined using bagging based active learner. This approach achieves improved accuracy in comparison to other semi-supervised models. Two new approach based on margin sampling strategy including SVM and ensemble of SVM-CRF is proposed for active learning by [6]. They tested their approach over corpus of 3 different languages and a biomedical domain corpus. A new active learning approach based on new clustering method is proposed by [7] which makes use of informativeness and representativeness of the documents. The proposed approach outperformed the other state of art methods. A smart seed set is used for high quality pre-annotation by [8]. Also, they compared time savings by active learning annotation using various active learning query strategies including random sampling and supervised learning. [9] proposed an active learning based approach with multiple query strategies to classify entities within social media text. Some other recent work in this field includes [10–12].

This paper contains following sections: Introduction and related works in Sects. 1 and 2, details of the corpus used in experiment in Sect. 3, Sect. 4 contains feature extraction and classifier details used in the experiment, Sect. 5 discusses workflow of active learning algorithm and steps used in our experiment, Sect. 6 is result and discussion followed by conclusion in Sect. 7 respectively.

3 Corpus Details

The corpus used in this paper is disease names and adverse effect corpus [13]. The corpus have 112634 tokens with total 4362 sentences and 5 different label classes which includes: ‘B-DISEASE’, ‘I-DISEASE’, ‘B-ADVERSE’, ‘I-ADVERSE’, and ‘O’. The corpus is present in CoNLL format and is divided such that 75% of the data is present in training set (i.e. 3270 sentences) and 25% of data is present in test set (i.e. 1092 sentences). The train set is treated as if it is unlabeled set i.e. U from which appropriate sentences will get selected to be labelled by oracle. In this work, we have assigned original labels assigned to the selected sentences instead of oracle assigning the correct labels.

4 Feature Extraction and Classifier Used

The quality of extracted feature have very important role in the performance of the machine learning based models. The pos tag information has been also captured as a feature for each word of the corpus. The pos tags has been added using the NLTK [14]. The feature which are extracted here includes: word type information (alphabet or digit), word case information, and context information of the previous and next word to the current word [15]. We could have used some more features that would have improve the performance of the CRF classifier.

However, our focus is towards comparing the performance of the active learning approaches with the supervise approach. So we have used same feature set for both type of approaches i.e. active learning approach and supervised approach. The classifier used in this experiment is the Conditional Random Field (CRF) classifier. The CRF classifier is commonly used for sequence labelling task [16]. We have used the sklearn-crfsuite for the purpose of implementation of CRF classifier [15]. We have also set parameters of the CRF model according to the [15].

5 Active Learning Algorithm Steps

The workflow used by the active learning algorithm can be understood with the help of the Fig. 1. First the active learner selects the most appropriate data instances to get labelled from oracle. These instances are then removed from the unlabelled set U and labelled by the oracle. These labelled sentences are then added to the existing labelled set L which is then used for the training of active learner. In this figure, x is the word and y is the correct label assigned to the word by the oracle. This research work is inspired from the work done by [6]. The active learning steps used in this work is as follows:

1. Initially some randomly labelled instances (here 2% sentences of the training set) are selected from U and given to labelled set L which was initially empty to train the active learner.
2. Test learner on unlabeled data pool U and calculate confidence value for each class for each word.
3. Compute confidence interval (CI) between the two most probable classes for each word.
4. Make list of the sentence which contains the word having CI below the threshold value

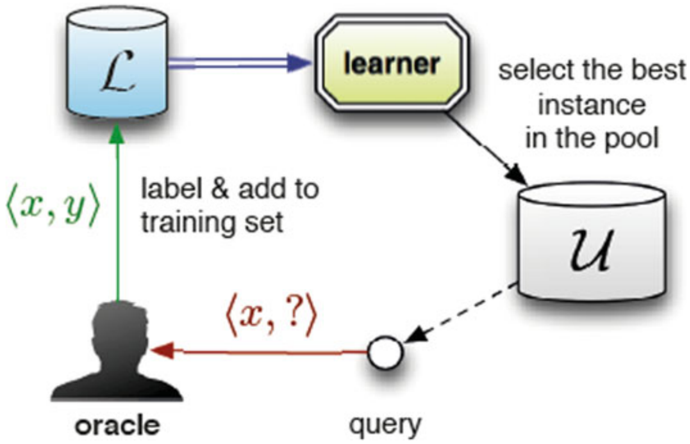


Fig. 1 Work flow of the pool based active learning approach. *Source* [17]

5. Pick top 10 sentences having word with lowest CI to be labelled by oracle.
6. Add the labelled sentences to set L and repeat step 2 until there is no sentence having word with CI value below threshold.

6 Result and Discussion

For comparison of the performance of the active learning based algorithm with the supervised learning algorithm, we have first trained the supervised CRF classifier over the train set carrying 75% of sentences and tested the result over the test set which is shown in Table 1.

For the active learning approach, the margin is calculated for each word using top two probability values for the classes. The word having minimum margin is the word for which the model is most uncertain. The active learning based approach is experimented using different threshold values for the margin including 0.1, 0.2, 0.3, and 0.4. The active learning algorithm is trained repeatedly using various features and CRF classifier and as discussed in earlier sections. The active learner is stopped

Table 1 Performance of the supervised CRF base classifier over test set of the corpus

Labels	Precision	Recall	F1-score
B-ADVERSE	0.679	0.333	0.447
I-ADVERSE	0.644	0.273	0.384
B-DISEASE	0.679	0.448	0.540
I-DISEASE	0.801	0.673	0.732
Average	0.707	0.457	0.547

Table 2 Performance of the CRF based active learner which uses margin query sampling with different margin thresholds over the test set of corpus

Margin threshold value	Total iterations before stopping	Total number of sentences used for training	Percentage of total sentence required for training (%)	F1-score
0.1	68	745	22.78	0.517
0.2	77	835	25.54	0.525
0.3	82	885	27.06	0.539
0.4	92	985	30.13	**0.558**

from training when there is no sentence left with word having margin less than the respective threshold value of the experiment. The final results obtained for the active learning algorithm using margin query sampling for each of the threshold values is reported in Table 2. We have experimented till 0.4 threshold value because it achieves f1-score of 0.558 (highlighted in bold in Table 2) which is greater than f1-score 0.547 achieved by the supervised classifier.

It is clear from the result table presented in Table 2 and Fig. 2 that the performance of the active learning model increases with the increase in the margin threshold value. The main reason is that the active learning model with higher threshold value trains the model with more training data which includes sentences with most uncertain words. The active learning model with higher threshold value provides the cumulative result i.e. it contains all the training examples that were earlier selected by the lower margin threshold values based active learning models in addition to the some more examples that have margin lower than the current higher threshold value and hence its performance is more. This thing can be observed clearly in Fig. 2 which shows the plot of the performance of active learning approach using margin sampling with different threshold values of margin.

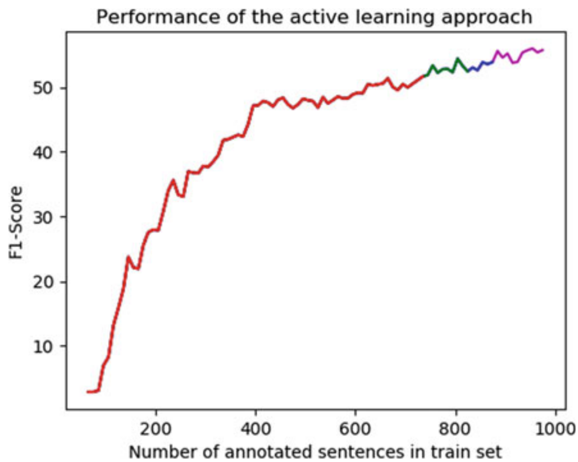


Fig. 2 F1-score versus number of annotated sentence in train set for active learning approach using margin sampling with different threshold values of margin

Table 3 Performance of the CRF based active learning approach with margin sampling having 0.4 margin threshold over the test set of corpus

Labels	Precision	Recall	F1-score
B-ADVERSE	0.685	0.342	0.456
I-ADVERSE	0.690	0.288	0.406
B-DISEASE	0.681	0.467	0.554
I-DISEASE	0.780	0.689	0.731
Average	0.710	0.472	0.558

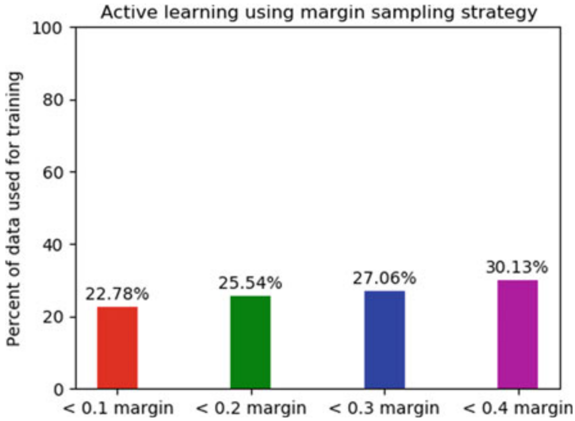


Fig. 3 Percentage of training sentences used by active learning approach based on margin sampling for different margin threshold values

Note that we have used same seed for random number generation so that the results are reproducible for each of the experiment and there is same setting for different experimental runs. It can also be seen in Tables 2 and 3 that the active learning based approach with margin threshold 0.4 outperforms the supervised approach while using only 30.13% of the training sentences. Total sentences used by active learning approach for training of active learner can be seen in Table 2 and Fig. 3.

7 Conclusion

In this research work, the active learning approach is used and compared with the supervised machine learning approach. For the experiment, various common features are extracted and the Conditional Random Field (CRF) classifier is used as a base classifier. The active learning approach uses margin sampling as query strategy with different margin confidence threshold values. The steps for the active learning approach are discussed in detail. The higher margin threshold tend to perform better as they train active learner over more training sentences which are selected using margin uncertainty sampling. Also, the active learner with margin sampling strategy with 0.4 threshold outperforms the supervised model while only using 30.13% of the sentences from the original training set for training. Thus, it is clear from this research work that in place of supervised learning algorithm, the active learning approaches can be readily adopted and applied for the task such as named entity recognition, which will require very less training data and will make no compromise in performance.

References

1. Nothman J, Ringland N, Radford W, Murphy T, Curran JR (2013) Learning multilingual named entity recognition from Wikipedia. Artif Intell 194:151–175
2. Angluin D (1988) Queries and concept learning. Mach Learn 2:319–342
3. Cui B, Lin H, Yang Z (2009) Uncertainty sampling-based active learning for protein–protein interaction extraction from biomedical literature. Expert Syst Appl 36:10344–10350
4. Chen Y, Lasko TA, Mei Q, Denny JC, Xu H (2015) A study of active learning methods for named entity recognition in clinical text. J Biomed Inform 58:11–18
5. Lee S, Song, Y, Choi M, Kim H (2016) Bagging-based active learning model for named entity recognition with distant supervision. In: 2016 International conference on big data and smart computing (BigComp), pp 321–324
6. Ekbal A, Saha S, Sikdar UK (2016) On active annotation for named entity recognition. Int J Mach Learn Cybernet 7:623–640
7. Han X, Kwoh CK, Kim J (2016) Clustering based active learning for biomedical named entity recognition. In: 2016 International joint conference on neural networks (IJCNN), pp 1253–1260
8. Kholghi M, Sitbon L, Zuccon G, Nguyen A (2017) Active learning reduces annotation time for clinical concept extraction. Int J Med Inform 106:25–31
9. Van Tran C, Nguyen TT, Hoang DT, Hwang D, Nguyen NT (2017) Active learning-based approach for named entity recognition on short text streams. In: Zgrzywa A, Choroś K, Siemiński A (eds) Multimedia and Network Information Systems. Springer International Publishing, Cham, pp 321–330
10. Tran VC, Nguyen NT, Fujita H, Hoang DT, Hwang D (2017) A combination of active learning and self-learning for named entity recognition on Twitter using conditional random fields. Knowl-Based Syst 132:179–187
11. Shen Y, Yun H, Lipton ZC, Kronrod Y, Anand kumar A (2017) Deep active learning for named entity recognition. CoRR. abs/1707.0
12. Fang M, Li Y, Cohn T (2017) Learning how to active learn: a deep reinforcement learning approach. CoRR. abs/1708.0
13. Gurulingappa H, Klinger R, Hofmann-Apitius M, Fluck J (2010) An empirical evaluation of resources for the identification of diseases and adverse effects in biomedical literature. In: 2nd Workshop on building and evaluating resources for biomedical text mining (7th edition of the language resources and evaluation conference) Valetta, Malta
14. Loper E, Bird S (2020) NLTK: the natural language toolkit. In: Proceedings of the ACL-02 workshop on effective tools and methodologies for teaching natural language processing and computational linguistics vol 1, Association for Computational Linguistics, Stroudsburg, PA, USA, pp 63–70
15. Korobov M, sklearn-crfsuite docs. https://sklearn-crfsuite.readthedocs.io/en/latest/tutorial.html
16. Yang L, Zhou Y (2014) Exploring feature sets for two-phase biomedical named entity recognition using semi-CRFs. Knowl Inf Syst 40:439–453
17. Settles B (2012) Active learning. Synth Lect Artif Intell Mach Learning 6:1–114

Design of QoS and Energy Efficient VM Consolidation Framework for Cloud Data Centers

Neha Songara and Manoj Kumar Jain

Abstract The virtualization and Virtual Machine (VM) Consolidation are the effective solutions for energy consumption reduction. The VM consolidation algorithms designed recently shows the efficiency in terms of energy consumption, Service Level Agreement Violations (SLAV) etc. The methods still suffered from the challenges to achieve the trade-off between QoS and Energy efficiency at VM consolidation. The design of VM consolidation should be efficient in terms of VM's allocation, migration, and workload management. In this paper, we proposed the design of novel VM consolidation approach in cloud data center to address the challenges of recent methods. First the Multi Resources based (CPU, RAM, network Traffic, bandwidth etc.) dynamic Overload Decision Algorithm (MR-ODA). For best VM selection, we used the policies of CloudSim framework. After selection of VMs to migrate, the efficient VM placement algorithm proposed based on optimization method called Particle Swarm Optimization (PSO). The algorithms designed to achieve the minimize energy consumption, QoS, reliability with minimum SLAV. The outcome of this paper is the new framework for VM consolidation to optimize the CloudSim tool.

Keywords Cloud computing · CloudSim · Data centers · Energy consumption · Physical machines · Virtual machine consolidation · Virtualization

1 Introduction

As the IT and non-IT organizations trend has been transformed from conventional to recent development for on demand services and resources provisioning from the available resource pool using well-known framework called cloud computing. Cloud computing consists of large number of data centers all over the world, however the

N. Songara (✉)
Research Scholar, Computer Science, Mohanlal Sukhadia University, Udaipur, India
e-mail: nehasongara.11@gmail.com

M. K. Jain
Professor, Computer Science, Mohanlal Sukhadia University, Udaipur, India
e-mail: manoj@cse.iitd.ernet.in

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_19

data centers becomes the big source of electricity consumption and hence significant emissions of unhealthy carbon dioxide (CO_2). It is essential for cloud service provider (CSP) to minimize the power and energy consumption at such data centers. This can be performed by minimizing the wastage of electricity and by improving the infrastructure of the data center. The management of resources inside data centers such as their allocation and utilization are also responsible for data center's inefficiency. Therefore, recent advancement in the resource management results significant improvement in the efficiency of data centers. For the improvement of energy inefficient data centers, it is very essential to understand how the power is distributed among the various components present inside data center. According to the report of EPA on data center's energy [1], the power consumption of a server is 40% of the data center's power and 80% of the total IT load. Similarly, the survey from open compute project [2] reported that 91% of energy consumption with in data center is only due to its computing resources. Thus, the source of high emission of CO2 is also due to inefficient usage of computing resources. Data provided in [3] claims that most of the times, server's works for more than 15–50% of their overall capacity which results over provisioning of the resources and leads to the higher Total Cost of Acquisition (TCA) [4]. According to National Resource Defense Council's report [5, 6], mostly the data centers are unused and the underutilization of servers is also inefficient for energy aware data centers. Solution of this problem is the consolidation of servers, by which fewer number of hosts will run the same application with lesser power consumption and it will further reduce the overall energy. Along with this, during the low server utilization, idle servers consume 70% of the power. Therefore, these idle servers should be turn off to reduce the energy consumption. Another important reason for energy wastage inside the data center is the lack of standard metrics. So, there should be some energy efficient metrics for servers so that servers can be arranged according to their energy efficiency in cloud data centers.

For energy efficiency in cloud computing data centers, there are three power saving methods such as VM Consolidation, powering down the servers, and Dynamic Voltage Frequency Scaling (DVFS). Automatic switching off or powering down the idle servers which are not in use is a very interesting method to reduce the total energy consumed. Most of the times, many servers remain in the idle position inside data centers. Thus, the dynamic provisioning of the selection of these types of servers and put them to sleep mode became a very challenging task. Various methods presented for dynamically turn off and turn on the servers inside data centers have been proposed to minimize the energy consumption [7–10]. The technique of dynamic VM consolidation introduced later on for the energy consumption minimization. The VM consolidation became useful approach for the server's selection that should be power down or power up. As the process of VM consolidation works in different steps, therefore, it is the key technique for the selection of most efficient servers. The VM consolidation process makes the use of less numbers of servers because it's VM live migration technique migrates the VMs from one physical machine (PM) to another and power down the underutilized servers. There are number of techniques proposed for efficient VM consolidation in cloud data center recently [11–17], but the concerns with challenges such as trade-off between energy efficiency and service (QoS) performance, live migration overhead, only based CPU utilization, un-optimized process for VM selection and placement etc.

Generally, the VMs those are running the application required to communicate efficiently with each other because of workload dependencies. Thus, effective monitoring of such VM's communications and accordingly efficient allocation of VMs to enough servers may lead to reduction in the overhead and energy consumption in cloud data center. We propose the new design for QoS and energy efficient VM consolidation framework based on the original CloudSim [18] framework of VM consolidation by contributing the algorithms for overload detection, Optimization based VM selection to migrate, and Optimization based VM placement. The proposed framework considers energy efficiency as well as QoS performance. In Sect. 2, we present the brief review of different methods for VM consolidation. In Sect. 3, we present the original CloudSim framework, its VM consolidation steps, and problems. In Sect. 4, the design of proposed VM consolidation framework presented with key contributions definition. In Sect. 5, the conclusion and future work presented.

2 Related Works

Recently the methods presented for efficient VM consolidation in cloud data center in terms of either energy efficiency or both energy and QoS efficiency [11–17].

In [11], two algorithms such as Improved Underloaded Decision (IUD) and Minimum Average Utilization Difference (MAUD) algorithm designed to achieve the energy and service efficiency in cloud data center.

In [12], first approach based multi-resources for multiple usage prediction for energy consumption minimization proposed. The prediction of future utilization performed based on multiple usages related to the horizon employed and resource types.

In [13], energy efficient VM placement algorithm designed to achieve the energy efficient VM consolidation in cloud data center based on prediction approach. The linear weighted method used to predict host load and classify hosts according to predicted host load for the VMs migration.

In [14], VM consolidation algorithms designed for cloud data center in category of dynamic decision making with load prediction with aim of improving the resource utilization and reduce the energy consumption with lower SLAV related to RAM, CPU, and bandwidth.

Very recently, in [15], the problem related to VM selection and allocation solved by residual available capacity (RAC) model. The method to achieve the trade-off between energy efficiency and QoS performance called energy-efficient and QoS dynamic VM consolidation (EQVC) method in [16]. In [17], the SLA aware and energy aware VM consolidation approach proposed in cloud data center. The algorithm for host overloading and underloading detection designed first, then VM placement algorithm proposed based on robust simple linear regression model.

Our work is different from above methods in which we consider both energy efficiency and service efficiency by using all resources inputs such as CPU, RAM, Network Traffic, Bandwidth etc. To achieve the service reliability, we designed multi-resources-based overload decision algorithm, then VM selection algorithm according to such metrics before their migration. Later, the VM placement needs to be optimizing rather than using some fuzzy logic based or fine-tuned approaches. The optimization methods (metaheuristic) are the effective ways to solve such problems.

In [19], Hybrid Genetic Algorithm (HGA) used for the energy efficient VM placement problem on PMs with communication network consideration in data centers.

In [20] the Ant Colony Optimization (ACO) based approach proposed. The multi-dimensional bin-packing developed in order to place VMs into the minimum number of PMs required for the current workload in cloud data center.

In [21], three objective functions proposed to apply multi-objective mimetic to address the VM placement problems. In [22], the Genetic Algorithms (GA) such as GA I and GA II designed for VM placement for energy minimization in cloud data center.

The above metaheuristic based VM placement solutions achieves the energy efficiency, but suffer from problems like more execution time and operational costs. In [23], particle swarm optimization (PSO) was designed to reallocate VMs in overloaded PMs and to consolidate underloaded PMs for energy efficiency. However, the VMs placement does not considered in this study. Further PSO based approach explored efficiently in [24] to allocate the many kinds of resources and to consolidate VMs across multiple data centers with minimum computation efforts.

In this work, we motivated from work reported in [24] and propose the PSO based VM placement algorithm by considering the multiple parameters.

3 CloudSim VM Consolidation

The conventional simulation-based tool for VM consolidation designed called CloudSim [18]. As our proposed VM consolidation method is based on CloudSim, in this section we present the working and problems of CloudSim.

- Host overloading detection
- VM Selection
- VM Placement
- Underloading host detection.

The hosts with minimum utilization frequently checked from the set of RUH. Finally, if all the VMs migrated, then PM will shut down to save power. The original CloudSim framework suffered from:

- The active node or host may have more states rather than just overloaded or under loaded.

- Original CloudSim framework considers objective of power consumption and hence may lead to imbalance data center.
- Imbalance data center leads to frequent VM migrations and hence degrades the QoS performance with poor energy efficiency.

The methods which we described in Sect. 2, are reported to solve above problems of CloudSim, but yet the challenges of VM consolidation are there for cloud computing data centers.

4 Proposed VM Consolidation System

Due to dynamic workload in data center, dynamical migration of VMs may result into the overloaded PMs, which resulting in server performance degradation with SLAV. Additionally, the VMs migrations should be controlled within specific amount. The frequent migrations in data center consume more bandwidth and affect the QoS of data center. Therefore, in this paper, we deal with the challenge of VM consolidation designing to satisfy the objective of *trade-off between energy efficiency and optimized service performance to meet Service Level Agreement (SLA).* The proposed VM consolidation framework is showing in Fig. 1 which is based on original CloudSim framework.

As observed in Fig. 1, the framework of proposed VM consolidation overlapping the key steps of CloudSim framework. However, the algorithms for overloaded detection, underloaded detection, VM selection, and VM placement are newly designed to improve the reliability, energy efficiency and QoS of data center. The key contributions to this framework are elaborated below.

4.1 Novel Load Detection Algorithm

We propose the Multi-Resource based overload decision algorithm (MR-ODA) which is used to determine whether a PM is overloaded or not and finally get a set of overloaded hosts by considering the different resources such as CPU utilization, RAM, bandwidth, network traffic, and storage capabilities. This is the first time; the multiple resource inputs are exploited to decide overload behavior of PMs. The use of multiple resources helps to boost the performance and reliability of VM consolidation framework. All available PMs are divided into two main groups overloaded and underloaded. The overloaded PMs are either idle or overloaded. We address the constrain of just two states of PMs, as we divide the underloaded hosts further in three sub states such as Severe Load (SL), Moderate Load (ML), and Low Load (LL) according to above set of resources using the Multi-Resource based underload decision algorithm (MR-UDA). When an underloaded host is at HL state, it has a high workload and is easy to become overload after accepting migrated VM.

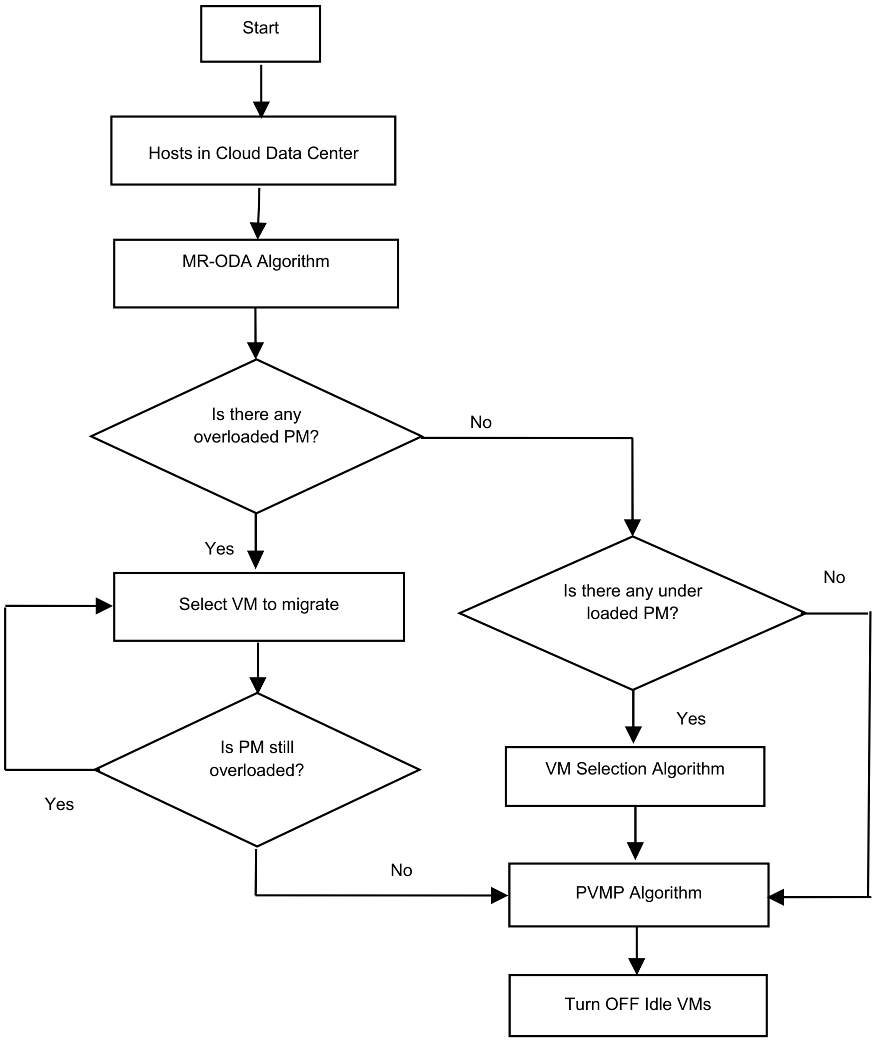


Fig. 1 Proposed VM consolidation framework in cloud data center

An underloaded host at LL state means that it has a low workload. Lastly, ML state is the intermediate state between HL and LL. The above sub-states of underloaded PMs helps to prevent the frequent migrations as we focus on minimum workload-based hosts for migration.

4.2 VM Selection Algorithm

For the overloaded hosts, one or more VMs on it should be selected to migrate out. The policies for VM selection are used to choose VM to migrate out from an overloaded host. Repeat this process until the host is not overloaded and finally obtain the list of VMToMigrate. We used the CloudSim VM selection policies in this framework.

4.3 PSO Based VM Placement

The VM placement is the task of mapping the VMs to hosts (PMs) in way that efficient utilization of hosts achieved. After the VM placement, the unused PMs will be shut down to save the power based on the load conditions. The selection of VM placement algorithms mainly depends on cloud user and CSP needs. The generalized algorithm is challenging task for VM placement. To address the efficient placement of VMs, we proposed the PSO based VM Placement Algorithm (PVMP) in this paper. For the VMs in the list of VMsToMigrate, we sorted them according to the input resources considered such as CPU utilization, BW, RAM, Storage etc. Then using the PSO algorithm the placement of VM performs to achieve the reliability and energy efficiency with minimum SLA violations.

Detailed algorithmic representation of proposed phases is out of scope for this paper, and will be part of our future work.

5 Conclusion and Future Work

The goal of this paper is to present the current problems associated with recent VM consolidation techniques as well as original CloudSim architecture. According to this, we present the initial methodology and contributions to overcome the challenges discussed. We presented the novel contributions towards the overload and underload decision phase in order to achieve the reliability and efficiency during the VM consolidation in cloud data center. Later in VM selection algorithms explored in proposed framework as well. For efficient VM placement, we propose the meta-heuristic approach in which the advantage of PSO technique used to estimate the best solution for VMs placement. For future works we will work on:

- Mathematical design and practical evaluations of MR-ODA and MR-UDA techniques with CloudSim framework.
- Mathematical design and practical evaluations of PVMP algorithm with CloudSim framework.
- Complete proposed framework implementation and evaluations with CloudSim framework and recent methods.

References

1. Report to congress on server and data center energy efficiency, Environmental Protection Agency (online). www.energystar.gov/ia/partners/prod_development/downloads/EPA_Datacenter_Report_Congress_Final1.Pdf. Accessed on 21 Nov 2016
2. Open Compute Project, "Energy efficiency" (online). http://opencompute.org/about/energy-efficiency/. Accessed on 3 Feb 2017
3. Barroso LA, Holzle U (2007) The case for energy-proportional computing. Computer 40(12):33–37
4. Fan X, Weber WD, Barroso LA (2007) Power provisioning for a warehouse-sized computer. In: Proceedings of the 34th annual international symposium on computer architecture (ISCA), pp 13–23
5. Natural Resources Defense Council (online): http://www.nrdc.org/energy. Accessed on 12 Jan 2017
6. Scaling up energy efficiency across the data center industry: evaluating key drivers and barriers, NRDC. http://www.nrdc.org/energy/files/data-center-efficiency-assessment-IP.pdf, accessed on 12/2/2017 (online)
7. Orgerie AC, Lefevre L (2011) ERIDIS: Energy-efficient reservation infrastructure for large scale distributed systems. Parallel Process Lett 21(2):133–154
8. Zhang Q, Zhani MF, Zhang S, Zhu Q, Boutaba R, Hellerstein JL (2012) Dynamic energy-aware capacity provisioning for cloud computing environments. In: ACM International conference on autonomic computing, ICAC, pp 145–154
9. Guenter B, Jain N, Williams C (2013) Managing cost, performance and reliability trade-offs for energy-aware server provisioning. In: Proceedings. of the 30th annual IEEE international conference on computer communications (INFOCOM), pp 1332–1340
10. Chen G, He W, Liu J, Nath S, Rigas L, Xiao L, Zhao F (2013) Energy-aware server provisioning and load dispatching for connection-intensive internet services. In: Proceedings of the 5th USENIX symposium on networked systems design and implementation, pp 337–350
11. He K, Li Z, Deng D, Chen Y (2017) Energy-efficient framework for virtual machine consolidation in cloud data centers. China Commun 14(10):192–201
12. Nguyen TH, Francesco MD et al (2017) Virtual machine consolidation with multiple usage prediction for energy-efficient cloud data centers. IEEE Trans Serv Comput
13. Zhou Z, Abawajy J, Chowdhury M et al (2017) Energy-efficient virtual machine consolidation algorithm in cloud data centers. J Cent South Univ 24(10):2331–2341
14. Khoshkholghi MA, Derahman MN, Abdullah A, Subramaniam S, Othman M (2017) Energy-efficient algorithms for dynamic virtual machine consolidation in cloud data centers. IEEE Access 5
15. Chang Y, Gu C, Luo F, Fan G, Fu W (2018) Energy efficient resource selection and allocation strategy for virtual machine consolidation in cloud datacenters. IEICE Trans Inf Syst E101–D(7):1816–1827
16. Liu Y, Sun X, Wei W, Jing W (2018) Enhancing energy-efficient and QoS dynamic virtual machine consolidation method in cloud environment. IEEE Access 6:31224–31235
17. Li L, Dong J, Zuo D, Wu J (2019) SLA-aware and energy-efficient VM consolidation in cloud data centers using robust linear regression prediction model. IEEE Access 7:9490–9500
18. Calheiros RN, Ranjan R, Beloglazov A, De Rose CA, Buyya R (2011) CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. Softw Pract Exp 41(1):23–50
19. Tang M, Pan S (2014) A hybrid genetic algorithm for the energy-efficient virtual machine placement problem in data centers. Neural Process Lett 41(2):211–221
20. Feller E, Rillin, L, Morin C (2011) Energy-aware ant colony-based workload placement in clouds. In: 12th IEEE/ACM international conference on grid computing, pp 26–33
21. Lòpez-Pires F, Barán B (2013) Multi-objective virtual machine placement with service level agreement: A memetic algorithm approach. In: Proceeding of the IEEE/ACM 6th international conference on utility and cloud computing, pp 203–210

22. Adamuthe AC, Pandharpatte RM, Thampi GT (2013) Multiobjective virtual machine placement in cloud environment. In: International conference on cloud & ubiquitous computing & emerging technologies, pp 8–13. IEEE
23. Dashtia SE, Rahmani AM (2015) Dynamic VMs placement for energy efficiency by PSO in cloud computing. J Exp Theor Artif Intell 28:1–16
24. Chou L, Chen H et al (2018) DPRA: dynamic power-saving resource allocation for cloud data center using particle swarm optimization. IEEE Syst J 12(2):1554–1565

Cyberbullying Detection on Multiple SMPs Using Modular Neural Network

Chinmay Patil, Shreya Salmalge and Purva Nartam

Abstract As reported in recent statistics, the rate of cyberbullying on social media platforms(SMPs) is increasing daily along with affected population in the teenage and young generation worldwide. There is an urgent need to curb cyberbullying online and help curb its adverse effects on young minds and their psychology. Deep Learning in Artificial Intelligence can prove to be a benefactor in detection of such instances. The existing cyberbullying detection techniques attempt to detect either similar type of bullying or target only one social media platform. The proposed model, hence, will attempt to curb both the given bottlenecks. We propose a model that detects varied cyberbullying instances on multiple SMPs, essentially using individual detection models of three social media platforms Wikipedia, Formspring, Twitter synergically, to make a single prediction.

Keywords Cyberbullying detection · LSTM · Neural networks · Modular neural network

1 Introduction

The necessity of building a model for cyberbullying detection can be stated from the known facts about the victims of cyberbullying in the youth generation due to humungous spread in technology and text-based anonymous communication modes on social media platforms. Existentially, cyberbullying cannot be prevented on run-time unless the victim reports about the bully. Therefore, to curb the bottlenecks of detection of cyberbullying instances on social media platforms with a large global population on different ways of cyberbullying with respect to caste, race, gender,

C. Patil (✉) · S. Salmalge · P. Nartam
Department of Computer Engineering, B.R.A.C.T's Vishwakarma Institute of Technology, Pune, India
e-mail: chinmay.patil16@vit.edu

S. Salmalge
e-mail: shreya.salmalge16@vit.edu

P. Nartam
e-mail: purva.nartam16@vit.edu

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_20

religion, country of origin, color of skin, personal attacks, abuse, offensive/brutal language and other such instances, there is a need for a cyberbullying detection model that encompasses both targets of targeting multiple social media platforms and detecting various types of cyberbullying.

It has been observed from recent studies that, a total of 37% of parents across India said their child was bullied online, with 14% of that total saying the bullying occurred on a regular basis. In the United States, over a quarter of parents said their child had also experienced the problem and that's up from just 15% in 2011. Cyberbullying rates are much more mixed in Europe with 17% of parents in the UK confirming it along with 14% in Italy and 9% in both Spain and France. Interestingly, the phenomenon seems to be virtually non-existent in Russia with no parents citing instances of cyberbullying. In the 2016 study, it did exist in Russia but the rate was quite low at just 9%.

Cyberbullying is a sensitive concern that involves sentiments expressed in the form of words either directly addressing an individual or a community as a whole. As noted in this experiment, Wikipedia datasets contained instances targeting personal attacks; Twitter datasets contained instances regarding racism, sexism i.e. community-centric offensive language. For clarity, models trained with Wikipedia dataset cannot detect instances of sexism, racism/community-centric bullying instances like '**Feminism is a weapon for women**'. Similarly, Twitter-trained models cannot detect personal attacks like '**If stupidity were a crime, you would end up in jail**'. A language model can predict the probability of the next word in the sequence, based on the words already observed in the sequence. Neural network models are a preferred method for developing statistical language models because they can use a distributed representation where different words with similar meanings have similar representation and because they can use a large context of recently observed words when making predictions. The proposed model is a prediction model which can be achieved by means of machine learning, deep learning using LSTMs neural networks. Long-Short Term Memory (LSTM) is a type of Recurrent Neural Networks (RNN). It takes sequences of information and uses recurrent mechanisms and gate techniques. LSTM has many advantages over other feed forward and recurrent neural networks in modeling of time series as in [1, 2]. Features of words in sentences are forwarded to LSTM models in the form of word embeddings [3]. Word embedding is the collective name for a set of language modeling and feature learning techniques in natural language processing (NLP) where words or phrases from the vocabulary are mapped to vectors of real numbers corresponding to unique factors/attributes.

A modular neural network is an artificial neural network characterized by a series of independent neural networks moderated by some intermediaries. Each independent neural network serves as a module and operates on separate inputs to accomplish some subtask of the task the network hopes to perform [4]. By means of modular neural networks, the intermediary will take outputs of each module and processes them to generate output of the network as a whole. Hence, modular neural networks [5] is a feasible construct for the proposed model. The models will not interact with each other (Fig. 1).

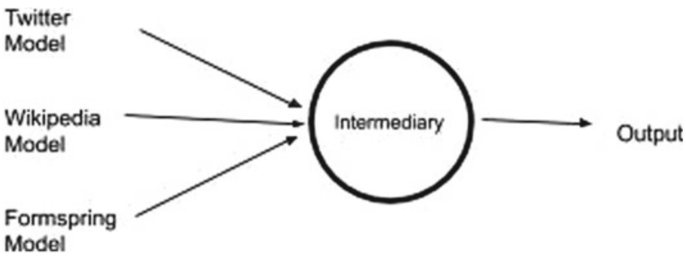


Fig. 1 Basic modular architecture

In order to combine the major features of individual models into a synergic model, three independent Wikipedia-trained, Twitter-trained, Formspring-trained models are merged together generating a combined effect on the complete structure and functionality of the model.

2 Related Work

Use of Social media platforms has increased with increased use of internet, cell phones and launching of new social media platforms. There has also been an increase in cyberbullying on SMPs. Interpretation of the message as instance is subjective. Though researchers have worked to come up with the models that will detect cyberbullying instances. Many of these models are trained on dataset from single social media platform. Some of them use datasets from multiple SMPs and combine them in one dataset to train a single model. These methods also involve pre-processing steps such as stemming, lemmatization, etc. Some machine learning methods also use handcrafted features such as swear words.

Several machine learning methods such as logistic regression, Support Vector Machine, Decision trees as well as Deep learning methods. are used for cyberbullying detection task [6].

3 machine learning methods were discussed—(1) Support vector machine with linear kernel trained on datasets from 3 different platforms. (2) J 48—Decision Tree algorithm. It divides data into classes by processing it through decision tree. It uses Twitter corpus. (3) Naive Bayes—Simple conditional probabilistic classifier which uses Bayes theorem with Naive independence assumptions between different features. According to various methods in [6], SVM is better in performance in comparison with other 2 methods.

Deep learning methods are proven to be better than machine learning methods in many tasks. The paper [7] investigate deep learning methods in detection of hate speech in tweets. It proposes 3 architectures—Convolutional Neural Networks, Long Short Term Memory(LSTM) Neural networks, LSTM with Gradient Boosted Decision Tree (GBDT). CNN and LSTM architectures use pre-trained GloVe word embeddings whereas LSTM with GBDT uses random embeddings. LSTM + GDBT gives best results as stated in [7].

Transfer learning is a technique where a model trained for one task is re-purposed on second related task [8]. It uses transfer learning in 2 ways—(1) Feature Level Transfer Learning where network is trained on one dataset and learned word embeddings are transferred to another dataset to train new model. (2) Model Level Transfer Learning where network is trained on one dataset and learned word embeddings along with learned network weights are transferred to another dataset to train new model. 3 datasets are used for these tasks—Formspring, Wikipedia, Twitter. These transfer learning methods give better result than individual models trained on single dataset. [9] reproduces and validates the models proposed in [8] and their results. Researchers are also working on unsupervised methods such as Self Organizing Maps [10], Clustering, etc. for detection of cyberbullying which are reported to be effective.

3 Proposed Model

3.1 Dataset Preprocessing

The considered datasets of the three SMPs were cleaned by removing unnecessary characters like punctuation marks, extra spaces, numbers and other such entities which possess no value in the training process. The model is trained in batches using mini-batch gradient descent algorithm. Therefore, input sentences in one batch are required to be of a fixed length, in our case 50. Sentences with length less than 50 were padded with <'PAD'> vectors; greater than 50 were truncated to size 50 words. Each dataset consists of 2000 negative (non-cyber bullying) and 750 positive (cyber bullying) samples.

The dataset was modified such that the positive and negative samples in one dataset aligned with positive and negative samples in other datasets, i.e. they shared the same index in the array. This is a necessary modification when training the entire network with 3 input datasets as input, and a single output. A single labels array is used, with outputs aligned with the positive and negative samples from each dataset.

3.2 Architecture

The modules of the model were trained independently on their respective datasets as

Wikipedia model—Wikipedia dataset; Twitter model—twitter dataset; Formspring model—Formspring dataset. Each module learns features specific to its dataset.

The architecture of all the above 3 models was the same—An input layer, followed by a custom SSWE (Sentiment Specific Word Embeddings) embedding layer, a LSTM layer with return sequences set to true, another LSTM layer with return sequences set to false, a dropout layer with probability = 0.5, and finally a Dense layer with 2 classes using Softmax Activation.

The reason we chained two LSTMs in the above models is that we need the hidden states of the LSTM at every time step. The first LSTM layer returns the hidden state for each input time step, then separately, the hidden state output and the cell state for the last input time step. We combine these states to make a single prediction. The hidden state outputs from the first LSTM layer of every model is retrieved using keras backend functions.

As mentioned above, the LSTM layer returns 3 hidden states in the form of arrays: The LSTM hidden state output for each input time step, the LSTM hidden state output for the last time step and the LSTM cell state for the last time step. Every type of hidden state from each model is sent as a combined input to a custom keras layer. There are three custom keras layers, one for each type of hidden state. Each layer combines the 3 inputs (one from each model) to output a single hidden state. These are A first layer to combine Hidden state output at every time step, A second layer to combine Hidden state output for the last time step, A third layer to combine cell state for the last time step.

So we obtain 3 new hidden states, each combined from the outputs of each model. These states act as the input to a new LSTM layer. We define a new LSTM layer that takes as input the combined hidden state output for each time step, and its hidden states are initialized with the combined hidden state output for the last time step and the combined cell state. This LSTM layer acts as the intermediary in our model.

The output of the first custom layer, is connected to the LSTM layer. This layer is initialized with the hidden state output and the cell state, i.e. the outputs of the second and third custom layer respectively. The LSTM layer is then connected to a dropout layer with probability = 0.5 followed by a dense layer with two numerical classes using Softmax Activation.

Before the intermediary and all the layers after it are trained, first the three modules are trained on their respective datasets. A single dataset is formed by stacking each individual dataset, i.e. $2750 \times 3 = 8250$ samples. Similarly, the labels are stacked. This dataset is then sent as input to each module to generate the input for the intermediary layer, on which it is trained. The model is fit on the combined labels, using Adam optimizer and Binary cross entropy loss function.

4 Experiments and Results

We experimented with 2 variants of modular architecture-one where we used pre-trained models to generate inputs for our main prediction model, i.e. the intermediary and second where we trained the independent modules from scratch along with the main prediction model.

To test these models, last 2500 samples each from the Wikipedia and Twitter dataset were combined to form a single test dataset. This combined dataset tries to simulate a social media platform with bullying instances of every type. Due to lack of samples in the Formspring dataset, it wasn't considered for testing.

Each of these individual models were first tested on this test dataset, to set a benchmark. The results were as follows:

Wikipedia model	71.1%
Twitter model	63.9%
Formspring model	71.0%

In each of the below modular architectures, we try to obtain an accuracy higher than each of the independent modules. The main prediction model is expected to learn to look at the predictions of each independent module, analyze them and make a single final prediction.

Proposition 1 This is the main model explained in the above section, where pre-trained independent models generate the inputs for the intermediary. During testing the test dataset is sent through every model to generate hidden states and these states are sent to the intermediary for prediction. This model achieved the highest accuracy in all our models of 73.4% which is higher than the accuracy of the independent modules, showing the success of combining modules to make a single prediction.

Another variation of this model is combining only the hidden state output of the last time step and the cell state for the last time step from each model, and initializing the hidden states of another LSTM with these states. This LSTM is then trained on all three datasets one by one. This variation is a combination of Transfer Learning with Modular Networks (Fig. 2).

Proposition 2 This is a 3 input 1 output single model. The independent modules are trained along with the intermediary. The model has three input layers, one for each

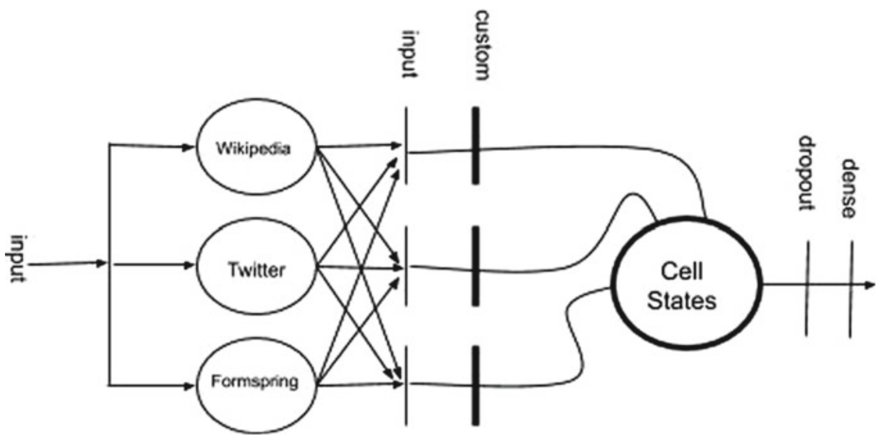


Fig. 2 Modular architecture for method 1

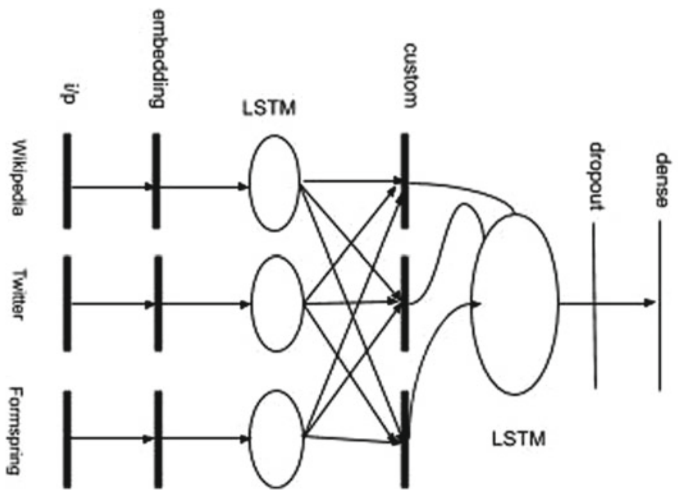


Fig. 3 Modular architecture for method 2

dataset. All the next layers for each module are the same with—input layer followed by a custom SSWE layer, followed by a LSTM layer. Every type of hidden states from each model is combined using custom keras layers. The rest of the architecture is same as explained above, with outputs of custom layers connected to a LSTM layer followed by a dropout and dense layer. The entire model is then trained with 2750 samples in each dataset and a single output array with 2750 labels. Every model reads a different input sample from its respective dataset, but weights of all three modules get updated at once, as there is a single output layer. During prediction, every test sample is sent parallel across all three models, before the states are combined to make a single prediction.

This model gave an accuracy of 67.2%.

Another variation of this approach is to preload weights into the layers of the independent modules, before starting the training of the whole network (Fig. 3).

5 Conclusion

In conclusion, the proposed modular architecture was successful in achieving a higher accuracy than each independent module. The model successfully learned to combine output predictions of the three individual models to make a single final prediction. Hence, the goal of creating a single model to predict cyberbullying across multiple social media platforms was achieved. Discovering other variants of the modular architectures and improving hyper-parameters of the model to increase its existing accuracy can be progressed. Further, when we get more datasets with more number of cyberbullying instances of different types of cyberbullying it will help the model learn better and more variant features. The modular architecture can also be used in other fields of machine learning where there is a need to combine predictions from multiple independent modules to make a single prediction.

Furthermore, data cleaning and refining the input datasets which prominently consists of slang words or misspelled words, not present in the dictionary of embedding matrix, should be done. Slogans in the form of slang words or hashtags are considered as unknown tags, however, these parameters form a major part in cyberbullying impacts.

Defining inputs require truncating or padding of input sentences to a fixed sized length of batches. This can, therefore, affect the overall performance of the model or even lead to improper results. Thus handling of variable length input sentences by using masking layers should help increase accuracy.

References

1. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
2. Understanding LSTM networks. https://colah.github.io/posts/2015-08-Understanding-LSTMs/
3. Introduction to word embedding and word2vec. https://towardsdatascience.com/introduction-to-word-embedding-and-word2vec-652d0c2060f_a
4. Azam F (2000) Biologically inspired modular neural networks. Ph.D. dissertation (PDF). Virginia Tech
5. Happel BLM, Murre JMJ (1994) The design and evolution of modular neural network architectures. Neural Netw 7:985–1004
6. Sugandhi R, Pande A, Chawla S, Agrawal A, Bhagat H (2015) Methods for detection of cyberbullying: a survey. In: IEEE, Dec 2015
7. Badjatiya P, Gupta S, Gupta M, Verma V (2017) Deep learning for hate speech detection in tweets. In: 26th International world wide web conference
8. Agrawal S, Awekar A (2018) Deep learning for detecting cyberbullying across multiple social media platforms. IIT Guwahati
9. Dadvar M, Eckert K, Cyberbullying detection in social networks using deep learning based models; a reproducibility study. Web-based Information Systems and Services, Stuttgart Media University Nebenstrasse 8, 70569 Stuttgart, Germany
10. Di Capua M, Di Nardo E, Petrosino A (2016) Unsupervised cyberbullying detection in social networks. In: 2016 23rd International conference on pattern recognition (ICPR)

Lightweight Two-factor Authentication Protocol and Session Key Generation Scheme for WSN in IoT Deployment

Ankit Attkan and Priyanka Ahlawat

Abstract In recent years, Wireless sensor networks (WSN) are extensively exploited in terms of their enormous applicability and are widely used from mere public branches like hospitals, banks, Institutes to defense and research wings of the country. Pointing that it can contain public or private (sensitive) information or both correspondingly which means authentication among user, nodes and gateways is a must based on a trust metric(s) or a common cryptographic key. WSNs act as a beehive of storming information that is continuous or timed periodically. An unsecure network is highly prone to outsider as well as insider attacks and then here comes the role of authentication protocols that act as a shield of defense because traditional security methods are not directly applicable on IOT's WSNs due to increasing heterogeneity of components. In this paper, We revisit the authentication protocol provided by Amin and Biswas in 2016 and we propose an improved authentication protocol based on two factor namely rabin cryptosystem and hashing that aim to provide enhanced security features with secure common session key generation and address some of the previous un-addressed attacks.

Keywords Sensor networks · IoT · Authentication · Authorization · Key distribution

1 Introduction

Wireless sensor networks (WSNs) are small piece of what IoT collaborates to the world of networking, smart objects that are sensors, sending and receiving data via a communication channel. A communication scheme is useless and unsecured if it doesn't maintains properties namely "Confidentiality"—which is violated when Unauthorized user gets protected information, "Integrity"—which is violated if an Unauthorized user modifies information and "Availability"—which is violated if system

A. Attkan (✉) · P. Ahlawat
Department of Computer Science and Engineering, NIT Kurukshetra, Kurukshetra 136119, India
e-mail: ankitzvma@gmail.com

P. Ahlawat
e-mail: priyankaahlawat@mail.nitkkr.ac.in

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_21

is not available for service to the authorized user. Sensors are resource constrained in terms of calculating power, memory, bandwidth, battery etc. Huge number of devices in WSN cannot be handled by IPv4 alone, so we needed IPv6 with wide range of IP addresses. Ipv6 demands heavy load of power and thus lightweight and ultra lightweight protocols like ZigBee [1] or 6LowPAN and hash based ECC—protocols are being used. An Unsecured WSN is full of vulnerabilities and loop holes that act as an opportunity for an adversary to exploit the same. Attacks possible may be active or passive in nature. Several problems related to WSNs have been already addressed but this doesn't means that new ones won't arise in future. Some of the major attacks in WSNs are sensor capture attacks [2], smart card forgery attack [3] where two factor authentication is used, replay attacks, denial of service attacks, session hijacking, user forgery attack, gateway forgery attack [4], Man in the middle attack etc. Major reasons for the cause of attacks on the networks were an inefficient algorithm for communication that failed to authenticate the communicating nodes.

Our Contribution: We are motivated with the work proposed in three major contributions by A. K. Das's three factor scheme, Jiang and Ma's Crypto secured three factor bio hashing based scheme & Amin and Biswas's hashing based on common session key generation which is lightweight. We propose an improved two factor rabin cryptosystem and hashing based protocol which includes not all but certain features from those protocols mentioned earlier. It provides enhanced security features with secure common session key generation as well as address some of the previous un-addressed attacks. We used Prover if cryptographic protocol verifier to prove the security features of our protocol. The Paper is organized as follows: In Sect. 2, discusses related work so far, Revisiting Amin and Biswas's protocol are subject of Sect. 3, In Sect. 4, we introduce our proposed protocol, while Sect. 5 gives analysis of security and compares the existing protocols and concludes this paper.

2 Related Work

In year 2009—Watro and Das deduced a two factor authentication protocol built on RSA cryptosystem. Watro et al. [5] gave an RSA and Diffie helmann scheme based on public key encryption/decryption which is a user authentication scheme but large key size is difficult to compute. In 2010—Khan's and Algathbar's [6] scheme came into picture but it doesn't provides mutual authentication among sensors. Wong et al. [7] presented a protocol that utilizes the hash function but still remains vulnerable to stolen-verifier attack. A password based scheme for user authentication was driven by Ashok kumar Das [8], which was incapable to address Denial of Service attacks and sensor node captures attacks. In 2013—Xue et Al. [9] gave their lightweight but being lightweight, its vulnerability to even weak attacks like offline Brute force attack was elevated. All these schemes do not provide any mutual authentication among WSN's components. Xu and Wang et al. [10] improved Das's [11] algorithm by introducing temporal credential based mutual authentication protocol. In 2014-Turkanovic et Al. [12] and 2016-Farash et Al. [13] gave a general scheme in which sensors worked in between user and server putting a heavy toll on battery power of the sensors. Jiang et al. in 2015, again derived a password-based authentication of user based scheme. Jiang et al. [14] also pointed that Xue's [15] method was vulnerable to Brute force offline guessing and User tracking attacks. Dynamic adding of sensors is an issue when it comes to Jiang's scheme [14]. Wu et al. in 2015 [16] proposed an enhanced scheme and its security was proven through provable security methods. 2016, Amin and Biswas proposed a hash based two-factor scheme for authentication which happened to be suffering from inefficient sensor registration. Huang et al. [17] introduced a multi-factor scheme based on smartcard, User-password and biometrics also known as 3-factor authentication scheme. According to Wang et al. [18], Huang's scheme is prone to stolen smart card attack, smart card forgery attack. Ma et al. [19] showed that a two-factor authentication scheme can be upgraded to a more secure Three-factor authentication protocol. Efficient algorithm working now days are hash based. Hash based algorithms [20–22] generally work on the basis of salts also known as nonces [23] generated at sensor, gateways and user side. Still these kind of algorithms are found prone to sensor capture attacks, User forgery attacks, gateway forgery attacks and even session hijacking, in which a sensor which was captured is used to compute the common secret key of sensor, gateway and user to access the data exchanging channel of the same [24, 25]. Its in Amin and Biswas's [23] algorithm which was proposed in 2016. To overcome the previous problems and flaws of earlier algorithms [26], a recent algorithm was proposed in 2017 by Wu, Xu, Das, Saru, Jian, Raymond, Wazid [27, 28]. In June 2017, Ahlawat et al. [29] proposed a scheme that asses the vulnerability of network towards adversarial attacks by building an attack matrix. It is shown that this matrix can be effectively hold in countering attacks. Whether it is symmetric key cryptosystem or asymmetric key cryptosystem, Even if a key is generated, Sharing the key secretly among components of a network without the adversary knowing it is an issue of security.

3 Revisiting Amin and Biswas's Algorithm

The previous protocol: There are 7 phases that constitute the scheme. It has initialization/system setup, user registration, sensor node registration, login and key agreement phases. (i) System setup and initialization phase, (ii) Sensor node registration, (iii) User registration, (iv) Login phase for ith user, (v) Authentication and key agreement phase, (vi) Password change phase and (vii) Session termination phase. For further details refer Amin and Biswas's (2016) [23].

4 Proposed Protocol

Our protocol utilizes the rabin cryptosystem and concept of random nonces to obtain enhanced security features that lack in the preceding protocol by Amin and Biswas.

Table 1 Notations used

Ui	User
xi	Key of the ith USer
Sj	SN
SAIH	System administrator, initiator and handler
SCID	Smart card number
GN	Intermediate gateway node
UPWi	U's password
UIDi	U's identity
UIDj	S's identity
Xgn	Gateway node's secret key
Xj	Sensor node's secret key
Ki, Kj	Random numeric values chosen by Ui and Sj, respectively
T	Timestamp
Fi	Ui's formal random number template
ΔT	Allowed time lag
SECK	Session key generated in the protocol
h()	One-way hash algorithm
FF()	The two-factor two way hashing formal function algorithm
||	Concatenation
⊕	XOR operation performed bitwise

Rabin cryptosystem is a public key based on integer factorization procedures. It has key generation phase where u, v prime numbers are selected to compute $N = u * v$. Encryption is $C = (m * m) \bmod N$ and (u, v) is the private key pair and N is the public key, whereas decryption is done by under rooting the cipher text and modulus by N. Table 1 depicts the Notations used in this paper and describes what these notations mean.

4.1 System Setup and Initialization Phase

Step 1 SAIH generates two primes, "u" and "v" to calculate $N = u * v$, and preserves (u, v) as private key and N as the public key. SAIH also selects X_{gn}, key for the gateway node and an integer "l" as the parameter for the verifier based on fuzzy logic.

Step 2 SAIH chooses a UID_j and computes the key $X_j = h\left(UID_j || X_{gn}\right)$ for every S_j $(1 < j < m)$.

Step3 SAIH arbitrarily chooses a number R which is shared between gateway node and S_j (jth sensor node). At last, S_j saves <UID_j, X_j, R> in its memory.

4.2 User Registration Phase

In our User registration phase, ith User chooses UID_i and credentials that are sent to SAIH via secure channel. At SAIH, necessary parameters α_i and β_i are computed, saved into database and a smartcard containing the information is issued to the user when necessary input of vector [UID_i, UPW_i, URN] is fed and the final contents stored in the smartcard contains $\mathbf{C_i}$, $\mathbf{f_i}$, $\mathbf{g_i}$, **SCID**, **l**, **n**, $\mathbf{r_i}$, **FF() and h()**, whereas $\boldsymbol{\alpha_i}$ and $\boldsymbol{\beta_i}$ parameters are deleted for security reasons.

4.3 Login Phase

(User logins successfully only if smartcard and credentials are available):

Step 1 U_i inserts the smartcard and inputs the identity UID_i', password UPW_i', and nonce F_i. Then, the smartcard calculates $F_i^* = FF(r_i, URN)$ and $C_i^* = h\left(h\left(UID_i^* \| UPW_i^* \| F_i^*\right) \bmod l\right)$. The smartcard discards U_i's login request if $C_i{*} \neq C_i$.

Step 2 The smartcard generates a random nonce K_i and a timestamp T_1, computes $\alpha_i^* = f_i \oplus h\left(UID_i' \| UPW_i' \| F_i'\right)$, $\beta_i^* = g_i \oplus h(UID_i^* \oplus UPW_i^* \oplus F_i^*)$, $M_1 = (UID_i \| SCID \| K_i)^\wedge 2 \bmod n$, $M_2 = h(\alpha_i * \| \beta_i * \| K_i || T_1)$.

Step 3 U_i chooses the identity UID_j of the sensor node that the user wants to access, then the smart card computes $CUID_j = UID_j \oplus h(UID_i \| K_i \| T_1)$ and sends $MSG_1 = \langle M_1, M_2, T_1, CUID_j \rangle$ to Gateway node.

4.4 Authentication and Session Key Generation Phase

Step 1 On receiving MSG_1 from U_i, Gateway node decrypts M_1 using "u" and "v" computed earlier to obtain UID_i', $SCID'$, K_i', then obtains x_i as per the UID_i' & checks whether $SCID'$ matches the value in the record. If there is no matching, then Gateway node discards the request and aborts the process; otherwise, Gateway node computes $\beta' = h\left(SCID' || X_{gn}\right)$, $\alpha' = h\left(UID_i' \| X_{gn} \| x_i\right)$, $K_i' = M_2 \oplus h\left(\alpha' || T_1\right)$, & $M_2' = h\left(\alpha' \| \beta' \| K_i' || T_1\right)$. Gateway node terminates the current session if $M_2' \neq M_2$; else, Gateway node calculates $UID_j' = CUID_j \oplus h(UID_i \| K_i \| T_1)$, $X_j' = h\left(UID_j' || X_{gn}\right)$,

$M_3 = h\left(UID'_i \| UID'_j \| ID_{gn} \| X'_j \| K'_i || T_2\right)$, $M_4 = UID'_i \oplus h\left(ID_{gn} \| X'_j \| T_2\right)$, $M_5 = K_i \oplus h\left(ID'_i \| ID'_j \| X'_j || T_2\right)$ & then sends $MSG_2 = <ID_{gn}, M_3, M_4, M_5, T_2>$ to S_j.

Step 2 S_j verifies whether $|T_3 - T_2| \leq \Delta T$ holds, where T_3 is the present timestamp. If found invalid, S_j immediately discards the session; else, it computes $UID''_i = M_4 \oplus h(ID_{gn} \| X_j \| T_2)$, $K''_i = M_5 \oplus h(UID''_i \| UID_j \| X_j || T_2)$, $M''_3 = h(UID''_i \| UID_j \| ID_{gn} \| X_j \| K''_i || T_2)$. S_j terminates the connection if M_3"! = M_3; else, it accepts that U_i and Gateway nodes are authentic. Next is, S_j calculates $SECK_i = h(UID''_i \| UID_j \| K''_i || K_j)$, $M_6 = h(SECK_j \| X_j \| K_j || T_3)$, and $M_7 = K''_i \oplus K_j$, where K_j is the randomly generated numeric value by S_j. Finally, S_j sends the $MSG_3 = <M_6, M_7, T_3>$ to Gateway node.

Step 3 Gateway node verifies whether $|T_4 - T_3| \leq \Delta T$ holds true, where T_4 is the present timestamp. If it is negative, Gateway node discards the session; else it computes $K'_j = M_7 \oplus K'_i$, $SECK_{gn} = h\left(ID'_{gn} || UID_j \| K'_i \| K'_j\right)$, & $M'_6 = h\left(SECK_{gn} \| X'_j \| K'_i || T_3\right)$. Gateway node aborts the present session if $M_6' \neq M_6$; else, it calculates $M_8 = h\left(SECK_{gn} \| UID'_i \| \alpha' || K'_j\right)$. At Last, gateway node sends $MSG_4 = <M_7, M_8>$ to U_i.

Step 4 U_i calculates $K_j = M_7 \oplus K_i$, $SECK_i = h\left(UID_i \| UID_j \| K_i || K'_j\right)$, also $M'_8 = h\left(SECK_i \| UID_i \| \alpha || K'_j\right)$. U_i discards the current session if M_8'! = M_8; else, U_i grants access that gateway node and S_j are legitimate. Finally, a session key $SECK_i = SECK_{gn} = SECK_j$ has been generated among U_i, S_j and GN.

5 Analysis of the Proposed Protocol

5.1 Prevents All Session Hi-Jacking Due to Sensor Capture Attack

Sensors are physical devices that can be personally tampered by the attacker via using hardware tools. If a sensor is compromised and attacker has physical access to the sensor node, the information like sensor ID, sensor's secret key, sensor's configuration might be revealed. If a common symmetric key is being used, sensor's key will be the same as gateway's and the user's key, So, a complete session can be hijacked by the attacker. Figure 1 shows properly how sensor capture attack occurs. Using separate salts/nonces to generate separate keys for every communicating channel between a user interacting with the sensor, where gateway node acts as the intermediate can avoid complete system's key pool being compromised, However, that one session, between captured sensor and the gateway, user its interacting with remains

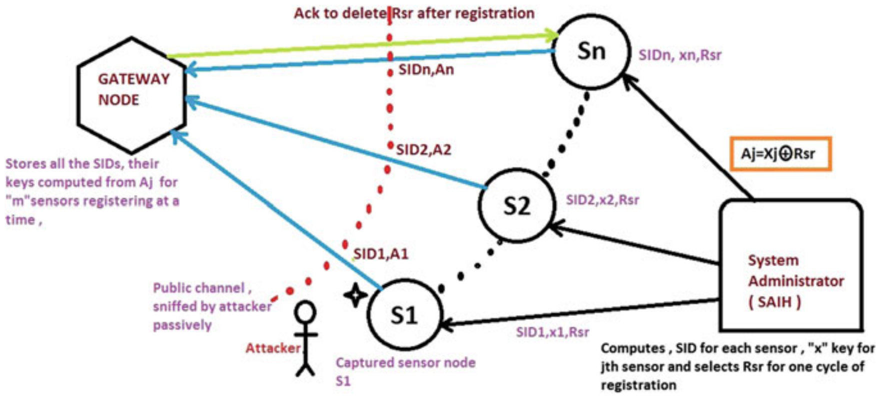


Fig. 1 Sensor capture attack in Amin and Biswas's algorithm [23]

slave of the attacker serving limited information to the attacker. Restarting the session again can fix the capturing of node to some extent. Discarding that sensor is another choice. Figure below shows how sensor capture attack in Amin & Biswas's Algorithm can be executed to hi-jack all the sessions henceforth compromising all the sessions. For instance say attacker captures sensor node 1 out of the three sensor nodes deployed. Attacker obtains system administrator to secretly share the Sensor IDs. Now, R_{sr} is common for all the sensors trying to register at a particular time, a registration cycle. Attacker side by side sniffs the network for the data packets being exchanged. Due to lack of security features, the random number obtained by capturing on sensor node can be used to exclusively OR with the various data packets captured via sniffing. Now, The packets that belong to the sensors that were registered in the same registering cycle will have the same crypto-reversing R_{sr}, which may expose the sensor IDs and private keys resulting in session hi-jacking. Once, the SIDs and keys are leaked, Sensor forgery attacks are possible using the IDs and keys which can create a severe problem for network. Our proposed protocol prevents this attack as for every group of sensors trying to register at a particular period of time, the SAIH generates a unique nonce for every sensor node using fuzzy verifier, instead of a common shared nonce R_{sr}. It leads to the feature that even if a sensor is captured, the complete Hijacking of the session of the group of sensor nodes trying to register at a particular having a common R_{sr} earlier is not possible. Only the session of the captured sensor node gets compromised in this case.

5.2 Preventing Gateway Impersonation Attack [23, 27]

In our method, the Adversary is unable to impersonate as Gateway node to either U_i or S_j. To impersonate as Gateway node to S_j, It needs to calculate a protected value $M_3 = h\left(UID_i \| UID_j \| ID_{gn} \| X_i \| K_j || T_2\right)$. However, without this value $X_j' = h\left(UID_j' || X_{gn}\right)$, it

Table 2 Comparison analysis

S. No.	Immunity against attack	Amin and Biswas (2016)	Our protocol (2019)
1	Type II SSLA attack	Secure	Secure
2	Gateway impersonation attack	Insecure	Secure
3	Sensor node impersonation	Insecure	Secure
4	Sensor node capture attack	Insecure	Secure

is impossible for Adversary to calculate M_3. Since, we use the hash based algorithm & timestamps, Adversary cannot obtain any utilizable information from the messages from the previous communicated authentication sessions. In general, to impersonate as Gateway node to either User U_i or S_j, Adversary needs to compute a protected value $M_8 = h(SECK_{gn} \| UID_i \| \alpha_i || K_j)$. Adversary needs to have information of K_i to compute the value $SECK_{gn} = h(UID_i \| UID_j \| K_i || K_j)$. To obtain K_i, Adversary has to know the secret key u and v of Gateway node. It is not feasible because the secret key is protected by the System administrator-initiator and handler very carefully.

5.3 *Avoiding Sensor Node Impersonation Attack [27]*

Adversary tries to impersonate as S_j after capturing the messages transferred during the previous communicated authentication protocol phase sessions. Adversary has to compute $MSG_3 = <M_6, M_7, T_3>$ to impersonate as S_j, where $SECK_j = h(UID_i \| UID_j \| K_i \| K_j)$, $M_6 = h(SECK_j \| X_j \| K_j || T_3)$, & $M_7 = K_j'' \oplus K_j$. Adversary must know K_i in aim to calculate $M_6 = h(SECK \| X_j \| K_j \| T_3)$. It is very similar to the analysis of the gateway impersonation attack analyzed earlier, So, Adversary is incapable of obtaining K_i. Adversary cannot execute the SN impersonation attack (Table 2).

6 Conclusion and Future Scope

One of the main Aim of WSN is to transmit the sensitive information over a network in a secure and efficient manner. To achieve this, a strong authentication protocol is the basic requirement. WSN are placed in hostile environment thus making them vulnerable to many attacks. Thus, to make the defensive mechanism effective, we should consider the possible attacks on it. The proposed protocol makes use of rabin cryptosystem and random nonces and fuzzy logic to secure channel, prevent session hijacking and add probabilistic randomness respectively. It is shown using prover if protocol verifier tool that proposed protocol is immune to Gateway impersonation, sensor node impersonation attacks and almost preventive against the sensor node capture attack. In future we intend to improve authentication scheme by integrating the voice command and speech recognition functionality to access the sensors, for the user authentication phase and same can be used to mutually authenticate user and gateway node among one another. This will provide enhanced security as multiple levels of abstraction and security will provide better hindrance to the adversary.

References

1. Gartner report from. https://www.gartner.com/en/newsroom/press-releases/2017-02-07-gartner-says-8-billion-connected-things-will-be-in-use-in-2017-up-31-percent-from-2016
2. Wu F, Kumari S, Li X, Shen J, Choo R, Wazis M, Das Ak (2017) An efficient authentication and key agreement protocol for multi-gateway wireless sensor networks. J Netw Comput Appl 89:72–85
3. Kumari S, Om H (2016) Authentication protocol for wireless sensor networks applications like safety monitoring in coal mines. Comput Netw 104:137–154
4. Das AK, Zeadally S, He D (2018) Taxonomy and analysis of security protocols for internet of things. Futur Gener Comput Syst 89:110–125
5. Watro R, Kong D, Securing sensor networks with public key technology, In: Proceedings of the 2nd ACM workshop on security of ad hoc and sensor networks, ACM, pp 59–64
6. Khan MK, Alghathbar K (2010) Cryptanalysis and security improvements of two-factor user authentication in wireless sensor networks. Sensors 10:2450–2459
7. Wong KHM, Zheng Y, Cao J, Wang S (2006) A dynamic user authentication scheme for wireless sensor networks. IEEE International conference on sensor-networks, ubiquitous & trust-worthy computing, p 8
8. Das ML (2009) Two-factor user authentication in Wireless sensor network IEEE Trans. Wireless Commun 8:1086–1090
9. Xue K, Ma C, Hong P (2013) A temporal-credential-based mutual authentication and key agreement scheme for wireless sensor networks. J Netw Comput Appl 36:316–323
10. Li X, Ma J, Wang W, Xiong Y, Zhang J (2013) A novel smart card and dynamic ID based remote user authentication scheme for multi-server environments. Math Comput Model 58(1):85–95
11. Das AK (2012) A random key establishment scheme for multi-phase deployment in large-scale distributed sensor networks. Int J Inf Secur 11(3):189–211
12. Turkanovic M, Brumen B, Holbl M (2014) A novel user authentication and key agreement scheme for heterogeneous ad hoc wireless sensor networks, based on the internet of things notion. Ad Hoc Netw 20:96–112
13. Farash MS, Turkanovic M, Kumari S, Holbl M (2016) An efficient user authentication and key agreement scheme for heterogeneous wireless sensor network tailored for the internet of things environment. Ad Hoc Network 36:152–176
14. Jiang Q, Ma J, Lu X, Tian Y (2015) An efficient two-factor user authentication scheme with unlinkability for wireless sensor networks. Peer-to-peer Netw. Appl. 8:1070–1081
15. Wu F, Xue L, Kumari S, Li X (2015) A novel and provably secure biometrics-based three-factor remote authentication scheme for mobile client-server networks. Comput Electr Eng 45:274–285
16. Wu F, Kumari S, Li X, Shen J, Choo R, Wazis M, Das Ak (2017) An efficient authentication and key agreement protocol for multi-gateway wireless sensor networks. J Netw Comput Appl 89:72–85

17. Huang X, Xiang Y, Bertino E, Zhou J, Xu L (2014) Robust multi-factor authentication for fragile communications. IEEE Trans Dependable Secur Comput 568–581
18. Yu J, Wang G, Mu Y, Gao W (2014) An efficient generic framework for 3-factor authentication with provably secure instantiation. IEEE Trans Inf Forensics Secur 9(12):2302–2313
19. Li X, Ma J, Wang W, Xiong Y, Zhang J (2013) A novel smart card and dynamic id based remote user authentication scheme for multi-server environments. Math Comput Model 58:85–95
20. Srivastava K, Awasthi AK, Kaul SD, Mittal R (2015) A hash based mutual RFID tag authentication protocol in telecare medicine information system. J Med Syst 39(1):153
21. Phan RC-W (2009) Cryptanalysis of a new ultra lightweight RFID authentication protocols. IEEE Trans Dependable Secur Comput 6(4):316–320
22. Ouafi K, Phan RC-W (2008) Traceable privacy of recent provably-secure RFID protocols. In: International conference on applied cryptography and network security, Springer, pp 479–489
23. Amin R, Biswas G (2016) A secure light weight scheme for user authentication and key agreement in multi gateway based wireless sensor networks. Ad Hoc Netw 36:58–80
24. Ammar M, Russello G, Crispo B (2018) Internet of things: a survey on the security of IoT frameworks. J Inf Secur Appl 38:8–27
25. Chi Q, Yan H, Zhang C, Pang Z, Da Xu L (2014) A recon_gurable smart sensor interface for industrial WSN in IoT environment. IEEE Trans Industr Inf 10(2):1417–1425
26. Wang D, Wang P (2014) On the anonymity of two-factor authentication schemes for wireless sensor networks: attacks, principle and solutions. Comput Netw 73:41–57
27. Wu F, Kumari S, Li X, Shen J, Wazis M, Das AK (2017) An efficient authentication and key agreement protocol for multi-gateway wireless sensor networks. J Netw Comput Appl 89:72–85
28. Adat V, Gupta B (2018) Security in internet of things: issues, challenges, taxonomy, and architecture. Telecommun Syst 67(3):423–441
29. Ahlawat P, Dave M (2017) A hybrid approach for path vulnerability matrix on random key predistribution for wireless sensor networks. Wirel Pers Commun 94(4):3327–3353

An Empirical Evaluation of Correlation Based Feature Selection for Tweet Sentiment Classification

Sanur Sharma and Anurag Jain

Abstract This paper presents a study on Twitter sentiment analysis where tweets are gathered and sentiments behind the tweet are evaluated by using various machine learning techniques. It presents an empirical evaluation of correlation based feature selection for sentiment classification on twitter data. The data is extracted from twitter in real time and text preprocessing and feature extraction is applied on the textual data. Correlation based attribute selection methods are used and machine learning classifiers (SVM, Naïve Bayes, Random Forest, Meta classifier, SGD, Logistic Regression) are compared on various performance parameters to show which classifier gives better results. The results show that when STWV with Attribute Selection methods are used together in the same setup, the classifiers give accuracy between 78 and 88% with about 0.88 true positive rate and 0.15 false positive rate which is far better when no attribute selection method is used.

Keywords Twitter analysis · Sentiment analysis · Correlation based feature selection · Machine learning techniques

1 Introduction

This Social media today has become a tool that manifests user's opinion on a public platform where anyone can share and interact with one another. Twitter is one such social media platform where users share their opinions about any topic or person openly. Twitter contains enormous amount of posts containing short messages which presents a valuable source of people's opinion. Businesses are using twitter posts to understand user opinions about their products and services. Sentiment analysis has become an integral part in analysis of textual data where user opinions can be categorized as positive, negative and neutral. Recent work in the field of twitter

S. Sharma (✉) · A. Jain
USICT, Guru Gobind Singh Indraprastha University, Delhi, India
e-mail: sanursharma@gmail.com

A. Jain
e-mail: anurag@ipu.ac.in

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_22

sentiment analysis revolves around analysis of twitter data to understand the sentiments and emotions of social media users [1]. In [2] Huang et al., presented twitter sentiment analysis and implemented three models—Naïve Bayes, Logistic Regression and SVM. In [3] two methods based on multi nominal Naïve Bayes algorithm have been proposed which gives better accuracy then existing methods. A unified framework based hybrid classification scheme is presented that incorporates four classifiers: Slang classifier, emotion classifier, sentiment word net classifier and a domain specific classifier. This method for tweet classification presents better results in terms of accuracy [4]. A study on twitter sentiment analysis is presented where tweets are gathered in real time and lexicon based and machine learning techniques are implemented to evaluate the sentiments of the social media text posted by the users [5]. Textual data contains humungous amount of information and extracting and preprocessing this data is difficult and time consuming. To extract useful information and make predictions from such data feature selection methods are used for sentiment classification.

The aim of feature selection methods is to remove redundant data. The feature selection methods used are correlation based and One-R [6–9]. One-R attribute selection method uses one rule for each feature and selects the rule with the smallest error [8, 10]. Correlation based feature selection methods creates a subset of features according to an evaluation function. The subset contains highly correlated features with the class and uncorrelated features are not considered [9, 11]. It can be categorized into two types: wrapper method and filter method [12, 13].

This paper presents an evaluation of correlation based feature selection methods that are used to extract only the relevant features form the data which is further analyzed for sentiment evaluation by using various machine learning techniques. The aim of this work is to compare the correlation based feature selection method with the OneR attribute eval Real time twitter data is extracted form twitter API and preprocessed to remove noisy and inconsistent data and vectorization is performed on that data, after which the feature extraction methods are applied to extract meaningful features and sentiment classification is done based on lexicon based and machine learning techniques.

2 Evaluation Methods and Model

The various steps involved in the evaluation model are dataset preparation, preprocessing and vectorization, feature selection and sentiment classification.

2.1 Dataset

The dataset is collected via twitter api where tweets were collected in real time. 'Cambridge Analytica' dataset is built by searching the keyword and tweets are collected

Table 1 Sentiment tweet data set summary

Dataset	Sentiment class	Tweets	Time span
Cambrige Analytica	Positive	1332	May 2018–June 2018
	Negative	2454	
	Neutral	3330	

for duration of one month. 7116 tweets are collected manually and sentiment classification is performed based on lexicon based and machine learning based techniques [5]. Table 1 presents the sentiment tweet data summary where tweets are classified into positive, negative and neutral.

Lexicon Dictionary: An opinion lexicon of 6789 words is used, out of which 4783 words are negative and 2006 words are positive [14].

2.2 Data Preprocessing

This step includes stemming, part of speech tagging, stop word removal, tokenization and vectorization. For classification features like unigrams, bigrams, n-grams, term presence and frequency measures like tf, tf-idf are used. For sentiment analysis lexicon based features are used like corpus based and dictionary based approaches are followed.

2.3 Feature Selection Methods

Feature selection methods extract subset of features from all the possible list of features in the dataset to present the prediction results. Feature selection methods works at two levels: first is selection of subset of attributes through an attribute evaluator algorithm and second is the evaluation of the search heuristics via a search algorithm [15]. Feature selection can be done through three ways:

- Directly applying the attribute evaluator and search algorithm on the dataset.
- Evaluating the performance of the set of attributes by using a specific classifier. This is also known as wrapper method.
- Selecting attributes as a filter in the pre-processing phase of data analytics.

Table 2 presents the evaluated feature method and subset searching method used for the evaluation of features.

Table 2 Evaluated feature method and subset searching method

	Ranker	Best-first
CorrEval	√	
cfsSubsetEval		√
OneREval	√	

Correlation Based Feature Selection Methods.

Correlation Attribute Evaluator. The correlation attribute evaluator measures the Pearson's correlation between the attribute and the class. A weighted average is calculated for nominal attributes on a value by value basis. The correlation is calculated between each attribute and the output variable. The attributes with high positive or negative correlation are selected and the ones having low values are not selected [12, 13]. Following is the standardized formula for Pearson's Correlation.

$$\rho_{\mathrm{A,B}} = \frac{Cov(A, B)}{\sigma_A \sigma_B} \tag{1}$$

CFS Subset Evaluator. The correlation-based feature subset evaluator ranks the subset of features based on a heuristic evaluation function. The subset features that are highly correlated with the classes but are not correlated with each other are considered and the rest are ignored. The redundant features and the overlapping features are not considered as they are highly correlated with other features. The feature acceptance depends on the prediction results that state whether the predicted instance is not already predicted by other features [9, 11, 16, 17]. Following is the CFS evaluation function:

$$Value_s = \frac{n\overline{C_{cf}}}{\sqrt{n + n(n-1)\overline{C_{ff}}}}, \tag{2}$$

where $\mathrm{Value_s}$ is the heuristic value of attribute subset 's' having n features, $\mathrm{C_{cf}}$ is the mean feature to class correlation, and $\mathrm{C_{ff}}$ is the average feature to feature correlation.

OneR Feature Selection. One-R feature selection was first proposed by Holte [8]. It is based on the one rule theory where each attribute in the training data has one rule and it selects the rule which has the smallest error rate. This method is used as a baseline performance evaluator for various other algorithms. This method divides the numerically valued features as continuous variable and breaks it into various disjoint intervals [7, 18, 19]. To predict class distribution based on attribute value, compute the most frequent class and evaluate the following rule:

$$\text{Class} \rightarrow \text{most Frequent Class} \tag{3}$$

$$\text{Rule} \rightarrow \text{Attribute} = \text{Value} \Rightarrow \text{Class} \tag{4}$$

2.4 Sentiment Classification

The main goal of analyzing sentiment is to analyze the text and examine the scores of sentiments in terms of the polarity (negative, positive and neutral). The sentiment analysis can be divided into many levels: word/term level, sentence level, document/message level and aspect level [20]. The sentiments can be characterized as subjective or objective. Objective characterization presents the factual information and subjective characterization deals with the polarity classification of the text which detects the polarity in terms of positive, negative, neutral and mixed. The opinion or sentiment types can further be classified as standard (which contains direct and indirect opinions) and comparative opinions that states the similarities and dissimilarities between objects under consideration. The nature of opinion can be explicit or implicit. Explicit opinions are subjective that expresses regular or comparative opinion. Implicit opinions on the other hand are objective and expresses regular and comparative opinions. The review structure of the sentiments can further be structured (formal text written by professionals), semi-structured or unstructured (informal text as found in reviews).

There are various sentiment challenges that arise in the analysis of social media data that are described in literature [21, 22]. A few of them are domain dependence which deals with topic oriented features, huge lexicon, Negation and extracting lexicon based features, POS tagging, Bag of words, term presence and frequency, opinion words and phrases and more. Finally, the sentiment analysis techniques include machine learning, Lexicon based and hybrid techniques.

3 Experimental Results

The experimental setup includes the dataset modelling, preprocessing and vectorization discussed in the previous section. The feature selection methods used in the evaluation model is evaluated and the statistics of the feature selection methods is shown in Table 3.

In general machine learning techniques like logistic regression, naive bayes, SVM give good prediction results for classification of sentiments. But in case where evaluation of sentiment classification is based only on positive and negative sentiments,

Table 3 Feature subset selection statistics

Subset evaluator	Search method	Number of features	Statistics			
			Min	Max	Mean	Std. dev.
Correlation	Ranker	1775	0	3.044	0.123	0.482
cfsSubset Eval	Best-First	53	0	0.525	0.206	0.162
OneREval	Ranker	1775	0	3.044	0.123	0.482

Table 4 Comparison between various classifiers for sentiment modelling without using attribute selection methods

String to word vec, tf-idf, n-gram				
Algorithm	Precision	Recall	F-measure	Accuracy
Naive bayes	77.3	65.1	51.6	65.1
SMO (SVM)	77.3	65.1	51.6	65.1
LR (simple logistic)	77.2	64.9	51.1	64.8
Bayes network	77.3	65.1	51.6	65.1
SGD	77.3	65.1	51.6	65.1
Meta attribute selected classifier	77.2	64.9	51.1	64.8
Bagging	77.3	65.1	51.6	65.1
LR	77.3	65	51.5	65
Random forest	77.3	65.1	51.6	65.1

logistic regression gives better results than Naive Bayes and other linear classifiers. Table 4 presents a comparison of various algorithms based on String to Word Vectorization (STWV) with tf-idf and ngram approach with no attribute selection method used. The classifiers run on 10-fold cross validation and the results are compared over nine classifiers. The results show that the accuracy of the various classifiers is close to 65% and there is no major variation among them. This method uses string to word vectorization with tf-idf and n-gram features and does not uses any attribute selection methods. The next section presents the attribute selection method and how it improves the accuracy of the model.

The method that has been used for the comparison of various classifiers is Correlation based feature selection. Correlation based features selects the attributes based on the ranking of subsets of features according to the correlation based heuristic function. It forms the subsets that has highly correlated features with the class and removes the redundant features. For the comparison among the classifiers we have used three attribute selection methods namely Correlation Attribute Eval, CFS Subset Eval and One R attribute Eval. Before applying these attribute selection methods, the dataset is preprocessed and string to word vectorization with tf-idf and n gram features are evaluated. Table 5 presents the comparison between various classifiers for sentiment modelling based on the attribute selection methods. The attribute selection methods vastly improve the performance results by providing better accuracy. On comparing the results from Tables 4 and 5 it can be seen that:

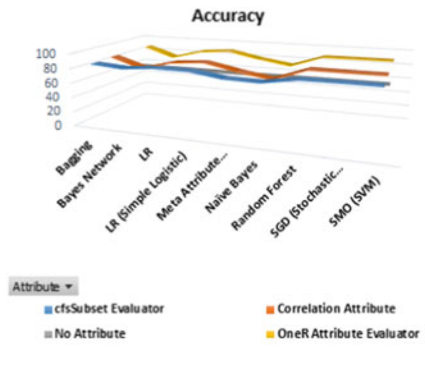
- Using only the STWV filter, we get an accuracy of the classifiers around 65% with about 0.65 true positive rate and 0.643 false positive rate.
- Using the STWV with Attribute Selection in the same setup, gives an accuracy of the various classifiers between 78 and 88% with about 0.88 true positive rate and 0.15 false positive rate.

In Table 5, on comparing the Correlation based attribute selection methods with the One R attribute evaluator which is used as a baseline performance benchmark

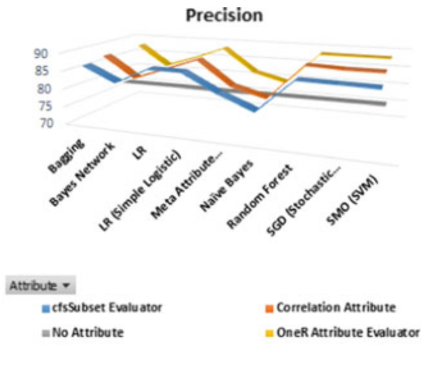
Table 5 Comparison between various classifiers for sentiment modelling based on the attribute selection methods

String to word vec, tf-idf, n-gram												
	Correlation attribute				CFS subset evaluator				OneR attribute evaluator			
Algo	P	R	F	A	P	R	F	A	P	R	F	A
Naive bayes	79	72	72.5	72	78.1	78.2	77	78.15	79	72	72.5	72
SMO (SVM)	88.1	88.2	88.1	88.2	86.4	85.4	84.7	85.4	88.1	88.2	88.1	88.2
LR (simple logistic)	88.1	88	87.8	88	86.6	85.6	84.8	85.5	88.1	88.1	87.8	88.1
Bayes network	81.5	73.1	73.6	73.1	82.3	82	82.1	81.9	81.5	73.1	73.6	73.1
SGD	88.1	88.1	88.1	88.1	86.6	85.6	84.9	85.6	88.1	88.1	88.1	88.1
Meta attribute	81.4	79	76.8	79	81.8	79.3	77	79.2	81.4	79	76.8	79
Bagging	87.4	87	86.5	86.97	86.3	85.3	84.5	85.2	87.5	87	86.6	87
LR	84.7	84.5	84.6	84.5	86.7	85.6	84.9	85.6	84.6	84.4	84.5	84.4
Random Forest	88.5	88.2	87.9	88.2	86.8	85.4	84.6	85.4	88.1	87.8	87.5	87.8

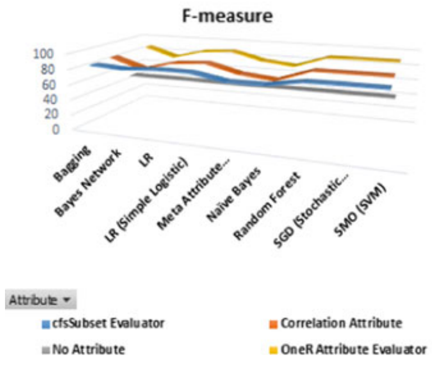
for other learning schemes. It was found that for classifiers—SVM, LR, SGD and Bagging, the correlation attribute evaluator gives similar results when compared to OneR Evaluator, but for random forest classifier the accuracy is improved with the correlation attribute. It gives an accuracy of 88.2 when compared to OneR which gives an accuracy of 87.8. Further for classifiers—Naive Bayes, Bayes Network, multinominal LR and Meta attribute selected classifier CFS subset evaluator gives better results than the other two. Figure 1a–d shows the various performance measures that compare the results of attribute selection methods like CFS subset evaluation method and correlation based evaluator with the OneR evaluator. The results show that the attribute selection methods perform better than the method that uses only STWV without any attribute selection.



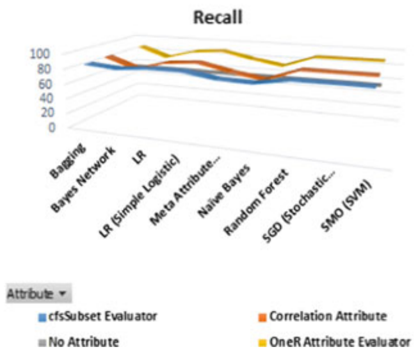
(a) Accuracy



(b) Precision



(c) F-measure



(d) Recall

Fig. 1 Classifier evaluation measures **a–d** accuracy, precision, recall and F-measure graphs of the various classifiers depicting the comparison between the various attribute selection methods and without any attribute selection

4 Conclusion

This study presents an evaluation framework for twitter sentiment analysis where tweets are collected in real time and the data is preprocessed and correlation based feature selection techniques are applied. A comparison between String to Word (STWV) vectorization without any feature selection method is compared with STWV with feature selection methods. The machine learning techniques that are applied on the dataset are: Naïve Bayes, SVM, LR, SGD, Meta attribute, Bagging, Random Forest. The results show that the STWV filter gives an accuracy of the classifiers around 65% and whereas STWV with Attribute Selection in the same setup, gives an accuracy of the various classifiers between 78% and 88%. On comparing the Correlation based attribute selection methods with the One R attribute evaluator which is used as a baseline performance benchmark for other learning schemes. It was found that for classifiers—SVM, LR, SGD and Bagging, the correlation attribute evaluator gives similar results when compared to OneR Evaluator, but for random forest classifier the accuracy is improved with the correlation attribute. It gives an accuracy of 88.2 when compared to OneR which gives an accuracy of 87.8. Further for classifiers—Naive Bayes, Bayes Network, multi-nominal LR and Meta attribute selected classifier CFS subset evaluator gives better results than the other two.

Acknowledgements This publication is an outcome of the R&D work undertaken project under the Visvesvaraya Ph.D. Scheme of Ministry of Electronics and Information Technology, Government of India, being implemented by Digital India Corporation and with the cooperation of GGSIP University.

References

1. Pak A, Paroubek P (2010) Twitter as a corpus for sentiment analysis and opinion mining. In: LREc. vol 10
2. Huang C-L, Hu Y-C, Lin C-H (2017) Twitter sentiment analysis. https://cseweb.ucsd.edu/classes/wi17/cse258-a/reports/a080.pdf. Accessed 20 Apr 2018 (Online)
3. Song J, Kim KT, Lee B, Kim S, Youn HY (2017) A novel classification approach based on na¨ıve bayes for twitter sentiment analysis. KSII Trans Internet Inf Syst (TIIS) 11(6):2996–3011
4. Asghar MZ, Kundi FM, Ahmad S, Khan A, Khan F (2018) Tsaf: twitter sentiment analysis framework using a hybrid classification scheme. Expert Syst 35(1):e12233
5. Sharma S, Jain A (2019) Cyber social media analytics and issues: a pragmatic approach for twitter sentiment analysis. In: Bhatia S, Tiwari S, Mishra K, Trivedi M (eds) Advances in computer communication and computational sciences. Advances in Intelligent Systems and Computing, vol 924. Springer, Singapore
6. Maurya S, Jain A (2019) Ranking based feature selection techniques for phishing website identification. In: IEEE international conference on machine learning and data science 2018 (in press)
7. Yildirim P (2015) Filter based feature selection methods for prediction of risk in hepatitis disease. Int J Mach Learn Comput 5(4)
8. Holte RC (1993) Very simple classification rules perform well on most commonly used datasets. Mach Learn 11:63–91

9. Hall MA (1999) Correlation-based feature selection for machine learning. In: Ph.D. Department of Computer Science, The University of Waikato, Hamilton
10. Novakovic J, Strbac P, Bulatovic D (2011) Toward optimal feature selection using ranking methods and classification algorithms. Yugoslav J Oper Res 21(1):119–135
11. Hall MA, Smith LA (1999) Feature selection for machine learning: comparing a correlation-based filter approach to the wrapper. In: Proceedings of the twelfth international Florida artificial intelligence research society conference, pp 235–239
12. Ashraf M, Chetty G, Tran D (2013) Feature selection techniques on thyroid, hepatitis, and breast cancer datasets. Int J Data Min Intell Inf Technol Appl (IJMIA) 3(1):1–8
13. Leach M (2012) Parallelising feature selection algorithms. University of Manchester, Manchester
14. Hu M, Liu B (2004) Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining ACM, pp 168–177
15. Zareapoor M, Seeja KR (2015) Feature extraction or feature selection for text classification: A case study on phishing email detection. Int J Inf Eng Electron Bus 7(2):60
16. Khonji M, Jones A, Iraqi Y (2013) An empirical evaluation for feature selection methods in phishing email classification. Int J Comput Syst Sci Eng 28(1):37–51
17. Bahl S, Sharma SK (2015) Improving classification accuracy of intrusion detection system using feature subset selection. In: 2015 Fifth international conference on advanced computing and communication technologies. IEEE, pp. 431–436
18. Bahl S, Sharma SK (2016) A minimal subset of features using correlation feature selection model for intrusion detection system. In: Proceedings of the second international conference on computer and communication technologies. Springer, New Delhi, pp. 337–346
19. Morariu D, Cretulescu R, Breazu M (2013) Feature selection in document classification. In: The fourth international conference in romania of information science and information literacy, ISSN-L pp 2247–0255
20. Liu B (2012) Sentiment analysis and opinion mining. Synth Lect Hum Technol 5(1):1–167
21. Medhat W, Hassan A, Korashy H (2014) Sentiment analysis algorithms and applications: a survey. Ain Shams Eng J 5(4):1093
22. Hussein DMEDM (2018) A survey on sentiment analysis challenges. J King Saud Univ Eng Sci 30(4):330–338

Automatic Traffic Sign Detection and Classification of Indian Traffic Signage's Based on Multi-feature Fusion

Shivanand S. Gornale, Ashvini K. Babaleshwar and Pravin L. Yannawar

Abstract Traffic Road sign detection and classification is a fundamental technology for building automated driver guidance system with the aid of computer vision technologies. Due to improper and irregular road signage's the accident are increasing day to day. Many practices are presented for identifying, recognition and classification of traffic signs from daces. To address this acute problem the study is conceded out for the detection and classification of signage's grounded on arrangement of matching and discriminative feature sets. The feature like LBP, HOG, and LPQ are used to extract the features. The Classification is carried using K-d Tree and KNN machine learning algorithms. The presentation of the anticipated technique is conceded on database which is created in heterogeneous climatic condition for Indian road systems. The tentative outcomes reveal that each feature yields high performance and the feature fusion gives good accompaniment and yielded good classification rate in competent with the prevailing algorithm.

Keywords Traffic signage's · KNN · K-d tree · Classification · Local binary pattern · Histogram of oriented gradients · Local phase quantization · Picture element

1 Introduction

Traffic road sign identification and recognition is a fundamental technology towards building automated driver management system by means of computer vision. Number of accidents is recounted annually due to inattention and non-ideal weather and

S. S. Gornale · A. K. Babaleshwar (✉)
Department of Computer Science, Rani Channamma University, Belagavi, Karnataka, India
e-mail: ashw.babaleshwar@gmail.com

S. S. Gornale
e-mail: shivanand_gornale@yahoo.com

P. L. Yannawar
Vision and Intelligent System Lab, Department of Computer Science & IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, MS, India
e-mail: pravin@visbamu.in

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_23

street circumstances like narrow bridge, dips, humps and poor law enforcement and improper maintenance of traffic signage's. Occluded, broken, and deteriorated signs may miss guide the driver and cause the risks. Hence the automated Traffic sign detection and classification plays an important role in the Traffic Sign Recognition (TSR) system all over the world. Traffic signage's are different from one country to other country. In this work the experiments are concentered on the classification of Indian traffic sign boards.

Traffic signage's in Indian road System. The Indian Traffic Signage are of mainly three categories, '*Mandatory*', '*Cautionary*', and '*Informatory*', that are represented in the Fig. 1. Previously various methods have been implemented for the detection and classification of traffic signage's, based on colour, and shape. Since the non-stationary environment the approaches may fail or may not properly classify the signage's. Hence this work adopts the fusion level feature and their classification to make the system robustness at certain level as equated to single feature classification.

The organization of the paper is as follows, the Sect. 2 contributes the related work of traffic signs, Sect. 3 gives the methodology, followed by Sect. 4 with experimental results, and lastly the Sect. 5 with the conclusion of the paper.

Fig. 1 Indian traffic signage's

2 Related Work

Fundamentally traffic signage acknowledgement is a multi-class classification problem that has a straightforward test for computer vision and machine learning approaches. In [1] authors have proposed a multi-level feature fusion using EML classifier for traffic sign recognition. The datasets were collected from GTSRB and BTSC correspondingly. The outcomes obtained were suitable for the combination of HOG, CLBP, and Gabor features. Also in [2] authors have proposed the K-d Tree and random forest methods for the classifying the signs boards. HOG and distance transform features are computed for classification of signage's. The GSTRB dataset are used for experimentation. In [3] authors used color threshold method for segmentation and shape analysis for detection of signage's which were classified using neural networks gaining good results. In [4] Method like Kalman Filter is used for motion features for tracking the sign boards. In [5] color segmentation, tamplet matching are adopted for classifying the signage's using neural network. In [7] the feature is extracted based on statistical features and morphological contours of the colour image using red band information. Further these features are classified using SVM classifier. Experiments were carried on domestic datasets with 4 red sign board classes. In [8] also proposed colour based segmentation using RGB, YCbCr and NTSC spaces. Blob detection using binarzation and Otsu thresholding by ROI extraction and shape classification. The datasets were collected using mobile captured videos. Techniques like CNN and RCNN for classification of road sign boards [9]. In [10] HOG features of different spectral bands of signage's are extracted and classified using SVM. Further the complimentary features like dense-shift, LBP and Gabor filter are used for classifying with good accuracy rate on GTSRB datasets.

3 Methodology

The diagrammatic representation of the proposed method is embodied in the Fig. 2. That represents the different phases like, preprocessing and segmentation, feature

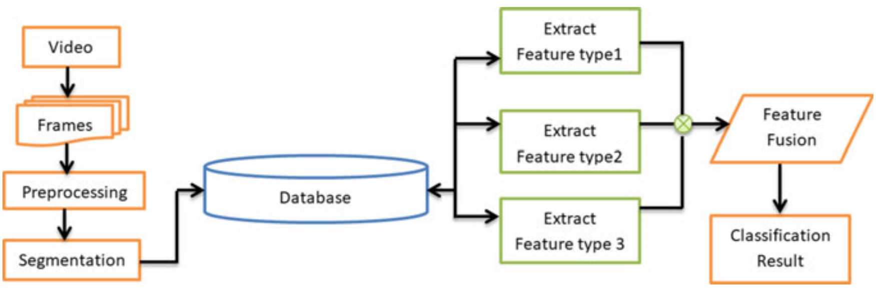


Fig. 2 Illustration of the proposed method

extraction, and classification. The proposed method gives the study of feature fusion that involves the association of LBP, HOG, and LPQ employed within a KNN as well as K-d tree classification methods.

Dataset: In this study a database of Indian road traffic signage's are collected in a heterogeneous climatic condition. The videos are captured from a non-stationary vehicle using 16 mega pixel mobile cameras. The videos are down sampled with each of them to 30 s duration with frame rate of 30fps. The videos contain 15 different classes of signage's.

An algorithm of a proposed method is as follows.

Input: video in mp4 format.
Output: classified signage's.
Step 1: Read video in mp4 format.
Step 2: Extract frames and select the interested candidate frames.
Step 3: Segmenting the ROI using by passing the ground truth values.
Step 4: Extract features1, feature 2, and feature 3.
Step 5: Combining these features and creating a new feature vector.
Step 6: Classification of signage's is done using image processing methods.

3.1 Preprocessing and Segmentation

The input video is break down to frames using image processing method in MATLAB 2018a, and is preprocessed. This phase involves the segmentation of ROI that eliminates the other noisy elements form the frames and extracts the traffic signage's. The color information of these segmented objects is discarded and the resulting gray scale images are included. By using region poly method the objects are segmented. The size and contrast of the ROI are normalized to 200 $\times$ 200 pixel resolution, as the scope of the transportation signage's differs across different scales.

3.2 Feature Extraction Methods

3.2.1 Local Binary Pattern Features (LBP)

It is effective texture descriptor for images. The LBP operative exchanges the picture element values of an image with the decimal numbers which are known as LBP codes that encodes the native structure all over the place of each pixel. Each adjacent picture element are compared with the central picture element, the adjacent picture element is greater than the central picture element make it as bit 1, and the neighbors with less than the central pixel will have bit 0. These gives an 8-digit binary number then convert it to decimal as presented in Fig. 3. Now evaluate the histogram over

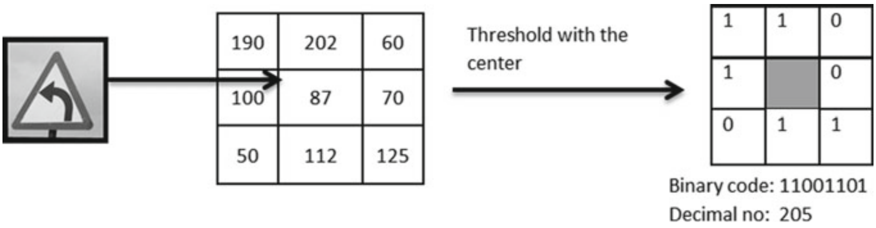


Fig. 3 LBP descriptor

the cell, of the frequency of each value occurring, which is used as feature vector and then stabilize the histogram that forms a feature vector for the entire image window.

For the different sizes LBP operator extends the use of neighborhoods. In general the LBP descriptor for every picture element is given as follows in Eq. 1.

$$LBP_{P,R} = \sum_{P=0}^{P-1} s(g_p - g_c)2^P \tag{1}$$

where gp and gc denote the intensity of the present and adjacent picture elements, P is the number of neighboring pixels chosen at radius R, respectively.

3.2.2 Histogram of Oriented Gradient Features (HOG)

HOG features are greatly known in image processing for pattern recognition problems. The segmented images of the traffic signage's is divided into non-overlapping chunks. Each chunk is subsequently divided into non-overlapping cells. For each picture element, the gradient magnitude and angle are calculated. A histogram is created of the orientation of each cell's gradients. The magnitude of the gradient is used as a vote weight by concatenating the feature vector of each block. The window is split into 16 × 16 overlapping blocks in this job to compute HOG characteristics. Figure 4. shows the HOG features on segmented ROI.

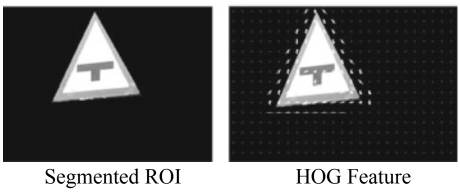


Fig. 4 Hog feature on a segemented ROI

3.2.3 Local Phase Quantization Features (LPQ)

The LPQ is grounded on Fourier stage spectrum's blur invariance property. It uses the local phase information extracted using the 2D DFT or STFT computed over the rectangular M-by-M neighborhood Nx at each pixel position x of the image f(x) defined as follows [6].

$$F(u, x) = \sum_{y \in Nx}^{n} f(\mathrm{x} - \mathrm{y}) e^{-j2\pi u T y} = W_u^T f_x \tag{2}$$

where Wu is the 2-D DFT base vector at frequency u and where fx is another vector that containing all M2 image samples from Nx.

The scalar quantizer is defined by Eq. 3 as follows

$$q(j) = \begin{cases} 1, \ gi \geq 0 \\ 0, \ otherwise \end{cases} \tag{3}$$

where q(j) is the jth component of the G. the scalar quantizer result in a 2 bit integer representation for a single frequency component at every print at x. the L quantized coefficient is presented at the histogram according tp the binary coding, and is given as in Eq. 4.

$$\mathrm{B} = \sum_{(j=1)}^{L} q_{(j)} 2^{(j-1)} \tag{4}$$

This ranges between 0 and 2L-1 and describe the local texture at x.

3.3 Feature Fusion

the methods for combining several confirmations, such as several outcomes or characteristics of classification can be split into three groups, "*feature level fusion*", "*decision level fusion*", and "*score level fusion*" [1]. This work presents the fusion at the feature level. Features like LBP, HOG, and LPQ are extracted individually and in different combinations to enhance the performances for the classification. The performance of these feature combination yielded high classification rate, with average computational cost.

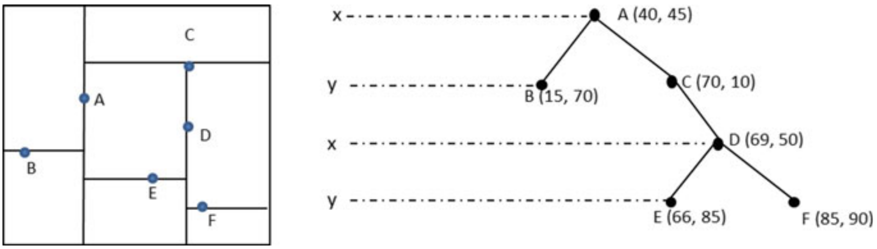


Fig. 5 Sample example of K-d tree

3.4 Classification

3.4.1 K Nearest Neighbor Classifier

KNN is a humble operation in which all the obtainable bags are stored and fresh bags are categorized based on similarity measures. K is the user defined constant and an unlabeled vector is categorized by allocating the best repeated tag to the closest k training sample at that query stage.

3.4.2 K-d Tree Classifier

K-d tree is a binary search tree organizing where l- dimensional points are organized. In the function I each non-leaf node divides the information into two subspaces with the greatest variance at that stage. Each node therefore includes a (k-1) dimensional hyper plane $Hi = \{x1, \ldots, xi-1, i, xi + 1, xk\}$, $1 \leq i \leq k$ parallel to the ith axis. The Fig. 5 shows the sample example of k-d tree.

Alternatively the 2D space is divided by 1-D lines parallel to the x and the y axes to ensure a balanced tree. The ith dimension average value vi of the is used for the splitting in each node. The sub-tree of the leaf includes data points smaller than vi and points larger than vi of the right sub-tree. This division is repeated until a leaf represents each data points. The K-d tree supports for the indexing of higher dimensional spaces like, HOG vector or other large descriptor [2].

4 Results

The experimentations were carried on the datasets collected from the mobile videos in Indian climatic conditions. Overall Dataset consists of 15 different classes of traffic signage's, each class consist of 50 samples, total segmented signage's are 750 traffic signage's taken from the video datasets. The features were classified for two classifiers, KNN and K-d tree. The features are extracted individually and the feature classification rate is represented in the Table 1 along with the result graph shown in the Fig. 6.

Further the individual features vectors are combined with different combinations for evaluating the performances. The outcomes gained are presented in the following Table 2 along with the chart presenting the accuracy gained by KNN and K-d tree in Fig. 7.

Table 1 presents the obtained classification rates for individual features where Table 2 shows the gained classification degrees for diverse combinations of features. From the tentative results it is demonstrated that the signage features based on HOG are more competitive than the LBP and LPQ features. Whereas the mixture of all the three features that is LBP, HOG and LPQ can increase the classification rate to good extent. Hence the fusion method based on mixture of three texture features gives

Table 1 Classification results

Type	Time (s) KNN	Accuracy KNN (%)	Time (s) K-d tree	Accuracy K-d tree (%)
LBP	47.14	87.6	46.44	91.33
HOG	55.6	92.6	47.31	97.1
LPQ	58.4	73.3	56.59	94.13

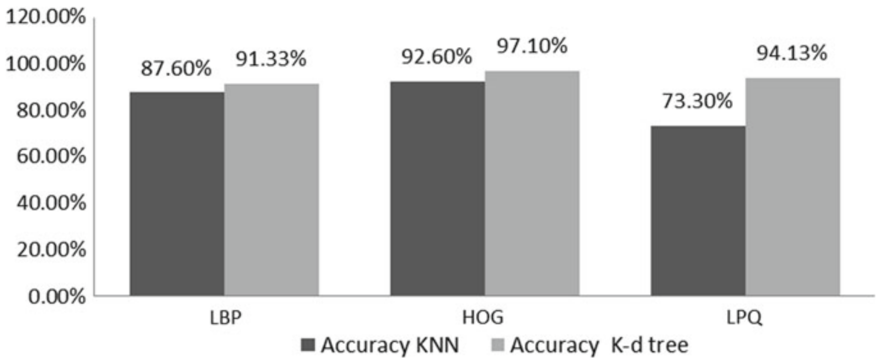


Fig. 6 Result analysis of classification rate obtained by individual features

Table 2 Classification results

Type	Time (s) KNN	Accuracy KNN (%)	Time (s) K-d tree	Accuracy K-d tree (%)
LBP + LPQ	64.83	79.3	68.62	95
LBP + HOG	56.59	92.8	65.94	97
HOG + LPQ	69.04	92.6	61.18	97.02
LBP + HOG + LPQ	81.76	96.8	74.17	97.3

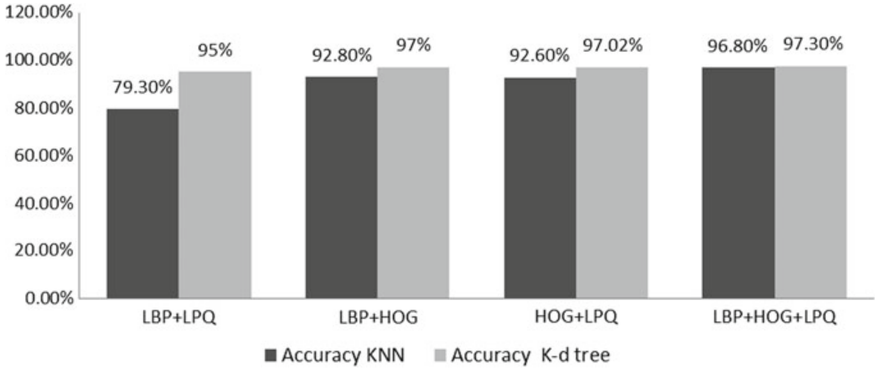


Fig. 7 Result analysis of classification rate obtained by mixtures of features

good feature level traffic sign classification accuracy for both classifiers by 96.8% using KNN and 97.3% using K-d Tree. Result analysis of classification rate obtained by Mixtures of features is shown in Fig. 7.

5 Conclusion

Real time road traffic signage's with different climatic conditions gives many challenges as for as image processing and analysis is concerned. Signage's taken in diverse illumination environments gives different results. Therefore the proposed work is carried for Indian traffic signage's for achieving good results in the detection and classification is apprehensive. The input of this work is to explore a computationally effective TSR technique based on feature of fusion characteristics by combining LBP, HOG, and LPQ with KNN and K-d tree classifiers. The experimental findings indicate that the suggested technique provides similar efficiency of state-of-art methods with excellent time complexity in calculation. Performance of the method is evaluated on self-captured dataset through a mobile camera. The future work of this approach is to concentrate on the detection and classification method more accurate within less time complexity.

References

1. Aziz S, El Aroussi M, Youssef F (2018) Traffic sign recognition based on multi-feature fusion and ELM classifier. In: The first international conference on intelligent computing in data sciences, vol 127, pp 146–153
2. Zaklouta F, Stanciulescu B, Hamdoum O (2011) Traffic sign classification using k-d tree and random forests. In: Proceedings of international joint conference on neural network, san Jose, California, USA, 31 July 2011

3. de la Escalera A, Moreno LE, Salichs MA, Armingol JM (1997) Road traffic sign detection and classification. IEEE Trans Ind Electron 44(6):848–859
4. Dhingra H, Diwakar S, Saranya G, Kumar M (2018) Fusion model for traffic sign detection, tracking and recognition. J Electron Syst 8(2)
5. Dean HN, Jabir KTV (2013) Real time detection and recognition of indian traffic signs using Matlab. Int J Sci Eng Res 4(5)
6. Ojansivu V, Heikkila J (2008) Blur intensitive texture classification using local phase quantization. ICISP, Springer, Berlin, pp 236–243
7. Khodadadzadeh M, Sarrafzade O, Ghassemian H (2010) A traffic sign detection and classification system based on fusion of color and morphological information. 978-1-4244-9708-9/10/$26.00 ©2010 IEEE
8. Arun K, Vasanthabalan R, Siranjeevi S (2018) Video based traffic sign detection by scale based frame fusion technique. Int Res J Eng Technol (IRJET) 5(3)
9. Zhu Z, Liang D, Zhang S, Huang X, Li B, Hu S, Traffic sign detection and classification in wild. IEEE
10. Ma Y, Huang L (2015) Hierarchical traffic sign recognition based of multi-feature and multi-classifier fusion. In: First international conference on information science and electronic technology (ISET), pp 56–59

Shivanand S. Gornale has completed M.Sc. in Computer Science. M.Phil. in Computer Science, Ph.D. in Computer Science from University of Pune, Maharashtra, India. He is recognized research guide for Ph.D. in Computer Science and Engineering from Rani Channamma University, Belagavi and Jain University Bangalore. He has published more than 85 research papers in various National and Inter-national Journals and conferences. He is a Fellow of IETE New Delhi, Life Member of CSI, Life Member of IPRA, Member of IARCS, Life Member of Indian Science Congress Association, Kolkata-India. Presently he is working as Professor and Chairman, Department of Computer Science, Rani Channamma University, Belagavi-Karnataka, India. His research area of interest is Digital Image Processing, Pattern Recognition, Computer Vision and Machine Learning, Video Retrieval and Biometric analysis.

Ashvini K. Babaleshwar completed M.Sc in Computer Science from Rani Channamma University, Belagavi, Karnataka, India in 2013. She is pursuing Ph.D. in Computer Science at Rani Channamma University, Belagavi, Karnataka, India. Her area of research includes Image Processing and Pattern recognition, Content based Image and Video retrieval.

Pravin L. Yannawar is Assistant Professor in Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS). He has completed M.Sc. (2001), SET(Computer Science & application) Ph.D. (2011). He is Member of IETE & life member of IAEng, CSTA, IUPRAI and IACSIT. He is recipient of DST Fast Track Young Scientist research project. His area of research includes Vision, Intelligent Systems, AVSR, OCR, Image Processing and pattern recognition.

Extended Local Ternary Pattern for Face Anti-spoofing

R. J. Raghavendra and R. Sanjeev Kunte

Abstract The Local Binary Patterns (LBP) finds its use in texture classification because of it is simplicity. The extension of LBP that is Local Ternary Pattern (LTP) is a three valued code which is better descriptor than LBP and resistant to noise. But LTP is not invariant to gray-level transformations. To overcome this deficiency of LTP, we propose to use Extended Local Ternary Patter (ELTP) for face anti-spoofing system. The ELTP uses an auto-adaptive method for threshold value selection to make it immune to gray-level variance. Experiments were performed by using all the three descriptors (LBP, LTP and ELTP) and the results are compared. The proposed face anti-spoofing method performs better than LBP and LTP.

Keywords Face Anti-spoofing · Extended local ternary pattern (ELTP) · Local ternary pattern (LTP) · Local binary pattern (LBP)

1 Introduction

In biometrics, human face plays major role, its features forms the basic characteristics of their biometric profile. The user data and privacy can be secured using the adoption of face biometrics facility. Now days, there will be increase in the identity theft have amplified usage of physical characteristics such as face, finger prints, signature and iris, whereas behavioral characteristics such as gait, voice, skin reflection, body odor and head movement [1, 2].

The spoofing attack is nothing but an intruder, pretending to be authorized user, attempt to obtain information of person or a system. In the most common types of spoofing attacks includes photos, videos and 3D mask or recorded videos. Whenever the intruder tries to spoof the system, it detects the biometric trait and detects spoof-attacks continuously and appropriate actions are taken to secure the systems. This type of system is able to distinguish between real and fake users.

R. J. Raghavendra (✉) · R. S. Kunte
J N N College of Engineering, Shimoga 577204, India
e-mail: raghavendra.r.j@jnnce.ac.in

R. S. Kunte
e-mail: sanjeevkunte@jnnce.ac.in

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_24

In this paper, we presents an Extended Local Ternary Pattern (ELTP) a novel descriptor for anti-spoofing system, which is based on extracting features from the face images in order to address problem of spoofing. In this work we use SVM as a machine learning model for real/fake face classification. The organization of the paper is as follows: Section 2 describes the related work on face spoofing techniques with different approaches. Section 3 presents the details face spoofing detection using LBP and LTP descriptor and explores the proposed work carried out using ELTP for face spoofing detection. Section 4 presents the results and discussion about proposed method. Section 5 concludes the work carried out.

2 Related Work

In previous work, the face liveness detection categorized into Intrusive and Non-Intrusive methods [3]. In the Intrusive technique, liveness can be identified by asking user to do some specific random actions, such as movement of head in certain directions according to the system generated instruction [4]. Another method is lips movement [5], where clients are asked to express particular digit to record the movement of the lips. Further, in non-intrusive methods user's interaction is not needed. It uses the unconstrained physiological exercises of face.

LBP is a descriptor used for analysis of texture. This descriptor uses the pixels of an image and represented in terms of binary numbers 0's and 1's. The Local Binary Pattern (LBP) and multi-scale Local Binary Patterns are used to dissect the texture elements of the face [6–8]. Micro-textures present in the face [9] are also identified using LBP. In [10] the photo image's characteristics such as contrast and texture have been used. In another work [11] an additional approach which uses three orthogonal planes along with LBP to combine space and time for photos and videos is presented. The face image is segmented into number of local regions to extract texture information [12].

Face recognition systems are easily spoofed [13, 14] using photograph of authorized user, may be displayed in photo-copy or on a screen. Many clues are used in anti-spoofing for face recognition system such as motion analysis, texture exploration and liveness detection [15]. Li et al. [16], suggests new method for liveness detection. This method concentrates on 3D depth information of an image, which is based on the analysis of single Fourier spectra.

Some of the category of anti-spoofing methods deals with live-face specific motion such as blinking of eyes, movements of lips and head rotation are proposed in [17, 18]. Some of the publications which are specifically concentrate on the properties of human head movement as 3D object in [19, 20]. Anjos et al. [21] shows that there will be strong relation between the movement of head in the context and the background of the scene. This algorithm achieved HTER rate of 8.98%.

3 Proposed Method

The printed photos used in spoofing almost looks similar to the live face images. To handle such a micro texture [8, 22] of photo images LBP descriptor was introduced. However, LBP was not dealt with variations of appearance in photo images caused noises, illumination and partial occlusions. The conventional LBP was more sensitive to noise, to overcome this a 3-valued pattern instead of a binary pattern was introduced by Tan and Triggs [23]. This 3-valued pattern was modified as Local Ternary Pattern (LTP) presented a texture operator which is more robust to noise. The major challenge of LTP is difficulty to set the right threshold value for a particular application. Also, in LTP when the threshold value becomes exactly the same as the difference of central and neighborhood pixels, in such cases generated code of LTP turns out to be zero.

In order to overcome these shortfalls recently noticeable extension of LTP known as Improved LTP, Local Adaptive Ternary Pattern, Dynamic LTP and Extended LTP have been proposed, which uses some automatic methods to set the threshold value in each local region.

In order to overcome the limitations and enhance the performance of anti-spoofing system, we proposes a liveness detection system using Extended Local Ternary Pattern.

3.1 Spoofing Detection Using ELTP

In this section, we addresses the anti-spoofing system using ELTP, which will differentiate real and spoof face. In the beginning, face will be identified using Viola-jones face detector [24]. It uses landmarks, to crop the face in given image. These landmarks are provided by Active Shape Models with STASM. Further, cropped faces will be pre-processed and converted from RGB to gray-scale. Then, apply ELTP descriptor to 3×3 overlapping regions of whole image. This process will be repeated for both real and spoof faces of NUAA dataset. SVM classifier is used finally to differentiate given input image is real or spoof.

I. **Extended Local Ternary Pattern (ELTP)**

LTP was more resistant to noise but, it is not having invariance to gray-level transformations. To resolve this, we are proposing Extended Local Ternary Pattern (ELTP) [25], which uses auto-adaptive method to select the threshold value by Median Absolute Deviation (MAD) instead of customized threshold value used in LTP.

In this work we are proposing to use ELTP for face spoof detection. The ELTP descriptor is made to invariant to gray-level transformations as follows. Gray-levels in each pixels of an image is quantized depends on the width of tolerance zone $\pm tz^e$ around g_c^e to zero, above the tz^{e} quantized as $+1$ and below it as -1. The conventional LBP indicator U(X) is replace with function contains 3 values as show in Eq. (1).

$$U^e(g_m, g_c^e, tz^e) = \begin{cases} 1 & g_m - g_c^e \geq tz^e \\ 0 & |g_m - g_c^e| < tz^e, m = 0, 1 \ldots P-1 \\ -1 & g_m - g_c^e \leq -tz^e \end{cases} \tag{1}$$

$$g_c^e = \text{mean(G)}, tz^e = MAD(G), \quad \text{G} = \{g_i | i = 0, 1, \ldots, 8\}$$

where M represents number of neighbor pixel, neighbor pixel's gray value is represented by g_m (m = 0, 1,..., M−1), pixel values of 3 × 3 local region is the set *G*, *mean* (*G*) is represents average set of *G*, MAD (*G*) is the median absolute deviation of gray values of the set *G*, and g_c^e is the center pixel gray value. In LTP customized threshold is used. In this work we propose to use a strategy to set the values of threshold dynamically by auto-adaptive method. In order to make ELTP more robust to noise and gray level transformation the threshold value is calculated by using Median Absolute Deviation (MAD) for all 3 × 3 block pixels. This method will not affect the model complexity.

For simplicity, each ternary pattern of ELTP is divided into upper half (ELTP_U) and lower half (ELTP_L) as shown in Fig. 1. Histogram of lower and upper half are combined to get final ELTP value. The ELTP descriptor is defined in Eq. (2).

$$\text{ELTP\_U}_{P,R} = \sum_{m=0}^{m-1} 2^m \left(e\left(\text{U}^e\left(\text{g}_m, \text{g}_e^c, tz^e\right), 1\right)\right),$$

$$\text{ELTP\_L}_{P,R} = \sum_{m=0}^{m-1} 2^m \left(e\left(\text{U}^e\left(\text{g}_m, \text{g}_e^c, tz^e\right), -1\right)\right), \quad \text{e(x, y)} = \begin{cases} 1, & x = y \\ 0, & x \neq y \end{cases} \tag{2}$$

II. **Classification**

For classification of images we have used SVM classifier. In our experiments, once the histograms are calculated and combined of each image it will be fed to the SVM classifier [21] for determining the given input image is live or spoof. The SVM

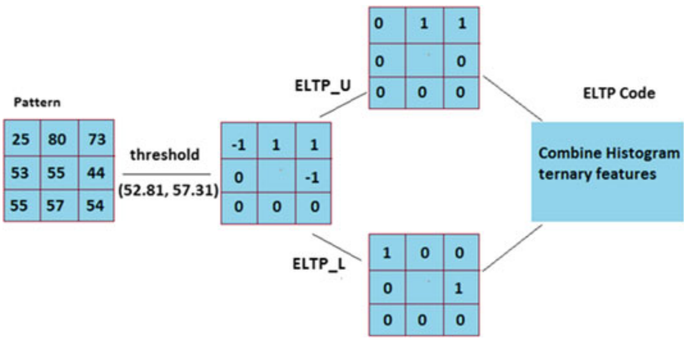


Fig. 1 Extended local ternary pattern calculation process on sample image

classifier is trained using samples of both real and spoofed images from NUAA dataset. The following pseudo code is used.
Pseudo code:

1. Read input images from NUAA dataset.
2. Resize the input image into 160 × 160 size and convert each image from RGB to gray scale image.
3. Apply Viola-Jones face detector to gray scale image to detect and extract face.
4. Segmented the image into number of blocks.
5. Apply ELTP descriptor to each block, which produces ELTP code from upper and lower half of the block.
6. Histogram is applied to combine the both 128 features and get 256 features from it.
7. Apply the SVM classifier to classify the fake and real faces.

4 Results and Discussions

We conducted three different experiments using the NUAA dataset [26] with 14 different classes of real and spoof images. ELTP features are extracted after from face regions including real and fake images. The extracted features are collected in a file and the features are given as input to SVM classifier which is trained with kernel function. During testing, the test image is also extracted features using ELTP descriptor as above and given as input to trained SVM classifier to decide give image is real/fake.

Experiment 1: Varying training images of same class

In this experiment, we have considered only one class (Class 1). The original image of dataset is resized into 160 × 160 pixels. During training of SVM classifier 10 images were considered in same class and then tested. Subsequently number of images was increased from 20 to 100 for training. The testing set comprising both real and spoof to a total of 200 images. For all images total number of ELTP features extracted are 256. It has been observed from Table 1 that as the number of training images in training set increases the percentage of recognition rate also increases (both for real and spoof images) as show in Fig. 2.

Experiment 2: **Comparison of ELTP with other descriptor**

In this experiment, along with ELTP descriptor we have considered LBP, LTP also for comparison. During training of SVM classifier 10 images were considered in each of the 14 class and then tested. Later numbers of images were increased from 20 to 100 in each class for training. The testing set comprising both real and spoof total of 1150 and 1949 images respectively. The evaluated results of LBP, LTP and ELTP is presented in the Table 2. For all images total number of ELTP features extracted are 256. It has been observed from Fig. 3 that ELTP performs far better

Table 1 Recognition rate for varying training images of same class

S. No.	No. of training images	Number of correct recognition ELTP	Overall recognition rate (%)
1	20	119	59.5
2	40	157	78.5
3	60	162	81
4	80	162	81
5	100	187	93.5
6	120	191	94.5
7	140	192	96
8	160	193	96.5
9	180	197	98.5
10	200	199	99.5

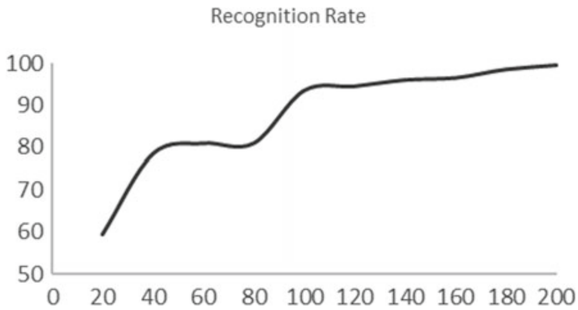


Fig. 2 Recognition rate of the proposed work for varying number of training images

Table 2 Comparisons of LBP, LTP, ELTP descriptor

S. No.	No. of training images	Overall recognition rate (%)		
		LBP	LTP (t = 2)	ELTP
1	280	52.34	80.5	85.6
2	560	66.9	80.9	85.8
3	840	67.18	81.6	86.2
4	1120	67.6	81.8	86.7
5	1400	70.28	82.8	87.5
6	1680	71.12	82.9	87.73
7	1960	74.6	83.1	88
8	2240	77	83.8	88.21
9	2520	79	84.2	89.3
10	2800	80	84.9	90

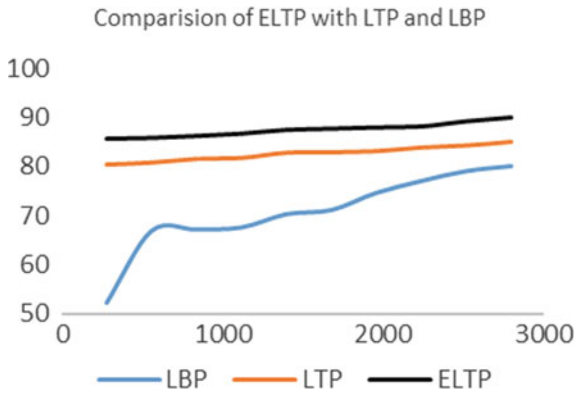


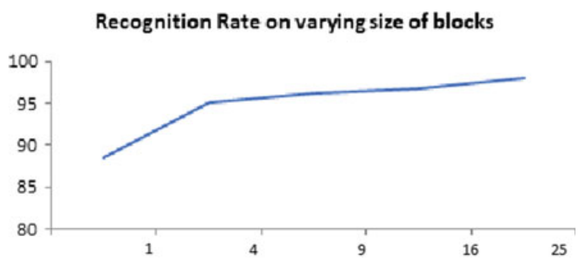
Fig. 3 Overall recognition rate of the ELTP with LBP and LTP

compare to other descriptors. As per observation, the number of training images in training set increases the percentage of recognition rate also increases (both for real and spoof images). The capability of the proposed ELTP compared to the LTP and LBP to distinguish the images into real or spoof is evident from this experiment.

Experiment 3: **Varying sizes of blocks of training images of same class**

In this experiment, we have considered NUAA anti spoofing face database comprising 14 different classes of real and spoof images in that we have considered only one class (Class 1) of 140 images of both real and spoof. Each face region of dataset is divided into different number of blocks of size from 160 $\times$ 160 pixels to 32 $\times$ 32 pixels and then ELTP descriptor is applied on each block. The testing set comprising both real and spoof total of 1150 and 1949 images respectively. For all images total number of ELTP features extracted is 256 from each block. It has been observed from Fig. 4. that as the number of block size of training images in training set increases the percentage of recognition rate also increases (both for real and spoof images).

Fig. 4 ELTP performance for varying block sizes of training images



5 Conclusion

In this work a face spoofing detection system using Extended Local Ternary Pattern descriptors is developed. The ELTP produces the features using dynamic threshold value, which makes the ELTP descriptor more insensitive to gray-scale transformation. We have conducted three different experiments on NUAA database. The obtained results from different descriptors (LBP, LTP and ELTP) with different class sizes were compared. It has been observed from the experiments that ELTP performance better than other two descriptors. The outcome of this works shows ELTP performance is better in face anti-spoofing system.

References

1. Li SZ (2009) Encyclopedia of biometrics: I-Z. In: Proceedings of springer science & business media, vol 1
2. Jain AK, Maltoni D, Maio D, Wayman J (2005) Biometric systems technology, design and performance evaluation. In: Proceedings of Spring, London. https://doi.org/10.1007/978-0-387-71041-9_1
3. Parveen S, Syed Ahmad SM, Hanafi M, Wan Adnan WA (2015) Face anti-spoofing methods. In: Proceedings of current science, pp 1491–1500
4. Frischholz RW, Werner A (2003) Avoiding replay attacks in a face recognition system using head-pose estimation. In: Proceedings of analysis and modeling of faces and gestures. IEEE. https://doi.org/10.1007/978-3-642-23678-5_12
5. Kollreider K, Fronthaler H, Faraj MI, Bigun J (2007) Real-time face detection and motion analysis with application in liveness assessment. In: Proceedings of information forensics and security. IEEE, pp 548–558
6. Hadid A (2008) The local binary pattern approach and its application to face analysis. In: Proceedings of Image processing theory, tools and applications, pp 1–9. http://dx.doi.org/10.1109/IPTA.2008.4743795
7. Chingovska I, Anjos A, Marcel S (2012) On the effectiveness of Local binary patterns in face anti-spoofing. In: Proceedings of international conference of the biometrics special interest group. IEEE, pp 1–7
8. Maatta J, Hadid A, Pietikainen M (2011) Face spoofing detection from single images using micro-texture analysis. In: Proceedings of international joint conference on biometrics (IJCB). IEEE, pp 1–7
9. De Freitas Pereira T, Komulainen J, Anjos A, DeMartino JM, Hadid A, Pietikäinen M, Marcel S (2014) Face liveness detection using dynamic texture. In: Proceedings of journal on image and video processing (EURASIP). IEEE, pp 1–15
10. Kose N, Dugelay JL (2012) Classification of captured and recaptured images to detect photograph spoofing. In: Proceedings of international conference on informatics, electronics & vision (ICIEV). IEEE, pp 1027–1032
11. De Freitas Pereira T, Anjos A, De Martino JM, Marcel S (2013) LBP—TOP based countermeasure against face spoofing attacks. In: Proceedings of computer vision-ACCV workshops. Springer, Berlin, pp 121–132
12. Housam KB, Lau SH, Pang YH, Liew YP, Chiang ML (2014) Face spoofing detection based on improved local graph structure. In: Proceedings of international conference on information science and applications (ICISA). IEEE, pp 1–4
13. Thalheim L, Krissler J, Ziegler P-M (2002) Body check: biometric access protection devices and their programs put to the test. Heise Online

14. Duc NM, Minh BQ (2009) Your face is not your password. In: Proceedings of black hat conference
15. Chakka MM, Anjos A, Marcel S, Tronci R (2011) Competition on counter measures to 2-d facial attacks. In: Proceedings of international joint conference on biometrics
16. Li J, Wang Y, Tan T, Jain AK (2004) Live face detection based on the analysis of Fourier spectra. In: Proceedings of defense and security. international society for optics and photonics, pp 296–303
17. Pan G, Wu Z, Sun L (2008) Liveness detection for face recognition: recent advances in face recognition. Austria, Vienna
18. Jee HK, Jung SU, Yoo JH (2008) Liveness detection for embedded face recognition system. Proc Int J Comput Inf Eng 2(6)
19. Kollreider K, Fronthaler H, Bigun J (2009) Non-intrusive liveness detection by face images. Proc J Image Vis Comput, 27(3)
20. Bao W, Li H, Li N, Jiang W (2009) A liveness detection method for face recognition based on optical flow field. In: Proceedings of international conference on image analysis and signal processing. IEEE
21. Anjos A, Marcel S (2011) Counter-measures to photo attacks in face recognition: a public database and a baseline. In: Proceedings of international joint conference on biometrics, pp 1–7
22. Pietikainen M, Hadid A, Zhao G, Ahonen T (2011) Computer vision using local binary patterns. Springer, London, UK
23. Tan X, Triggs B (2010) Enhanced local texture feature sets for face recognition under difficult lighting conditions. Proc Trans Image Process 19(6):1635–1650
24. Ojala T, Pietikainen M, Maenppaa T (2002) Multi resolution gray-scale and rotation invariant texture classification with local binary patterns. Proc Pattern Anal Mach Intelligence 24(7):971–987
25. Yuan J-H, Zhu H-D, Gan Y, Shang L (2014) Enhanced local ternary pattern for texture classification. In: Proceedings of springer international publishing switzerland, pp 443–448
26. Joshi T, Dey S, Samanta D (2009) Multimodal biometrics: state of the art in fusion techniques. Proc Int J Biom 1(4):393–417

Enhancing Data Security in Cloud Computing Using Threshold Cryptography Technique

K. Suresha and P. Vijaya Karthick

Abstract Modern technological advances have given rise to the popularity and success of cloud in organizations and industries because it provides data storage facility and computing services at minimal cost. Data security is one of the important requirements for outsourced data in a cloud storage environment. It also brings new security vulnerabilities for assuring the data confidentiality, data integrity and data access control. Data security of a cloud system depends heavily on the cloud service provider, data owner and cloud users. So, there needs to be some methodology to monitor security vulnerabilities and providing data security in cloud environment. Now a day's providing data security in cloud storage is one of the most critical tasks. For the past few years many works have been done to assure data security but they are failed in providing data security such as information leakage, secret-key management, cloud-user authentication, breach of data confidentiality, because of collusion attack and lots of computational activities (because of large number of secret keys).In this paper, we used threshold cryptography technique for providing data security in cloud storage system. Proposed approach addresses the data security issues in cloud system more efficiently. In this Threshold cryptography technique data owner create users in groups based on place or location, project, and department, data owner allocates single secret key to each group of users for decryption of data and every users in the group shares parts or pieces of the secret key. Proposed methodology also uses data access control list to control the data access. Proposed approach provides data security more efficiently, it also increases the performance of the system and also it minimizes the number of secret keys.

Keywords Threshold cryptography · Capability list · Collusion attack · Malicious user · Message digest

K. Suresha (✉)
Department of Computer Science and Engineering, DRR Government Polytechnic, Davanagere, Karnataka, India
e-mail: skdrrgpt@gmail.com

P. Vijaya Karthick
Department of Information Science & Engg., Sir M Visvesvaraya Institute of Technology, Bangalore, Karnataka, India
e-mail: vijay691988@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_25

1 Introduction

Cloud computing is a type of Internet-based computing that provides shared computer processing resources and data to computers and other devices on demand. Cloud computing is a model for enabling ubiquitous, on demand access to a shared pool of configurable computing resources (e.g., Computer networks, Servers, Storage, Applications and Services) which can be provisioned and released with minimal management effort. Cloud computing and storage solutions provide users and enterprises with various capabilities to store and process their data in either privately owned or third-party data centers that may be located far from the user ranging in distance from across a city to across the world. Cloud computing relies on sharing of resources to achieve coherence and economy of scale, similar to a utility over an electricity network.

Cloud Storage provides computing services based on three fundamental cloud computing service models: Software as a service-SaaS, Platform as a service-PaaS, Infrastructure as a service-IaaS. The five features of cloud computing are: measured scale service, rapid elasticity, location independent, Self Service, On-demand service. These features make cloud computing is very significant advancement. Companies are make full use of and derive benefit from all features of cloud system and maximize their finance or profit or revenue. So that all companies are moving their day to day activities towards cloud storage.

Data security is a major obstacle in cloud storage. Industries hesitate to use the cloud storage system. Many IT persons believes cloud computing is not safe place to store valuable data at cloud service provider. You lost complete access control over data, absolutely they are thinking correct. Valuable data is processing and storing at remote server system. That's way data confidentiality, data integrity, data access control may breach. Cloud server systems are maintained by external cloud service providers, so data owner don't believe on cloud service provider. Cloud service provider can use user's data for their financial benefits and can do spoil or destroy day to day transactions of data owner. Even though data owner also don't trust on users of each group, because users also disclose the group data to cloud service provider for financial benefits. Data confidentiality may breach because of collusion attack by users and also cloud service provider. Many methodologies are proposed to assure data security but they are failed because of collusion attack and lots of computational activities (because of large number of secret keys).

Proposed approach provides solution to the data security problems in cloud storage, it is composed of Users, Data Owner and Cloud Service Provider. Group-Users are created on criteria such as place or location, project work, department. Data owner allocates single secret key to every group for doing encrypt as well as decrypt on data. Every user in the group allocates pieces of the secret key. Data may be decrypted when at least threshold number of users will present. This methodology provides data security very efficiently, it also increases the performance of the system also it minimizes the no. of secret keys. To successfully get the data access control, the methodology is used data access control capability list. It contains registered user details, user data and what operations are allowed for an each user is specified. It is good method than Access Control List (ACL), Access Control List maintains users and their permitted operations for data as well as file. In this proposed approach D-H algorithm is used to create KT-one time session-key between Cloud Service Provider and group user to protect the data from attackers or outsiders. To assure data integrity the proposed system is used Secure Hash algorithm-SHA-512.

2 Problem Statement

Cloud storage system brings many security vulnerabilities for assuring the data confidentiality, data integrity and data access control. Data security of a cloud system depends heavily on the cloud service provider, data owner and users as a whole, but there needs to be some methodology to monitor security vulnerabilities and provide data security. Now a day's providing data security in cloud storage is one of the most critical activities. For the past few years many works have been done to assure data security but they are failed in providing data security such as data confidentiality breach because of collusion attack and lots of computational activity(because of large number of secret keys).

In this paper, we used threshold cryptography technique for providing data security in cloud storage system. Proposed system address the data security issues in cloud system more efficiently. In this Threshold cryptography technique data owner create users in groups based on place or location, project, and department, data owner provides single secret key to each group of users for decryption and every users in the group shares parts or pieces of the secret key. Proposed system uses data access control list to control the data access. Proposed approach provides data security more efficiently, it maintains the performance of the system more efficiently and also minimizes the number of secret keys.

3 Literature Survey

Confidentiality and Access control of data are two important and critical data-security issues for outsourced data in cloud environment. Most of the IT people concentrate more on security of data, but neglecting the performance of the systems. For illustration purpose, consider if we want to provide security to data, common method is to use more secret keys. Everyone knowing secret-keys are very confidential, security and key management is extra overhead task. This overhead task impact on the overall performance of the cloud system. We need a security system to minimize the number of secret-keys to provide data security and also increase the performance of the cloud system. For the past few years many works have been done as in [1–14] to assure data security but they are failed in providing data security such as data confidentiality breach because of collusion attack and lots of computational activity (because of large number of secret keys).

3.1 Existing Systems

1. The scheme proposed as in [3] allocates a single secret key to each group so that all users must know the secret key to decrypt the data. It is also called as group key mechanism.
   Drawbacks: Here, no. of secret-keys is minimized but possibility of collusion attack because a single harmful user can disclose complete data of the group to Cloud Service Provider. Because we can't believe on Cloud Service Provider. Cloud service provider can use the outsourced data for his financial profits.
2. The mechanism proposed as in [2] provides data confidentiality and access control of data. Basically what happens here is, data are encrypted by symmetric keys and this symmetric key is shared by both data owner and data users. Data encrypted and storing in cloud service provider. Here cloud service provider also not able to see the data because it is encrypted. In this mechanism a onetime session key is used between the cloud service provider and user to protect data from outside attackers and also they used Diffie-Hellman algorithm to generate one time session key.
   Drawbacks: Here it provides overall data security but each user allocates a separate secret key. Here what happens if the users become large so secret keys also becomes large. So it creates new overhead to secure and management of secret-keys.
3. The scheme proposed as in [1] uses threshold cryptography to provide data security. This scheme reduces the number of secret keys by giving single symmetric secret key to each group.
   Drawbacks: In this scheme user has to forward encrypt message and user part of secret key to next user to enter shared secret key parts to decrypt the data. So there is a possibility of message and secret key compromised with malicious user in the group. This scheme uses computational complexity to decrypt the data. So it affects the performance of the system, it requires an improvement. Second drawback of this scheme is privileges of access rights, parent group can see the data of child group, and here also system needs an improvement.

3.2 Proposed System

To address the drawbacks of existing systems we have proposed an approach using Threshold Cryptography Technique in this technique data owner creates users in groups and allocates single secret key to each user group to decrypt the data and each user in the group shares pieces or parts of the secret key. Proposed approach, also using access control capability list to control the data access. Proposed approach provides data security more efficiently and also it minimizes the no. of secret and also it increases the performance of the system. The proposed approach is applicable in team work and group work, suitable in IT companies where work can be done in group or team.

4 Objectives

The objectives of the proposed research work:

(a) To understand and evaluate the existing data security schemes in cloud computing and analyze their performances under different parameters and identify the deficiencies.
(b) To identify and analyze the security issues, challenges, vulnerabilities and threats those are currently present in cloud computing and find the counter measures with possible solutions.
(c) To Design and implement the proposed data security framework to enhance the data Security in cloud computing using cryptographic techniques, encryption algorithms and Secure hash algorithms.
(d) To test and analyze the performances of the developed data security framework with different parameters and compare the results with existing systems.

5 Methodology and Assumptions

Proposed Security Scheme consists of: Cloud Service Provider, Data Owner and Users. First of all, all Users must be register in data owner, while doing registration they must provide userid, passwords and related information. Here user-id, passwords and related information must be obtained in secure fashion, afterwards data owner creates users in group and provide secret and other related information for secure exchange of communication between cloud service provider and user. After successful registration, group user can request the data from cloud service provider by providing valid credentials after completes authentication process at Cloud Service Provider he/she can obtain the data from Cloud Service Provider in secure manner. Cloud Service Provider having a more storage capability and lots of computing facility. And also no one can breach the security at Cloud Service Provider. And also secret key generation algorithm is secure at data owner. Data Owner also having storing capability to store information and also he can do execute programs at Cloud Service

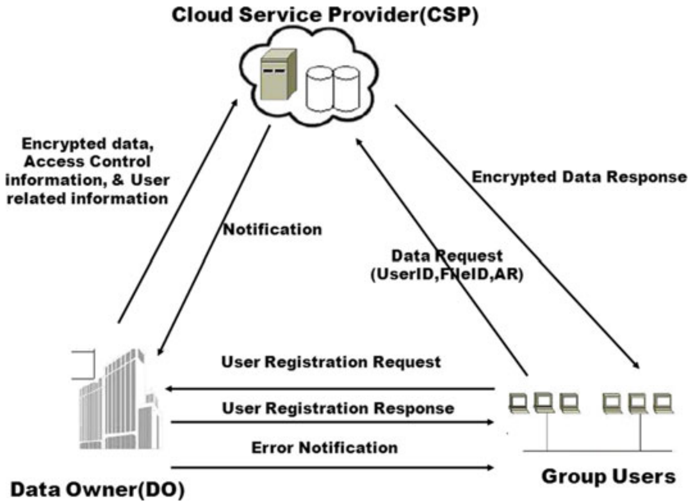


Fig. 1 Proposed system architecture

Provider to maintain data. In the proposed approach we using DH and asymmetric algorithm to establish communication between data owner and cloud service provider in secure fashion. Figure 1 demonstrate the proposed system communication model and Architecture. Figure 2 shows proposed system symbols Table.

We assume that the following techniques will be used in the implementation of the proposed research work:

(a) Threshold Cryptography technique, in this technique data owner creates users in groups and allocates single secret to each group for decryption of data and every user in the group shares pieces of the secret-key.
(b) Shamir secret sharing scheme algorithm to divide Secret Key S into n pieces (S1, S2, S3, …, Sn) and secret key reconstruction.
(c) Advanced Encryption Standard algorithm (AES) to ensure encryption and decryption of the data and capability list.
(d) SHA-512 message digest algorithm to ensure data integrity.
(e) Asymmetric Cryptography algorithm to establish secure communication between (a). Data Owner and User (b). Data Owner and Cloud Service Provider (c). Cloud Service Provider and User.
(f) DH algorithm is used to create session key between Cloud Service Provider and User.

The following diagram illustrates the proposed system architecture.

The following tables illustrates the proposed system symbol table.

Fig. 2 Proposed system symbol table

Description	Symbol
Data Owner	DO
Cloud Service Provider	CSP
Public Key	Pu
Private Key	Pr
Symmetric secret Key	Kt
Encryption	E
Decryption	D
Cloud Service Provider Pubic Key	CSP-Pu
Cloud Service Provider Private Key	CSP-Pr
Data Owner Public Key	DO-Pu
Data Owner Private Key	DO-pr
User Public Key	User-Pu
User Private Key	User-Pr
User Identity	UID
File Identity	FID
Access Rights	AR
Capability List	CPList
Message	Msg
Hash Algorithm	SHA-512
Choosen-secret key	XA/B
Calculated Public Key	YA/B
Shared Session Key	Ks
ith File	Fi
Message Digest	MD

6 Implementation

To implement the proposed system we have proposed four algorithms. In this section a possible implementation of the proposed approach is given.

Algorithm 1 Illustrates what Cloud Service Provider does whenever he obtains owners-data and access control list. Cloud Service Provider do decryption by his secret and data Owners-key. Cloud service provider storing scrambled data Access control list in Cloud Service Provider. Cloud Service Provider cannot see data because everything is encrypted with Symmetric secret key.

Step 1: Cloud Service Provider stores the received encrypted data and CPList.

Array = Rece (EkPuCSP (EkPrDO ((Fi)) − (CPList)) CPList − EKt (Fi) DkPrCSP (DkPuDO ((Array)).
Step 2: Cloud Service Provider do modification on encryption data.
Encrypted-data = Encrypted-data (FileID, BaseAdds).
Step 3: Cloud Service Provider updates Capability List
Capability List = CapabilityList (UserID, FileID, AccessRights).

Algorithm 2 Illustrates what data owner do when fresh file creates, Data Owner does fill the entries in Capability List containing UserID, FileID and AccessRights. Data Owner creates a secret-KT, do encryption on file using secret-KT. Then, Data Owner do encryption on modified access control list data using secret-KT with its own secret after the Cloud Service Provider secret then transmits to Cloud Service Provider. Whenever Cloud Service Provider obtains the files, it does modify on access control list and encryption data.

Step 1: Data Owner updates Capability List.
Capability List = Add(CapabilityList, (UserID, FileID, AccessRights)).
Step 2: Data Owners do encryption on Capability List,
Encryption File, secret transmits to Cloud Service Provider.
Sends(EPuCSP(EPrDO(CPList, (Fi), EPrDO(EPuUSR(KT, N + 1TimeStamp))))).

Algorithm 3 Illustrates the secure communication with cloud service provider and user with D-H technique.

Step 1: Registered user transmits data request to Cloud-Service-Provider.
Send(UserID, FileID, AccessRights)).
Step 2: Cloud Service Provider matches UserID, FileID, AccessRights with CapabilityList.

If Condition is successful return to Step 3 otherwise return to Step 6.

Step 3: Cloud Service Provider initiates Diffie-Hellman algorithm along user and securely transmits one time session
key(KS).
Step 4: Cloud Service Provider do encryption on data by using one time session key and transmits to user.
Step 5: User do decryption on obtained data generates MD of data.

If generated MD and obtained MD equal user confirms data is accurate no modification takes place.

Otherwise data is corrupted, so user immediately transmits data corrupted information to Data owner.

Step 6: Cloud Service Provider transmits illegal request message to User.

We called it D-H algorithm to do encryption using DH characteristics to prevent the outsider's security attack. Whenever user obtains authenticated credentials, the group-User can ask data from Cloud Service Provider. CSP-Cloud Service Provider initiates Diffie-Hellman with asking groupuser, if it is user is genuine KS is securely transmitted with cloud service provider and group-user. Cloud service provider do encryption using KS on MD and data and transmits to group-user.

Algorithm 4 Illustrates the threshold cryptography technique, moreover it illustrates decryption at user side. Whenever user obtains scrambled data from Cloud Service Provider, one user only cannot do decryption, first user has to transmit encrypted request message to all users of same group to enter shared secret key part components to decrypt the message. Proposed methodology not requires all key pieces (Because of threshold no. of key pieces).when threshold number of users entered shared secret key parts then User-1 can decrypt the data.

Step 1: authorized User1 receives Encrypted data File. Msg = Received ((Fi)).
Step 2: authorized User1 sends request message to other users in the group to enter secret key parts to decrypt.
Step 3: when threshold number of users entered secret key parts.
User1 decrypt the encrypted message-Msg to obtain the data.
File = DKt(Msg).
Otherwise he alone cannot decrypt the received encrypted data file.

To understand proposed approach, consider real time applications i.e. IT-Company is considered as data owner, he can do data storage at Cloud Service Provider, company employees considered as users they want to see the valuable data at Cloud Service Provider. Data Owner creates employees considered as users and make group based on conditions i.e. department wise and data owner generates secret key and allocates one single secret key to one group and also data owner do encryption using allocated secret key and that secret key is break into small pieces each piece of secret key is allocated to each group users. Here, Data owner creates MD of each data using secure hash algorithm (SHA-512), the generated MD is appended with data, so obviously it provides more data security and integrity. Data Owner then entrees the User ID, File ID and Access Rights in access control List to all users. Data Owner then do encryption on access control list with his own key after the key of Cloud Service Provider. Data owner transmit the encrypted information to Cloud Service Provider

The following tools will be used in the implementation of the proposed research work:

- *Zscaler Tool*
- *Cloud Sim version 3.0.3 Tool*
- *Java pairing based cryptography (jpb-1.2.0) Tool*
- *FOSS-Cloud Tool*
- *Mind Touch Tool*
- *Riak Tool.*

7 Security and Performance Analysis

Proposed approach enhances the data security in Cloud Storage in terms of.

7.1 Data Confidentiality

In the proposed approach, Data Owner stores its data at Cloud Service Provider in scrambled format. Here, the algorithm is used to generate secret key is very secure at data owner, and that secret key is shared in secure fashion between the group users and data owner. The cloud service provider is not able to read or do modification on data owner's valuable data. If any user wants data, first he/she has to request the data from cloud service provider by providing authentication credentials, afterwards cloud service provider validates user credentials if authentication is passed then cloud service provider again encrypts the data by onetime session key afterwards data will be send to the user. The onetime session key is securely exchanged between the group users by cloud service provider. Here, double encryption takes place, first encryption with secret key and second encryption with onetime session key. In the proposed approach single user not able to decrypt the data because he/she don't have complete secret key, he/she have only piece or part of the secret key. So that collusion attack never takes place in proposed approach.

7.2 Entity Authentication

In Proposed approach all three participating entities are authenticated systematically. First, User is authentication takes place at cloud service provider, second, data owner authentication takes place at cloud service provider when data owner transmitting data and access control list, third, user authentication takes place at cloud service provider when he/she has provide proper credentials to obtain the data, those credentials stored at cloud service provider database.

7.3 Data Integrity

Proposed approach uses SHA-512 hash algorithm to generate the MD of each data, data owner doing encryption on data and encapsulating data and MD, afterwards encapsulated data is send to the cloud service provider. Here, whenever any particular user obtains the encrypted form of data and MD. First he/she has to do decryption on data and MD, afterwards generate the MD and compare the MD with obtained MD, if it equals data integrity service is assured.

7.4 Data Access Control

Proposed approach assures access control by using the new method called access control list, it contains the details of User ID, File ID and Access Rights. Only Data Owner can do any task or computation on it. Cloud service provider has only read for the sake of secure access and also it checks the each user access rights details in access control list and act accordingly.

7.5 Performance Analysis

Proposed approach mainly deals with ensure data security over cloud storage when minimum ‘n’ number of users are available. In our proposed approach Data Owner transferred maximum of its load and computation to Cloud Service Provider and did only important activities by itself. Number of secret keys has minimized in the proposed approach (because in threshold cryptography, there is a single secret key corresponding to each group). Hence, minimizes the management and security concerns of secret keys and minimizes the extra overhead of computation time.

8 Conclusion

Cloud computing has received a lot of popularity in the last few years however, it also raises some security problems which may slow down its use. Cloud Computing is an emerging trend of provisioning scalable and reliable services over the Internet as computing utilities. Devising a way to provide data security for outsourced data at cloud storage system. Proposed approach mainly deals with ensure data security over cloud storage when minimum ‘n’ number of users are available. Proposed approach makes use of threshold cryptography technique at the user side. In this paper it is tried to provide data security more efficiently and also minimizes the use of more number of secret keys. So that proposed approach increases the performance of the system very efficiently. In this approach only one user not able to decrypt the data it requires at least minimum number of users otherwise decryption is not possible. Proposed approach is simple and user friendly. Future scope of work is that the approach requires to be studied in detail for efficient and optimized communication and data security among the cloud service provider, data owner, users. Proposed approach is better compared to previous solutions in terms of providing security and performance of the system and access control of the data. Proposed approach minimizes the usage of more number of secret keys by allocating one single secret key to entire group, so that all users must obtain the pieces of the secret key and whenever decryption takes place user has to enter the obtained pieces of secret key to decrypt the data.

References

1. Saroj SK, Chauhan SK, Sharma AK, Vats S (2015) Threshold cryptography based data security in cloud computing. In: 2015 IEEE international publication on computational intelligence communication technology
2. Chauhan SK, Sharma AK (2016) Secure data access in cloud computing. IOSR J Comput Eng (IOSRJCE) 13(5):35–42. e-ISSN: 2278-0661, p-ISSN: 2278-8727
3. Zhong H, Zhen H (2016) An efficient authenticated group key agreement protocol. 2007 41st Annual IEEE international Carnahan publication on security technology, pp 250–254, 8–11 Oct 2016
4. Madhavi Popli, Gagandeep (2016) Addressing security and privacy issues in cloud. In: Proceedings of TECHNOARETE-IAF International Publication, 13th Mar 2016, Chandigargh, India. ISBN: 978-93-85973-68-0
5. Ghule JJ, Khebade JR, Landage SA, Wavhal DN (2016) Cloud computing based data security by using threshold cryptography. Int J Innov Res Comput Commun Eng 4(2)
6. Choudharri NP, Biradar S (2016) A survey paper on data security in cloud computing using threshold cryptography and user revocation. Int J Eng Res Gen Sci 4(1). ISSN: 2091-2730
7. Suryawanshi R, Shelke SN (2015) Improving cloud storage security using public-key based homomorphic linear authenticator (HLA) and threshold cryptography. Int J Comput Trends Technol (IJCTT) 29(2). ISSN: 2091-2730
8. Mishra M, Bhatele M (2015) Improved cloud security approach with threshold cryptography. Int J Master Eng Res Technol 2(5)
9. Lomte S, Dudhani S (2015) Secure key for authentication and secret sharing in cloud computing. Int J Adv Res Comput Sci Softw Eng 5(6)
10. Pir RM (2014) Data integrity verification in cloud storage without using trusted third party auditor. Int J Eng Dev Res 2(1). ISSN: 2321-9939
11. Patel H, Patel P, Patel K, Achieving data integrity in cloud storage using BLAKE Hash function. Int J Eng Dev Res 2(2). ISSN: 2321-9939
12. Durga CBV, Venkata Narayana C, Reddy SSS (2013) Secure cloud storage with controlled data access assured deletion and data forwarding. Int J Sci Res Publ 3(10)
13. Do J-M, Song Y-J, Park N (2011) Attribute based proxy encryption for data confidentiality in cloud computing environments. In: IEEE 2011 first ACIS/JNU international conference on conference on computers, networks, systems and industrial engineering
14. Huang Y, Rine D, Wang X, A JCA based implementation framework for threshold cryptography

Multi-resolution Hierarchical Clustering by Vector Quantization

Mainak Pal, Preeti Mukherjee and Amit Konar

Abstract Clustering aims at grouping of objects or data-points based on a certain measure of similarity. Existing clustering algorithms estimate the measure of similarity of expected data-points to fall in a cluster with respect to presumed or computed cluster centroids. Such approach of distance measure between cluster centroid and possible data points to lie in cluster often results in misclustering, particularly for points equidistant to multiple cluster centroids. This paper offers an interesting solution to this problem by quantization of the attributes of the preferred cluster centroids and then checking the existence of the respective attributes of data-points within the quantized intervals for possible inclusion of data-point in the cluster. This approach, referred to as vector quantization offers additional merits of clustering at different user-defined resolutions of data-points of varying local density. Experiments undertaken confirm the superior performance of the proposed clustering over the state-of-art algorithms with respect to Jaccard coefficient on breast cancer dataset.

Keywords Clustering · Multi-resolutional · Vector quantization · Pattern recognition · Intelligent computing

1 Introduction

Quantization refers to truncation or round off of an Analog signal to represent it in discrete levels of fixed amplitudes. Clustering, on the other hand, is concerned with grouping of data-points based on certain similarity. A look into standard benchmark dataset reveals that data-points having close proximity in individual attributes often

M. Pal and P. Mukherjee contributed equally to this paper.

M. Pal (✉) · P. Mukherjee · A. Konar
Jadavpur University, Kolkata, India
e-mail: mainak.pal08@gmail.com

P. Mukherjee
e-mail: preetimukherjee08@gmail.com

A. Konar
e-mail: konaramit@yahoo.co.in

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_26

fall in the same cluster. Unfortunately, the existing clustering algorithms[1, 2] compare the datapoint generally by a distance measure between the cluster centroids and data-points to check the possible existence of data-points within a cluster. Quantization offers the freedom to examine the possible inclusion of individual attributes of a data-point within bounds of the respective attributes of the cluster centroids. This observation inspired the authors to propose a novel algorithm for automatic clustering by quantization of the attributes in the data-points. In this paper, the boundaries of attributes of the data-points in a cluster are selected by a fraction (such as α) of the dynamic range of the respective attributes of all the data-points. The choice of α is left to the user to cluster data at desired resolutions. Thus, the choice of increasing α offers the freedom of multi-resolutional hierarchical data-clustering where the lower value of α selects high density clusters (at higher level of the hierarchy) and the larger value of α returns clusters of relatively low data density (at the lower level of hierarchy). In this paper, each data-point is given the freedom to be a trial cluster centroid. However, a few of these data-points, whose (quantized) bounds of attributes include respective attributes of a large number of data-points, concurrently, are selected as the final cluster centroids. Because of the quantized bounds of attributes, two selected cluster centroids always have significant spacing in n-dimensional hyperspace, where n denotes the dimension of the data-point. The proposed algorithm thus facilitates automatic clustering with minimum risk of having false data-points in the clusters. The merit of the proposed algorithm is twofold. First, with the help of quantization of attributes of data-points, it eliminates the chance of false clustering (unexpected data-points) in a given cluster. Second, data-points of different density are clustered at different levels of the hierarchy, thereby providing significant information of data-points within a cluster, obtained at different resolutions. For instance, the high resolution clusters in certain applications, say image-processing, identify objects of interest, whereas the low density clusters offer the inter-relation among objects in the image.

2 Principles and Methodology

2.1 The Proposed Clustering Algorithm for Vector Quantization

The pseudocode of Vector quantization based clustering is presented below. The pseudocode includes 5 steps. In step 1, the dynamic range of each data dimension/attribute is computed by taking the difference of maximum and minimum values of each attributes in a given set of data-points.

In step 2, each data point is considered as trial cluster centroid, and the possibility that other data-points belong to the cluster of a presumed trial cluster centroid is computed. To accomplish this neighborhood of each attribute of a selected data-point is considered and same attribute of other data-points are examined for possible inclusion in the selected range around the selected attribute of the trial cluster centroid. The interval around the attribute of the trial cluster centroid is chosen as “fraction” of the dynamic range of the selected attribute. A set L_k is constructed for the k-th cluster centroid, where the list of point close enough to the cluster centroid are recorded. Here, by close enough, we mean that the attributes of the data-points must lie in the interval chosen for the respective attribute.

Step 3 is used to set the list of data points in descending order based on cardinality of the set L_k for all k, and top η% cluster of points are selected in descending order of cardinality.

The top η% selected clusters thus obtained may have overlap. Step 4 is used to merge clusters, if there is an overlap of the cluster pairs by more than 90% of either. The process is repeated for all pairs of clusters. The resulting clusters thus obtained are made free from the overlap.

In Step 5, the top ten clusters without overlap are printed.

Algorithm 1 Vector Quantization Based Clustering

Input: A set of l-dimensional data-points $V_k = [V_{k1}, V_{k2}, \ldots, V_{kl}]_{1Xl}$ for k = 1 to m; Scale factor $\alpha \in [0,1]$
Output: Cluster centroids $C_1, C_2, \ldots, C_n$ and corresponding data-points in the cluster.
Step 1:
for each data dimension j in V_k **do**
    $DR_j = \max_{1 \le k \le m} V_{kj} - \min_{1 \le k \le m} V_{kj}$ ▷ Where DR is the dynamic range
end for
Step 2:
for each data point $V_k : k = 1 \to m$ **do**
    for each cluster j in V_k (i.e. V_{kj}) **do**
        Initialize $L_k \leftarrow \phi$
        If $V_{k'j} \in V_{kj} \pm \alpha * DR_j$ **then** ▷ where $\alpha \in [0,1]$ is the Scale factor
            Then data-point k’ supports data-point k
            $L_k \leftarrow L_k U \{k'\}$ ▷ Save data-point k’ in the set
        end If
    end for
end for
Step 3: Sort L_k
for $k = 1 \to m$ **do**
    Sort in descending order of $|L_k|$, the cardinality of L_k
    Take top η% of the sorted clusters.
end for ▷ Rename L_k as $L_{\alpha 1}, L_{\alpha 2}, \ldots L_{\alpha r}$ where $r = (\eta * m)/100$

Step 4:

```
w ← 1
for j = 1 → r − 1 do
    for k = j + 1 → r do
        if |L_αj ∩ L_αk| ≥ 0.9 ∗ |L_αk| then
            merge L_αj , L_αk into L_w
        else
            L_w ← L_αj
            L_w+1 ← L_αk
        end if
        w ← w + 1
    end for
end for
```

Step 5: Print the elements in clusters L_w $\forall w$

2.2 Complexity Analysis

In every steps, inner loop operations like addition, subtraction or multiplication with constant predefined entities yields O(1) time complexity.

Now, coming to for loops, as the iterative variables have been incremented by unity, i.e., a constant quantity, so they have O(n) time complexity. Accordingly it can be inferred that step 1 has O(n), while steps 2, 4, 5 have O(n^2) complexities.

Step 3 is a modified version of Sorting algorithm, which has worst case complexity O(nlogn). Thus, we can conclude that the complexity of our algorithm is O(n^2). Where n denotes the number of data-points present in the dataset.

3 Multi-resolution Clustering

The clustering algorithm proposed is extended in multi-resolution settings like the well known DBSCAN (Density-based spatial clustering of applications with noise) algorithm [3]. The control in resolution is performed by suitable values of parameter α. The α value is initialized at 0.2 arbitrarily, and the data-points with high data density are clustered. The resulting clusters obtained for the smallest (possibilities) value of α have highest resolution. After the high resolution data points are clustered, the clustering algorithm is re-invoked again for the second pass with increased value of $\alpha = \alpha + 0.1$. Each time a set of data point of a selected resolution is clustered they are taken out from the list of data-points. The algorithm of multi-resolution terminates, when the number non-clustered data points goes below a user defined threshold. The proposed multi-resolution algorithm is called hierarchical as at different levels of the hierarchy the data-points are clustered at different resolutions.

4 Experiment and Results

The prepared Vector-Quantization (VQ) based algorithm has been tested on 32-dimensional breast cancer dataset [4] for a two class classification problem: malignant and benign. The experiment was repeated for different settings of α in [0.2, 0.5], and it was noted that the proposed VQ clustering out-performs the well known k-means clustering algorithm [5] with respect to Jaccard coefficient.

Results of 32-dimensional data is reduced to 2-dimension using T-SNE (T-distributed Stochastic Neighbor Embedding) technique for visualization and plotted for Ground Truth, KMeans [5] and VQ Clustering for different values of α in Table 1. A table for the Jaccard indices obtained for different values of α for VQ Clustering has been compared with other methods in Table 2.

The multi-resolutional hierarchical version of the VQ clustering is tested on synthetic data of sand and sand with white powder, sprayed at different density. It is observed that for $\alpha = 0.2$ to 0.4 all the clusters: pure sand, sand with low density of white powder and sand with high density of white powder can be clustered easily. For high dimension, this proposed algorithm outperforms the existing state of art algorithms. The highest Jaccard coefficient has been obtained as **0.83** for $\alpha = \mathbf{0.24}$.

5 Conclusion

The paper provides a novel and interesting approach of data clustering using vector quantization. Here, the attributes of the data points are quantized into intervals so as to accommodate the respective attributes of other close data-points within the quantized interval of the true cluster centroids. The proposed VQ algorithm has also been extended for multi-resolution hierarchical clustering like DBSCAN [3]. Experiments undertaken on benchmark standard benchmark breast cancer dataset [4] reveals that the proposed technique outperforms the state of art algorithms [1, 2] with respect to standard cluster validation index.

Table 1 Jaccard indices at various scale factors

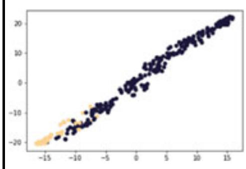
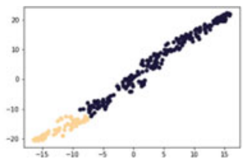
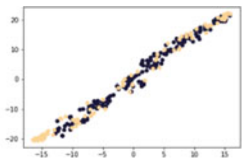
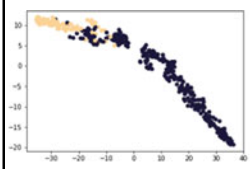
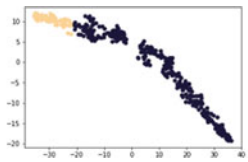
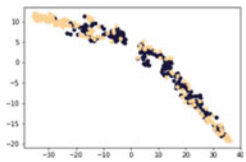
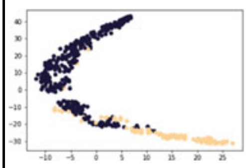
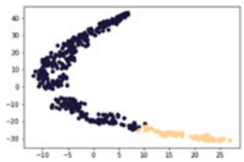
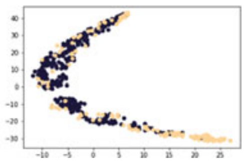
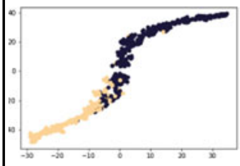
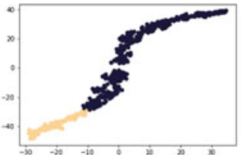
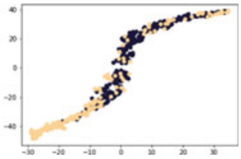
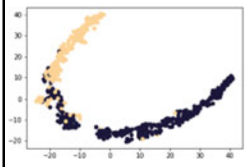
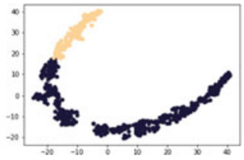
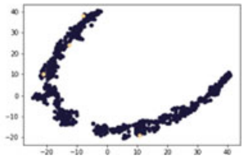
α	Ground Truth	KMeans	VQC
0.20		 Jaccard index = 0.82	 Jaccard index = 0.72
0.22		 Jaccard index = 0.82	 Jaccard index = 0.73
0.24		 Jaccard index = 0.82	 Jaccard index = 0.83
0.26		 Jaccard index = 0.82	 Jaccard index = 0.82
0.30		 Jaccard index = 0.82	 Jaccard index = 0.65

Table 2 Jaccard indices at different values of scale factor

α	Jaccard coefficient		
	K-Means	Fuzzy C-means	VQ clustering
0.2	0.82	0.81	0.7200
0.22	0.82	0.81	0.7300
0.24	0.82	0.81	0.8297
0.26	0.82	0.81	0.8176
0.28	0.82	0.81	0.8127
0.30	0.82	0.81	0.6490

References

1. Wu X (2007) Top 10 algorithms in data mining. In: Wu X, Kumar V, Ross Quinlan J, GhoshJ, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Yu PS, Zhou Z-H, Steinbach M, Hand DJ, Steinberg D (eds). Springer, London. http://www.cs.uvm.edu/icdm/algorithms/10Algorithms-08.pdf
2. Abdul Hameed W, Basha SS, Fuzzy C-means algorithm to diagnose breast cancer. Int J Recent Innov Trends Comput Commun
3. Ester M (1996) A density-based algorithm for discovering clusters. In: Ester M, Kriegel H-P, Sander J, Xu X (eds) AAAI, KDD-96 proceedings. https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf
4. Wolberg WH, Breast cancer Wisconsin (diagnostic) data set—diagnostic Wisconsin breast cancer database. In: Wolberg WH, Nick Street W, Mangasarian OL (eds). https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)
5. Radha R (2014) Using K-means clustering technique to study of breast cancer. In: Radha R, Rajendiran P (eds) World congress on computing and communication technologies. IEEE. https://ieeexplore.ieee.org/document/6755142

Certificate Verification System Using Blockchain

Harshita Khandelwal, Kritika Mittal, Shreeyaa Agrawal and Harsh Jain

Abstract The large number of counterfeit certificates in circulation is a huge problem that has been prevalent for a long time. Issuing of such certificates has become a business stemming from the need/want of the people for employment. Hard working people with legitimate degrees/certificates have to suffer the consequences of this phenomenon since they are denied what could have been theirs by the holders of these fake credentials. It a lot of cases this can prove extremely dangerous for example, consider a surgeon operating on the basis of a fake degree. It is situations like these that highlight the need for a system which can provide verification and authentication of the certificates, its issuer and the holder. In this paper, leveraging the tamperproof and non-repudiation property of blockchain, we propose and implement a certificate issuing and verifying model.

Keywords Blockchain · Fake certificate · Counterfeit

1 Introduction

With the exponentially growing population, the need for employment is also increasing. The current employment sector is not able to keep up with the demand and thus the competition for the limited jobs that exist has risen significantly. To filter through the numerous applications, the qualifications for the jobs have become more demanding. For example, a job previously requiring an undergraduate degree does not entertain anybody with less than a master's degree now. This has led to unnecessary pressure which in turn has led to people seeking the easy way out i.e. buying/forging certificates [1]. This need for qualification also inspired the establishment of several businesses which sell counterfeit certificates. A popular example would be the Pakistani software company, Axact. Under the alias of a software company, Axact was running a fake degree mill. An investigation into the company showed that over

H. Khandelwal (✉) · K. Mittal · S. Agrawal · H. Jain
Dwarkadas J. Sanghvi College of Engineering, Mumbai, India
e-mail: harshitaskh@gmail.com

K. Mittal
e-mail: kritikamittal17@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_27

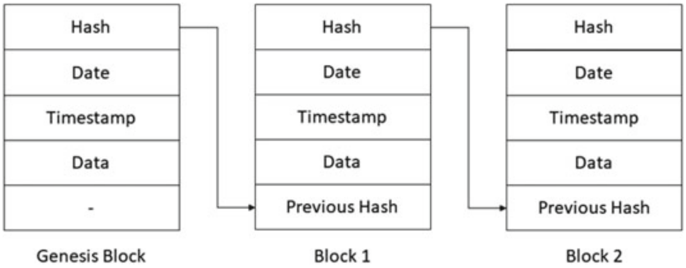


Fig. 1 Blockchain

200,000 fake certificates were sold and many of the holders of these certificates were holding high paying jobs [2]. There are many such websites which issue fake degrees and certificates. They exploit the lack of a system that checks the authenticity and validity of degrees/certificates. In this paper, we propose to build a decentralized record holding system using a blockchain.

1.1 Blockchain

Blockchain was first introduced to the world by Satoshi Nakamoto to support the idea of Bitcoin [3, 4]. Blockchain is a decentralized online ledger used for data sharing purposes. It can be monitored by the various nodes present in the network [5]. A blockchain typically has a certain number of blocks connected to each other by the means of a cryptographic hash (see Fig. 1). Each block present on the chain contains the hash of its previous block. Due to this property if any change is made within a block it's hash will change accordingly. This will cause a problem in the block to which it is linked to it. If this block is tried to be updated, its hash will change and the same problem will occur with the next block linked to it. Thus, changing one block in the chain would mean making changes to all the blocks after that block. This means making changes to a block is not possible and thus a blockchain is immutable. If any change is required then it is inserted as a new block in the chain.

1.2 Ethereum

Ethereum was first proposed by Vitalik Buterin, a programmer and cryptography researcher, in a white paper [6]. It is an open-source platform based on blockchain. It provides a distributed platform with which developers can build decentralized applications. Before the introduction of Ethereum, implementing a blockchain was a complicated process as Bitcoin [3] only focused on a peer to peer electronic cash system i.e. limited functionality. But now, Ethereum has made it possible to implement any function thereby form any decentralized application. Ethereum does so with the help of its tokens and Ethereum virtual machine (EVM). It has two types of tokens—Ethers, used for transactional fees and services and earned by mining and gas, used for smart contract (next section) execution. Ethereum virtual machine is a Turing complete software that uses gas to deploy and run smart contract functions thereby making applications more efficient.

1.3 Smart Contract

Smart contracts, also called self-executing contracts, are contracts whose terms and conditions are written as computer code [7]. They are stored and managed in a distributed and decentralized manner by a network of computers that run the blockchain. Transactions can be carried out in a traceable, conflict-free and a transparent way among various anonymous parties. No middleman is required to oversee the transactions. They are executed on virtual machine like the EVM in Ethereum software.

2 Related Works

Blockcert [8], is an open standard software built to help applications issue, publish and verify certificates. Blockcert selects a blockchain network like Ethereum or Bitcoin and uses its transactional data to its advantage to store certification records in JSON DL format. Being an open standard, it allows the user to have total control their records. But connecting to a Bitcoin network makes certificate issuing in Blockcert dependent on the Bitcoin prices which vary unstably.

0xcert [9], a similar work like Blockcert, creates a private blockchain network incorporating different types of cryptocurrencies. This leads to two types of charges being levied—for certification and transaction.

CertChain [10] is a certificate management platform which leverages the blockchain technology to provide certificate authentication and prevents counterfeiting of certificates. It uses a public blockchain network with bookkeepers and certificate authorities at each node. Bookkeepers are responsible for accessing the blockchain for recording the certificate operations. The data layer, network layer, extension layer and application layer make up the four-layer architecture which this system uses.

OpenCerts [11] is a platform which uses the Ethereum blockchain to overcome the problem of counterfeit certificates. Using this system educational institutes can create digital copies of the academic certificates that will be issued or have been previously issued. These digitized copies will be published onto the blockchain which provides the certificates with immutability.

All of the above systems provide an excellent solution for eliminating counterfeit certificates but none of them tackle the issue of identity theft. Also, cryptocurrencies are illegal in most countries, including India.

In this paper, we implement a blockchain based web portal to overcome the issues mentioned above.

3 Proposed System

The proposed system uses Ethereum blockchain along with a web portal. It would have three major identities—user, organization and the company that verifies the certificates. The organizations are responsible for creating the certificates and uploading them onto the blockchain. These certificates are issued to the user. The verifier can cross check the authenticity of the certificate and the user claiming it.

3.1 Process

There are 3 main users of the system: The user, Organization, the company that verifies. The user is the person to whom the certificates belong and he/she can share it with the companies. The organizations are responsible for issuing the certificates. The verifier can check the authenticity of the certificates and proceed accordingly. The access to the chain is provided to the authorised companies (the organisations and the verifiers). The organisations can upload the certification along with some other necessary details onto the chain and the verifiers can view the data on the chain to authenticate the user and his/her claimed certifications. The user receives the certificates along with the proof of certification from the organisation which he/she can then send to any company (verifier).

Publishing. Each organization is verified before it is provided entry to the chain and a public-private key pair is assigned to it upon inclusion.

Once a user completes a particular course/diploma/degree, the organization produces a digital version of the certificate. This certificate contains all the relevant details along with the user id of the organization assigned to it in the chain.

After the certificate is produced, the organization converts the copy to a base-64 string [12] and hashes it using the SHA-512 algorithm [13]. This hash is signed using the private key of the organisation is sent to the blockchain. The blockchain automatically verifies the signature before proceeding with the transaction. If the signature is verified, a transaction is run to add the hash to the blockchain. A transaction ID is generated after its success.

The transaction ID obtained and the digital copy of the certificate are sent to the user. If the user is present in the system, they are uploaded to the portal under the user's account and is visible on their dashboard. These files can be downloaded by

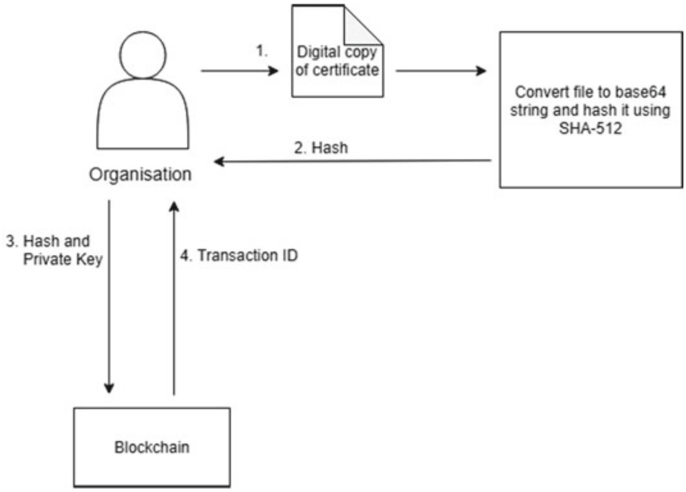


Fig. 2 Publishing certificates

the user. If the user is not present in the system, an email is sent to the user which contains the certificate and the transaction ID.

If the user uses the system to send the certificates to a company, they do not need to send the transaction ID along with it. The transaction ID is automatically sent along with the certificate. If the user uses some other method to send the certificates to a company then they need to send the transaction ID along with the certificate (Fig. 2).

Verify. A company is sent the digital copy of the certificate by the user. If the user used the system then all the received certificates are displayed to the company on its dashboard along with a verify option. The presence of the user on the system and the visibility of the certificates in their respective profiles is proof enough for the authentication of the certificates. But still if the company wants to verify the certificates, they can select the verify option. If the company is sent the certificate and the transaction ID using some other method, then the company can upload the certificate and enter the transaction ID manually before selecting the verify option.

On choosing to verify, a function present in the smart contract is executed. This function converts the certificate copy to a base 64 string and hashes it once again. The transaction ID received previously is used to query the blockchain to retrieve the hash which was added to the chain by the organization that issued the certificate to the user. This hash obtained from the blockchain is compared to the hash that was produced from the certificate copy [14]. If the hash match it means that the certificate copy is authentic i.e. the copy was not tampered with. If even one pixel of the certificate is changed, the hash of that copy changes. So, it is important that the user does not tamper with the file.

If the user exists on the system and the system is used for sending the certificates, the company does not need to be concerned with identity theft. The organisation has

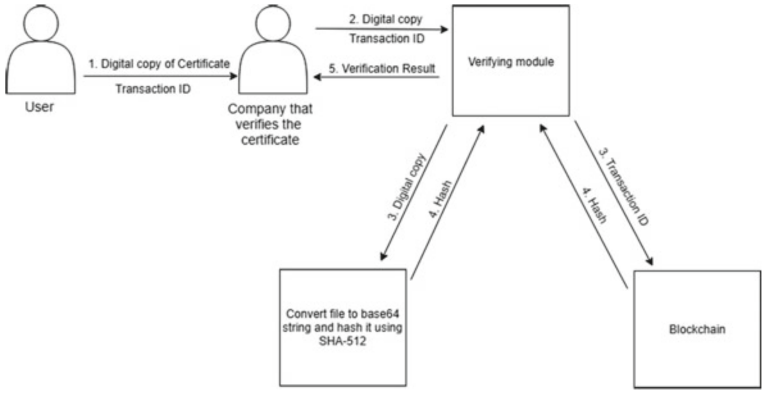


Fig. 3 Verifying certificates

issued the certificate to a particular account and an account can be held by only one user. If the user is not present on the system, protection against identity theft cannot be guaranteed (Fig. 3).

4 Conclusion

Counterfeit certificates are a long existing problem in our employment system due to lack of communication between the certifier and employer as well as due to lack of a proper record system. Blockchain is an emerging technology that provides us with a distributes, tamper-proof ledger which is an appropriate solution for the problem. Hence, we use blockchain to create an ideal system for certifiers to record published certificates and the employers to check their authenticity and validity. The system also ensures authentic identity of all three entities involved—user, organization and verifier.

References

1. Rampell C (2014) The college degree has become the new high school degree. The Washington Post (Online). https://www.washingtonpost.com/opinions/catherine-rampell-the-college-degree-has-become-the-new-high-school-degree/2014/09/08/e935b68c-378a-11e4-8601-97ba88884ffd_story.html?noredirect=on&utm_term=.36ba2e673449. Assessed on 6 May 2019
2. Walsh D (2015) Fake diplomas, real cash: Pakistani company axact reaps millions. The New York Times (Online). https://www.nytimes.com/2015/05/18/world/asia/fake-diplomas-real-cash-pakistani-company-axact-reaps-millions-columbiana-barkley.html. Assessed on 6 May 2019

3. Nakamoto S (2019) Bitcoin: a peer-to-peer electronic cash system (Online). https://bitcoin.org/bitcoin.pdf. Accessed on 6 May 2019
4. Narayanan A, Bonneau J, Felten E, Miller A, Goldfeder S (2019) Bitcoin and cryptocurrency technologies. Available at: https://d28rh4a8wq0iu5.cloudfront.net/bitcointech/readings/princeton_bitcoin_book.pdf?a=1. Accessed on 6 May 2019
5. Blockchain (Online). https://blockchainwolf.net/blockchain-technology/what-is-blockchain-easiest-explanation-ever/. Assessed on 6 May 2019
6. Wood G (2014) Ethereum: a secure decentralized generalized transaction ledger. Ethereum Proj Yellow Pap 151:1–32
7. Tonelli R, Destefanis G, Marchesi M, Ortu M (2018) Smart contracts software metrics: a first study. arXiv preprint arXiv:1802.01517
8. Blockcert (Online). https://www.blockcerts.org/. Assessed on 6 May 2019
9. 0xcert (Online). https://0xcert.org/use-cases/certification/. Assessed on 6 May 2019
10. CertChain (Online). https://certchain.io/. Assessed on 6 May 2019
11. OpenCerts (Online). https://opencerts.io/. Assessed on 6 May 2019
12. Sumartono I, Siahaan APU, Arpan A (2016). Base64 character encoding and decoding modeling. Int J Recent Trends Eng Res 2:63–68
13. Cheng J-C, Lee N-Y, Chi C, Chen Y-H (2018) Blockchain and smart contract for digital certificate. In: 2018 IEEE international conference on applied system invention (ICASI)
14. Huynh TT, Huynh TT, Pham DK, Ngo AK (2018) Issuing and verifying digital certificates with blockchain. In: 2018 International conference on advanced technologies for communications (ATC)

IoT Based Automatic Plant Watering System Through Soil Moisture Sensing—A Technique to Support Farmers' Cultivation in Rural India

Syed Musthak Ahmed, B. Kovela and Vinit Kumar Gunjan

Abstract Agriculture and Cultivation of paddy, wheat and vegetables basically takes place in Rural areas where technology isn't available to that extend where this kind of huge production of grains and vegetables can be automated to help farmers. Farmers spend most of their time in the Agriculture field for watering the crop by leaving other works. Hence to help farmers from staying on the field whole day, we came up with a project which senses soil moisture and based on the data this system automatically turns ON the water pump into the field, and when the soil reaches enough moisture level, then water pump automatically gets turned OFF. Hence this concept may provide a long term solution to the farmers for a maintenance free agriculture where farmers don't have to stay on the field breathing toxic chemicals and spoiling their health. The proposed project also have other features like sensing the Ambient temperature and humidity in the Agricultural field, sensing daylight intensity and rainfall detection on the field. Hence this inexpensive project can provide a solution for many agricultural and health related problems. To implement this project in a ease of access and updated way, we have incorporated IoT platform where the farmer can monitor all these field parameters over internet on their smart phone application. Therefore this could be a milestone in rebuilding the future.

Keywords Nodemcu · DHT11 · IoT · Blynk · Soil moisture · Relay

1 Introduction

The Internet of Things (IoT) has transformed the way we live in this world [1]. Now we have Smart connected homes, cars, more integrated industries, smart cities, smart villages [2] all these because of IoT systems.

S. M. Ahmed (✉) · B. Kovela
S R Engineering College, Warangal, Telangana, India
e-mail: syedmusthak_gce@rediffmail.com

V. K. Gunjan
Department of Computer Science and Engineering, CMR Institute of Technology, Hyderabad, India

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_28

Our Earth will reach a population of 9.2 billion by 2048. Hence to feed this huge population, the old agricultural practices needs to be combine with IoT [3]. The dramatic change in climate and weather has left a environmental impact on farming, hence smart farming through the use of IoT will help farmers to reduce waste and increase productivity [4–6].

IoT application help farmers to collect data regarding health and condition of their crop [7]. By using internet of things (IoT) and sensors network technology we can control water wastage by adapting green farming methods in irrigation [8].

Every second, the field conditions like soil moisture, rainfall, humidity, temperature and light intensity are sensed and this data is sent to IoT cloud server and will get updated on smart phone of the farmer, so that the farmer can monitor the field without being present on the field [9].

This IoT system will be very helpful in solving a wide range of human related problems in upcoming days [10].

2 System Design

The designed system consists of microcontroller (nodemcu) which has both WiFi and microcontroller capabilities and is a best suit for IoT application projects. The representation of such a system is shown in Fig. 1

The system further consists of bunch of sensors like soil moisture sensor to sense the moisture level in soil, DHT11 sensor for temperature and humidity monitoring, LDR sensor for light intensity and finally a raindrop sensor for rainfall detection. Depending upon the moisture content in soil, AC water pump is switched ON/OFF by a relay which is inturn controlled by NodeMCU. A logic level converter is added because NodeMCU (microcontroller) is a 3.3 V device which cannot drive or receive

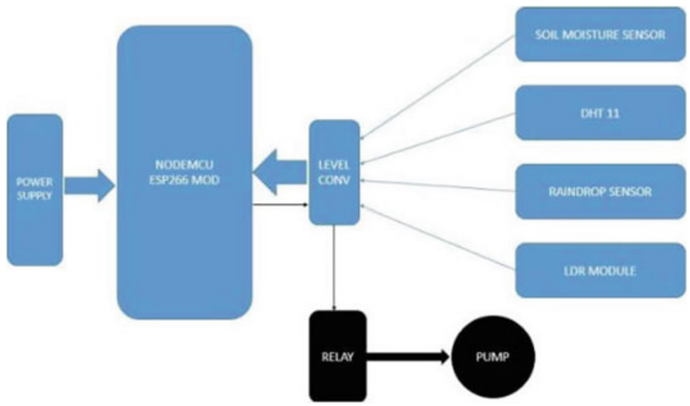
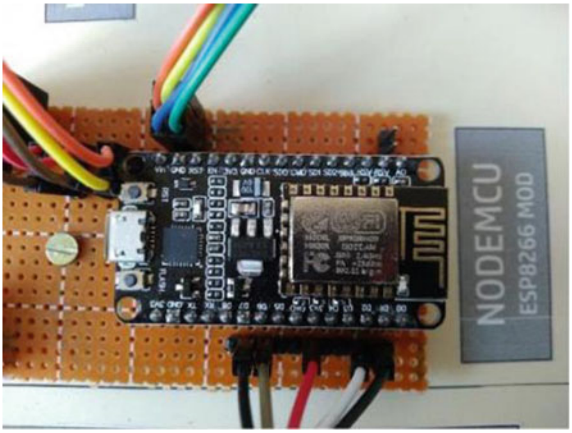


Fig. 1 Block diagram representation of system design

Fig. 2 NodeMCU module



data from a 5 V device. Hence to convert 5 V logic level to 3.3 V we used a 4 channel logic level converter.

3 Hardware Requirements in Implementation of Module

The following components are used for the Implementation of project.

3.1 NodeMCU

It is an open source platform having firmware (which runs on the ESP8266 WiFiSoC) and hardware (based on the ESP-12 module). The "NodeMCU" has several features such as programable WiFi module, Arduino-like (software defined) hardware IO, PCB antenna, WiFi networking, Event-driven API. The NodeMCU board is shown in Fig. 2.

3.2 Soil Moisture Sensor

Farmers can manage their crops more effectively and efficiently by measuring the soil moisture content. This is achieved using a soil moisture sensor. This sensor measures the amount of water contained in the soil by knowing the soil properties such as 'electrical resistance, dielectric constant, or interaction with neutrons, as a proxy for the moisture content' [11]. Such a sensor is shown in Fig. 3.

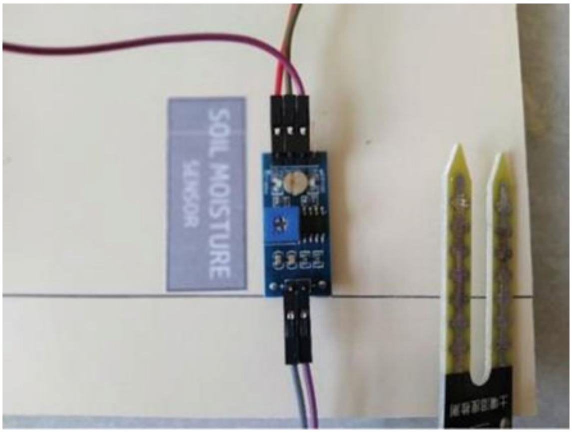


Fig. 3 Soil moisture sensor

By finding the soil moisture contents, farmers will be able to feed significant amount of water to their crops and also be able to improve the quality of the crop.

3.3 DHT11 Sensor

It is an ultra low-cost digital temperature and humidity sensor. It operates between 3 and 5 V power supply and I/O with humidity readings accuracy of 5%, temperature with ±2 °C accuracy. It measures the surrounding air humidity and temperature and gives out a digital signal. The structure of such a Sensor is shown in Fig. 4.

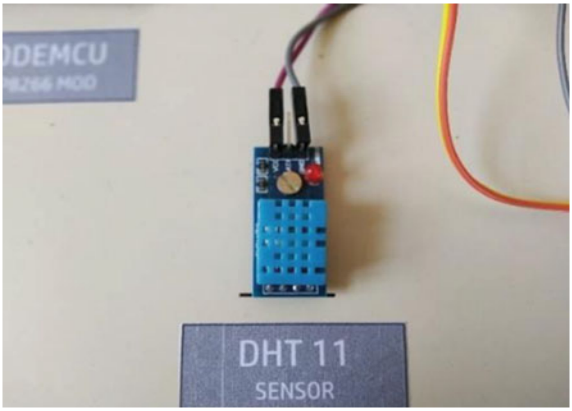


Fig. 4 DHT11 sensor

3.4 Water Pump

Water pump is used for controlling the supply of water to the field. The water pump system used in our project is shown in Fig. 5. The operation of the pump is controlled by a relay to water the crop. Whenever the field gets dried up i.e. moisture content gets reduced, the system makes the relay to operate which intern switches ON the motor till the required conditions of the soil is reached. Once the moisture level reaches the desired water for the selective crop the motor is switched OFF via the relay. The relay circuitry is shown in Fig. 6.

Fig. 5 Water pump system

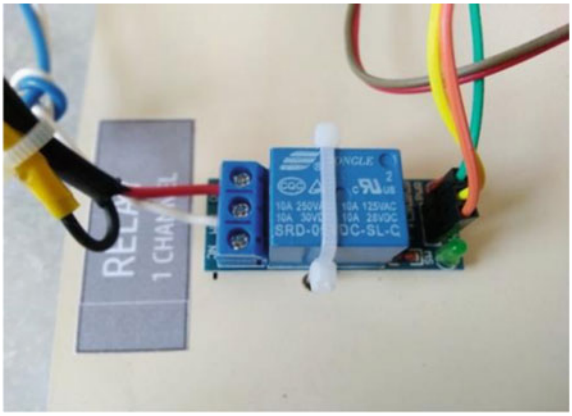


Fig. 6 Relay module for switching motor

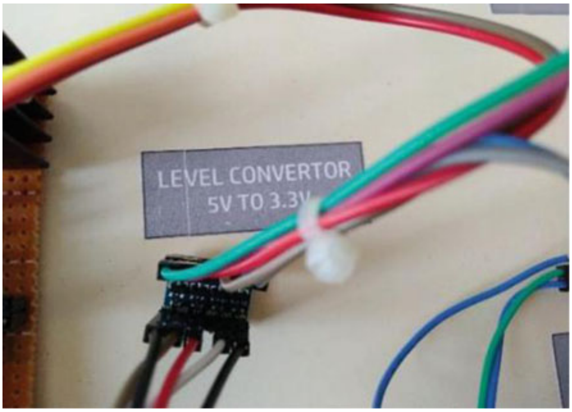


Fig. 7 Logic level converter

3.5 Logic Level Converter

The logic level convertor is used to step up or step down the voltage levels as required for the circuit. Here SparkFun bi-directional logic level converter is used in our application. This device steps down 5 V signals to 3.3 V and or steps up 3.3–5 V at the same time. This is shown in Fig. 7.

3.6 LDR Module

The Light Dependent Resistor (LDR) is a light-controlled variable resistor. The resistance of a photoresistor varies with variation of light intensity falling on it. In other words, it exhibits the property of photoconductivity. A photoresistor finds applications in light-sensitive detector circuits, and light-activated and dark-activated switching circuits as shown in Fig. 8.

In the dark, a photoresistor will have a high resistance of several megohms (MΩ), while in the light a low resistance few hundred ohms.

3.7 Raindrop Module

The rain sensor module is shown in Fig. 9. The function of this sensor is to shut down the system in the event of rainfall. It consists of a water conservation device to control the water supply to the fields during rain fall. It causes the Irrigation System to automatically shut down in the event of rain. This saves power in system operation and time of a farmer.

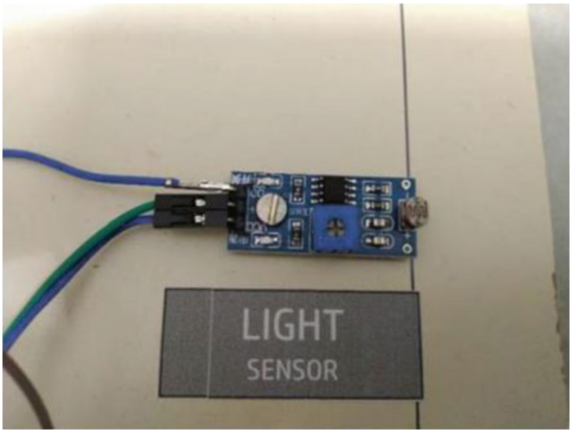


Fig. 8 LDR module to detect light sensitivity

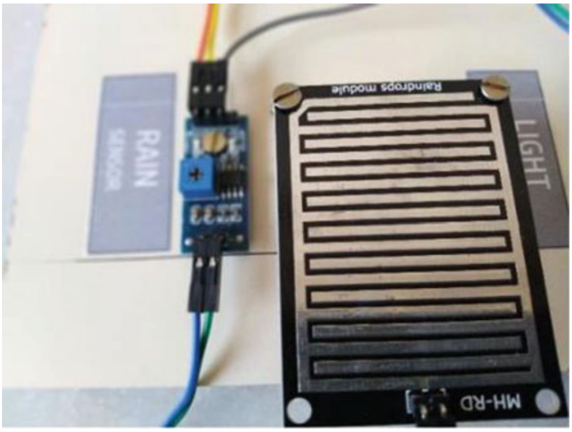


Fig. 9 Raindrop sensor circuit

4 Software Requirements in Implementation

The software used for monitoring sensors data over IoT is based on Blynk IoT Platform. It is nothing but an IoT platform with customizable mobile Apps, private cloud, rules engine, and device management analytics dashboard. The Blynk displayed on App for various conditions of soil moisture, humidity, temperature, rainfall and light intensity are shown in Fig. 10.

5 Results

The complete Test set up for Automatic Plant Watering System Through Soil Moisture Sensing is shown in Fig. 11.

Fig. 10 Blynk App display

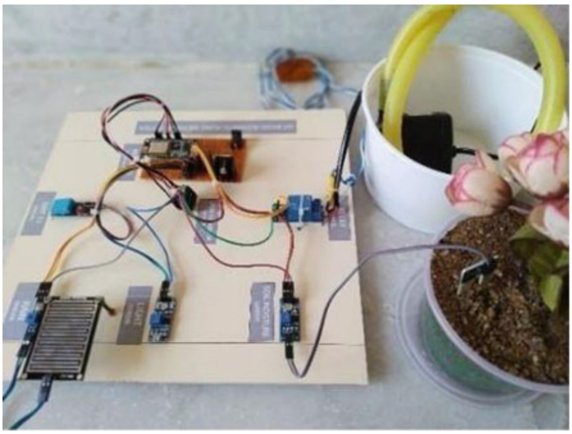


Fig. 11 Setup of the proposed system

For the project to function over IoT, Nodemcu should be provided with WiFi credentials like SSID and Password to access the cloud server. After powering everything, Nodemcu collects data from all sensors. Soil moisture sensor is an active low device which gives a low signal if the soil is wet and a high signal if the soil is dry. We also have a 1 channel relay module interfaced with Nodemcu, which is also an active low device.

When the soil in the field gets dried, the soil moisture sensor sends a high signal to Nodemcu, then according to the program, nodemcu gives a low signal to relay which turns it ON. Hence water pump starts pumping water to the field.

Now when the soil gets enough wet the soil moisture sensor sends a low signal to nodemcu which intern sends a high signal to relay which turns it OFF. Hence water pumping to the field is stopped.

All the data from sensors and the status of water pump is sent to the BLYNK cloud server, so that the farmer can monitor it from the Android Application.

6 Conclusions

Thus, IoT based automatic plant watering system has been implemented. NodeMCU collects data from all sensors and uploads to Blynk IoT server. The soil moisture sensor senses the moisture content present in the soil. If the soil is dry it sends a high signal to nodemcu which intern switches ON the water pump via the relay. If the soil gets enough wet then the sensor sends a low signal to Nodemcu, hence the water pump gets turned OFF automatically. Thus the complete system is tested. It is more efficient with affordable cost for implementation in small home gardens and can be extended for large scale cultivation of farmers.

References

1. Sandeep CH, Kumar SN, Kumar PP (2018) Security challenges and issues of the IoT system. Indian J Public Health Res Dev (IJPHRD) 9(11):748–753
2. Ahmed SM, Sai Chandu T, Rohit U, Naveen G, Naveen S (2019) IoT based garbage disposer for educating rural India. In: International conference on data sciences, machine learning and applications (DSMLA 2019), 29–30 Mar 2019, Springer Conference
3. Patil MSS, Malvijay AV (2014) Review for arm based agriculture field monitoring system. Int J Sci Res Publ 4(2)
4. Tani FH, Barrington S (2005) Zinc and copper uptake by plants under two transpiration rates. Part I. Wheat (*Triticum aestivum* L.). Environ Pollut 138:538–547
5. Gutierrez J, Villa-Medina JF, Nieto-Garibay A, Porta-Gándara MA (2013) Automated irrigation system using a wireless sensor network and GPRS module. IEEE
6. Patil SS, Malvijay AV (2016) Review for ARM based agriculture field monitoring system. Int J Sci Res Publ 4(2). Nagothu SK (2016) Weather based smart watering system using soil sensor and GSM. In: 2016 World conference on futuristic trends in research and innovation for social welfare (startup conclave). IEEE Xplore: 06 Oct 2016

7. Jury WA, Vaux HJ (2007) The emerging global water crisis: managing scarcity and conflict between water users. Adv Agron 95:1–76
8. Arun C, LakshmiSudha K (2012) Agricultural management using wireless sensor networks—a survey. In: 2nd International conference on environment science and biotechnology (IPCBEE), vol 48. IACSIT Press, Singapore
9. Nagothu SK, Anitha G, Annapantula S (2014) Navigation aid for people (Joggers and runners) in unfamiliar urban environment using Inertial Navigation. In: 2014 Sixth international conference on advanced computing (ICoAC), pp 216–219
10. Yuan G, Luo Y, Sun X, Tang D (2004) Evaluation of a crop water stress index for detecting water stress in winter wheat in the North China Plain. Agric Water Manage 64(1):29–40
11. Nagothu SK (2016) Weather based smart watering system using soil sensor and GSM. In: 2016 World conference on futuristic trends in research and innovation for social welfare (startup conclave). IEEE Xplore: 06 Oct 2016

IoT Enabled Detection of Suspicious Human Behavior for ATM Environment

Vaishnavi R. Mali, Anil R. Surve and V. R. Ghorpade

Abstract Nowaday's security is the main thing in all industries. The most serious thing facing the industries which provide financial services and the retail markets is ATM physical attack. The industry which provides financial services and the retail markets lose their lots of money due to physical attack. Today, however, ATMs give a considerable measure of administrations to the clients; they are still dominatingly utilized for their essential capacity of pulling back cash. ATM security is the main issue that has been addressed in the proposed work by studying current scenarios where the traditional way is to take actions after an attack has occurred. But using the CCTV cameras present in the ATM we can prevent these attacks from happening. We can take actions before an incident is going to happen. The proposed system is an attempt to analyze the human in the ATM with the various parameters mentioned in this paper and send alert to the respective entity utilizing IOT platforms if he/she found suspicious. So that actions can be taken before the loss.

Keywords Suspicious behavior · ATM environments · IOT platforms · Context-aware environment · Image processing · Deep learning

1 Introduction

Detection of irregular visual examples in pictures and in the video is helpful for a number of assignments. Recognizing Suspicious behavior or surprising items is critical for observation and checking. Examination of suspicious behavior alludes to the program identification of suspicious behavior through a computer or other

V. R. Mali (✉) · A. R. Surve
Department of Computer Science and Engineering, Walchand College of Engineering, Sangli, Maharashtra 416415, India
e-mail: vaishnavimali16@gmail.com

A. R. Surve
e-mail: anil.surve@walchandsangli.ac.in

V. R. Ghorpade
Bharati Vidyapeeth College of Engineering, Kolhapur, Maharashtra 416012, India
e-mail: vijayrghorpade@hotmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_29

equipment gadgets. In a video observation framework, the prior those suspicious practices are discovered, the less unsafe they will be. Along these lines, how to discover a strategy that can consequently investigate suspicious behavior and can completely or in part replace manual monitoring have turned into a debated issue in the field of computer vision.

From many applications, one application computer vision is utilized in surveillance dependent on strategies. Enormous progressions have been accomplished and numerous new strategies have been actualized in this area to carry out automatic abnormality discovery. Abnormality identification strategies can work into three, i.e. supervised, semi-supervised, unsupervised anomaly detection. Procedures prepared in supervised mode expect the accessibility of a training dataset which has marked examples for ordinary and anomaly.

Banks ATM is open places that interest high security, and the utilization of video reconnaissance inside them is a generally acknowledged practice. The expanding number of ATM assaults is a developing concern. Assaults are going on a regular routine. There is a number of cameras located for the security purpose, but they act as a blind and just record the video for post-incident manual analysis. To reduce the manual analysis and to take preventive measures before the loss is the main objective of the proposed systems. Video Surveillance framework with Active Deterrence gives Surveillance in ATM by checking ATM premises day and night. A context-aware system utilizes context to deal with relevant information and services for the intended user [1]. IoT platforms available now that provide the option to deploy data on clouds and send the notification or data through the platform [2].

In this paper, we propose an IoT enabled detection of suspicious human behavior for ATM environment with minimal effort. After the first two sections which is introduction and applications, the architecture is characterized in Sect. 4, the applications and the research problems are characterized in Sect. 2, Related work in Sect. 3 and project methodology are a reference to Sect. 5. Furthermore, conclusion and future work of this paper is in the last section.

2 Applications

A. ATM Environments

Implementing this system in ATM environments will detect suspicious behavior of human and immediately alert can be sent to the respective entity so preventive actions can be taken and it will reduce the manual monitoring as well.

B. Other Environments

Security is the main concern in public/crowded area, home, school, hospitals, etc. Tragedies can happen everywhere, so the security of these places is also important. Further, this system can be implemented in crowded environments also [3].

3 Related Work

In this section, we have briefly reviewed related works which have done to topics.

The expression “irregular” relies upon the setting in which the “normal” or “substantial” are characterized. Districts in the watched information which can be created utilizing substantial touching lumps of information from the database are viewed as likely, while locales in the watched information which can’t be formed from the database (or can be made, yet just utilizing little divided pieces) are viewed as impossible/suspicious [4].

In foggy conditions, permeability diminishes and causes numerous issues. Less visual because of foggy conditions while driving builds the danger of street mishaps. VESY (Visibility Enhancement Saliency YOLO) sensor and YOLO (You Only Look Once) are used. The proposed combination calculation gives the bounding boxes of the association of the items identified by Saliency Map and YOLO Algorithm in this manner demonstrating to be a practical answer for constant applications [5].

Article identification is one of the key programming segments in the up and coming age of self-sufficient vehicles. Old style PC vision and AI approaches for item discovery as a rule experience the ill effects of the moderate reaction time. YOLO (You Only Look Once) calculation, take care of this issue without exactness misfortune, the exhibit of the utilization of the most up to date YOLOv3 calculation for the recognition of traffic members are given [6].

A strategy to recover the distinguished question from the database for protest acknowledgment and distinguishing proof utilizing content-based picture recovery. At last, a product based recreation utilizing MATLAB was performed and the outcomes of the led tests demonstrated an amazing reconnaissance framework that can all the while play out the following, scene learning, abnormality location in a scholarly domain with no human intercession [7].

4 Architecture

As surveillance winds up ubiquitous, a key objective of surveillance is to recognize irregular pattern. The explored investigations are investigated over five perspectives: surveillance targets, abnormality definition and suspicions, sorts of sensors utilized and element extraction forms, learning strategies, and displaying calculations [8].

Figure 1 represents the architecture of the proposed system. In that there are various modules, in the first module, we will capture the data in the form of video or will use the dataset, in the second module, using suspicious detection mechanism we will detect any suspicious activity. In the third module, if the suspicious activity is detected then alert will be sent to the particular entity or person using IOT platforms.

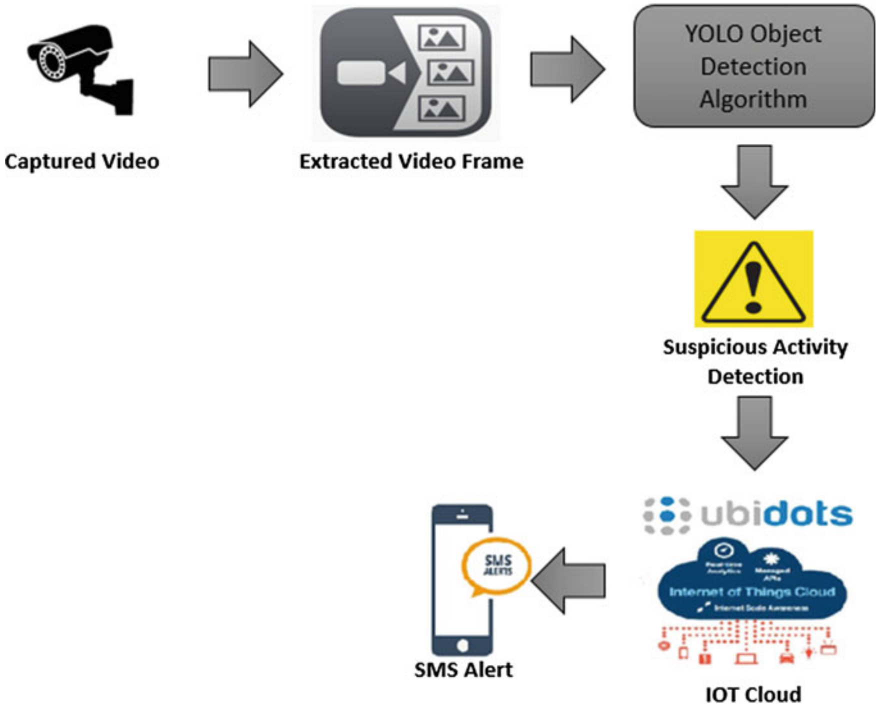


Fig. 1 System architecture

5 Methodology

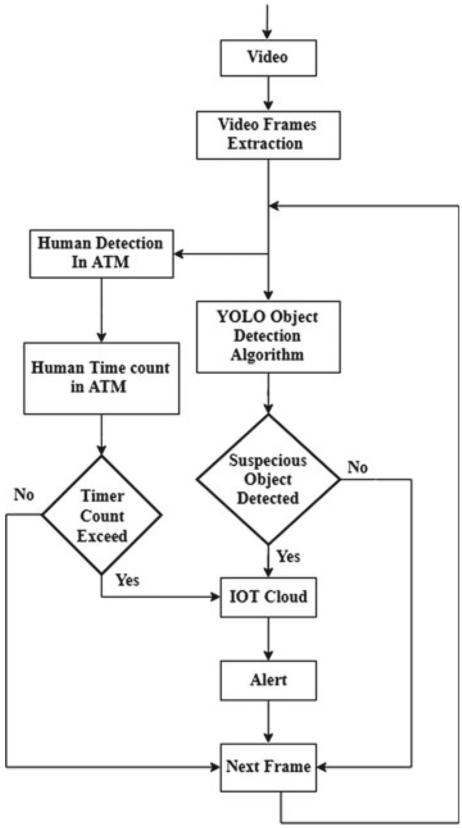
5.1 YOLO Object Detection Model

The calculation applies a neural system. The system partitions the frame into a S × S matrix and thinks of jumping boxes, which are boxes drawn around pictures and anticipated probabilities for every one of these locales. The bounding boxes are weighted by the related probabilities. For class forecast, free calculated classifiers are utilized [9]. Yolo engineering is progressively similar to FCNN (completely convolutional neural system) and passes the picture (n x n) once through the FCNN and yield is (m × m) forecast [10].

Steps in Fig. 2 are as follows:

1 We will be using YOLO object detection pre-trained model. We will give a video input to the system.
2 Video will be passed to the YOLO object detection pre-trained model.
3 After object detection, we will look for objects like a knife and gun. Simultaneously, we will count time for time spent in the ATM room.

Fig. 2 Flowchart



4 If a knife or gun is detected then alert will be sent to the respective entity. And if the human spent more time in the ATM room than threshold time then alert will be sent through IOT cloud.
5 If any suspicious behavior is not detected then it will process the next frame.

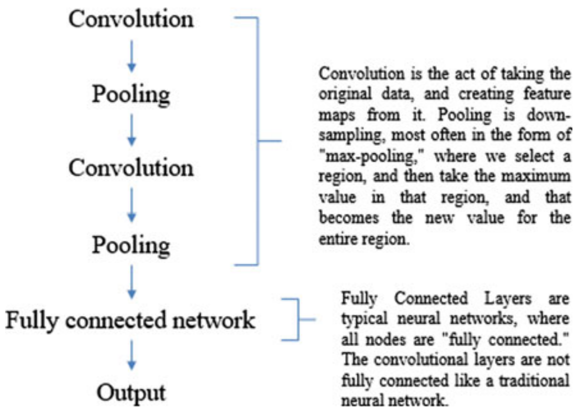
Extraction and classifications will be done. CNNs will train using Feed forward and back propagation algorithm [11].

5.2 CNN Algorithm

Convolutional neural networks layer includes three types, First Convolutional, pooling, and fully-connected layer [12, 13].

The basic structure of CNN Algorithm is given in Fig. 3.

Fig. 3 Structure of CNNs algorithm



5.3 Ubidots: An IOT Platform

Ubidots is an IoT Platform which is code less intended to engage us to model and scale our IoT activities to creation while improving and streamlining our general surroundings with sensor information. Utilize the Ubidots stage to send information to the cloud, then configure activities and alarms for ongoing information. Interface equipment to Ubidots cloud effectively with in excess of 200 clients demonstrated libraries, SDKs, and instructional exercises to direct your combination over HTTP, MQTT, TCP, UDP, or by Parsing custom/mechanical protocols [14, 15].

6 Experimental Result

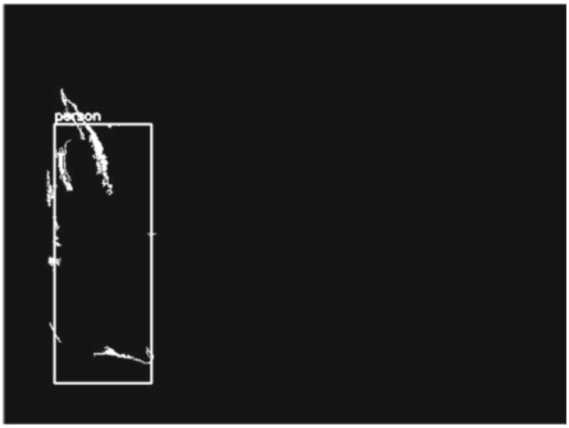
A. Out Frame Detection

We count the time human spent in the ATM room. We have set a threshold if human spend more time than the threshold the alert will send trough Ubidots IoT platform to the respective entity like security department. We convert the extracted video frames in binary format. In that frames Human will be in white pixels. So we define a bounding rectangle as a non suspicious zone and if human spent more time outside of that defined area then the system will check the time and send the alert. To find out whether a human is in inside or outside, we count the number of white pixels. If white pixels are more in the outside than inside then the system will consider human is in the suspicious area and start counting the time. If the time exceeds than the threshold then the alert will be sent.

Figure 4 video frame in which a person is detected outside of the defined rectangle area. Figure 5 the binarized frame of the same video. With the help of this, we will count the number of white pixels inside the defined rectangle area as well as outside defined rectangle area.

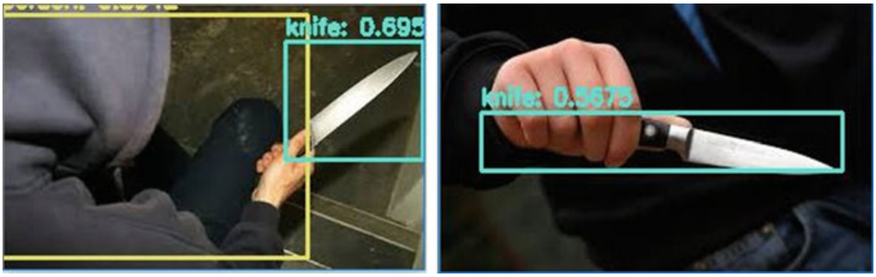
Fig. 4 Human Detection

Fig. 5 Human Detection



B. Suspicious Object Detection

If suspicious objects like knife, gun, etc. are detected in the system then immediately the system will send alert to the respective entity like the security department of ATM.



7 Conclusion

We have presented a one of the crucial security aspects ATM security with IoT Enabled Detection of Suspicious Human Behavior for ATM Environment system. In this system we use deep learning algorithm for the detection of human suspicious behavior and IOT platform to send notification to the respective entity when suspicious behavior is detected.

8 Future Work

Suspicious behavior once successfully detected by using the system. We can further expand the system to the gender and age of the person who will act suspicious. Human behavior is frequently indeterminate, and here and there it is influenced by feeling or condition. Recognition can be helped to perceive human behavior which later can aid in examining suspicious occasions [16].

Acknowledgements We would like to express our gratitude to all the authors of research papers. Also towards the Computer Science and Engineering Department of Walchand College of Engineering, Sangli for the support and guidance.

References

1. Surve A, Ghorpade V (2017) Pervasive context-aware computing survey of context-aware ubiquitious middleware systems. Int J Eng Res Technol 10. ISSN 0974-3154
2. Patil S, Talele K (2015) Suspicious movement detection and tracking based on color histogram. In: 2015 International conference on communication, information & computing technology (ICCICT), Mumbai, 2015, pp 1–6. https://doi.org/10.1109/iccict.2015.7045698
3. Mehran R, Oyama A, Shah M (2009) Abnormal crowd behavior detection using social force model. In: 2009 IEEE conference on computer vision and pattern recognition, Miami, FL, 2009, pp 935–942. https://doi.org/10.1109/cvpr.2009.5206641
4. Boiman O, Irani M (2006) Detecting irregularities in images and in video. In: Tenth IEEE international conference on computer vision (ICCV'05), vol 1, Beijing, 2005, pp 462–469. https://doi.org/10.1109/iccv.2005.70
5. Katyal S, Kumar S, Sakhuja R, Gupta S (2018) Object detection in foggy conditions by fusion of saliency map and YOLO. In: 2018 12th International conference on sensing technology (ICST), Limerick, 2018, pp 154–159. https://doi.org/10.1109/icsenst.2018.8603632
6. Ćorović A, Ilić V, Durić S, Marijan M, Pavković B (2018) The real-time detection of traffic participants using YOLO algorithm. In: 2018 26th telecommunications forum (TELFOR), Belgrade, 2018, pp 1–4. https://doi.org/10.1109/telfor.2018.8611986
7. Al-Nawashi M, Al-Hazaimeh OM, Saraee M (2017) A novel framework for intelligent surveillance system based on abnormal human activity detection in academic environments. https://doi.org/10.1007/s00521-016-2363-z
8. Sodemann AA, Ross MP, Borghetti BJ (2012) A review of anomaly detection in automated surveillance. IEEE Trans Syst Man Cybern Part C (Appl Rev) 42(6):1257–1272. https://doi.org/10.1109/tsmcc.2012.2215319

9. YOLO Object Detection Introduction. https://towardsdatascience.com/an-introduction-to-implementing-the-yolo-algorithm-for-multi-object-detection-in-images-99cf240539
10. You Only Look Once (YOLO). https://towardsdatascience.com/yolo-you-only-look-once-real-time-object-detection-explained-492dc9230006
11. Dobhal T, Shitole V, Thomas G, Navada G (2015) Human activity recognition using binary motion image and deep learning
12. Guo T, Dong J, Li H, Gao Y (2017) Simple convolutional neural network on image classification. In: 2017 IEEE 2nd international conference on big data analysis (ICBDA), Beijing, 2017, pp 721–724. https://doi.org/10.1016/j.procs.2015.08.050
13. Nwekea HF, Teha YW, Al-garadia MA, Alob UR (2018) Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: state of the art and research challenges, 27 Mar 2018. https://doi.org/10.1016/j.eswa.2018.03.056
14. Enciso L, Vargas A (2018) Interface with Ubidots for a fire alarm system using WiFi. In: 2018 13th Iberian conference on information systems and technologies (CISTI), Caceres, 2018, pp 1–6. https://doi.org/10.23919/cisti.2018.8399327
15. Ubidots: an IOT Platform. https://ubidots.com/platform/
16. Wang X, Mohd Ali A, Angelov P (2017) Gender and age classification of human faces for automatic detection of anomalous human behaviour. In: 2017 3rd IEEE international conference on cybernetics (CYBCONF), Exeter, 2017, pp 1–6. https://doi.org/10.1109/cybconf.2017.7985780

Analyzing Driving Skills in Adults: A Cognitive Measure Using Eye Tracker

Syed Musthak Ahmed, Shankaranand Jha and Ankit Maurya

Abstract Driving is complex task that requires amalgamation of various cognitive processes along with coordinated sensory and motor skills. For proper coordination, both gross motor skills and precise motor skills are necessary. There are multiple factors that lead to accidents while driving. These factors could be cognitive, sensory or physical. Accidents happen because of failure in either or all of three domains of the driver itself rather than the colliding person. In present paper, we have discuss in detail specifically about the various cognitive factors responsible for impaired driving. Brain regions associated with driving task were also described in neuroimaging studies. There are various qualitative and quantitative techniques that can measure driving abilities of a person, but recent researches have emphasized more on measuring cognitive abilities of the drivers using eye tracking technique. It is a robust technique that provides insights into the cognitive capabilities of the drivers. This paper also discussed about various metrics of eye tracker and eye tracking indicators while driving. This paper support many researchers to identify the mindset of drivers, attitude while driving and correct the measures to be followed to make driving effective. This paper is a platform for researchers who are working on cognitive domain.

Keywords Cognitive process · Sensory · Impaired driving · Neuroimaging · Eye tracker

S. M. Ahmed (✉)
S R Engineering College, Warangal, Telangana, India
e-mail: syedmusthak_gce@rediffmail.com

S. Jha
Department of ECE, S R Engineering College, Warangal, Telangana, India

A. Maurya
JRF, Centre for Creative Cognition, S R Engineering College, Warangal, Telangana, India

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_30

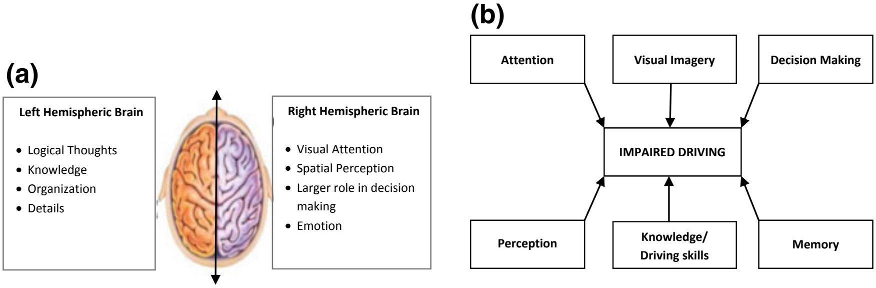


Fig. 1 **a** Hemispheric lateralization of human brain, **b** Outline of the cognitive factors responsible for impaired driving

1 Introduction

Mobility is important for any individual for their independent living. In present world, for economic development of any country, transport sector is one of the important factors. One needs to interact with different societies to have better social and personal relationship. Driving is one of the important ways to achieve this mobility between and within societies. Nowadays, vehicles have become necessary for people rather than luxury and it has been largely acknowledged in scientific literatures. Ergonomists of cars are working day and night hard to make people comfortable with driving, but human factors have some limitations. Despite the best designing of automobiles, accidents do happen. Drivers are not well trained and have impaired driving skills, thus severely impacting the safety of driver himself as well as others on roads. Driving is a complex task and its complexity increases when cardinal task i.e. driving hinders with other auxiliary tasks like talking on cell phone, playing music, looking for objects, operating radio, and so on [1]. To ensure proper safety while driving, one need to have better knowledge of driving rules, selectively attending to the relevant stimulus and simultaneously ignoring the irrelevant stimuli, ensuring the proper functioning of hardware part of vehicle at the very beginning of driving, etc. For proper navigation, co-ordination of motor action and cognitive thinking is very essential. Physical strength and decision making while driving precludes one from getting involved into lethal accidents [2, 3]. Figure 1a shows the hemispheric lateralization of human brain dealing with consciousness, emotion, cognition and behaviour while Fig. 1b shows outline of the cognitive factors required while driving.

2 Cognitive Factors Governing the Driving Behaviour

There are several factors responsible for driving. Some of the factors include age, physical health, experience, driving skills, mental abilities and many more. All these factors could be broadly divided into three main domains: Cognitive domain, sensory domain and motor domain. In the present paper, the study is confined to cognitive domain and the ways to improve the cognitive abilities in a person to perform safe and secure driving.

Cognitive domain refers to the cognitive capabilities required for driving. Here, the term cognitive capabilities means various cognitive processes like attention, perception, memory, decision making, reasoning, problem-solving, language etc. Integration of these various cognitive processes are very essential for proper driving. Disruption in any of these components while driving can lead to fatal accidents. The major cognitive factors responsible for impaired driving are cognitive workload, stress, risk perception and visual attention.

(A) **Cognitive workload**: High cognitive workload while driving is one of the leading factors for crashes. A study conducted in China, assessed traffic safety by analysing driver's behaviour and performance under cognitive workload in areas of complex environment. Cognitive workload was assessed by different quantities of information presented on traffic sign. Results showed, greater amount of information presented on traffic signs lead to high cognitive workload. Reaction time increased in relation to greater amount of information on traffic signs. There were no significant effect of gender and driving experience. Another experiment was performed on cognitive load in driving simulator in the same study. Cognitive workload was assessed by reading information on traffic signs varied at four different levels. Results showed different speed and lane deviations under four different conditions of cognitive workload. Effects of gender and driving experience were significant [4].

(B) **Stress**: Stress is very common cognitive state and people generally feel stressed in high pressure situations or while performing cognitively demanding task like driving. Moderate amount of stress is very common in routine driving. There could be multiple factors that lead to increased level of stress. Some of the common factors include unhealthy interpersonal relations, workplace and job related problems, stressful life events, socio-economic status etc. Increased levels of stress are very harmful for persons while driving. High levels of stress decrease the thinking capability of a person and his/her attention gets distracted very frequently that can ultimately lead to fatal accidents. Personality characteristics like personal maladjustment, anxiety, depression and aggression is likely to increase the rate of accidents [5, 6]. An experiment was performed to study the effects of stress on driving performance using simulator. Driving Behaviour Inventory (DBI) was developed by Gulian et al., that measures vulnerability to driver stress. Aggression, dislike of driving and alertness were the three dimensions identified that is related to driving performance. Aggression factor predicted frequent and error-prone overtaking; dislike of driving predicted better requirement of attention and caution, reduced control skills and mood disturbance; alertness factor predicted speeded reactions to pedestrian hazards [7].

(C) **Risk Perception**: Perception of risk is another important cognitive factor that can prevent fatal and non-fatal accidents. Risk perception is better in experienced drivers than novices because of better anticipation abilities while driving. Earlier researches showed a large number of young drivers are involved in accidents because of their inexperienced driving skills. This inexperienced driving skills is manifested in many ways like inability to handle vehicles on roadways, acquiring insufficient information from environment, not able to attend the incoming sensory information etc. Inability to focus attention while driving is one of the major factors that result into fatal accidents [8, 9]. An eye tracking study was performed in driving simulator between three groups of 'young novice drivers', 'young experienced drivers' and 'older experienced drivers'. Results showed engagement of novice drivers in behaviours that are indicative of risk perception 35.1% of time. The other two groups showed 50.3% and 66.2% respectively. The identified factors included inattention, ineffective visual search and lack of ability to scan relevant areas that may be a potential threat, less scanning of side by side areas [10, 11]. Another driving simulator study was performed to explore the effect of risk perception training on novice drivers. Risk perception training involved personal computer (PC) based training to identify high risk locations. Risk Awareness and Perception Training (RAPT) program is computer based presentation of plan views of risky situations along with information related to each of these risky situations. It also indicates where we should focus our attention in order to detect risky situations. This work was carried out at 'Human Performance Laboratory' at University of Massachusetts. The training process involved three broad categories of risk: obstruction, sign ahead and visible pedestrian and vehicle. The training program basically has five sections: (a) instructions (b) pre-test (c) training (d) questions and (e) post-test. Earlier studies already showed effectiveness of RAPT program on scanning behaviour immediately within 1 h. in driving simulator. This study explored its effectiveness after an avg. of four days after training. Two groups were formed: trained drivers (experimental) and matched controls. Eye tracking results showed more fixations on risky regions by experimental group thereby indicating anticipation of hazardous situations. Experimental group scanned risky regions almost twice than matched controls. Results also indicated the effects of the training program were long-lasting [12].

(D) **Visual Attention**: Visual attention refers to attention towards visual stimuli surrounding our environment. This is very important for navigation and for extracting out relevant visual information necessary for driving. Visual attentional resources are shared and consequently reduced for each task when drivers perform both primary task (driving) and secondary task (attending calls or using cell phone, using entertainment systems etc.) simultaneously. Sharing of these available attentional resources could lead to fatal results and one may end up with serious accidents. Earlier studies showed reduction in lateral control, specifically for visual tasks [13–16], higher variations in speed [15, 17, 18], higher steering effort [14, 19] and delay in reaction times to sudden unexpected events [13, 20]. Secondary tasks that require visual attention often interferes with driving. Eye tracking studies showed rapid switching of gazes between primary and secondary visual tasks [21–23]. Another eye tracking study performed in driving simulator compared the two visual secondary tasks. One of the tasks was highly demanding and externally paced and another was self-paced menu system task. Participants performed either of them. Before beginning of secondary tasks, larger fixation duration of gaze and smaller standard deviation were observed for both tasks. Comparison showed that drivers used the possibilities of self-paced task and they fixated longer for driving scene and SD was greater for gaze position [24].

3 Neuroimaging Studies on Driving Behaviour

Very few neuroimaging studies were performed that involved driving simulator to study the brains regions associated with driving [25–29]. Certain network of brain regions like- 'parieto-occipital cortices, cerebellum and cortical regions associated with perception and motor control' are more active during driving in comparison to the resting period. Driving is a sensory-cognitive highly demanding task in terms of visual attention, motor coordination and visuo-motor integration and these sensory-cognitive processes are responsible for activation of above listed brain regions.

Neuroimaging studies involved blocks of 30–60 s averaged over time, but it is very difficult to timestamp each and every cognitive process during driving. One can't associate which brain regions' are associated with which cognitive processes during driving. Therefore to pinpoint certain brain regions and to measure the dynamicity of brain activity while driving, Calhoun et al. performed ICA (Independent Component Analysis) [25]. By employing ICA, hidden temporal signals can be specifically determined and they were able to find that different brain regions showed different dynamic patterns during driving. Average driving speed was found to be correlated with activity in frontal, parietal, occipital and thalamic regions [25, 27]. Activity in the posterior cingulate region was found to be negatively correlated with number of car crashes [27] and activity in anterior cingulate region was negatively correlated with maintaining safe distance while driving [28].

Spiers and Maguire [30] performed fMRI (functional magnetic resonance imaging) study on driving and they identified events related to driving and the underlying brain regions associated with it. Figure 2 shows networks of premotor, parietal and cerebellar regions were found to be associated with preparatory actions such as starting, turning, reversing and stopping [30]. Other brain regions also showed some activity with these preparatory actions. Activation in insula, greater posterior region in medial premotor cortex (m-pMC) and lateral occipital and parietal cortex were associated with avoiding collision and swerving (unexpected events). Also, found activations in superior parietal, lateral occipital cortex and cerebellum for future action planning. Visual processing of traffic rules was found to be specifically associated with activation in right lateral prefrontal cortex.

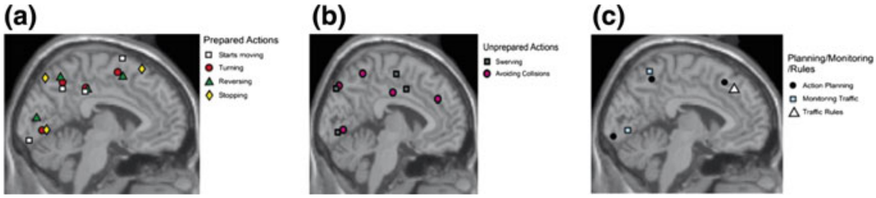


Fig. 2 Activated brain regions associated with prepared, unprepared and planning actions. **a** Brain regions associated with preparatory actions, **b** brain regions associated with sudden unexpected actions and **c** brain regions associated with future action planning

4 Behavioural Study of Drivers Using Eye Tracker

Researches on eye tracking dates back to early eighteenth century but modern studies on eye tracking technique began in late 1970s [31]. Eye tracking studies in relation to driving behaviour also began in 1970s. A survey of reports on drivers was published by Soliday in 1971 [32]. Japanese developed the NAC Marker recorder which was the first device used in driving studies. Visual strategies used by the experienced and novice drivers were studied at that time. Different kinds of visual patterns were found with the TV eye movement systems in experienced and novice drivers [11]. Rayner published lot of papers on eye tracking studies in relation to cognitive task of reading and focussed primarily on analysing eye movements in studies of information processing and process of perception [33].

Various types of eye movements as shown in Fig. 3 provide insight into the cognitive abilities of drivers. Some of the measures defined are saccades, smooth pursuit, and vergence eye movement Saccades refer to rapid movement of eyes used in repositioning of fovea to the new location in the visual environment. The duration of a saccade in general is 20–200 ms, depending upon saccadic amplitude. Saccadic amplitude doesn't go beyond 8°. Speed of a saccade is 500°/s but may go even up to 800°/s. Saccadic latency (the time between appearance of target and the beginning of eye movement) varies from 100 to 300 ms. Smooth pursuit refers to the movement

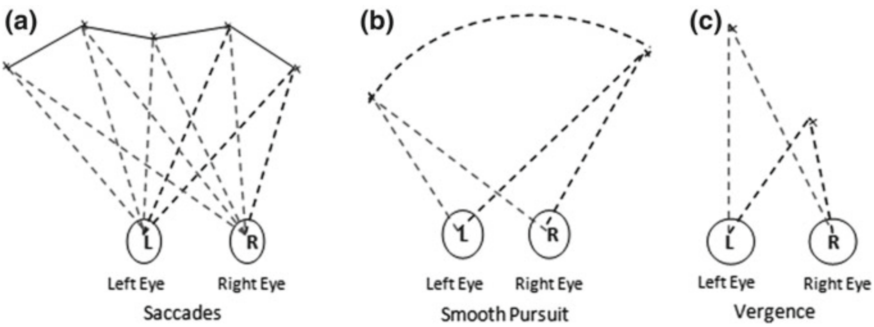


Fig. 3 Types of eye movements. **a** Saccades; **b** smooth pursuit; **c** vergence eye movement

of eyes along with or in synchrony with moving objects. Speed of smooth pursuit is below 30°/s. If the velocity of the moving object is very fast, then smooth pursuit is often accompanied by saccade to match up with the velocity of the moving object. According to many researchers, it is very important to analyse smooth pursuits in driving studies [34]. Vergence eye movement occurs when we need to focus on near objects and the eye movements are directed towards the centre of the visual field. Eye movements become parallel once the target object is viewed at a span greater than 6 m and converge when the target object is located at <1 m, so to bring the image of the object onto the foveal region.

Other types of eye movements include nystagmus, drifts and micro-saccades. They have higher frequency and lower amplitude. Of all these types of eye movements, saccades are most important one to be studied. The pauses between the ending of a saccade and beginning of another saccade are referred to as fixations and its duration varies from 200 to 600 ms depending upon factors like cognitive load, scanning area complexity and intensity of luminance contrast. Saccades could be of shorter duration of less than 200 ms during driving [35].

5 Eye Tracking Methods

The method that allows us to measure eye movements and gaze directions is called oculometry. Basically there are three types of oculometry: (a) electrooculography (EOG) (b) galvanometric (c) corneal reflection technique. EOG measures bioelectric potential difference caused by corneal-retinal eye movements. Electrodes were placed around the eyes to record these electric potentials and artefacts were eliminated using filtering technique. It has poor spatial resolution and better temporal resolution. It allows us to measure eye movement only when the head is kept fixed and not gaze direction. Another limitation of that it is not applicable for long term recording. Second techniques called galvanometer measures changes in electromagnetic field. Contact lenses equipped with induction coil were applied to the cornea. Its invasive nature precluded its usage in field studies. Lastly, corneal reflection technique involves infrared light. Infrared light emitted from a source is reflected from cornea and is tracked in a field of view. It is non-invasive technique used for measuring gaze direction. It is currently the only technique that is used in eye tracking studies and is conducive for long term recording. This technique also fails in the condition of solar radiations of very high intensity because of inability to track infrared rays emitted from the source. The criteria for section of the above mentioned methods depends on accuracy, resolution, reliability, choice of filter, programming and associated interfacing circuitry [36].

6 Visual Strategies Employed by Drivers During Driving

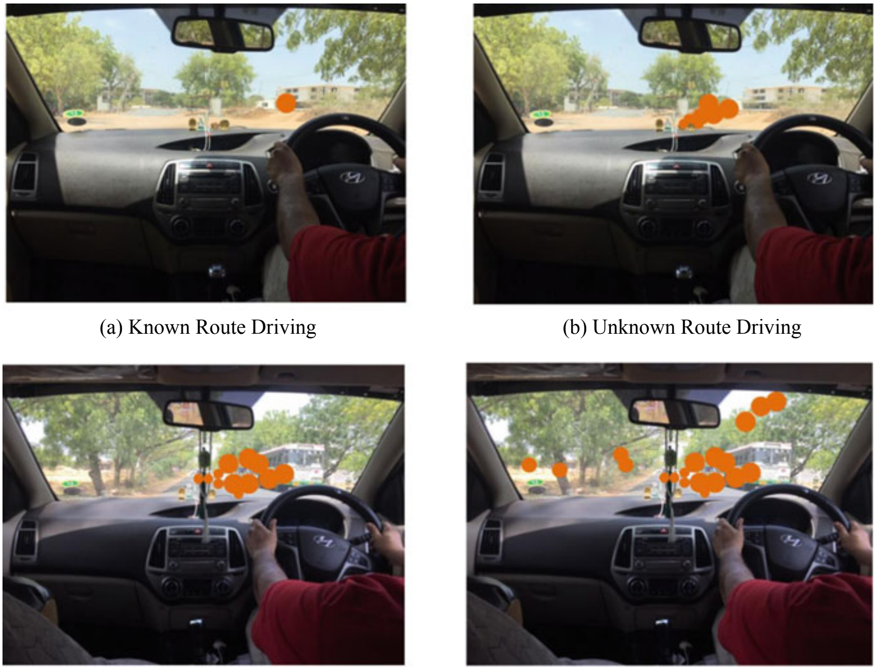
When drivers look ahead while driving, their visual field is generally classified into 3 main regions: ‘foveal, parafoveal and peripheral regions’ corresponds to 2°, 5° and greater than 5° respectively from central axis of vision. Fovea is the region of excellent visual acuity and acuity decreases moving from foveal region to peripheral region. Saccades are very essential as it covers whole field of view and fixations are very important to extract out the desired information. Usually, large objects are recognized in the parafoveal region without performing any specific eye movement indicating presence of control system. Human’s functional visual field of view can be categorised into 3 parts [37] viz., Stimulus identification without eye movement. Eye movement is essential for stimulus identification and stimulus identification requires head movement.

7 Eye Tracking Indicators While Driving

Common eye movement parameters that researchers assessed while driving are—(i) number of fixations, (ii) average fixation time and (iii) areas of functional visual field. Few studies on comparative studies of visual strategies employed by experts and novices, older and younger people and by motorcyclists and car drivers had been performed earlier and the above listed parameters were analysed. Results showed visual strategies are richer in motorcyclists than cars drivers indicated by larger no. of fixations on the objects on the road. Driving performance deteriorates seriously while performing others tasks while driving. It is indicated by avoiding viewing in rear mirror, lacking control, increased no. of brakes [38]. Another study performed in driving simulator showed older drivers assessed traffic parameters very slowly when compared to young drivers indicated by larger fixation duration. Their precision while driving was also less indicated by their greater field of observation of object [39]. Eye movements and head movements should not be measured simultaneously rather individual analysis would lead to accurate assessment of strategies [40]. In case of normal driving, drivers usually focus on the stationary objects in straightforward direction. They also make few eye movements onto the sign boards present on the corner side of the roads [41]. The cited authors told that duration fixation increases after occurrence of dangerous events. Larger eye movements are made while driving heavier vehicles like bus, truck i.e. there is an increase in amplitude [42] (Fig. 4).

8 Conclusion

Driving is both physically and mentally taxing task and requires amalgamation of all the three domains viz., cognitive, sensory and motor domain. Nowadays, vehicles



(a) Known Route Driving

(b) Unknown Route Driving

(c) Conscious Driving

(d) Unconscious Driving

Fig. 4 Eye movement while driving indicating, **a** known route, **b** unknown route, **c** conscious driving, **d** subconscious driving

are not symbolizing luxury only but it has become essential part of our daily routines. To ensure proper independent living and to maintain our social relations, mobility is very important. Mobility can be achieved through proper driving skills, but often we hear cases of impaired driving skills and repercussion to this is lethal and fatal accidents. There are various cognitive factors responsible for this impaired driving viz., cognitive workload, stress, risk perception, visual attention etc. Eye movement plays a very important role while performing strenuous task like driving and these eye movements can be tracked through the usage of a device called eye tracker. Eye tracker provides insight into cognitive abilities of humans. Recently, there is a surge in study of eye tracking in relation to driving abilities. With the usage of eye tracker, one can identify the both specific and numerous cognitive factors responsible for impaired driving. Neuroimaging studies showed activations in network of brain regions like 'parieto-occipital cortices, cerebellum and cortical regions' for perceptual-motor activities while driving. Activations in 'superior parietal, lateral occipital cortex and cerebellum' is associated with future action planning. Frontal, parietal, occipital and thalamic regions are associated with average driving speed. Most studies focussed on corneal reflection technique to study eye movements as it is the most feasible one among other techniques. Drivers use various visual strategies from foveal, parafoveal and peripheral regions of human visual apparatus to extract out the desired and relevant information required to perform driving.

9 Future Scope of Study

In this paper, we have discussed thoroughly about the various cognitive factors responsible for impaired driving. But there are other major factors responsible for the same. The other factors include sensory and physiological factors. Integration of all these three factors is very important from driving perspective. Thus, future work can be carried out by considering

1. Physical strength and physical activity to make better driving.
2. Integration of cognitive, sensory and physical for better motor action for safe driving.

Acknowledgements The authors acknowledge Department of Science and Technology, Government of India for financial support vide Reference No.: DST/CSRI/2017/457 under Cognitive Science Research Initiative (CSRI) to carry out this work. The authors acknowledge Centre for Creative Cognition, S R Engineering College, Warangal for the platform provided to take up the research work. The authors also acknowledge the Management and Principal of S R Engineering College, Warangal Urban for their continuous support by providing all the necessary facilities.

References

1. Nabatilan LB (2007) Factors that influence visual attention and their effects on safety in driving: an eye movement tracking approach. LSU Doctoral Dissertations, 3014
2. Ahmed SM, Yasmeen A, JagadeeshBabu B (2018) Wheelchair with auto navigation for Adults with Physio and cognitive impairments. Int J Innov Technol explor Eng (IJITEE) 8(2s2):458–462
3. Ahmed SM, Shireen A, JagadeeshBabu B, Shruti, Powered wheelchair for mobility with features to address physical strength, cognitive response, and motor action development issues. In: International conference on data science, machine learning and applications (ICDSMLA), CMRIT Hyderabad, 29–30
4. Lyu N, Xie L, Wu C, Fu Q, Deng C (2017) Driver's cognitive workload and driving performance under traffic sign information exposure in complex environments: a case study of the highways in China. Int J Environ Res Public Health
5. Gulian E, Matthews G, Glendon AI, Davies DR, Debney LM (1989) Dimensions of driver sress. Ergonomics 32:585–602
6. Selzer ML, Vinokur A (1975) Role of life events in accident causation. Ment Health Soc 2:36–54
7. Matthews G, Dorn L, Hoyes TW, Glendon AI, Davies DR, Taylor RG (1998) Driver stress and performance on a driving simulator. Human Factors Ergon Soc 136
8. McKnight JA, McKnight SA (2003) Young novice drivers: careless or clueless? Accid Anal Prev 35:921–925

9. Treat JR et al (1979) Tri-level study of the causes of traffic accidents. Final report. Causal factor tabulations and assessments, vol I. Indiana University Institute for Research in Public Safety, Bloomington
10. Crundall D, Underwood G (1998) Effects of experience and processing demands on visual information acquisition in drivers. Ergonomics 41:448–458
11. Mourant RR, Rockwell TR (1972) Strategies of visual search by novice drivers and experienced drivers. Hum Factors 14:325–335
12. Pradhan AK, Fisher DL, Pollatsek A (2006) Risk perception training for novice drivers evaluating duration of effects of training on a driving simulator. J Transp Res Board 1969:58–64
13. Törnros JEB, Bolling AK (2005) Mobile phone use-effects of handheld and handsfree phones on driving performance. Accid Anal Prev 37:902–909
14. Engström J, Johansson E, Östlund J (2005) Effects of visual and cognitive load in real and simulated motorway driving. Transp Res Part F 8:97–120
15. Horrey WJ, Wickens CD (2002) Driving and side task performance: the effects of display clutter, separation, and modality. Technical report AHFD-02-13/ GM-02-2. Savoy: University of Illinois, Aviation Human Factors Division
16. Jamson AH, Merat N (2005) Surrogate in-vehicle information systems and driver behavior: effects of visual and cognitive load in simulated rural driving. Transp Res Part F 8:79–96
17. Burns PC, Parkes A, Burton S, Smith RK, Burch D (2002) How dangerous is driving with a mobile phone? Benchmarking the impairment to alcohol. TRL report TRL547. Berkshire, United Kingdom: TRL Limited
18. Zheng XS, Tai Y-C, McConkie GW (2003) Effects of cognitive tasks on driver's eye behavior and performance. Paper presented at the 3rd international symposium on human factors in driver assessment, training & vehicle design
19. Boer ER (2000) Behavioral entropy as an index of workload. Paper presented at the IEA 2000/HFES 2000 congress, San Diego
20. Ishida T, Matsuura T (2001) The effect of cellular phone use on driving performance. International association of traffic safety sciences (IATSS) research, 6–14
21. Sodhi M, Reimer B, Llamazares I (2002) Glance analysis of driver eye movements to evaluate distraction. Behav Res Methods Instrum Comput 34(4):529–538
22. Victor T (2005) Keeping eye and mind on the road. Unpublished dissertation, Universitet Uppsala, Uppsala
23. Wierwille WW (1993) Visual and manual demands of in-car controls and displays. In: Peacock B, Karwowski W (eds) Automotive ergonomics. Taylor & Francis, London, pp 299–320
24. Metz B, Schömig N, Kruger H (2011) Attention during visual secondary tasks in driving: adaptation to the demands of the driving task. Transp Res 14F(5)
25. Calhoun VD, Pekar JJ, McGinty VB, Adali T, Watson TD, Pearlson GD (2002) Different activation dynamics in multiple neural systems during simulated driving. Hum Brain Mapp 16:158–167
26. Graydon FX, Young RA, Tdss MD, Tdss B, Genik II RJ, Posse S, Hsieh L, Green CC (2004) Visual event detection during simulated driving: identifying the neural correlates with functional neuroimaging. Transp Res Part F Traffic Psychol Behav 7:271
27. Horikawa E, Okamura N, Tashiro M, Sakurada Y, Maruyama M, Arai H, Yamaguchi K, Sasaki H, Yanai K, Itoh M (2005) The neural correlates of driving performance identified using positron emission tomography. Brain Cognition 58:166–171
28. Uchiyama Y, Ebe K, Kozato A, Okada T, Sadato N (2003) The neural substrates of driving at a safe distance: a functional MRI study. Neurosci Lett 352:199–202
29. Walter H, Vetter SC, Grothe J, Wunderlich AP, Hahn S, Spitzer M (2001) The neural correlates of driving. NeuroReport 12:1763–1767
30. Spiers HJ, Maguire EA (2006) Spontaneous mentalizing during an interactive real world task: an fMRI study. Neuropsychologia 44:1674–1682
31. Young LR, Sheena D (1975) Survey of eye movement recording methods. Behav Res Methods Instrum 7(5):397–429

32. Soliday M (1971) Driver's eye movements: a literature review. Appendix D. In: Johns TR, Allen JA (eds) Driver license and testing. Highway Safety Research Center, University of North Carolina, North Carolina, pp 15–23
33. Rayner K (1988) Eye movements in reading and information processing: 20 years of research
34. Shin D, Sakai H, Uchiyama Y (2011) Slow eye movement detection can prevent sleep-related accidents effectively in a simulated driving task. J Sleep Res 20(3):416–424. https://doi.org/10.1111/j.1365-2869.2010.00891.x
35. Salvucci DD, Goldberg JH (2000) Identifying fixations and saccades in eye-tracking protocols. In: Proceedings of the eye tracking research and applications symposium. ACM Press, New York, pp 71–8
36. Drewes H (2010) Eye gaze tracking for human computer interaction (dissertation). München: LFE Medien-Informatik der Ludwig-Maximilians-Universität
37. Sanders AF (1993) Processing information in the functional visual field. In: d'Ydewalle G, van Rensbergen J (eds) Perception and cognition: advances in eye movement research. North Holland, Amsterdam, pp 3–22
38. Muttart JW, Peck LR, Guderian S, Bartlett W, Ton LP, Kauderer C et al (2011) Glancing and stopping behavior of motorcyclists and car drivers at intersections. Transp Res Rec 2265:81–88. https://doi.org/10.3141/2265-09
39. Fildes B, Charlton J, Muir C, Koppel S (2007) Driving responses of older and younger drivers in a driving simulator. Annu Proc Assoc Adv Automot Med 51:559–572
40. Doshi A, Trivedi MM (2012) Head and eye gaze dynamics during visual attention shifts in complex environments. J Vis 12(2):9
41. Chapman PR, Underwood G (1989) Visual search of dynamic scenes: event types and the role of experience in viewing driving situations. In: Underwood G (ed) Eye guidance in reading and scene perception. Elsevier, Oxford, pp 369–394
42. Kito T, Haraguchi M, Funatso H, Sato T, Kondo M (1989) Measurements of gaze movements while driving. Percept Mot Skills 68(1):19–25. https://doi.org/10.2466/pms.1989.68.1.19

Sensory-Motor Deterioration in Older Drivers and Their Amelioration Through Various Training Strategies: A Study

Syed Musthak Ahmed, M. Sheshikala, Ankit Maurya and Vinit Kumar Gunjan

Abstract Physical strength is one of the major factors for safe driving habits. But as people get older, there is continuous degeneration in their sensory, cognitive and motor abilities. Deficits in any of these abilities can result into fatal accidents. In this paper, we will emphasize upon sensory-motor abilities essential for driving, deterioration in these abilities and their amelioration through proper training strategies. Deficits in sensory system, especially in visual and auditory domain while driving can have lethal repercussions. For proper driving skills, coordination of both gross and fine motor skills is needed. Notably, in India, geriatric population is on the surge and is expected to be around 173 million by 2026. Therefore, careful researches and observations are needed for this segment of population, especially in relation to safe driving.

Keywords Sensory-motor abilities · Geriatric population · Training strategies · Visual · Auditory · Gross and fine motor skills

1 Introduction

Driving is very important for daily activities, societal participation and interaction as well as to maintain independent living [1]. In our day to day life, integration of sensory-motor skills is very essential. But as people get older temporally and naturally, there is observed progressive declination in their sensory-motor abilities.

S. M. Ahmed (✉)
Department of ECE, S R Engineering College, Warangal, Telangana, India
e-mail: syedmusthak_gce@rediffmail.com

M. Sheshikala
Department of CSE, S R Engineering College, Warangal, Telangana, India

A. Maurya
JRF, Centre for Creative Cognition, S R Engineering College, Warangal, Telangana, India

V. K. Gunjan
Department of Computer Science and Engineering, CMR Institute of Technology, Hyderabad, India

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_31

Sensory deterioration refers to degradation in abilities of sense organs for visual, auditory, haptic, olfactory and gustatory perception. Motor deterioration refers to declination in motor control and strength i.e. abilities and skills required for proper coordination of movements. Proper coordination of these sensory-motor abilities is very essential while driving. One can't control the process of natural aging and the issues associated while driving, but with the help of appropriate training strategies, one can improve their sensory-motor skills and physical strength. Visual and auditory impairment is very familiar issue associated with driving. Motor skills are very vital in daily routines [2–4]. Motor impairments occur in form of decrease in reflexive actions i.e. slowness in making sudden movements, critical decline in fine motor hand [5], and leg movements etc. However, a vivid explanation in terms of biological basis of this slowness in reflexive action and fine motor hand and leg movements is still pending and not well understood.

2 Factors Leading to Poor Driving Performance

Aging is the major factor for poor driving performance due to deterioration of organ functioning leading to reduced physical strength, decision making and navigation. These deficits can be broadly classified into sensory deficits and motor deficits.

2.1 Sensory Deficits

Sense organs play a vital role in gathering sensory information while driving. Especially, visual and auditory information are the most important ones. Between visual and hearing apparatus, most of the incoming information while driving, we receive from former one with minor contribution from later one. Older populations are potential targets and have increased risk of impairments in these sensory domains. Generally, these impairments are often overlooked in our daily life, but it would be a huge mistake to ignore these impairments in relation to driving. It can prove to be fatal for our life as well as for other drivers on roads. As explained above, impairment in auditory and visual apparatus are the most important sensory deficits among others. Many people have Dual sensory impairment (DSI) i.e. impairment in both auditory and visual apparatus. In the age group of 80 and above, the percentage of visual, auditory and DSI impairment was found to be 31%, 78% and 25% respectively [6]. All these impairments are associated with functional declination [7], loss in societal interaction [8], cognitive decline [9], degradation of quality of life [10, 11], depressive thoughts and emotions [12–15]. Other sensory deficits include impairment in olfactory, haptic and gustatory domain. One has to surpass various sensory tests to obtain driving license for driving. If people are having any kind of sensory deficit in visual, auditory or dual sensory domain, they will be disqualified for driving.

2.1.1 Hearing Impairment and Driving Performance

Hearing impairment in relation to the driving researches is defined as loss of hearing that should be significant and affect the private or commercial vehicle's driver. The range is from 40 db loss of average voice frequency in one ear to total deafness in both ears. Here deafness is referred to complete hearing loss or up to the level which we can't hear even with the use of hearing aids. Deaf drivers are safer and better drivers than hearing drivers in the sense that, former one are not able to listen any sound, therefore they are less distracted and have better concentration abilities while the later one have to face more distractions [16] and there is positive opine about safe driving abilities of deaf drivers [16–18].

Coppin and Peck performed couple in studies in 1963 and 1964 respectively. In the first study in 1963 [19] they studied the no. of accidents and how many times the deaf and non-deaf were convicted. Results showed that the deaf drivers involved in 1.78 times more in no. of accidents and 1.26 times more in being convicted than non-deaf drivers, however both the groups of drivers differed on many other variables like their job, mileage driven and sex. Second study in 1964 [20] involved large sample size than earlier they were matched on many additional variables like age, annual mileage and residential area. Results compared deaf and non-deaf females on accidents and violations and it was found that there was no significant difference between both groups. When compared between deaf and non-deaf males, it was found that rate of accident was 1.8 times higher in deaf males than non-deaf males.

Another survey was performed on 220 drivers by Roydhouse in 1967 [21] in New Zealand. It was a self- report study. Survey report showed drivers <50 years of age had fewer accident rates than drivers >50 years old.

A review study performed in 1978 [22] discussed the effects of impairment in auditory apparatus on driving performance. According to this study, the literature is very limited in this domain in order to determine the specific auditory requirements for safe driving. More researches need to be done to establish this assertion. The study also emphasized the importance of coming up with the standards to define hearing loss as well as the kinds of restrictions that need to be imposed by State Department of Motor Vehicles. There is no direct relation between auditory requirements and safe driving performance for bus and truck driving [23].

2.1.2 Visual Impairment and Driving Performance

Eyes are the most dominant sense organs among all and most of the information processing while driving is performed by our eyes, with little contribution from ears. Vision is function of our eyes. Driving performance depend lot on visual information processing, but when if there is impairment in visual apparatus, driver may face unprecedented consequences. Wikipedia defines driving impairment as reduced ability to see to extent that causes problems not fixable by usual means, such as glasses. There are several kinds of impairment in visual domain which are generally overlooked. Some of the common visual impairment is complete blindness, myopia, hyper-metropia, presbyopia, cataract, glaucoma, macular degeneration etc. A study [24] showed deteriorated driving performance in normal older drivers than the younger drivers who were put in simulator which made the younger drivers felt like they are visually impaired. Researchers indicated the reason for the same is due to slow Reaction time and slow information processing.

Another study [25] suggest the different types of accidents that elderly people often go through are accidents at turns, ignoring stop sign, traffic issues and accidents due to inattention. Other study results were in accordance with the previous one [26].

Some aspects of vision are very important and is considered to be very useful while driving are as follows:

- Contrast Sensitivity: To determine contrast threshold, one of the widely used test is Pelli-Robson chart. The chart consists of letters of identical size, but there is gradual decrease in the contrast of letters and one has to recognize the letter with lowest contrast for test scores [27]. Wood and Troutbeck in 1995 [24] showed high correlation between Pelli-Robson scores and driving task performance and another study [28] showed positive correlation between Pelli-Robson scores and self-reported problems associated with both day and night-time driving.
- Stereopsis: Stereopsis refers to depth perception. There is no consensus among researchers on viewpoints of whether there is a relationship between stereopsis and driving performance. According to a study in 1994 [28] there is no relationship between stereopsis and poor driving performance among elderly drivers. Another study [29] suggests that stereopsis is related to driving performance but it is of very much importance.
- Colour Vision: Norman's [30] study suggested that impaired colour vision is closely related to poor driving performance. Another study [31] showed deficit in colour vision causes not only inability to recognize signals but other vehicles also. Researches on red colour blind people show they require nearly 4 times intensity than normal people to perceive red colour [32].

2.2 *Motor Deficits*

Motor deficits refer to complete loss or slowness in motor movements. It is very natural process as age progresses and is much more prominent in older adults. Older drivers suffer maximally due to decrease in reflexivity. There is deterioration in both motor control and strength as age progresses.

Motor control was classically quantified with Motor Output Variability (MOV) which refers to unintended variation in output of voluntary contractions [33] and functional significance of motor output variability is inaccurate movements [34].

Reactive driving is very essential in car driving [35] which requires responding to sudden unprecedented stimuli in accurate and consistent manner. Another study was performed to examine whether declination in the above stated two variables impairs reactive driving in older adults. Reactive driving task requires control of both gas pedal and brake pedal and was examined previously by other researchers in their laboratory [36–39].

Gas pedal control is visuo-motor tracking task and brake pedal control is a goal directed movement task. Motor output variability in former one is greater in older adults at very low force levels [40–42] and high amount of visual information [38, 43].

Motor output variability in later one was also found to be greater in older adults at all force levels [44]. Existing literatures suggest that both motor control and strength are independent of each other in older peoples [45].

In the same study, another question of interest arises whether declination of strength with aging impairs reactive driving performance, independent of MOV. They hypothesized that strength could influence reactive driving performance, independent of MOV. Results showed, reactive driving performance was impaired by greater MOV, not because of declination in strength [46]. Physio operations have been conducted on aging adults to improve their motor abilities [47, 48].

3 Training Programs and Guidelines

Driving is a complex task involving integration of coordinated sensory-motor system. Old age drivers are most prone to road accidents and therefore need an immediate training programs and suggestions in order to drive safely. Traditional age based road training programs are of no use in safety [49–51] and drivers need classroom based education training programs and visual-attentional training programs. Another study [52] involving training program of 'cognitive stress management program' (for speed of information processing) and 'range of motion exercise program' was performed. The training period of the study was 8 weeks. The former training program wasn't effective but later was very effective in 'trunk rotation and shoulder flexibility'.

WHO provided some guidelines in relation to promote physical activity among older adults of 50 or above years of age. According to the guidelines, formal exercise programs need to be established. Exercise programs should include exercises like walking, dancing, swimming, cycling, stair climbing, chair and bed exercises etc. Exercise should be simple, relaxable, enjoyable and regular [53].

Useful field of view (UFOV) training program showed promising results in studies of older drivers. UFOV was found to be highly correlated with accidents and specific predictor of crashes [54]. Other study suggested possible expansion of UFOV with practice [55]. With UFOV training practice, drivers showed better choice reaction time. Similar to UFOV, another training program was employed known as VM-UFOV (Visual-Motor Useful Field of View). VM-UFOV training program showed improvement in psycho-motor abilities of older drivers. Researches at University of Toronto and University of Missouri showed the efficiency of this training program in improving information processing and attentional allocation in studies of older drivers who suffered from stroke. Those who had undergone these training programs showed better visual, perceptual, cognitive and physical abilities.

Yoga based training programs are very important and useful for maintaining mental and physical health in older adults [56]. Another study [57] showed Ashtanga-Based Yoga Therapy increased postural stability in visually impaired persons who are at risk of falls.

Yoga also contribute in improvement of hearing performance. Various types of yoga useful in checking hearing loss are *Greevachalan* (neck exercise), *Skandhchalan*, *Brahmari pranayama* and *Kumbhak*. Under *Greevachalan*, there are sub-exercises which are neck flexion extension exercise, lateral flexion exercise and head rotation exercise. *Shankhanaad aasana* and *shunya* mudra is also very important yoga postures for ameliorating hearing performance [58].

4 Expected Outcomes of Performing (Yoga) Training

The results of training improves the physical health, cognitive and motor skills as discussed in Table 1.

5 Conclusion

Driving is very strenuous task and its complexity is very prominent in older drivers. Older drivers are prone to getting involved in accidents more than younger adults because of age related sensory-motor deficits. Driving requires integration of sensory-motor and cognitive skills. Sensory deficits mean impairment in visual or auditory apparatus or dual sensory impairment. Some aspects of vision is very important while driving like depth perception, contrast sensitivity, colour vision, visual acuity etc. Similarly, incoming auditory information is also important, but less than visual information. There is still no scientific consensus on driving abilities of deaf and non-deaf drivers. Akin to sensory deficits, motor deficits could also lead to fatal and lethal consequences. Motor deficits refer to slowed body reflexivity, critical decline in fine hand finger movements. Motor deficits are not only restricted to motor control but it also incorporates motor strength. Research study on reactive driving performance showed impairment in reactive driving is due to greater motor output variability, rather than decreased strength in older adults. Various training programs were also established to make older adults drive safely and independently. Traditional training programs were not beneficial. Researchers suggest the need to establish classroom based age related training programs that provides education about driving skills. World Health Organization (WHO) also provided some guidelines to promote physical activity in older adults. Physical activity should be simple exercises like walking, dancing, swimming etc. which should be enjoyable.

Other training programs were also employed like Useful field of view training program (UFOV), Visual-Motor useful field of view training program (VM-UFOV). Yoga aasana and postures like *Greevachalan*, *Skandhchalan*, *Brahmari pranayama*,

Table 1 Various Yoga aasanas, their benefits along with the postures

S. No.	Yoga aasana	Benefits	Posture
1	Ashtanga Yoga	Maintaining postural stability Coordinates mind and body Can do any work effectively and with skills Can enjoy every moment in life Can manage ourselves and others	
2	Greevachalan	Prevents spondylitis (neck) problem and prevent the neuro problems like giddiness	
3	Brahmari pranayama	Vibrate the brain, empties the mind, stops blood-clots and rupture of blood vessels Provides relief in long term tinnitus De-stressing Prevent the paralysis	
4	Kumbhak	Increase the blood circulation and vitalize the body Helps in IQ increment Also clears the sinuses and prevents or delays the age related decay of aging body and body organs	

(continued)

Table 1 (continued)

S. No.	Yoga aasana	Benefits	Posture
5	Shankhanaad (Shanmukhi-mudra)	Supplies enough oxygen and activate the sensory organs like nose, ears, eyes, tongue and skin Helps in preventing the aging hearing loss	
6	Shunya mudra	Increase the power of hearing in normal person and for spiritual awakening	
7	Trataka	Traditionally believed to remove any disease from the eyes Increases eye sight Prevent wearing glasses Decreases mental lethargy and increases *buddhi* (intellect)	
8	Training the third eye	Improves concentration power Can forestall the need for reading glasses	

Kumbhak, *Shankhanaad aasana* and *shunya* mudra showed improvement in sensory, motor as well as mental health of older drivers.

Acknowledgements The authors acknowledge Department of Science and Technology, Government of India for financial support vide Reference No.: DST/CSRI/2017/457 under Cognitive Science Research Initiative (CSRI) to carry out this work. The authors acknowledge Centre for Creative Cognition, S R Engineering College, Warangal and SREC Yoga Club for the platform provided to take up the research work. The authors also acknowledge the Management and Principal of S R Engineering College, Warangal Urban for their continuous support by providing all the necessary facilities.

References

1. Interrante J (1986) The road to autopia: the automobile and the spatial transformation of American culture. In: Lewis DL, Goldstein L (eds) The automobile and american culture. The University of Michigan Press, Ann Arbor, p 89
2. Diermayr G, McIsaac TL, Gordon AM (2011) Finger force coordination underlyingobject manipulation in the elderly—a mini-review. Gerontology 57:217–227. https://doi.org/10.1159/000295921
3. Kornatz KW, Christou EA, Enoka RM (1985) Practice reduces motor unit discharge variability in a hand muscle and improves manual dexterity in old adults. J Appl Physiol 2005(98):2072–2080
4. Darling WG, Cole KJ, Abbs JH (1988) Kinematic variability of grasp movements as a function of practice and movement speed. Exp Brain Res 73:225–235
5. Smith CD, Umberger GH, Manning EL, Slevin JT, Wekstein DR, Schmitt FA, Markesbery WR, Zhang Z, Gerhardt GA, Kryscio RJ, Gash DM (1999) Critical decline in fine motor hand movements in human aging. Neurology 53(7):1458–1458
6. Vaal J, Gussekloo J, de Klerk MM, Frijters DH, Evenhuis HM, van Beek AP, van Nispen RM, Smits C, Deeg DJ (2007) Combined vision and hearing impairment: in an estimated 30,000–35,000 people aged 55 years or over in The Netherlands. Ned Tijdschr Geneeskd 151(26):1459–1463
7. Brennan M, Su YP, Horowitz A (2006) Longitudinal associations between dual sensory impairment and everyday competence among older adults. J Rehabil Res Dev 43(6):777–792
8. Crews JE, Campbell VA (2004) Vision impairment and hearing loss among community-dwelling older Americans: implications for health and functioning. Am J Public Health 94(5):823–829
9. Lin FR, Yaffe K, Xia J, Xue QL, Harris TB, Purchase-Helzner E, Satterfield S, Ayonayon HN, Ferrucci L, Simonsick EM (2013) Hearing loss and cognitive decline in older adults. JAMA Int Med 173(4):293–299
10. Dalton DS, Cruickshanks KJ, Klein BE, Klein R, Wiley TL, Nondahl DM (2003) The impact of hearing loss on quality of life in older adults. Gerontologist 43(5):661–668
11. Wahl HW, Heyl V, Langer N (2008) Quality of life by limited vision in old age: the example of age-related macula degeneration. Ophthalmologe 105(8):735–743
12. Hallberg LR, Hallberg U, Kramer SE (2008) Self-reported hearing difficulties, communication strategies and psychological general well-being (quality of life) in patients with acquired hearing impairment. Disabil Rehabil 30(3):203–212
13. Lamoureux EL, Fenwick E, Moore K, Klaic M, Borschmann K, Hill K (2009) Impact of the severity of distance and near-vision impairment on depression and vision-specific quality of life in older people living in residential care. Invest Ophthalmol Vis Sci 50(9):4103–4109
14. Lupsakko T, Mantyjarvi M, Kautiainen H, Sulkava R (2002) Combined hearing and visual impairment and depression in a population aged 75 years and older. Int J Geriatr Psychiatry 17(9):808–813
15. McDonnall MC (2009) Risk factors for depression among older adults with dual sensory loss. Aging Ment Health 13(4):569–576
16. Finesilver SG (1970) A study of driving records, licensing requirements and insurability of physically impaired drivers (No. Rpt. No. RD 2283 G
17. Fox HG (1969) Driven Silent but skilled. Police Chief, 360
18. Mirkin A (1967) Medical conditions and driving ability. Md State Med J 16(54):59–62
19. Coppin RS, Peck RC (1963) The totally deaf driver in California: part I. Sacramento, California: California Department of Motor Vehicles, Division of Administration, Report No. 15, July, 1963
20. Coppin RS, Peck RC (1964) The totally deaf driver in California: part 11. Sacramento, California: California Department of Motor Vehicles, Division of Administration Report No. 16, Dec 1964
21. Roydhouse N (1967) Deafness and driving. N Z Med J 66:878–881

22. Booher HR (1978) Effects of visual and auditory impairment in driving performance. Hum Factors 20(3):307–319
23. Henderson RL, Burg A (1973) The role of vision and audition in truck and bus driving. Santa Monica, California: System Development Corporation, Thl(L)-5260/ OOWOO, Dec 1973
24. Wood JM, Troutbeck R (1995) Elderly drivers and simulated visual impairment. Optom Vis Sci 72:115–124
25. Keltner JL, Johnson CA (1987) Visual function, driving safety, and the elderly. Ophthalmology 94:1180–1188
26. Hakamies-Blomqvist LE (1993) Fatal accidents of older drivers. Accid Anal Prev 25:19–27
27. Pelli DG, Robson JG, Wilkins AJ (1988) The design of a new chart for measuring contrast sensitivity. Clin Vis Sci 2:187–199
28. Rubin GS, Roche KB, Prasada-Rao P, Fried LP (1994) Visual impairment and disability in older adults. Clin Vis Sci 71:750–756
29. Humphriss D (1987) Three South African studies on the relation between road accidents and drivers' vision. Ophrhal Physiol Opt 7:73–79
30. Norman LG (1960) Medical aspects of road safety. Lancet 1:989–994
31. Faraldi I (1975) Colour vision and road safety. In: Reports of 1st international conference on vision and road safety, Paris, 1974, Prevention Rout & eInternationale, Linas-Montlhery, pp 267–269
32. Coles BL, Brown B (1966) Optimum intensity of red road-traffic signal lights for normal and protanopic observers. J Opt Sot Am A 56:516–522
33. Christou EA (2011) Aging and variability of voluntary contractions. Exerc Sport Sci Rev 39:77–84. https://doi.org/10.1097/JES.0b013e31820b85ab
34. Woodworth RS (1899) The accuracy of voluntary movement. In: Baldwin JM, Cattell JM (eds) The psychological review, vol III. The Macmillan Company, New York, pp 1–114. https://doi.org/10.1037/h0092992
35. Toledo T, Koutsopoulos HN, Ben-Akiva M (2007) Integrated driving behaviour modeling. Transp Res C Emerg Technol 15:96–112. https://doi.org/10.1016/j.trc.2007.02.002
36. Lodha N, Coombes SA, Cauraugh JH (2012) Bimanual isometric force control: asymmetry and coordination evidence post stroke. Clin Neurophysiol 123:787–795. https://doi.org/10.1016/j.clinph.2011.08.014
37. Kwon M, Chen YT, Fox EJ, Christou EA (2014) Aging and limb alter the neuromuscular control of goal-directed movements. Exp Brain Res 232:1759–1771. https://doi.org/10.1007/s00221-014-3868-2
38. Kennedy DM, Christou EA (2011) Greater amount of visual information exacerbates force control in older adults during constant isometric contractions. Exp Brain Res 213:351–361. https://doi.org/10.1007/s00221-011-2777-x
39. Chen YT, Pinto Neto O, de Miranda Marzullo AC, Kennedy DM, Fox EJ, Christou EA (2012) Age-associated impairment in endpoint accuracy of goal-directed contractions performed with two fingers is due to altered activation of the synergistic muscles. Exp Gerontol 47:519–526. https://doi.org/10.1016/j.exger.2012.04.007
40. Enoka RM, Christou EA, Hunter SK et al (2003) Mechanisms that contribute to differences in motor performance between young and old adults. J Electromyogr Kinesiol 13:1–12
41. Sosnoff JJ, Newell KM (2006) Information processing limitations with aging in the visual scaling of isometric force. Exp Brain Res 170:423–432
42. Vaillancourt DE, Larsson L, Newell KM (2003) Effects of aging on force variability, single motor unit discharge patterns, and the structure of 10, 20, and 40 Hz EMG activity. Neurobiol Aging 24:25–35
43. Baweja HS, Kwon M, Christou EA (2012) Magnified visual feedback exacerbates positional variability in older adults due to altered modulation of the primary agonist muscle. Exp Brain Res 222:355–364. https://doi.org/10.1007/s00221-012-3219-0
44. Kwon M, Baweja HS, Christou EA (2011) Age-associated differences in positional variability are greater with the lower limb. J Mot Behav 43(5):357–360

45. Christou EA, Yang Y, Rosengren KS (2003) Rapid communication. Taiji training improves knee extensor strength and force control in older adults. J Gerontol Ser A: Biol Sci Med Sci 58(8):M763–M766
46. Lodha N, Moon H, Kim C, Onushko T, Christou EA (2016) Motor output variability impairs driving ability in older adults. J Gerontol Ser A: Biomed Sci Med Sci 71(12):1676–1681
47. Ahmed SM, Yasmeen A, Jagadeesh Babu B (2018) Wheelchair with auto navigation for Adults with Physio and cognitive impairments. Int J Innov Technol Explor Eng (IJITEE) 8(2s2):458–462
48. Ahmed SM, Shireen A, Jagadeesh Babu B, Shruti (2019) Powered wheelchair for mobility with features to address physical strength, cognitive response, and motor action development issues. In: International conference on data science, machine learning and applications (ICDSMLA), CMRIT Hyderabad, 29–30 Mar 2019
49. Hull M (1991) Driver licence review: functionally impaired and older drivers (no. DP91/1)
50. Lange JE, McKnight AJ (1996) Age-based road test policy evaluation. Transp Res Rec 1550(1):81–87
51. Torpey SE (1986) Licence retesting of older drivers
52. McPherson K, Ostrow A, Shaffron P, Michael J (1988) Physical fitness and the aging driver. Phase I (no. HS-040 558)
53. World Health Organization (1996) The Heidelberg guidelines for promoting physical activity among older persons
54. Ball K, Owsley C, Sloane ME, Roenker DL, Bruni JR (1993) Visual attention problems as a predictor of vehicle crashes in older drivers. Invest Ophthalmol Vis Sci 34(11):3110–3123
55. Roenker DL, Cissell GM, Ball KK (2001) The effects of useful field of view and driver simulator on driving performance. Unpublished manuscript
56. Büssing A, Michalsen A, Khalsa SBS, Telles S, Sherman KJ (2012) Effects of yoga on mental and physical health: a short summary of reviews. Evid Complement Altern Med
57. Jeter PE, Moonaz SH, Bittner AK, Dagnelie G (2015) Ashtanga-based yoga therapy increases the sensory contribution to postural stability in visually-impaired persons at risk for falls as measured by the wii balance board: a pilot randomized controlled trial. PLoS ONE 10(6):e0129646
58. Taneja MK (2015) Improving hearing performance through yoga. J Yoga PhysTher 5(194):2

Artificial Intelligence Oriented Security System Using Alexa

Gandu Harilatha and K. Riyazuddin

Abstract An Amazon Alexa and Raspberry Pi that enables a user to grasp World Health Organization specifically is knock at their door mistreatment biometric authentication. “Alexa, World Health Organization is at the Door?” is associate Amazon Alexa talent set that utilizes Raspberry Pi in conjunction with a biometric authentication API to allow you to apprehend World Health Organization is knock at your door. After Alexa gets the approval to require an image of the visitor’s face, it sends a command to the Raspberry Pi to require an image of them victimization the camera module. the image is then processed through a automatic face recognition that then determines whether or not or not the face is recognized, unknown, or not there in the least. when Alexa is tutored a replacement face, the person is given a reputation or ID by the user that is then saved within the information. It is re-taught many times in every occasion that the visitant involves your home. This way, the visitor’s face is recognized in numerous conditions like brightness, darkness, or a small modification in look of the individual. This conjointly improves Alexa’s speed at that she acknowledges every new friend that’s more to the information. This undertaking is planned to be an entire framework for face acknowledgment: easy to make, ease cost and compelling. Primary object is to be set as A caution for home visitors and supply information concerning the guests during a dynamic site and telephone application. It might be used in elective fields like businesses, workplaces and even air terminals for particular wished people. Among the inverse biometric strategies, face acknowledgment approach offers one decent favorable position that is ease of use.

Keywords Alexa Voice Service (AVS) · Artificial intelligence · Raspberry Pi (RPI)

G. Harilatha (✉) · K. Riyazuddin
Department of ECE, AITS, Rajampet, India
e-mail: harilathayadav1996@gmail.com

K. Riyazuddin
e-mail: shaik.riyazuddin7@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_32

1 Introduction

Alexa Voice Service (AVS) is Amazon's suite of services designed around its voice-controlled AI assistant for the house and alternative environments. AVS and Alexa were 1st introduced with Echo, the company's intelligent speaker, that allows voice interaction with numerous systems within the atmosphere and on-line. After Alexa gets the approval to require an image of the visitor's face, it sends a command to the Raspberry Pi to require an image of them exploitation the camera module. The image is then processed through a biometric identification that then determines whether or not or not the face is recognized, unknown, or not there in the least. Once Alexa is schooled a replacement face, the person is given a reputation or ID by the user that is then saved within the info. It will be re-taught many times in every occasion that the traveller involves your home. This way, the visitor's face will be recognized in several conditions like brightness, darkness, or a small modification in look of the Individual [1]. This conjointly improves Alexa's speed at that she acknowledges every new friend that's else to the info.

2 Literature Survey

2.1 Smart Door System for Home Security Using Raspberry Pi3

Keen advanced entryway lock could be a framework to watch and the board numerous gadgets inside the home. Our great computerized entryway lock framework works over net system by exploitation raspberry pi3.

Information area expects to include the key by a newcomer and if the mystery's legitimate or the picture of the newcomer is coordinating with the picture inside the data, the entryway is opened, else the entryway isn't opened and a photo is taken and sent to the proprietor of the house by email thus if the proprietor of the house needs to allow him to enter, the key and in this way the picture of the individual is keep inside the information [2]. The individual is allowed authorization to enter whenever by supporting the verification of the catchword inside the information or his/her picture.

2.2 IoT based Smart Home Design utilizing Power and Security Management Model uses temperature sensor and smoke sensors to check for flame at the client's home, PIR movement Sensors to see for the undesirable nearness at their homes related furthermore screen and the board the $64,000 time pursue and change of all their electrical gadgets through a mechanical man based for the most part portable application. The framework is associated with the present mechanical man application misuse net property for higher and snappy correspondence. The model has partner plausibility of predominant gadgets by either causation voice directions or by simple tap-to-flip framework, making the general framework easy to use and easy to oversee. along these lines moving into insight about the model, we have a temperature identifier that works along the edge of smoke sensor to see the nearness of chimney gathering, PIR movement finder to locate the human development inside the house, and hand-off associated gadgets so they'll be essentially flipped by the microcontroller. The cerebrum of our model is partner neighborhood based for the most part Intel uranologist ordinal Generation Board that let our gadgets and sensors associated with the net. The ordinal age Intel uranologist board gives one board that depends on the Intel Quark SoCX1000, a 32-bit Intel Pentium processor-class framework on a chip (SoC). It is Arduino-certificated and intended to be equipment, programming, and stick perfect with monster change of Arduino Uno R3 shields [3].

2.3 Automatic Service Request System for Security in reasonable Home exploitation IoT Automatic administration demand framework is implemented by utilizing Advanced diminished guidance set registering machine ARM) controller utilizing net of things (IoT).

These Raspberry Pi based for the most part frameworks will be basically implemented much, and a large portion of them will be utilized with accomplishment in reasonable applications. The Arm Controller might be a 64-bit controller at that point it's a great deal of gifts contrasted with common 8-bit or 16-bit microcontrollers. The Arm configuration is predicated on Reduced Instruction Set pc (RISC) rather than run of the mill propelled guidance set PC (CISC) all together that it will be utilized for ongoing application since it has high code thickness and brilliant reaction to intrudes. This chips away at pipelining instrument, at that point the presentation of controller is high. In pipelining instrument, headings are handled in the meantime exploitation three entirely unexpected cycles of bring, modify and execute very surprising bearings for example when guidance is dead by processor consequent guidance is in rework state and option next guidance is brought from memory to processor [4].

3 Existing Method

Savvy advanced entryway lock framework works over net system by exploitation raspberry pi3. the framework structure comprises of the ensuing 3 stages: (i) input, (ii) procedure and (iii) yield. Information area expects to include the key by a newcomer and if the key's substantial or the picture of the newcomer is coordinating with the picture inside the data, the entryway is opened, else the entryway isn't opened and picture is taken and sent to the proprietor of the house by email so the house needs to allow him to enter, the key and furthermore the picture of the individual is hang on inside the data. The individual is conceded consent to enter whenever by favoring the verification of the slogan inside the information or his/her picture this paper plans to blessing a totally one of a kind security home framework upheld Bluetooth arrange. the most objective is to build up an embodiment that has the adaptability to recreate the remote undertakings just as perception and prevailing advanced entryway lock. Such a framework would have the bent to deliver secure and controlled home apparatuses [5].

4 Problem Definition

As in our regular day to day existences we have information concerning everything readily available in light of the headway of grouped innovations. however, we'd like to handle concerning the various visitors visiting places (like working environment, home, airplane terminals, bank-storage spaces). We have to follow data concerning visitors for various inevitabilities like

1. Section of unapproved staff at desk work area.
2. Section of staff gathering.
3. At the spot of bank storage spaces.
4. Abusing the security standards at flying field, Malls.
5. Stealer distinguishing proof gathering, Jewelers search, Banks, Warehouse.
6. Investigation research centers and so on.

5 Proposed Method

Utilizing Alexa and counterfeit keen security framework will be made which may be utilized for face distinguishing proof. the most goal of this strategy is to execute for a chosen face and particular it from assortment of hang on appearances with some ongoing varieties moreover. this technique functions as clarified underneath Whenever a voyager presses the catch, a Pi camera catches picture of it [6]. On this picture it takes exclusively facial a piece of it that is required for highlight extraction, that is then increasingly handled and sends to the information. It attempts to coordinate the picture dotty the hang on data in it. When it's coordinated, it offers notice to the proprietor exploitation IOT for giving the entrance. On the off chance that the picture isn't coordinated proprietor will store the photos he can store the pictures inside the information by then. instructing process: we have prepared a few pictures of incredible people in databases. At the season of instructing it gathers most assortment of photos of the individual. Bigger the amount of tests, extra will be the exactness of the framework [7].

Block Diagram (Fig. 1).

Recognition process: As a little while later as voyager presses catch, camera catches picture and attempts to coordinate with the keep information in it. On the off chance that match discovered, at that point it sends voyager name to the proprietor of the house.

Conceding authorization: once accepting mail of explorer data, proprietor chooses whether or not to give get to or not by causing message as appeared in fig. On the off chance that he sends positive, at that point entryways are opened and in the event that sends no, at that point entryways can remain shut exclusively [8].

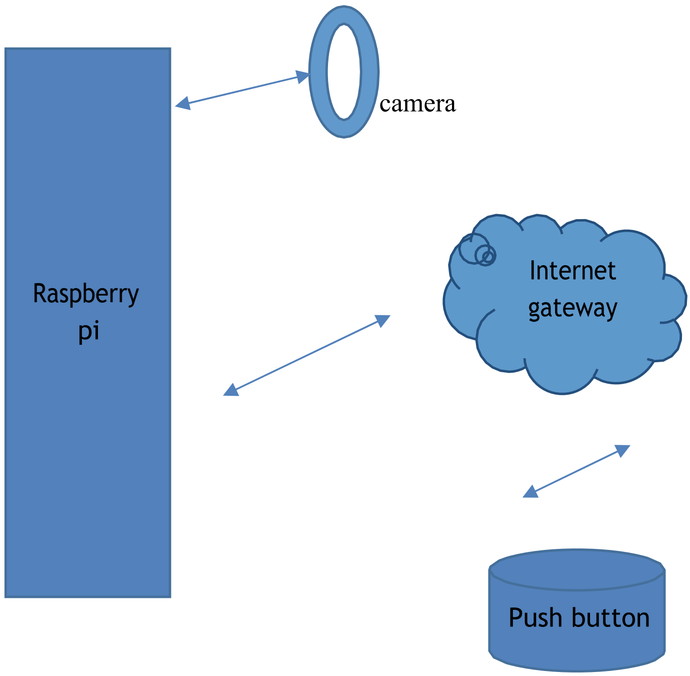


Fig. 1 Artificial intelligence oriented security system using Alexa

Calculation: Alexa fundamentally based security framework misuse AI and PI camera

Stage 1: Image are caught.
Stage 2: Detection of face is finished from caught picture.
Stage 3: acquiring required parameters from picture.
Stage 4: Face acknowledgment happens.
Stage 5: Name of explorer is conveyed to the proprietor through web, golem notice and so forth.

Flow Chart (Fig. 2).
Explanation of Flow Chart:

- once the voyager press the switch or activity catch, location of face happens. The recognized face will be caught by exploitation Raspberry Pi Camera.
- The caught picture is then handled through a biometric validation that at that point decides if or not the face is perceived obscure, or not there in the scarcest degree.
- once Alexa thought a pristine face is given a notoriety or ID by the client that is then spared in dat.

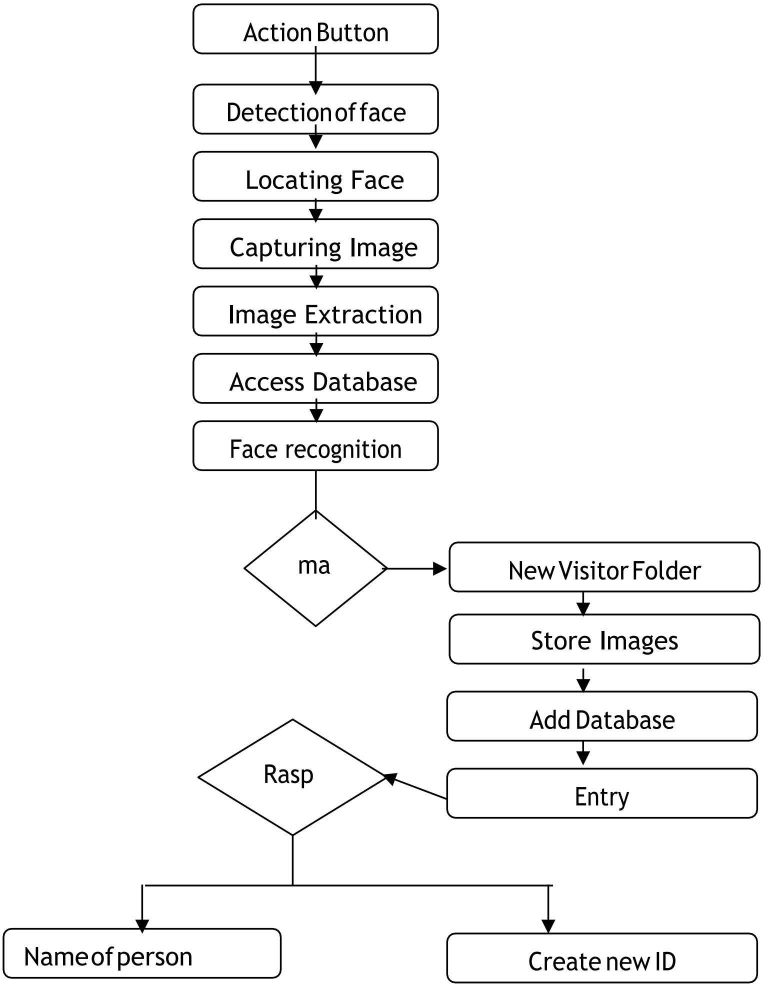


Fig. 2 Design flow of artificial intelligence of oriented security system using Alexa

Result and Discussion (Fig. 3).

To know the information of visitors already the database should be present to check the person is known person or not, that database is creating by using python (Figs. 4 and 5).

To create the database images of known persons should be captured, the captured images are Trained and saved with separate name.

When ever visitant involves visit home once Alexa gets the approval to require an image of the visitor's face, it sends a command to the Raspberry Pi to require an image of them mistreatment the camera module, If the image is already hold on

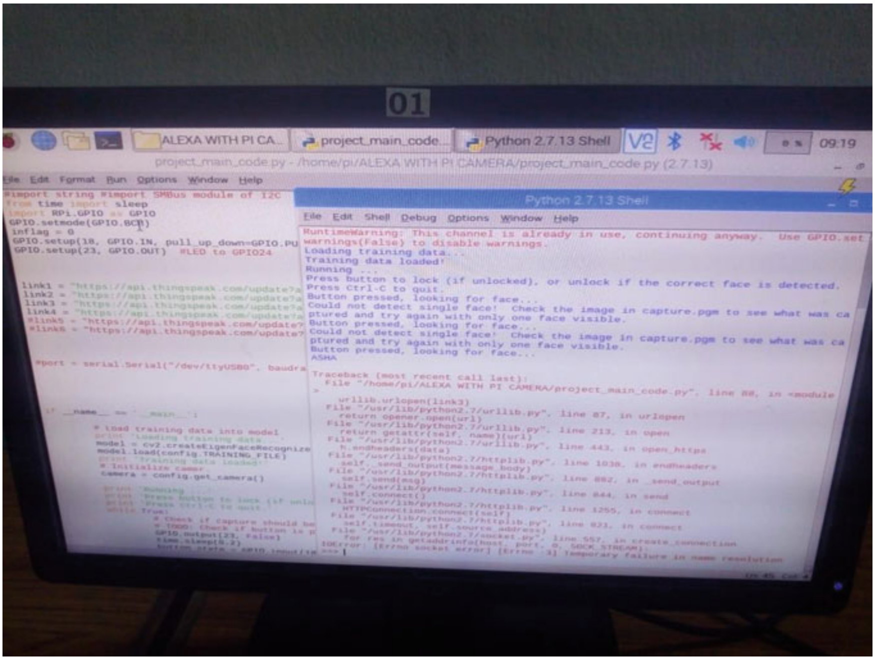


Fig. 3 Creating the database using python

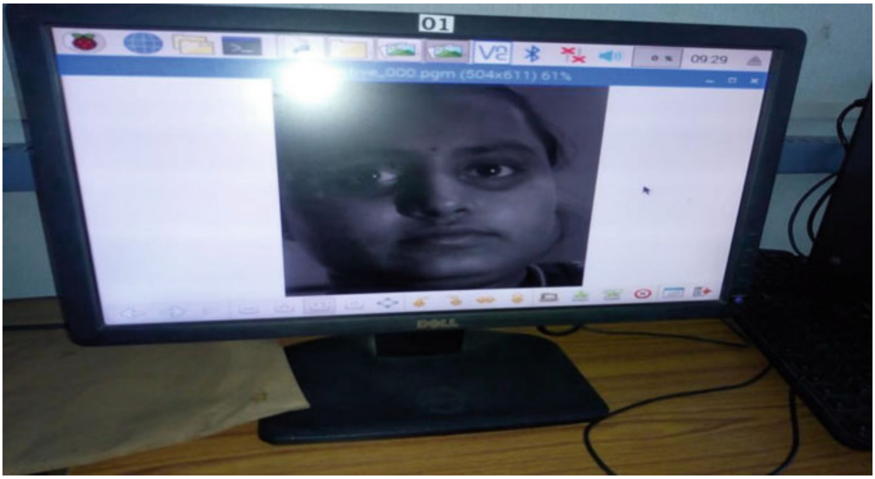


Fig. 4 Training of images

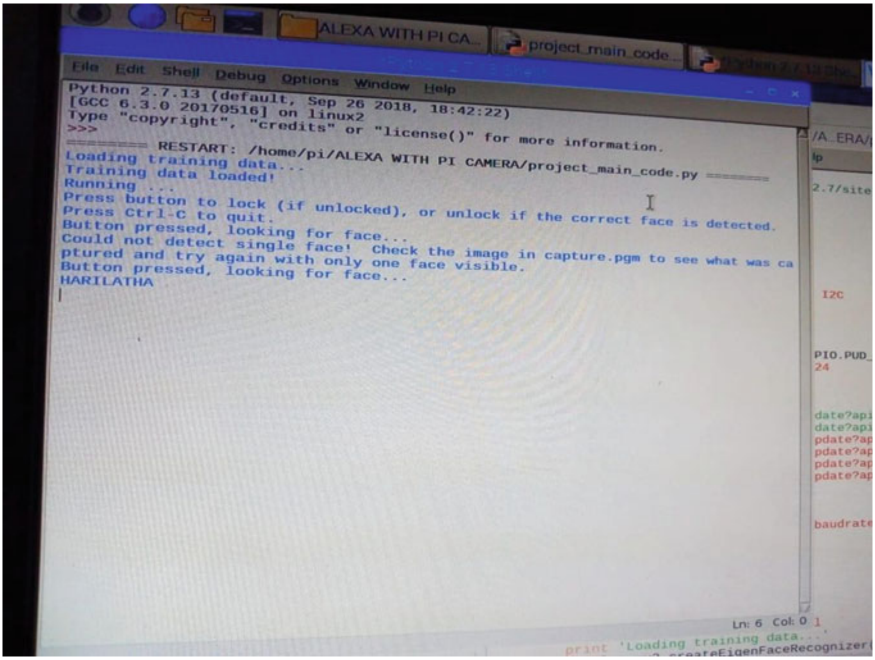


Fig. 5 Sending the visitor name to house owner

within the information base his name are going to be sent to accommodate owner (Fig. 6) [9].

We are creating the data base if the number images are more than name will be given accurately to house owner.

6 Future Scope

The first potential future for Alexa is one wherever the present trends principally continue and even accelerate. Amazon's own Alexa-based merchandise still sell well, with Dot in all probability taking a larger share of sales going forward relative to Echo (or Tap), marketing into the tens of uncountable put in units within the next few years. On prime of that, the adoption by third parties that was thus evident at CES continues, with even additional devices giving integration. significantly, Alexa starts to create Associate in Nursing look in robot smartphones, creating it as pervasive and omnipresent as existing smartphone-based assistants, presumably even creating Associate in Nursing look in another spherical of Amazon smartphones.

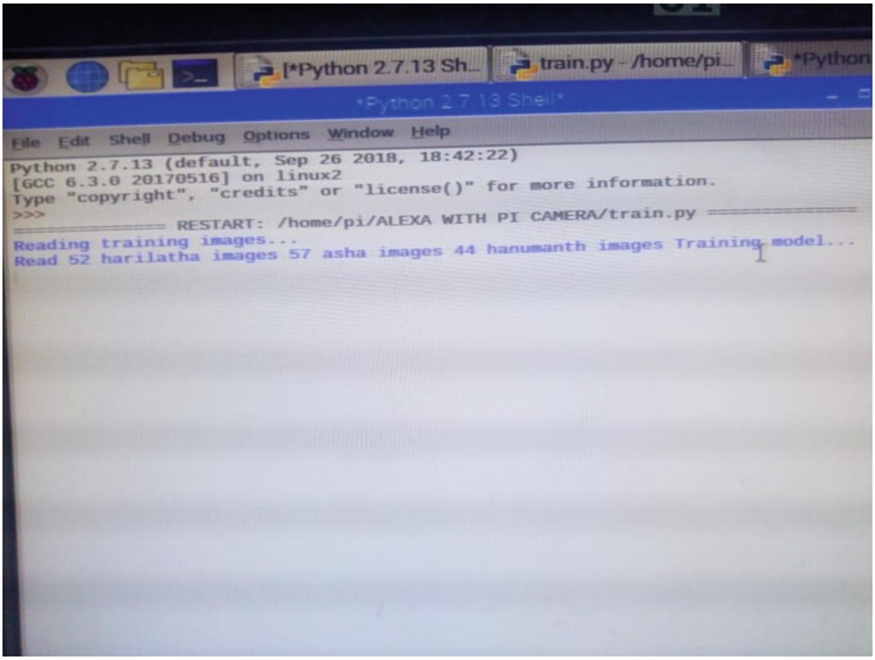


Fig. 6 Number of images trained

7 Conclusion

The objective of our undertaking is to benefit and guard human lives what's more on get the information with respect to visitors of wanted area to the proprietor abuse processing. The framework might be utilized at various areas like living arrangements, workplaces, ventures, bank storage spaces, air terminals for various capacities like crime recognizable proof, police examination and so forth.

References

1. Thabet AB, Amor NB, Enhanced smart doorbell system based on face recognition
2. Vermesan O, Friess P (2013) Internet of things: converging technologies for smart environments and integrated ecosystems (PDF). River Publishers, Aalborg, Denmark. ISBN 978-87-92982-96-4
3. Paul LC, Sumam AA, Face recognition using principal component analysis method
4. Sutawijaya IGP, Yudohusodo A, Jatmika AH, Real time face recognition engine using compact features for electronics key
5. Brown E (2016) Who needs the internet of things? Linux.com. Retrieved 23 Oct 2016
6. Niveditha GV, Sharmila BP, Manikantan K, DWT-based face recognition using fast walsh hadamard transform and chiral image superimposition as pre-processing techniques

7. Hitesh Babu S, Shreyas HR, Manikantan K, Face recognition using active illumination equalization and mirror image superposition as pre-processing techniques
8. An introduction to the Internet of Things (IoT) (PDF). Cisco.com. Lopez Research., San Francisco, California (2013). Retrieved 23 Oct 2016
9. Internet of things: science fiction or business fact? (PDF). Harvard Bus Rev (2014). Retrieved 23 Oct 2016

Performance of LoRa Network for IoT Applications

Ch. Nagaraju and Pandillapalli Asha Jyothi

Abstract System correspondence is most significant in our every day existence's without this correspondence we cont ready to arrive at our objectives. So correspondence is essential occupation in our universes transactions. But now a days because of the constrained range we can transmits the information at certain range only. To over come this issue we go for LoRa systems for IoT applications. This venture incorporates different highlights facilities. It utilizes remote sensor systems for taking note of the dirt properties and ecological factors ceaselessly. Different sensor hubs are sent at various areas in the fields. Observing these parameters are through the any showcase gadget and the tasks are performed by interfacing sensors. By utilizing this idea we can convey and transmits the information up to long range to better places one after another. The helpful thing in this LoRa systems it expends less power this is one of the preferred position to the drifting IoT applications.

Keywords LPWN · IoT · Sensors · Microcontroller · Wi-Fi

1 Introduction

It is intended to empower low-fueled units to speak with shrewd Internet-associated applications over long range Wi-Fi associations. LoRa (short for long range) Technology permits astute IoT applications that cure the absolute greatest difficulties managing earth: vitality the board, home grown guide decrease, foundation productivity and then some. A solitary portal or base station can offer transporter to a total city or a great many rectangular kilometers [1, 2].

Ch. Nagaraju · P. Asha Jyothi (✉)
Department of ECE, AITS, Rajampet, India
e-mail: ashajyothip11@gmail.com

Ch. Nagaraju
e-mail: chrajuaits@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_33

1.1 Why LoRa?

Dissimilar to other problematic applied sciences that can be delayed to accomplish worldwide reception, LoRa Technology is currently not a guarantee of a future potential, however it is accessible now daily's inside and out the globe. LoRa Technology has collected more than 600 realized utilizes occasions for cunning urban communities, savvy properties and structures, brilliant agribusiness, shrewd metering, keen production network and coordinations, and that's just the beginning. With more than 50 million gadgets identified with systems in 95 global areas and developing, LoRa Technology is the DNA of IoT, to developing of our Smarter Planet. In this paper we are observing the readings of the savvy IoT applications and the utilization of Arduino IDE hub MCU.

The LoRa topology offers us the one of a kind sorts of IoT programming focused observing readings. For future better Wi-Fi verbal trade here we have to place into impact the exhibition of LoRa systems for IoT applications.

Here we have to upgrade remote availability to assemble the smart planet the utilization of long range correspondences. Here the correspondence is at some phase in the middle of the end hubs of the LoRa gadgets.

Here in this venture we are utilizing Arduino IDE with inbuilt LoRa arrange the three LoRa hubs associated with three unmistakable sensors it will gives the detecting data to the LoRa handsets. These handsets give the readings to the door and spared there as it were. The spared records can take the readings by methods for ace Arduino and showed in the LCD. By utilizing this readings we can screen the particular sorts of IoT works as where are in sitting position [3–6].

2 Existing Methods

In this project we are mainly talking about four existing methods those are [7–10]

1. Bluetooth
2. GSM
3. Zigbee Technology
4. NB IOT

In the above existing 4 methods are having very short range with variations of distances and limited nodes (Fig. 1).

By using Bluetooth we can able to transfer the data up to certain distance only with only one node at a time.

By using GSM module we can transmit the data with same bandwidth to limited nodes at a time.

By using Zigbee we can transmit more data with short distance, and it is not secured, and in Zigbee the transmission speed is very slow. Network stability is also low.

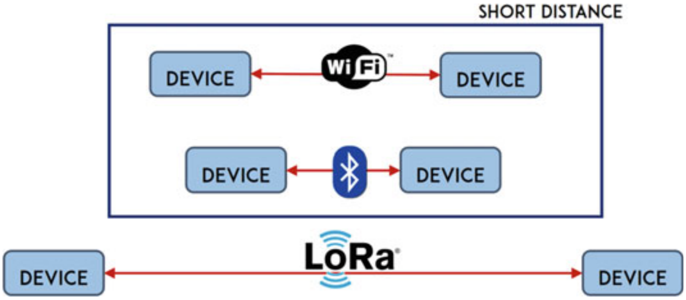


Fig. 1 Block diagram showing short and long range distances

Nb IoT is the short form of the narrow band internet of things the data rate is low and it does not support volte and it is not support for roaming also.

To overcome all the above problems we go for LoRa networks. It gives the long range for transactions with sufficient nodes at a time.

3 Proposed Method

In the existing method they are all giving short range communication here in the proposed method we are using long range by using LoRa topology. In this paper we are observing the readings in display device by using different sensors and observed the performance of Lora networks in different IoT applications [11–14].

3.1 LoRa Topology

- It is intended to empower low-fueled units to talk with Internet-associated applications over long range Wi-Fi associations.
- LoRa (short for long range) Technology empowers shrewd IoT applications that cure the absolute biggest difficulties confronting earth: power the board, normal guide decrease, two foundation productivity and the sky is the limit from there.

A solitary passage or base station can gives the range to a whole town or several square kilometers (Fig. 2).

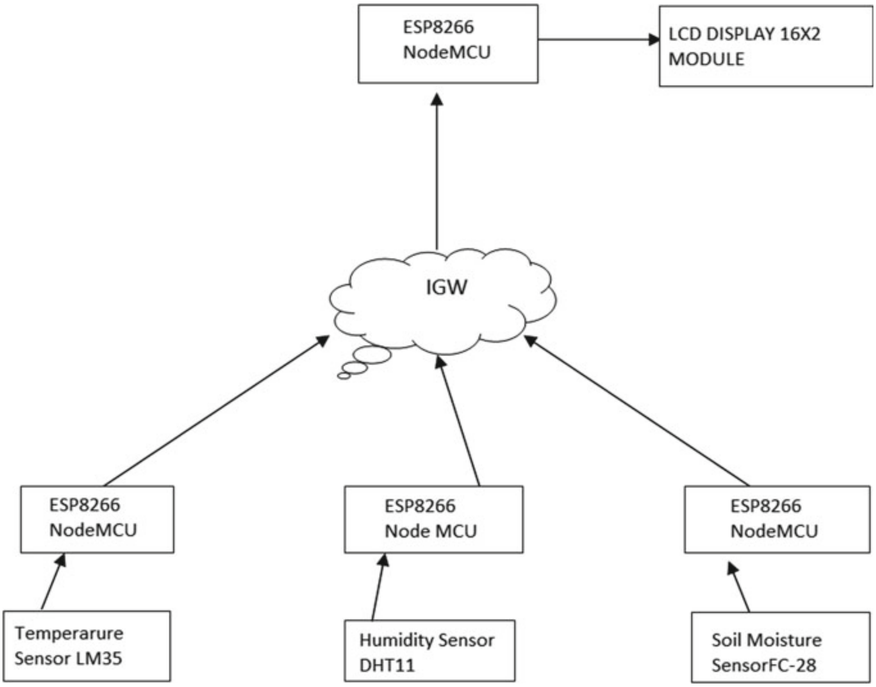


Fig. 2 Block diagram of proposed network

4 Hardware Used

In this undertaking we are utilized and sent in the fields like temperature sensor, dampness sensor and soil dampness sensor, LCD show (16 × 2 module), and micro-controller hubs. The information gathered from the sensors are associated with the microcontroller through the ADC converter.

4.1 ESP8266 Node MCU

The ESP8266 Wi-Fi Module is an independent SOC with inbuilt TCP/IP convention stack it is supply to our Arduino microcontroller to get right of section to our Wi-Fi arrange. The ESP8266 is the name of a smaller scale controller structured by Espressif Systems. The ESP8266 itself is an independent Wi-Fi organizing arrangement offering us an extension from existing smaller scale controller to Wi-Fi and it is additionally fit for running independent applications. This module is utilized with

Fig. 3 Printed circuit board of ESP8266 node MCU

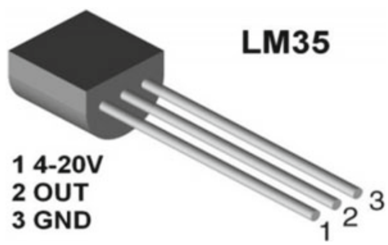
an implicit USB connector and a rich arrangement of stick outs. With a small scale USB link, here we are connected Node MCU devkit to our workstation and flashed it with no issue, much the same as Arduino. It is additionally quickly breadboard inviting.

In this project we are using this mainly to transfer the data from one place to another place with wireless connectivity. Here the Arduino IDE software is using for the ESP8266 node MCU. The sensors are connected to these node MCU and gives the sensed data to the gateway (Fig. 3).

4.2 Temperature Sensor

The LM 35 sensor is exceptionally utilized on the grounds that its yield voltage is straight with the Celsius scaling of temperature. It doesn't give any outer cutting. It has a wide working extent. The most extreme yield is 5 V. The yield will increment 10 mV for each one degree ascend in temperature. The range is from −55° to + 150°. There are three terminals as Vcc, Ground and the simple sensor. It expends least measure of power. Along these lines, it is vitality proficient. It is productive in cultivation. It is easy to use to utilize (Fig. 4).

Fig. 4 Pin configuration of LM35 temprature sensor



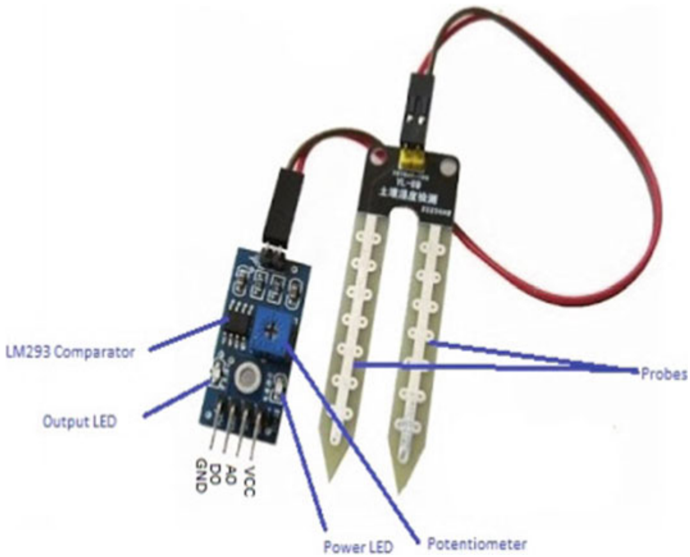


Fig. 5 Configuration of soil moisture sensor

4.3 Soil Moisture Sensor

Soil moisture sensor is a sensor which detects the dampness substance of the dirt. The sensor has both the simple and the computerized yield. The computerized yield is fixed and the simple yield limit can be fluctuated. It chips away at the guideline of open and short out. The yield is high or low shown by the LED. At the point when the dirt is dry, the present won't go through it thus it will go about as open circuit. Henceforth the yield is said to be greatest. At the point when the dirt is wet, the present will go from one terminal to the next and the circuit is said to be short and the yield will be zero. The sensor is platinum covered to make the effectiveness high. The scope of detecting is additionally high. It is hostile to rust thus the sensor has long life which will manage the cost of the rancher at any rate cost (Fig. 5).

4.4 Humidity Sensor

A moistness sensor faculties and measures the stickiness noticeable all around and it reports the level of dampness noticeable all around. The proportion of dampness noticeable all around to the most noteworthy measure of dampness at a specific air temperature is called relative stickiness relative mugginess is a significant factor when searching for solace. Mugginess sensor work by distinguishing changes that caution electrical flows noticeable all around (Fig. 6).

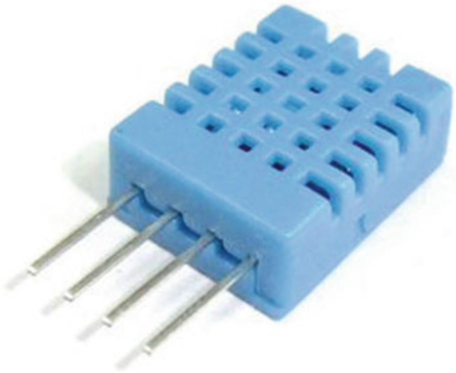


Fig. 6 Configuration of wetness device

DHT11 is a moistness sensor witch creates aligned advanced yield. Here we are interfacing DHT11 sensor with the Arduino microcontroller and get immediate outcome. DHT11 is a minimal effort dampness which gives high unwavering quality and long haul solidness.

4.5 LCD Display

We are using 16 * 2 lines of LCD display to monitor the output of the project. The six data pins will connect to the Arduino microcontroller board and remaining two will connect to the ground and power supply respectively (Fig. 7).

5 Software Used

Here we are using Arduino IDE and Embedded C. An Arduino IDE is the best simulation platform by using Embedded C programming it is the simple programming language to the any microcontrollers to perform the device according to our requirement. It can be used to test programs and embedded designs for electronics before actual testing.

6 Experimental Results and Discussion

The hardware is interfaced with all the sensors in the board. The hardware components includes the microcontrollers and LCD display, ADC converter and all the sensors interfaced. Here the LCD display gives the data through the IIC communication. So in the IIC communication is working based on master and slave concept. The

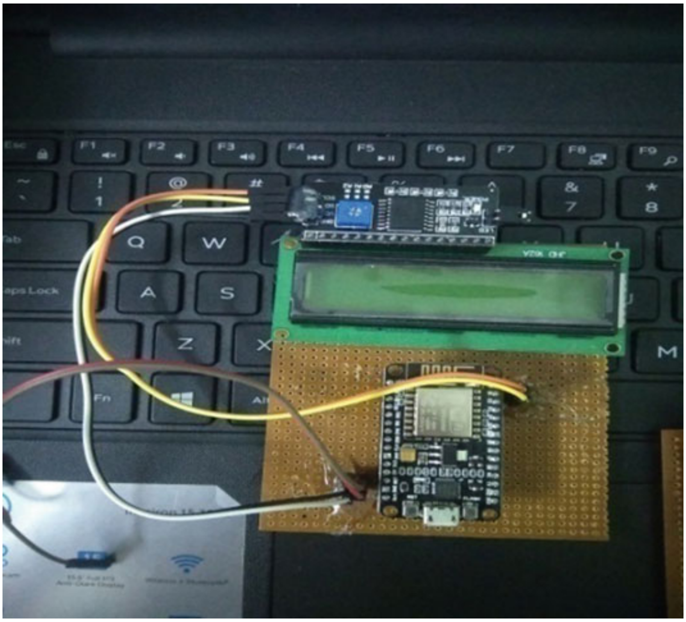


Fig. 7 Liquid crystal display with I^2C interface

three node microcontrollers are working as slaves and interfaced with three sensors those are humidity, soil moisture and temperature sensors and one microcontroller is acting as master which is connected to the LCD display and shows the respected sensor readings (Fig. 8).

Once the hardware setup is finished we have to restart to clear the previous data once the power is given to the board the slave nodes are gives the sensed data to the master node. Between this process the data is stored first of all in the gateway. This gateway transmits the data between the master node and slave nodes and also this gateway is gives the updated values at every time (Fig. 9).

The LCD display shows the updated values (Fig. 10).

The liquid crystal display is showing information with completely different readings once add time with sensing by the various sensors.

7 Conclusion and Future Scope

The sensors and microcontrollers are effectively interfaced and remote correspondence and long range is accomplished between different nodes. All perceptions and test tests are demonstrates that this venture is finished answer for the long separation interchanges. Execution of such a technique is in conveyed IoT applications

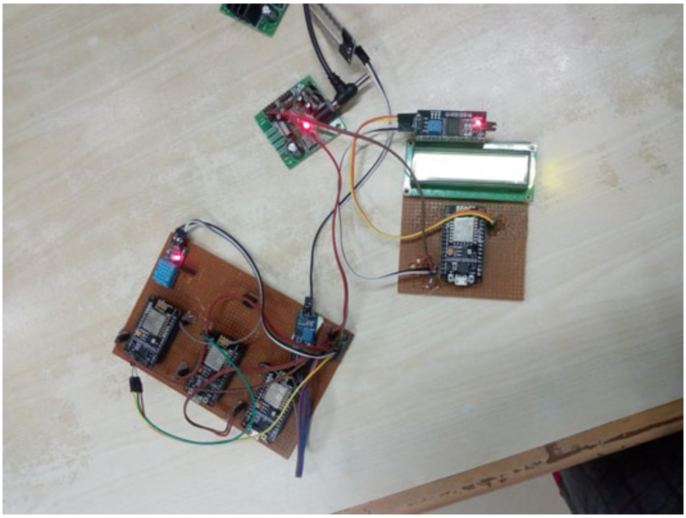


Fig. 8 Circuit connections between hardware modules

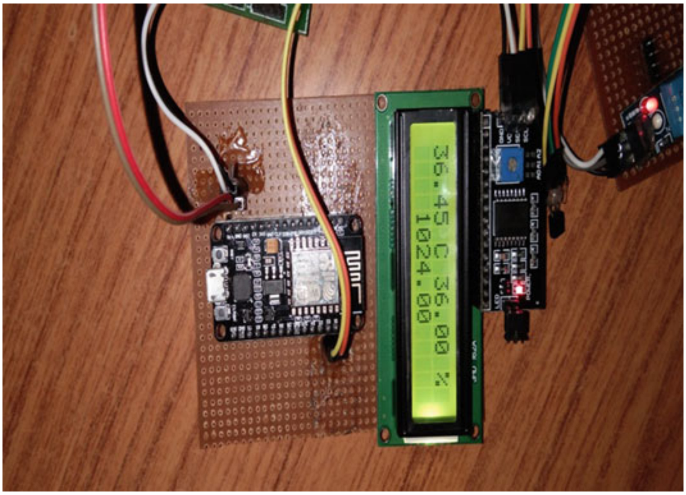


Fig. 9 Output of proposed project showing the various parameter values

improves the yield of the correspondence and over all systems. This LoRa topology offers the long range as much as ten kilometers.

The LoRa topology offers us the unique kinds of IoT software oriented monitoring readings. For future better Wi-Fi verbal exchange here we need to put into effect the performance of LoRa networks for IoT applications.

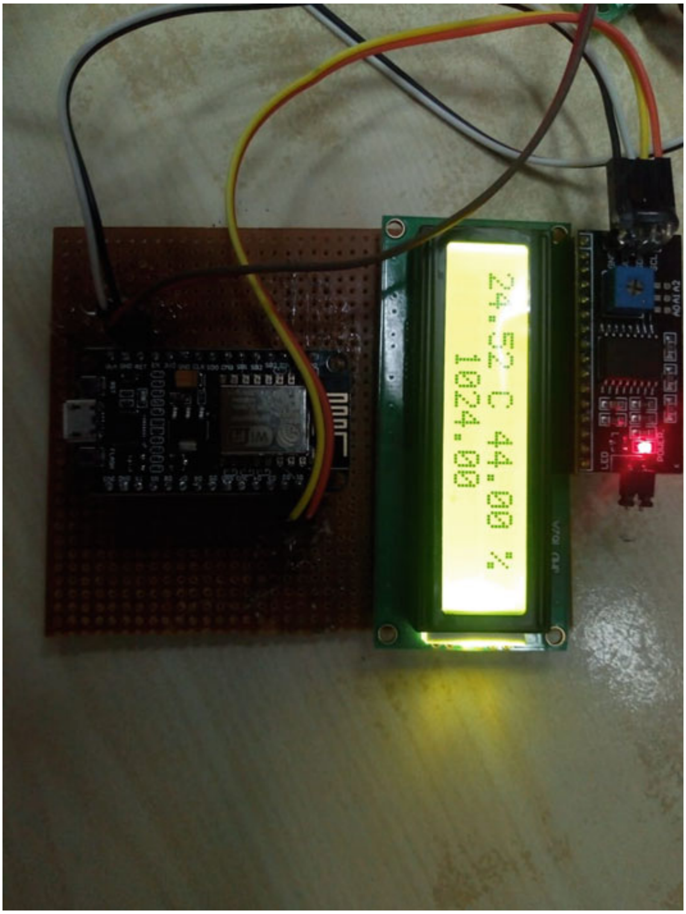


Fig. 10 Output of proposed project showing different parameter values w.r.t variations in the input

References

1. LoRa. https://www.lora-alliance.org. Accessed 7 Nov 2015
2. Sigfox. http://www.sigfox.com. Accessed 7 Nov 2015
3. Weightless open commonplace. http://www.weightless.org. Accessed 7 Nov 2015
4. Bachir A, Dohler M, Watteyne T, Leung KK (2010) MAC necessities for wireless device networks. IEEE Commun Surv Tutorials 12(2):222–248
5. Dutta P, Dawson-Haggerty S, Chen Y, Liang C-JM, Terzis A (2010) Style and analysis of a flexible and economical receiver-initiated link layer for low-power wireless. In: Proceedings of the eighth ACM conference on embedded networked device systems, SenSys ’10. ACM, New York, NY, USA, pp 1–14
6. Raza U, Kulkarni P, Sooriyabandara M (2017) Low power wide space networks: an outline. IEEE Commun Surv Tutorials

7. Centenaro M, Vangelista L, Zanella A, Zorzi M (2016) Long-range communications in unaccredited bands: the rising stars within the IoT and sensible town situations. IEEE Wirel Commun 23(5):60–67
8. LoRa Alliance (2015) LoRaWAN what's it? A technical summary of LoRa and LoRaWAN. https://www.lora-alliance.org/portals/0/documents/whitepapers/LoRaWAN101.pdf (Nov)
9. LoRa Alliance (2016) LoRaWAN specification (V1.0.2) (July)
10. The Items Network (2015) The issue Network Mission. https://github.com/TheThingsNetwork/Manifest/blob/master/Mission.md
11. Reynders B, Meert W, Pollin S (2017) Power and spreading factor control in low power wide area networks. In: IEEE ICC 2017 SAC symposium internet of things track (ICC'17 SAC-7 IoT), pp 1–5
12. Varsier N, Schwoerer J (2017) Capacity limits of Lo-RaWAN technology for smart metering applications. In: 2017 IEEE international conference on communications (ICC). IEEE
13. Augustin A, Yi J, Clausen T, Townsley WM (2016) A study of LoRa: long range & low power networks for the internet of things. Sensors 16(9):1466
14. Kim D-Y, Kim S, Hassan H, Park JH (2017) Adaptive data rate control in low power wide area networks for long range IoT services. J Comput Sci 33:171–178

Region Based Image Compression with Deep Learning and Binary Plane Difference Methods

B. P. Santosh Kumar and K. Venkata Ramanaiah

Abstract Image coding is treated as one of the important factor in multimedia communication and storage systems. In recent years with the rapid improvement in the usage of mobile internet there is a huge demand for faster transmission of the multimedia data. Medical image compression has received greater importance due to its heavy demand for decreasing its size without disturbing the quality of it. Binary plane technique has attained attention in recent time due to its capability of compressing the image both in lossy and lossless modes and achieving higher quality than existing JPEG and SPIHT methods. This paper presents an effective approach of extracting the region (ROI) and non-region (NROI) of interest from MR brain images and then compress the ROI in loss less mode and NROI in lossy mode.

Keywords Medical image compression · Region of interest · MR brain image · Deep learning

1 Introduction

Medical imaging is widely used in diagnosing the diseases and help doctors in surgical planning. Medical image acquisition devices generate tens of images per patient and it is observed that more than 500 Mb of storage space is allocated for one patient data. Hence this type of huge data requires a large storage space and requires larger bandwidth for its transmission [1, 2]. In Telemedicine, the acquired image with a sophisticated machinery has to be transmitted to a remote village or data centre for its research and analysis. In general, to transmit such data a huge bandwidth network is required and also the large data is subjected for multiple interferences. For all the above mentioned limitations, the medical images are supposed for efficient compression that could able to cut off the storage space and provide faster transmission with in a given bandwidth while preserving the quality of it.

B. P. Santosh Kumar (✉) · K. Venkata Ramanaiah
Department of Electronics and Communication Engineering, Y. S. R Engineering College of Yogi Vemana University, Proddatur, Kadapa, Andhra Pradesh 516360, India
e-mail: santoshyvuce@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_34

Medical image compression is treated as one of the solution for transmitting and storing these images [3]. The compression methods are broadly categorized into lossy and lossless methods however for a medical image a 10% lossy compression is not acceptable as it distorts the texture content of it leading to the wrong diagnosis. But performing a lossless mode operation may not be able to reduce the size to a considerable level.

Inorder to treat the limitations of both lossy and lossless modes of compression, the region of interest based compression is introduced. In this approach a selected region is processed for lossless compression and the rest in the lossy mode thereby providing the reduction in size while preserving its quality [4]. This paper presents an effective approach of extracting the brain region from MR images using deep learning methods and then apply the binary plane difference approach in lossless mode for it. The paper is organized as follows, Sect. 1 presents the introduction to region based compression model and scope of this work proceeding with related work in Sect. 2 where several works that were presented earlier were discussed. In Sect. 3, the concept of BPD is discussed with appropriate references, Sect. 4 discuss about the proposed methodology and the inclusion of deep networks in extract the ROI from medical images following with experimental results that were conducted on several medical images.

2 Related Work

Several region based compression algorithms were proposed so far in the literature but four of them are widely useful and suitable for practical applications [5]. It is evident that in region based image compression technique; segmentation or partition of particular region for the given image is the foremost step and later the corresponding compression algorithm are applied for the respective regions.

In literature there is no standard or universal accepted algorithm rather researchers are interested to propose and invent new segmentation algorithms [6] and these algorithms are relied on several factors. In [7] authors presented K-means and Chan-Veese algorithms for segmentation of the medical images. In this work they have used DICOM images and applied JPEG 2000 image codec for the segmented images.

In [1] Ravi et al., presents a block based PCA (principal Component analysis) approach for implementing their hybrid model of compression with lossy and lossless modes for ROI and NROI regions. The method uses canny edge operator for extraction of mask region which produces segmentation blobs and fails to neglect the skull region specifically for MR brain images.

Salih et al. in [8] uses a hybrid compression model in which the authors have applied lossless compression algorithm for region of interest and the other regions with lossy compression methods. In this analysis they applied the approach for CT images and compared the proposed approach with earlier algorithms with discrete cosine transforms and vector quantization.

In [9] Vikas et al., presented am fractal image compression approach for non-region of interest and also concentrated on mitigating the noise content in the image. In [10], Krishna et al. presented a skull stripping approach for extracting the brain region from MR image and then apply binary plane difference coding approach for both ROI and non-ROI in lossless and lossy modes and compared the proposed approach with traditional JPEG and SPIT algorithms and achieved high quality compressed at the same compression ratio.

This papers considers the idea of Krishna in [10] but it is observed that this method involved several morphological operations that involves several thresholds and structural elements of multiple size and shapes, this makes the methods more cumbersome and the same constants are not suitable for different images. Hence there is a scope and need to propose an adaptive approach that could able to extract the brain regions effectively.

So in this paper we include deep learning methods to extract the brain regions and then apply binary plane difference technique for ROI and non-ROI in lossless and lossy modes.

3 Binary Plane Difference Technique (BPD)

The generalized block diagram of BPD is depicted in Fig. 1.

This method can be applied in both modes (i) Lossless and (ii) Lossy modes. The loss less compression method is depends upon the spatial domain of the image and they are very much apt for the compression of medical images. The Lossy Binary Plane technique introduces little loss to achieve more compression rate. In this processes the compressed file is partitioned into two, the first is bit plane and the second is Data table. Usually the 1's and 0's are collection of bit plane, they represent pixel is repeating or not. Coming to the data table, it preserves only the required pixel values. Both bit plane and data table were combined into one file. The other loss less method of compression Huffman coding is applied for further compression, and then

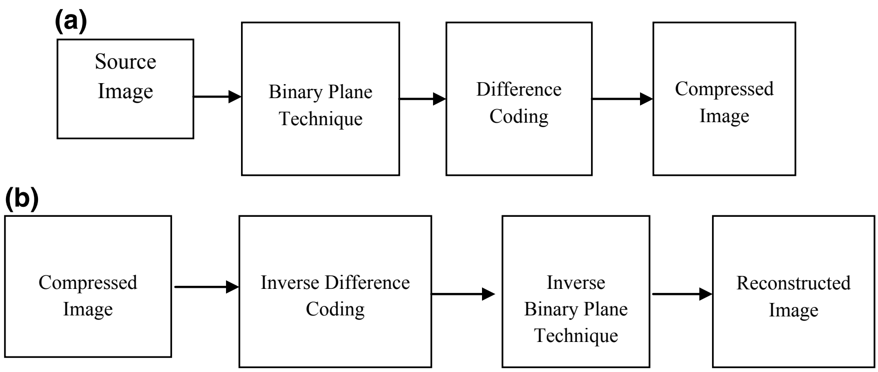


Fig. 1 Encoding and decoding process of the BPD approach

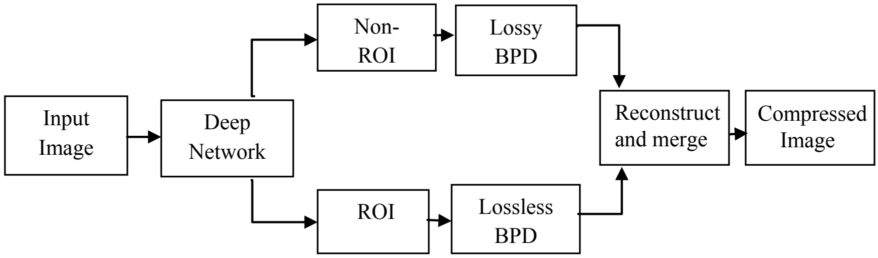


Fig. 2 Block diagram of the proposed compression model

it generates the final result of the compression. The main purpose of this method is to take improvement of repeated values in uninterrupted pixel positions. For a group of repeated successive values only one value is retained [10–12].

If current pixel is differing to the previous pixel code a '1' is used. It moves the current pixel to Data Table. If current pixel is similar to the previous pixel we use 'Code 0'. This removes the storage of current pixel. After generating and combining the Bit Plane and Data Table, Huffman coding is applied and final form of compressed file is produced. In the lossy mode a quantizer is incorporated in the approach that considers a threshold value (TV) which varies in [4–32] to control the degree of quantization. The approximate value of the current pixel is predicted based on this threshold value. The current pixel is an approximate value of the previous pixel if the current pixel lies within the stipulated range [13].

4 Proposed Approach

The block diagram of the proposed approach is depicted in Fig. 2.

As presented in our previous papers [14, 15], we proposed region of interest extraction with morphological operations and deep learning techniques. In this work we have incorporated the deep learning concepts that were developed for region extraction. The results of extracted region are presented in the next section. Thus, obtained ROI and Non-ROI are further processed for BPD in lossless and lossy modes with a varying threshold value. After decoding process of BPD, the respective data tables are merged together to form a compressed image.

5 Experimental Results

For evaluating the process of the proposed approach the algorithm is tested with several medical brain MR images with size $224 \times 224 \times 3$ which are taken from [16]. The input images are processed for deep networks, in this work we considered a fully connected convolutional neural network of 5 layers. The first convolutional

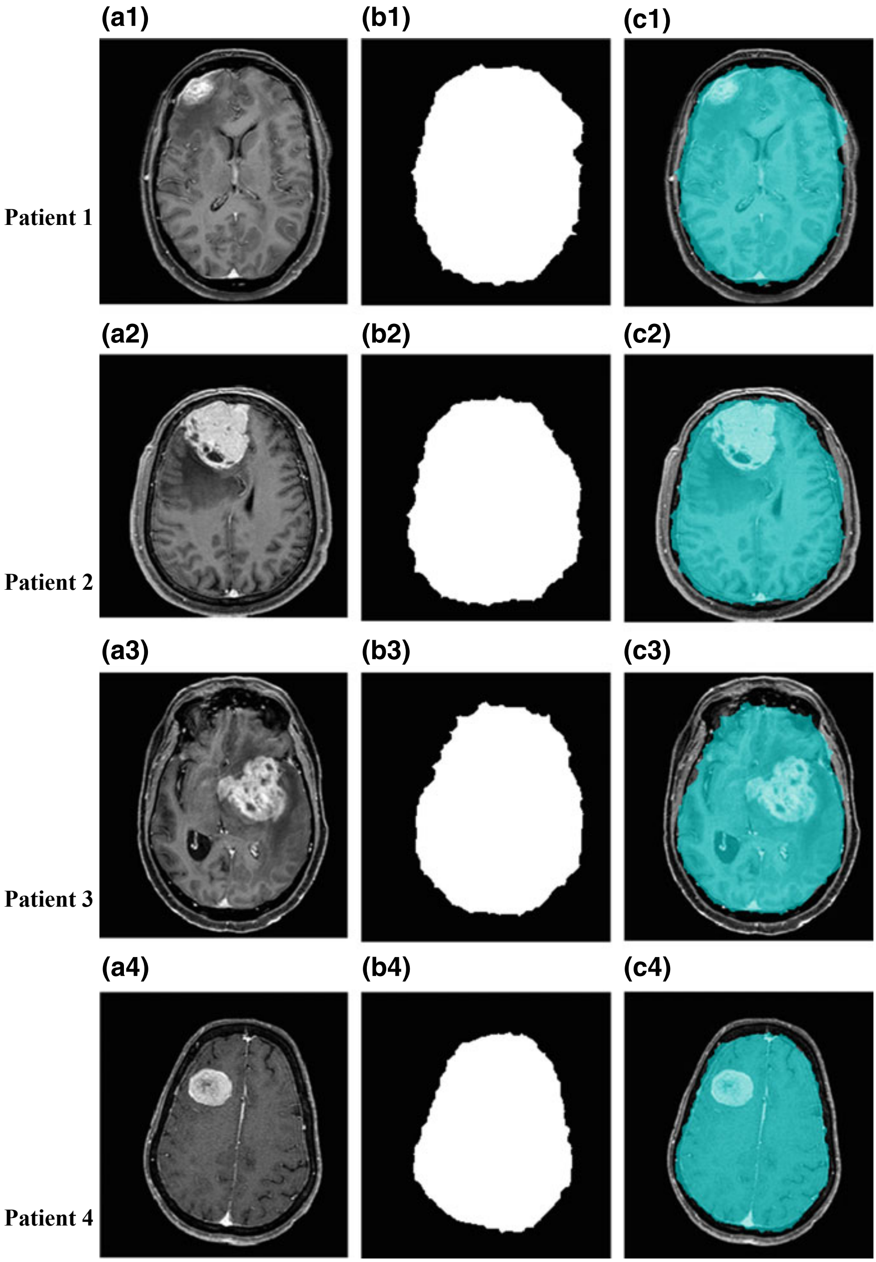


Fig. 3 Extracted regions using deep networks (a1–a5) input images (b1–b5) extracted ROI (c1–c5) overlapped with original images

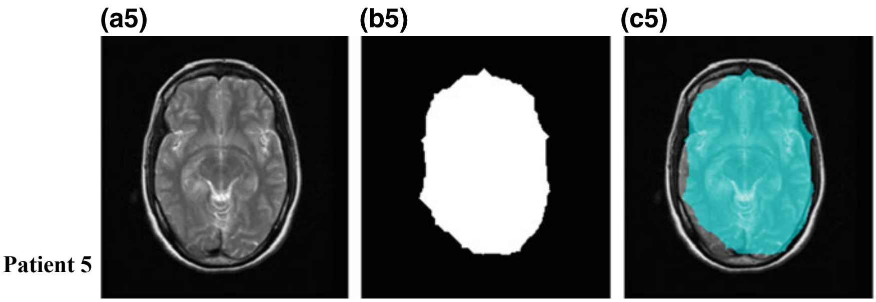


Fig. 3 (continued)

Table 1 Performance analysis of the proposed approach and its comparison

Image with size (bits)	Algorithm-used	Compression ratio	PSNR
Patient 1 (21900)	JPEG	1.103	37.189
	SPIHT	1.107	33.97
	Proposed	1.137	48.0324
Patient 2 (19200)	JPEG	1.061	31.121
	SPIHT	1.102	30.989
	Proposed	1.116	45.0824
Patient 3 (20100)	JPEG	1.124	33.580
	SPIHT	1.135	33.124
	Proposed	1.177	44.452
Patient 4 (16700)	JPEG	1.504	36.278
	SPIHT	1.65	35.162
	Proposed	1.454	42.254
Patient 5 (21400)	JPEG	1.173	35.45
	SPIHT	1.181	33.356
	Proposed	1.234	42.229

layer contains 64, 3 × 3 filters and likewise 128, 256 and 512. The extracted ROI from brain images are presented in Fig. 3.

To evaluate the performance of the proposed approach, several metrics were calculated and compared against the traditional JPEG [17] and SPIHT [18]. The results were tabulated in Table 1 (Fig. 4).

6 Conclusion

A region based hybrid compression model with deep networks and binary plane difference approach is proposed in this paper. The images are first subjected for

Fig. 4 PSNR analysis of the proposed approach and its comparison for multiple brain images

region extraction using deep networks and then further processed for compression using BPD in lossless and lossy modes. The method is evaluated with several metrics like compression ratio and PSNR and these are compared against the traditional compression model metrics. From the observation it is analyzed that the proposed approach could able to attain approximately same compression ratio with high quality which is a good achievement in meeting the medical image compression objective.

References

1. Kiran R, Kamargaonkar C (2016) Region separation techniques for medical images. Int J Innov Res Sci Eng Technol 5(2)
2. Ting LS, Weng DYF, Manap NBA (2015) A novel approach for arbitrary-shape ROI compression of medical image using principal component analysis (PCA). Trends Appl Sci Res 10:68–76 (Academic Journals Inc)
3. Miaou S, Ke F, Chen S (2009) A lossless compression method for medical image sequences using JPEG-LS and interframe coding. IEEE Trans Inf Technol Biomed 13(5):818–821
4. Maglogiannis I, Doukas C, Kormentzas G (2009) Pliakas T (2009) Wavelet-based compression with ROI coding support for mobile access to DICOM images over heterogeneous radio networks. IEEE Trans Inf Technol Biomed 13(4):458–466
5. Ansari MA, Anand RS (2009) Context based medical image compression for ultrasound images with contextual set partitioning in hierarchical trees algorithm. Adv Eng Softw 40(7):487–496
6. Sharma N, Aggarwal LM (2010) Automated medical image segmentation techniques. J Med Phys 35(1):3
7. Valdes AL, Trujillo M (2013) Medical image compression based on region of interest and data elimination
8. Gokturk SB et al (2001) Medical image compression based on region of interest, with application to colon CT images. In: Proceedings of the 23rd annual international conference of the IEEE engineering in medicine and biology society, 2001, vol 3. IEEE
9. Kaur M, Wasson V (2015) ROI based medical image compression for telemedicine application. Procedia Comput Sci 70:579–585
10. Krishna V, Purna Chnder Rao V (2017) Region based medical image compression with Binary Plane Coding. ARPN J 12(17)

11. Ashok M, Bhaskar Reddy T (2012) Color image compression based on luminance and chrominance using Binary Wavelet Transform (BWT) and Binary Plane Technique (BPT). Int J Comput Sci Inf Technol Secur (IJCSITS) 1(2):2249–9555
12. Subhash Chandra N (2008) Loss less compression of images using binary plane, difference and Huffman coding (BDH technique). J Theor Appl Inf Technol 3(1):3–56
13. Ashok Babu P, Prasad KVSR (2012) A lossy color image compression using IWT and BPT. Graph Vis 12(15)
14. Santosh Kumar BP, Venkata Ramanaiah K (2019) Effective ROI extraction methods for hybrid medical image compression. IJATCSE 8(2)
15. Santosh Kumar BP, Venkata Ramanaiah K (2019) Medical image compression and region of interest extraction using deep learning methods. IJIES 5(4)
16. http://braintumorsegmentation.org/
17. Pennebaker WB, Mitchell JL (1993) JPEG: still image data compression standard. Van Nostrand Reinhold
18. Said A, Pearlman WA (1996) A new, fast, and efficient image codec based on set partitioning in hierarchical trees. IEEE Trans Circuits Syst Video Technol 6(3):243–250

A Comprehensive Study of Sentiment Analysis in Big Data Applications

Midde Venkateswarlu Naik, Mohd Dilshad Anasari, Vinit Kumar Gunjan and Santosh Kumar

Abstract Sentiment analysis [SA] is a computational analysis of sentiments or opinions, emotions, views, subjectivity expressed in text or associated with big data such as reviews, blogs, discussions, news, comments, feedback etc., about things such as electronic products, movies, public or private services, organizations, individuals, issues, events, topics, and their attributes. It represents a large problem space and this field has become a very active area. This survey paper deals with a comprehensive analysis of the research work carried out for the duration of 2010–2019 in the relevance of conventional methods, enhanced methods and associated techniques employed for the sentiment analysis research problems. Evidently, this article includes an analysis of the majority of the research work done by earlier researchers with respect to the related domain. However, we anticipate that the references mentioned will wrap up the most important theoretical aspects, and this review would be eventually helpful to subsequent researchers for upcoming research trends.

Keywords Big data · Sentiment analysis · Emotion extraction · Machine learning

M. V. Naik (✉) · M. D. Anasari
Department of Computer Science & Engineering, CMR College of Engineering & Technology, Hyderabad, India
e-mail: vvnaikcse@gmail.com

M. D. Anasari
e-mail: m.dilshadcse@gmail.com

V. K. Gunjan
Department of Computer Science and Engineering, CMR Institute of Technology, Hyderabad, India
e-mail: vinitkumargunjan@gmail.com

S. Kumar
Department of Electronics and Communication, Invettis University, Bareilly, India
e-mail: santu.thoughts@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_35

1 Introduction

In the information era, enormous data have become available on hand to decision-makers. Big data refers to datasets that are not only big but also high in volume, variety, and velocity, which makes it difficult to handle using traditional tools and techniques. Due to the rapid growth of such data, solutions, novel technical architectures, analytics, and innovative tools need to be studied and provided in order to handle and extract value and knowledge from these datasets. Furthermore, decision-makers need to be able to gain valuable insights from such varied and rapidly changing data, ranging from daily transactions to customer interactions and social network data and it has an extent from computer science to managing sciences [1] Such value can be provided using big data analytics, which is the application of advanced analytics techniques on big data. Apart from this, there are many applications have been produced outperformed results using machine learning and biological related techniques for various domains [2–9].

The most favorable solution for big data analytics is SA defined as the process of examining a text or speech to find the opinions, views, beliefs, convictions, and feelings the author or speaker has about a determined topic. Sentiment analysis at three-word phrases occurrence can also improve the intensity of the classifier accuracy [10]. It is tricky to understand the opinions of the people and summarizes the state about a specific product due to highly subjective and ambiguous concepts. This makes quit a challenge for a computer program to be extracted from the huge text as some words or the expression has different meanings depending on the knowledge domain. However, the treatment of a news article is different from that of a tweet. An article follows all grammatical rules of a language but tweet lacks in sentence structure, uses slang, emotional (☺, ☹), lengthened repeating words to increase emotional feelings, and due to these, the possibility of misspelling is higher. Due to these problems, automate the extraction of information with a computer system increases the complication, as the software programs eliminate require well-defined boundaries to eliminate ambiguity. As per the literature survey, sentiment analysis deals with different levels of the analyzed texts such as word or phrase, sentence, document level, and user level. Although, there are various approaches employed towards sentiment analysis, among them, the machine learning approach is the most appropriate one as per the survey. However, in SA text classification from the ambiguous text is the vital challenging task.

Consequently, the purpose of SA is to discover opinions, recognize the sentiments they express, then categorize their polarity as presented in Fig. 1. Three classification levels in SA, document level, sentence level, and aspect level. Document-level SA objective to categorize an opinion document as expressing positive or negative sentiment. It reflects the entire document as one topic. Sentence level SA purposes to classify sentiment uttered in each sentence. The primary stage is to recognize whether the sentence is subjective or objective. If the sentence is subjective, Sentence level SA will conclude whether the sentence expresses positive or negative sentiment.

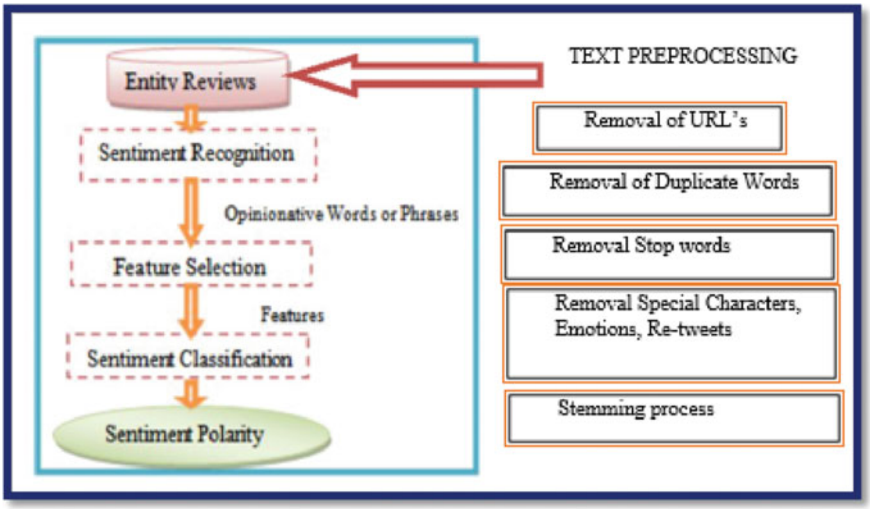


Fig. 1 Abstract sentiment analysis on entity reviews

There are many applications adopted to SA associated with big data [11] such as Face book, Google, LinkedIn, Netflix, Pandora, Twitter etc., Online publicity applications such as retail Sales which is employed using Salesforce.com, Picturing applications such as Tableau, QlikView, Commercial Cleverness Applications such as Operative Intelligence applications Splunk, Data as a provision applications such as FICO financial data, BlueKai PROVIDES Consumer profile, Data cleaning applications such as Trifacta, Data Privacy applications. This review goal to give enhancements and to consolidate and classify some papers offered in this area according to the numerous SA techniques.

Opinions are fundamental to practically all human activities since they are indispensable key influencers of our practices. At whatever point we have to make an end, we need to know others' suppositions. In the present world, business associations forever need to find a client or network suppositions about their items and administrations.

Singular buyers likewise need to know the sentiments of existing clients of an item before acquiring it, and further assessments about political people preceding making a casting a ballot determination in a political race. Previously, when an individual required conclusions, one can ask colleagues. At the point when an association required open or client conclusions, it directed an audit/study, assessment survey.

With the short-tempered expansion of social media comments, feedbacks, remarks, blogs, Twitter, Instagram, Facebook on the Web, persons and governments and private firms are gradually added and using the content in these media for decision determinations. Currently, if one wants to acquisition an end-user product, people will not ask any opinion to neighbors because there are numerous user reviews and deliberations in public discussions on the Web concerning the product. For corporate organizations, there is no need to conduct surveys, opinions in order to collect community opinions. Every site typically contains an enormous quantity of opinion text that is not always simply interpret in long blogs, forums. Usually, a person who reads relevant sites and extracting and concluding the opinions in them is so complex job. In such situations, computational sentiment analysis systems are necessary.

Because of these application areas, commercial activities have grown in current years. SA applications have prolonged to roughly every likely field, from customer products, economic services to social events and political elections, services, healthcare, etc… These truthful applications and trade welfares have provided motivation for investigation in sentiment analysis.

Along with real-time applications, numerous application-based research articles have been published [12], a sentiment model was projected to forecast sales performance, and analysis was utilized rank goods. The Twitter opinion was related with public and private opinion surveys, the social media sentiment was also employed to prediction election results and he proposed, a technique was reported for forecast political blogs and various websites would compute movie reviews forecast box office revenue for movies. Apart from these, Fig. 2 demonstrates classification techniques employed for sentiment analysis research problems until 2014.

The rest of the paper organized as the following style. Section 2 gives survey data about methodology and published articles and in respective tasks associated with data resources information. Section 3 describes the need and technical information about sentiment analysis. Section 4 explains the importance of big data. Section 5 gives emphasis to the benefits of big data strategy and possible obstacles in big data implementation. Section 6 describes discussion and analysis as a result of the survey.

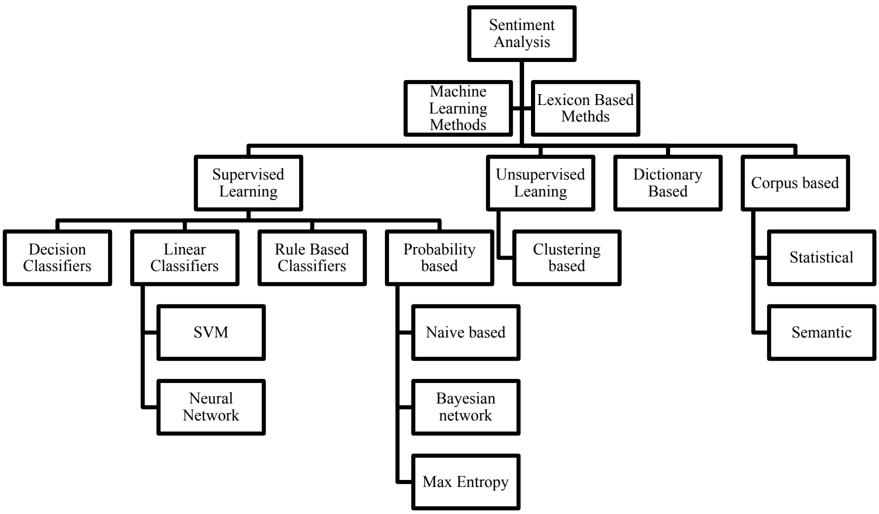


Fig. 2 Taxonomy of various methods and techniques employed on SC

Finally, Sect. 7 gives information about the conclusion and future scope of sentiment analysis.

2 Methodology

The proposed work presented papers in the survey are summarized in Table 1. The

Table 1 Techniques and datasets employed for SA

Reference/year	Technique/methods	Dataset	Language	Task
[13] 2019	NLP techniques,	NLPCC2014	Chinese	Word Embedding in SA
[14–32] 2010–2014 [33] 2019	ML techniques, Rule-based, LSA, CRF, SVM, Chi-square, Statistical, Context NLP, Lexicon based, NLP, SVM and K-nearest Neighbor, KNIME, HDP-LDA, Prediction Algorithm, Dictionary ML, Bayesian Ensemble,, survey, senticNet, SVM and PCA and Max entropy, Fuzzy Set	automativeforum.com, CNETD, Mobile Customer Review, Camera Reviews, Education, epinions.com, Book-reviews, Restaurant, News, amazon, Media Analysis, Tripadvisor.com hotel reviews, Restaurant, Twitter, Geographic region, Cornell Movie, NOAA Climate, elections, facebook	English	SA for classifier accuracy enhancement
[34] 2019	AI Techniques	Dravidian	Dravidian	SA in Dravidian Languages
[10] 2018	Unsupervised learning	Carnel Movie	English	Discovery of SA Phrases
[35] 2017	Using Supervised Sentiwordnet	News Articles	Telugu	SA with 74% accuracy and SC with 81%
[36–39] 2010–2014	CSETR, PMI, S-HAL and SO-PMI, SO-SD, Word2Vec	Election, Product Reviews, News	Chinese	SA on Web blog or social network
[40] 2012	Semantic Lexicon	Dutch WordNet	Dutch	SA analysis for classifier enhance

intentions of the papers are specified in the last column. They are divided into six classes which are *Sentiment Analysis (SA), Sentiment Classification (SC), Emotion Detection (ED), Feature Selection (ES)*, *Building Resource (BR)*, *Cross-language (SLSA)* and Multilingual SA (MSA). The BR groups can be categorized to Dictionary, lexica, Corpora. The current work divided the articles that resolve SC. Rest of the papers that resolve SA. The papers feature selection groups as FS. The articles characterized to represent the SA connected fields such as ED, BR and Transfer Learning (TL).

Table 1 emphasizes information about Sentiment analysis task using various machine learning techniques and NLP techniques with the help of various datasets on distinguishing languages.

Table 2 emphasizes the summarized information about sentiment classification (SC) tasks resolved with various levels and scientific techniques such as NLP, Machine learning and optimization techniques. Table 2 give information about on which dataset and language the sentiment classification task employed with obtaining several accuracies. Usually, Sentiment classification is the process of determination of the orientation of sentimentality of specified text in two or more classes. SC has been done in numerous classes such as binary, ternary, etc…

Table 2 List of techniques employed for SC

Reference/year	Technique/methods	Dataset	Language	Task
[41] 2018	Linear SVM and clustering	Unlabeled IMDB	English	Sentiment classification
[42] 2018	ML classifiers using n-gram technique	Twitter	English	Sentiment classification
[43–51] 2010–2014	SVM and NB and ME, Semi supervised classification, Lexicon based, Statistical (MM), NB and SVM, Semi-supervised, NB and Rule-based, NB, SVM and C4.5	Movie and news, Multi-domain sentiment reviews, Movie, amazon.com, Restaurant reviews, artificial data, 2 sided debates, Movie, film reviews	English	Sentiment classification accuracy Enhancement
[52, 53] 2010 [54] 2011	Graph approach and SVM and NB, CRF	Chinese sentiment data, computer review	Chinese	Sentiment classification accuracy Enhancement
[51] 2013	SVM, NB, C4.5	Film review-data	Spanish	Sentiment classification accuracy Enhancement

Tables 3, 4, 5, 6 and 7 emphasizes the summarized information about ED, FS, BR, CLSA, MLSA tasks resolved with various levels and scientific techniques such as NLP, Machine learning and optimization techniques. They gives information about on which dataset and language the sentiment classification task employed with obtaining several accuracies.

Table 3 List of techniques employed for Emotion Detection (ED)

Reference/year	Technique/methods	Dataset	Language	Task
[55–57] 2010	Rule based method and semantic label, Lexicon based	Health data, Experience project,	English	Emotion detection for sentiment analysis
[58, 59] 2013	SVM and Lexicon based, SVM	ISEAR Data, Weibo crawled	English	Emotion detection for sentiment analysis
[57] 2012	Blogs	Live journals	Spanish	Emotion detection for sentiment
[60] 2019	Survey	Twitter	English	Emotion detection for sentiment analysis
[61] 2019	Dynamic models	Mobile data	English	Emotion detection for sentiment analysis
[62] 2019	Bi-directional LSTM	Web dialogs	English	Emotion detection

Table 4 List of techniques employed for Feature Selection (FS)

Reference/year	Technique/methods	Dataset	Language	Task
[63] 2012	HMM LDA Statistical technique	Movie	English	Features extraction for sentiment analysis
[64] 2018	Chi-square, information gain, DF	Movie reviews	English, Turkish	Feature extraction
[65] 2013	BMNB	Book, DVD reviews	English	Feature extraction
[66] 2010	Document frequency difference, NLP based techniques	Web reviews	English	Feature extraction

Table 5 List of techniques employed for Building Resources (BR)

Reference/year	Technique/methods	Dataset	Language	Task
[67] 2011	Random-walk	Chinese Hotel review text	Chinese	To build resource
[68] 2012	Semantic approach	Restaurant-reviews	Spanish	To build resource
[69] 2012	Natural Language Processing	ISEAR text	Italian	To build resource
[70] 2012	Triangulation tech	Opinion dictionary	Arabic	To build resource
[71] 2018	NLP, Sarcasm detection	Greek reviews	Greek	To build resource
[72] 2016	Normalized chi-square based feature weight generation	SentiWordnet	English	To build resources
[73] 2017	Corpora, lexicons in Arabic	Dialectal Arabic	Arabic	To build resources

Table 6 List of techniques employed for Cross-Language SA

Reference/year	Technique/methods	Dataset	Language	Task
[74] 2011	Entropy based technique	Chinese domain text	Chinese	Cross language SA
[75] 2011	Ranking based technique	Restaurant review text	Chinese	Cross language SA
[76] 2011	SAR Technique	MPQA and RIMDB text	Chinese	Cross language SA
[77] 2011	Shared-learning approach	Flicker and CNN BBC text	English	Cross language SA
[78] 2017	English Spanish vector space model	Non English context review	English, Spanish	Cross language SA
[79] 2017	Machine translation	English, Arabic reviews	English, Arabic	Cross language SA

3 Inspiration: Need for SA

The passion for sentiment analysis has not just been users necessitate driven but also it has been expanded to the market needs. This would be a field in which we found applications in the user domain and vendor domain. The following information describes how sentiment analysis essential for the users and manufacturers perspective.

Table 7 List of techniques employed for Multilingual SA

Reference/year	Technique/methods	Dataset	Language	Task
[80] 2018	SVM and CRM, LR	NTCIR-6	English, Japanese MPQA	SA using multilingual
[81] 2019	Information gain, SVM with sentence level	Cell phone reviews in English, Chinese	Chinese, English	SA using multilingual
[82] 2014	NB, ME, SMO-SVM	Turkish multi-domain product reviews	Turkish	SA using multilingual
[83] 2010	Hybrid meta classifier	Spanish Muchocine (MC)	Spanish	SA using multilingual

3.1 A User's and Vender's Angle

An intellectual person usually learns some things from other errors. Consequently, what other customers regarding an entity (product) carry a huge influence on one's decision about any product (entity). In big data environments, it is very difficult for any person to actually go through this huge amount of information (unstructured, semi-structured, structured) in order to extract a useful knowledge (opinion) regarding a product. In this situation, for even easy queries, the current system retrieves millions of results on current search engines. Therefore, according to the user's perspective, it needs a system that would automatically go through such social media reviews (big data) and conclusion with a useful opinion regarding an entity or organization being evaluated.

3.1.1 Applications of SA in Big Data Environment

Of late, SA has found applications in a variety of domains. The major applications such as review related applications, business development applications in various commercial and public domains. Following are the comprehensive list of the most important areas.

- Big Data Applications: Google, LinkedIn, Netflix, Twitter, Instagram
- Online publicizing applications
- E-commerce applications: salesforce.com
- Imagining applications: Tableau, QlikView
- Commercial Intelligence-applications
- Operative intellect applications: Splunk
- Data as a facility applications: FICO, BlueKai, INRIX, LexixNexis
- Data cleansing applications
- Data privacy applications.

3.1.2 Issues in Sentiment Analysis

The issues in SA would be separated based on various criteria. The first way is to distinguish based on the task whether it is a classification process or scoring or ranking process. The second way to distinguish based on the granularity level at which the classification takes place. Though, there may exit various ways to define research problems in SA, mainly we have emphasized on the two distinguished ways of SA tasks.

Classification Based Tasks

- Recognition text subjective or objective polarity
- Recognition text positive or negative polarity
- Recognition text positive or negative or objective
- Recognition score or ranking of text positive or negative polarity.

Level Based Sentiment Analysis

- Word level analysis
- Phrase level analysis
- Sentence level analysis
- Document-level analysis.

Complex Issues in SA

- Aspect sentiment study
- Lexicon based sentiment generation
- Extraction of comparison opinions on the product
- Co-reference resolution at semantic level
- Opinion spam detection,
- Cross field SC
- Multilingual SC
- Entity recognition
- Identifying opinion holders
- Contextual-polarity disambiguation.

4 Significance of Big Data

To recognize the importance of the big data, it is essential to know (i) major sources for the big data for the analysis (ii) key advantages of adopting big data approach for the analysis (iii) key technical problems to implementing big data approach for decision-maker for an analysis. According to the survey [84] following is the information extracted as a literature survey for during 2015 in the significance of big data.

4.1 Key Sources for Big Data

In Fig. 3 shown various percentages of the data source as part of the big data. However, semi-structured and unstructured data has become 58% as part of big data. To conduct sentiment analysis on such data with a mix of distinguishing languages is a key challenge. In this regard, the category of data captured would depend on the business organization or related entity. Therefore, as per the survey, there is a scope to perform research on unstructured data.

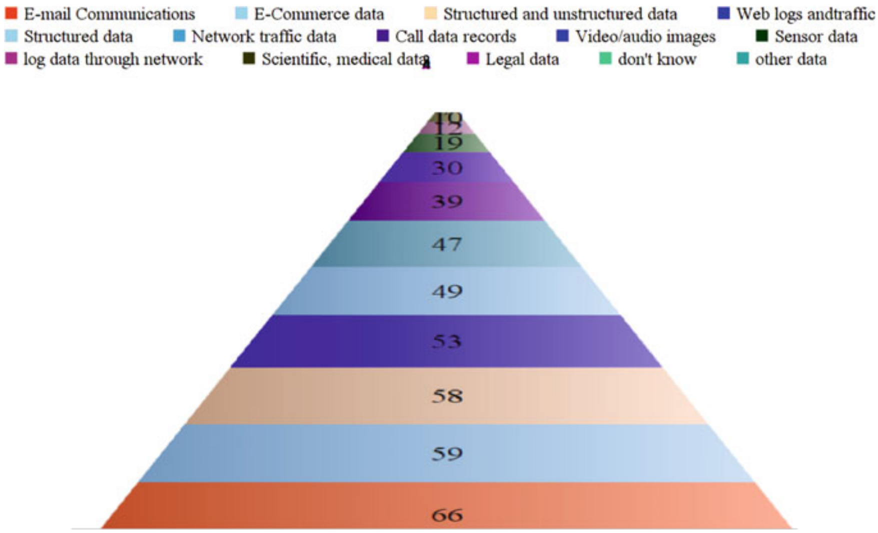


Fig. 3 Major sources of big data from several areas

Table 8 Merits of SA to adopt big data strategy

Improved Decision making	52%
Better Operational Efficiency	42%
Better data quality	40%
Enhanced customer intelligence	37%
Identify new trends/ opportunities	37%
Enhanced overall business insight	36%
Improved customer service	31%
Quicker queries and analysis	29%
Reduced Duplication Information	28%
Reduced Operating Costs	25%
Improved Risk Management	20%
Enhanced Sales Revenue	17%
Optimized marketing and branding	16%
Better Compliance	14%
Better Collaboration with business partners	13%

5 Benefits of Sentiment Analysis in Big Data Strategy

In Table 8 listed major advantages by adopting sentiment analysis for big data analytics [84] in distinguish domains and associated applications. Although, listed benefits applicable to sentiment analysis with respect to users and the organization's perspective.

According to the survey, in Table 8 decisions making would be improved 52% to the organization by adopting sentiment analysis as a big data strategy. 42% obtained operational efficiency. Similarly, the rest of the factors would helpful for an organization by adopting sentiment analysis in big data scenarios.

According to the survey [32], the variety of data structures are occurred by 51% and linking unstructured and structured data by 40% as technical obstacles in big data strategy. This obstacle addresses the process of converting structured data from unstructured data. Subsequently, this task would come under sentiment analysis on cross-language or independent domains. Although, rest of the issues also yet to be resolved and they have shown the percentage of effect as difficulty to be resolve in big data strategy (Table 9).

6 Noticeable Analysis and Discussions

Tables 1, 2, 3, 4, 5, 6 and 7 presented information about the earlier authors carried works in the seven categories of sentiment extraction or determination tasks. Figure 4 visualizes summarization of articles published year wise to demonstrates the progress

Table 9 Technical issues appear in big data due to following reasons

Large kind of data structures	51%
Association of structured with unstructured data	40%
Determination of useful and discard data	37%
Privacy of data issue	31%
Huge amount of files to be incorporate	25%
Velocity of changes in the data to be analyze	21%
Communicating high volume of data through network	19%
Authentication and authorization data concern	18%
Lack of hardware to accommodate huge info	16%
Computational problems with existing architectures	13%
Backup storage issue	10%
Lack of technology transformations	4%

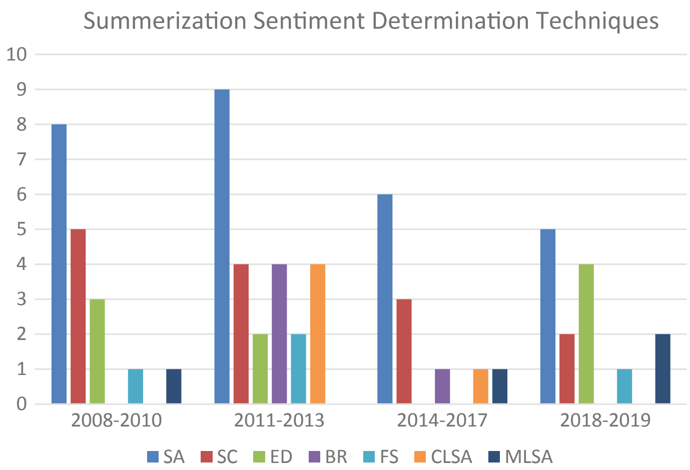


Fig. 4 Sum of articles published with 7 categories of tasks from 2008 to 2019

of work carried out by previous authors for sentiment extraction task in various domain.

Figure 4 demonstrates the number of the papers which gives impact about Sentiment sub tasks according to the categories listed. This categories of tasks have been calculated according to the year of publication. Here proposed survey have been noticed that Emotion detection, Sentiment analysis in cross domain, resource building is still there would be huge scope when we consider local languages and Dravidian languages.

In this survey most of the machine learning techniques and hybrid techniques employed for sentiment analysis and classification very precisely and accurately, rest of the SA sub tasks need to be resolved with innovative algorithms to produce best results depends on the language concern challenges.

Figure 5 demonstrates about datasets and corpus used for several task to be resolved as part of sentiment analysis domain as part of big data concern. As per the survey data sets have been identified as reviews, web blogs, news articles, social media, and other datasets. The pictorial graph shown in Fig. 5 would give clear picture about most frequently used datasets by most of the authors earlier used in their works for sentiment analysis.

Figure 6 shows the employed techniques or algorithms or methods year wise in which classified four types of approaches. As shown in the graph year wise usage of the techniques haven changing drastically due to type of problems to tackle and

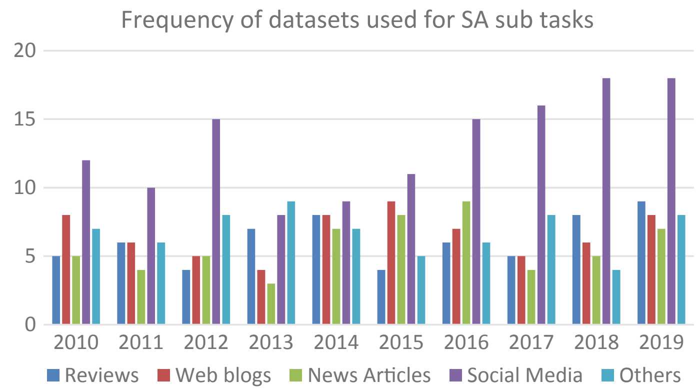


Fig. 5 Sum of distinguished data resources used for SA tasks

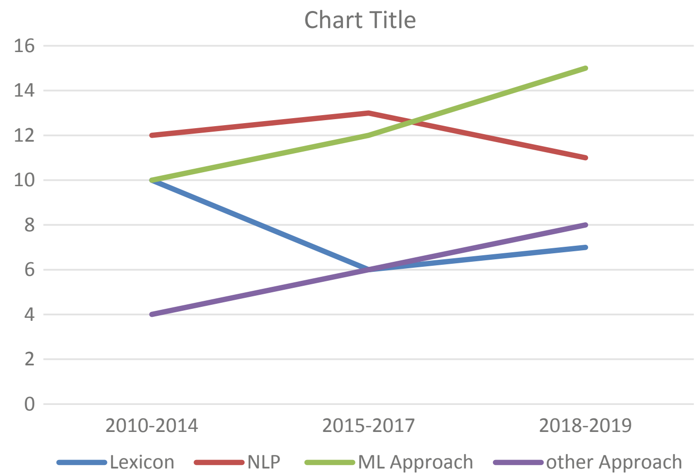


Fig. 6 Consolidated techniques employed according to category wise and year wise

resolved. For sentiment analysis and classification most of the authors used Lexicon based or corpus based or ML techniques used for solving certain level of works. For few works NLP techniques have been used for resolving certain tasks as part of the sentiment analysis works depends on level of granularity levels. Apart from these advanced approaches such as fuzzy set, bio logical techniques and hybrid approaches can also outperform the results which were produced by earlier techniques.

7 Conclusion and Future Scope

The current literature survey paper presented summarized information about various algorithms to analyzing various levels of sentiment in several languages. The referred articles have contributed most of the research issues related to sentiment analysis at various granularity levels and in several languages. After all instigation about status of the sentiment related sub tasks such as SA, SC, ED, BR, MSA, CSA have scope to enhance existing classifier accuracies.

The focus on sentiment analysis has been done at all levels in English language and Chinese languages as all resources are available to the researchers. Although building required resources at various granularity levels of sentiment analysis in Dravidian languages using natural languages yet to be enhanced.

Therefore, it is concluding the applying enhanced machine learning approach, evolutionary computing approach, expert system approach and fuzzy logic set approach for current research would enhance the overall accuracy of the sentiment analysis over various environments.

Acknowledgements I would like thank to the CMR College of Engineering & Technology for their continues support.

References

1. Chen Y, Xie J (2008) Online consumer review: word-of-mouth as a new element of marketing communication mix. Manage Sci 54(3):477–491
2. Naik MV, Mohanty R (2014) An expert system approach for legal reasoning in acquiring immovable property. IEEE conference ICNSC, pp 370–374
3. Naik MV, Reddy SS (2017) An innovative optimized model to anticipate clients about immigration in telecom industry. In: IEEE conference ICATCCT, pp 232–236
4. Narsimlu M, Naik MV, Sree MR (2017) An innovative research on augmented reality to control status of physical light using MAYA as part of IOT. In: IEEE conference SmartTechCon, pp 1221–1227
5. Naik MV, Lokhanday S (2012) Building a legal expert system for legal reasoning in specific domain—a survey. Int J Comput Sci Inf Technol 4(5):175
6. Guda V, Srujana I, Naik MV (2011) Reasoning in legal text documents with extracted event information. Int J Comput Appl 28(7):8–13

7. Rao VN, Naik MV. An enhanced model to extract expected answer for temporal questions over question answering system. IJIRTS 2(5):27–34
8. Rani KJ, Naik MV. Machine learning intelligent techniques and bio inspired computing techniques applications for software engineering problems—a survey. IJAERSPL3 10(81):130–152 (republication)
9. Zhang W, Zhang Y, Yang K (2019) Optimizing word embedding for fine-grained sentiment analysis. In: ICAIS. Springer, pp 276– 286
10. Mohanty R, Naik V, Mubeen A (2014) Software reliability prediction using ant colony optimization technique. In: IEEE conference ICCSNT, pp 496–500
11. Feinleib D (2014) Big data bootcamp. www.it-books.info
12. Liu B (2012) Sentiment analysis and opinion mining. In: Hand book. Morgan & Claypool Publishers (May)
13. Dolianiti FS et al (2018) Sentiment analysis and applications in education: a survey. In: TECH-EDU. Springer, pp 413–427
14. Qiu G, He X, Zhang F, Shi Y, Bu J, Chen C (2010) DASA: dissatisfaction-oriented advertising based on sentiment analysis. Expert Syst Appl 37:6182–6191
15. Cao Q, Duan W, Gan Q (2011) Exploring determinants of voting for the "helpfulness" of online user reviews: a text mining approach. Decis Support Syst 50:511–521
16. Xu K, Liao SS, Li J, Song Y (2011) Mining comparative opinions from customer reviews for competitive intelligence. Decis Support Syst 50:743–754
17. Chien CC, Tseng Y-D (2011) Quality evaluation of product reviews using an information quality framework. Decis Support Syst 50:755–768
18. Fan T-K, Chang C-H (2011) Blogger-centric contextual advertising. Expert Syst Appl 38:1777–1788
19. Gupta SK, Phung D, Adams B, Venkatesh S (2012) Regularized non negative shared subspace learning. Data Min Knowl Disc 26:57–97
20. Kang H, Yoo SJ, Han D (2012) Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews. Expert Syst Appl 39:6000–6010
21. Zhang W, Xu H, Wan W (2012) Weakness finder: find product weakness from Chinese reviews by using aspects based sentiment analysis. Expert Syst Appl 39:10283–10291
22. Ortigosa-Hernández J, Rodríguez JD, Alzate L, Lucania M, Inza I, Lozano Jose A (2012) Approaching sentiment analysis by using semi-supervised learning of multi-dimensional classifiers. Neurocomputing 92:98–115
23. Lui B, Blasch E, Chen Y, Shen D, Chen G (2013) Scalable sentiment classification for big data analysis using Naïve Bayes classifier. In: IEEE international conference on big data, pp 99–104
24. Ding W, Song X, Guo L, Xiong Z, Hu X (2013) A novel hybrid HDP-LDA model for sentiment analysis. In: IEEE/WIC/ACM international conference on web intelligence (WI) and intelligent agent technology (IAT), pp 329–336
25. Makrehchi M, Shah S, Lia WH (2013) Stock prediction using event-based sentiment analysis. In: IEEE/WIC/ACM international conference on web intelligence (WI) and intelligent agent technology (IAT), pp 337–342
26. Nguyen VD, Varghese B, Barker A (2013) The royal birth of 2013: analysing and visualising public sentiment in the UK using Twitter. In: IEEE international conference on big data, pp 46–54
27. Huang S, Han W, Wue X, Wang W (2013) Polarity identification of sentiment words based on emotions. In: 9th International conference on computational Intelligence and security, pp 134–138
28. Gandomi A, Haider M (2015) Beyond the hype: big data concepts, methods, and analytics. Int J Inf Manage 35:137–144
29. Weichselbraun A, Gindl S, Scharl A (2014) Enriching semantic knowledge bases for opinion mining in big data applications. Knowl. Based Syst. 69:78–85
30. Anjaria M, Guddeti RMR (2014) Influence factor-based opinion mining of twitter data using supervised learning. IEEE

31. Xue B, Fu C, Shaobin Z (2014) A study on sentiment computing and classification of Sina Weibo with Word2ve. In: IEEE international congress on big data, pp 358–363
32. www.talend.com, talend-big-data, research paper on big data (2015)
33. Rani S, Kumar P (2019) A journey of Indian languages over sentiment analysis: a systematic review. Artif Intell Rev 52(2):1415–1462
34. Naidu R, Bharti SK, Babu KS (2017) Sentiment analysis using Telugu Sentiwordnet. In: IEEE WiSPNET. https://doi.org/10.1109/WiSPNET.2017.8299844
35. Mudunas A, Zhang D, Levene M (2018) Bootstrap Domain specific sentiment classifiers from unlabeled corpora. Trans Assoc Comput Linguistics 6:269–285
36. Yang H, Yang B, Zhou X, Zhou C, Chai Z (2010) Community discovery and sentiment mining for Chinese BLOG. In: IEEE seventh international conference on fuzzy systems and knowledge discovery (FSKD), pp 1740–1745
37. Min H-J, Park JC (2012) Identifying helpful reviews based on customer's mentions about experiences. Expert Syst Appl 39:11830–11838
38. Steinberger J, Ebrahim M, Ehrmann M, Hurri-yetoglu A, Kabadjov M, Lenkova P, Steinberger R, Tanev H, Vázquez S, Zavarella V (2012) Creating sentiment dictionaries via triangulation. Decis Support Syst 53:689–694
39. Mukkamala RR, Hussain A, Vatrapu R (2014) Fuzzzy–set based sentiment analysis of big social data. In: IEEE 18th international enterprise distributed object computing conference, pp 71–80
40. Lane PCR, Clarke D, Hender P (2012) On developing robust models for Favorability analysis: model choice, feature sets and imbalanced data. Decis Support Syst 53:712–718
41. Naz S, Sharan A, Malik N (2018) Sentiment classification on Twitter data using Support vector machine. In: IEEE international conference on web intelligence. https://doi.org/10.1109/WI.2018.00-13
42. Kim E, Klinger R (2018) A survey on sentiment and emotion analysis for computational literary studies. arXiv:1808.03137v1 [cs.CL] (9 Aug 2018)
43. Bai X (2011) Predicting consumer sentiments from online text. Decis Support Syst 50:732–742
44. He Y, Zhou D (2011) Self-training from labeled features for sentiment analysis. Inf Process Manage 47:606–616
45. Heerschop B, Goossen F, Hogenboom A, Frasincar F, Kaymak U, de Jong F (2011) Polarity analysis of texts using discourse structure. In: Presented at the 20th ACM conference on information and knowledge management (CIKM'11)
46. Zirn C, Niepert M, Stuckenschmidt H, Strube M (2011) Fine-grained sentiment analysis with structural features. In: Presented at the 5th international joint conference on natural language processing (IJCNLP'11); 2011
47. Moreo A, Romero M, Castro JL, Zurita JM (2012) Lexicon-based comments-oriented news sentiment analyzer system. Expert Syst Appl 39(91):66–80
48. Boldrini E, Balahur A, Martínez-Barco P, Montoyo A (2012) Using Emotional Blog to annotate and analyze subjectivity in the new textual genres. Data Min Knowl Disc 25:603–634
49. Balahur A, Hermida JM, Montoyo A (2012) Detecting implicit expressions of emotion in text: a comparative analysis. Decis Support Syst 53:742–753
50. Li Y, Jain A (1998) Classification of text documents. Comput J 41:537–546
51. Banic L, Mihanovic A, Brakus M (2013) Using big data and sentiment analysis in product evaluation. In: MIPRO, pp 1149–1154
52. Zhao Y-Y, Qin B, Liu T (2010) Integrating intra- and inter-document evidences for improving sentence sentiment classification. Acta Autom Sin 36:1417–1425
53. Hu Y, Li W (2011) Document sentiment classification by exploring description model of topical terms. Comput Speech Lang 25:386–403
54. Jiao J, Zhou Y (2011) Sentiment Polarity Analysis based multi-dictionary. In: Presented at the 2011 international conference on physics science and technology (ICPST'11)
55. Lu C-Y, Lin S-H, Liu J-C, Cruz-Lara S, Hong J-S (2010) Automatic event-level textual emotion-sensing using mutual action histogram between entities. Expert Syst Appl 37:1643–1653

56. Neviarouskaya A, Prendinger H, Ishizuka M (2010) Recognition of affect, judgment, and appreciation in text. In: Proceedings of the 23rd international conference on computational linguistics (Coling 2010), Beijing, pp 806–14
57. Xu T, Peng Q, Cheng Y (2012) Identifying the semantic orientation of terms using S-HAL for sentiment analysis. Knowl Based Syst 35:279–289
58. Maks I, Vossen P (2012) A lexicon model for deep sentiment analysis and opinion mining applications. Decis Support Syst 53:680–688
59. Fersini E, Messina E, Pozzi FA (2014) Sentiment analysis: Bayesian ensemble learning. Decis Support Syst 68:26–38
60. Dupré D, Bland B, Bolster A et al (2017) Dynamic model of athletes emotions based on wearable devices. Springer International Publishing, Cham, pp 42–50
61. Chatterjee A, Narahari KN, Joshi M, Agarwal P (2019) SemEval-2019 task 3: Emocontext Contextual Emotion detection in text. In: 13th International workshop on sematic evaluation (SemEval-2019), pp 39–48
62. Parlar T, Ozel SA, Song F (2018) QER: a new feature selection method for sentiment analysis. Hum Centric Comput Inf Sci 8:10
63. Robaldo L, Di Caro L (2012) Opinion Mining-ML. Comput Stand Interfaces
64. Agarwal B, Mittal N (2013) Optimal feature selection for sentiment analysis. In: CICLing. Springer, pp 13–24
65. Nicholls C, Song F (2010) Comparison of feature selection methods for sentiment analysis. In: Canadian AI. Springer, pp 286–289
66. Tsakalidis A et al (2018) Building and evaluating resources for sentiment analysis in the Greek language. Lang Resour Eval 52(4):1021–1044
67. Tan S, Wu Q (2011) A random walk algorithm for automatic construction of domain-oriented sentiment lexicon. Expert Syst Appl 120:94–100
68. Di Caro L, Matteo G (2012) Sentiment analysis via dependency parsing. Comput Stand Interfaces
69. Keshtkar F, Inkpen D (2012) A bootstrapping method for extracting paraphrases of emotion expressions from texts. Comput Intell
70. Walker MA, Anand P, Abbott R, Fox Tree JE, Martell C, King J (2012) That is your evidence?: Classifying stance in online political debate. Decis Support Syst 53:719–729
71. Khan FH, Qamar U, Bashir S (2016) Senti-CS: building a lexical resource for sentiment analysis using subjective feature selection and normalized Chi-Squared feature weight generation. Expert Syst. 33(5):489–500
72. Itani M, Roast C, Al-Khyatt S (2017) Developing resources for sentiment analysis of informal Arabic text in social media. Procedia Comput Sci 117:129–136
73. Abdalla M, Hirst G (2017) Cross lingual sentiment analysis without good translation. In: AFNLP, pp 506–515
74. Tan S, Wang Y (2011) Weighted SCL model for adaptation of sentiment classification. Expert Syst Appl 38:10524–10531
75. Wu Q, Tan S (2011) A two-stage framework for cross-domain sentiment classification. Expert Syst Appl 38:14269–14275
76. Lambov D, Pais S, Dias G (2011) Merged agreement algorithms for domain independent sentiment analysis. In: Presented at the Pacific association for computational linguistics (PACLING'11)
77. Duric A, Song F (2012) Feature selection for sentiment analysis based on content and syntax models. Decis Support Syst 53:704–711
78. Al-Shabi A et al (2017) Cross lingual sentiment classification from English to Arabic using machine translation. IJACSA 8(12):434–440
79. Seki Y, Kando N, Aono M (2009) Multilingual opinion holder identification using author and authority view points. Inf Process Manage 45:189–199
80. Wang H, Yin P, Zheng L, Liu JNK (2013) Sentiment classification of online reviews: using sentence-based language model. J Exp Theor Artif Intell 26:13–31. https://doi.org/10.1080/0952813x.2013.782352

81. Demirtas E (2013) Cross lingual sentiment analysis with machine translation, utility of training corpora and sentiment lexica. Master thesis, University of Technology
82. Martin-valdivia MT et al (2013) Sentiment polarity detection in Spanish reviews combining supervised and unsupervised approaches. Expert Syst Appl 40:3934–3942
83. Khan M, Ansari MD (2019) Security and privacy issue of big data over the cloud computing: a comprehensive analysis. Int J Recent Technol Eng (IJRTE) 7(6S):413–417
84. Tsytsarau M, Palpanas T (2012) Survey on mining subjective data on the web. Data Min Knowl Disc 24:478–514

A Comparative Case Study on Machine Learning Based Multi-biometric Systems

Santosh Kumar, Mohd Dilshad Ansari, Midde Venkateswarlu Naik, Vijender Kumar Solanki and Vinit Kumar Gunjan

Abstract Biometric has been an extensive method for years, which allow the comprehensive recognition and identification of an individual on the basis of various biometric traits such as face, fingerprint, iris, and retina and gait etc. although these biological and behavioral information of an individual is unique to oneself yet they failed to protect against the intruder attack sometimes. A Multibiometric system which can be formed by clubbing two or more traits has shown the better result in developing a more secure and reliable authentication system. In this paper we have presented a comparative study of various machine learning based unimodal and multimodal biometric systems.

Keywords Unimodal biometrics · Multimodal biometric · Iris · Face · Fingerprint · Gait · Machine learning

Nomenclature

AE Auto encoder

S. Kumar
Department of Electronics and Communication Engineering, Invertis University, Bareilly, India
e-mail: santu.thoughts@gmail.com

M. D. Ansari (✉) · M. V. Naik
Department of Computer Science & Engineering, CMR College of Engineering & Technology, Hyderabad, India
e-mail: m.dilshadcse@gmail.com

M. V. Naik
e-mail: vvnaikcse@gmail.com

V. K. Solanki · V. K. Gunjan
Department of Computer Science and Engineering, CMR Institute of Technology, Hyderabad, India
e-mail: spesinfo@yahoo.com

V. K. Gunjan
e-mail: vinitkumargunjan@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_36

ANN	Artificial neural network
BM	Boltzmann machine
BPNN	Back propagation neural network
CCA	Canonical Correlation Analysis
CFB	Correlation Filter Bank
CNN	Convolution neural network
EER	Equal error rate
ELM	Extreme learning machine
ESFI	Enhanced side face image
FAR	False acceptance rate
FFNN	Feed forward neural network
FRR	False rejection rate
FTC	Failure to Capture
FTE	Failure to enroll
GAE	Gait energy image
GAR	Genuine acceptance rate
GRU	Gated recurrent unit
HMM	Hidden Markov model
HN	Hopfield network
ICP	Iterative Closet Point
KN	Kohonen network
LDA	Linear Discriminant analysis
MDHND	Multimodal deep hashing neural decoder
MLP	Multi layer perceptron
NIR	Near Infrared region
NN	Neural network
PCA	Principle component analysis
PSO	Particle Swarm Optimization
QG-MSVM	Q-Gaussian multi support vector machine
RBF	Radial basis network
RNN	Recurrent neural network
SLP	Single layer perceptron
SVM	Support vector machine

1 Introduction

The first and the foremost step in protecting a highly confidential network from the intruders are to secure the network it with behavioral or physiological characteristics of the passersby [1]. Generally, it takes any one of the biometric identity such as face, finger, iris, retina, signature, gait, hand geometry or palm prints in single/unimodal biometric recognition system which first enroll the users in advance and testify them during security checks [2]. There are many reasons that have made biometric security a better option over manual checks as these system permits a higher degree of universality which ranges up to 98%, long term permanence as fingerprint and iris patterns remain same even after certain cut and surgeries and individuality based on the differences in biological and behavioral characteristics [3, 4]. Biometric has emerged as inseparable part of all modern systems which requires moderate to high security in applications such as cell phone, workplace, and confidential areas like military base and nuclear plants. The state of the art also shows that most of the electronic and commercial products are equipped with artificially intelligent biometrically secures systems.

The essential parameter in the implementation of a biometric system is its performance, acceptability, and circumvention. However these parameters in a unimodal biometric system are affected by several limitations such as presence of noise in data collection at sensors, difficulties in grabbing the identity from some users (non-universality), problem of individuality (similar faces of twins) and lack of invariant representation etc. [5, 6].

Therefore, unimodal biometric systems suffer to fulfill the requirement in commercial and security environment, and hence they may not be able to achieve a high degree of accuracy. However if we mix two or more biometrics modalities then we can overcome the issues raised in unimodal biometric systems [7].

2 Hindrances of Unimodal Biometric Systems

In every biometric security system, role of sensors is not limited to successfully acquire the data, it is also important that it should be collected noise free. However, it is quite difficult to read input at sensor end without noise. There may be several reasons which create hindrances in noise free data collection.

One of the instances that commonly observed for inclusion of noise along with data is the presence defective acquisition conditions. Improper Illumination conditions, dirt on lens or improperly maintained camera may lead to capture distorted or noisy face image. Same is the case with fingerprint as if any residue remains on the sensor it may affect the next sample collected without cleaning it [8]. On the other hand, not every person in this world is eligible to register or enroll oneself on the basis of who he is. It means that he is one who cannot present any biometric clue for authentication purpose. It raises the issue of non-universality in the field of biometric recognition [9, 10].

It is believed, that every biometric trait has a significant variation in every individuals, but when the feature vectors are constructed out of it, similarities may exist in the feature set. Therefore, the degree of freedom provided by the biometric evidences, are limited by the problem of homogeneity among the feature vector sets. Study shows that, the discrimination capability while using the fingerprints as biometric trait is approximately 10^4–10^5, it is also consider as keyspace. In addition to it, the keyspace for a randomly set password of 8-character is about 6.6×10^5 [11].

The concept of discrimination is not only limited to the fingerprints, but to every biometric trait.

The accuracy in biometric systems is affected, if the system is not able to discriminate between two biometric evidences of different persons. This problem occurs due to genital factor, for example, father and son or twin brothers can have the same facial expression. The problem of lack of individuality may help in increasing the false acceptance rate and low inter-class variations means that biometric evidences of different individuals may appear to be similar [12]. Sometimes, there exists a problem of lack of homogeneity when the biometric data acquired for the template generation is not similar to the data acquired for verification. This problem is also known as intra class variation, and generally it is occurred due to the improper use of the sensors by users during enrollment or verification.

Last and the foremost issue with biometric security occur when they are tested against the spoof attacks by the intruders. Although Biometric systems are made to provide a barricade against the imposter attacks, but some time they fail to distinguish between intruders and genuine users. Therefore to alleviate the effects of the unimodal biometric systems, a consolidated multimodal system can be made by adjoining two or more sources of biometric traits [12]. Hence, a powerful database can be created by reducing FTE/FTC error rates and spoofing the multiple sources at a time is not as easy as it is for single source.

Biometric systems which utilize two or more traits create a barrier for the user who he has to cross it by offering multiple evidences of his own biometric identities. Therefore, multimodal biometric system gives a higher degree of accuracy and its improved performance helps to design the more powerful and accurate security and personal identification systems. The process of combining two or more biometric modalities is called the fusion. An application which works on biometric security comprises of an architecture which allow the fusion at any of the four levels of the integration of the user information [13]. Every level of the system offers a mixture of pros and cons depending upon the difficulties faced during data fusion.

Fusion makes system robust, highly secure against the spoof attack and increases its performance. A fusion model for multiple traits is shown in Fig. 1, which shows that the information acquired from the sensor module can be fused at matching score level or at decision level with ease. Although sensor module and feature extraction modules have maximum amount of raw/redundant data yet preprocessing and autocorrelation among the data affects the accuracy of the fusion at these levels [14, 15].

The raw information acquired by the at the source of the sensors is a raw form of the data and can only be fused if it found to be in a common domain and compatible to each other [16]. However the presence of redundant and high dimensional data disqualifies the accurate data fusion at sensor module [17]. Data that reaches at feature extraction level is a selective set of feature vectors extracted from previous state which show the distinctive characteristic of the individuals. Many researchers contributed comprehensively in developing the fusion strategies at feature extraction level. Score level fusion is the most studied part of the fusion process in multimodal biometric system as the problem of dimensionality can be solved at this level with

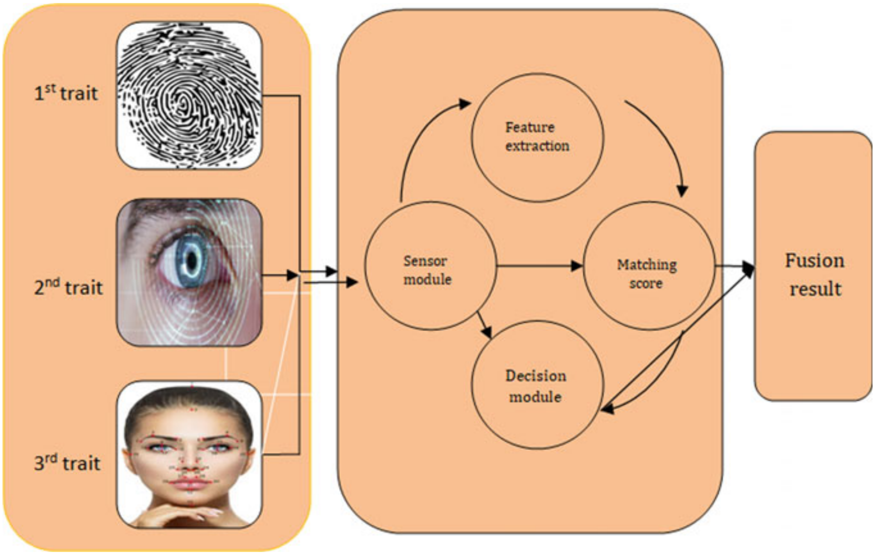


Fig. 1 Different levels fusion of Multibiometric traits

ease by the use of data normalization and classification [18]. Fusion at decision level does not require transformation technique as there is no such issue of data dimensionality and outliers. But the use of the decision level for fusion is beneficial only if the numbers of classifier are more than two, because the problem that may occur with decision level fusion is the possibility of ties. Decision at this level can be combined in serial and parallel form, through AND and OR fusion rules respectively and rule helps to reduce FAR and OR rule helps to reduce the FRR [19].

Moon et al. [20] created a composite image through ICP registration algorithm to showcase the sensor level fusion for the minutiae samples from two different fingerprint images. Recently, Raghavendra et al. [21] has demonstrated a novel approach of sensor level fusion for face and palm print images applying Particle Swarm Optimization (PSO). Singh et al. [22] conducted a sensor level fusion using multispectral imaging concept to integrate the face images from visible and NIR.

Fusion of the features which are obtained from the feature extraction processor became the successor of the sensor level fusion. Rattani et al. [2] presented a multimodal biometric system for the fusion of face and fingerprint images based upon a compatible feature extraction technique to receive corresponding features from the raw information. Zhang et al. [23] proposed a feature level fusion using Canonical Correlation Analysis (CCA) and Linear Discriminant analysis (LDA) which is an ideal technique to formulate and assess the linear relationship between two data sets. Kumar et al. [24] talks about the effect of feature subset choice and its potency in a distinctive bimodal biometric system. The extracted feature vectors of the hand image are compared and combined with Bayes transformation theory and sets are fused with learning rules and decision trees such as k-NNs, SVM, and FFNN. Zhou et al. [25] presented a principle component analysis (PCA) based feature selection method to choose main features from enhanced side face image (ESFI) and gait energy image (GEI), respectively. The separate feature vectors are then combined to get a complete and final feature vector and fused them at feature level only. Yan et al. [23] demonstrated a Correlation Filter Bank (CFB) analysis for the fusion of face and palm print at feature level.

The fusion at matching score level turns out to be the best and it encourages many researchers to work on this level comprehensively. Kuncheva et al. [26] worked on decision templates based method for multiple classifier fusion, and compared it with Behavior knowledge space method, majority voting, Naïve Bayesian, probabilistic product, fuzzy integral and linear discriminant classifier. For the fusion of fingerprint and face at score level Toh et al. [27] developed a Hyperbolic function network with the help of feed forward neural network (FFNN) and SVM. Jain et al. [28] studied the fusion of three modalities; face, fingerprint and hand geometry at the matching score level through autonomous Weighted Linear combination of similarity matching scores.

A polynomial model developed by Toh et al. [29] aimed to work upon a multivariate and high dimension data to reduce the complexity for matching score fusion. Frischholz et al. [30] considered two different fusion terminologies of matching scores of speech, face, and lip motion using Sum rule and Majority voting. A Weighted Sum rule based fusion technique used by Pavesic [31] for the fusion of three biometric traits, hand geometry, palm and fingerprint.

Literature work has shown the superiority of the Multibiometric system over unimodal biometric. NNs have played a vital role in implementation of complex and high dimensional data to achieve high accuracy in every possible way. The purpose of this study is to discuss the applicability of the NNs to the analysis of the Multibiometric based authentication systems.

3 Structure of the Neural Networks and Deep Learning

Neural networks are the computation models and have the functioning model of human brain. The NNs are trained to understand the data obtained through multiple sensors in multi-dimensions and later on used to cluster or classify them in meaningful information. Although NNs mosaics the brain yet they process the real world data such sound, images, sensory data and every other subject under processing in a computational form [32]. A more precise explanation of the NNs involves the consideration of neurons that exist in the form of nodes in a highly complex environment of interconnected network [33, 34]. Every neuron in the network has connecting medium called synapse through which information transfer takes place from one node to the other. This synapse contains the electric impulses in the form of the weights that carries all the information to be processed by perceptrons. The learning and outcome of artificially trained NNs are mainly depends on the weights of the network.

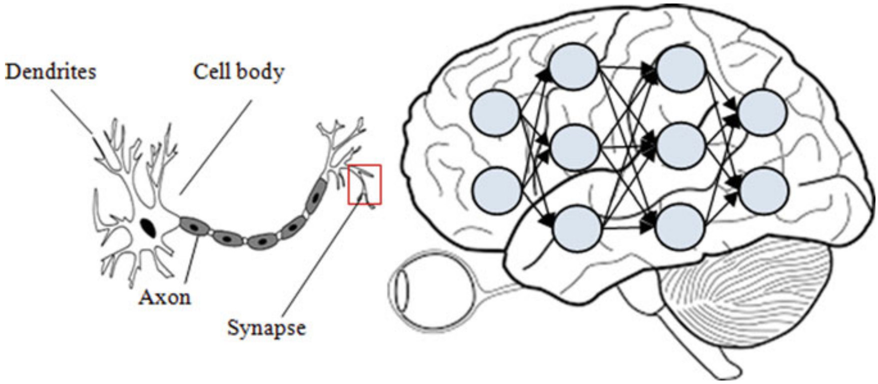


Fig. 2 Representation of the artificial neural network

Figure 2 shows the structure of the neuron and various elements involve in the mechanism of it. An equivalent description of ANNs is also depicted which has two hidden layers.

In context to fusion of multiple biometrics traits with the help of NNs can achieved with ease. For this we can develop a NN with single layer perceptron (SLP) or a multi layer perceptron (MLP) with one input layer, one output layer and one or more hidden layers depending upon the type of the perceptrons [35].

4 Comprehensive Detail of Available NNs

In this section we have presented the major category of the NNs available for the research purpose. The performance analysis of all the NNs has been studied and demonstrated in Table 1. The comparative study shows that as the size of the NNs increases in terms of layers and data to work upon its performance also increases.

In Table 1, a major category of the existing machine learning and ANN based methodologies are mentioned. However, every method has a wide range of application in the field of biometric security based on single and multimodal biometric traits but the knowledge of specific field can help in providing better result.

Furthermore, these methods can now be compared on single and multimodal biometric trait to know their applicability separately. A comparative study of the available NNs applied on single/multimodal trait has been presented in Tables 2 and 3 respectively.

In Table 2 a brief comparative study presented on the work demonstrated by various authors on different and single biometric traits. Table 2 comprises of the result of the single biometric traits such as retina, iris, finger, face and gait and are given in the form of GAR, FAR, FRR and EER. Although all the work reported here succeeded to achieve the GAR up to 90% or above but on the same time few of them

Table 1 List of all the available NNs and their utility index

NNs model and learning algorithms	Introduced by and year	Application	Mathematical expression
Perceptron	Minsky and Paert (1969)	Classification	$f(x) = \begin{cases} 1 \; if \; w.x + b > 0 \\ 0 \; otherwise \end{cases}$
BPNN	Rumelhart (1974)	Pattern recognition and classification	Error function: $E(X, \varphi) = \frac{1}{2N}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2$
FFNN	Oriented from Back Propagation (1982)	Pattern recognition and classification	Error Function: $E(X, \varphi) = \frac{1}{2N}\sum_{i=1}^{N}(g(\vec{w}.\vec{x} + b) - y_i)^2$
RBF	Broomhead and lowe (1988)	Function approximation, and machine control	$\varphi(X) = \sum_{i=1}^{N} a_i \rho(\Vert X - c_i \Vert)$
RNN	Based on Rumelhart work (1986)	Text analysis	$o^t = \left(h^t; \theta\right)$ $h^t = g\left(h^{t-1}; x^t; \theta\right)$
GRU	Kyunghyuncho (2014)	Sound and speech processing	$z_t = \sigma_g(W_z x_t + U_z h_{t-1} + b_z$ $r_t = \sigma_g(W_r x_t + U_r h_{t-1} + b_r$ $h_t = (1 - z_t) \, o \, h_{t-1} + z_t \, o \, \sigma_h$ $(W_h x_t + U_h(r_t \, o \, h_{t-1}) + b_h)$
AE	LeCun (1987)	Dimensionality reduction and feature learning	Reconstruction error: $£\left(\boldsymbol{x}, \boldsymbol{x}'\right) = \Vert \boldsymbol{x} - \boldsymbol{x}' \Vert^2$
HN	John Hopefield I (1982)	Denoising and restoration	$f(x) = \begin{cases} +1 \; if \; \sum_j w_{ij} s_j \geq \theta_i \\ -1 \; otherwise \end{cases}$
BM	Geofferyhinton (1975)	Feature detection and extraction	Global Energy: $E = \left(\sum_{i<j} w_{ij} s_i s_j + \sum_i \theta_i s_i\right)$
CNN	Yann LeCun (1994)	Recognition and data reduction in image processing	$x_{ij}^{l} = \sum_{a=0}^{m-1}\sum_{b=0}^{m-1} w_{ab} y_{(i+a)(j+b)}^{l-1}$
ELM	Guang Bin Huan (2004)	Network complexity reduction	Optimize form of ELM: $L_{P_{ELM}} = \frac{1}{2}\Vert \beta^2 \Vert + C\,\frac{1}{2}\sum_{I=1}^{N}\Vert \xi_i \Vert^2$ Subjected to: $h(X_i)\beta = e_i^T + \xi_i^T, \Delta_i$

(continued)

Table 1 (continued)

NNs model and learning algorithms	Introduced by and year	Application	Mathematical expression
KN	Kohonen (1982)	Classification	$m_i = \sum_j n_j h_{ji} \bar{x}_j / \sum_j n_j h_{ji}$
SVM	Vladimir N. Vipnik, Alexey Ya. Chervonenkis (1963)	Binary classification	Hard-margin: $y_i\,(\vec{w}\,.\,\vec{x} - b \geq 1),\ for\ all\ 1 \leq i \leq n$ Soft-margin: $\left[\frac{1}{n}\sum_{i=1}^{n} \max(0, 1 - y_i(\vec{w}\,.\,\vec{x} - b)\right] + \lambda\|\vec{w}\|^2$

Table 2 List of Unimodal Biometrics system implemented through NNs

Traits	Author and Year	NN model	Number of hidden layers	Accuracy rate (GAR, FAR, FRR, and EER) in %
Retina	Sadikoglu et al. (2016), [36]	BPNN	8–35	GAR: 85–97.5%
Iris	Abiyev et al. (2009), [37]	Gradient Based NN	2	GAR: 98.62
Finger vein	Yang et al. (2019), [38]	BDD-ML-ELM	2	GAR: 93.09
Hand Geometry	Franco et al. (2005), [39]	Unsupervised SOM, Supervised SOM and LVQ	2	FAR: 0.35 FRR: 0.15
Face	Er et al. (2002), [40]	RBF	1	GAR: 99.96
Palm print	Dian et al. (2018), [41]	CNN	8	EER: 0.044
Gait	Yeoh (2016), [42]	CNN	2	GAR: (CNN + SVM) 91.38 GAR: (CNN + Softmax) 87.80

could achieve near about 85%. Along with GAR, few methods failed to identify the imposters as a result the false acceptance rate and false rejection rate of these systems increases to a value of 0.15–0.35 This shows that the systems made with single biometric trait are always prone to security breach on low threshold and if we increases the threshold they might fall victim of less GARs.

Similarly, Table 3 summarizes the study of the work done in the field of security systems made with Multibiometric traits with different type of NNs. It is obvious from Table 3 that the obtained values of GARs are much better than the obtained in single biometric traits.

Table 3 List of Multimodal Biometrics system implemented through NNs

Traits	Author and year	NN model	Number of hidden layers	Accuracy rate % (GAR)
Face, ear and gait	Lathika (2014), [43]	FFNN	1	GAR: 99.21
Skeletal and Face	Priya (2019), [44]	BPNN	2	GAR: 98.62
ECG and Fingerprint	Hammad (2018), [45]	QG-MSVM and CNN	2	GAR: 98.66
Finger vein and finger shape	Kim et al. (2018) [46]	CNN	2	EER: 3.3653(SDU-DB) 1.0779 (Poly U-DB)
Speech and gait	Hamid et al. (2012), [47]	HMM and SVM	2	EER: 0.1548 (sum rule) 0.1487 (weighted sum rule)
Face and iris	Talreja (2019), [48]	MDHND	2	EER: 0.84

5 Conclusion

Formerly, NNs had a limited application areas and single biometric security system were one of them. With the emergence of security threats NNs are also upgraded their security parameters. The fusion of multiple biometric traits with the help of NNs has shown incredible capabilities in adapting the new challenges of developing the highly secure and reliable systems.

In this paper, we have presented an extensive comparative case study of the available literature in the field of single and multimodal biometrics. Since the result obtained through multimodal biometric traits are more accurate in terms of high GAR and low FAR and FRR. In future, the development of more reliable and more secure system can be achieved through deep learning.

6 Future Scope

Although the use of NNs in combination with multimodal biometric has shown the promising results in the field of biometric security, yet the use of the MLP with deep learning is the future of the biometric security. The deep learning is an advance technique for the data that is available in large amount and high in dimension. In deep learning the MLP has perceptron model of more than three layers which can handle a highly complex nature of data.

References

1. Faundez-Zanuy M (2006) Biometric security technology. IEEE Aerosp Electron Syst Mag 21(6):15–26
2. Miller B (1994) Vital signs of identity. IEEE Spectr 31:22–30
3. Ross A, Nandakumar K, Jain AK (2006) Handbook of multibiometrics. Springer, Berlin
4. Chellin G, Chandra J, Rajesh RS (2009) Performance analysis of multimodal biometric system authentication. Int J Comput Sci Netw Secur 9(3)
5. Nanavati S, Thieme M, Nanavati R (2002) Biometrics: identity in a networked world. Wiley, New York
6. Naik MV, Vasumathi D, Siva Kumar AP (2018) An enhanced unsupervised learning approach for sentiment analysis using extraction of tri-co-occurrence words phrases. In: ICCII. Springer, Berlin, pp 17–26
7. Hong L, Jain AK, Pankanti S (1999) Can multibiometric improves performance? In: Proceedings AutoID'99, Summit, NJ, USA, Oct 1999, pp 59–64
8. Chen Y, Das SC, Jain AK (2005) Fingerprint quality indices for predicting authentication performance. In: Proceeding to the fifth international conference on audio and video-based biometric person authentication (AVBPA), New York, USA, July 2005
9. NIST report to the United State of Congress. Summary of the NIST Standard for Biometric Accuracy, Temper Resistance, and Interoperability, Nov 2002. ftp://sequoyah.nist.gov/pub/nist_internal_report/NISTAPP_Nov2.pdf
10. Kukula EP, Sutton MJ, Elliott SJ (2010) The human–biometric-sensor interaction evaluation method: biometric performance and usability measurements. IEEE Trans Instrum Measur 59:784–791
11. O'Gorman L (2002) Seven issues with human authentication technologies. In: Proceedings of workshop on automatic identification advanced technologies (AutoID), Tarrytown, New York, Mar 2002, pp 185–186
12. Wechsler H et al (eds) (1998) Face recognition: from theory to applications. NATO ASI series. SERS. F. Springer, Berlin
13. Ross A, Jain AK (2004) Multimodal biometrics: an overview. In: Proceedings of the twelfth european signal processing conference, pp 1221–1224
14. Deriche M (2008) Trends and challenges in mono and multi biometrics. IEEE
15. Maltoni D, Maio D, Jain AK, Prabhakr S (2003) Hand book of pattern recognition. Spring, 1, 3, 34, 61, 101, 102
16. Iyengar SS, Prasad L, Min H (1995) Advances in distributed sensor technology. Prentice Hall
17. Jain AK, Ross A (2002) Fingerprint mosaicking. In: Proceedings of IEEE international conference on acoustics, speech, and signal processing (ICASSP), Orlando, Florida, 13–17 May 2002
18. Jain AK, Hong L, Kulkarni Y (1999) A multimodal biometric system using fingerprint, face, and speech. In: Second international conference on AVBPA, Washington, DC, USA, pp 182–187
19. Willem M, Fmet P (1997) Layped biometric verification, White paper, Keywm Technologies, Dec 1997
20. Moon YS, Yeung HW, Chan KC, Chan SO (2004) Template synthesis and image mosaicking for fingerprint registration: an experimental study. Int J ICASSP 4:409–412
21. Raghavendra R, Rao A, Kumar GH (2010) Multisensor biometric evidence fusion of face and palmprint for person authentication using Particle Swarm Optimization (PSO). Int J Biometrics (IJBM) 2(1)
22. Singh R, Vatsa M, Noore A (2008) Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition. Pattern Recogn Spec Issue Multimodal Biometrics 41(3):880–893
23. Zhang Y, Yan Y (2008) Multimodal biometrics fusion using correlation filter bank. Int J IEEE 8:1–4
24. Kumar A, Zhang D (2005) Biometric recognition using feature selection and combination. Int J ICIP 4:1–10

25. Zhou X, Bhanu B (2007) Feature fusion of side face and gait for video-based human identification. Pattern Recogn 41:778–795 (Center for Research in Intelligent Systems, University of California, Riverside, USA)
26. Kuncheva LI, Bezdek JC, Duin RPW (2001) Decision templates for multiple classifier fusion: an experimental comparison. Pattern Recogn 34:299–314
27. Toh K, Yau W (2004) Combination of hyperbolic functions for multimodal biometrics data fusion. IEEE Trans Syst Man Cybern Part B Cybern 34(2):1196–1209
28. Jain AK, Ross A (2002) Learning user-specific parameters in a multibiometric system. In: Proceedings of the IEEE international conference on image processing (ICIP), vol 1, pp 57–60
29. Toh K, Yan W (2002) Multi-modal biometrics fusion: beyond optimal weighting. In: 7th international conference on control, automation, robotics, and vision (ICACV'02), Singapore, Dec 2002, pp 788–792
30. Frischholz R, Dieckmann U (2000) BioID: a multimodal biometric identification system. Computer 33:64–68
31. Pavesic N (2003) A multimodal biometric user-identification system for network-based applications. IEEE Proc Vis Image Sig Process 150:409–416
32. Hassbis D, Kumaran D, Summerfiled C, Botvinivk M (2017) Neuroscience inspired artificial intelligence. Neurons Rev 95(2):245–258
33. Parisi GI, Kemker R, Part JL, Kanan C, Wermter S (2019) Continual lifelong learning with neural networks: a review. Neural Netw 113:54–71
34. Thrun S, Mitchell T (1995) Lifelong robot learning. Robot Auton Syst 15:25–46
35. Marinin F, Bucci R, Magri AL, Magri AD (2008) Artificial neural network in chemometrics: history, examples and persperctive. Microchem J 88(2):178–185
36. Sadikoglu F, Uzelaltinbulat S (2016) Biometric retina identification based on neural network. Procedia Comput Sci 102:26–33 (12th international conference on application of fuzzy systems and soft computing, ICAFS 2016, 29–30 Aug 2016, Vienna, Austria)
37. Abiyev RH, Altunkaya K (2009) Neural network based biometric personal identification with fast iris segmentation. Int J Control Autom Syst 7(1):17–23. https://doi.org/10.1007/s12555-009-0103-1
38. Yang W, Wang S, Hu J, Zheng G, Yang J, Valli C (2019) Securing deep learning based edge finger-vein biometrics with binary decision diagram. IEEE Trans Ind Informatics 15:4244–4253. https://doi.org/10.1109/tii.2019.2900665
39. Franco L, Jerez JM, Bravo JM (2005) Role of function complexity and network size in the generalization ability of feedforward networks. Lecture notes in computer science, pp 1–8. https://doi.org/10.1007/11494669_1
40. Er MJ, Wu S, Lu J, Toh HL (2002) Face recognition with radial basis function (RBF) neural networks. IEEE Trans Neural Netw 13(3):697–710. https://doi.org/10.1109/tnn.2002.1000134
41. Dian L, Dongmei S (2016) Contactless palmprint recognition based on convolutional neural network. In: 2016 IEEE 13th international conference on signal processing (ICSP). https://doi.org/10.1109/icsp.2016.7878049
42. Yeoh T, Aguirre HE, Tanaka K (2016) Clothing-invariant gait recognition using convolutional neural network. In: 2016 international symposium on intelligent signal processing and communication systems (ISPACS). https://doi.org/10.1109/ispacs.2016.7824728
43. Lathika BA, Devaraj D (2014) Artificial neural network based multimodal biometrics recognition system. In: 2014 international conference on control, instrumentation, communication and computational technologies (ICCICCT). https://doi.org/10.1109/iccicct.2014.6993100
44. Priya S, Mukesh R (2018) Multimodal biometric authentication using back propagation artificial neural network. https://doi.org/10.5013/ijssst.a.19.06.29. ISSN: 1473-804x online, 1473-8031 print
45. Hammad M, Liu Y, Wang K (2018) Multimodal biometric authentication systems using convolution neural network based on different level fusion of ECG and fingerprint. IEEE. https://doi.org/10.1109/access.2018.2886573
46. Kim W, Song JM, Park KR (2018) Multimodal biometric recognition based on convolutional neural network by the fusion of finger-vein and finger shape using near-infrared (NIR) camera sensor. Sensors 18(7):2296. https://doi.org/10.3390/s18072296

47. Hamid LA, Ramli DA (2012) Performances of qualitative fusion scheme for multi-biometric speaker verification systems in noisy condition. J Appl Sci 12(12):1282–1289
48. Talreja V, Valenti MC, Nasrabadi NM (2017) Multibiometric secure system based on deep learning. In: 2017 IEEE global conference on signal and information processing (GlobalSIP), Montreal, QC, pp 298–302. https://doi.org/10.1109/globalsip.2017.8308652

Retinal Blood Vessel Extraction Using Weighted Contrast Images with Deep Learning Methods

Shafiulla Basha Shaik and Venkata Ramanaiah Kota

Abstract Extraction of blood vessels from retinal fundus region is of great importance; as many of the eye diseases and pathological problems are diagnosed. Retinal vascular analysis is very important in diagnosing several diseases like diabetes, heart stroke and other health ailments that can be diagnosed with retinal images. This paper presents an effective approach of extracting blood vessel by applying weighted contrast approach as a pre-processing step and then extract the vessel using deep learning method. The method is compared with several existing methods experimented with standard dataset images and observed that this method could able to attain an average accuracy of 98.6%.

Keywords Blood vessel extraction · Weighted contrast adjustment · Deep learning method

1 Introduction

The main aim of any blood vessel extraction algorithm is to retain the orientation of blood vessel automatically, accurately and precisely. The automatic extraction algorithms reduce the cost and time involved in localizing the blood vessel in a very precise manner [1]. Whereas manual segmentation is a cumbersome process and precision is solely depending on the experience of the person [2].

Development of automatic blood vessel segmentation algorithms are very useful in diagnosing and localizing the abnormalities in the blood vessels that are prominently observed in diabetic retinopathy. Several authors have proposed multiple ways of extracting these vessels, Zhang in [3] extracted a multi-dimensional vector with morphological operators and using self-organizing map and Ostu-thresholding they could mark the outliers of the vessel tree. Ergen and others in [4] proposed a diagnosis approach with deep learning methods. They employed AlexNet and LeNet for classifying the abnormal images and predict the type of the disease. Eladawi and

S. B. Shaik (✉) · V. R. Kota
Department of Electronics and Communication Engineering, Y. S. R Engineering College of Yogi Vemana University, Proddatur, Kadapa, Andhra Pradesh 516360, India
e-mail: shafiulla@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_37

others in [5] developed a CAD (computer aided diagnosis) system that could able to detect and analyse the diabetic retinopathy by using feature based optical coherence tomography. However, feature based recognition algorithms not yield more accurate results when tested with unknown or untrained samples.

In [6] Ahmad et al., presented a VGG19 network based classification approach, where they observed that NASnet achieving highest accuracy and precision values of about 95% but the localization is efficiently done using VGG19 with minimum mean of 0.45.

To overcome the above mentioned shortcomings, this work presents a hybrid supervised deep learning method in which the images are pre-processed with weight contrast approach and thus obtained enhanced images are used for segmentation using convolutional neural network.

2 Weighted Contrast Adjustment (WCA) for Vessel Enhancement

The images that were imported from the databases are usually unprocessed and their intensities may vary unevenly but in order to locate the orientation of blood vessels accurately this images are to be processed. The proposed contrast adjustment algorithm involves conventional CLAHE [7] algorithm with adaptive weighting operation (Fig. 1).

The blood vessels are usually accompanying a reflex light which has to be minimized. This reflex runs through the central length of the blood vessel and this contrast adjustment process may be termed as central light reflex removal. As in [8], and in most of the retinal pre-processing approaches CLAHE algorithm is used but it often

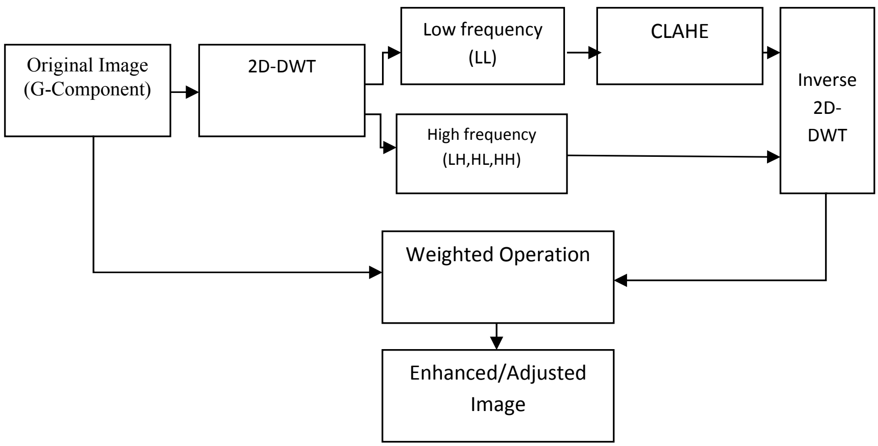


Fig. 1 Contrast adjusted process for the retinal fundus image

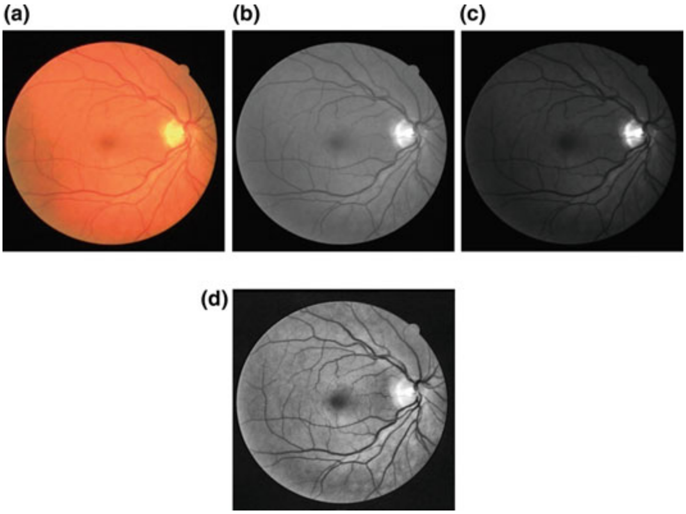


Fig. 2 **a** Original color fundus image. **b** Green component. **c** WCA approach. **d** CLAHE

observed that over enhancement occurs in some portions of the image which is to be treated. The results obtained with proposed approach and conventional CLAHE algorithms were shown in Fig. 2.

The original Green component (I_o) image is decomposed into low and high frequency components with 2D discrete wavelet transform and thus obtained low frequency components are applied with CLAHE operation and the high frequency components remains unchanged. The modified low frequency components along with high frequency components are subjected to inverse wavelet operation to obtain a reconstructed image (I_R). An adaptive weighting Operation is performed for the original and reconstructed image to obtain an enhanced image (I_E)

$$I_E = I_o * F + \beta * I_R * (I - F) \tag{1}$$

where "I" is the identity matrix and $1 < \beta < 2$ and "F" can be obtained as $F = f(k)^{\alpha}$ where $0 < \alpha < 1$ and $f(k) = \left[k(x, y)^{\alpha}\right]$. It can be observed from Fig. 4 that some of the regions in CLAHE are over enhanced but the contrast values are to be limited for accurate segmentation.

3 Deep Network for Vessel Segmentation

The block diagram of the proposed vessel training approach for CNN is depicted in Fig. 3. The approach includes training and testing phases, the training phase block diagram is depicted in Fig. 3.

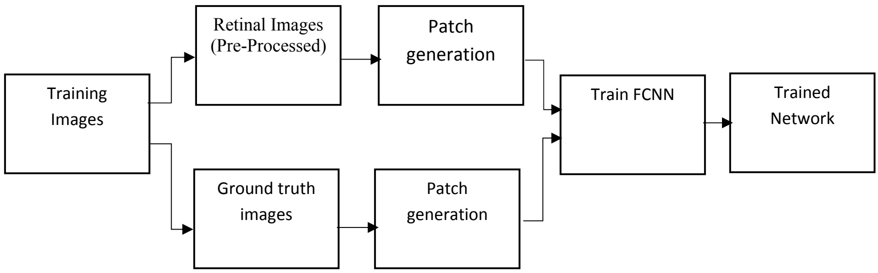


Fig. 3 Training mechanism for the proposed approach

In this phase, the training samples and their ground truth images are partitioned into non overlapped blocks of size $m \times n$ and these patches are directed for training the fully connected convolutional neural network (FCNN). The purpose of the convolutional layer is to extract features from the input data. It learns image features using small squares of input image and creates a feature map by maintaining spatial relation between pixels. After giving input to the convolutional layer, convolution is performed between input and the features learned by the network. Each convolutional layer output (Feature map) is passed through ReLU layer. ReLU is a Non-linear operation used to normalize the output of convolutional layer. It is applied per pixel and replaces all negative pixel values in the feature map by Zero. Output feature map of ReLU layer also have same resolution as input image. Pooling reduces the dimensionality of each feature map but retains the most wanted information. In our approach Max pooling is used in which the largest pixel value from the rectified feature map within the selected window is taken. Here we stack all the output feature maps of pooling layer and give as input to fully connected layers. The output of the final pooling layer acts as an input to the fully connected layer [9, 10].

4 Proposed Methodology

Figure 4 depicts the vessel classification approach presented with rained network.

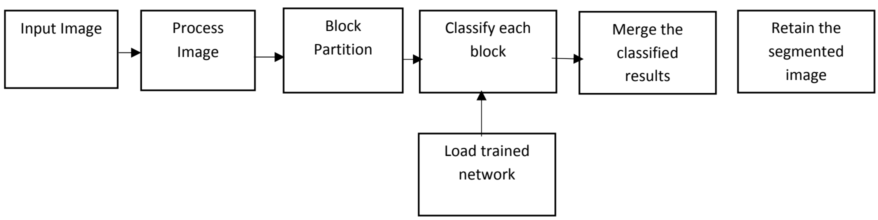


Fig. 4 Vessel classification approach

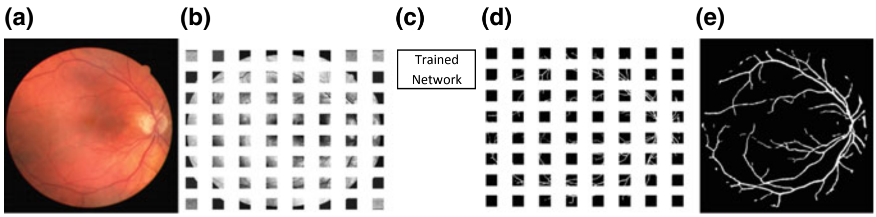


Fig. 5 Block wise outputs of the proposed approach. **a** Input test sample. **b** Block partitioned processed patches. **c** Trained network. **d** Classified patches. **e** Merged segmented output

In the testing phase the given input image is processed first with WCA and it is partitioned in $m \times n$ blocks. The trained network is loaded and tested for each patch; the respective classification result is retained. Thus, obtained patches are merged together to form a total segmented image. The step by step output of Fig. 3 is represented in Fig. 5.

5 Results

The experiments were conducted on publicly available datasets [11–13] from which the DRIVE dataset which comprises of 40 different subjects with their respective ground truth images. Here for our convenience the images are scaled to 512 × 512 with the block size of 64 × 64. The trained network is a 91-layered fully connected CNN with 100 connections. Initially we have considered 10 images for training which are scaled and block partitioned giving rise to 640 training patches. These patches on trained on "CPU" with parallel processing rather than using a GPU. Thus trained network is tested with reaming samples and the results are tabulated in the paper. The performance of the proposed approach is evaluated with metrics like Sensitivity (Se), Specificity (Sp), Positive Predicted Value (PPV), Negative predicted value (NPV) and Accuracy (ACC), for more information on the formulation of the metrics please refer [14]. The method here is compared against the start of art method like SAUCE [15] and principal curvature methods [16] and Kirchlet's template method [17] (Figs. 6 and 7; Tables 1, 2, 3, 4 and 5).

It can be observed from the above tabulated results that the proposed approach is performing well and attained an accuracy of 98.5% on average which is increased by 2.41%, which states that the proposed approach has reached the objectives of the current research work. It is also observed that the value of SP and PPV are "1" in every case that shows that the approach could retain all the vessel components when compared with ground truth images.

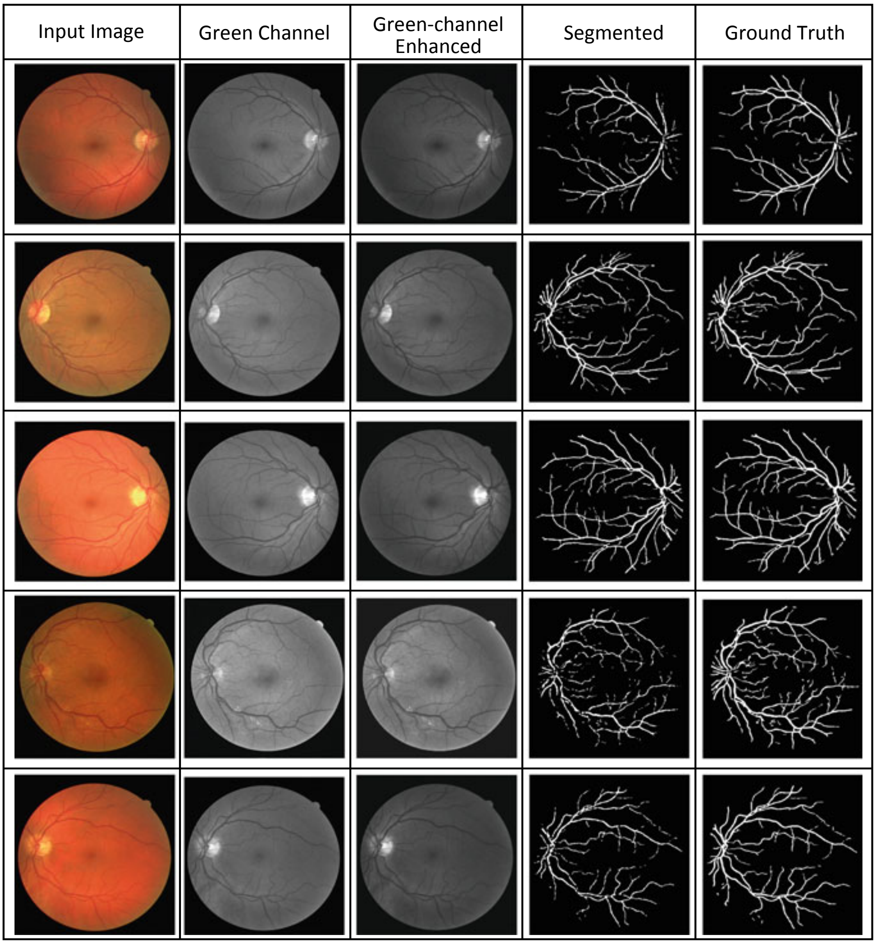


Fig. 6 **a** Column 1: input image. **b** Column 2: G-component. **c** Colum 3: Enhanced G-component. **d** Colum 4: Segmented Output with DNN. **e** Ground truth

6 Conclusions

This paper presents a hybrid approach of extracting retinal blood vessels which are more important in diagnosis of any retinal diseases. The method includes a weighted contrast approach that enhance the orientation of the blood vessel which helps the deep network to localize it accurately. The designed approach could able to attain an accuracy of about 98.6% which is improved accuracy when compared with state of art traditional methods. From the experimental results it was observed that the proposed approach accuracy is 2.5–4% improved performance with earlier that makes this approach to more suitable for real time deployment.

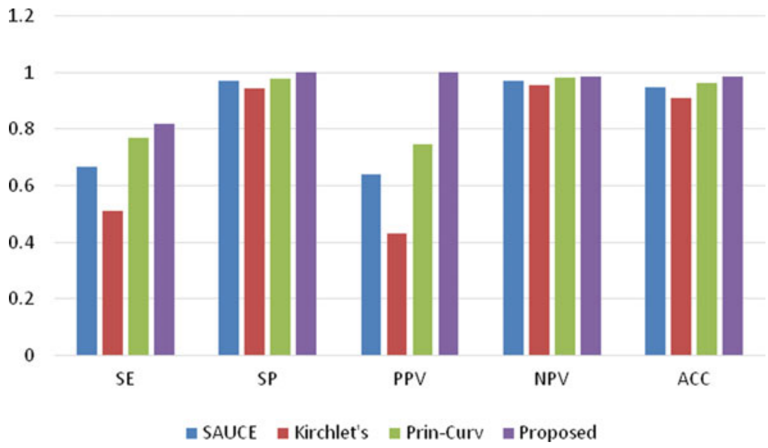


Fig. 7 Performance analysis of the blood vessel extraction algorithms

Table 1 Performance analysis of the SAUCE method [10]

	SE	SP	PPV	NPV	ACC
1	0.6653	0.9711	0.6747	0.9699	0.9458
2	0.6631	0.9716	0.697	0.967	0.9439
3	0.6291	0.9695	0.6579	0.9656	0.9405
4	0.5916	0.9753	0.6524	0.9683	0.9475
5	0.7422	0.9682	0.6356	0.9805	0.9524
6	0.6191	0.9739	0.6247	0.9732	0.9501
7	0.6791	0.9708	0.6368	0.9757	0.9503
8	0.6522	0.9782	0.6453	0.9564	0.9452
9	0.6078	0.9728	0.5771	0.9762	0.9518
10	0.8109	0.9595	0.5823	0.9865	0.9499
AVG	0.6660	0.9710	0.6383	0.9712	0.9478

Table 2 Performance analysis of the principal curvature [13]

	SE	SP	PPV	NPV	ACC
1	0.7742	0.9824	0.7986	0.9797	0.9652
2	0.7251	0.9905	0.8823	0.9734	0.9667
3	0.8415	0.933	0.5394	0.9844	0.9252
4	0.6004	0.9958	0.9182	0.9695	0.9671
5	0.7757	0.9893	0.8445	0.9883	0.9744
6	0.7329	0.9916	0.8604	0.9812	0.9744
7	0.7522	0.9732	0.6787	0.9812	0.9577
8	0.8162	0.9673	0.5784	0.9897	0.9594
9	0.8203	0.9836	0.7539	0.989	0.9742
10	0.844	0.9653	0.6284	0.9889	0.9574
AVG	0.768	0.9772	0.748	0.9825	0.9621

Table 3 Segmentation efficiency metrics for Kirch's template [15]

	SE	SP	PPV	NPV	ACC
1	0.55	0.932	0.4347	0.958	0.903
2	0.566	0.937	0.472	0.956	0.9043
3	0.346	0.951	0.397	0.939	0.899
4	0.576	0.935	0.411	0.966	0.9092
5	0.487	0.952	0.433	0.9613	0.9199
6	0.492	0.948	0.407	0.9636	0.9188
7	0.582	0.9393	0.4199	0.9676	0.9143
8	0.4764	0.9406	0.3062	0.9073	0.9164
9	0.499	0.952	0.367	0.9659	0.9327
10	0.486	0.9502	0.404	0.9636	0.92
AVG	0.506	0.943	0.405	0.954	0.912

Table 4 Performance analysis of the proposed approach

	SE	SP	PPV	NPV	ACC
1	0.9304	1	1	0.993	0.994
2	0.9312	1	1	0.993	0.9938
3	0.7890	1	1	0.979	0.9813
4	0.8134	1	1	0.985	0.986
5	0.823	1	1	0.987	0.987
6	0.765	1	1	0.983	0.9842
7	0.782	1	1	0.982	0.9847
8	0.8108	1	1	0.987	0.990
9	0.695	1	1	0.975	0.9764
10	0.8665	1	1	0.990	0.989
AVG	0.8206	1	1	0.9854	0.9861

Table 5 Average metrical values for the proposed and earlier methods

Method	SE	SP	PPV	NPV	ACC
SAUCE [10]	0.666	0.971	0.6383	0.9712	0.9478
Kirchlet's	0.512	0.9442	0.4318	0.956	0.9112
Princ-Curv [13]	0.768	0.9772	0.748	0.9825	0.9621
Proposed	0.8206	1	1	0.9854	0.9861

References

1. Pal AL (2015) Extraction of retinal blood vessels from retinal fundus image for computer aided diagnosis. In: Proceedings of national joint conference on innovations in engineering & technology. ISBN: 97-8-93-81195-82-6
2. Niemeijer M, Staal J, van Ginneken B, Loog M, Abramoff MD (2004) Comparative study of retinal vessel segmentation methods on a new publicly available database. In: Medical imaging, pp 648–656
3. Zhang J, Cui Y, Jiang W et al (2015) Blood vessel segmentation of retinal images based on neural network. In: International conference image and graphics, Cham, Aug 2015, pp 11–17
4. Sertkaya ME, Ergen B, Togacar M (2019) Diagnosis of eye retinal diseases based on convolutional neural networks using optical coherence images. In: 2019 23rd international conference electronics, Palanga, Lithuania, pp 1–5
5. Eladawi N et al (2019) Diabetic retinopathy early detection based on OCT and OCTA feature fusion. In: 2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019), Venice, Italy, pp 587–591
6. Ahmad M, Kasukurthi N, Pande H (2019) Deep learning for weak supervision of diabetic retinopathy abnormalities. In: 2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019), Venice, Italy, pp 573–577
7. Pizer SM, Amburn EP, Austin JD et al (1987) Adaptive histogram equalization and its variations. Comput Vis Graph Image Process 39(3):355–368
8. Setiawan AW, Mengko TR, Santoso OS, Suksmono AB (2013) Color retinal image enhancement using CLAHE. In: 2013 international conference on ICT for smart society (ICISS), pp 1–3

9. Chen X, Yang X, Wang M, Zou J (2017) Convolution neural network for automatic facial expression recognition. In: Proceedings of the 2017 IEEE international conference on applied system innovation
10. Ucar A (2017) Deep convolutional neural networks for facial expression recognition. In: IEEE international conference on innovation in intelligent system and applications (INISTA)-2017
11. https://www.isi.uu.nl/Research/Databases/DRIVE/
12. http://cecas.clemson.edu/~ahoover/stare/
13. https://blogs.kingston.ac.uk/retinal/chasedb1/
14. Shafiulla Basha S, Venkata Ramanaiah K (2019) An effective approach for blood vessel tree extraction from retinal images with adaptive weighted contrast adjustment approach. IJATCSE 8(2):285–292
15. Bandara MRR, Giragama PWGRMPB (2017) A retinal image enhancement technique for blood vessel segmentation algorithm. In: 2017 IEEE international conference on industrial and information systems (ICIIS), Peradeniya, pp 1–5
16. Fraz MM, Basit A, Remagnino P, Hoppe A, Barman SA (2011) Retinal vasculature segmentation by morphological curvature, reconstruction and adapted hysteresis thresholding. In: 2011 7th international conference on emerging technologies, Islamabad, pp 1–6
17. Li H, Chutatape O (2000) Fundus image feature extraction. In: Proceedings of the 22nd annual EMBS international conference, Chicago, IL, July 2000, pp 3071–3073

Nurse Alarming Device for Bedridden Patients Using Hand Gesture Recognition System

Vijaya Kumar Koppula, D. Sai Soumya and Suresh Merugu

Abstract It has always been a challenge for hospitals to eliminate the miscommunications between healthcare providers and those they serve. Whether it is to locate a nurse or responding to a patient's request for more pain medication, ask for any help or call in emergency—it has been outrageously time consuming and difficult to streamline and improve the workflow. Since patients need round-the-clock assistance or sometimes only need help now and then for certain tasks, there should be some solution which uses such technology that can give them peace of mind even when they encounter a situation that they cannot handle independently, or if they have a health issue and need help. Hand gesture based alarming device is a nurse calling alerting system that can keep patients safe since they will be able to call for assistance when required using their own hand gestures. This paper is about suggesting a way of interaction between nurse and patients along with real time gesture recognition. Previously, for some systems, excessive wiring in patient rooms and nurses' areas was a must, which made installation expensive and hard quite cumbersome. This technology which is purely software based and works with very less equipment seems very promising. As it can be used in hospitals, assisted living centres, or rehabilitation centres, which helps healthcare providers serve their patients more attentively and efficiently, it improves responsiveness, levels of care, operational efficiency, and patient or client satisfaction. This paper proposes the method or algorithm for an application which would help in recognising the different gestures of patient's hand. The images are of the palm side of right and left hand and are loaded at runtime. The method has been developed with respect to single user. The real time images will be captured first and then stored in directory and on recently captured image and feature extraction will take place to identify which gesture has been articulated by the user. The comparisons will be performed in arrears and then after comparison the result will be produced in accordance through matched key points from the input image to the image stored for a specific gesture already in the directory or the database the outputs for the following can be seen in below sections.

V. K. Koppula (✉) · D. Sai Soumya · S. Merugu
CSE Department, CMR College of Engineering & Technology, Medchal Road, Kandlakoya, Hyderabad, Telangana, India
e-mail: hodcse@cmrcet.org

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_38

Keywords Indian sign language · Feature extraction · Key point matching · Sign/gesture recognition

1 Introduction

Gestures are the oldest way of communication. It can be defined in various ways. If we want to sum up in simple words, Gesture is a physical action posed by a person to convey a message. Its roots are deeply penetrated in our communication system as people continuously pose even while they speak normally. Those who can't speak and hear, also communicate with gestures. It is really difficult finding a proper person for translating the sign language every time and everywhere but human-computer interaction system made this very easy as if it can be installed anywhere possible. This idea emerged with a motivation from the fact that it would prove to be of utmost productive for patients in their difficult times [1]. The god gifted ability of human vision is to recognize hand gestures which is mainly noticeable in differently abled people. In this paper, we take up one of the social problems to suggest a solution which can make a difference in reducing annual death rate caused by negligence or inattention by the attendants near patients [2]. This alarming device is based on human hand gesture recognition which uses a web cam to capture the images.

The research paper sheds light on recognizing hand gestures as a sign to map it with various notification sounds and speech. Detection of hand gestures is done using techniques of Counter analysis and feature extraction.

2 Literature Survey

So far, many techniques have been used to recognize hand gestures, convert them into speech and incorporate in into various applications. However, these came out to be limited in their functionalities [3]. Since many required gloves with sensors, some required lots of wiring, and some others used complex cables. These applications make the system more complicated as well as expensive [4]. Along with these difficulties, the system got limited to a particular boundary [5]. Clearing noise and removing disturbance has always been a challenge for these systems. There were some ideas which were totally dependent on GPUs hence making it very problematic for a common man to use them [6, 7]. Some more advanced projects were suggested which necessarily required certain skin tone for recognition. Although there are numerous systems to convert recognized gestures into texts, very few of them focus on converting into text and speech both that to with confined scope and restricted properties [8, 9]. Accelerometer and tactile sensors are another method used in the past. This is specifically to recognize and track the hand movement and gather information based on the path followed and gesture posed at the end. EMG sensors are used which generates various signs are sent as input to the system and produce text or speech as output.

Major algorithms used in common are Hidden Markov Mode (HMM), Convolutional Neural Networks (CNN), Artificial Neural Networks (ANN) and Sum of Absolute Difference (SAD) in specifically vision based gesture recognition systems. The main objective of this paper is to provide a tool to support patients by retrieving the features of their hands [10]. The feature recognition is done using Morphological Segmentation and CNN for Classification and image recognition which is developed with the help of OpenCV.

3 Proposed Methodology and Architecture

The process that is been implemented in this paper can be briefly summed up in a flow diagram Fig. 1. There are basically six major steps:

1. Input Image Pre-processing
2. Feature Extraction
3. Training
4. Classification
5. Image Recognition
6. Speech conversion.

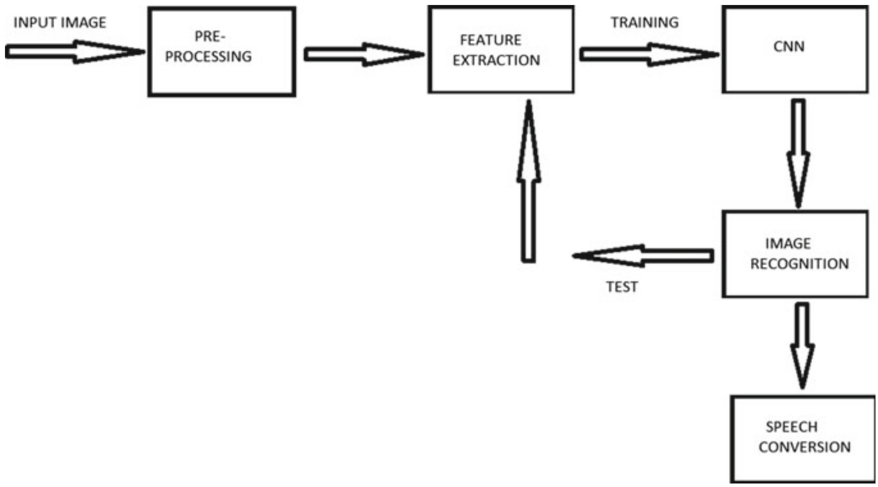


Fig. 1 Flow diagram of idea implemented

3.1 Input Image Pre-processing

This is the first and foremost important step that has to be taken care of. The input image, which is of 300 * 300 pixels is converted into grey scale image. Random variation and disturbances in image captured in context with brightness, or colour is termed as noise. This noise is undesirable and compulsorily removed. Though various filtering methods are available for de-noising the images, which includes Morphological filter, Kuan filter, Gaussian filter etc. we are here interested in Gaussian filtering of images. This uses Wavelet transformation to de-noise the image is very successful methodology.

3.2 Feature Extraction

It is important to separate the gesture captured with the background so as to make it easier for the process of recognition. This process is based on different pixel intensity of hand image and background. We allocate important pixels of hand regions and the contrasting one with different ranges. Like 255 (white) and 0 (black). We are using OpenCV inbuilt methods to do this. The function returns a set of coordinate values which when joined together makes the boundaries clearly visible.

3.3 Classification Using CNN

Convolution is the process of taking data in its original form, and creating feature maps from the same. Pooling is major part of this Convolution process. Max-Pooling is majorly used, where we select a region, then derive the maximum value of that region. This becomes the new value for the whole region. When we talk about Fully connected Layers, all the nodes are "completely connected". The convolutional neural network is partially connected with the nodes unlike traditional neural networks. Let us have a look on the algorithm and flow of how does this methodology works.

1. Take an image and convert it into pixels.
2. For the convolution, let us take a portion and define it as window and find features within that.
3. Slide the window over another and continue this process.
4. Overlapping will occur and we can define it in prior how much overlapping is allowed. This factor is must as it won't allow any pixel to skip.
5. After covering the entire image, we have to feature map it.
6. Next step is of pooling; in which we simply take the maximum value in the matrix.
7. Each convolution and pooling step becomes a hidden layer. By the end of this process, we get a fully connected layer followed by an output layer (Fig. 2).

Fig. 2 Contour extraction

3.4 Training and Testing

Once this classification is done using convolution neural networks, we can either test the data or we can train the data. When trained, we need multiple images of the same gesture so as to make the machine smart enough to recognize the gesture, even in cases on variations.

The more training of data sets is done, more efficient will me the application. To sum up, we can say that efficiency of a system is totally proportional to training images.

3.5 Speech Conversion

Once a gesture is recognized, we map it with a speech sentence which is to be delivered right after the image is perfectly identified. This image mapping to speech takes a unit of time that tells us about the robustness of code. To measure this, we have certain standard or expected output which is predefined by the system. The actual output is then compared with this expected one to find the margin of deviation.

Architecture:

This system majorly consists of two basic parts, front end part and back end part. In front end we have:

1. Camera module
2. Speech module.

And in the back end we have:

1. Detection module
2. Interface module
3. Speech conversion module.

Brief Explanation about module implementation is below.

1. *Camera module*:

The module helps in connecting and capturing the gesture image using different types of cameras or image detectors. Here, in our paper, we are using Simple web cam which is cost effective as well as readily available. It is user friendly as it is easy to setup. It recognizes both static and dynamic gestures.

2. *Speaker Module*:

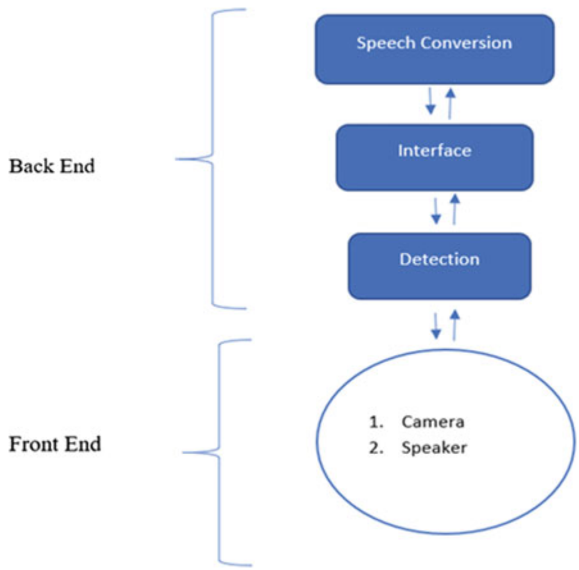
This module is to deliver the output speech which came out as a result after processing the image. These can be placed anywhere audible to nurse or sometimes they can be installed as a notification band and wrapped to the wrist of attendants.

3. *Detection Module*:

This module helps in image processing. The Image captured from the camera is undergone carious image processing techniques which includes—Colour detection, colour conversion, removal of noise, defining threshold values, etc. to extract the contour image. Removing convexity defects to recognize the main image plays the most important role in the application. Depending on the convexity hulls, classification algorithm can be determined.

Haar Cascade algorithm is useful in case of less or no convexity hulls. Here we have used Morphological Image Segmentation algorithm (Fig. 3).

Fig. 3 Flow diagram of modules



4. *Interface Module*:

This Module is responsible for training the system with various image data sets using some algorithm and mapping the trained images with the assigned functionality as of converting it to text or speech. The application consists of two windows:

1. One which captures the image using web camera,
2. The other which browses the image from the saved files, recognizes the result and delivers the speech.

4 Implementation and Results

In this paper, we have taken six static gestures, each of which shows six various messages as notification in the form of speech. First approach was of background subtraction and refinement. As the name itself shows, this is the process of separating the main gesture from background. It is majorly used in separating the static objects from the complex backgrounds. When implementing background subtraction in our recognition systems, we came across many challenges. It cannot deal with sudden or drastic changes because of which we encountered many inconsistencies. Relative parameters are also to be defined which we have to decide carefully. Because of these complications, we used Morphological segmentation for feature extraction and CNN for classification. This grouping helped us to achieve greater accuracy levels and also astounded the difficulties.

To measure the accuracy, we tested this two times and each time we changed the backgrounds. First, we had simple plain background (wall), and for the next time, we had a background having many complications. Each gesture was performed 4 times in each background and the number of times it was correctly recognized is shown here in Table 1. It was concluded that when the background was plain the code came out to be more robust and easily recognizing the images. This accuracy was maintained regardless of colour, skin tone, position etc. (Fig. 4).

Table 1 Results

Plain	Background		Heavy	Background
Gesture	Accuracy (%)	Speech Notification	Gesture	Accuracy (%)
1	94	I need help	1	50
2	84	I am in emergency	2	40
3	92	Please call the nurse	3	48
4	82	I want to go for a walk	4	42
5	95	Take me to the washroom	5	52
6	92	I need some water	6	80

Fig. 4 Gestured used in this alarming device

Expected output according to our algorithm is in-between the range of 75–85%. The actual output is defined as E_0.

$$E_0 = \text{Time Taken(in seconds)}/3 * 100$$

We limited it to 3 s so as to reduce the number of comparisons. Based on this, we underwent four rounds of tests.

As we observe that the Gestures 2 and 4 have low accuracy rates in comparison to others. This is because of the training sets we have used to train the particular gesture. For each gesture, we have used 140 images for training. So total training data is 140 * 6 i.e. 840 images. But gestures 2 and 4 required more training sets. So it ended up with less accuracy. The recognition rate of this alarming device is roughly 86.12%.

5 Conclusion and Future Scope

This alarming tool idea is entirely new and promising. Though many devices are existing in market with same goal of working, this paper suggests a method which bridges the interaction between human and computers. In future we would like to boost the accuracy and precision along with the implementation of video or motion capturing feature. We also aim to extend this idea to the next level which could help other needy persons when they require any assistance.

References

1. Azoz Y, Devi L, Sharma R (1998) Reliable tracking of human arm dynamics by multiple cue integration and constraint fusion. In: Proceedings of IEEE computer vision and pattern recognition (CVPR), Santa Barbara, CA, pp 905–910
2. Ahmad S (1994) A usable real-time 3D hand tracker. In: Asilomar conference on signals, systems and computers, Pacific Grove, CA, pp 1257–1261
3. Argyros AA, Lourakis MIA (2006) Vision-based interpretation of hand gestures for remote control of a computer mouse. In: ECCV workshop on HCI, Graz, Austria, May, pp 40–51

4. Argyros AA, Lourakis MIA (2006) Binocular hand tracking and reconstruction based on 2D shape matching. In: Proceedings of international conference on pattern recognition (ICPR), Hong-Kong, China
5. Azarbayejani A, Pentland A (1996) Real-time self-calibrating stereo person-tracker using 3-d shape estimation from blob features. In: Proceedings of international conference on pattern recognition (ICPR), Vienna, Switzerland, pp 99–108
6. Argyros AA, Lourakis MIA (2004) 3D tracking of skin-colored regions by a moving stereoscopic observer. Appl Opt 43(2):366–378
7. Argyros AA, Lourakis MIA (2004) Real-time tracking of multiple skin-colored objects with a possibly moving camera. In: Proceedings of European conference on computer vision, Prague, Czech Republic, May, pp 368–379
8. Athitsos V, Sclaroff S (2001) 3D hand pose estimation binding appearance-based matches in a large database of training views. In: IEEE workshop on cues in communication, pp 100–106
9. Athitsos V, Sclaroff S (2002) An appearance-based framework for 3D hand shape classification and camera viewpoint estimation. In: IEEE conference on face and gesture recognition, Washington, DC, pp 45–50
10. Athitsos V, Sclaroff S (2003) Estimating 3D hand pose from a cluttered image. In: Proceedings of IEEE computer vision and pattern recognition (CVPR), Madison, WI, vol 2, pp 432–439

Heat and Mass Characteristics of Magneto-Newtonian Fluid Through Upright Porous Plate

P. Chandra Reddy, P. V. Sanjeeva Kumar, L. Rama Mohan Reddy and M. C. Raju

Abstract An examination has performed to explain the flow characteristics of an unsteady MHD Newtonian fluid past over a vertical porous plate with rotation under the existence of heat and mass transfer. The governed expressions of the flow pattern are solved by using finite difference scheme. The impact of diverse parameters on the fluid velocity, temperature and species concentration is depicted in the form of numerical results and graphical presentations. The obtained results are having the close agreement with the existed literature results and promising the trueness of the numerical method. The enrichment of rotation parameter causes to decline the primary velocity of the fluid and also raises its secondary case velocity.

Keywords Rotating fluid · Thermal radiation · Chemical effect · Soret number and Dufour effect

1 Introduction

Rotating flows along porous media has received extensive importance in the modern research on computational fluid dynamics. Tremendous treatises on this topic with advantages in planetary sciences and geophysics have been published before the year 1950 onwards. The shared impact of heat transfer and mass transfer is much inspired to analysts in dynamic applications especially in chemical and manufacturing processes industries.

The theoretical concepts of revolving fluids were described by Greenspan [1]. Hydrodynamic resistivity and heat thrashing of revolving solids were established by

P. Chandra Reddy (✉) · P. V. Sanjeeva Kumar
Annamacharya Institute of Technology and Sciences, Rajampet, Andhra Pradesh, India
e-mail: chandramsc01@gmail.com

L. Rama Mohan Reddy
Department of Mathematics, IIIT Ongole Campus, RGUKT, Ongole, Andhra Pradesh, India

M. C. Raju
JNTUA College of Engineering, Pulivendula, Andhra Pradesh, India

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_39

Dorfman [2]. Seth et al. [3, 4] explained hydromagnetic Couette flow in an oscillatory rotating system by taking various parameters. MHD ephemeral flow along a precipitately started permeable flat plate considering Hall current has discussed by Ahmed and Sarmah [5].

To inspect the flow pattern under the control of cross diffusion effects we have gone through the articles which involve these effects. Tsai and Huang [6] measured Soret and Dufour effects with Hiemenz type flow onto a stretchable plane surrounded by holey medium. Narayana and Murthy [7] established these consequences in a Darcy interior. Narayana and Sibanda [8] analyzed the same along a vertical wavy surface. Alam et al. [9] illustrated cross diffusion impact of stable natural convective mass transport flow along holey medium. Makinde and Olanrewaju [10] considered these effects under mixed convection through a double blend of fluid under chemically reaction. Mishra et al. [11] discussed in detail the Soret and chemical parameter effects on hydro compelling micro polar fluid. Rashidi et al. [12] evaluated and explained differential transforms and group hypothesis of miscellaneous convective transfer commencing a level surface. Kumar and Singh [13] established the consequences of stimulated compelling region under the continuation of Newtonian heating/cooling. Mythili and Sivaraj [14] found the impact levels of higher sort chemical response and non-stable heat supply/sink on Casson flowing along a plane plate and upright cone. Sandeep et al. [15] selected 3D-Casson flowing embedded by a shell at complete zero under modified viscosity model. Abbasi and Shehzad [16] gave conclusions for heat transport scrutiny for Maxwell fluid flow in three-dimensional form by applying Christov-Cattaneo heat flux model. Deka et al. [17] analyzed MHD flow under transitory natural convection embedded with pores and gradually decomposing wall temperature and radiation. Further many researchers are given the studies on the flow characteristics of rotating channels [18, 19].

The above inspired findings gave a path to investigate the heat and mass characteristics for magneto-Newtonian fluid through upright porous plate with rotary motion. This study is the extension of the article by Deka et al. [17] with the novelty of including rotation parameter, heat generation, chemical reaction and cross-diffusion effects

2 Mathematical Formulation

The mathematical formulation is developed for MHD laminar fluid flow through rotating frame on a holey plate. The magnitude of temperature and concentration of the conducting field is adjustable by considering the chemical reaction, radiation and transverse diffusion into account. The plate is considered such that its sides are in vertical direction and choosing this as the direction of *X-axis* and *Y-axis* is considered in normal direction to the face of the plate. The fluid and the plate are considered in position as stiff body and rotate about Y-axis with angular velocity Ω. Initially at time $t^* \leq 0$, the fluid and the plate were at rest position with standardized temperature and concentration T_∞ and C_∞ respectively. At time $t^* > 0$, the plate starts moving in

X direction with uniform velocity $U_0 a^* t^*$. The temperature and concentration raises to $T_\infty + (T_s^* - T_\infty)\left(\frac{t^*}{t_0}\right)$ and $C_\infty + (C_s^* - C_\infty)\left(\frac{t^*}{t_0}\right)$ respectively.

Under these considerations the equations that govern the flow regime are as follows.

$$\frac{\partial u^*}{\partial t^*} + 2\Omega V^* = \nu \frac{\partial^2 u^*}{\partial y^{*2}} - \frac{\sigma B_0^2 u^*}{\rho} + g\beta_T\left(T^* - T_\infty\right) - \frac{\nu}{k}u^* + g\beta_C\left(C^* - C_\infty\right) \tag{1}$$

$$\frac{\partial V^*}{\partial t^*} - 2\Omega u^* = \nu \frac{\partial^2 V^*}{\partial y^{*2}} - \frac{\sigma B_0^2 V^*}{\rho} - \frac{\nu}{k}V^* \tag{2}$$

$$\rho C_p \frac{\partial T^*}{\partial t^*} = k_T \frac{\partial^2 T^*}{\partial {y^*}^2} + Q^*\left(T^* - T_\infty\right) - \frac{\partial q_r^*}{\partial y^*} + \frac{D_m k_T \rho}{C_s}\frac{\partial^2 C^*}{\partial {y^*}^2} \tag{3}$$

$$\frac{\partial C^*}{\partial t^*} = D\frac{\partial^2 C^*}{\partial {y^*}^2} - K_r^*(C^* - C_\infty) + D_1\frac{\partial^2 T^*}{\partial {y^*}^2} \tag{4}$$

The corresponding primary and edge conditions are

$$\left.\begin{array}{ll} u^* = 0, V^* = 0, T^* = T_\infty, \mathrm{C}^* = C_\infty & for\ all\ y^*, t^* \le 0 \\ t^* > 0 : u^* = U_0 e^{a^* t^*}, V^* = U_0 e^{a^* t^*}, & \\ \mathrm{T}^* = T_\infty + (T_s^* - T_\infty)\left(\dfrac{t^*}{t_0}\right), & at\ y^* = 0 \\ \mathrm{C}^* = C_\infty + (C_s^* - C_\infty)\left(\dfrac{t^*}{t_0}\right) & \\ u^* = 0, T^* = T_\infty, \mathrm{C}^* = C_\infty & as\ y^* \to \infty \end{array}\right\} \tag{5}$$

The term $\frac{\partial q_r^*}{\partial y^*}$ in Eq. (3) represents the alteration in the radiative flux with the distance perpendicular to the plate. For an optically slim gray gas, the local beaming is given by the result

$$\frac{\partial q_r^*}{\partial y^*} = 4a'\sigma^*\left(T_\infty^{*^4} - T^{*^4}\right)$$

where σ^* stands for Stefan-Boltzmann constant and a' denotes mean absorption coefficient. The differences contained by the flow are chosen as adequately minor. As a result T^{*^4} is possibly written as a linear function in T^* terms excluding greater order terms. The following estimate will be derived: $T^{*^4} + 3T_\infty^{*^4} \cong 4T_\infty^{*^3}T^*$. Then Eq. (2) can be modified and takes the subsequent form:

$$\rho C_p \frac{\partial T^*}{\partial t^*} = k\frac{\partial^2 T^*}{\partial y^{*2}} + q_0\left(T^* - T_\infty^*\right) - 16a'\sigma^* T_\infty^{*^3}\left(T^* - T_\infty^*\right) \tag{6}$$

The associated quantities are expressed in non-dimensional form and are given as:

$$u = \frac{u^*}{U_0}, \quad V = \frac{V^*}{U_0}, \quad t = \frac{t^* U_0^2}{\nu}, \quad t_0 = \frac{\nu}{U_0^2}, \quad y = \frac{y^* U_0}{\nu},$$
$$\theta = \frac{T^* - T_\infty}{T_s^* - T_\infty}, \quad \mathrm{C} = \frac{C^* - C_\infty}{C_s^* - C_\infty}, \quad a = \frac{a^* \nu}{U_0^2}$$

By using the above terms, the equations from (1) to (4) are condensed as

$$\frac{\partial u}{\partial t} + 2k_1^2 V = \frac{\partial^2 u}{\partial y^2} + Gr\theta + GmC - Mu - \frac{1}{K}u \tag{7}$$

$$\frac{\partial V}{\partial t} - 2k_1^2 u = \frac{\partial^2 V}{\partial y^2} - MV - \frac{1}{K}V \tag{8}$$

$$\frac{\partial \theta}{\partial t} = \frac{1}{\mathrm{Pr}}\frac{\partial^2 \theta}{\partial y^2} + Q\theta - F\theta + Df\frac{\partial^2 C}{\partial y^2} \tag{9}$$

$$\frac{\partial C}{\partial t} = \frac{1}{Sc}\frac{\partial^2 C}{\partial y^2} - KrC + So\frac{\partial^2 \theta}{\partial y^2} \tag{10}$$

where $Gr = \frac{\nu g \beta_T (T_s^* - T_\infty)}{U_0^3}$, (Grashof number) $Gm = \frac{\nu g \beta_C (C_s^* - C_\infty)}{U_0^3}$, (Modified Grashof number)
$M = \frac{\sigma B_0^2 \nu}{\rho U_0^2}$, (Magnetic parameter) $k_1^2 = \frac{\Omega \nu}{U_0^2}$, (Rotation parameter)
$K = \frac{k U_0^2}{\nu^2}$, (Permeability of the porous medium) $\mathrm{Pr} = \frac{\rho \nu C_p}{k_T}$, (Prandtl number)
$Q = \frac{Q^* \nu^2}{k_T U_0^2}$, (Heat absorption) $F = \frac{4\nu I^*}{\rho C_p U_0^2}$, (Radiation parameter)
$Kr = \frac{K_r^* \nu^2}{U_0^2}$, (Chemical reaction parameter) $S_0 = \frac{D_1 (T_s^* - T_\infty)}{\nu (C_s^* - C_\infty)}$ (Soret number)
$Df = \frac{D_m k_T (C_s^* - C_\infty)}{\nu C_s C_p (T_s^* - T_\infty)}$ (Dufour number) $Sc = \frac{\nu}{D}$, (Schmidt number)
The corresponding initial and boundary conditions are

$$\left.\begin{array}{ll} u = 0, V = 0, \theta = 0, \quad \mathrm{C} = 0 & \text{for all y, t} \le 0 \\ t > 0 : u = e^{at}, V = e^{at}, \theta = t, \mathrm{C} = t & at\ y = 0 \\ u = 0, V = 0, \theta = 0, \quad \mathrm{C} = 0 & as\ y \to \infty \end{array}\right\} \tag{11}$$

3 Methodology of Solution

The differential equations (7)–(10) are to be solved subjected to the preliminary and edge conditions (11). The computations for exact solution are complicated and therefore these equations are solved by choosing explicit form of finite-difference

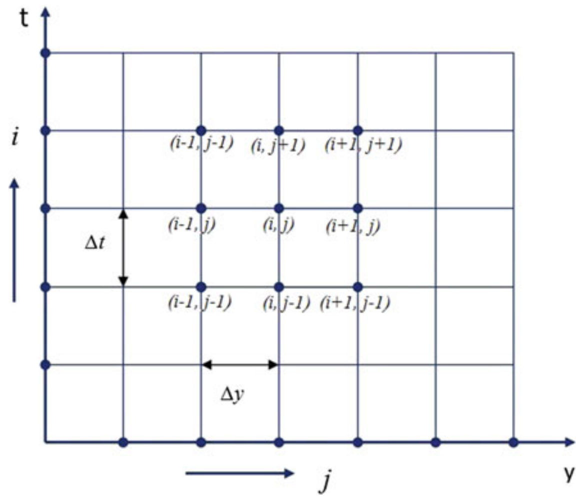


Fig. 1 Finite difference space grid

method. In connection to this, a rectangular domain which is suitable to the flow field is selected, and the region is alienated into a grid of straight lines parallel to X and Y-axes, as shown in Fig. 1.

Based on the forward finite differences the Eqs. (6)–(9) are articulated as:

$$\frac{u_{i,j+1} - u_{i,j}}{\Delta t} + 2k_1^2 V_{i,j} = Gr\theta_{i,j} + GmC_{i,j} + \frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{(\Delta y)^2} - Mu_{i,j} - \frac{1}{K}u_{i,j} \quad (12)$$

$$\frac{V_{i,j+1} - V_{i,j}}{\Delta t} - 2k_1^2 u_{i,j} = \frac{V_{i-1,j} - 2V_{i,j} + V_{i+1,j}}{(\Delta y)^2} - MV_{i,j} - \frac{1}{K}V_{i,j} \quad (13)$$

$$\frac{\theta_{i,j+1} - \theta_{i,j}}{\Delta t} = \frac{1}{\Pr}\frac{\theta_{i-1,j} - 2\theta_{i,j} + \theta_{i+1,j}}{(\Delta y)^2} + Q\theta_{i,j} - F\theta_{i,j} + Df\frac{C_{i-1,j} - 2C_{i,j} + C_{i+1,j}}{(\Delta y)^2} \quad (14)$$

$$\frac{C_{i,j+1} - C_{i,j}}{\Delta t} = \frac{1}{Sc}\frac{C_{i-1,j} - 2C_{i,j} + C_{i+1,j}}{(\Delta y)^2} - KrC_{i,j} + So\frac{\theta_{i-1,j} - 2\theta_{i,j} + \theta_{i+1,j}}{(\Delta y)^2} \quad (15)$$

In these schemes, the suffix "i" relates to "y-axis" and "j" to t-axis. The interior of grid is alienated into small meshes with the chosen value $\Delta y = 0.1$. The primary and edge constraints considered in (10) will get the subsequent form:

$$\begin{aligned} &u(i,0) = 0, V(i,0) = 0, \theta(i,0) = 0, C(i,0) = 0 \quad \forall i \\ &u(0,j) = e^{at}, V(0,j) = e^{at}, \theta(0,j) = t, C(0,j) = t \quad \forall j \\ &u(i_{\max}, j) = 0, u(i_{\max}, j) = 0, \theta(i_{\max}, j) = 0, C(i_{\max}, j) = 0 \quad \forall j \end{aligned} \quad (16)$$

($i_{\max}$ was chosen as 200).

The principal ending stage velocity $u(i, j + 1)$ $\{i = 1200\}$ with respect to time is calculated from Eq. (12) and that of secondary velocity $V(i, j + 1)\{i = 1200\}$ is derived by the scheme (13). The temperature θ $(i, j + 1)$ is evaluated with scheme

(14) and later C $(i, j+1)$ can be calculated from (15). Computations are continued up to t = 0.5 $\{j = 500\}$. Δt has been taken as 0.001 throughout the computations.
Skin-friction: The (gradient of velocity) skin-friction near the plate is computed by the relation

$$\tau = \left(\frac{du}{dy}\right)_{y=0}$$

Heat transfer rate: The (gradient of velocity) Nusselt number is given by

$$Nu = -\left(\frac{d\theta}{dy}\right)_{y=0}$$

Mass transfer rate: The rate of heat transfer is expressed in terms of Sherwood Number and is written as;

$$Sh = -\left(\frac{dC}{dy}\right)_{y=0}$$

4 Results and Discussions

The results obtained from the present investigation helps in drawing the influence of relevant physical parameters on friction, Nusselt and Sherwood numbers, molecular concentration, variations in velocity and temperature. The obtained results are tabulated and represented in graphical manner. The reliability of obtained results is having good agreement with the reference results [17]. This is obviously noticed in Fig. 2. The influences of magnetic parameter on both primary and secondary velocities are clearly illustrated in Figs. 3 and 4. These illustrations show that the increase in the magnitude of this parameter causes the decrement in velocity. This occurred physically because of the implementation of magnetic impact transversely. Figures 5 and 6 shows the impact of porousness on two types of velocities. Figures 7 and 8 represents the variations in the primary and secondary velocities under the rotation parameter and it is seen that the primary one is lessen with the increment of rotation parameter where as the secondary one is increased. The rich in magnitude of coriolis force is the main reason for the counterproductive of the primary velocity.

Figure 9 shows how the temperature is influenced by the source and sink of heat energy. This notices that the temperature is decreased due to the presence of heat inclusion parameter. It also revealed that improvement of temperature takes place under the generating heat. It is predicted that this variation is mainly due to diminish in thermal and kinetic energies of the fluid. This leads the decrease in the momentum thickness and thickness of thermal boundary layers in heat absorbing fluids. The temperature verses Dufour is shown in Fig. 10 and it clearly give the impact of

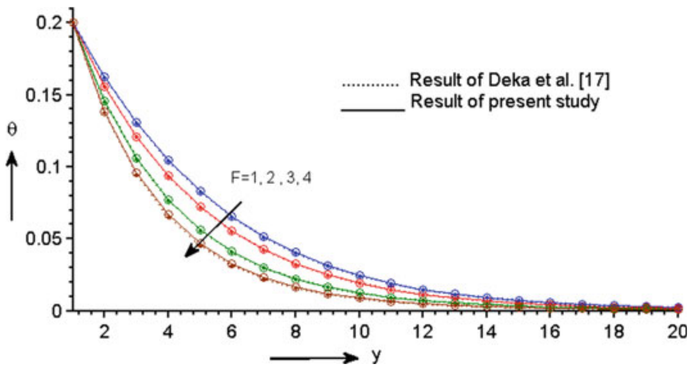


Fig. 2 Comparison of present result with that of Deka et al. [17] in the absence of porosity parameter, thermal diffusion, radiation, mass diffusion and rotation parameter

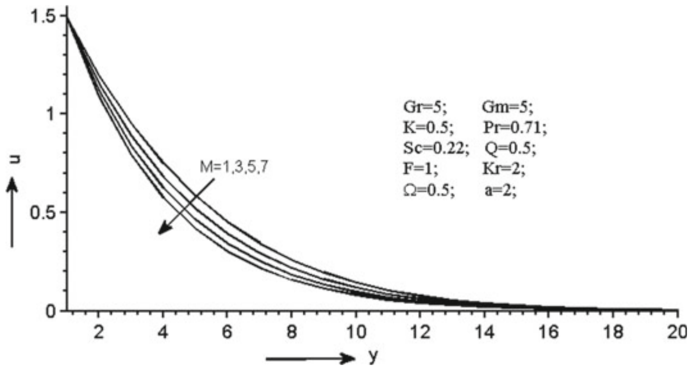


Fig. 3 Magnetic field impact on primary velocity

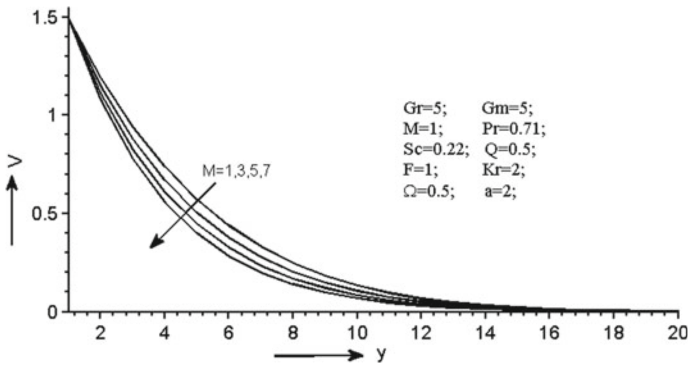


Fig. 4 Consequence of magnetic field on secondary velocity

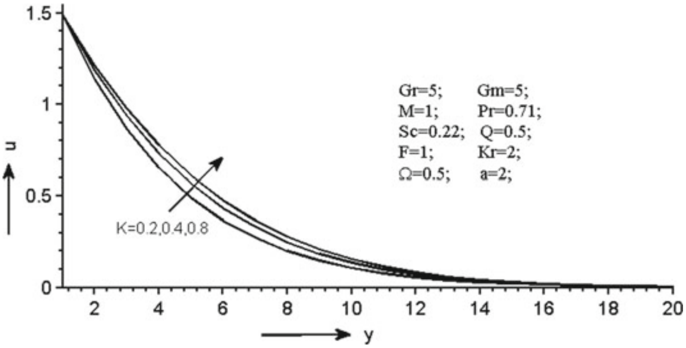


Fig. 5 Outcome of porosity parameter on primary velocity

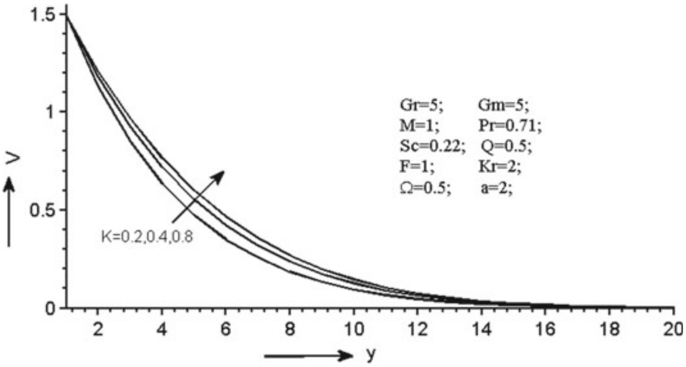


Fig. 6 Outcome of porosity parameter on secondary velocity

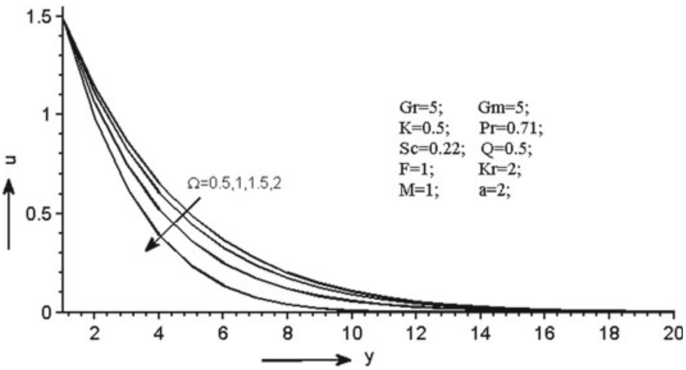


Fig. 7 Impact of rotation parameter on primary velocity

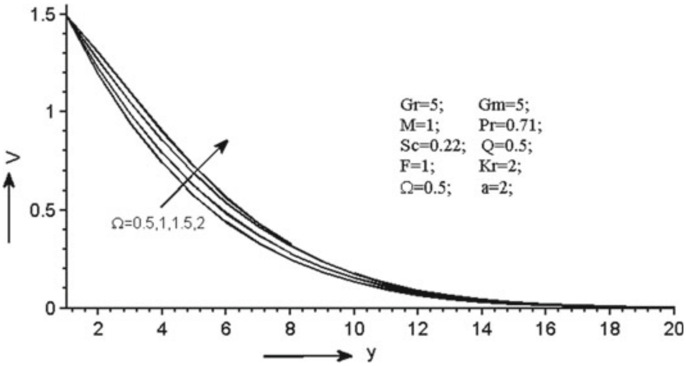


Fig. 8 Impact of rotation parameter on secondary velocity

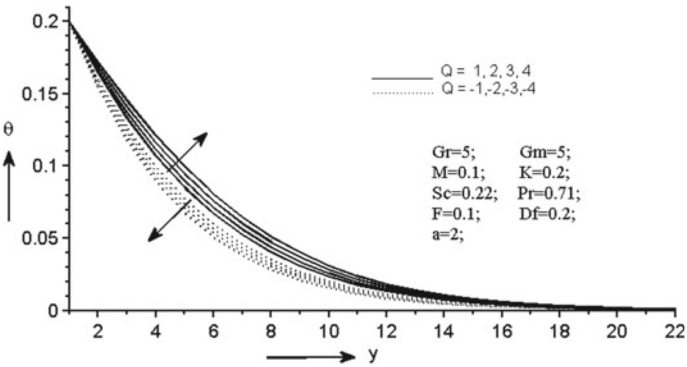


Fig. 9 Heat source and sink existence on temperature distribution

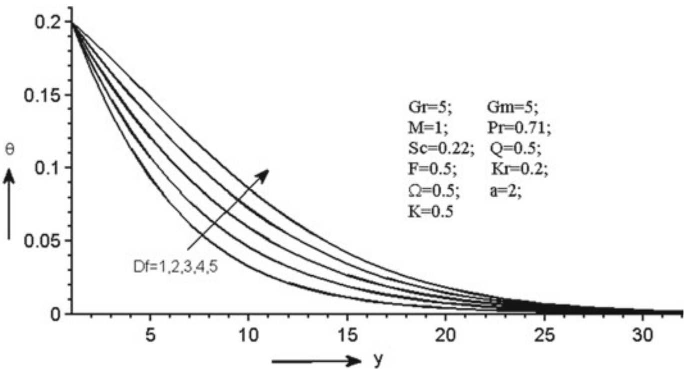


Fig. 10 Consequence of Dufour number on temperature

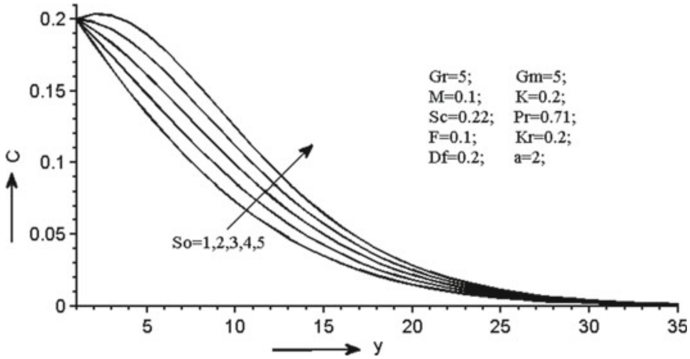


Fig. 11 Consequence of Soret number on concentration profile

thermal diffusion on the temperature. Figure 11 represents the concentration of fluid verses the thermal diffusion. In the presence of chemical reaction how the species concentration is varied is shown in Fig. 12. It is concluded that concentration of fluid is decrease due to the presence of chemical reaction.

The results showed in Table 1 are primarily belongs to consequences created by the physical parameters on skin friction. Impact of Grashof and modified Grashof numbers on skin friction is tabulated and it is pragmatic that the skin friction is decreased in both primary and secondary studies for increased values of these two parameters. Also it is observed that the skin friction is increased if the increased values of magnetic and porosity parameters. The skin friction increased in the presence of rotation parameter in primary case but it let down in the rotational situation.

The results recorded in Table 2 are belongs to the gradients of velocity and temperature with the increments of Prandtl Number, Dufour number, radiation parameter and heat source parameter. From table values it is noticed that the primary and secondary skin frictions are decreased if radiation parameter increased where as it is found that those are increased with increased values of diffusion thermo effect. Also

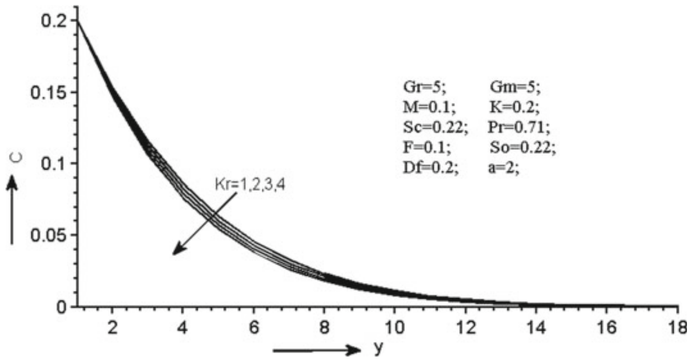


Fig. 12 Impact of chemical reaction on concentration

Table 1 Consequences in skin friction under the impact of considered parameters

Gr	Gm	M	K	Ω	τ_x	τ_y
3	4	8	1.5	2.8	3.8445	0.1049
5	4	8	1.5	2.8	3.6970	0.1036
7	4	8	1.5	2.8	3.3895	0.1024
9	4	8	1.5	2.8	3.0921	0.1012
3	6	8	1.5	2.8	4.5279	0.3227
3	8	8	1.5	2.8	4.5809	0.2669
3	10	8	1.5	2.8	4.4838	0.1565
3	4	9	1.5	2.8	1.2990	0.1100
3	4	10	1.5	2.8	1.3919	0.1142
3	4	11	1.5	2.8	1.6287	0.1183
3	4	8	2	2.8	3.0292	0.1015
3	4	8	2.5	2.8	3.0900	0.1016
3	4	8	3	2.8	3.1803	0.2018
3	4	8	1.5	3.4	3.2262	0.2015
3	4	8	1.5	3.8	3.2864	0.2014
3	4	8	1.5	4.2	3.3969	0.2013

Table 2 Variations in skin friction and Nusselt number

Pr	Q	F	Df	τ_x	τ_y	Nu
1.5	2.5	0.32	2	3.3192	1.0211	3.9248
2.5	2.5	0.32	2	3.0133	1.0387	5.4645
3.5	2.5	0.32	2	2.5327	1.0478	6.4965
1.5	2.5	0.32	2	2.3544	1.0485	4.4963
1.5	3.5	0.32	2	2.0724	1.0225	2.7262
1.5	4.5	0.32	2	1.8336	1.0139	0.4322
1.5	2.5	0.60	2	4.7318	1.1602	2.0368
1.5	2.5	0.78	2	4.7896	1.1523	2.0374
1.5	2.5	0.96	2	4.8322	1.1359	2.0388
1.5	2.5	0.32	2	3.9546	1.1024	2.3479
1.5	2.5	0.32	2.5	4.0256	1.1742	2.5476
1.5	2.5	0.32	3	4.3589	1.3965	2.8463

it is noticed that the heat transfer rate is increased for increased values of Dufour number, radiation parameter and Prandtl number where as the heat transfer rate is increased with decreased values of heat source parameter. The numerical results showed in Table 3 reveals that the variations in Sherwood number. This number gets

Table 3 Effect of chemical reaction parameter, Soret number and Schmidt number on Sherwood number

Kr	Sc	So	Sh
3	1	2	0.8081
5	1	2	1.2738
6	1	2	1.3198
3	1	2	0.1903
3	2	2	0.4004
3	3	2	0.9918
3	1	2	8.4572
3	1	4	7.9326
3	1	6	7.1524

improved under growing values of Schmidt number and chemical parameter but a reverse nature is found in the being of Soret effect.

5 Conclusions

The consequences of highlighted physical parameters, like thermal radiation, heat source and sink, chemical field reaction, thermo diffusion and Dufour number are considered and the flow characteristics are examined. The key points of this work can be acknowledged as follows:

- The velocity of the fluid diminishes in primary case and improves in secondary case for enhancing values of rotation parameter.
- Fluid temperature augments when the heat source and diffusion thermo effect is improved whereas a quash tendency is initiated in the existence of heat sink.
- The continuation of thermal transmission consequences in improving the concentration of the fluid whereas the chemical effect leads to lessen it.
- Skin friction of primary case grows with the impact of rotation parameter and it falls down in secondary case.
- Heat transfer gradient rises when Prandtl number, Dufour number and radiation parameter increases and it decays under the existence of heat source.

References

1. Greenspan HP (1968) The theory of rotating fluids. Cambridge University Press, New York
2. Dorfman LA (1963) Hydrodynamic resistance and the heat loss of rotating solids. Oliver & Boyd, Edinburgh
3. Seth GS, Ansari MS, Nandkeolyar R (2011) Unsteady Hartmann flow in a rotating channel with perfectly conducting walls. Int J Appl Mech Eng 16:1129–1146

4. Seth GS, Hussain SM, Sarkar S (2014) Hydromagnetic oscillatory Couette flow in a rotating system with induced magnetic field. Appl Math Mech 35:1331–1344
5. Ahmed N, Sarmah HK (2011) MHD transient flow past an impulsively started horizontal porous plate in a rotating system with Hall current. Int J Appl Math Mech 7(2):1–15
6. Tsai R, Huang JS (2009) Heat and mass transfer for Soret and Dufour's effects on Hiemenz flow through porous medium onto a stretching surface. Int J Heat Mass Transf 52:2399–2406
7. Narayana PL, Murthy P (2007) Soret and Dufour effects in a doubly stratified Darcy porous medium. J. Porous Media 10:613–624
8. Narayana PL, Sibanda P (2010) Soret and Dufour effects on free convection along a vertical wavy surface in a fluid saturated Darcy porous medium. Int J Heat Mass Transf 53:3030–3034
9. Alam MS, Ferdows M, Ota M, Maleque MA (2006) Dufour and Soret effects on steady free convection and mass transfer flow past a semi-infinite vertical porous plate in a porous medium. Int J Appl Mech Eng 11(3):535–545
10. Makinde OD, Olanrewaju PO (2011) Unsteady mixed convection with Soret and Dufour effects past a porous plate moving through a binary mixture of chemically reacting fluid. Chem Eng Commun 7:920–938
11. Mishra SR, Baag S, Mohapatra DK (2016) Chemical reaction and Soret effects on hydromagnetic micropolar fluid along a stretching sheet. Eng Sci Technol Int J 19(4):1919–1928
12. Rashidi MM, Rahimzadeh N, Ferdows M, Uddin MJ (2012) Anwar Bég O (2012) Group theory and differential transform analysis of mixed convective heat and mass transfer from a horizontal surface with chemical reaction effects. Chem Eng Commun 199(8):1012–1043
13. Kumar D, Singh AK (2015) Effect of induced magnetic field on natural convection with Newtonian heating/cooling in vertical concentric annuli. Procedia Eng 127:568–574
14. Mythili D, Sivaraj R (2016) Influence of higher order chemical reaction and non-uniform heat source/sink on Casson fluid flow over a vertical cone and flat plate. J Mol Liq 216:466–475
15. Sandeep N, Koriko OK, Animasaun IL (2016) Modified kinematic viscosity model for 3D-Casson fluid flow within boundary layer formed on a surface at absolute zero. J Mol Liq 221:1197–1206
16. Abbasi FM, Shehzad SA (2016) Heat transfer analysis for three-dimensional flow of Maxwell fluid with temperature dependent thermal conductivity: application of Cattaneo-Christov heat flux model. J Mol Liq 220:848–854
17. Deka RK, Kalita N, Paul A (2015) Transient free convection MHD flow past a vertical plate immersed in porous medium with exponentially decaying wall temperature and radiation. In: Proceedings of international conference on frontiers in mathematics, pp 158–161
18. Seth GS, Kumbhakar B, Sharma R (2015) Unsteady hydromagnetic natural convection flow of a heat absorbing fluid within a rotating vertical channel in porous medium with Hall effects. J Appl Fluid Mech 8(4):767–779
19. Seth GS, Kumbhakar B, Sarkar S (2015) Soret and Hall effects on unsteady MHD free convection flow of radiating and chemically reactive fluid past a moving vertical plate with ramped temperature in rotating system. Int J Eng Sci Technol 7(2):94–108

IoT Based MEMS Crash Sensor for Airbag System

Shameem Syed, Viswa Teja Challa, Yaswanth Maddi and Hari Kishore Kakarla

Abstract Technology is increasing exponentially day by day, besides this increasing of traffic takes place which leads to increasing of accidents. During accidents if airbags are not there then people may die because of severity of accidents. If airbags are present then people can be safe comparatively. To deploy airbags properly we need a system. If airbags can't deploy properly then people get injured or in some cases they may die. If airbags deploys with high speed then old people and infants may die because of that force of deploying airbags. Beside this if airbags can't deploy in time then people may die. To control the deployment of airbags we are having different sensors life g sensor, crash sensor, accelerometer sensor. Using any of these sensors it can send alert message to airbag system and we can share our location to our friends or family members and we can send message to ambulance because this is help to save lives of victims. It share's accident taken place location with the help of GSM sim 300 module by knowing the latitude and longitude of location with GPS NEO 6m module. Here GSM module used to send message and GPS module used to detect the location of accident taken place area. To interface all these we require a micro controller, we used Arduino UNO R3 to interface GPS module, GSM module and crash sensor. For future development we can interface camera to capture the accident area and capture the position of victims in car.

Keywords Crash sensor · Airbag system · Sensors · IoT · GPS · GSM

1 Introduction

MEMS are defined as micro electro mechanical systems which is a promising technology and most developing technology in present world. It is a technology where used to design and manufacture micro sized devices such as IC's. It's a combination of electrical and mechanical components. Here in this technology we no more use any mechanical devices but this imitates the properties of mechanical devices.

S. Syed (✉) · V. T. Challa · Y. Maddi · H. K. Kakarla
Department of ECE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, India
e-mail: shameemsyed@yahoo.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_40

To build any MEMS [1] devices there will be a separate process. MEMS devices like IC's can be manufactured through LIGA process. MEMS are not confined to single area but it was used in many areas such as mechanical engineering, electrical, biology, chemistry, optical, petroleum and so on.

Airbag [2] system is a device which is used to deploy air bags when an accident is taken place. Airbag system controls the airbag. When airbag system get signal from crash sensor then it deploys airbags. Airbag is a vehicle safety component that inflates air into flexible fabric bag (Fig. 1).

The hardware components for this project is crash sensor, Arduino UNO R3, GSM sim 300 and GPS NEO 6m beside this we need jumper wires, bread board, 12 V adapter for power supply and USB cable to connect Arduino with pc.

In this Arduino is interfaced with GPS, GSM and sensor. According to Fig. 2 we are interfacing those modules with Arduino. To run properly we need to write a code in Arduino IDE according to our requirement and required functionality.

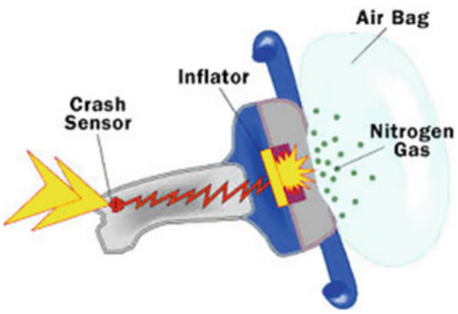


Fig. 1 Working of airbags system in a vehicle

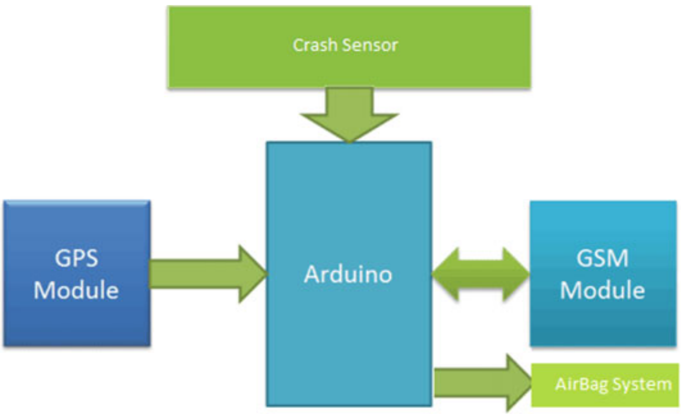


Fig. 2 Block diagram

In this project we are interfacing GPS, GSM and Crash sensor with Arduino. When any accident happens then crash sensor that placed front part of the vehicle send signal to the controller then that controller responds to it and activates GPS and GSM to get the location and share the location. Using GPS we will get latitude and longitude and GSM send that message to the saved number. Beside this controller sends signal to airbag system for deployment of airbags within time.

2 Literature Review

Prabha et al. [3] discussed about vehicle accident detection using piezo electric sensor, ARM 7 TDMI is a microprocessor for interfacing devices, EEPROM to store mobile number and send alert message. They used GPS and GSM alert based algorithm and implemented with LPC2148 MCU. They have proposed a vehicle accident detection system that which can track the accident location accurately. Beside this they have mentioned about some applications named field sales, fleet management, asset tracking and stolen vehicle recovery.

Zaldivar et al. [4] discussed about the connecting mobile with vehicle to monitor the vehicle condition. In this they have combined mobile to vehicle with the help of OBD-II interface. This gives alert message to mobile whenever vehicle condition is not proper and when accident is takes place then it will send message to saved numbers and send e-mail. Behalf of this they also give information to the emergency call to hospitals. They also tracking the exact location of incident and sharing those data to the saved members. They have developed for android platform only.

Bhumkar et al. [5] (Accident Avoidance and Detection on Highways) discussed about increasing of present technology and increasing of traffic day by day. Due to this accidents also increasing. So they prepared an algorithm with the help of eye blink, alcohol, impact, gas sensors. Using this sensors they will detect the fatigue and get alert when it detected. Using this they can save the lives of people responding them immediately. They have used MQ-3 gas sensor, MQ-6 gas sensor and GPS Receiver.

Song et al. [6] discussed about increasing of traffic leads to increasing of overweight vehicles collision with highway bridges. They developed a system using piezo ceramic transducer to detect collision, impact detection and health monitoring system. They have developed a circuit to determine impact of collision and activation of cameras to take photos of collision of truck with concrete bridge. Buy impact test they came to know that output of piezo ceramic material is proportional to the severity of collision. With the help of impact sensor signal they can estimate the impact level. Using PZT health monitoring system they can estimate the growth of cracks inside the concrete structure.

Deotare et al. [7] (Intelligent Car System for Accident Prevention Using ARM-7) discussed about increasing of present technology and increasing of traffic day by day. Due to this accidents also increasing. So they prepared an algorithm with the help of eye blink, alcohol, impact, gas sensors. Using this sensor they will detect the fatigue and get alert when it detected. Using this they can save the lives of people responding them immediately. They have used MQ-3 gas sensor, MQ-6 gas sensor and GPS Receiver.

Shameem et al. [8] discussed about Design and simulation of crash sensor in Comsol Metaphysics tool and they have check the performance of that crash sensor using different materials. According to results they came to know among all materials gold has good efficiency and performs very well comparatively. When crash or accident takes place then the proof mass of the sensor displace and develops a capacitance across plates which leads to development of voltage and sends signal to the connected devices such as microcontroller or microprocessor.

Prasad et al. [9, 10] discussed designed and analyzed a crash sensor in Comsol Metaphysics tool using different types of materials like poly-silicon, gold, silicon and aluminum oxide and their design parameters are, radius of 120 μm and thickness of 1 μm. Using this they constructed sensor and simulated and come to know the required load is 16N. Here they calculated Eigen frequency and displacement of above different materials. Using semiconductors and insulators materials they tested or simulated and come to know that gold is the best material for that design.

3 Hardware

In this project we used Arduino UNO R3, GPS NEO 6m, GSM sim 300 and crash sensor. Arduino shown in Fig. 3 is comes under embedded system which has platform on both software and hardware system. It is a open source hardware and software

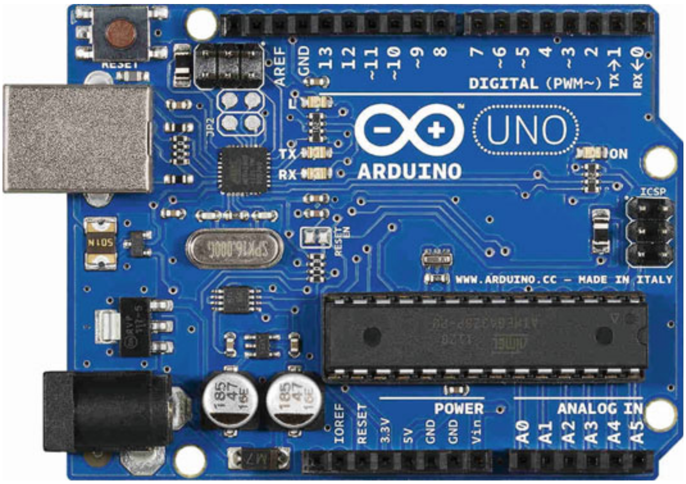


Fig. 3 Arduino UNO R3

company. It is used for IOT applications and using this we can implement or do many hardware projects. It is mostly used by students [11, 12].

In Arduino there are many types line mega, UNO, node MCU and so on. In this project we use Arduino UNO R3, it has an AT mega 328p AVR microcontroller in it. It has Analog and Digital pins. There are 20 digital input/output pins among them there are 6 PWM generate pins. In this we have power pins which consist of GND, 3.3, 5 V. To operate voltage of Arduino UNO is 5 V.

To run this system according to our requirement we need to write code for it. Requirements of writing code are Arduino IDE and features for programming are c and c++.

GSM stands for Global System for Mobile. It is developed by European Telecommunication Standard Institute. First it is 2G network used in mobiles and tablets. Now we using this modules for real time projects for sharing information. In GSM there many models such as GSM sim 300, GSM sim 900A, GSM sim 808, GSM sim 800L and so on.

In this project we used GSM sim 300 Fig. 4. Using RS232 we can interface GSM. Power on and Power off control pin for auto reset with the help of microcontroller. It is very helpful to auto rest functionality. For additional safety they connect antenna connector to separate ground plane. GSM works on frequencies dcs 1800 MHz, pcs 1900 MHz and egsm 900 MHz. Range of baud rate to operate GSM sim 300 is 1200 to 115,200 BPS. GSM sim 300 has 4 pins they are Vin, Tx, Rx and GND. In GSM sim 300 it has 4 pin connector for mic and speaker. Input power is 9–12 V/15A. Features of GSM sim 300 is GPRS multi slot and GPRS coding schemes.

Fig. 4 GSM sim 300

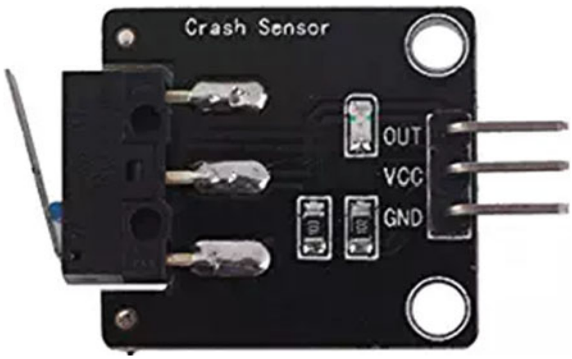
GPS is defined as Global Positioning System. Actually GPS is a space based satellite navigation system. It is used to get the location and time. In all weather conditions it works. GPS project is launched by U.S. Department of defense. Now a days GPS play vital role in telecommunication and used in mobile phones. GPS has many applications such as military, commercial and civil.

In GPS there are different types among them GPS NEO 6m Fig. 5 is used in this project. It is used to get location using external antenna connected to it. Features of GPS NEO 6m is, Its baud rate is 9600. Operating voltage is 5 V. It has 4 pins there are Vcc, Tx, Rx and GND. While interfacing this with Arduino in code we need to add library named Tiny GPS or Tiny GPS++.

Crash senor Fig. 6 is a switch which is also known as micro switch. It acts as a electrical switch but it needs a little physical force only. Operating voltage is 5 V. It has either high or low which represents digital pin. In this there are only 3 pins they are Digital OUT, Vcc and GND. There will be on board LED which represents as status indicator.

Fig. 5 GPS NEO 6m

Fig. 6 Crash sensor



Crash sensor is used for detecting the collision and sends signal through digital output. Because of this reason it is also called as collision signal sensor. It should send signals within milliseconds. When collision occurs the accelerating force acting on sensor is 100 times greater than gravitational force.

Previously we have designed and analyzed a crash sensor [8] in Comsol Multiphysics FEM tool. In Comsol we have designed the crash sensor Fig. 7 with parameters of, for outer cylinder are radius of 100 μm and thickness is 5 μm. For inner cylinder is radius of 20 μm and thickness is 5 μm. It has menders in between outer and inner cylinder their design parameters are length is 40 μm, width is 10 μm and height or thickness is 5 μm (Fig. 8).

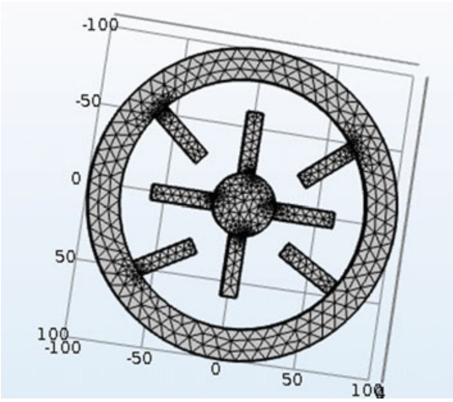


Fig. 7 Structure and parameters of crash sensor that we designed

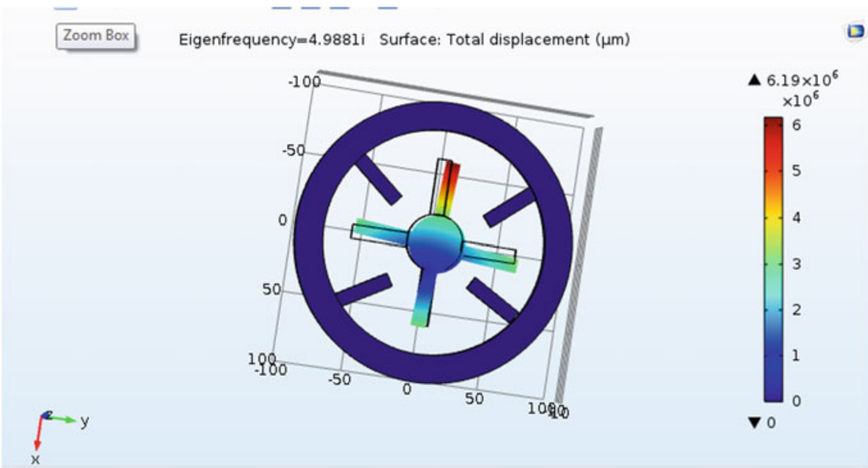


Fig. 8 Displacement of proof mass when a force is applied

Table 1 Displacement with respect to their Eigen frequency

Eigen frequency (Hz)	Displacement (μm)
1.16	2.65×10^7

We have implemented different materials like silicon, gold, Al_2O_3 while designing. We analyzed them according to their Eigen frequency. Among all the materials silicon performs very well.

In this project we have used that sensor for real time application. This sensor works very well and it sends signal in milliseconds and it is helpful for deploying air bags in right time to save human lives.

From Table 1 we come to know that according eigen frequency value of 1.16 Hz it displaces 2.65×10^7 μm which leads to generation of voltage. When switch is pressed then proof mass is displaced. While displacing of proof mass then it develops capacitance in between plates which leads to generation of voltage. This voltage is considered as logic "HIGH", and according to code when it is high micro controllers recognizes it and proceed for further compilation of code i.e.: GPS and GSM activates and message is sent to recipient or saved number (Fig. 9).

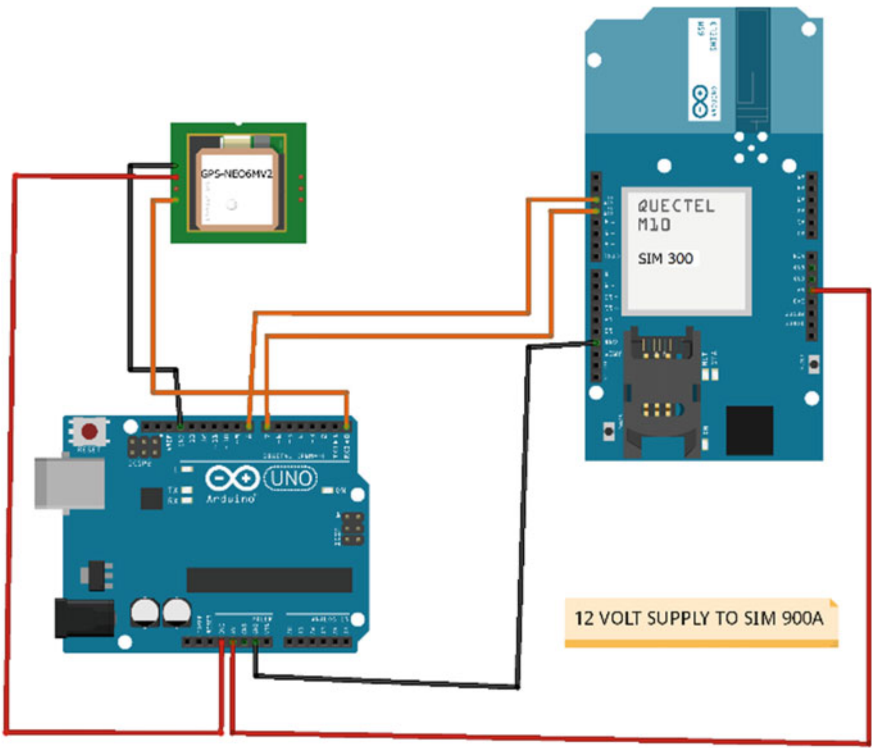


Fig. 9 Hardware connections

Crash Sensor to Arduino:

OUT to D10
VCC to VCC
GND to GND

GSM to Arduino:

Rx to 8
Tx to 7
GND to GND

For VCC using external voltage with the help of adapter
GPS to Arduino:

Rx to Tx
Tx to Rx
VCC to VCC
GND to GND
IV. Software

Arduino IDE (Integrated Development Environment) is a software platform used for writing code for hardware functionality of Arduino. Arduino is an open source software anyone can download it freely from website of Arduino.

Arduino IDE supports the languages c and c++. In this we write programs named as sketches. These sketches are written in text editor and saved as .ino file. This software is used to compile those written programs and upload those codes to hardware (Arduino). In this we can include library files for special hardware's and we have a serial monitor to display the output that was printed. In this there are two functions mainly, they are void setup () and void loop ().

In Arduino IDE avrdude is helpful to convert the executable code into a text file with hexadecimal encoding. This is loaded into Arduino board by a loader program in the board's firmware (Fig. 10).

In Arduino IDE first we include required libraries for perfect functioning of code. The libraries that we used in this project was TinyGPS.h used to get GPS location and SoftwareSerial.h used because in Arduino UNO R3 we have only one serial port but we need two serial ports. So, we take one port as software serial port and other one as hardware serial port. To get software serial port we used SoftwareSerial.h library.

Implementation of Libraries in Arduino IDE:

```
#include <TinyGPS.h>
#include <SoftwareSerial.h>
```

From crash sensor we need to take data because when it is pressed we have know about it so we need to check it continuously.

Fig. 10 Arduino IDE

```
int Led = 10; //define LED Interface
int buttonpin = 8; //define the collision sensor interface
int val; //define numeric variables val
void setup ()
{
pinMode (Led, OUTPUT); //define LED as output interface
pinMode (buttonpin, INPUT); //output interface defines the collision sensor
}
void loop ()
{
val = digitalRead (buttonpin); //digital interface will be assigned a value of 12 to
read val
if (val == HIGH) //When a collision sensor detects a signal, LED flashes
{digitalWrite (Led, HIGH);}
else
{digitalWrite (Led, LOW);}}
```

This is code for crash sensor when it is pressed then LED is ON condition. Similarly when it is pressed we need to get the GPS location and that data should be send to the saved number. To send location to saved number we need to place the preexisted code of GPS and GSM module in if and else conditions according to our requirement.

4 Result

After uploading code into the board then it runs continuously because of void loop() in code. When switch of crash sensor is pressed then it shows the data of latitude and longitude in serial monitor Fig. 9. With help of GSM we will send location data to the saved number then recipient receives data in form of url and it redirect to Google map with that location, same as Figs. 11 and 12.

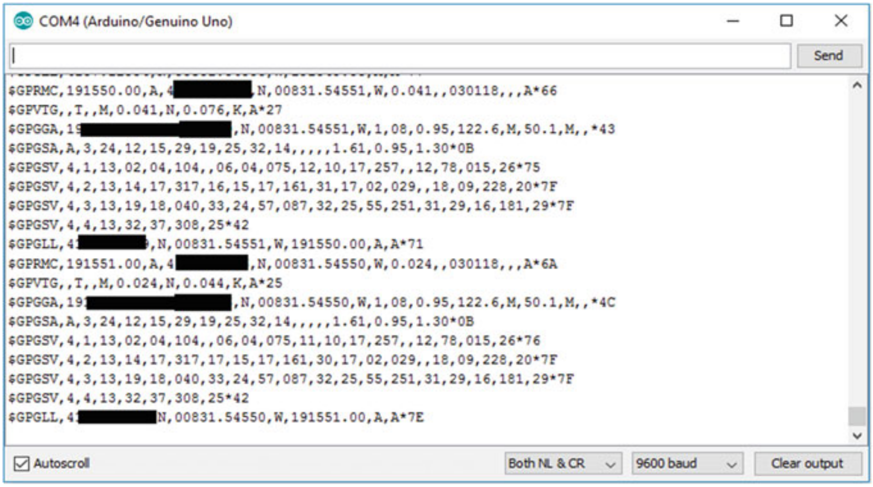
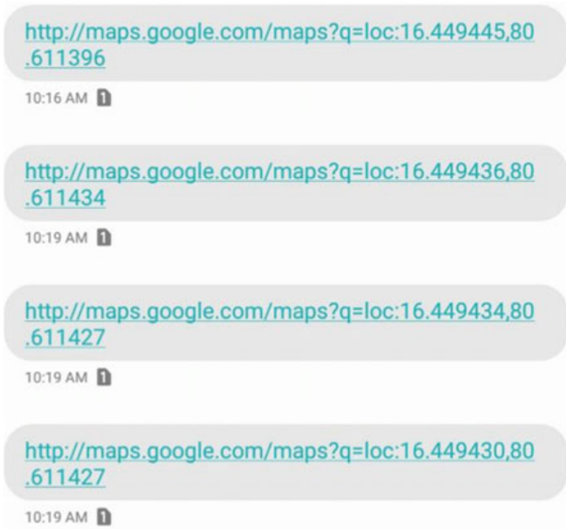


Fig. 11 Output in serial monitor

Fig. 12 Message send to saved number



In serial monitor we will get the latitude and longitude location. Here we got latitude as 16.449445 and longitude as 80.611396 (Fig. 13).

In this project code we make a URL link of Google maps which consist of location data. By pressing that link we will redirected to Google maps with that location.

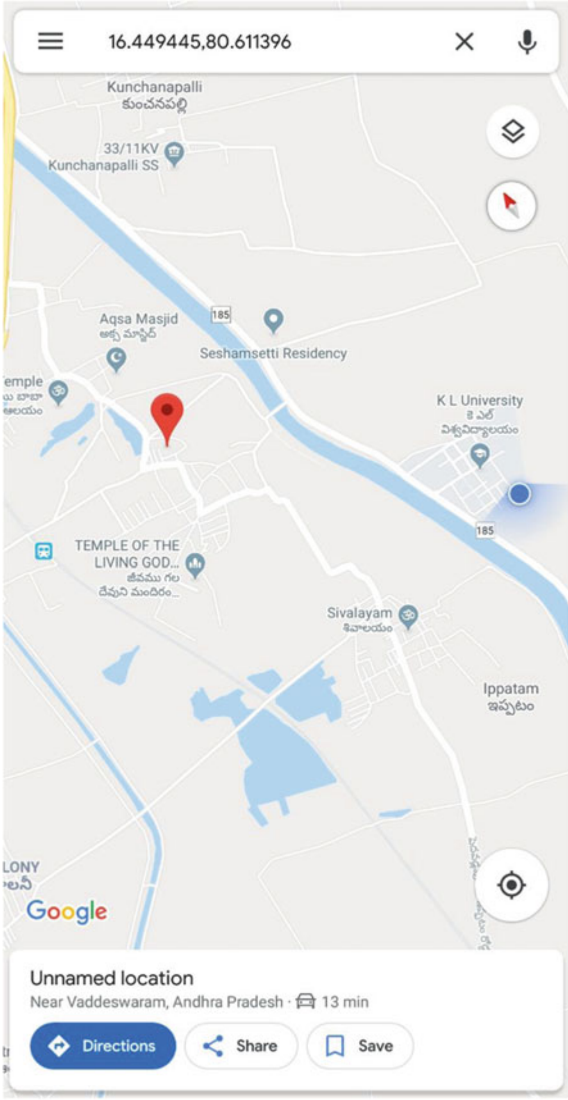


Fig. 13 URL redirected to Google maps

5 Conclusion

Using this hardware project we have advantages like, we can know the location of accident and we can respond to it very immediately and main advantage is we can save the lives of people who are injured while accident happens. Using this we can send emergency alert to the hospital or ambulance centers. Using this we can deploy airbags with in time and according to the severity of accident the crash sensor develops the voltage. According to voltage the micro controller send signal to airbag system then airbag system immediately responds to it and saves life. If deployment of airbags in wrong time or late it leads to heavy pay or loos of life to overcome such problem we used crash senor.

References

1. Shameem S (2018) Design and analysis of MEMS based capacitive pressure sensor for the detection of colon cancer. Int J Eng Technol 7(1.5):77–83
2. Kumar A, Madhu S (2015) A research review on airbag in automobile safety system. Int J Appl Eng Res 10(33):26815–26819. ISSN 0973-4562
3. Prabha C, Sunitha R, Anitha R (2014) Automatic vehicle accident detection and messaging system using GSM and GPS modem. Int J Adv Res Electr Electron Instrum Eng 3(7):10723–10727
4. Zaldivar J, Calafate CT, Cano JC, Manzoni P (2011) Providing accident detection in vehicular networks through OBD-II devices and android-based smartphone's. In: 5th IEEE workshop on user mobility and vehicular networks. IEEE, 978-1-61284-928-7, pp 813–819
5. Bhumkar SP, Deotare VV, Babar RV (2012) Accident avoidance and detection on highways. Int J Eng Trends Technol 3(2):247–252
6. Song G, Olmi C, Gu H (2007) An overweight vehicle–bridge collision monitoring system using piezoelectric transducers. IOP Publishing Ltd, pp 462–468. 0964-1726/07/020462
7. Bhumkar SP, Deotare VV, Babar RV (2012) Intelligent car system for accident prevention using ARM-7. Int J Emerg Technol Adv Eng 2(4):527–530. ISSN 2250-2459
8. Shameem S, Prasad GRK, Challa VT, Yaswanth M. Design simulation and analysis of crash sensor for air bag system
9. Sai G, Santosh K, Siva Kumar KPM, Prasad GRK, Srinivasrao K (2018) Design and simulation of spinning wheel type crash sensor for the airbag system in car. J Adv Res Dyn Control Syst 10(6-Special Issue):1419–1425
10. Stuetzler F-J, Century K (2000) Advanced frontal crash sensing with peripheral sensors. In: International body engineering conference Detroit, Michigan, 3–5 Oct 2000
11. Sara MRA, Abuseeni AA (2019) A web development system for producing an interactive electronic book. Int J Adv Trends Comput Sci Eng 8(1)
12. Rao SVRK (2018) Wireless sensor network based industrial automation using internet of things (IoT). Int J Adv Trends Comput Sci Eng 7(6):82–86

Implementation of Dynamic Decoders for Security Applications Using EG-LDPC Codes

J. Chinna Babu, C. Chinnapu Reddy and M. N. Giri Prasad

Abstract In this paper, the comparison of dynamic algorithms has been proposed and the advancements in low-density parity-check (LDPC) codes resulted in reducing the delay and decoding complexity. The Dynamic algorithms include Belief Propagation and Weighted Soft-Bit-flipping Decoders. This WSBF algorithm uses both the check relationships and the reliability of received message, therefore obtains a better decoding performance when compared with BF algorithm. WSBF offers a good trade-off between performance, complexity and speed. The main advantage of WSBF is the facts that, at the every iteration of the decoding process have weighted sums and the same values are computed using soft decision decoding. Hence these sums can be pre-processed, so that the speed advantages of WSBF are preserved. On the other hand, BP decoder displays the blunder amending frameworks predicated on PG (Protographic) LDPC codes. This paper extends and we develop protograph predicated Protographic less thickness equality scrutinize hubs through combining a progression of Variable Length coupled LDPC codes. Specifically, we demonstrate that equality of peeling interpreting (PD) and idea engendering for LDPC system of words across the parallel eradication channel is examined. Altering and abstracting the variable hubs (VNs) in each cycle with Belief Propagation. The decrementation of cancelled VNs provides the deciphering procedure and it is inferred with diminishing in involution and fast decoder. PG-LDPC codes are using the built up for BP decoder calculation. These LDPC codes are used in Signal Processing, Medial and Cyber security Applications. This proposed algorithm detects and corrects the number of errors that can be occurred in signal transmission and security applications.

J. Chinna Babu (✉)
JNTUA, Ananthapuramu, India
e-mail: jchinnababu@gmail.com

Department of ECE, AITS, Rajampet, India

C. Chinnapu Reddy
Department of ECE, Government Polytechnic, Adoni, India
e-mail: ccreddyece@gmail.com

M. N. Giri Prasad
Department of ECE, JNTUA, Ananthapuramu, India
e-mail: mahendragiri1960@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_41

Keywords Low density parity check codes · WSBF · SPD · PPD · Protographic-LDPC codes · Belief propagation decoder · Signal transmission · Security

1 Introduction

Low-density parity-check (LDPC) codes are Good error correcting codes and proposed by Gallager in 1962, and these codes are came into existence after 35 years. Low-density parity-check (LDPC) codes are a class of LBC codes. This name can be considered from the characteristics of the parity matrix which only having few number of one's than zeros. The major advantage of these codes is to provide a sophisticated performance which is very near to the channel capacity of different channels, which is also known as Shannon limit. LDPC codes are among the most central limit moving towards blunder redressing codes. The primary energy of LDPC codes lies in their pseudo-erratic nature, guaranteeing great properties for code, and the accessibility of low-multifaceted design disentangling calculations. As of late, a lot of research has been dedicated to the plan of LDPC codes with short to relieve square lengths, which compare to a large portion of the application situations of these codes in remote gauges. LDPC codes are categorized into two types and the categorized types are binary LDPC codes and NB-LDPC codes [1–4].

Binary LDPC codes are said to be uneven LDPC codes as their column and row weights are dissimilar. NB-LDPC codes are said to be Non-Binary LDPC codes and these are having all the rows and columns have equal weights. Hence these are known to be usual LDPC codes.

A. Construction of Binary LDPC Codes

1. Calculate the primeval polynomial
2. Calculate the degree terms
3. Galois addition/multiplication
4. Calculate the generator polynomial G(x)
5. From G(x), determine H-matrix.

LDPC codes are described using two types of representations. They are Matrix and Tanner graph forms. Let us consider one of the examples of a LDPC parity check matrix with (n, k) representations are (8, 4). Where we can define 'n' as total number of bits and 'k' can be represented as the number of message bits. The weights of this given parity matrix are represented using Wr, Wc. Depends on these two parameters, we may increase or decrease the weights of the parity matrix. The conditions to satisfy for adding the weights are Wc $\ll$ n and Wr $\gg$ m. Where Wc is number of ones in each and every column and Wr is the number of ones in each and every row [5, 6] (Fig. 1).

Tanner introduced Tanner graph, which may contains Check nodes and Variable nodes. These nodes are connected together in graphical representation form. This is bipartite graph, the connection was established in between variable nodes and check

Fig. 1 Matrix representation of LDPC codes

$$H=\begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix}$$

nodes depends on the number of one's present in the parity matrix. Where Variable nodes can be indicated by V_nodes and Check nodes are indicated as C_nodes. Fig. 2 is the representation of the Tanner graph, and the connections from f0 to c1, c3, c4, c7 can be established depends on the number of ones in the first row of the parity matrix. This Tanner graph consists of m number of C_nodes and followed by 'n' V_ nodes. C_nodes fi is connected to variable node Cj if the element of Hij is a 1 [7–10].

In general case, two types of decoding schemes are there for the decoding of LDPC codes. They are Soft-Decision-Decoding and Hard-Decision-Decoding techniques. In HDD, the code word which is received compared with all the remaining code words which results the Hamming Distance which is of minimum value. In HDD, the code word which is received compared with all the remaining possible code words which results Euclidean Distance which is of minimum value. Therefore, we may conclude that SDD improves the reliability of the LDPC codes by properly improving the decision making [11–14].

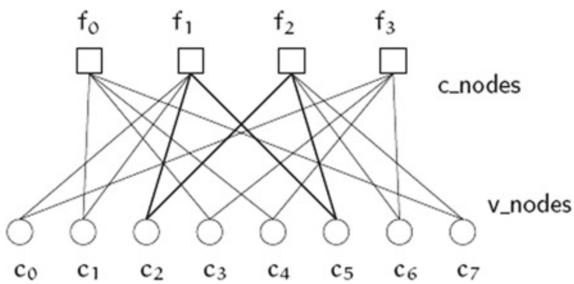
B. PG-LDPC Code Construction

Consider transmission of an arrangement of L code words V1, V2, …, VL of length n. In traditional piece coding, each of these code words is encoded freely by betokens of some given code. Gathering, for instance, a rate R = n * k LDPC code characterized through an equality check grid H, these system of words need to satisfy the condition

$$V_T \cdot H_T = 0 \text{ for all } t = 1 \text{ to } L.$$

This graph is developing a coupling of four LDPC codes in a solitary chain. Here every one of the hubs is associated with mix of check hubs (CNs) to variable hubs (VNs) and then the message are passed from CNs to VNs. Protographic LDPC codes are presenting the two sorts of developments, such as (1) Desultorily built

Fig. 2 Pictorial demonstration of parity check matrix



Protographic-LDPC code and (2) Protograph development. In the wake of characterizing the BEC, we characterize the remaining chart degree appropriation after transmission (Fig. 3).

1. Discretionarily built (l, r, L) Protographic-LDPC Codes

Here it considers the two course VNs and one square CNs. In this upper VNs are bits in first course and lower VNs are bits in second flow. In here all VNs are associated with pertinent CNs of diagram development. In this chart development each VNs consider the 3 edges VN degree and each check hub comprises the 4 edges of CN degree and the aggregate length is L = 4. This is bad diagram by the translating procedure, why since it experiences gigantically enormous unpredictability caused by a sizably voluminous arrangement of handling units and mind boggling interconnection (Fig. 4).

2. Protograph development (l, r, L) codes

A protograph is bipartite chart characterized through an arrangement of VNs and an arrangement of CNs and set of edges. Protographs are spoken to by using bi-nearness lattice B called base framework (Fig. 5).

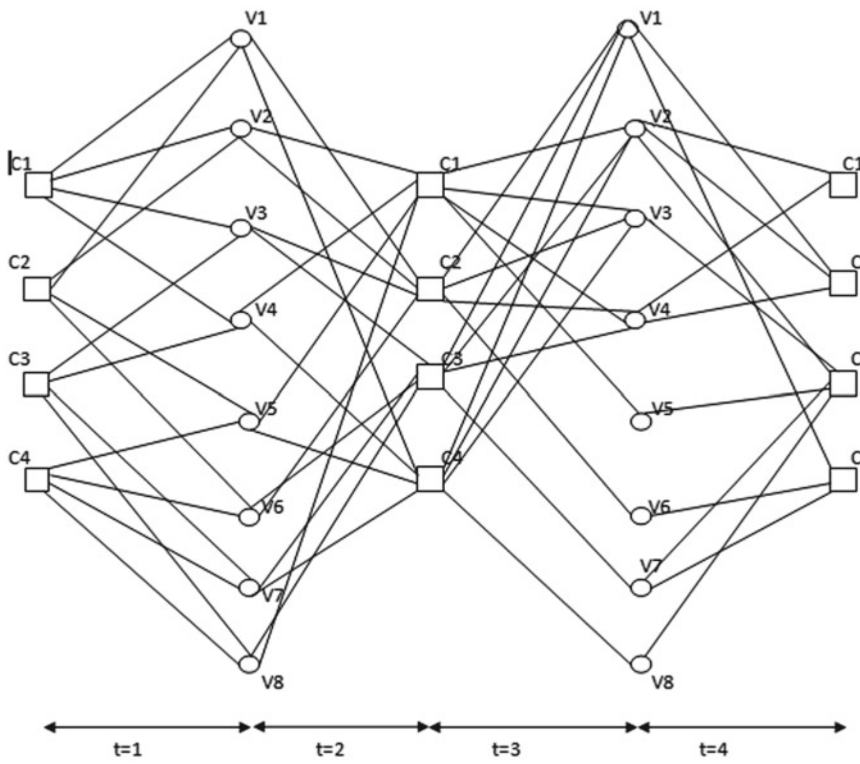


Fig. 3 Development of SC-LDPC codes by using coupled protographs

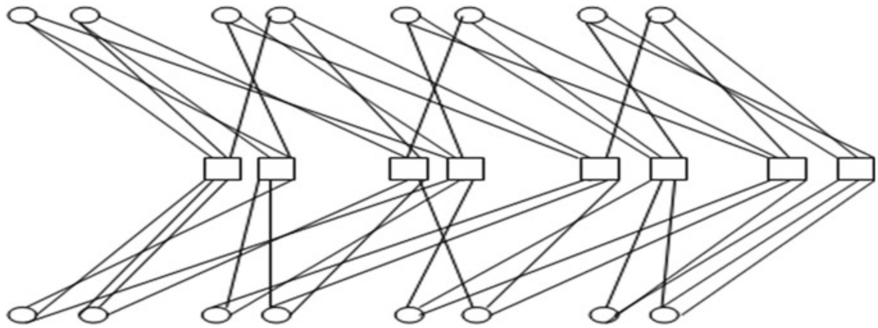


Fig. 4 Example construction of (l, r, L) = (3, 6, 4) coupled diagram

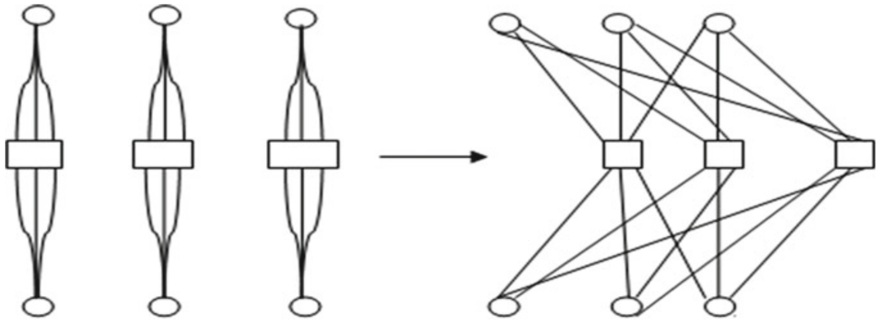


Fig. 5 Case development of the (l, r, L) = (3, 6, 3) coupled protograph

In this protograph development each VN consider the 3 edges of VNs degree and each CN comprises the 6 edges of CV degree. Protographs are then associated with a (l, r, L) coupled as appeared in Fig. 3, for L = 3.

2 Decoding Algorithms

In general scenario of decoding algorithms, several numbers of decoding algorithms has been developed for decoding binary LDPC and NB-binary LDPC codes based on soft and hard decision decoding methods. The standard block diagram of decoding algorithms is as given in Fig. 6.

The following metrics are included in the basic procedure for decoding LDPC codes.

1. Check nodes place certain limits on the nodes of the function.
2. If these constraints are met by the variable nodes, the input code is believed to be valid to the decoder and transmits the information to the processor.

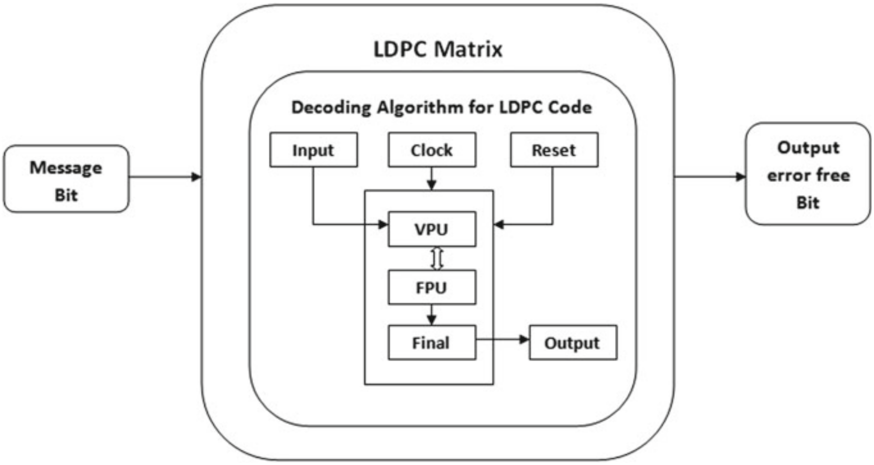


Fig. 6 Standard block illustration of decoding algorithm used for LDPC codes

3. Using the syndrome, the error is detected and the received code word is corrected and then passed to the output node.

The decoding of LDPC codes is an iterative process. The algorithms considered in this paper for comparison are Weighted Soft-Bit Flipping (WSBF) Algorithm, Belief Propagation Decoding Algorithm, which uses both Parallel Peeling and Sequential Peeling Decoding algorithms [15].

A. Weighted Soft-Bit-Flipping Algorithm (WSBF)

Weighted Soft Bit Flipping algorithm is the most efficient LDPC decoding algorithm with good performance of error and less complexity of hardware. The following steps provide the decoding procedure for LDPC codes using the WSBF algorithm.

(1) Parity check sums are the constraints of check nodes that depend on the weights of rows and columns to be satisfied by variable nodes. The resulting bits are called Fi syndrome bits, all computed syndrome bits are zero, then all parity check sum equations are satisfied, and then the subsequent cycle of decoding is stopped.
(2) The resulting syndrome bits are detected for each and every bit code position, referred to as Fi. Where i = 0 to n − 1 is concerned.
(3) Notify the array of bits with a total Fi.
(4) The bits in the album package should be included.
(5) Steps from 1 to 4 should be rotated until the whole parts of the condition are zero. The iterations will be stopped at this point (Fig. 7).

Figure 4 shows the serial hybrid decoder of WSBF in this paper. The decoder consists of buffer input (which holds the code bits), a rollback buffer (which holds the product of flipped bits), a buffer output (holds the error-free code bits), check

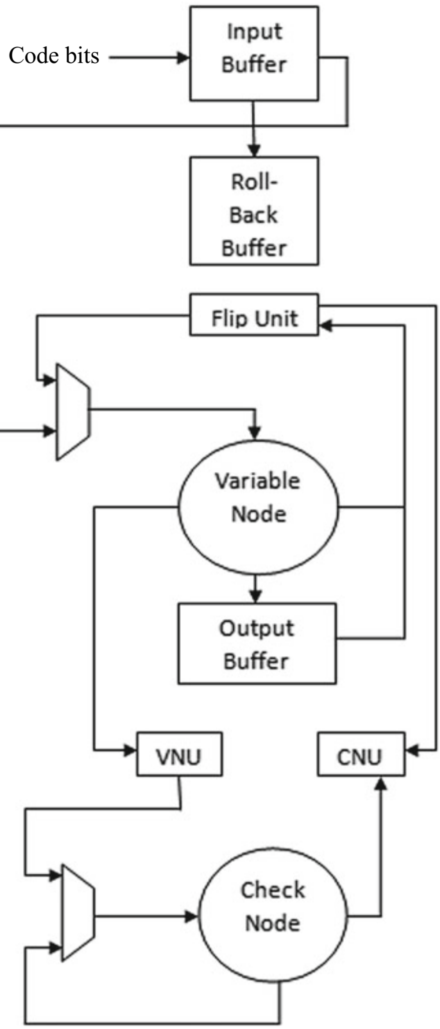


Fig. 7 Block schematic of WSBF decoder

nodes, variable nodes, a flip unit (flips the bits when the constraints are not met), a variable node processing unit (VNU) and a check node processing unit (CNU).

B. Belief Propagation Decoding Algorithm

Consider transmission over a BEC with eradication likelihood. In this area, consecutive peeling disentangling (SPD), PPD and BPD are presented. The figures SPD and PPD as far as the messages sent from variable hubs to check hubs into front missive, if scrutinize hub is associated with lingering chart. Then rearward messages if all different variable hubs associated with the particular check hubs are now settled.

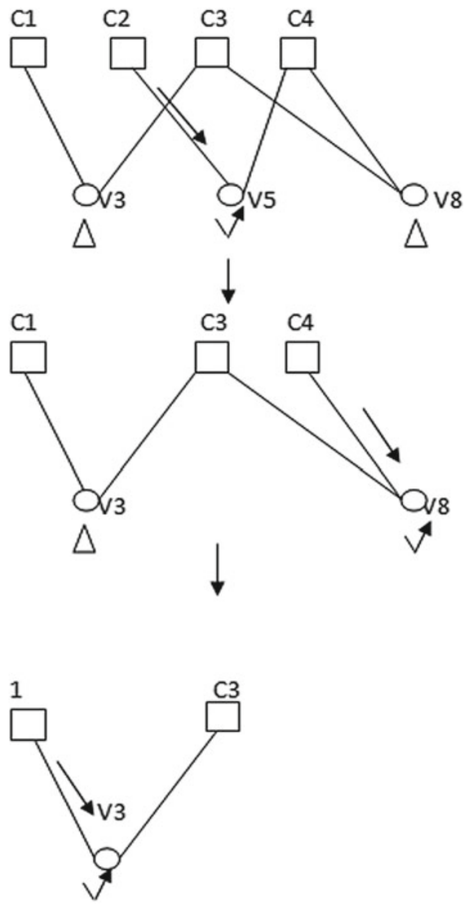
1. Sequential Peeling Decoder

All kenned VNs of the chart and their associated edges are dreamy to get the remaining diagram. A deg-1 CN is a CN of this lingering diagram with just a single associated obscure VN. In each emphasis, a solitary deg-1 CN c is separated and the associated VN v is settled. Figure 8 represents the SPD for the graph which is residual is as shown in Fig. 9. Since at every iteration the deg-1 CN to be resolved would picked up in random manner, and the sequence of these residual graphs for several decoding realizations may differ for a given transmission realization [16–23].

2. Parallel Peeling Decoder

It is a one of iterative message passing algorithm. This decoding algorithm also working as the same as SPD, but in this decoder choose the all available degree-1 check hubs per every repetition of sequence of the process and then all grade-1 check hubs are settled.

Fig. 8 Decoding of Sequential Peeling Decoder



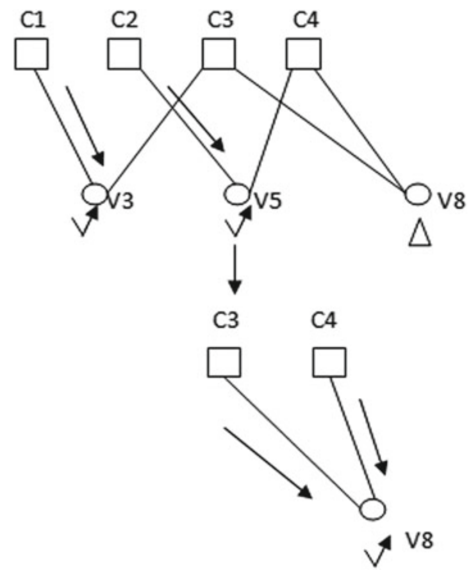


Fig. 9 Decoding of Parallel Peeling Decoder

In Fig. 9 it describes the PPD iterations for the graph which is residual. Where the PPD is deterministic, from a given transmission realization, all the available deg-1 Check Nodes are resolved in each and every step and thus the sequence of graphs which are residual does not differ in large extent [7, 24–28].

3. Belief Propagation Decoder

BPD is also one of the finest decoding algorithms. The BPD is defined by in an iterative process in which neighbouring VNs are interacting to each other. In BP, the messages are translate from both directions. Here the messages are transmits the variable hubs to check hubs into forward direction and back again to transmitted the messages in rearward direction in every repetition of the decoding process represented in Fig. 8. The CN work is indistinguishable to PPD. In the case of BP decoder each variable hub to any adjoining check hubs depends just on messages gotten from all another neighbouring check hubs. Here the messages sent in both ways can vary. BP for Fig. 5 is delineated in Fig. 8. Since every single other message are settled, we just demonstrate the messages in the staying lingering diagram [29–31] (Fig. 10).

In this section, we discussed methodology and simulation results of decoding algorithms are presented. The proposed research study, WSBF decoding algorithm results are compared with existing decoding algorithms in terms of latency, hardware utilization or complexity and power consumption. The following three metrics are used to measure the decoding latency. The simulation values are obtained from Xilinx Synthesis report and SPARTAN 3e. The resultant values are tabulated in Table 1.

The performance metric latency is computed using Eq. (1).

$$Latency = CC * CP \tag{1}$$

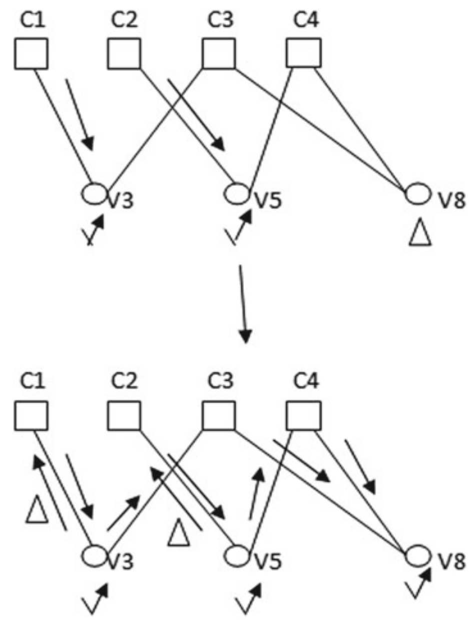


Fig. 10 Decoding of belief propagation

Table 1 Comparison of dynamic decoding algorithms

Logic utilization	WSBF	BPD	PPD	SPD
Usage of hardware components	Moderate	Moderate	Low	High
Number of slices	46	54	56	120
Number of slice flip flops	16	65	65	129
Number of 4-input LUTs	94	97	101	213
Number of bonded IOBs	53	195	251	387

where

CC = Total number of Clock Cycles to obtain output
CP = Minimum required Clock Period.

The total number of clock cycles referred to as CC is computed from Eq. (2).

$$CC = \frac{Time\ period\ needed\ for\ decoding\ the\ output}{Time\ period\ Interval} \tag{2}$$

It is evident from Tables 1, 2, and 3 and Fig. 11; the proposed WSBF took less latency to produce decoded data. The proposed research study, WSBF decoding algorithm has taken 778 ns time to decoding the data transmission over a communication channel comparing with other decoding algorithms.

Table 2 Simulation parameters versus decoding algorithms

Parameters	Algorithms					
	SBF	BPD	SPD	PPD	MLDD	WSBF
Minimum clock period (in ns)	29.61	23.64	24.34	21.38	26.49	22.23
Total required cycles to obtain output	36	37	39	41	34	35
Time interval (in ns)	200	200	200	200	200	200

Table 3 Latency performance metric over decoding algorithms

Performance metric	Algorithms					
	SBF	BPD	SPD	PPD	MLDD	WSBF
Latency (ns)	1065.9	874.68	949.26	876.58	900.66	778.00

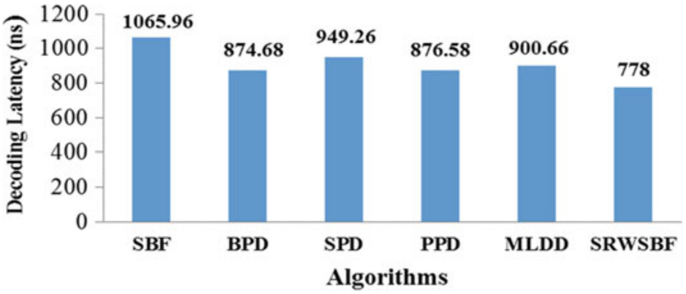


Fig. 11 Decoding latency comparison over various algorithms

From Table 3, it is observed that WSBF algorithm reduced 27.01% of decoding latency comparing with SBF algorithm, 11.05% comparing with BPD, 18.04% comparing with SPD, 11.24% comparing with PPD and 13.62% when comparing with MLDD algorithms. Hence, the proposed research study is reduced the decoding latency over naïve decoding algorithms such as SBF, SPD, PPD, BPD and MLDD.

3 Conclusion

In many fields such as Medical Applications, Signal Processing Applications, Barcode Generations and Cyber Security applications, the strong error correction capabilities of Geometric LDPC codes are gaining more popularity. As the technology progresses, several decoding algorithms have been created for the LDPC codes. In this paper the Weighted Soft Bit Flipping (WSBF) algorithm, Belief Propagation Decoding. It is concluded from these checks that the SPD algorithm's hardware complexity exceeds PPD, BPD, and WSBF algorithms. WSBF decoder have less device hardware and less latency than other two components. Whereas for the WSBF algorithm the delay is less and followed by BPD, PPD and SPD decoders. The best decoder can be selected according to the desired application, taking into account the hardware complexity and speed. We have suggested SBF, MLDD algorithms in our previous work [31] and these algorithms are very complex compared to WSBF and BP decoders.

References

1. Gallager RG (1962) Low density parity check codes. IRE Trans Inf Theory IT-8:21–28
2. MacKay DJC, Neal RM (1996) Near Shannon limit performance of low density parity check codes. Electron Lett 32(18):1645–1646
3. MacKay DJC (1997) Good error correcting codes based on very sparse matrices. IEEE Trans Inf Theory 45:399–431
4. MacKay DJC (1998) Gallager codes that are better than turbo codes. In: Proceedings of 36th Allerton conference on communication, control, and computing, 1998
5. Kou Y, Lin S, Fossorier M (1999) Low density parity check code based on finite geometries: a rediscovery and more. IEEE Trans Inf Theory
6. Fossorier MPC, Mihaljevic M, Imai H (1999) Reduced complexity iterative decoding of low-density parity check codes based on belief propagation. IEEE Trans Commun 47(5):673–680
7. Kschischang FR, Frey BJ, Loeliger H-A (2001) Factor graphs and the sum–product algorithm. IEEE Trans Inf Theory 47(2):498–519
8. Zhang J, Fossorier MPC (2004) A modified weighted bit-flipping decoding of low-density parity-check codes. IEEE Commun Lett 8(3):165–167
9. Chen J, Dholakia A, Eleftheriou E, Fossorier MPC, Hu X-Y (2005) Reduced-complexity decoding of LDPC codes. IEEE Trans Commun 53(8):1288–1299
10. Palanki R, Fossorier MPC, Yedidia JS (2007) Iterative decoding of multiple-step majority logic decodable codes. IEEE Trans Commun 55(6):1099–1102
11. Savin V (2008) Min-Max decoding for nonbinary LDPC codes. In: Proceedings of IEEE international symposium on information theory (ISIT), 2008, pp 960–964
12. Cho J, Sung W (2009) High-performance and low-complexity decoding of high-weight LDPC codes. J Korea Inf Commun Soc 34(5):498–504 (in Korean)
13. Zhang C, Wang Z, Sha J, Li L, Lin J (2010) Flexible LDPC decoder design for multi-Gb/s applications. IEEE Trans Circuits Syst I Reg Paper 57(1):116–124
14. Zhang X, Siegel PH (2012) Quantized min-sum decoders with low error floor for LDPC codes. In: IEEE international symposium on information theory, Jul 2012, pp 2871–2875
15. Chinna Babu J, Chinnapu Reddy C, Giri Prasad MN (2016) Comparison of various technologies for the implementation of SBF decoder for geometric LDPC codes. INDJST J 9(30):1–9
16. Lentmaier M, Sridharan A, Costello DJ, Zigangirov KS (2010) Iterative decoding threshold analysis for LDPC convolution codes. IEEE Trans Inf Theory 56(10):5274–5289
17. Mitchell DGM, Pusane AE, Lentmaier M, Costello DJ (2011) Exact free distance and trapping set growth rates for LDPC convolution codes. In: IEEE international symposium on information theory (ISIT), 2011, pp 1096–1100
18. Mitchell DGM, Lentmaier M, Costello DJ (2015) Spatially coupled LDPC codes constructed from protographs. IEEE Trans Inf Theory 61(9):4866–4889
19. Kasai K, Sakaniwa K (2011) Spatially-coupled MacKay-Neal codes and Hsu-Anastasopoulos codes. In: Proceedings of IEEE international symposium on information theory, St. Petersburg, Russia, 2011
20. Stinner M, Olmos PM (2015) On the waterfall performance of finite-length SC-LDPC codes constructed from protographs. IEEE J Sel Areas Commun PP(99):1

21. Ezri J, Montanari A, Urbanke R (2007) A generalization of the finite-length scaling approach beyond the BEC. In: IEEE international symposium on information theory (ISIT), June 2007, pp 1011–1015
22. Ezri J, Montanari A, Oh S, Urbanke RL (2008) The slope scaling parameter for general channels, decoders, and ensembles. In: IEEE international symposium on information theory (ISIT), July 2008, pp 1443–1447
23. Thorpe J (2013) Low-density parity-check (LDPC) codes constructed from protographs. Tech. Rep., JPL IPN
24. Richardson TJ, Urbanke RL (2008) Modern coding theory. Cambridge University Press, New York, NY
25. Di C, Proietti D, Telatar IE, Richardson TJ, Urbanke RL (2002) Finite-length analysis of low-density parity-check codes on the binary erasure channel. IEEE Trans Inf Theory 48(6):1570–1579
26. Aref V, Schmalen L, Ten Brink S (2013) On the convergence speed of spatially coupled LDPC ensembles. In: Allerton conference on communication, control, and computing, 2013, pp 342–349
27. Gardiner C (2009) Stochastic methods: a handbook for the natural and social sciences. Springer, Berlin, Heidelberg
28. Karatzas I, Shreve SE (1991) Brownian motion and stochastic calculus, 2nd edn. Springer, New York
29. Gillespie DT (1996) Exact numerical simulation of the Ornstein-Uhlenbeck process and its integral. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics 54(2):2084–2091
30. Nobile AG, Ricciardi LM, Sacerdote L (1985) Exponential trends of Ornstein-Uhlenbeck first-passage-time densities. J Appl Probab 22(2):360–369
31. Chinna Babu J (2017) VLSI implementation of decoding algorithms for EG-LDPC codes. Procedia Comput Sci 115(3):143–150

Low Power Mixed Signal with CVNS Based Neuron Chip

Syed Shameem, P. Kanakaraja, N. L. S. P. Sairam, Sameer Kalyan, Govind Krishna, Anil Kumar and K. Hari Kishore

Abstract The usage of a CVNS neuron of blended flag of low power and configuration is introduced. The straight guess by parts in the simple area is utilized for the actuation work. The usage of the simple neural system has a constrained accuracy, so this neuron is proposed. Less simple digits of the nonstop esteem numerical framework (CVNS) are supplanted by general advanced yields of a sigmoid neuron, while in the meantime the most extreme methodology blunder continues as before as computerized structures. The execution of ASIC and the learning of chips for fitting neuronal chips is the aftereffect of this proposed CVNS neuron. Utilizing circuits of current mode, the VLSI usage of the neuron will be completed. The consequences of the usage contrast positively and the structures recently created regarding zone, postponement and vitality utilization.

Keywords Continuous valued number system (CVNS) · ASIC · Neural networks · Neuron chips

1 Introduction

CVNS represents Numeric System of Continuous Value. This numerical framework is a simple number framework with various simple digits that is reasonable for actualizing high exactness simple bits and blended flag circuits [1].

The low-control blended flag is a blended flag incorporated circuit: any IC that has both simple circuits and advanced circuits in a solitary semiconductor pass on. Toward the day's end, applications with blended flag structures are all over the place, for instance, a keen versatile. Regardless, it is progressively exact to call them blended flag frameworks [1, 2]. Blended flag ICs likewise process simple and computerized flags together. For instance, a simple to computerized converter is a blended flag circuit. Circuits or frameworks are frequently beneficial responses for the development of any cutting edge equipment application for the shopper [3, 4].

S. Shameem (✉) · P. Kanakaraja · N. L. S. P. Sairam · S. Kalyan · G. Krishna · A. Kumar · K. Hari Kishore
Department of ECE, Koneru Lakshmaiah Educational Foundation, Vaddeswaram, Guntur, India
e-mail: shameemsyed@yahoo.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_42

A neural system is a movement of counts that attempts to see major associations in an arrangement of data through a strategy that duplicates the manner in which the human cerebrum works. Neural frameworks can be adjusted to evolving data, so the framework makes the best outcome without holding on to refresh the execution criteria. The starting point of neuronal frameworks is quickly picking up unmistakable quality in the trading of enhancements to the structure [5, 6].

A counterfeit neuron is a scientific capacity considered as a model of Biological neurons. The exchange capacities generally have a sigmoid shape, however they may likewise appear as other non-linear functions, piecewise linear capacities, or step capacities. CVNS is used to meet the high accuracy and before going deep into this number system we have to discuss about the Approximation systems they are: Piecewise Linear approximation, Piecewise Non-Linear approximation, Bit level Mapping, Lookup Table (LUT) and Hybrid Method [7, 8].

The above are the approximation methods that have some errors, some of them are old methodologies. These methodologies are not giving exact performance and consumption of power is uneven. Some of the methods are not so much well developed.

The neurons are of two types they are Analog Neurons and Digital Neurons. These neurons have lower consumption area and limited accuracy but it cannot meet the low approximation error requirements. These neurons have better power consumption but it occupies more area and high cost. As these neurons cannot meet the certain characteristics so CVNS methods are proposed for the better outputs (Fig. 1).

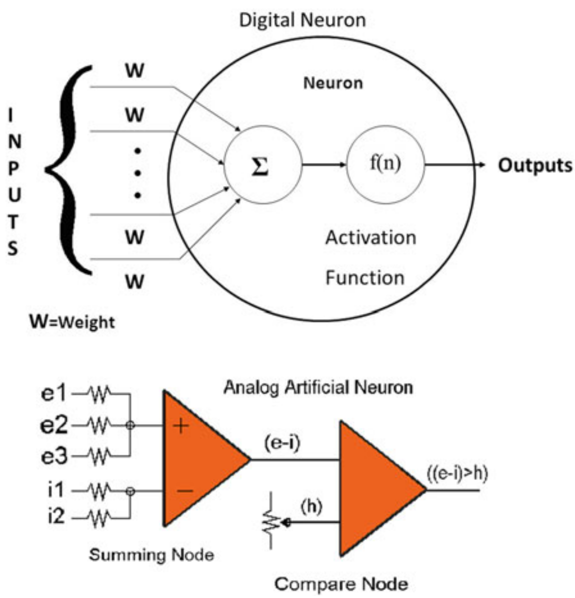


Fig. 1 Analog neuron and digital neuron

CVNS Analog Digits
This arrangement sets the required number if the CVNS simple digits for the sigmoid neuron. The CVNS portrayal of the sigmoid capacity is upgraded to make a proficient region plan. One methodology is to diminish the quantity of CVNS simple digits in a set, contingent upon the necessities of the framework.

CVNS Networks
In a CVNS neural system, the information sources are first changed over to CVNS. At that point, the CVNS esteems that speak to the sections are increased by the loads put away in the neurotransmitter registers. This should be possible through the created CVNS augmentation calculation and the multiplier yields are included. The expansion is made through the CVNS expansion calculation. This calculation can give the contributions to the proposed structure in the twofold and CVNS position. Subsequently, as long as the proposed structure is utilized in a CVNS neural system, no paired is required for CVNS or CVNS for parallel changes.

CVNS Sigmoid
The VLSI execution of the assessment of the proposed CVNS work is done utilizing the present circuit mode circuits. Neuron comprises of four essential units that incorporate an information go decoder, a present decoder, a present generator, a sigmoid estimate and a yield task. The info extend decoder recognizes the information go, while the present generator gives the signs required to the yield distribution and sigmoid methodology units.

2 Theoretical Analysis

A. *CVNS Arithmetic*

CVNS means Continuous number value system and it is represented in the form of analog number system to produce greater efficiency and lower space occupation of circuits.

The supreme estimation of a genuine number X utilizing a settled point parallel number framework organization can be appeared as

$$x = \sum_{i=-Nf}^{Ni-1} xi * 2^i$$

CVNS analog of X radix-2 for is as shown below:

$$x = \sum_{i=-Nf}^{Ni-1} xi * 2^{i}$$

$$((\boldsymbol{x})) = (\boldsymbol{x} * 2^{-\boldsymbol{m}})$$

where $\boldsymbol{m}$ is the index of the CVNS analog digits.

$-\boldsymbol{Nf} \le \boldsymbol{m} \le \boldsymbol{Ni}$ and mod 2 are the continuous modular reduction operation.

CVNS digits share data, not the majority of the digits in a set are required for number-arithmetic and computing. The architect can profit by this adaptability to decrease the region and power.

Both of these functions are implemented in CVNS system

$$((x))m = \sum_{i=m-\phi+1}^{m} xi * 2^{i-m}$$

where ϕ is environmental resolution.

The conversion of binary digits is as below (Fig. 2):

$$\boldsymbol{xm} = \boldsymbol{1}((\boldsymbol{x}))\boldsymbol{m} \ge \boldsymbol{1}\ \boldsymbol{and}\ \boldsymbol{xm} = \boldsymbol{0}\ \boldsymbol{only}((\boldsymbol{x}))\boldsymbol{m} < 1$$

B. *CVNS Neural Network*

Prior to experiencing the numerical inference of the proposed sigmoid neuron three essential properties that will be utilized these highlights will be considered as properties one, two and three in whatever remains of this segment:

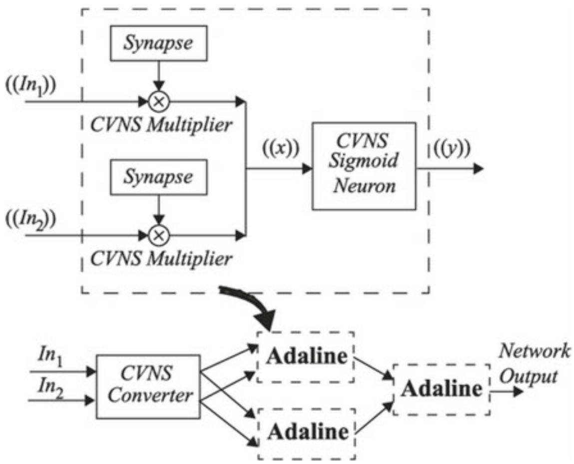


Fig. 2 The gate level configuration of a 2-2-1

1. According to discussion of the equation which is based on area and power minimum value represent table by each CVNS digit is equal to $2 - (\phi - 1)$.
2. Here value of ϕ is 4 and other values like (negative) are neglected so it shown as

$$\boldsymbol{i \le -4 \Rightarrow 2^i xj = 0 \; and \; mod \; 2 = 0}$$

   where $\boldsymbol{i}$ and $\boldsymbol{j}$ are the integer values and $\boldsymbol{xj}$ can assume values 0 or 1.
3. Using the definition of continues modular reduction this is expressed as

$$\boldsymbol{i \ge 0 \Rightarrow 2^i xj \; and \; mod \; 2 = 0}$$

C. *CVNS Neurons*

In a CVNS neural framework, the passages are changed first in the CVNS. Around then, the CVNS gauges speaking to the passages are duplicated by the put away loads and these are put away in the neural connection registers. This can be tended to through the CVNS increase figuring made. The multiplier returns are incorporated. The development is affirmed through the count of the CVNS extension. This figuring can offer commitments to the proposed structure in both the CVNS and the paired design. Along these lines, since the proposed structure is utilized in a CVNS neural framework, it doesn't require any paired for CVNS or CVNS for parallel changes.

D. *VLSI implementation of the CVNS sigmoid Neuron*

In this fragment, the proposed structure and VLSI execution of the sigmoid neuron dependent on CVNS are analyzed. Since in the present mode circuits, the extension is performed without exertion, the VLSI execution of the proposed CVNS work assessment is completely utilized in the present mode circuits. The square form of the proposed neural structure is made out of four fundamental units that incorporate an info run decoder, a present generator, a sigmoid estimation and a yield allotment unit. The data go decoder recognizes the data that is executed, while the present generator gives the normal signs to the sigmoid methodology and yield units. Since there are four subunits, the sigmoid estimation unit of four subunits is built up. It ought to be noticed that the subunits must be actuated by the flag, which is started by the yield signs of the info r1–r4. The yield unit assigns the yield contingent upon the information signals. On the off chance that the section is sure, it passes your data without changes. Something different, a subtraction is finished. The action of each square is cleared up in detail (Fig. 3).

E. *Input Range Decoder*

Four signals r1, r2, r3, and r4 are utilized to translate the input range. Truth table is shown in Fig. 4.

F. *Current Generator*

With the last goal of executing the sigmoid capacity that relies upon the relations made in consistent current streams, it is required. The primary structure square of the

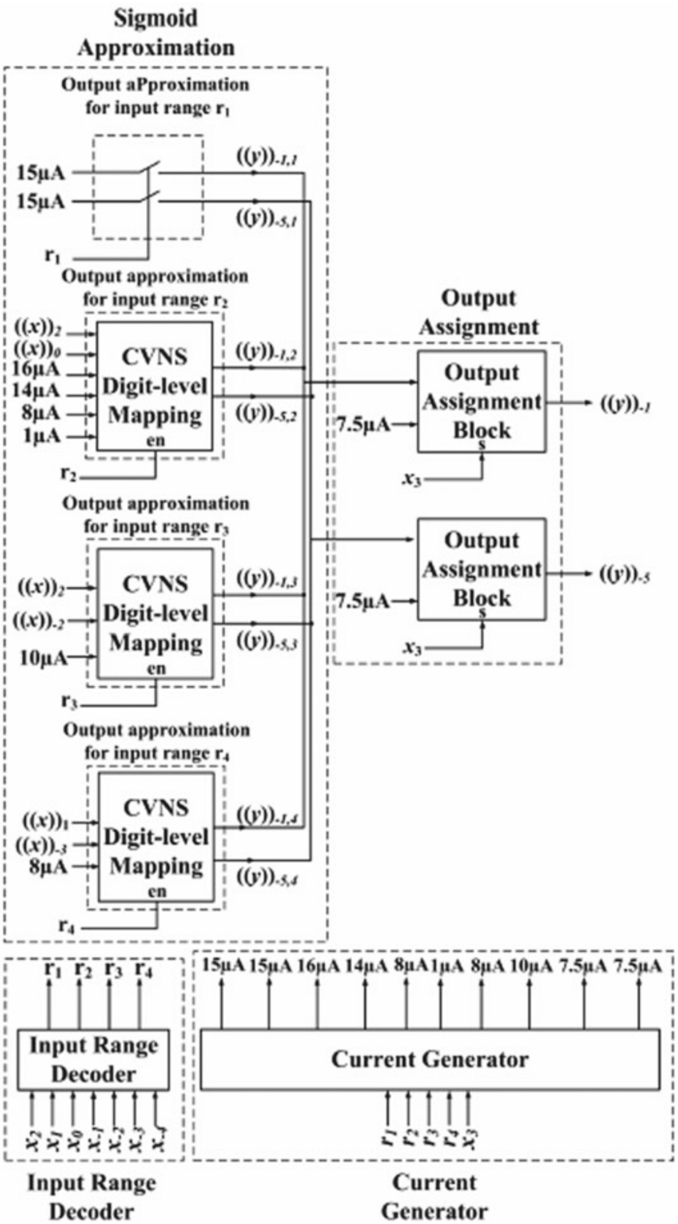


Fig. 3 Block diagram of evolution structure

Input range	r_1	r_2	r_3	r_4
$x \geq 5$	0	1	1	1
$2.375 \leq x < 5$	1	0	1	1
$1 \leq x < 2.375$	1	1	0	1
$0 \leq x < 1$	1	1	1	0

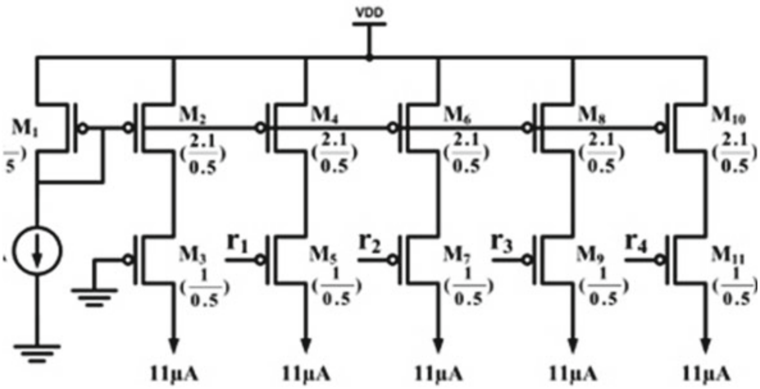


Fig. 4 Current reference circuit with various (w/l) ratios

present generator unit is appeared in Fig. 4. By properly estimating the transistors, a current of 11 μA is sent to transistors M1 to M11. The current of 11 μA sent to transistors M2 and M3 is utilized for the vestige of the present flag required for the yield unit. The current is sent to transistors M4–M11 since the comparing ri flag is dynamic. Along these lines, contingent upon the info extend, the current of 11 μA will be made. In this manner, the current replicated to the M4–M11 transistors is utilized to make the steady current signs required for four subunits of the sigmoid methodology unit. The age of a few current characteristics is conveyed utilizing current mirrors. The present characteristics made by the present generator are for a few scopes of data appeared as follows. It ought to be noticed that the yield unit utilizes the 7.5 μA made in all info ranges (Table 1).

Table 1 Input range and generated current (μA)

Input range	Generated current (μA)
$x \geq 5$	7.5, 15
$2.375 \leq x < 5$	1, 7.5, 8, 14, 16
$1 \leq x < 2.375$	7.5, 10
$0 \leq x < 1$	7.5, 8

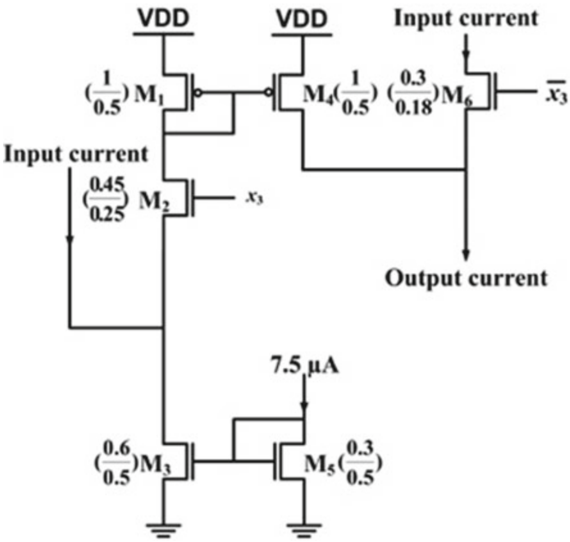


Fig. 5 Assignment block

G. *Output Assignment Block*

The yield portion square doles out the yield contingent upon the info flag. In the slim chance that the data is valid, the yield of the sigmoid methodology unit is passed explicitly to the exit (Fig. 5).

The proposed structure utilizes the features of CVNS to run a gainful straightforward sigmoid neuron. While applying CVNS number-crunching, the neuron requires fewer yield digits with a lower region utilization. What's more, to expand the productivity of the zone, circuits in current mode were utilized. Since the VLSI execution of the proposed structure is utilized altogether in simple circuits, process assortments and cross-connecting impact their execution. The execution of the VLSI usage of the proposed CVNS sigmoid neuron was confirmed by ensuing Monte Carlo configuration plan reenactments for various contributions inside the scope of (−8, 8).

The reenactment results after the plan of the impacts of the neuron in four example input appraisals of 0.875, 1.875, 3.625 and 7.9375, where each information is inside the expansions r4, r3, r2 and r1, individually. These data sources are associated at interims of 10 ns. The ordinary yield approaches identified with the associated data sources. The recreational outcomes after the organization completely match with the logical surmising appeared. The plan of the sigmoid neuron dependent on CVNS utilizing the TSMC 180-nm innovation is appeared in Fig. 6.

To assess the effects of the minor deviation procedure of the proposed structure, Monte Carlo reproductions have been done. Monte Carlo recreations demonstrate that the best assortment happens for the yield ((y))—5.2 for the most extraordinary gauge of the information ((x)) 0. The after effects of the Monte Carlo recreations

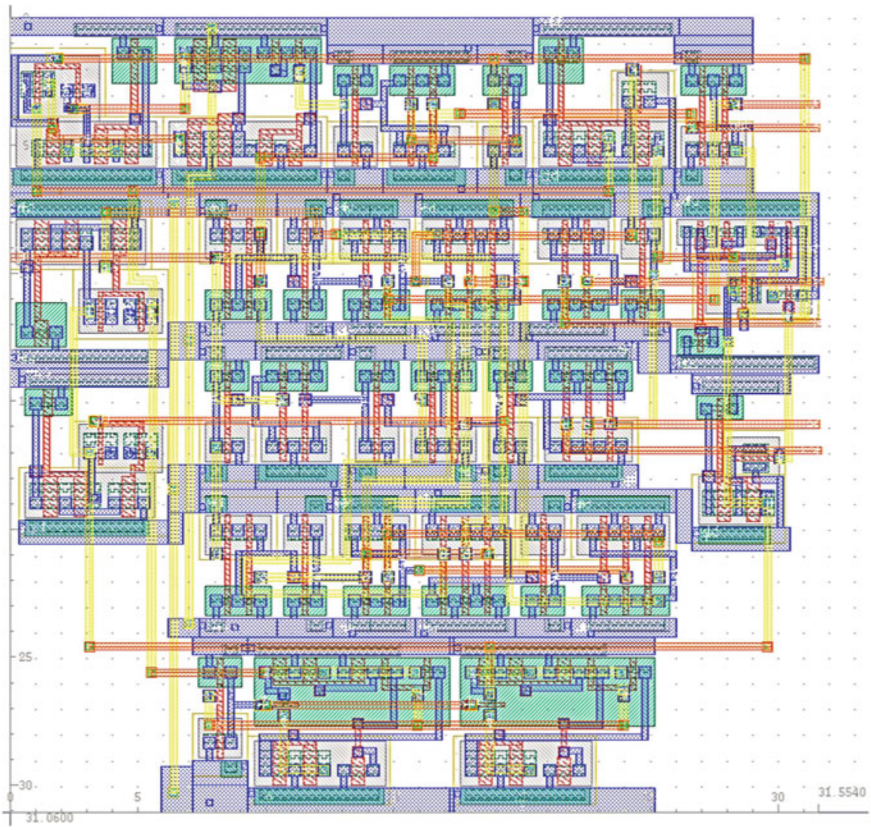


Fig. 6 180-nm sigmoid neuron layout

for ((y))—Output 5.2 with an outrageous gauge of the ((x)) 0 input are appeared as follows. The assortment of execution current dimensions is under 1 μA. Since the qualification between each two dimensions of CVNS yield consecutive is identical to 1 μA, the vast majority of the yield streams that are somewhere close to the two dimensions speak to a comparative yield. In this sense, the proposed structure can bolster the assortments of methodology. This makes the proposed structure sensible for refreshing the accurate simple sigmoid neuron for neurochips with on-chip learning (Fig. 7).

Addressable LUT (RALUT), bit-level mapping, and cross breed techniques. Examinations are performed regarding required number of digits, most extreme mistake, region, power, and defer where data is accessible. It ought to be called attention to that CVNS is simple and its design is done physically. To give meet grounds to examinations with the digitalized plans as far as zone effectiveness, the structures created are reproduced and simulations in a TSMC 0.18-μm technology.

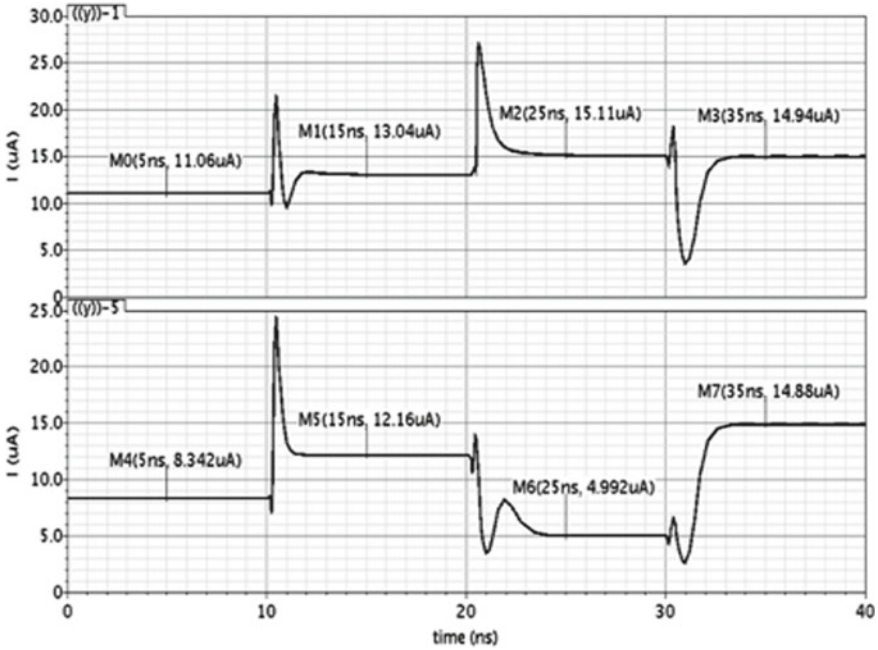


Fig. 7 Sigmoid neuron's simulation results

N_{in} and N_{out} in truth table are comparable elements utilized exclusively. These are the marker of the quantity of information and yield bits for every one of the sigmoid limit assessments. For the proposed work, there are eight double digits and eight digits of CVNS input. After the number-crunching arrangement, the proposed structure requires just two simple CVNS digits of yield. These two CVNS simple digits give 8-bit focal points. The bit-level mapping method made does not meet the data and execution destinations required for chip neurons with on-chip learning. Moreover, in spite of the way that this work has the scarcest deferral, it disregards the requirement for the goals. It has the most astounding estimation mistake among every one of the structures.

The non-direct methodology by parts requires multipliers, which possess an impressive piece of the area and have extensive deferrals. The structure made fulfills the requirements of the neural frameworks on the chip as for the augmentation of the data, yield by structure for the most elevated estimation blunder of the ((x)) 0 appeared as follows. This structure has the most astounding least mistake to the detriment of usage, deferral and vitality utilization of the region. Point by point delays is the essential type of postponement of related structures. At the point when all is stated, systems dependent on PWL require multipliers, which expand their region and deferrals. With a definitive objective of defeating this issue, structures dependent on PWL duplicated less are delivered. These structures are perceived utilizing duplicating coefficients that are powers of two. The reported structure has a postponement of 2.07-ns. Anyway, its info augmentation is limited to just and its suspicion mistake is to some degree high. The initiation work relies upon a less increasing PWL procedure and gives a high simple methodology mistake. The passage rates for the proposed structure and structures are created, and they are set at a perfect coupled speed with the goal that these structures can work. The data time span for these structures is set at 3.71 ns, its info speed is set at 64.87 ns. A uniform sporadic piece stream that incorporates 10,000 data sources is associated with every one of these circuits to evaluate the power utilization. The proposed CVNS structure fulfills every one of the prerequisites of the application including most extreme error, the quantity of digits, and input extend. The most notes worthy number of arithmetic activities for the proposed structure happens for the input to the info run r2, which incorporates expansion, mod2 task, and output assignment block. The proposed structure has the most minimum area and power consumption. The delay of the developed structure is equivalent to 2.53 ns.

Tanner tool is a software tool that is used for simulating the proposed method. A small description about the steps involved in tanner tool and design specifications are discussed here.

Tanner Tools:

The present semiconductors and electronic frameworks are mind boggling that planning them would be impossible without electronic Design Automation (EDA). This groundwork gives a thorough outline of the electronic design process, and after that portrays how configuration groups utilize Cadence instruments to make the most ideal plan at all measure of the time (Fig. 8).

Design Specifications:

This progression included expressing in positive terms the performance of the chip. Like on the off chance that we are making a processor, data size, processor speed, special functions, control and so forth is obviously expressed now. Likewise to some degree it is chosen, the best approach to implement the design. In this way, it manages architectural part of the Design at highest level possible (Fig. 9).

Experimental Results:

The schematic shown below done by using sigmoid neuron approximation.

Sigmoid main objectives are:

1. Decide the relationships between the biological and artificial neural networks
2. Usage of fake neural systems—The OR example
3. The exemplary XOR issue.

When all is said in done, a sigmoid capacity is real-valued approximate and differentiable, having a negative or positive first derivative, one nearby least, and one neighborhood most extreme. The logistic sigmoid function is identified with the hyperbolic digression as pursues (Figs. 10 and 11).

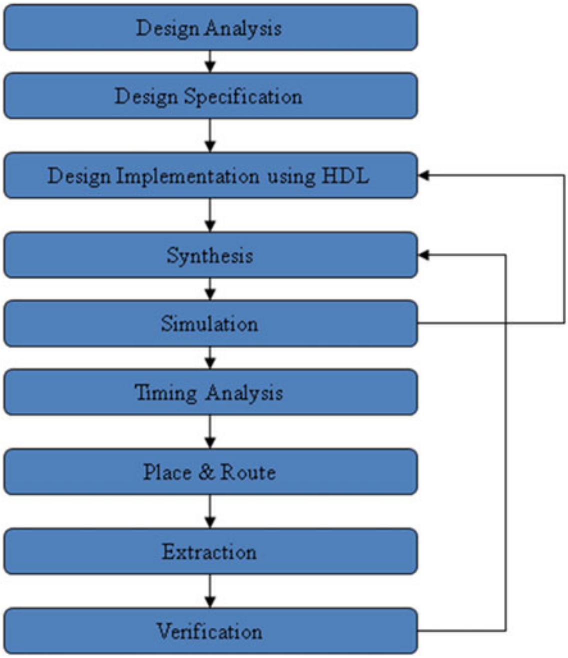


Fig. 8 Steps involved in tanner tools

3 Discussion of Results

The execution results contrast positively and recently created structures as far as zone, delay, and power utilization. The proposed neuron network possesses 28% less area contrasted and the state of-the-art techniques and it has multiple times brought down power and delay.

Moreover, it is critical that the half and half system has been made to execute the hyperbolic digression enactment work. This is pertinent just for cutting edge frameworks, while this record is gone for blended and simple flag frameworks. The ASIC of the proposed structure is utilized in a development TSMC 0.18-µm. The proposed plan has slight estimate mistakes, essentially indistinguishable to advanced frameworks.

Moreover, it has been tried under fluctuating information conditions what's more, corners with victories.

4 Summary

To comprehend the sigmoid limit, another structure dependent on CVNS is proposed. The consequences of the reenactment in a TSMC 0.18-µm innovation demonstrate that the proposed structure differentiates emphatically and forefront innovation. The

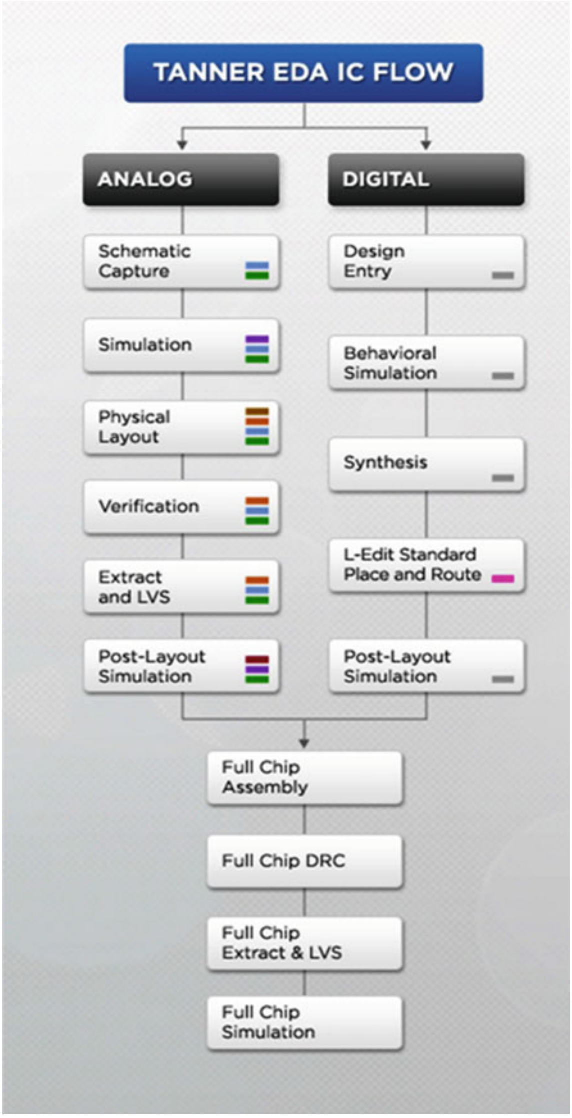


Fig. 9 Tanner EDA IC flow

main recommendation in this project is we have to give the low power mixed CVNS as input to the circuit. Another factor is giving the software used tanner tools will work only when we give low power CVNS as input and we can get the desired output which is power consumption, area and delays.

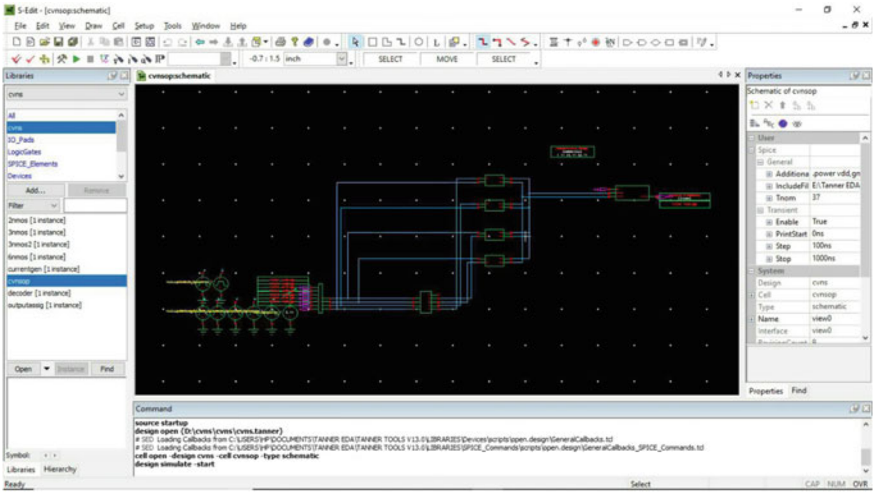


Fig. 10 Schematic diagram of CVNS

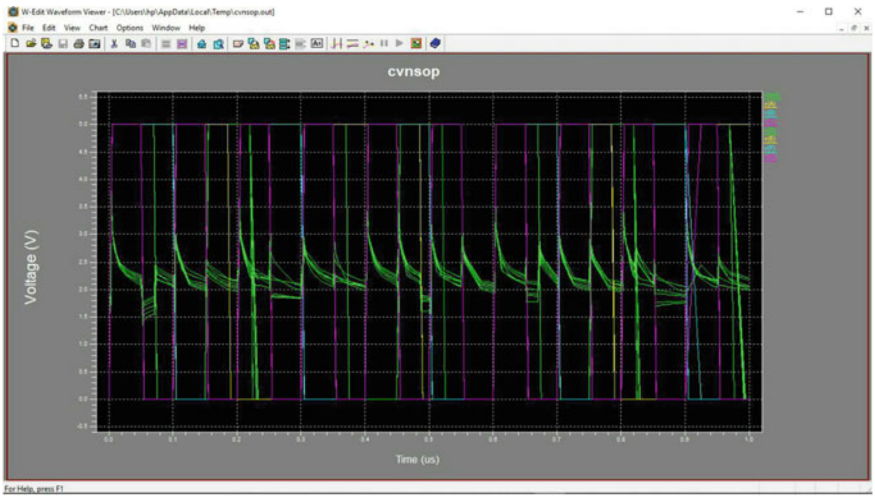


Fig. 11 Results for the schematic CVNS

5 Conclusion

Another basic sigmoid neuron dependent on CVNS is delivered, which is sensible for neurological chips with on-chip learning. The assessment of the assessment plan of the proposed capacity contrives the maltreatment of the PWL estimation procedure and relies upon a logical reasoning that utilizes the CVNS features. What's more, in perspective on the higher methodology blunder, the quantity of basic CVNS digits of info and yield required for the VLSI reproduction of the proposed sigmoid limit assessment strategy is illuminated.

References

1. Zamanlooy B, Mirhassani M (2017) An analog CVNS-based sigmoid neuron for precise neurochips. IEEE Trans Very Large Scale Integr (VLSI) Syst 25:894–906
2. Tsai C-H, Chih Y-T, Wong WH, Lee C-Y (2015) A hardware efficient sigmoid function with adjustable precision for a neural network system. IEEE Trans Circuits Syst II Express Briefs 62(11):1073–1077. https://doi.org/10.1109/tcsii.2015.2456531
3. Zamanlooy B, Mirhassani M (2015) CVNS synapse multiplier for robust neurochips with on-chip learning. IEEE Trans Very Large Scale Integr (VLSI) Syst 23(11):2540–2551
4. Khodabandehloo G, Mirhassani M, Ahmadi M (2011) CVNS-based storage and refreshing scheme for a multi-valued dynamic memory. IEEE Trans Very Large Scale Integr (VLSI) Syst 19(8):1517–1521
5. Khodabandehloo G, Mirhassani M, Ahmadi M (2012) A prototype CVNS distributed neural network using synapse-neuron modules. IEEE Trans Circuits Syst I Reg Papers 59(7):1482–1490
6. Amin H, Curtis K, Hayes-Gill B (1997) Piecewise linear approximation applied to nonlinear function of a neural network. IEE Proc Circuits Devices Syst 144(6):313–317
7. Prasad GRK (2018) Design and verification of AXI APB bridge using system verilog. JARDCS 10(6):1401–1408
8. Sankar Rao DG (2018) Analysis of static and dynamic CMOS low power high speed NP domino logic. Int J Adv Trends Comput Sci Eng 7(6):99–102

A Systematic Synthesis and Analysis of Linear Antenna Arrays Using MATLAB

K. Prasad and R. P. Singh

Abstract Antennas are the major components of a communication system to provide coverage over a wide area. Gain obtained from a single antenna is not sufficient to provide required coverage for effective communication. In order to fulfill this requirement in communication systems antenna arrays are desired for high gain and optimal coverage. In this paper synthesis of linear antenna arrays is studied and synthesis is performed by using Mat lab synthesizer. Several Antenna Synthesis Methods are available in literature and the main purpose of this paper is to create a computer tool for such Antenna Synthesis Methods. With the help of MATLAB, a graphical user interface (GUI) is created, which is very much required to the user to choose the one among the different synthesis methods. Finally in this paper different Antenna Synthesis Methods are presented theoretically and different MATLAB test results were discussed. In this research work, a tool is made for Linear Array synthesis. Enticingly, two popular methods namely, general method and Schelkunoff methods were investigated in this paper.

Keywords Linear arrays · Array synthesis · Schelkunoff method

1 Introduction

Antennas demonstrate a property known as reciprocity, which means that an antenna will maintain the same characteristics regardless if it is transmitting or receiving. Most antennas are resonant devices, which operate efficiently over a relatively narrow frequency band. When a signal is fed into an antenna, the antenna will emit radiation distributed in space in a certain way [1, 2].

In transmission, a radio transmitter supplies an electric current oscillating at radio frequency (i.e. a high frequency alternating current (AC)) to the antenna's terminals, and the antenna radiates the energy from the current as electromagnetic waves (radio waves). In reception, an antenna intercepts some of the power of an electromagnetic

K. Prasad (✉) · R. P. Singh
Department of ECE, AITS, Rajampet, AP, India
e-mail: kasigariprasad@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_43

wave in order to produce a tiny voltage at its terminals, which is applied to a receiver to be amplified. An antenna array (or array antenna) is a set of multiple connected antennas which work together as a single antenna, to transmit or receive radio waves [3]. The individual antenna elements are connected to a single receiver or transmitter by feed lines that feed the power to the elements in a specific phase relationship.

It is often necessary to design an antenna system that will yield desired radiation characteristics for establishing an effective communication system. The most commonly required aspect in designing an antenna is to choose the one in which the far-field pattern of antenna possesses nulls only in certain directions. Some of the other requisites are the antenna radiation pattern that has to exhibit desired distribution, narrow beam width and low side lobes, decaying minor lobes, and many others. To an antenna engineer in general, the task is not only to determine the antenna configuration; instead it is more important to examine the antennas geometrical dimensions and current excitation distribution [4, 5].

2 Objective—Proposed Work

The objective of this paper is to create a synthesizing tool for different antenna synthesis methods. The tool is designed using MATLAB software. A graphical user interface (GUI) is created with four linear array synthesis methods included in it. By using this tool user can find the excitation current coefficients for all elements of linear array antenna. The obtained current coefficients produce the desired radiation characteristics. Finally the results were compared with desired values. This comparison enables user to select best results to give desired radiation characteristics. The methodology implemented in this paper is described in the following. At the outset two synthesis methods were analyzed, later the implementation and analysis of individual methods were performed by using MATLAB.

Initially synthesizing antenna model is considered, to produce uniform sectorial pattern in desired direction by using both Fourier Transform and Woodward-Lawson methods, is implemented. Next after, a generalized program code is written for each synthesizing antenna methods [6–8]. Later a graphical user interface is created for each and every method. For each method, all possible cases are tested and outputs are compared with the expected values. In the final stage all the programs are combined to create one GUI. For those Antenna synthesis methods, in this work the researchers develop a computer tool, which easily interfaces with the user. When user gives the desired array factor and other inputs, it produces normalized current coefficients and elements positions [9]. In the case of Woodward-Lawson and Fourier method, it provides the facility of comparison of desired array factor and approximated array factor.

Generally the desired antenna patterns consists of determining the full null positions, beam shaping of the discrete arrays, beam shaping of the continuous arrays and reduced side lobe level. To determine those antenna characteristics different methods are involved such as Schelkunoff Polynomial method for determining the null positions, Woodward-Lawson method for beam shaping of the discrete arrays, Fourier transform method for beam shaping of continuous arrays and Dolph-Tschebyscheff method for reduction of the side lobe level. The following procedure is used in determining the desired antenna parameters by using the antenna synthesis models in mat lab. Depending upon the different method chosen, inputs such as array factor, side lobe level (in dB), and null positions (in radians) are provided in the user space. After entering these values, tool calculates and displays the excitation current coefficients in all array elements [10]. These currents are used to synthesize array factor which is compared with the desired array factor. By varying the number of elements and spacing between the elements one can get array factor which is close to the desired array factor. The tool has been tested using typical synthesis problems and the results are observed to be satisfactory.

3 Results and Discussion

In order to study the outcomes of the synthesized linear antenna arrays using mat lab the following procedure is to be executed. The general method has four steps. Open MATLAB.

3.1 General Method

Select the file dipole Array with .m as extension, Compile the source file and debug the errors (if any), Give the number of elements for the array, input current to each element and phase, number of elevation points and simulate, and Output plot is obtained and the corresponding values of gain, directivity, HPBW, Side Lobe Level are also obtained in the result. The sample inputs to the system is Enter the current amplitude 1, Number of elements 3, and Number of elevation points 2000. The respective input and output is shown in Table 1.

Table 1 General method input output relation

Input parameter	Value	Output—parameter	Value
Enter the current amplitude	1	Total elements input current	3.0 A
Number of elements	3	Total radiated power	158.3 W/m^2 (22.0 dBW)
Number of elevation points	2000	Peak gain	3.4 (5.3 dBi)
–	–	Half-power beam width	33.0°
–	–	Side-lobe level	18.7 dB

From the above simulation Procedure, there is no chance to select the null positions, Reduction of side lobe level etc., which can be overcome by using the following methods of the proposed system. It can also be shown that by choosing appropriate values for the parameters, Gain, Directivity can be increased and the grating lobes can be reduced to some extent compared to the general method. Grating is the major problem for the antenna arrays. They occur when the distance between the elements is greater than half the wavelength of the EM waves radiated by the Antenna Radiating elements.

3.2 Proposed Method—Schelkunoff Polynomial Method

The procedure for the proposed method is given as follows. Open MATLAB and from that select the file named as Antenna_project with .m extension and then run it. The next step is selection of suitable method among many others that already discussed in Sect. 2. Provide inputs to the respective methods, later select the type of the plot, followed by selection of the array factor, and finally interface the outputs.

The output obtained from each method commonly includes the elemental locations, Current magnitudes and the phases of the respective elements. The plot obtained will be either polar or XY depending on the input. The interpretations from the obtained plot can be carried out. In view of the polar plot of Schelkunoff method, gain can be determined. Sample Output of this method is given in Fig. 1.

Circular dotted lines indicate the gain of the lobes, which are demarcated in Fig. 1c. The first Null beam width (FNBW) or Half Power Beam Width (HPBW), from which the beam area and directivity can be estimated. The circle corresponding to 3-dB indicates the HPBW. From that the beam area can be estimated as

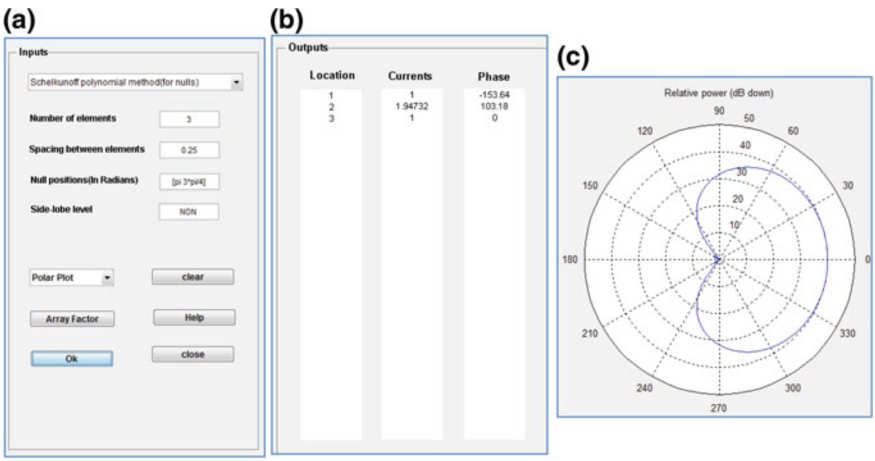


Fig. 1 Sample output of Schelkunoff polynomial method. **a** Inputs, **b** outputs and **c** polar plot

Table 2 Summary of inputs, outputs and calculations of the system

Method	Inputs	Outputs	Calculations
Schelkunoff polynomial method	Elemental spacing NULL positions Polar plot	Number of elements Locations Currents Phases	$HPBW_\theta$ Beam area = $HPBW_\theta$ * $HPBW_\phi$ Gain
Woodward Lawson method	Number of elements Elemental spacing Array factor XY-plot	Locations Currents Phases	Side lobe level
Fourier transform method	Number of elements Elemental spacing Array factor XY-plot	Locations Currents Phases	Side lobe level
Dolph-Tschebyscheff method	Number of elements Elemental spacing Side lobe level Polar plot	Locations Currents Phases	Gain

$HPBW_\theta$ * $HPBW_\phi$. Generally, it is assumed that $HPBW_\theta = HPBW_\phi$ for simplicity. Theoretically, directivity is calculated as 32,400/($HPBW_\theta$ * $HPBW_\phi$) (for small radiation angles) and as 41,253/($HPBW_\theta$ * $HPBW_\phi$) (for large radiation angles) (Table 2).

The results obtained are described as follows. Maximum gain from the pattern is obtained as 40 dB from the circular ring of polar plot. HPBW is obtained as 70° − (−70°) = 140° Beam area is 19,600 square degrees and Directivity = 41,253/19,600 = 2.104 = 3.23 dB.

The results obtained from the method are enhanced compared to the general method. For the same number of elements the current obtained is 3 A to all the elements. But the total magnitude of the current for this method is 3.54 A which is slightly greater. The gain obtained from the general method is 3.4 dBi but the gain of this method is 40 dB which has increased enormously. HPBW of the general method is 33.0° but the HPBW of this method is 140°. The beam area of general method is 1089 square degrees where as that of this method is 19,600 square degrees. The directivity of the general method is 14.73 dB but that of this method is 3.23 dB which is low because of increased beam area and gain. Thus, there exists a trade-off between the parameters of the system; one must choose the required parameter.

3.3 Experiment 1

In this synthesis experiment the Inputs are Spacing between elements 0.25λ, Nulls position is 0 and π/2. The expected outputs are Number of elements 3; Expected

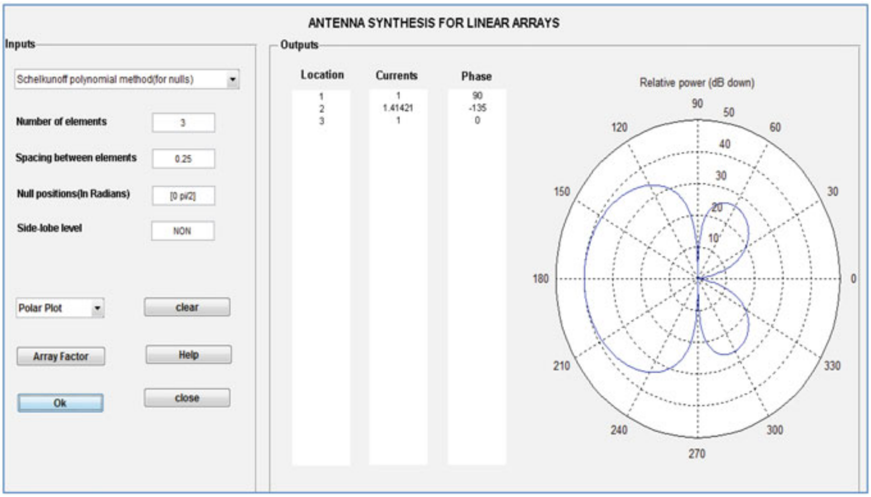


Fig. 2 Output of Schelkunoff polynomial methods with nulls at 0 and π/2, spacing 0.25λ

output contains no radiation at 0 and π/2. The obtained output is shown as follows (Fig. 2).

3.4 Calculations

From the output Panel of the above figure, the total current required is given as 3.4142 A. The gain of the pattern is given as 40 dB. The beam width (HPBW) at the 3 dB down circular ring is calculated approximately to be 130°. The beam area is calculated to be 16,900 square degree. And the directivity is calculated as 3.8 dB.

Tool produces currents for Radiation at 0 and π/2 also since the spacing between elements is not suitable for the array factor requirement. The spacing between the elements plays important role. Depending upon the spacing the pattern changes. If the spacing is not appropriate then it shows the error i.e. it will not produce the nulls in the desired direction, "If we give 0.5λ as spacing between elements, it produce desired array factor (nulls at 0 and π/2)". This is shown in the below output (Fig. 3).

From the output Panel of the above figure, the total current required is given as 2.9 A. The gain of the pattern is given as 40 dB. The beam width (HPBW) at the 3 dB down circular ring is calculated approximately to be $80° - 20° = 60°$ per each lobe. The beam area is calculated to be 3600 square degree per each lobe. And the directivity is calculated to be 10.8 dB. The number of elements remained same though the spacing is changed. As the spacing between the elements is 0.5λ, nulls are produced at [0, π/2] and also at their respective counter clock-wise angles also i.e., π and 3π/2.

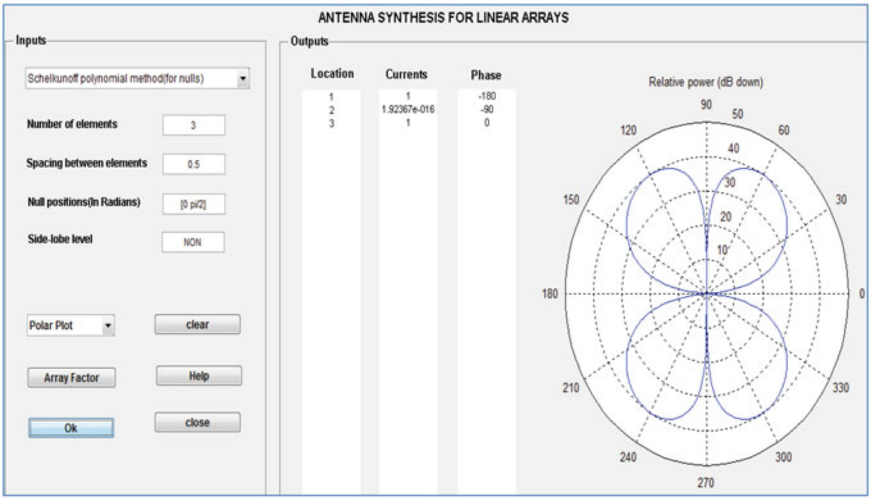


Fig. 3 Output of Schelkunoff polynomial methods with nulls at 0 and π/2, spacing 0.5λ

3.5 *Experiment 2*

In this synthesis experiment the Inputs are Spacing between elements 0.5λ, Nulls position is 0. The expected outputs are Number of elements 3; Expected output contains no radiation at 0.

The obtained output is shown in Fig. 4.

From the output Panel of the above figure, the total current required is given as 2 A. The gain of the pattern is given as 40 dB. The beam width (HPBW) at the 3 dB down circular ring is calculated approximately to be $150° - 30° = 80°$ per each lobe. The beam area is calculated to be 6400 square degree per each lobe. And the directivity is calculated to be 8.09 dB.

Tool produces currents for no Radiation at 0 and π also. Because of the spacing between elements produces extra null at pi. Refer to the figure. "If we give 0.25λ as spacing between elements, it produce desired array factor (nulls at 0 only)". This is shown in Fig. 5.

From the output Panel of the above figure, the total current required is given as 2 A. The gain of the pattern is given as 40 dB. The beam width (HPBW) at the 3 dB down circular ring is calculated approximately to be 300°.

Tool produces currents for no Radiation at 0 only. Because of the spacing between elements 0.5λ produces extra null at pi, we have reduced the spacing between the elements to 0.25λ, exact output is obtained i.e., null only at 0 positions. From the previous test results, we can comparatively say that Nulls are produced in pairs when the spacing between the elements is given as greater than the half of the wave length (operating). So we must also see that positions for the nulls are to be selected very carefully.

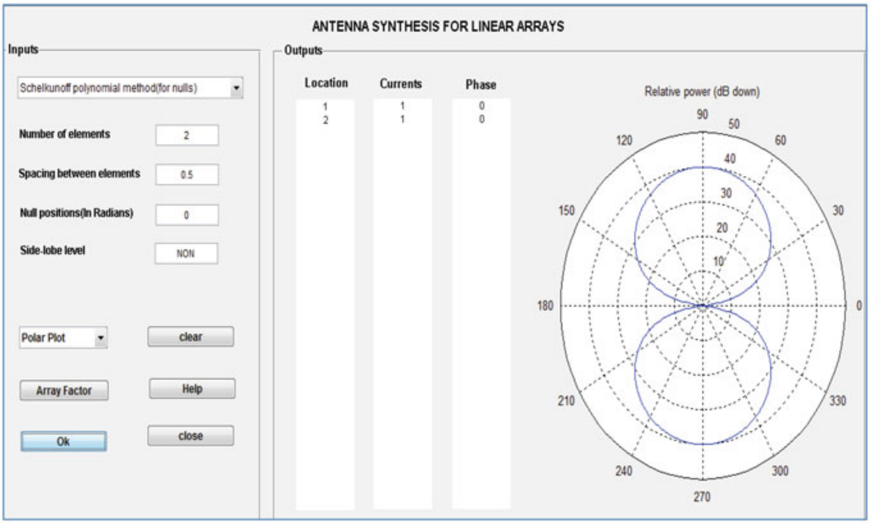


Fig. 4 Output of Schelkunoff polynomial methods with null at 0, spacing 0.5λ

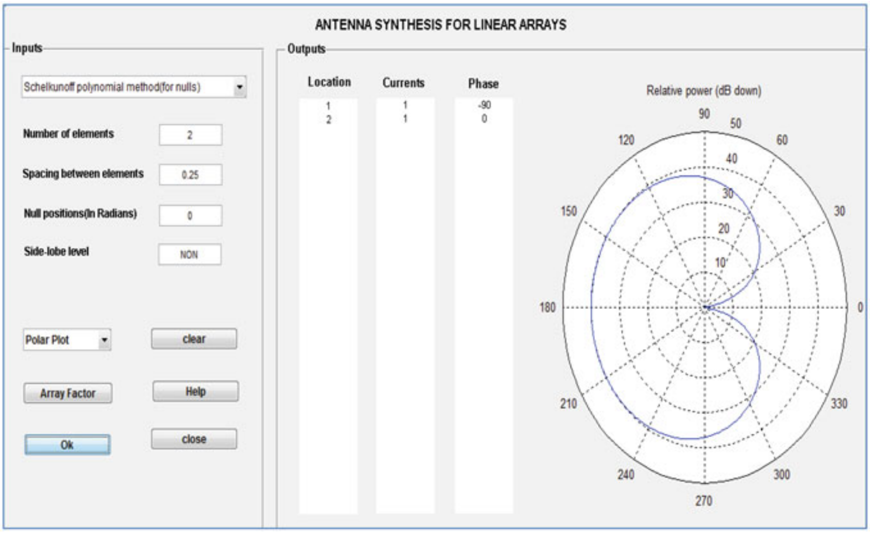


Fig. 5 Output of Schelkunoff polynomial methods with null at 0, spacing 0.25λ

4 Conclusion

We propose an approach to achieve the nulls in the desired direction by adjusting the spacing between the elements and number of elements in the proposed array. It is comparatively concluded that the nulls are produced in pairs when the spacing between the elements is given as greater than the half of the wave length (operating).

References

1. Khalid A, Sheikh SA, Ul Hasnain Shah I, Khan QU (2015) Synthesis of linear antenna array using genetic algorithm to reduce peak sidelobe level. In: 2015 9th international conference on electrical and electronics engineering (ELECO). IEEE
2. Elgetun B, Holm S (1995) Optimization of the beampattern of 2D sparse arrays by weighting. In: IEEE ultrasonics symposium proceedings, vol 2
3. Leahy RM, Jeffs BD (1991) On the design of maximally sparse beamforming arrays. IEEE Trans Antennas Propag 39(8):1178–1187
4. Cen L, Yu ZL, Rahardja S, Cen W (2010) Linear sparse array synthesis with minimum number of sensors. IEEE Trans Antennas Propag 58(3):720–726
5. Aboul-Seoud AK, Hafez A-DS, Hamed AM, Abd-El-Latif M (2014) A conformal conical phased array antenna for modern radars. In: IEEE aerospace conference, 2014
6. Jeyali Laseetha TS, Sukanesh R (2011) Synthesis of linear antenna array using genetic algorithm to maximize sidelobe level reduction. Int J Comput Appl 20(7)
7. Shreniand V, Raikwar P (2012) Optimization of reduction in side lobe level using genetic algorithm. Int J Emerg Technol Adv Eng 2(12)
8. Stutzman WL, Thiele GA (1998) Antenna theory and design, 2nd edn. Wiley, New York
9. Balanis CA (2008) Antenna theory: analysis and design, 2nd edn. Wiley, New York (student edition)
10. Hahn BD, Valentine DT (2007) Essential MATLAB for engineers and scientists, 3rd edn. British Library Cataloguing in Publication

Dr. Kasigari Prasad is a noted academician, and a well-known researcher in the fields of Antenna Theory and wave propagation. Dr. K. P. is the Alumni of Annamacharya Institute of Technology & Sciences, one of the prestigious institutions. He did his Doctorate of Philosophy on Blade Antennas under the esteemed guidance of versatile personality, great academician Dr. R. P. Singh. He is International Young Investigator Award winner in the year 2012 for his research work on orthogonal frequency division multiplexing. He is the Best Engineering Award winner in the year 2015 from Institution of Engineers for his outstanding contributions in the Engineering. He is also honored with The Jewel of Rajampet Award for his outstanding contribution in promotion of values ethics into the young minds. He is a noted Antenna Researcher, Graphologist, and Writer & Motivational Speaker. He has delivered many keynote addresses and acted as session chairs for national and international conferences. He has published chapter books and monographs. With a rich background of Engineering, Political science, English Literature, Journalism and Personality development areas and with wide teaching experience, he has been deeply involved in research, writing, teaching and promotion of many branches of Engineering and Education for the past 15 years. He has associated with antenna design projects for many for institutions Dr. Prasad has presented papers on Antenna design, Interference reduction in OFDM, Artificial Intelligence, wave Propagation, Electromagnetic fields, Graphology, Political marketing and in many national and international conferences and journal publications.

Selecting Electrical Billing Attributes: Big Data Preprocessing Improvements

Amelec Viloria, Jesús García Guiliany, Nataly Orellano Llinás, Hugo Hernandez-P, Ernesto Steffens Sanabria and Omar Bonerge Pineda Lezama

Abstract The attribute selection is a very relevant activity of data preprocessing when discovering knowledge on databases. Its main objective is to eliminate irrelevant and/or redundant attributes to obtain computationally treatable issues, without affecting the quality of the solution. Various techniques are proposed, mainly from two approaches: *wrapper* and *ranking*. This article evaluates a novel approach proposed by Bradley and Mangasarian (Machine learning ICML. Morgan Kaufmann, Sn Fco, CA, pp. 82–90, 1998 [1]) which uses concave programming for minimizing the classification error and the number of attributes required to perform the task. The technique is evaluated using the electric service billing database in Colombia. The results are compared against traditional techniques for evaluating: attribute reduction, processing time, discovered knowledge size, and solution quality.

Keywords Electric billing · Concave programming · Data mining · Electric service billing

---

A. Viloria (✉)
Universidad de la Costa, St. 58 #66, Barranquilla, Atlántico, Colombia
e-mail: aviloria7@cuc.edu.co

J. G. Guiliany
Universidad Simón Bolívar, Barranquilla, Colombia
e-mail: jesus.garcia@unisimonbolivar.edu.co

N. O. Llinás
Corporación Universitaria Minuto de Dios—UNIMINUTO, Barranquilla, Colombia
e-mail: Nataly.orellano@uniminuto.edu

H. Hernandez-P · E. S. Sanabria
Corporación Universitaria Latinoamericana, Barranquilla, Colombia
e-mail: hhernandez@ul.edu.co

E. S. Sanabria
e-mail: esteffens@ul.edu.co

O. B. P. Lezama
Universidad Tecnológica Centroamericana (UNITEC), San Pedro Sula, Honduras
e-mail: omarpineda@unitec.edu

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_44

1 Introduction

Data mining is mainly applied when large amounts of historical data are stored with the purpose of exploiting them, seeking the knowledge implicit in this information. However, when the databases to be mined become very large, mining algorithms work very slow, requiring too much time to process the information, so the problem becomes very difficult to face. One way for addressing this issue is to reduce the data before applying mining [2]. Particularly, the selection of attributes applied as a pre-processing stage to mining has been useful as it seeks to eliminate the irrelevant and/or redundant attributes that cause mining tools to become inefficient, but without affecting the quality of classification performed by the mining algorithm. Moreover, sometimes, the percentage of properly classified instances becomes higher when using attribute selection because the data to be mined are free of noise or data that cause the mining tool to generate more nodes than necessary [3].

Attribute selection was performed using *wrapper* and *filter* methods. The wrapper methods, although effective for removing irrelevant and redundant attributes, are very slow because they apply the mining algorithm several times, varying the number of attributes in each execution, following any search and stop criteria [4]. Filter-type methods employ some kind of measure of information gain, distance, or consistency between the attribute and the class, these being much more efficient than the wrapper [5]. However, since they measure the importance of each attribute independently, they cannot detect if there are redundant attributes, nor determine whether the combination of two or more attributes, seemingly irrelevant in isolation, can become relevant [6].

This article evaluates the use of optimization techniques for attribute selection, as an alternative to traditional methods. Particularly, the research evaluates an interesting and novel approach by Bradley and Mangasarian [1, 7] which uses concave programming for minimizing the classification error while minimizing the number of attributes needed to perform the task. Bradley's approach, according to the results reported by this author, seems to count on the efficiency of the filter methods and the effectiveness of the wrapper methods, hence the importance of this evaluation.

2 Method

2.1 Data

The database used for the study in conformed by the electric service billing in the period 2017–2018, of three main cities in Colombia: Medellin, Cali, and Bucaramanga, obtaining a mine with 5411 records with attributes such as: Permanent User Registration (PUR), Year, Month, Type-Debit, Digit, kWh, Energy, Cve-invoice, Total, Status, Draft, Tariff, Name, Load-Installed, Load-Contracted, and others for a total of 21, where the Type-Debit attribute indicates illicit use, and represents the class column or focus of attention. For confidentiality reasons, the specific meaning of each attribute is not reported in this paper.

2.2 Methods

2.2.1 Concave Programming

Mangasarian and Bradley widely worked on finding and applying optimization methods for Data Mining issues [8, 9]. Among their studies, the scheme for classifying and selecting attributes stands out, obtaining excellent results when compared against a method based on Support Vector Machines (SVM). The evaluation carried out in this research is based on the model presented in [1, 10] and other publications by the same authors [9] that support such a model.

The essential idea of Bradley's model, called Feature Selection via concaVe minimization (FSV), consists of two parts: instance classification and attribute selection. Regarding the classification, the mathematical programming model seeks to minimize the average distances of instances incorrectly classified by the separation of a line, plane, or hyperplane. As for attribute selection, the purpose is to minimize the number of attributes required to perform a good class separation, under the assumption that it is possible to obtain a hyperplane, in less dimensions than the original data, such that the classification keeps being acceptable [11, 12].

These widely exposed, demonstrated, experimented, and discussed ideas in [1, 13, 14] were implemented in this study, specifically to be used and executed with the optimization tool called General Algebraic Modeling System (GAMS) [15] (unlike Bradley who used CPLEX in his experiments).

The objective of this experimentation was to observe the behavior of FSVGAMS in terms of solution quality and response time in a real-world and impartial problem. For this purpose, 25 experiments were designed with different values of P and ALFA: although Bradley obtained good results with values of 0.05 and 5 respectively, he warns that it may vary with the domain [16–18].

3 Results

The results obtained when varying the ALFA and P parameters are shown in Table 1, which reports the number of selected attributes (# Ats.), the value of F, that is, the value of the target function to be minimized (Val.Fun.Obj.), the processing time in seconds (Time (s)), and the number of iterations that were required to reach the final solution (Iters). Table 1 shows that:

- The number of attributes selected varies from 1 to 16, indicating that ALFA and P significantly affect FSV-GAMS in this regard.

Table 1 Results for various values of P and ALFA

Alpha	P	# Ats.	Val.Fun.Obj.	Time (s)	Iters
0.1	0.01	15	0.6724	215.14	4231
0.1	0.05	16	0.6589	264.28	5010
0.1	0.20	11	0.5478	269.52	5301
0.1	0.50	12	0.3652	217.54	4270
0.1	0.90	8	0.1004	194.52	4214
1	0.01	15	0.6412	475.25	8654
1	0.05	11	0.6457	278.12	5014
1	0.20	8	0.6101	240.15	4132
1	0.50	7	0.4973	230.21	4065
1	0.90	5	0.1800	271.85	4751
5	0.01	12	0.6741	289.25	5895
5	0.05	3	0.6995	179.25	2360
5	0.20	7	0.9245	227.14	4344
5	0.50	1	0.7658	193.49	2336
5	0.90	8	0.1990	255.17	4578
20	0.01	12	0.7738	170.29	3658
20	0.05	8	0.7140	194.25	3043
20	0.20	1	0.7547	175.25	2636
20	0.50	5	1.3758	190.25	3751
20	0.90	6	1.8449	244.17	4438
100	0.01	9	0.7389	287.78	10,000
100	0.05	4	0.7599	173.17	2456
100	0.20	8	2.1645	187.90	3645
100	0.50	5	0.8587	245.39	5087
100	0.90	8	7.2701	165.99	2999

- The value of the target function was not seriously affected by ALFA and P, as well as the number of iterations and consequently processing time, which kept at acceptable levels, most of the time.
- The number of iterations obtained contrasts with what was reported by Bradley, who mentions that it required 5–7 iterations, although it should be considered that the databases used by Bradley were, at best, three and a half times smaller.
- Just for ALFA $= 1$, the number of selected attributes was reduced as the P value was increased as expected. For the other ALFA values, a rather irregular behavior was obtained as P was increased.

Next, attributes selected by FSV-GAMS were used to feed the decision tree induction algorithm J4.8 used in the Weka tool [16], which is the latest version of C4.5, which in turn is one of the induction algorithms known and used in data mining [7].

Table 2 shows the following results: the size of the generated tree is reported as the number of nodes in the tree (TmArb); the percentage of successfully classified instances, using 10-fold cross validation (AccTest); the time required to generate the tree (Time (s)); the cost-benefit obtained (C/Ben) assuming a profit of +97.5 when an "illicit" and a cost or loss of equal magnitude are correctly predicted otherwise, and a profit of 2.5 when a "non-illicit" is correctly predicted (so there is savings by not inspection) and a cost or loss of equal magnitude otherwise. The last column (C/Ben 1000) is the cost-benefit obtained standardized with respect to the cost-benefit that is obtained when the 21 attributes are fed to J4.8, that is, when no attribute selection is made.

Table 2 shows that:

Table 2 Results of *J4.8* using the attributes selected by FSV-GAMS

Alpha	P	TmArb	AccTest	Time (s)	C/Ben	C/Ben 1000
0.1	0.01	39	96.36	2.33	54,620	997.8
0.1	0.05	39	96.36	2.54	54,810	1010.3
0.1	0.2	15	93.33	1.64	52,555	961.1
0.1	0.5	25	93.41	1.77	48,570	867.7
0.1	0.9	3	90.01	0.85	29,850	505.0
1	0.01	39	97.87	2.33	54,775	1003.5
1	0.05	21	95.47	1.78	48,175	885.7
1	0.2	9	90.35	1.07	28,560	522.3
1	0.5	3	90.44	0.84	27,770	508.1
1	0.9	1	79.48	0.77	−51,180	−923.4
5	0.01	21	95.54	1.88	47,105	874.1
5	0.05	3	90.52	0.24	27,035	521.3
5	0.2	19	90.78	0.78	26,485	476.2
5	0.5	3	90.00	0.07	28,285	520.4
5	0.9	7	79.23	1.10	−48,835	−930.0
20	0.01	55	90.14	1.34	24,564	438.2
20	0.05	3	90.11	0.63	27,780	516.9
20	0.2	3	90.11	0.01	27,883	520.6
20	0.5	3	90.23	0.23	27,895	510.4
20	0.9	3	90.11	0.46	27,890	517.2
100	0.01	15	90.46	1.29	24,755	434.6
100	0.05	3	89.36	0.65	25,845	443.9
100	0.2	19	90.38	1.25	24,240	427.8
100	0.5	3	90.45	0.14	29,690	517.1
100	0.9	19	90.12	0.98	24,240	427.4

- The largest tree size is 55 nodes, while the smallest one has just 1 node; however, neither of these gets the best solution quality.
- The percentage of correctly classified instances ranged from 80 to 97%, i.e. an acceptable prediction percentage was achieved in all experiments.
- The processing time did not exceed 3 s.
- Four experiments achieved a cost-benefit that was very similar to that obtained using all attributes.

4 Conclusions

In this case study (detection of electrical energy illicits), the processing time of FSV-GAMS was reasonable (no more than 5 min) compared to the wrapper scheme which typically requires hours or days of execution. However, the results suggest that, for databases with the largest number of instances, the FSV-GAMS runtime will grow exponentially.

Comparing FSV-GAMS results against various ranking methods, it is noted that, in some cases, the quality of the solution is similar although the *ranking* methods offer the advantage of linear complexity with respect to the number of attributes. However, FSV-GAMS offers the advantage of finding subsets of attributes in which the possible inter-dependency between variables is framed, which is something that *ranking* methods cannot offer.

References

1. Bradley P, Mangasarian O (1998) Feature selection via concave minimization and support vector machines. In: Shavlik J (ed) Machine learning ICML. Morgan Kaufmann, Sn Fco, CA, pp 82–90
2. Hu C, Du S, Su J et al (2016) Discussion on the ways of purchasing and selling electricity and the mode of operation in China's electricity sales companies under the background of new electric power reform. Power Netw Technol 40(11):3293–3299
3. Xue Y, Lai Y (2016) The integration of great energy thinking and big datas thinking: big data and electricity big data. Power Syst Autom 40(1):1–8
4. Wang Y, Chen Q, Kang C et al (2017) Clustering of electricity consumption behaviour dynamics toward big data applications. IEEE Trans Smart Grid 7(5):2437–2447
5. Liu R, Feng G, Ding W (2011) Statistical analysis and application of SAS. China Machine Press, China
6. Ozger M, Cetinkaya O, Akan OB (2017) Energy harvesting cognitive radio networking for IoT-enabled smart grid. Mob Netw Appl 23(4):956–966
7. Isasi P, Galván I (2004) Redes de Neuronas Artificiales. Un enfoque Práctico. Pearson, London. ISBN 8420540250
8. Mangasarian O (1997) Arbitrary-norm separating plane. Technical report 97-07, Computer Science Dept., Univ. Wisconsin Madison
9. Bradley P, Fayyad U, Mangasarian O (1999) Mathematical programming for data mining: formulations and challenges. INFORMS J Comput 11:217–238

10. Rahmani AM, Liljeberg P, Preden J, Jantsch A (2018) Fog computing in the internet of things. Springer, New York ISBN: 978-3-319-57638-1, ISBN: 978-3-319-57639-8 (eBook)
11. Gangurde HD (2014) Feature selection using clustering approach for big data. Int J Comput Appl Innov Trends Comput Commun Eng (ITCCE):1–3
12. Abualigah LM, Khader AT, Al-Beta MA, Alomari OA (2017) Text feature selection with a robust weight scheme and dynamic dimension reduction to text document clustering. Expert Syst Appl 84:24–36
13. Sanchez L, Vásquez C, Viloria A, Cmeza-estrada (2018) Conglomerates of Latin American countries and public policies for the sustainable development of the electric power generation sector. In: Tan Y, Shi Y, Tang Q (eds) Data mining and big data. DMBD 2018. Lecture notes in computer science, vol 10943. Springer, Cham
14. Sánchez L, Vásquez C, Viloria A, Rodríguez Potes L (2018) Greenhouse gases emissions and electric power generation in Latin American countries in the period 2006–2013. In: Tan Y, Shi Y, Tang Q (eds) Data mining and big data. DMBD 2018. Lecture notes in computer science, vol 10943. Springer, Cham
15. Perez R et al (2018) Fault diagnosis on electrical distribution systems based on fuzzy logic. In: Tan Y, Shi Y, Tang Q (eds) Advances in swarm intelligence. ICSI 2018. Lecture notes in computer science, vol 10942. Springer, Cham
16. Silva V, Jesús A (2013) Indicators systems for evaluating the efficiency of political awareness of rational use of electricity. In: Advanced materials research, vol 601. Trans Tech Publications, Switzerland, pp 618–625
17. Perez R, Inga E, Aguila A, Vásquez C, Lima L, Viloria A, Henry MA (2018) Fault diagnosis on electrical distribution systems based on fuzzy logic. In: International conference on sensing and imaging, June. Springer, Cham, pp 174–185
18. Perez R, Vásquez C, Viloria A (2019) An intelligent strategy for faults location in distribution networks with distributed generation. J Intell Fuzzy Syst 36(2):1627–1637 (Preprint)

Prediction of Electric Consumption Using Multiple Linear Regression Methods

Amelec Viloria, Hugo Hernandez-P, Omar Bonerge Pineda Lezama and Jesús Vargas

Abstract In the new global and local scenario, the advent of intelligent distribution networks, or Smart Grids, allows the collection of data about the operational state of the electric network in real time. Based on this data availability, the consumption prediction becomes feasible and convenient in the short term, from a few hours to a week (temporary variables). The research proposes that the method used to present the temporary variables for a system to predict electrical consumption affects the results. To verify this hypothesis, different methods for representing these variables are considered, applied to the problem of predicting daily values of electricity consumption in the city of Bogota, Colombia.

Keywords Energy consumption · Short term load forecasting · Variable selection · Linear regression

1 Introduction

The short-term prediction is closely related to the problem of consumption peaks, in which there is a high increase in demand in a short period of time. This issue results in a serious problem that, if not properly planned and managed, causes economic losses, damage to equipment at different levels, and service cuts or price increases (both as a way to discourage consumption in economies that allow consumption by time slots, and due to the need to repair or replace damaged equipment) [1, 2]. Thus, this

A. Viloria (✉) · J. Vargas
Universidad de la Costa, St. 58 #66, Barranquilla, Atlántico, Colombia
e-mail: aviloria7@cuc.edu.co

J. Vargas
e-mail: jvargas41@cuc.edu.co

H. Hernandez-P
Corporación Universitaria Latinoamericana, Barranquilla, Colombia
e-mail: hhernandez@ul.edu.co

O. B. P. Lezama
Universidad Tecnológica Centroamericana (UNITEC), San Pedro Sula, Honduras
e-mail: omarpineda@unitec.edu

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_45

prediction plays an important role for companies in the electricity sector, especially in terms of distribution, because it allows a better approach to make strategic and operational decisions [3].

With regard to the temporal factor, there are several methods to represent variables related to the date used in the prediction models of electricity consumption. Among the most used methods [4–6], there are several variations of DOY (Day-Of-Year), such as the sample number, sequential numbering and non-linear variables derived from the date. Based on this, the hypothesis establishes that the method used to include temporary variables in a system for predicting electricity consumption influences the results [7–9]. Then, the proposal includes a comparison of different methods for representing temporary variables to determine which is the most convenient for the prediction of electricity consumption in the short term. As a case study, the electricity consumption of the city of Bogota between 2017 and 2018 was chosen, using a simple linear regression method to validate the proposed hypothesis, since the objective of this article is to determine the impact on the way in which temporary variables are presented, and not to build a prediction system.

2 Method

2.1 Data

The data used in the research correspond to the time period between 01-01-2017 and 29-10-2018 and contains daily samples of the following variables: power (MWh), Day Of Year (DOY), Year, Month, Day Of Month (DOM), Day Of Week (DOW). In addition, the minimum and maximum temperatures are added to the square belonging to each of the stations in a similar way to that observed in [10]. Preliminary tests determined that cubic variables do not influence this case.

In the procedure of initial analysis of the data, as happens in networks of distributed sensors, registries with zero or missing values were found, both in the climatic data and those of electrical consumption [11]. This is possibly due to problems in the measurement equipment or in the information acquisition processes. Since the amount of affected data is not significant, it was decided to perform a preprocessing process through which the records that show any anomaly were eliminated. The percentage of cells with missing data is 7%, resulting 20,595 usable samples out of a total of 22,145.

2.2 Methods

This research proposes the hypothesis that the way to include temporary variables affects the quality of electrical consumption predictions. To validate it, a comparison is made between the results of the STLF experiments using different methods to represent temporary variables. The analysis is carried out on real data about the city of Bogota.

Based on the recurrent use of the DOY variable, this approach will be used as the first method and starting point in order to establish the first point of comparison. As a secondary method, a separation of the date variable is proposed, breaking it down into four individual variables: year, month, day, and day of week. Finally, an adaptation of the method described in [12] is used, where the author applies a pre-processing (non-linear operations) on the temporary input variables. This last approach resembles the previous one, since it denotes an addition of temporary variables derived from the original ones. In this way, three groups of variables were formed for the different tests. The fourth group corresponds to the climatic variable and variables with information of the atypical days that will be used in all the tests. Finally, the combination of these three methods was tested following the procedure of [13].

Since electricity consumption continuously varies over time, it is feasible to apply different techniques and time series methods to predict consumption based on available historical data. One of these techniques consists in the use of current and/or previous information to predict future variables [14]. In this case study, the use of data from previous days is proposed, both about the demand for electricity and the weather variables of the region and data related to dates. In turn, in order to improve the prediction accuracy, the researchers took advantage of the fact that there are variables whose future values can be known in advance. This is the case of the variables like the date and related data (holidays or special regional events). It should be noted that the presence of atypical days such as weekends and holidays influence the consumption of electricity [15]. For this reason, this information should be added as common input variables for all tests.

In order to evaluate the results obtained with the selected model, the following errors obtained through different metrics are analyzed: the Root Mean Squared Error (RMSE) which, when expressed in the same unit of measurement as the variable to be estimated, facilitates its interpretation; the Relative Error (RE) expressed as a percentage; the Mean Bias Error (MBE) that allows the analysis if there is an underestimation or an overestimation in the prediction of electrical consumption; and the Pearson's Linear Correlation Coefficient (R) which helps determine the degree to which the data follow the general trend of the model [16].

3 Results

In the tests, daily samples were analyzed with the aim of predicting a day forward (One-Day Ahead) on the variable named electrical consumption. To improve the estimation, in addition to taking current samples, samples from several consecutive previous days or delays (self-regressive time series) are considered.

Figure 1 shows the error variation using the variable DOY and atypical days, and the climatic variables, for the 4 types of errors used as comparison metrics.

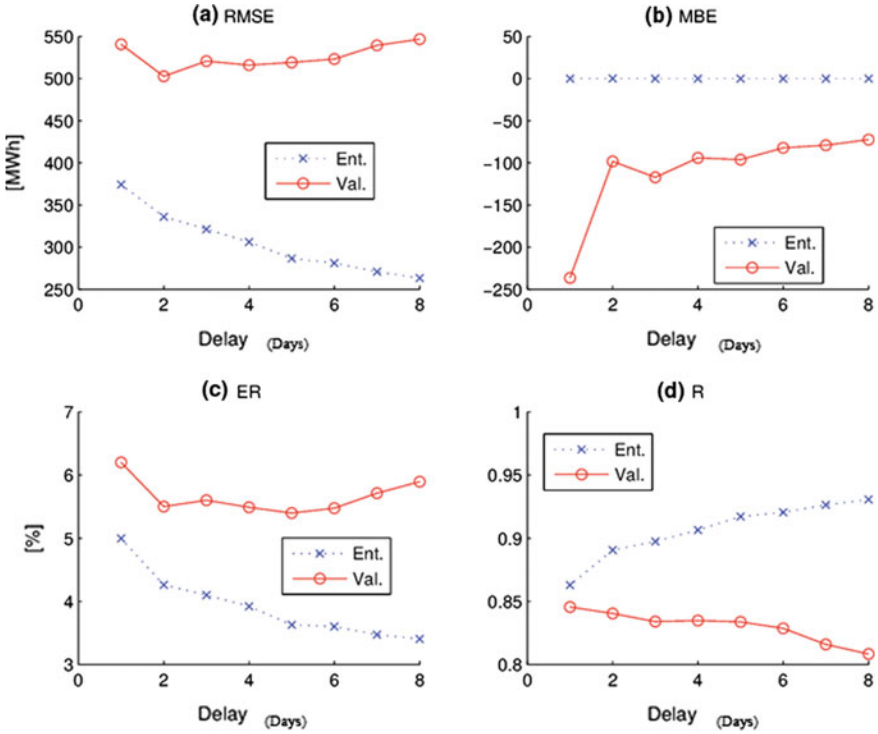


Fig. 1 Errors obtained for different delay values

It can be noted that the best result is achieved with a delay 2, that is, considering current samples and those of 1 day backwards. However, for higher values of delay, the training error decreases while the validation error tends to increase. It occurs because the increase in variables that implies a large delay causes an increase in the complexity of the linear regression model, which in turn generates a system overfitting, as explained in [17].

Tables 1 and 2 show the errors obtained at the province level, for the training and validation sets, respectively. It can be denoted, both in training and in validation, that the best result was obtained by using the variables of groups 2–4 as input to the system. An improvement of 10.56% in the Relative Error was achieved regarding the non-use of temporary variables. The other alternatives do not provide a significant improvement. As can be seen in Fig. 2, the trend of electrical consumption can be correctly estimated by the linear system.

Figures 2 and 3 show that the system correctly captures the linearity of the problem and that very few data are outside the error range of 10% (area between dotted lines) what are expected in hot weather in Tucumán. These external values correspond to the load peaks. The treatment of these anomalous data requires the development of specific systems that are out of the objective of this study.

Table 1 Training errors

Groups of variables	Description	SSE	MBE	R	RMSE	Err. Rel.
4	Basic variables	156,985,214.3	0.0335	0.8899	334.3812	4.33
2–4	Decomposition on dates	124,872,145.6	0.0000	0.9061	311.7453	3.82
3–4	Non-linear components	167,825,963.4	0.0168	0.8799	354.9917	4.35
1–4	DOY	148,693,256.4	0.0958	0.8908	337.0126	4.28
1–2–3–4	All	125,478,035.3	0.0002	0.9199	310.5817	3.79

Table 2 Validation errors

Groups of variables	Description	SSE	MBE	R	RMSE	Err. Rel.
4	Basic variables	141,107,458.6	−87.0713	0.8399	511.5355	5.51
2–4	Decomposition on dates	118,948,725.0	4.0456	0.8628	460.4421	4.92
3–4	Non-linear components	148,212,330.8	−121.7175	0.8458	513.982	5.64
1–4	DOY	141,968,197.4	−98.4265	0.8475	502.7834	5.64
1–2–3–4	All	118,098,505.7H	12.1687	0.8640	458.7604	4.89

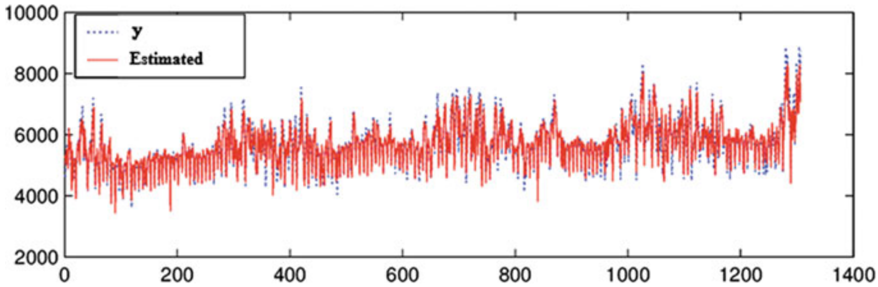


Fig. 2 Curve profile—prediction at city level (training sample)

4 Conclusions

This research proposed the hypothesis that the way in which temporary variables are presented to a system for predicting electrical consumption influences the quality of the results. This hypothesis was validated using different methods to represent these variables, applying it together with climate data and electricity consumption of the city of Bogota. The division of the temporary variable into day, day of week, month

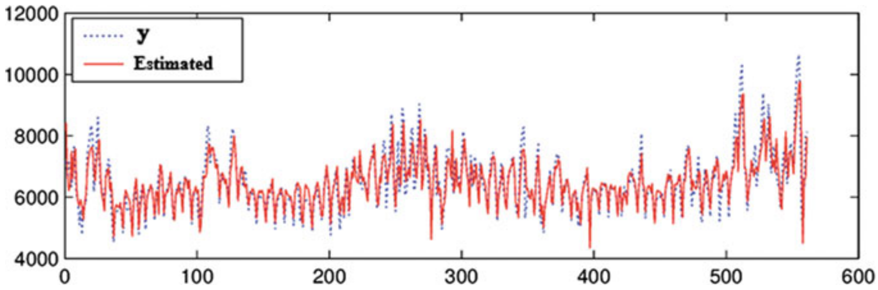


Fig. 3 Curve profile—prediction at city level (validation sample)

and year individually for each period involved in the problem, turned out to be the most convenient method, obtaining an improvement of up to 10.56% with respect to other methods. The use of all temporary variables available together, resulted in a reduced training error and validation. This implies that none of the temporary variables introduce noise in the prediction for the used data set.

References

1. Perez R et al (2018) Fault diagnosis on electrical distribution systems based on fuzzy logic. In: Tan Y, Shi Y, Tang Q (eds) Advances in swarm intelligence. ICSI 2018. Lecture notes in computer science, vol 10942. Springer, Cham
2. Silva V, Jesús A (2013) Indicators systems for evaluating the efficiency of political awareness of rational use of electricity. In: Advanced materials research, vol 601. Trans Tech Publications, Switzerland, pp 618–625
3. Perez R, Inga E, Aguila A, Vásquez C, Lima L, Viloria A, Henry MA (2018) Fault diagnosis on electrical distribution systems based on fuzzy logic. In: International conference on sensing and imaging, June. Springer, Cham, pp 174–185
4. Perez R, Vásquez C, Viloria A (2019) An intelligent strategy for faults location in distribution networks with distributed generation. J Intell Fuzzy Syst 36:1627–1637 (Preprint)
5. Xue Y, Lai Y (2016) The integration of great energy thinking and big datas thinking: big data and electricity big data. Power Syst Autom 40(1):1–8
6. Wang Y, Chen Q, Kang C et al (2017) Clustering of electricity consumption behaviour dynamics toward big data applications. IEEE Trans Smart Grid 7(5):2437–2447
7. Liu R, Ding W (2011) Statistical analysis and application of SAS. China Machine Press, China
8. Ozger M, Cetinkaya O, Akan OB (2017) Energy harvesting cognitive radio networking for IoT-enabled smart grid. Mob Netw Appl 23(4):956–966
9. Isasi P, Galván I (2004) Redes de Neuronas Artificiales. Un enfoque Práctico. Pearson, London. ISBN: 8420540250
10. Mangasarian O (1997) Arbitrary-norm separating plane. Technical report 97-07, Computer Science Dept., Univ. Wisconsin Madison
11. Andersson M (2009) A comparison of nine PLS1 algorithms. J Chemometr 23:518–529
12. Xu Q-S, Liang Y-Z (2001) Monte Carlo cross validation. Chemometr Intell Lab 56:1–11
13. Li H-D, Xu Q-S, Liang Y-Z (2017) A phase diagram for gene selection and disease classification. Chemometr Intell Lab 167:208–213
14. Cao DS, Liang YZ, Xu QS, Li HD, Chen X (2010) A new strategy of outlier detection for QSAR/QSPR. J Comput Chem 31:592–602

15. Tian W (2017) The research into methods of map building and path planning on mobile robots. In: IEEE 2nd information technology, networking, electronic and automation control conference (ITNEC 2017), pp 1087–1090
16. Dugas M et al (2014) Missing semantic annotation in databases. Methods Inf Med 53(6):516–517
17. Silberschatz A, Korth HF, Sudarshan S et al (1997) Database system concepts, vol 4. McGraw-Hill, New York

Forecast of the Demand for Hourly Electric Energy by Artificial Neural Networks

Amelec Viloria, Alberto Roncallo Pichon, Hugo Hernandez-P, Osman Redondo Bilbao, Omar Bonerge Pineda Lezama and Jesús Vargas

Abstract Obtaining an accurate forecast of the energy demand is fundamental to support the several decision processes of the electricity service agents in a country. For market operators, a greater precision in the short-term load forecasting implies a more efficient programming of the electricity generation resources, which means a reduction in costs. In the long term, it constitutes a main indicator for the generation of investment signals for future installed capacity. This research proposes a prognostic model for the demand of electrical energy in Bogota, Colombia at hourly level in a full week, through Artificial Neural Network.

Keywords Forecasting · Electric load · Artificial neural networks

A. Viloria (✉) · J. Vargas
Universidad de la Costa, St. 58 #66, Barranquilla, Atlántico, Colombia
e-mail: aviloria7@cuc.edu.co

J. Vargas
e-mail: jvargas41@cuc.edu.co

A. R. Pichon
Corporación Universitaria Minuto de Dios—UNIMINUTO, Barranquilla, Colombia
e-mail: alberto.roncallo@uniminuto.edu

H. Hernandez-P · O. R. Bilbao
Corporación Universitaria Latinoamericana, Barranquilla, Colombia
e-mail: hhernandez@ul.edu.co

O. R. Bilbao
e-mail: oredondo@ul.edu.co

O. B. P. Lezama
Universidad Tecnológica Centroamericana (UNITEC), San Pedro Sula, Honduras
e-mail: omarpineda@unitec.edu

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_46

1 Introduction

The study of forecast models for electric power demand is wide and varied, including methods such as the classical statistical analysis using ARIMA models, exponential smoothing, and regression with main component analysis [1]. Also, heuristic methods such as artificial neural networks are used [2–5], as well as hybrid methodologies between neural networks and wavelet analysis [6] and neural networks together with information theory [7]. Other more recent approaches include vector support machines [8], combining them with diffuse clustering [9], and the use of wavelet decomposition and Kalman filters [10].

In general, the studies on this field are developed for forecasts with data available with a week of lag and seek to forecast energy demand for a single day, or monthly [11–15]. In contrast, the model proposed in this paper considers a lag in the information of about three weeks according to the regulation and, at the same time, usually takes time to get the actual information from the recording areas to the system operator in order to forecast the time demand for a whole week (168 h). For this purpose, the study proposes a neural network structure that retrains each time a forecast is run and that considers both quantitative historical information (associated with the lag data of the series) and qualitative (associated with calendar events that produce significant changes in the energy demand throughout the year).

2 Method

2.1 Designing the Forecast Model

The forecast model proposed in this paper consists of an Artificial Neural Network with a Multilayer Perceptron structure (described in number 2) with a single hidden layer using the tangent—sigmoidal function as activation function of that layer, specified in Eq. (1) [16].

$$f(x) = \text{tansig}(x) = \frac{2}{1 + \exp(-2x)} - 1 \tag{1}$$

Energy demand is defined as output at a specific time of day, which means that, to predict a full week, the model must be executed once during each period of the day and during each day of the week, that is, 168 times.

2.2 Definition of Inputs

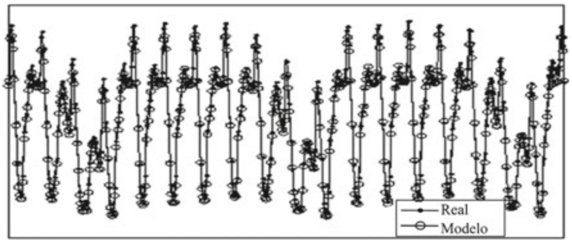
For the inputs of the model, two sources of available information, one quantitative and the other qualitative, stand out for helping in the characterization of the demand along the prediction horizon. The first one refers to the historical data related to the time series considering the analysis of the lags made in number 3, as well as the actual availability of the data. The second one is a tag assignment to each type of day depending on its location in the week or the particular event that is celebrated on this date. Such labels correspond to a classification of the days of the year according to their typical behavior (identified from expert knowledge), taking values such as "normal week day", "Sunday", "Holy Friday", "first of January", etc. In total, 42 labels were considered, which for confidentiality reasons cannot be mentioned in detail [17, 18].

In total, the selected entries were 11, of which 3 are return entries so that the seasonal dynamics of the series can be captured; 7 are categorical so that the qualitative information of special events can be consistently interpreted by the model and the remaining input is a normally distributed random variable that acts as a persistent excitation to the model. Formally, if $L_{h,d}$ is the demand of the hour h of the day d that will be forecasted, the input variables are [19, 20]:

- $L_{h,d-21}$: Demand of the same time three weeks ago
- $L_{h,d-364}$: Demand for the same time 52 weeks ago, that is approximately one year
- $L_{h^*,d-21}$: Average demand of the day, three weeks ago
- d: Type of day you want to forecast
- $d - 1$: Type of day previews to the forecast
- $d + 1$: Type of day after the forecast
- $d - 21$: Type of lag day three weeks before to the day forecast
- $d - 364$: Type of lag day 52 weeks before the day forecast
- h: Period of the day to be forecasted
- Week number of the year in which the forecast will be developed
- Normally distributed random variable N(0,1).

Since the model intends to calculate the forecast of a full week, and an important part of the inputs to the proposed model are lags, it is advisable to obtain demand information about the immediately preceding week. However, there are several operational and regulatory restrictions that prevent the acquisition of this data set. For this reason, data with a minimum lag of three weeks are considered in the input variables.

Fig. 1 Forecast result (portion of validation data)



3 Results

3.1 Training Algorithm

The proposed model was applied in Matlab using the toolbox of Neural Networks and, to define the best training algorithm and the number of neurons that the hidden layer should contain, the procedure described below was used.

The available historical data (around five years) for a specific PCU were taken and divided into two sets: training data and validation data, in a proportion of 80% and 20% respectively. With these data organized in input-output tuples defined according to the previous section, the following four training algorithms were tested to verify which generated the best results in terms of the series adjustment: Retro-propagation, Levenberg-Marquardt, Bayesian Regulation, and Gradient Conjugate scaling. Models were trained for each of these algorithms with a different number of neurons in the hidden layer, starting in 4 neurons and then increasing this value by to reach 40 neurons [21, 22].

This exhaustive and iterative procedure allowed to optimize the number of neurons balancing the performance and execution time. The results obtained are shown in Fig. 1, considering the Absolute Average Percentage Error (MAPE) as performance criterion showing the average percentage deviation of each hourly forecast of the validation data in relation to the real demand values [6]. From the data analysis, it was possible to establish that the best combination training algorithm—number of neurons is the Levenberg-Marquardt algorithm with 16 neurons in the hidden layer, being this configuration the one selected for the proposed forecast model.

3.2 Validation of the Proposed Model

To validate the proposed model, real historical data of 5 years were taken and forecasts were made at different points of interest, dividing them into training and validation data to the same proportion. From the results obtained, shown in Fig. 1, it can be

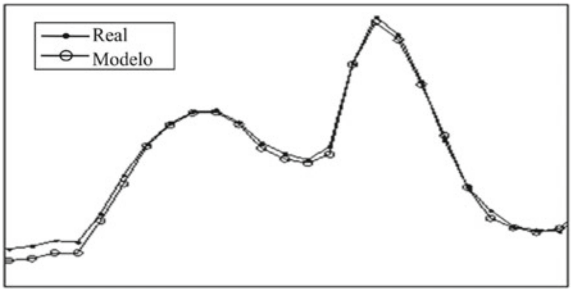


Fig. 2 Forecast of a holiday Monday

Table 1 Results of the forecast models

Forecast method	MAPE (training) (%)	MAPE (forecast) (%)
SR	3.77	4.13
SRX	2.88	3.24
Neural networks	1.96	2.33

concluded that the model captures the seasonal and weekly components of the series. Another important point is that the model can accurately approximate special days that are not found in seasonal patterns, such as holiday Mondays (see Fig. 2).

To provide more validity to the results, two additional models were applied in order to contrast them with the proposed model. The first one consists of a self-regressive model (SR) whose inputs are the first two variables of the RNA model (the lagged variables three weeks and a year), while the second is a self-regressive model with exogenous variables (SRX) whose inputs are exactly the same as the RNA model. Both models were adjusted by means of least squares.

To make the comparison, a forecast of about 320 days was generated with a delay of three weeks. Table 1 shows the results obtained in terms of the MAPE for the three models.

When performing the corresponding statistical analysis for the proposed RNA model, an asymmetry coefficient of 19.9 and a kurtosis coefficient of 281.5 were determined. These data together with the information provided by the mustache box plot presented in Fig. 3 indicate that the errors are not normal. Figure 4 shows the behavior of the errors for the validation data of the RNA model. The series of errors show a behavior around the mean, however, the presence of peaks, systematic behaviors (indicated by lines), and presence of periods of variable volatility (indicated by circles) are noted.

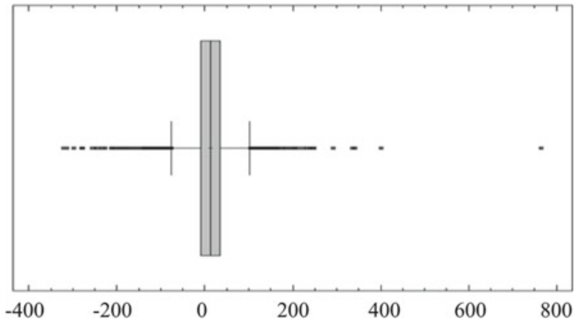


Fig. 3 Mustache box plot for the errors of the neural networks model

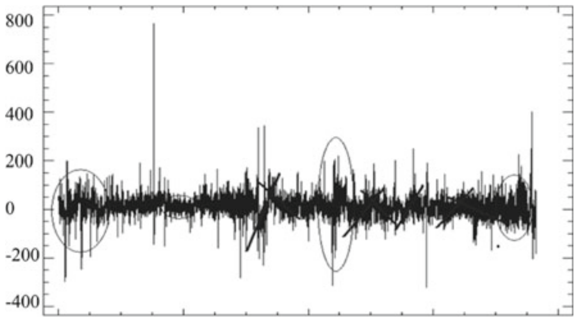


Fig. 4 Residuals of the validation set of the proposed model

4 Conclusions

After defining the model entries, an iterative procedure was performed to find the best combination of training algorithm and number of hidden layer neurons, resulting in the Levenberg-Marquardt algorithm with 16 neurons. With the defined model, several validation tests were carried out, the first ones regarding the adjustment of the model and its forecasting capacity, and the second ones to contrast them with traditional models such as SR and SRX. In both cases, MAPE was used as a measure of performance and the usefulness of the proposed model was demonstrated graphical and quantitatively.

As a conclusion regarding the input variables, it is important to add additional information to the regressors of the same series of demand in the form of explanatory variables to make a more adjusted forecast of special days.

References

1. Perez R et al (2018) Fault diagnosis on electrical distribution systems based on fuzzy logic. In: Tan Y, Shi Y, Tang Q (eds) Advances in swarm intelligence. ICSI 2018. Lecture notes in computer science, vol 10942. Springer, Cham

2. Silva V, Jesús A (2013) Indicators systems for evaluating the efficiency of political awareness of rational use of electricity. In: Advanced materials research, vol 601. Trans Tech Publications, Switzerland, pp 618–625
3. Perez R, Inga E, Aguila A, Vásquez C, Lima L, Viloria A, Henry MA (2018) Fault diagnosis on electrical distribution systems based on fuzzy logic. In: International conference on sensing and imaging, June. Springer, Cham, pp 174–185
4. Perez R, Vásquez C, Viloria A (2019) An intelligent strategy for faults location in distribution networks with distributed generation. J Intell Fuzzy Syst 36:1627–1637 (Preprint)
5. Isasi P, Galván I (2004) Redes de Neuronas Artificiales. Un enfoque Práctico. Pearson, London. ISBN: 8420540250
6. Kulkarni S, Haidar I (2009) Forecasting model for crude oil price using artificial neural networks and commodity future prices. Int J Comput Sci Inf Secur 2(1):81–89
7. Mazón JN, Trujillo J, Serrano M, Piattini M (2005) Designing data warehouses: from business requirement analysis to multidimensional modeling. In: Proceedings of the 1st international workshop on requirements engineering for business need and IT alignment, Paris, France
8. Ben Salem S, Naouali S, Chtourou Z (2018) A fast and effective partitional clustering algorithm for large categorical datasets using a k-means based approach. Comput Electron Eng 68:463–483. https://doi.org/10.1016/j.compeleceng.2018.04.023
9. Chakraborty S, Das S (2018) Simultaneous variable weighting and determining the number of clusters—a weighted Gaussian algorithm means. Stat Probab Lett 137:148–156. https://doi.org/10.1016/j.spl.2018.01.015
10. Abhay KA, Badal NA (2015) Novel approach for intelligent distribution of data warehouses. Egypt Inform J 17(1):147–159
11. Abdul Masud M, Zhexue Huang J, Wei C, Wang J, Khan I, Zhong M (2018) I-nice: a new approach for identifying the number of clusters and initial cluster centres. Inf Sci. https://doi.org/10.1016/j.ins.2018.07.034
12. Rahman MA, Islam MZ, Bossomaier T (2015) ModEx and Seed-Detective: two novel techniques for high quality clustering by using good initial seeds in K-means. J King Saud Univ Comput Inf Sci 27:113–128. https://doi.org/10.1016/j.jksuci.2014.04.002
13. Maren AJ, Harston CT, Pap RM (2014) Handbook of neural computing applications. Academic Press, San Diego
14. Baughman DR, Liu YA (2014) Neural networks in bioprocessing and chemical engineering. Academic Press, San Diego
15. Fast M, Assadi M, De S (2009) Development and multi-utility of an ANN model for an industrial gas turbine. Appl Energy 86(1):9–17
16. Chatterjee S, Sarkar S, Hore S, Dey N, Ashour AS, Balas VE (2016) Particle swarm optimization trained neural network for structural failure prediction of multistoried RC buildings. Neural Comput Appl 28:2005–2016
17. Tüfekci P (2014) Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods. Int J Electr Power Energy Syst 60:126–140
18. Samanta S, Acharjee S, Mukherjee A, Das D, Dey N (2013) Ant weight lifting algorithm for image segmentation. In: 2013 IEEE international conference on computational intelligence and computing research (ICCIC), pp 1–5
19. Jagatheesan K, Anand B, Dey N, Ashour AS (2015) Artificial intelligence in performance analysis of load frequency control in thermal-wind-hydro power systems. Artif Intell 6(7)
20. Kalogirou SA (2001) Artificial neural networks in renewable energy systems applications: a review. Renew Sustain Energy Rev 5:373–401
21. Laha P, Chakraborty B (2017) Energy model—a tool for preventing energy dysfunction. Renew Sustain Energy Rev 73:95–114
22. Moldes O, Mejuto J, Rial-Otero R, Simal-Gandara J (2017) A critical review on the applications of artificial neural networks in winemaking technology. Crit Rev Food Sci Nutr 57:2896–2908

Electric Consumption Pattern from Big Data

Amelec Viloria, Hugo Hernandez-P, Omar Bonerge Pineda Lezama and Victoria Diago Orozco

Abstract From the concept of smart grid, reaching an efficient and reliable network is a task that implies several stages and sub-stages with a defined and specific mission. In this way, the intelligent measurement stage conformed by the smart meters obtains the information of electrical consumption from the users or consumers (residential, commercial, and industrial). For this purpose, a smart metering infrastructure made of wireless telecommunications and fiber optic has been generated allows to guarantee the connectivity of the smart meters and the central office of electric companies. This paper aims to describe the use of MapReduce as a technique to obtain information about the load curve at an appropriate time to obtain trends and statistics related to the pattern of residential electricity consumption.

Keywords Big data · MapReduce · Meter data management system (MDMS) · Smart grid · Smart metering

1 Introduction

The smart metering infrastructure is closely related to the system of measured data management. It is where the amount of information takes a real importance and a nontrivial problem that requires a treatment of information with the lowest cost for the resources used in the process of information processing. To transport the information from the smart meters, heterogeneous wireless network planning models have been

A. Viloria (✉) · V. D. Orozco
Universidad de la Costa, St. 58 #66, Barranquilla, Atlántico, Colombia
e-mail: aviloria7@cuc.edu.co

V. D. Orozco
e-mail: vdigo2@cuc.edu.co

H. Hernandez-P
Corporación Universitaria Latinoamericana, Barranquilla, Colombia
e-mail: hhernandez@ul.edu.co

O. B. P. Lezama
Universidad Tecnológica Centroamericana (UNITEC), San Pedro Sula, Honduras
e-mail: omarpineda@unitec.edu

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_47

generated, represented in a multilayer-multigraph to bring information with the best coverage of the smart meters and consider low cost technologies in relation to other options such as the cellular network, which is limited by the lack of coverage in suburban and rural areas [1–4].

In this sense, the optimal deployment of a wireless network through multi-haul has been considered as the best option, even in relation to initial studies that propose the hybrid use of fiber and wireless network where the deployment cost of a buried fiber optic network is expensive to be taken into account for an intelligent measurement of electrical energy, which is studied in [5]. In this way, an intelligent measurement infrastructure, articulated with the data management system measured in its different topologies such as: centralized, semi-distributed, and fully distributed, allows optimizing the use of the communication network capacity since, at this stage of MDMS, can reduce information that is not relevant for the decision-making process and does not require transportation. In the MDMS, the processes of encryption and information security are also articulated due to the use of important information that could be used for criminal purposes because of the importance of knowing the details of a user's energy consumption pattern [6, 7].

Thus, the amount of information stored with users' energy consumption exceeds gigabytes, which affects the need to look for a parallel computing process that reduces the cost in terms of resources used for the time spent in this type of information processing. In other words, a data mining technique allows to reduce the times in information processing and obtain a quick report for a subsequent decision making to improve the response of the demand according to what is proposed by Sun et al. [8].

An MDMS integrates and stores information from a smart metering, so, what is observed is a need for scalability in terms of the storage and processing of information and the growth of physical resources. But the problem of this study focuses on the management of information previously stored in silos. This information processing requires a certain time in order to obtain reports and/or statistics to determine the behavior of electricity consumption, and the use of MapReduce as a technique for handling big data which through parallel programming increases computational performance and reduces processing costs.

2 Method

The issue of big data management in MDMS requires a recursive process in parallel that allows to acquire all or part of the information according to the requirements for the analysis or desired reports for the later decision-making process for responding to electricity demand. Therefore, the big data analysis from the smart metering of electrical energy is not a trivial problem which requires an information management system that responds in a timely and scalable way and that allows an allocation of resources at lowest costs, considering the Capacity constraints of each MDMS,

Table 1 Variables algorithm—GDM-MapReduce

First name	Interpretation
Bigdata.csv	
Inputs	Data file
ds	Data variable
Field1	Field of the data file
Field2	Field of the data file
sumLen	Internal mapreduce summer
Outputs	
Outds	Data variable
KeyValueStore	Internal variable to form the packages

through an adequate report manager that allows to obtain trends in the behavior of electricity consumption [9–11].

The GDM-MapReduce algorithm takes an extension file.csv that is approximately 11 GB in size as an input and stores it in the variable called ds. This variable contains the built-in search criteria. From that moment, all the programming references are oriented towards the ds variable. Subsequently, these data must be prepared to be mapped, that is, the data should be prepared (without blank spaces within the fields) so that it forms smaller data packages and can work in a timely way. The MapReduce functions are used to create the packages using the key-value method. Table 1 shows the variables presented by the algorithm [12–14].

Once the package is formed (mapped), the next step is the step reduce that is responsible for carrying out the operations and save the partial results for later calculations. This process is iterated again until it is completed with all the data stored in the ds. The pseudocode of the GDM-MapReduce algorithm is shown in Fig. 1. It is important to bear in mind that the restrictions or selection criteria that will form a subset of the records that need to be analyzed, will require a greater time of MapReduce management, that is to say, more search restrictions the algorithm will use more processing time and, therefore, greater consumption of energy used in the process.

3 Results

The construction of a recursive process in parallel has allowed the computational time to be reduced depending on the treatment of the desired information. In this way, it is estimated that there are processes that can take a longer or shorter time according to the analysis to be performed. This study found, as a relevant contribution, that a process that delays or minimizes the processing time is in the filtering stage of the information. Therefore, the more restrictions on the search space, the more the processing time increases. The study was carried out by using order data of 11 GB,

```
Entrada: Bigdata.csv, ds, campo1, campo2, sumLen
Salida: outds, KeyValueStore
Paso 1: Carga de datos
ds= Almacenar(bigdata.csv)
Paso 2: Proceso de datos
leer_por_grupos(ds)
Visualizar cabeceras (ds)
ds=Convertir datos en blanco a NaN (bigdata.csv,NA)
ds=Convertir datos faltantes/perdidos a
NaN(bigdata.csv,missingvalue)
Paso3:  Clasificación de datos (por clave_valor)
add(intermKV, campo ,ds.campo)
addmulti(intermKV,{campo1 campo2},{ds.campo1 ds.campo2})
Paso 4: Mapeo
Utilizar las funciones hasnext y getnext para recuperar los valores
asociados por una clave única denominada KeyValueStore
While hasnext(sumLenIter)
sumLen = sumLen + getnext(sumLenIter)
End
Paso 5: Reducción
Determinación de variables por cada grupo
outds = mapreduce(ds)
volver a iterar para otro grupo
Go paso 2
Paso 6: Resultados
Agrupación de resultados por grupos
Determinación de resultados globales
```

Fig. 1 Pseudocode of the GDM-MapReduce algorithm, the original language of the code is preserved (Spanish)

which was processed under the MATLAB environment, obtaining clear and relevant results for the response analysis of the electric power demand, as detailed in Fig. 2.

Figure 2 shows the behavior of average consumption in KWh with respect to the variation in the timeline, obtained from the MapReduce technique which allows knowing the variation of energy consumption from historical data. It can be applied to the types of customers of each electricity distribution company and generate an analysis of energy demand based on historical data, thus achieving a deterministic trend line for the response to the demand for electricity. In turn, it would allow to reduce the indices of electrical generation by working just on the demand because the costs for not consumed energy is significant for a generating company.

Table 2 shows the processing time and energy consumption of the MapReduce stage for a big data of 11 GB, obtained from a case study applied to the free-use big data obtained from the London Data Store website (see Networks, 2017). The table shows the time required by the MapReduce technique to explore and extract information from three types of electric power users referred by type 1 UT1-2017, type 2 (UT2)-2018. Additionally, the energy consumption is presented in each process P1-2017, P2-2018 in watts. The registry number evaluated were 1,074,204 unevenly distributed among twelve months. In this period, it was identified that the analyzed

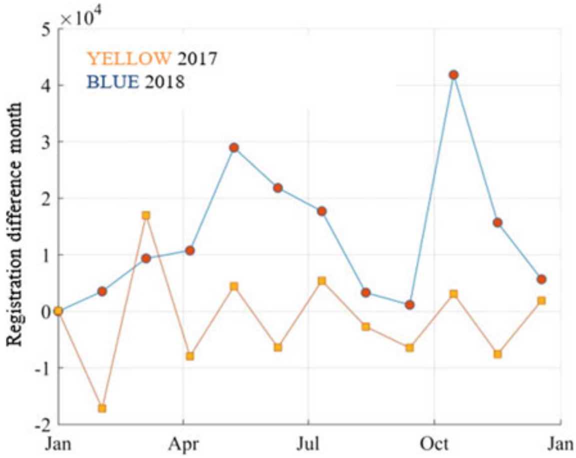


Fig. 2 Variability analysis of electricity consumption

big data had a record every 30 s, many of which had a value of 0, due to the failures of the communication system or due to a fault in the smart meter. The computer used to execute the algorithm has an I7 processor, 3.4 GHz processor, 8 MB RAM, 19.5 V, 3.34 Amp and developed under MATLAB R2016b environment [15–17].

4 Conclusions

A behavior pattern of the electrical consumption can be analyzed from the management of the stored historical data, obtained from the smart metering of electrical energy. The behavioral pattern allows to deterministically project the future behavior of electric power consumption. The research shows how the MapReduce technique, using the process named "divide and conquer" with a data of 11 GB of information, manages to obtain results of large amounts of data that is not read in a simple process due to memory limitations.

The MapReduce technique, applied to the management of big data built by historical data of smart metering of electric power, presents a great option to improve the performance and thus reconstruct the pattern of energy consumption, optimizing the information management process required in the decision-making and response to electric power demand.

Table 2 MapReduce tasks

		MapReduce Task							
		CPU Time				Consume of Energy Watts			
		UT1 2017	OR	UT2 2018		P1 2017		P2 2018	
Areas T	T1	54 h	6	58 h	3.14	76.95	9	51.3	7

References

1. Sanchez L, Vásquez C, Viloria A, Cmeza-Estrada (2018) Conglomerates of Latin American countries and public policies for the sustainable development of the electric power generation sector. In: Tan Y, Shi Y, Tang Q (eds) Data mining and big data. DMBD 2018. Lecture notes in computer science, vol 10943. Springer, Berlin
2. Perez R, Inga E, Aguila A, Vásquez C, Lima L, Viloria A, Henry MA (2018) Fault diagnosis on electrical distribution systems based on fuzzy logic. In: International conference on sensing and imaging. Springer, Berlin, pp 174–185
3. Perez R, Vásquez C, Viloria A (2019) An intelligent strategy for faults location in distribution networks with distributed generation. J Intell Fuzzy Syst 1–11
4. Chakraborty S, Das S (2018) Simultaneous variable weighting and determining the number of clusters—a weighted Gaussian algorithm means. Stat Probab Lett 137:148–156. https://doi.org/10.1016/j.spl.2018.01.015
5. Bontempi G, Ben Taieb S, Borgne YA (2013) Machine learning strategies for time series forecasting. In: Aufaure MA, Zimányi E (eds) Lecture notes in business information processing, vol 138, no 1. Springer, Heidelberg, pp 70–73
6. Abdul Masud M, Zhexue Huang J, Wei C, Wang J, Khan I, Zhong M (2018) Inice: a new approach for identifying the number of clusters and initial cluster centres. Inf Sci. https://doi.org/10.1016/j.ins.2018.07.034
7. Sánchez L, Vásquez C, Viloria A, Rodríguez Potes L (2018) Greenhouse gases emissions and electric power generation in Latin American Countries in the Period 2006–2013. In: Tan Y, Shi Y, Tang Q (eds) Data mining and big data. DMBD 2018. Lecture notes in computer science, vol 10943. Springer, Berlin
8. Sun M, Konstantelos I, Strbac G (2017) C-vine copula mixture model for clustering of residential electrical load pattern data. Power Syst IEEE Trans On 32(3):2382–2393
9. Perez R et al (2018) Fault diagnosis on electrical distribution systems based on fuzzy logic. In: Tan Y, Shi Y, Tang Q (eds) Advances in swarm intelligence. ICSI 2018. Lecture notes in computer science, vol 10942. Springer, Berlin
10. Silva V, Jesús A (2013) Indicators systems for evaluating the efficiency of political awareness of rational use of electricity. In: Advanced materials research, vol 601. Trans Tech Publications, pp 618–625
11. Kim M, Park S, Han K, Kim N, Kyun Choi J (2018) Dynamics of electricity consumers for classifying power consumption data using PCA. In: Big data and smart computing (BigComp) 2018 IEEE international conference on, pp 697–700
12. Deng ZH, Yu H, Yang Y (2016) Identifying sentiment words using an optimization model with L1, regularization. In: Thirtieth AAAI conference on artificial intelligence. AAAI Press pp 115–121
13. Alshareef SM, Morsi WG (2017) Probabilistic commercial load profiles at different climate zones. In: Electrical power and energy conference (EPEC) 2017 IEEE, pp 1–7
14. Gao XZ, Kaarna A, Lensu L, Honkapuro S (2017) A hybrid method for short-term electricity consumption prediction. In: Industrial electronics society IECON 2017—43rd annual conference of the IEEE, pp 7393–7398
15. Chen T, Qian K, Mutanen A, Schuller B, Järventausta P, Su W (2017) Classification of electricity customer groups towards individualized price scheme design. In: Power symposium (NAPS) 2017 North American, pp 1–4
16. Lv Z, Han Z, Yang L, Gang W (2017) An improved CFSFDP algorithm with cluster center automatically selected based on weighted average method. In: 2017 IEEE 7th annual international conference on CYBER technology in automation control and intelligent systems (CYBER), pp 955–959
17. Adiyoso W, Krisnadhi A, Wibisono A, Purbarani SC, Saraswati AD, Putri AFR, Saladdin IR, Anwar SRC (2018) Time performance analysis of multi-CPU and multi-GPU in big data clustering computation. In: 2018 international workshop on big data and information security (IWBIS), pp 113–116

Forecast of Operational Data in Electric Energy Plants Using Adaptive Algorithm

Amelec Viloria, Jesús García Guiliany, Hugo Hernandez-P, Luis Cabas Vásquez and Omar Bonerge Pineda Lezama

Abstract Traditional time series methods offer models whose parameters remain constant over time. However, industrial supply and demand processes require timely decisions based on a dynamic reality. A change in configuration, turning off, or on a production line or process, modifies the problem and the variables to be predicted. Decision support systems must dynamically adapt in order to respond quickly and appropriately to operations and their processes. This methodology is based on obtaining, for each period, the model that best fits the data, evaluating many alternatives and using statistical learning techniques. In this way, the model will adapt to the data in practice and make decisions based on experience. With three months of testing for the estimation of variables associated with supply and demand processes, predictions that differ less than 8 hundredths (less than 0.08) or 0.1% of the measured value were obtained. This indicates that data science and statistical learning represent an important area of research for variable prediction and process optimization.

Keywords Time series models · Estimation · Forecasts · Data analysis · Data mining · Statistical learning · Decision trees

---

A. Viloria (✉)
Universidad de la Costa, St. 58 #66, Barranquilla, Atlantic, Colombia
e-mail: aviloria7@cuc.edu.co

J. G. Guiliany
Universidad Simón Bolívar, Barranquilla, Colombia
e-mail: jesus.garcia@unisimonbolivar.edu.co

H. Hernandez-P · L. C. Vásquez
Corporación Universitaria Latinoamericana, CUL, Barranquilla, Colombia
e-mail: hhernandez@ul.edu.co

L. C. Vásquez
e-mail: lcabas@ul.edu.co

O. B. P. Lezama
Universidad Tecnológica Centroamericana (UNITEC), San Pedro Sula, Honduras
e-mail: omarpineda@unitec.edu

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_48

1 Introduction

Traditionally, the identification and selection of models is carried out by describing the series in its trend and seasonality components. The data scientist calculates and analyzes means, variances, autocovariance functions, and simple and partial autocorrelation [1–4], to raise and evaluate a set of transformations for the series. To determine the coefficients of the components Self-Regressive (p), degree of differentiation (d) and moving averages (q), of an Integrated Self-Regressive Moving Averages, ARIMA (p, d, q), the number of lags with out-of-band values in simple and partial autocorrelation functions is analyzed, for each of the differentiations and transformations applied to the original series, and a few tentative models are proposed to be evaluated [5, 6].

Once the parameters (p, d, q) are selected and the model is constructed, they are fixed and constant in time for all the future forecasts. But the reality says that this does not always work that way. The phenomena to be modeled change over time, and therefore, models must be adjusted as well [7, 8].

Analysts, process engineers, and operational managers need to make timely decisions based on a very dynamic reality. A change in the configuration of a plant or one of its processes changes the conditions of the problem and the variables to be predicted, and the systems that support the decision-making must be dynamically adapted in order to respond in a timely and appropriate way to the dynamics of operations [9, 10].

For this purpose, those responsible for controlling the processes, need agile and adaptive tools to analyze thousands of alternatives in a short time and present the best at the different levels where decision-making is required. All of this has motivated to design and develop a decision support system, based on time series analysis and decision trees to predict the value of industrial variables [11, 12].

This paper proposes a method based on adaptive algorithms that allow the selection of the best method of time series estimation and the dynamic variation of parameters. This method was designed and tested to estimate average monthly cycle values for certain industrial processes in the energy field, supporting tactical and operational decision-making. Currently, its adaptation for daily cycle forecasts is studied. Respecting to the natural criterion of confidentiality of the industrial data associated with the processes to be modeled, this paper seeks to document and expose the method used for the construction of this novel approach.

2 Method

From each series studied, the best models were selected in each execution, whether daily, weekly, monthly or another period, depending on the characteristics of the data and the series. For each period to be predicted, the selected models of each series and their forecasts are stored in a database to evaluate their effectiveness by retrospectively comparing the forecasts with the actual data. In this way, the effectiveness of all models selected by the machine and those selected by the analyst in each cycle were evaluated. This data is what the learning algorithm analyzes to provide the user more information about the best forecasting model based on their own previous experience [13].

2.1 Exhaustive Search

Having defined the series to be evaluated, a set of Models of Autoregressive Integral of Moving Averages, ARIMA (p, d, q) was selected for each, for $0 \leq p \leq 12$, $0 \leq d \leq 5$, $0 \leq q \leq 12$. This set originally provides a total of 1,014 models for each series, which reduces depending on the models that the machine eliminates or by decision of the expert when it comes to adjust the algorithm search rules for a series. It is also possible that, based on the experience and evolution that is noticed in each cycle, some transformations will be discarded and/or others will be incorporated into the test set [14–16].

2.2 Discarding Models

It is possible that, for a particular model, that is, a triplet (p, d, q), the coefficients of the model cannot be calculated. In other words, for these conditions there is no possible solution, so the model is discarded for all future runs. This is performed with the intention to 'prune' the search set as the forecasts evolve. Finally, the prevailing models will be those defined for all periods calculated from the first run for this set of series [13].

2.3 Segmented Selection

For each evaluated series, the algorithm will select the four best models, based on the criteria above. In the first instance, select the ones with the best AIC coefficient. The ones that are best in the Box-Pierce test are then selected. Then, from this small group, it intercepts the set of those that were best in the residue normality tests against those closest to the last value in the series (element n, last known value) [16].

2.4 Final Selection by the Data Analyst

Finally, these four pre-selected models for each series are presented to the data analyst to make the final decision and make the forecasts. The analyst can evaluate the machine's recommendation and validate some results, such as the histogram of the residuals of each preselected model [15].

2.5 Database Aspects

The developed program reads the source data from a database with the historical values associated with the processes to be worked on and builds the set of series with the transformations previously defined by the expert. These configuration data are also brought into the database [13].

2.6 Statistical Learning

Whether the study provides a lot of historical data to run the algorithm with previous training data, or the history is built with the forecasts made since its launch, the use of learning techniques to improve the final selection of models is a very valuable aspect [5].

3 Results

With this method and the algorithm developed and implemented, monthly average estimates were made for operational data from some plants in Colombia's energy sector.

During the test period, results were obtained with absolute differences of less than eight hundredths (<0.08) and relatives less than 0.1% with respect to the actual value, measured and verified by the department responsible for the corresponding operational area (see Fig. 1).

Table 1 shows some periods of learning and how the algorithm was adjusting the parameters. Likewise, the table presents how some models are reselected over time and, in some cases, the models are very close to each other.

Once the models are pre-selected, the analyst can observe the different results obtained and compare them to each other. Additionally, they can perform more tests if they deem it necessary. In Table 2, an example of some models preselected by the algorithm and the comparison of the resulting values can be seen in one of the applied tests.

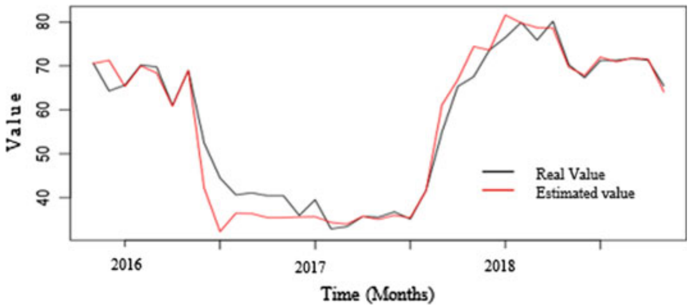


Fig. 1 Real data versus forecast performed by algorithm

Table 1 Model evolution during training

Year	Series	Q	D	Q	Difference	AIC
2016	2	11	0	8	1.36	836.12
2017	4	12	0	6	1.1	869.14
2018	3	10	0	12	0.85	865.43
2019	5	11	0	10	0.32	865.47

Table 2 Partial table of results for selection

Year	Q	D	Q	AIC	I'm slob. standard	Box-pierce	Forecast	Difference
2016	6	0	7	835.1	8.14	0.88	52.72	0.025
2017	7	1	8	835.1	7.25	0.89	52.74	0.048
2018	8	1	10	836.4	7.29	0.91	53.10	0.045
2019	5	0	5	862.1	8.14	0.95	53.01	0.035

For these resulting models, all the tests and analysis that is traditionally performed to the data can be applied, both to select the likely parameters of the model, and to statistically check the model. However, this subsequent analysis is performed not to select the model parameters but to validate those selected by the machine.

Figure 1 shows the behavior of a measure-based actual variable, against the forecasts that the algorithm was making over time. For each period, the algorithm changed the model to adapt to the new data.

Likewise, for the criterion called "Efficiency", referring to how close this model can predict the last value of the series with the n−1 elements above, provides a very good reference. However, when the standard deviation of the waste is smaller, too close, it will not be good either. On the other hand, when the standard deviation of the estimator is smaller, being very close to the last element of the series is again a very good signal. It is important to note that the rules that arise from this statistical learning establish new non-trivial relationships between these values that may be applicable to future scenarios in this same context.

Finally, for more detail, the following image presents the classification tree, with the criteria detailed in text mode (Algorithm 1).

Algorithm 1
Decision Tree

1) Root
2) AIC <739.4034 ?> Bad
3) AIC 739.4034
6) AIC 874.9885 > Bad
7) AIC <874.9885
6) Efficiency 2.793151 ?> Bad
14) Effectiveness <2.793151
30) Dev.Estd.Resíduos <8.549181
60) AIC <843.4874 ?> Bad
61) AIC 843,4874
122) Efficacy <0.8773045 ?> Bad
123) Efficiency 0.8773045 ?> Excellent
31) Dev.Estd.Resíduos 8.549181
62) Dev.Estd.Estimator 12.41628 ?> Bad
63) Desv.Estd.Estimator <12.41628
126) Efficiency 0.9645337
252) AIC 771.5318 ?> Bad
253) AIC <771.5318 ?> Excellent
127) Efficiency <0.9645337 ?> Excellent

4 Conclusions

This innovative and practical method allows to quickly obtain predictive models of great value and efficiency for the estimation of operational data, so analysts, engineers and operational managers can use better tools to make timely and more informed decisions. Storing these results in a forecast history offers wider analysis possibilities to continue building better business rules and better prediction and optimization methods in the future. The construction of rules using decision trees is not only adequate and easy to implement within decision-making algorithms, but are understandable by the functional user of the model and applicable in operational scenarios, helping to create new business rules.

References

1. Sanchez L, Vásquez C, Viloria A, Cmeza-Estrada (2018) Conglomerates of Latin American Countries and public policies for the sustainable development of the electric power generation sector. In: Tan Y, Shi Y, Tang Q (eds) Data mining and big data. DMBD 2018. Lecture notes in computer science, vol 10943. Springer, Berlin
2. Perez R, Inga E, Aguila A, Vásquez C, Lima L, Viloria A, Henry MA (2018) Fault diagnosis on electrical distribution systems based on fuzzy logic. In: International conference on sensing and imaging. Springer, Berlin, pp 174–185
3. Perez R, Vásquez C, Viloria A (2019) An intelligent strategy for faults location in distribution networks with distributed generation. J Intell Fuzzy Syst 1–11
4. Chakraborty S, Das S (2018) Simultaneous variable weighting and determining the number of clusters—a weighted Gaussian algorithm means. Stat Probab Lett 137:148–156. https://doi.org/10.1016/j.spl.2018.01.015
5. Bontempi G, Ben Taieb S, Borgne YA (2013) Machine learning strategies for time series forecasting. In: Aufaure MA, Zimányi E (eds) Lecture notes in business information processing, vol 138, no 1. Springer, Heidelberg, pp 70–73
6. Abdul Masud M, Zhexue Huang J, Wei C, Wang J, Khan I, Zhong M (2018) Inice: a new approach for identifying the number of clusters and initial cluster centres. Inf Sci. https://doi.org/10.1016/j.ins.2018.07.034
7. Sánchez L, Vásquez C, Viloria A, Rodríguez Potes L (2018) Greenhouse gases emissions and electric power generation in Latin American Countries in the period 2006–2013. In: Tan Y, Shi Y, Tang Q (eds) Data mining and big data. DMBD 2018. Lecture notes in computer science, vol 10943. Springer, Berlin
8. Sun M, Konstantelos I, Strbac G (2017) C-vine copula mixture model for clustering of residential electrical load pattern data. Power Syst IEEE Trans On 32(3):2382–2393
9. Perez R et al (2018) Fault diagnosis on electrical distribution systems based on fuzzy logic. In: Tan Y, Shi Y, Tang Q (eds) Advances in swarm intelligence. ICSI 2018. Lecture notes in computer science, vol 10942. Springer, Berlin
10. Silva V, Jesús A (2013) Indicators systems for evaluating the efficiency of political awareness of rational use of electricity. In: Advanced materials research, vol 601. Trans Tech Publications, pp 618–625
11. Kim M, Park S, Han K, Kim N, Kyun Choi J (2018) Dynamics of electricity consumers for classifying power consumption data using PCA.In: 2018 IEEE international conference on big data and smart computing (BigComp), pp 697–700
12. Chen S, Liu CC (2016) From demand response to transactive energy: state of the art. J Mod Power Syst Clean Energy
13. Sun M, Teng F, Konstantelos I, Strbac G (2018) An objective-based scenario selection method for transmission network expansion planning with multivariate stochasticity in load and renewable energy sources. Energy
14. Capizzi G, Sciuto GL, Napoli C, Tramontana E (2017) An advanced neural network based solution to enforce dispatch continuity in smart grids. Appl Soft Comput
15. Melzi F, Same A, Zayani M, Oukhellou L (2017) A dedicated mixture model for clustering smart meter data: identification and analysis of electricity consumption behaviors. Energies 10:1446
16. Chen T, Alsafasfeh Q, Pourbabak H, Su W (2017) The next-generation US retail electricity market with customers and prosumers—a bibliographical survey. Energies 11:8

Prediction of Academic Dropout in University Students Using Data Mining: Engineering Case

Jesús Silva, Luisa Fernanda Arrieta Matos, Claudia Medina Mosquera, Carlos Vargas Mercado, Rosio Barrios González, Nataly Orellano Llinás and Omar Bonerge Pineda Lezama

Abstract Student dropout is considered an important indicator for measuring social mobility and reflecting the social contribution that universities offer. In economic terms, there is evidence that students attribute their decision to defect from their academic programs because of their economic situation. Dropout causes significant waging gaps among people who complete their tertiary studies compared to those who do not, leading to a lack of skilled human capital that pays greater productivity to economic development of a country. Given the above, the objective of this study is to present a tree-based classification of decisions (CBAD) with optimized parameters to predict the dropout of students at Colombian universities. The study analyses 10,486 cases of students from three private universities with similar characteristics. The result of the application of this technique with optimized parameters achieved a precision ratio of 88.14%.

J. Silva (✉)
Universidad Peruana de Ciencias Aplicadas, Lima, Peru
e-mail: jesussilvaUPC@gmail.com

L. F. A. Matos
Universidad Simón Bolívar, Barranquilla, Colombia
e-mail: luisa.arrieta@unisimonbolivar.edu.co

C. M. Mosquera
Universidad de la Costa, St. 58 #66, Barranquilla, Atlántico, Colombia
e-mail: cmedina4@cuc.edu.co

C. V. Mercado
Corporación Universitaria Latinoamericana, Barranquilla, Colombia
e-mail: cvargas@ul.edu.co

R. B. González · N. O. Llinás
Corporación Universitaria Minuto de Dios—UNIMINUTO, Barranquilla, Colombia
e-mail: rosiobarriosgo@hotmail.com

N. O. Llinás
e-mail: nataly.orellano@uniminuto.edu

O. B. P. Lezama
Universidad Tecnológica Centroamericana (UNITEC), San Pedro Sula, Honduras
e-mail: omarpineda@unitec.edu

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_49

Keywords Student dropout · Classification based on decision trees · Optimization

1 Introduction

Several authors [1–5] define university dropout as a failure to achieve a desired goal in pursuit of which a subject entered a higher education institution. University dropout creates social and economic impacts in a country. Kulkarni and Haidar [6] details the social consequences of dropout, including the expectations of students and their families and the emotional consequences on students.

This issue is one of the major concerns of the Higher Education system in Colombia, and of the Ministry of Education in particular [7]. For this reason, higher education institutions make efforts to identify the factors that influence student dropout and to study ways to predict dropout to take corrective and timely action [8, 9].

In this context, this study aims to present a classification based on decision trees with optimized parameters to predict the desertion of university students. The following sections about materials and methods detail the source of information and the tool used for analysis. The Results section presents the optimization process and the application of decision tree-based classification. The discussion of the results is provided and, finally, the conclusions of the study are presented.

2 Method

A database of 10,486 students from three private universities in Colombia in the city of Bogota was used as a source of information for the analysis. This data corresponds to four consecutive cohorts of students belonging to 4 undergraduate programs in the areas of Electrical and Electronic Engineering. Table 1 describes the academic desertion of the sample in relation to the years of progress.

The attributes selected for the analysis relate to student demographic variables (age and gender), background of their admission to college (college selection test score and score associated with middle school scores), approximations to their financial situation (family income level and type of middle school), and data on their academic performance (years of progress, average grades and standard deviation of grades). Table 2a, b describe the attributes used in the CBAD analysis [10, 11].

Table 1 Academic dropout of the sample

Dropout	Year 1	Year 2	Year 3	Year 4	Total (%)
No	2415	2310	1895	1224	7844 (75)
Yes	485	790	725	642	2642 (25)
Total	2900	3100	2620	1866	10,486 (100)

Table 2 Attributes for CBAD analysis

Attribute	Type	Median	Stand. dev
a			
Years of advancement	Numerical	2.47	1.12
Age	Numerical	19.85	2.34
Family income level (1–6)	Numerical	1.35	0.75
Selection test score	Numerical	568.9	40.7
Score of middle school grades	Numerical	577.4	83.3
Average grades	Numerical	4.2	0.8
Standard deviation of grades	Numerical	1.0	0.4
b			
Gender	Nominal	N	%
• Female		5767	55
• Male		4719	45
Middle School	Nominal	N	%
• Private		3670	35
• Public		1573	15
Subsidized		5243	50
Dropout	Nominal	N	%
• No		7844	75
• Yes		2642	25
Total		10,486	100.0

The RapidMiner Studio 7.5 tool [12, 13], which implements the C4.5 algorithm for machine learning, was used to perform this CBAD analysis. As noted by [14, 15], based on an expert assessment, the number of citations in the literature, and a survey of the academic community, the C4.5 algorithm was selected as the most influential in Data Mining (DM).

3 Results

This section shows the results of two procedures required for the analysis. First, the findings of the process of finding optimal parameter values for CBAD are delivered, and then the results of applying the CBAD with those parameters in the sample data are detailed.

Table 3 Optimized parameters

Parameter	Rank/steps	List	Result
Attribute selection criteria for division		Gini index accuracy Earnings ratio information gain	Gini index
Maximum depth	1–20/20		16
Confidence level used for the calculation of pruning pessimistic error	0.05–0.5/9		0.15

3.1 Optimization

The optimal values were calculated for the parameters associated with criteria to select the attributes for the division, the maximum depth of the tree, and the confidence level used for the calculation of the pessimistic pruning error. Table 3 shows, in column 1, the parameters analyzed by the optimization process; in column 2, the range (start and term) of the values analyzed for the numeric parameters, and the number of steps used within that range; in column 3, the possibility list for non-numeric parameters; and finally in column 4, the result associated with the optimal value among all 800 possibilities evaluated [16, 17].

3.2 CBAD Results

For predictive analysis purposes, the data was randomly divided into two subsamples. The first one with 70% of the records which served to determine the CBAD, and the second one with the remaining 30% of records to test the predictability of the model.

Based on the confusion matrix presented in Table 4, the prediction accuracy ratio reached 88.14%. Specifically, the CBAD implementation indicates three academic factors that explain the student dropout: the average of grades, the years of progress in the career, and the score on the selection test. Therefore, demographic factors such as gender and age, or economic, such as the level of household income and the type of middle school, do not explain this desertion. Similarly, academic score, the scores of middle teaching, and standard deviation of grades do not affect the prediction of student dropout in the sample [18, 19].

Table 4 Confusion matrix for dropout prediction

		Dropout prediction		Total
		Yes	No	
Actual dropout	Yes	402	232	634
	No	385	2127	2512
Total		787	2359	3146

4 Conclusions

In relation to the results two elements highlight: the prediction accuracy achieved in this study and the attributes found relevant in the prediction. About prediction accuracy, the results of this study outweigh those obtained in other researches that used the C4.5 algorithm through the CBAD technique [20].

The accuracy obtained (88.14%) exceeds studies that have considered medium or large sized samples and were briefly described in the literature review [2, 14, 21]. Associated with this accuracy, it is necessary to indicate that other classification techniques could be a possibility to improve this accuracy. Models of artificial neural networks are an alternative to CBAD to predict university dropout.

In this context, the study of [21], which compares the application of a model of artificial neural networks with the application of a CBAD to estimate the determinants of university student dropout, points out that the CBAD has a right classification higher than the model of artificial neural networks (82 vs 80%), but a 1% lower accuracy (72 vs. 73%). The second element to highlight concerns the attributes that impact the student dropout.

As in the studies of [22, 23], grade average is one of the factors that most explains the dropout phenomenon in tertiary education students. In a similar way, [24] found that the amount and result obtained in the approved subjects greatly impact student dropout. The other important attributes found in this research correspond to the years of career advancement and college admission score. Finally, and like the findings of [25, 26], socioeconomic conditions have only a marginal effect on the explanation of the phenomenon of student dropout.

References

1. Zhang GP (2003) Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50(1):159–175
2. Duan L, Xu L, Liu Y, Lee J (2009) Cluster-based outlier detection. Ann Oper Res 168(1):151–168
3. Haykin S (1999) Neural networks a comprehensive foundation, 2nd edn. Macmillan College Publishing, Inc., New York. ISBN: 9780023527616
4. Haykin S (2009) Neural networks and learning machines. Prentice Hall International, London, NJ
5. Isasi P, Galván I (2004) Redes de neuronas artificiales. Un enfoque Práctico. Pearson, London. ISBN: 8420540250
6. Kulkarni S, Haidar I (2009) Forecasting model for crude oil price using artificial neural networks and commodity future prices. Int J Comput Sci Inf Secur 2(1):81–89
7. Mazón JN, Trujillo J, Serrano M, Piattini M (2005) Designing data warehouses: from business requirement analysis to multidimensional modeling. In: Proceedings of the 1st international workshop on requirements engineering for business need and IT alignment. Paris, France
8. Izquierdo NV, Lezama OBP, Dorta RG, Viloria A, Deras I, Hernández-Fernández L (2018) Fuzzy logic applied to the performance evaluation. Honduran coffee sector case. In: Tan Y, Shi Y, Tang Q (eds) Advances in swarm intelligence. ICSI 2018. Lecture notes in computer science, vol 10942. Springer, Berlin

9. Pineda Lezama O, Gómez Dorta R (2017) Techniques of multivariate statistical analysis: an application for the Honduran banking sector. Innovare: J Sci Technol 5(2):61–75
10. Viloria A, Lis-Gutierrez JP, Gaitán-Angulo M, Godoy ARM, Moreno GC, Kamatkar SJ (2018) Methodology for the design of a student pattern recognition tool to facilitate the teaching—learning process through knowledge data discovery (big data). In: Tan Y, Shi Y, Tang Q (eds) Data mining and big data. DMBD 2018. Lecture notes in computer science, vol 10943. Springer, Berlin
11. Ben Salem S, Naouali S, Chtourou Z (2018) A fast and effective partitional clustering algorithm for large categorical datasets using a k-means based approach. Comput Electronic Eng 68:463–483. https://doi.org/10.1016/j.compeleceng.2018.04.023
12. Chakraborty S, Das S (2018) Simultaneous variable weighting and determining the number of clusters—a weighted gaussian algorithm means. Stat Probab Lett 137:148–156. https://doi.org/10.1016/j.spl.2018.01.015
13. Abhay KA, Badal NA (2015) Novel approach for intelligent distribution of data warehouses. Egypt Inform J 17(1):147–159
14. Aguado-López E, Rogel-Salazar R, Becerril-García A, Baca-Zapata G (2009) Presencia de universidades en la Red: La brecha digital entre Estados Unidos y el resto del mundo. Revista de Universidad y Sociedad del Conocimiento 6(1):1–17
15. Bontempi G, Ben Taieb S, Borgne YA (2013) Machine learning strategies for time series forecasting. In: Aufaure M-A, Zimányi E (eds) Lecture notes in business information processing, vol 138, no 1. Springer, Heidelberg, pp 70–73
16. Parthasarathy S et al (2001) Parallel data mining for association rules on shared-memory systems. Knowl Inf Syst 3(1):1–29
17. Grossman RL, Bailey SM, Sivakumar H, Turinsky AL (1999) Papyrus: a system for data mining over local and wide area clusters and super-clusters. In: Proceedings of ACM/IEEE conference on supercomputing, Article No 63
18. Chattratichat J, Darlington J, Guo Y, Hedvall S, Kohler M, Syed J (1999) An architecture for distributed enterprise data mining. In: Proceedings of 7th international conference on high performance computing and networking, Netherlands, 12–14 Apr, pp 573–582
19. Wang L et al (2013) G-hadoop: MapReduce across distributed data centers for data-intensive computing. Futur Gener Comput Syst 29(3):739–750
20. Butenhof DR (1997) Programming with POSIX threads. Addison-Wesley, Boston
21. Bhaduri K, Wolf R, Giannella C, Kargupta H (2008) Distributed decision-tree induction in peer-to-peer systems. Stat Anal Data Min 1(2):85–103
22. Rafailidis D, Kefalas P, Manolopoulos Y (2017) Preference dynamics with multimodal user-item interactions in social media recommendation. Expert Syst Appl 74:11–18
23. Vásquez C, Torres M, Viloria A (2017) Public policies in science and technology in Latin American countries with universities in the top 100 of web ranking. J Eng Appl Sci 12(11):2963–2965
24. Aguado-López E, Rogel-Salazar R, Becerril-García A, Baca-Zapata G (2009) Presencia de universidades en la Red: La brecha digital entre Estados Unidos y el resto del mundo. Revista de Universidad y Sociedad del Conocimento 6(1):1–17
25. Torres-Samuel M, Vásquez C, Viloria A, Lis-Gutiérrez JP, Borrero TC, Varela N (2018) Web visibility profiles of top 100 Latin American universities. In: Tan Y, Shi Y, Tang Q (eds) Data mining and big data. DMBD 2018. Lecture notes in computer science, vol 10943. Springer, Berlin
26. Caicedo EJC, Guerrero S, López D (2016) Propuesta para la construcción de un índice socioeconómico para los estudiantes que presentan las pruebas Saber Pro. Comunicaciones en Estadística 9(1):93–106 (85–97 English)

Prediction of Psychosocial Risks in Teachers Using Data Mining

Amelec Viloria, Jorge Rodríguez López, Nataly Orellano Llinás, Carlos Vargas Mercado, Luz Estela León Coronado, Ana María Negrete Sepulveda and Omar Bonerge Pineda Lezama

Abstract Integrated management systems aim to improve these everyday situations that are inherent to work and cause for concern. In search for continuous improvement, it is necessary to innovate with techniques in areas that are not yet explored and that contribute to strategic decision-making processes, such as machine learning techniques or machine learning. In occupational safety and health management systems, it is important to carry out the proper follow-ups and process controls in any type of industry and organization whose relationship is direct. This paper presents the application of three methods related to data mining: Support Vector Machine algorithms, Naïve Bayes, and Genetic Algorithms to identify the degree of psychosocial risk in university teachers of the Mumbai University in India. The use of SVM easily recognizes physiological variables and the best prediction performance was achieved with 96.34% accuracy efficiency.

A. Viloria (✉)
Universidad de la Costa, St. 58 #66, Barranquilla, Atlántico, Colombia
e-mail: aviloria7@cuc.edu.co

J. R. López
Universidad Simón Bolívar, Barranquilla, Colombia
e-mail: jrodriguez2@unisimonbolivar.edu.co

N. O. Llinás · L. E. L. Coronado
Corporación Universitaria Minuto de Dios—UNIMINUTO, Barranquilla, Colombia
e-mail: nataly.orellano@uniminuto.edu

L. E. L. Coronado
e-mail: luzleoncoronado@gmail.com

C. V. Mercado
Corporación Universitaria Latinoamericana, Barranquilla, Colombia
e-mail: cvargas@ul.edu.co

A. M. N. Sepulveda
Universidad Cooperativa de Colombia Montería, Cordoba, Colombia
e-mail: ana.negrette@campusucc.edu.co

O. B. P. Lezama
Universidad Tecnológica Centroamericana (UNITEC), San Pedro Sula, Honduras
e-mail: omarpineda@unitec.edu

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_50

Keywords Support vector machine · Naïve bayes · Genetic algorithms

1 Introduction

As described by the International Labor Organization (ILO), work-related stress-related diseases are developed due to the lack of worker capacity to meet the conditions and demands that arise from the requirements of the organization of work, of relations at work, of environmental conditions that become risk factors because of the poor adequacy provided by organizations [1, 2].

The application of machine learning techniques such as Support Vector Machine algorithms (SVM), and the Naïve Bayes algorithm are used as a tool for predicting psychosocial conditions in order to evaluate their efficiency in the degree of accuracy with respect to identifying the type of risk degree and improving human resource management and management processes for risk prevention [3–6]. Genetic algorithms are used for the reduction of dimensionality, which have already been tested as artificial intelligence techniques in other processes to optimize different objectives.

The development and validation of new instruments from the engineering branch that contributes to the detection of psychosocial risks in organizations of all kinds is possible thanks to the patterns that facilitate the classification of the variables (attributes) that denote a result (class), currently, of new algorithms are being developed to facilitate the work for the administration and management of occupational safety and health in all kinds of activities [7, 8].

This paper presents the steps for applying the Naïve Bayes and Support Vector Machine (SVM) algorithms to the database for assessing the intra-labor psychosocial risk factors in Teachers at Mumbai University in India, as well as the physiological variables (electroencephalography, electrocardiogram, heart rate, electrical skin activity, and electromyography).

2 Method

For the development of the study, the paper provides information concerning the database to be intervened, the variables that comprise it, the description of the classification models, Support Vector Machines (SVMs), and the method of reducing the dimensionality of the data concerning this study: genetic algorithm, leaving the forty attributes of greater impact for the correct classification and with lower error rate as variables to intervene [9, 10]. The analyzed dataset was obtained from teacher's data, sociodemographic, physical, psychosocial, environmental, and physiological variable information, as shown in Table 1. The fields are checked with the correct information in the database with 10,452 records.

Table 1 Variables of intra-labor psychosocial risk factor domains

Items	Domains	Variables
S1	Demographic	S1… S10
D1	Demands from work	D1… D50
W1	Control over work	W1… W21
L1	Leadership and social relations in the workplace (workplace)	L1… L32
R1	Rewards	R1… R11
P1	V. Physiological	P1… P5

3 Results

Firstly, the Datasets were organized, secondly, the algorithms were applied, thirdly, the dimensionality reduction is carried out, the algorithms were then applied, and finally the stability analysis is performed [11, 12].

3.1 Step 1. Datasets and Information Checking

129 variables were considered, with quantitative and qualitative values and 10,452 records. The evaluation application was made directly to each of the people, containing 129 predictor variables and 1 class variable.

3.2 Step 2. Applying Machine Learning Techniques

After applying the Naïve Bayes algorithms and Support Vector Machines to the training dataset as a first stage, the results shown in Table 2 were obtained, presenting the correct instances generated for each algorithm [13, 14].

Where TP (True Positives) = Positive correctly classified cases; TN (True Negatives) = Properly classified negative cases; FP (False Positives) = Incorrectly classified negative cases; FN (False Negatives) = Incorrectly classified positive cases; and Sensitivity (True Positive Fraction) is the probability that a diagnostic test is positive, since the person presents the condition. Specificity (true negative fraction) is

Table 2 Comparison between classification algorithms with initial data

Technical	TP	TN	FP	FN	Sensitivity ± SD	Specificity	Accuracy ± SD
Naïve Bayes	32 ± 1	16	0	2 ± 1	91.5 ± 2.6	100	90.58 ± 1.83
SVM	30 ± 1	16	0	2 ± 1	92.3 ± 2.5	100	94. 3 ± 1.90

the probability that a diagnostic test is negative, because the person does not present the disease. Accuracy is the probability that a diagnostic test will succeed [15, 16].

This comparison shows that the SVM system offers high classification accuracy and less computing due to feature extraction. In fact, experimental results show a lower error rate and the sensitivity rate ranges from 89.5 to 94.8% with an average ± Standard Deviation of 91.5 ± 2.6%.

3.3 Step 3. Reducing Dimensionality

The dimensionality of the data was reduced using genetic algorithm. From the method, the common variables were extracted to form a new organization of variables (42). In the Genetic Algorithm, population subsets of 20 individuals with 42 features were selected, from which, the most relevant were contained in the parents, according to the information extracted from the database [17].

3.4 Step 4. Evaluating the New Dataset with Machine Learning Techniques

60 replicates were performed with each algorithm and the top 10 results were selected to be exposed. The results obtained are set out in Table 3. The obtained values were used to perform the stability analysis using the variance analysis with a 99% confidence interval. The evaluation of the dataset was again carried out with the two algorithms, Naïve Bayes and SVM, to obtain an average accuracy of 94.36% with a standard deviation of 1.85 and 96.34% respectively, with a standard deviation of 1.85 in the grade prediction of psychosocial risk, as shown in Table 4.

3.5 Step 5. Stability Analysis

Based on the solutions found in step 4 (Table 4), the variance analysis was performed with a 99.5% confidence interval. Since in Table 5, the tabulated F is greater than the calculated F, it can be expressed that there are no significant differences in the results provided by the system; that is, under a 99.5% confidence interval the data is statistically equal.

The model demonstrates that the use of Support Vector Machines in psychosocial risk classification has a great future for the implementation of already tested models and for the development of a new method based on Support Vector Machines that is adaptable to the data model generated in the identification of psychosocial risk in

Table 3 Replicates with each algorithm and its overall average

Algorithms	Replicates										Overall average
SVM	96.5	96.53	97.1	97.46	93.99	95.99	96.24	96.87	96.55	96.36	96.34
Naïve Bayes	90.2	90.33	94.01	93.55	93.47	94.83	96.78	96.47	96.99	96.99	94.36

Table 4 Comparison between classification algorithms

Technical	TP	TN	FP	FN	Sensitivity ± SD	Specificity	Accuracy SD
SVM	36 ± 1	18	0	4 ± 1	94.57 ± 2.67	100	96.34 ± 1.85
Naïve Bayes	35 ± 1	18	0	3 ± 1	91.63 ± 2.68	100	94.36 ± 1.85

Table 5 Analysis of variance (anova)

Source variation	SS	DF	MS	F	Prob > F
Columns	248.57	2	114.92	108.26	−1.87E–17
Error	31.62	28	1.173		
Total	271,077	30			

teaching population at Mumbai University, in India. The use of physiological variables in the study and their use with algorithms such as Support Vector Machines allows to generalize and infer that they easily adapt to the treatment of the mathematical model treated by SVM. It is recommended to continue to use of such algorithms with such physiological variables for implementation in SVM.

It is possible, for future research, to further implement Support Vector Machines (SVM) including new variables other than the physiological ones already demonstrated here, such as recognition of visual patterns or image patterns, developing a new method based on physiological variables along with organizational variables, and variables associated with *lean* techniques with machine learning techniques to automate the process of recognition of the presence of psychosocial risk in the organizational culture of universities and companies.

Optimizing the search space using mathematical optimization methods such as optimization algorithms belonging to the kind of local search techniques (hill climbing) and global search (simulated annealing) in addition to SVM and NBC techniques, can improve the development and application of classification techniques for psychosocial risk identification in different organizational models. As can be observed, disciplines related to machine learning techniques, they include psychological study of human learning, the study of evolution, the theory of adaptive control, the study of educational practices, neuroscience, organizational behavior and economics. Although during the last decade has seen greater growth with these other fields, researchers are beginning to take advantage of the potential synergies and diversity of formalisms and experimental methods used in these multiple fields to study systems that improve experience [1, 9, 13, 18] such as psychosocial and organizational methods.

4 Conclusions

Based on the results and their discussion it is concluded that: (1) Support Vector Machine algorithms can be used to develop better models that allow greater accuracy for the prediction of psychosocial risk in this sector. (2) The use of Naïve Bayes models could also be fruitful in predicting this type of risk with fewer variables included in the studies. (3) Experimentation with epidemiological data of psychosocial risk and physiological variables will allow the development of technological tools that facilitate the improvement of working conditions through prediction. (4) The implementation of little-used heuristics (Hill Climbing, Grasp) optimizes the search space achieving the best possible local solution, which can be improved by iterated local searches to favor ranking models, such as vector support machines. (5) The creation of a research line related to the implementation of SVM-like algorithms will improve the development of engineering and the application of soft computing techniques at the organizational, psychosocial, and ergonomic levels of companies in the country. (6) The implementation of machine learning in organizational decision-making processes will promote the improvement of occupational safety and health conditions in different economic sectors of the country. (7) The best prediction performance was obtained with the SVM method with an accuracy efficiency of 96.34%.

References

1. Viloria A, Bucci N, Luna M (2017) Historical development of psychosocial risk assessment models. J Eng Appl Sci 12(11):2915–2919. ISSN: 1816-949X. Medwell Journals. https://www.medwelljournals.com/journalhome.php?jid=1816-949x
2. Viloria A, Bucci N, Luna M (2017) Comparative analysis between psychosocial risk assessment models. J Eng Appl Sci 12(11):2901–2903. ISSN: 1816-949X. Medwell Journals. https://www.medwelljournals.com/journalhome.php?jid=1816-949x
3. Instituto Sindical de Trabajo, Ambiente y Salud (ISTAS). http://www.istas.net/web/index.asp?idpagina=1235
4. Organización Mundial de la Salud (2007) Commission on social determinants of health. A conceptual framework for action on the social determinants of health. (Discussion paper. Geneve: Retrieved from http://www.who.int/social_determinants/resources/csdh_framework_action_05_07.pdf
5. Moncada S, Llorens C, Navarro A, Kristensen T (2005) Versión en lengua castellana del cuestionario psicosocial de Copenhague (COPSOQ). La Societat Catalana de Seguretat i Medicina del Treball. España. http://www.scsmt.cat/pdf/8n1orig1.pdf
6. Serra J (2011) Pautas para la intervención Psicosocial en las organizaciones. Taller para gestionar el estrés y otros riesgos psicosociales. Reto laboral del siglo XXI. http://www.asepeyo.es/apr/apr0301.nsf/ficheros/PSI0904001%20Intervenciones%20medio%20laboral%20riesgo%20psicosocial.pdf/$file/PSI0904001%20Intervenciones%20medio%20laboral%20riesgo%20psicosocial.pdf
7. Viloria A, Bucci N, Luna M (2017) Comparative analysis between psychosocial risk assessment models. J Eng Appl Sci 12(11):2901–2903. ISSN: 1816-949X. Medwell Journals
8. Caamaño AJ, Echeverría MM, Retamal VO, Navarro CT, Espinosa FT (2015) Modelo predictivo de fuga de clientes utilizando minería de datos para una empresa de telecomunicaciones en chile. Universidad Ciencia y Tecnología, 18(72)

9. Mark Hall y otros 5 autores (2009) The WEKA data mining software: an update; SIGKDD explorations 11(1)
10. Bucci N, Luna M (2012) Contrastación entre los Modelos de Estudio del Estrés como Soporte para la Evaluación de los Riesgos Psicosociales en el Trabajo. Revista Digital de Investigación y Postgrado de la Universidad Nacional Experimental Politécnica "Antonio José de Sucre", Vicerrectorado Barquisimeto. Venezuela 2(1):21–38 abril 2012. ISSN: 2244-7393. http://redip.bqto.unexpo.edu.ve
11. Agarwal B, Mittal N (2014) Text classification using machine learning methods—a survey. In: Proceedings of the second international conference on soft computing for problem solving (SocProS 2012), Springer, New Delhi, 28–30 Dec 2012, pp 701–709
12. Larrañaga P, Inza I, Moujahid A (2016) Tema 6. Clasificadores Bayesianos. Departamento de Ciencias de la Computación e Inteligencia Artificial. En línea: http://www.sc.ehu.es/ccwbayes/docencia/mmcc/docs/t6bayesianos.pdf Acceso: 9 de enero de 2016, Universidad del País Vasco-Euskal Herriko Unibertsitatea, España
13. Quinlan JR (1993) C4.5: programs for machine learning. Elsevier, Burlington
14. García DA (2007) Algoritmo de discretización de series de tiempo basado en entropía y su aplicación en datos colposcopicos. Tesis de Maestría en Inteligencia Artificial. Universidad Veracruzana, México
15. Corso CL (2009) Alternativa de herramienta libre para la implementación de aprendizaje automático. En línea: http://www.investigacion.frc.utn.edu.ar/labsis/Publicaciones/congresos_labsis/cynthia/Alternativa_de_herramienta_para_Mineria_Datos_CNEISI_2009.pdf. Acceso: 10 de agosto de 2015, Argentina
16. Anon D (2016) Búsqueda exhaustiva. En línea: http://dis.um.es/~domingo/apuntes/AlgBio/1213/exhaustiva.pdf. Acceso: 2 de agosto 2015, Universidad de Murcia, España
17. Hepner GF (1990) Artificial neural network classification using a minimal training set. Comparison to conventional supervised classification. Photogramm Eng Remote Sens 56(4):469–473
18. Bucci N, Luna M, Viloria A, García JH, Parody A, Varela N, López LAB (2018) Factor analysis of the psychosocial risk assessment instrument. In: Tan Y, Shi Y, Tang Q (eds) Data mining and big data. DMBD 2018. Lecture notes in computer science, vol 10943. Springer, Berlin

A Method to Rationalize the Product Portfolio in Retail Stores

Jairo R. Coronado-Hernandez, Holman Ospina-Mateus, Javier Baron-Villamizar, Lucelys Vidal-Pacheco, Aurora Piñeres-Castillo, Rafael H. Rojas-Millán and Jesús Silva

Abstract Retail store operations face a variety of challenges and complexities. Determining the best assortment is the main problem in the retail store. This re-search presents a practical methodology for the analysis of products in the assortment with the goal of reducing the excess items and improve sales and profit margin of a retail store without affecting customer satisfaction. The methodology integrates 6 steps that allow to optimize products of a portfolio in categories, sub categories and segments, through Pareto analysis and clustering analysis using the BCG matrix. The methodology was applied in an independent supermarket. The results in the case of the application for non-perishable products, allowed to identify a set of different products (n = 152), of which they were prioritized in a subcategory (oils) in which 90 products were prioritized. In the example, it shows how 21 products have significant results in the variety of products. The combination of the global and local category of the product, the net profit, the inventory rotation and the participation of the growth

J. R. Coronado-Hernandez (✉) · L. Vidal-Pacheco · A. Piñeres-Castillo · R. H. Rojas-Millán
Universidad de la Costa, Barranquilla, Colombia
e-mail: jcoronad18@cuc.edu.co

L. Vidal-Pacheco
e-mail: lvidal2@cuc.edu.co

A. Piñeres-Castillo
e-mail: apineres2@cuc.edu.co

R. H. Rojas-Millán
e-mail: rrojas@cuc.edu.co

H. Ospina-Mateus
Universidad Tecnológica de Bolívar, Cartagena, Colombia
e-mail: hospina@unitecnologica.edu.co

J. Baron-Villamizar
Escuela Naval de Cadetes "Almirante Padilla", Cartagena, Colombia
e-mail: dfam@enap.edu.co

J. Silva
Universidad Peruana de Ciencias Aplicadas, Lima, Peru
e-mail: jesussilvaUPC@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_51

provides a multifactorial analysis in the decision-making to supply with products a retail store seeking to increase the level of service and maximizing profits.

Keywords Rationalization · Portfolio optimization · Clustering analysis · Retail

1 Introduction

The retail trade comprises a large part of the economy that directly represents about 6% of the GDP of the United States in 2016. Retail store operations face a variety of challenges and complexities. Determining the best assortment is the main problem in the retail store. The optimal store assortment contributes directly to the margins, sales and efficiency in customer service. The assortment is a key factor for the customer to choose the desired products and the store can promote their products at the time of purchase. Daily retailers seek assortment strategies [1]. One of the strategies to improve the assortment is reducing the number of items on the shelf to address the interest of customers about a product [2]. Previous research has shown that reducing items have effects on sales, assortment perception, satisfaction in product selection, loyalty to supermarket and store image [3]. In a study of customers in the Consumer Trends Report 2005 [4], is show a list of six fundamental factors that significantly irritate frequent visitors to supermarkets, like this: (1) long wait in line, (2) products not available because of the reduction of assortment, (3) replacement products during service time in the store, (4) regular items out of stock, (5) no chance to pack products after purchase or (6) promotional items out of stock.

Among the factors evidenced many are related to the assortment, at least from the consumer's perspective, should have been available. Therefore, the assortment has a direct relationship with customer satisfaction in a supermarket. Traditional research shows that when a product is out of stock or unavailable, it has a direct negative impact on the sales and profitability of the retail category, and indirectly on customer satisfaction, store loyalty and company image [5–8].

2 Method

The proposed method is based on ABC analysis, GMROI and clustering using the Boston Consulting Group (BCG) growth-investment matrix. The ABC analysis or the Pareto's law seeks to measure the irregular distribution of demand or the way in which the total sales of each product are distributed within the portfolio [9]. Many companies that manufacture or distribute a range of products found in Pareto's law a practical application that can be exploited to determine supply strategy [10, 11]. The GMROI is a measure of inventory productivity that is widely used in retail stores [12, 13]. GMROI provides a measure of the speed with which inventory is rotating and the return on the investment. BCG analysis is mainly used for companies that handle

different products in the business portfolio where each product can move at a different pace and with different strategies. The BCG Matrix makes it possible to represent the business portfolio, locating all the products (SUB) and its specific contribution to the growth and profitability of the company, classifying the products into four categories: Star when referring to SUB with great growth and great participation, Unknowns for SUB that have great growth and little participation, the Dows with low growth and high participation, and finally the Dogs in those SUB that do not have growth and the participation of is low [14–20].
In general, the method follows a series of steps that are shown below:

Step 0: "Pre-processing the data". This step consists of preparing and debugging the information system of the retailer, which allows reliable business data.
Step 1: "Global ranking using an ABC analysis of net sales": An ABC analysis of net sales of all products is performed on a global basis. Sales are taken for the ABC analysis because they show how each product contributes to the total sales of the entire company and the relative share of total net sales. This filter, each product is categorized according to their overall participation.
Step 2: "Classification by Category through an ABC analysis of net sales": Like step 1, a second ABC analysis is applied to net sales but for all items in each Group. The purpose is to study the share of each item in the Group's net sales. In this filter the item will receive a second rating. For example, A double C for an article in the Global and by Groups, is an indicative for a more detailed analysis of a product.
Step 3: "Sort by Category using an ABC analysis of Contribution Margin": A third filter is applied through an ABC analysis in the margin of each item by category. In this analysis, the contribution of each element to the profit generated by the category is studied.
Step 4: "Classification by Subcategory through an ABC analysis of the GMROI": A fourth filter is applied through an ABC analysis in the GMROI of each item by subcategory. In this analysis, the contribution of each item to how the investment in the inventory is generating profit by subcategory is studied. GMROI is calculated as shown in formula (1):

$$GMROI = \frac{Gross\ Margin}{Cost\ of\ Average\ Inventory} \tag{1}$$

Step 5: "Grouping to determine individual performance with the BCG matrix": In this step, the matrix of the Boston Consulting Group is used by means of Growth Participation charts to abstractly represent the portfolio of products of each segment to identify the specific contribution that each element makes in terms of growth and profitability. The growth rate is calculated by means of net sales and the relative share of sales in the article that generates the greatest profit for the retail store. Then, the classification of the items according to the matrix is applied as a criterion.
Step 6: "Design and analysis of the report": Finally, the report is presented to be analyzed in a session with the Category Manager. Basically, the structure of the report is done by category. In each category type there are five columns. The first

column refers to the product code, the second to the general sales ranking, the third to the sales ranking by category, the fourth to GMROI ranking in the segment and finally the classification of the growth-share matrix. From the report it is transferred to the debugging step of the portfolio when the product from the system is removed from the store assortment.

3 Results

The methodology was applied in an independent supermarket extracting from the ERP system (SAP) the data corresponding to sales, inventory levels, costs and sales prices during each day of the year 2017. In this retail store, the products can be classified into two large groups: Perishable and non-perishable. Within each group there are categories, which in turn are subdivided into subcategories, and these, in turn, into segments. For this research will apply the methodology in the non-perishable product group to reach segments of the article. As there is a classification by nature, there is a classification within each group by categories shown in Fig. 1.

Step 1: "Global ranking using an ABC analysis of net sales": In the global analysis, 21142 product references were identified. The articles categorized in A are 9.74% of all products generated 80% of sales. Type B items are 21.58% of articles generated 15% of net sales. The type C items are 68.68% of the articles generated only 5% of net sales for the year 2017. Table 1 shows the ABC classification of products in the portfolio according to their share in net sales. According to this analysis, each product is classified as A, B or C.

Fig. 1 Categories within the non-perishable products

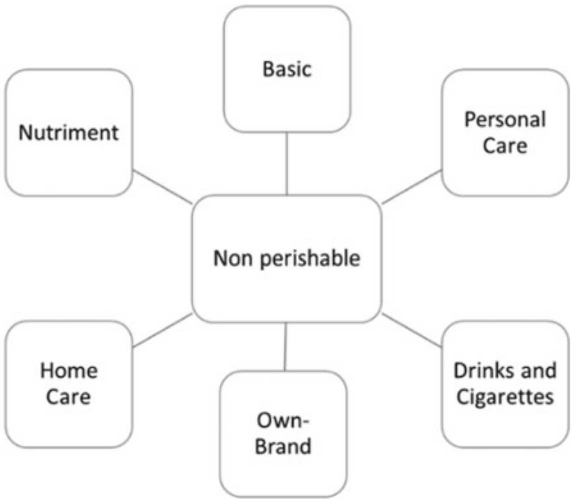


Table 1 Global ABC classification of products by net sales

Type	No of products	% ABC	% Contribution to the sales
A	2060	9.74	80
B	4562	21.58	15
C	14520	68.68	5
Total	21142	100	100

Step 2: "Classification by Groups through an ABC analysis of net sales": This step corresponds to the proportion of each product in the group it belongs to according to the universe of the product. To demonstrate the method, the calculation was performed only in the group of non-perishable products with a total of 11628 products. In this group is found that type A products that are approximately 30% in the ABC mix and generate 80% of the net sales of the non-perishable group, type B products with approximately 20% generating 15% of sales and finally type C products with 51% generating only 5% of sales.

Step 3: "Classification by Category through an ABC Analysis of Contribution Margin": in his step, the basic category was selected within non-perishables group, which represents a 21% of the sales of a company. To show the method a total of 970 product. The method is conducted upon the Contribution Margin of each product reference. Using previous filters, the results are used to show different behaviors: Type A represents approximately a 12% of products, which generate a 80% of the Contribution Margin in the Basic category; Type B can only contribute with a 22% of the company's utility; and Type C represents a 5% of the company's utility with a 66% of the products.

Step 5: "Grouping to determine individual performance with the BCG matrix": continuing the methodology within the category of oils, selected segment superfine oils to illustrate the application of the growth-share matrix. When applying the method, it is found that the products defined as "Dogs" correspond to 38% and do not have much margin for the segment, as seen in Fig. 2.

Step 6: "Design and analysis of the report": This part presents the final report for analysis by segment, subcategory, category, group, and overall participation resulting from the use of the methodology. It can be observed that some products with C-C-C-C-DOG classification are candidates to withdraw from the portfolio, since they do not have an important participation in global sales or in their group. Likewise, the products within their category are not generating a good margin of contribution, nor do they have a good internal rate of return on the money invested in the inventory, nor a good margin and participation in comparison with other products in their segment.

The results in the case of the application for non-perishable products, allowed to identify a set of different products (n = 152), of which they were prioritized in a subcategory (oils) in which 90 products were prioritized. In the example, it shows how 21 products have significant results in the variety of products. The combination of the global and local category of the product, the net profit, the rotation of inventories and the participation of growth provides a multifactorial analysis in the decision

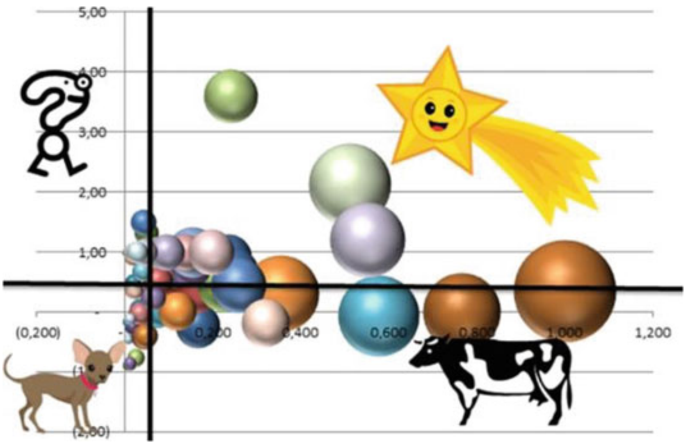


Fig. 2 Application of the BCG matrix to the extra-fine oils segment in the subcategory of basic category oils in the non-perishable group

making. This methodology allows the participation of new products within the market and respects the trajectory in the market of the products. Additional considers the consumption rate and profit-sharing business. In this way, the need to increase the level of service and maximize business profits is covered.

4 Conclusions

The analysis and rationalization of the product portfolio in retail stores is a very important task in retail store. Through this type of analysis, retail stores can easily determine the products that generate the most profits, sales and those that need priority, in terms of supply, since they are strategic for the market and the business. The proposed methodology allows solutions in decision-making in the short term. Decisions in the assortment of retail stores and supermarkets are a priority because of their direct impact on the level of service, inventories, internal distribution in the store, profitability and replenishment times. The methodology application requires careful analysis but not complex ensuring a holistic view of the problem. In the management of operations, practical tools are required to streamline the analysis of information as allowed by the proposed methodology. In the diversification of products, manufacturing companies seek to find new markets. This diversification becomes a challenge for distributors and retailers, considering that the volume should expand their inventories and balance the total cost of the goods. This diversification could lead to shortages of some products or otherwise over-inventories others.

References

1. Campo K, Gijsbrechts E, Nisol P (2004) Dynamics in consumer response to product unavailability: do stock-out reactions signal response to permanent assortment reductions? J Bus Res 57(8):834–843
2. Sloot LM, Fok D, Verhoef PC (2006) The short-and long-term impact of an assortment reduction on category sales. J Mark Res 43(4):536–548
3. Beneke J, Cumming A, Jolly L (2013) The effect of item reduction on assortment satisfaction—a consideration of the category of red wine in a controlled retail setting. J Retail Consum Serv 20(3):282–291
4. EFMI and CBL (2005) ConsumentenTrends 2005. Rotterdam and Leidschendam: EFMI and CBL
5. Zinn W, Liu PC (2001) Consumer response to retail stockouts. J Bus Logist 22(1):49–71
6. Valero-Herrero M, García-Sabater JP, Hernández JRC, Maheut J (2011) Planteamiento dinámico del problema de secuenciación en líneas de montaje con mezcla de modelos. In: V international conference on industrial engineering and industrial management, pp 288–296
7. Coronado-Hernández JR (2016) Análisis del efecto de algunos factores de complejidad e incertidumbre en el rendimiento de las Cadenas de Suministro. Propuesta de una herramienta de valoración basada en simulación. Universitat Politècnica de València, Valencia (Spain). https://doi.org/10.4995/Thesis/10251/61467
8. Garvis M (2010) SKU rationalization: determining the proper depth & breadth of an assortment. http://quantumretail.com/2010/08/17/the-profit-lab-new-seriesassortment-and-range-planning/
9. Downs B, Metters R, Semple J (2001) Managing inventory with multiple products, lags in delivery, resource constraints, and lost sales: A mathematical programming approach. Manage Sci 47(3):464–479
10. Irion J, Lu J-C, Al-Khayyal F, Tsao Y-C (2012) A piecewise linearization framework for retail shelf space management models. Eur J Oper Res 222(1):122–136
11. Van Herpen E, Pieters R (2002) The variety of an assortment: an extension to the attribute-based approach. Mark Sci 21(3):331–341
12. Gruen TW, Corsten DS, Bharadwaj S (2002) Retail out-of-stocks: a worldwide examination of extent, causes and consumer responses. Grocery Manufacturers of America, Washington, DC
13. Cachon GP, Terwiesch C, Xu Y (2005) Retail assortment planning in the presence of consumer search. Manuf Serv Oper Manag 7(4):330–346
14. Kök AG, Fisher ML (2007) Demand estimation and assortment optimization under substitution: methodology and application. Oper Res 55(6):1001–1021
15. Anderson SP, De Palma A, Thisse JF (1992) Discrete choice theory of product differentiation. MIT press, Cambridge
16. Christopher M, Towill D (2001) An integrated model for the design of agile supply chains. Int J Phys Distrib Logist Manag 31(4):235–246
17. Rius-Sorolla G, Maheut J, Coronado-Hernandez JR, Garcia-Sabater JP (2020) Lagrangian relaxation of the generic materials and operations planning model. Cent Eur J Oper Res 28(1):105–123
18. Mattila H, King R, Ojala N (2002) Retail performance measures for seasonal fashion. J Fash Mark Manag
19. Amelec V (2015) Validation of strategies to reduce exhausted shelf products in a pharmaceutical chain. Adv Sci Lett 21(5):1403–1405
20. Amelec V (2015) Methodology to increase the adaptability and flexibility of the supply chain of automotive company through lean manufacturing. Adv Sci Lett 21(5):1517–1520

Need of iSchools in Developing Countries

P. K. Paul, Vijender Kumar Solanki and Vinit Kumar Gunjan

Abstract Information is one of the valuable and important facets these days. It is responsible for the development of the society many ways. Information has created the Digital Revolution and that has changes the science of mechanics, and whole electronic technology. It is important to note that the digital electronics and its importance were started in 1950s and later grown in the 1970s. Moreover the development and proliferation of digital computers changes the entire information systems and process. Gradually digital record keeping and information management lead the organizations, institutions etc. for their information and technological solutions. The integration of Computing with the Management, Electronics, Information Studies lead the development of many information related domains internationally in many universities under various departments and faculties. That continues to the present day with more other interdisciplinary subjects. Information treated as power these days and also for the development. Information is for everything. Today most of the organizations, institutions are moving towards healthy and better knowledge and technology management; and in all cases Information Sciences and Technologies play a lead role. Internationally many academic departments, which are deals with the information and computing subjects moving towards a new academic innovation for solid and real development. That academic model is called *iSchools* or Information Schools. There are many opportunities of *iSchools* in business, healthcare, management and corporations, education etc. and there are huge possibilities of offering such subjects in the universities around the world especially in the developing countries. This paper discusses about the development projects of the *iSchools* and its potentialities in the developing countries in strategic and educational and cognitive context.

P. K. Paul (✉)
Department of Computer & Information Science, Raiganj University, Raiganj, West Bengal, India
e-mail: prantoshkpaul@gmail.com

V. K. Solanki · V. K. Gunjan
Department of Computer Science and Engineering, CMR Institute of Technology, Hyderabad, India
e-mail: spesinfo@yahoo.com

V. K. Gunjan
e-mail: vinitkumargunjan@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_52

Keywords Computing · Information science · Information systems · *iSchools* · Information schools · Information and communication technology for development (ICT4D) · Interdisciplinary development · Academic models · Universities · Emerging courses · Cognition

1 Introduction

iSchools is a kind of interdisciplinary school of university level responsible for advancement of information and technologies in different context apart from the technical context. *iSchools* may also refer as a School of Information, Department of Information Sciences, or "Information Science and Technology Department". *iSchools* academician normally engage in the research activities of the basics of information as well as related technologies in several contexts. In *iSchools* academic degrees, research degrees, training programs etc. Information Schools basically educates IT Professionals, Computing and Communications, Information Professionals, Researchers, as well as scholars who are interested to build a technology driven information enriched world. [1, 2] In restricted manner *iSchools* may also deemed as a member of *iSchools* Organization (earlier, *iSchools* Project or *iCaucus*) and comprised from the schools, colleges, as well as departments which are formerly newly established as *iSchools* or coming with the programs focused on Information Systems, Information Technology, Library Science, Informatics, Computer Science, Library and Information Science and other Information and Computing related fields [3–5].

With strong interdisciplinary approaches *iSchools* believes in opportunities and challenges of information management and information technology solutions, with more focus on universal access and user-centered information systems creation and design [6–8]. *iSchools* are purely committed to investment in design-preservation of information spaces; from digital to manual. Creation of the virtual spaces such as online communities, WWW based systems, network systems, cloud systems and databases etc. and moreover several physical spaces such as libraries, museums, archives including some other repositories. *iSchools* and its area now also moving towards designing and development of eco friendly and green information systems creation in all the settings and establishments.

Development of a nation is depends on several criteria which includes the industrialized country, with advanced technological infrastructure [9–11]. With good amount of the economic development which includes Gross Domestic Product, Gross National Product, the Per capita Income etc. Moreover development is also depends on the level of industrialization, widespread infrastructure, general standard of living—And for all these, Information and Technology both played a vital role. In conventional educational systems Information related and Technology related subjects are taught and practiced in the different departments while in the *iSchools,* all the related programs of IT, Information and Computing combined in one circle. Hence proper information and technological development are possible in many contexts with the establishment of *iSchools* in the developing countries.

2 Agenda and Objectives

As a theoretical, conceptual and policy based research work the present study has been mapped with the aim and objectives of as follows:

- To learn the *iSchool or Information School* formation, establishment and evolution.
- To find out the core objective of *an iSchools or whole iSchools* foundation and *iSchools* organization.
- To dig out the basic and emerging features as well as functions of an *iSchools* or the consortium of *iSchools* for the development.
- To get a good and sophisticated list of *iSchool listed by the iSchools* Organization and current nomenclature they are offering.
- To find out the popular courses and programs on information and computing under the Information Schools or *iSchool* with proper study and analysis of result.
- To know about the developed and developed countries and factors important for creating such disparities.
- To learn about the healthy and sophisticated possibilities of the *iSchool*s and similar programs offering in the developing countries.
- To formulate the proposed strategies, methods and models for creating industry-academia enhanced information, computing and complete interdisciplinary subjects.

3 Methodologies

This is a kind of conceptual study and research work and deals with several theoretical studies for its proper compilation. Among the methods of study few adopted are includes the review of literatures *(such as journals, conference proceeding)* in the field of Information Science, Information Systems, Computer Science, Computer Engineering, Computer Applications etc. Moreover the internet sources also have been used to analysis the current scenario of Information Systems and Knowledge Sciences. For more current analysis of the information schools and its current strategies the consortium of *iSchools* Organization (Caucus) and its current website has been utilized. Moreover general search engine (i.e. Google) has been also used with search term *iSchools*, Information and Development, Developing Countries, Information Sciences—and in all cases results up to ten pages was used to compile the work.

4 *iSchools*: What it is?

Fundamental interest regarding the relationships in between the information, people and technology are mainly performed by the *iSchool*. The main aim as well as objective of this paper is includes the Information Foundations and its essential areas of information management. *iSchool* Organization plays a valuable role for sophisticated Information and Technological Infrastructure including the promotion of interdisciplinary information, technologies as well as allied fields. India is the biggest education sector in the world. Though, the information-technology-people or society centric approaches and more clearly the *iSchool* related category still not available. With introducing and integrating some of the educational management principles, educational policies in the exiting educational platform it is very easy to start an *iSchool*. *iSchool,* internationally offers several programs in the interdisciplinary domains which includes Information-Architecture, Information Design, Economics, Information and IT Policy, Information and Cyber Security, and Telecommunication Science, Knowledge Management, User Experience Design, and Usability Engineering, Conservation and Preservation (with digital preservation, librarianship etc.), Human Computer Interaction, Information Systems etc. [12].

iSchool and its revolution was started in United States and gradually the concepts as well as strategies have been adopted by internationally by many universities. Academic units, such as Schools, Department, Centre etc. have been for creation of a full-fledged *iSchools* and established around the world [9, 11, 13]. Internationally many such academic units become part of the *iSchools* Organization, US for the cooperation, technical and academic support, membership etc. around the world. Academic units under the Faculty/School of Engineering and Technology, Faculty/School of Sciences, Faculty/School of Innovative Technologies, Faculty/School of Interdisciplinary Sciences etc. though internationally many universities have established the Faculty/School of Information Sciences and Technology; either by merging old departments into one umbrella or newly created institutes.

5 Growing *iSchools*: The Integration of components of the Information Sciences (*Computing, Information Studies, Information Technologies, Management, Humanities* etc.)

Information is treated as most important and valuable domain these days. Information is associated with the many domains, subjects and fields which include Information Science, Information Management, Information Technology, Communication Science, Information and Communication Technology, Computer Sciences, Media Studies, Archival-Library Science, and Information Studies [10, 14, 15]. To keep all these types of subjects and fields under one umbrella *iSchool* caucus was established in US. Sharing ideas, broadcasting information about technological development to the society and also formation of *information Schools* around the world are the core target of an *iSchool*.

In 1988 eminent Information Scientist, Toni Carbo from the School of Information Science; *University of Pittsburgh*, Donald Marchan of School of Information Studies; *Syracuse University* and also Richard Lytle from the College of Information Science and Technology of *Drexel University* involved for establishment of an association of international level i.e. Information association called *iSchool* Caucus. Within three years, the *iSchool* Caucus move to the 'Gang of Five'. Gradually many dean/s of several Information and computing related schools had joined the *iSchool* Caucus; due to its novel approaches. Among such schools important are University of Illinois, Florida State University, University of North Carolina, Indiana University, University of Texas and thus this way ten institutes become member of the caucus [12, 16–18].

This way the big demands of *iSchool* inspire the team to change the nomenclature from *iSchool* caucus to simply *iSchool or Information Schools Organization* with much wider agendas and activities [19, 20]. In the initial age of *iSchool Organization* many other fields such as Computing, IT, Electronics and Telecommunication Schools etc. have joined with the *iSchool* Organization. Thus internationally many schools and colleges had been joined the *iSchool* Organization for the interaction of information systems and technologies with applications to the people and organizations. This way more and more interdisciplinary skills, degree holders, and researchers joined together. Progressively the institutes with similar agendas and objectives joined from various countries of Asia, Europe, Australia and latest in the list is Africa (from the 2015). This way the programs, research agendas, academic events along with similar activities of Information Sciences, Information Technology, Communication Sciences etc. meet the *iSchool* Organization and established separate unit of *iSchool as an academi*c department in the respective universities [12].

6 Developing Countries, iSchools and Rapid Scenario

This way combination of *iSchool* are internationally growing for the betterment of information systems, information technology and complete information infrastructure development. Ultimately information and technology are responsible for development in many respects and thus it results development of a country in many ways. *iSchool* listed schools and departments now have increased and classified with several tiers. According to the *iSchool* Organization Tier 1 Schools are 27. While 26 are categorized as Tier 2. Those are depicted in Table 1 with nomenclature of the Schools/Departments etc. [21, 22].

As mention is Wikipedia that, "*A developed country, industrialized country, or more economically developed country MEDC), is a sovereign state that has a highly developed economy and advanced technological infrastructure relative to other less industrialized nations. Most commonly, the criteria for evaluating the degree of economic development are gross domestic product GDP), gross national product GNP),*

Table 1 Information schools fall under the iSchools organization members list (Tier 1 and 2)

University	Information schools
Information schools (Tier 1)	
University of California, USA	Berkeley School of Information
University of California, Irvine, USA	The Donald Bren School of Information and Computer Sciences
University of California, Los Angeles, USA	Graduate School of Education and Information Studies
Carnegie Mellon University, USA	School of Information Systems and Management, Heinz College
University of Copenhagen, Denmark	Royal School of Library and Information Science
Drexel University, USA	College of Computing & Informatics
Florida State University, USA	College of Communication and Information
Georgia Institute of Technology, USA	College of Computing
Humboldt University of Berlin: Germany	Berlin School of Library and Information Science
University of Illinois, USA	School of Information Sciences
Indiana University, USA	School of Informatics and Computing
University of Kentucky, USA	College of Communication and Information
University of Maryland, USA	College of Information Studies
University of Michigan, USA	School of Information
University of North Carolina, USA	School of Information and Library Science
University of North Texas, USA	College of Information
Pennsylvania State University, USA	College of Information Sciences and Technology
Rutgers, The State University of New Jersey: USA	School of Communication and Information
University of Pittsburgh, USA	School of Information Sciences
University of Sheffield, UK	Information School
Singapore Management University, Singapore	School of Information Systems
Syracuse University, USA	School of Information Studies
University of Tampere, Finland	School of Information Sciences
University of Texas, Austin, USA	School of Information
University of Toronto, Canada	Faculty of Information
University of Washington: USA	Information School
Wuhan University, China	School of Information Management
Information schools (Tier 2)	
University of Amsterdam, Netherlands	Graduate School of Humanities, Archives and Information Studies.

(continued)

Table 1 (continued)

University	Information schools
University of Boras, Sweden	The Swedish School of Library and Information Science.
University of British Columbia, Canada	School of Library, Archival and Information Studies.
Open University of Catalonia, Spain	Information and Communications Science Studies.
Charles Sturt University, Australia	School of Information Studies
Cornell University, USA	Faculty of Computing and Information Sciences
Hacettepe University, Turkey	Department of Information Management
University College London, UK	Department of Information Studies
University of Maryland, USA	College of Information Studies
University of Maryland, Baltimore County, USA	Department of Information Systems
University of Melbourne, Australia	Melbourne School of Information
University of Michigan, USA	School of Information
University of Missouri, USA	School of Information Science and Learning Technologies
Nanjing University, China	School of Information Management
Northumbria University, UK	Information School
NOVA University of Lisbon, Portugal	School of Statistics and Information Management
University of Porto, Portugal	Faculty of Engineering in cooperation with the Faculty of Arts
Robert Gordon University, UK	Department of Information Management of Aberdeen Business School
Sun Yat-sen University, Guangzhou, China	School of Information Management China
Sungkyunkwan University, Seoul	Library and Information Science Department
Télécom Bretagne, France	Department of Logic Uses, Social Sciences and Information
University of Tennessee, Knoxville, USA	School of Information Sciences
University of Tsukuba, Japan	Graduate School of Library, Information and Media Studies
University, Waikato, New Zealand	Faculty of Mathematics and Computational Sciences
University of Wisconsin, Madison	School of Library and Information Studies
University of Wisconsin, Milwaukee	School of Information Studies

Table 2 Few developed countries at a glance

Few developed countries							
Norway	Australia	Switzerland	Denmark	Netherlands	Germany	Ireland	United States
Canada	New Zealand	Singapore	Hong Kong	Liechtenstein	Sweden	UK	Iceland
South Korea	Israel	Luxembourg	Japan	Belgium	France	Austria	Finland

the per capita income, level of industrialization, amount of widespread infrastructure and general standard of living".

Developed Countries are basically comes after the post-industrial economies and normally focuses on service sector than the industrial sector. The developing countries normally engaged in the process of industrialization and an undeveloped country is in the stage of pre-industrial situation. According to the *International Monetary Fund*, it is important to note that advanced economies comprise with 60.8% of Global Gross Development Products GDP) based on nominal values while depending upon purchasing-power parity PPP) it is about i.e. 42.9%), in the context of 2015. Largest advanced economies top ten according to the GDP) in both nominal as well as PPP are include: Australia, Canada, France, Germany, Italy, Japan, South Korea, Spain, the United Kingdom, and the United States [21, 22]. Few Developed countries are listed in Table 2.

The third world countries are the less developed country and these are also classified as underdeveloped country. The developing countries are also deemed as this category. Countries having developed industrial base and importantly with higher Human Development Index (HDI) compare to other countries are also categorized into this. From the 1990s the growth rate of developing countries are better than the developed ones.

Nation's GDP per capita compared to other nations basically treated as main criteria, though there are no other criteria as such. The general term *less-developed country* should not be confused with similar term 'least developed country'. Few countries under the Developing tag are mention in Table 3.

Hence ultimately following are responsible for the information development which includes but not limited to the following:

- People and Country/States/Territories have lower life expectancy.

Table 3 Few developing countries at a glance

Few developing countries							
Argentina	Bangladesh	Brazil	Bulgaria	Colombia	China	Hungary	Malaysia
Mauritius	Mexico	Nigeria	Oman	Pakistan	Philippines	Poland	Qatar
Russia	South Africa	South Korea	Taiwan	Thailand	Ukraine	UAE	India

- People/Country/States/Territories have less education.
- People/Country/States/Territories have less money or income.

Information has the power to solve all these facets. Internationally such countries are treated as developed countries which cultivate huge and sufficient information. The concept of digital industry has also creates a knowledge society supported with high-tech global economy [23–25]. More over this results change of industrial sectors to the service sectors including information, Information Technology industries etc. Importantly information sectors and ICT have also played a leading role for the development of other business too. The information industry is responsible for the individuals to identify their personalized needs and this results the growth of development in many ways [24, 26, 27]. Ultimately individual development helps to create a whole development in many prospects. Thus *information transfer cycle* including the collection, selection, organization, processing, management of the entire information and their uses for the development and economic activities may lead developed economics and countries and that called as knowledge economy.

7 Future Course of Action

In generally Information related programs such as Computer Science, Information Technology, Information Systems, Information Science, Computing, Computer Technology, Library and Information Science, Information Management etc. normally practiced in the different schools and department internationally in most countries and specifically in the under developed countries [19, 28, 29]. Though, the combination and integration of these subjects and departments have wider potentialities in diverse mode. All these subjects and departments are deal with *information and technology*—thus purely responsible for the development of the education of every subjects as information is the core for any other subjects, similarly the development of information culture-transparencies-development are also the big names for the sustainable development and thus *iSchool* establishment are the important steps for development of every kind. Thus developing countries needs to develop *iSchool* either by merging related departments (which are mention above in this section of paper) or building the full-fledged schools dedicated to the information-technology and social interaction [30–32]. The following Fig. 1 Showing the Developing Countries (Green), newly developed nation (Red), Newly Industrialized country (Blue) [21, 22].

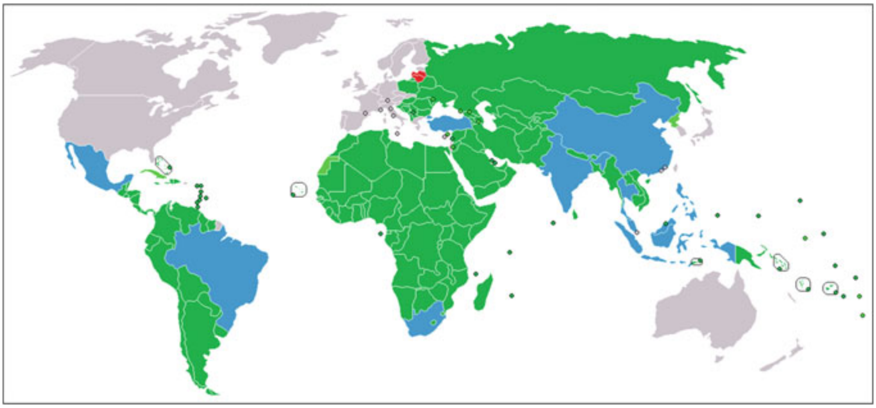


Fig. 1 Showing the developing, newly developed and newly economic countries

8 Findings and Outcome

- *iSchool* are combining not only from the Computer Science, Information Technology, Information Systems, Information Science, Computing, Computer Technology, Library and Information Science, Information Management but also some other required domains and disciplines such as *Cognitive Science, Mathematics and Statistics, Management Science, Social and Policy Studies* etc.
- *iSchool* earlier only treated as consortium and association that deals for the development of information and technologies; but gradually it is now become an important academic department in many universities, internationally.
- *iSchool* Organization listed members now Tier-1 is 27, Tier-2 is 26; while Tier-3 holds 20 institutes. Apart from these 5 institutes have been listed as an Associate Members. The Founding members are listed in *Tier-1 Schools*.
- *iSchool* internationally played a leading role for the academic development and collaboration with other departments etc. and these ultimately results the better culture in interdisciplinary research and innovation.

9 Recommendation

- Among the emerging areas of *iSchool* few technology oriented are include the Cloud Computing, Green Computing, Usability Engineering, Human Computer Interaction etc. Thus in the *iSchool* of the developing countries these areas may be provided as a full-fledged Academic discipline, as a specialization of the related subjects, research areas with technical, social, business and other domain based focuses etc.

- Governmental initiatives, planning, efforts from the educational ministries and departments are very much important in many cases.
- With combination of Information and Computing related departments it is may also possible to establish a full-fledged university on Information Science or Information Science.
- *iSchools* are basically combined with the Computer Science, Information Technology, Information Systems, Information Science, Computing, Computer Technology, Library and Information Science, Information Management and some other disciplines such as Cognitive Science, Mathematics and Statistics, Management Science, Social and Policy Studies—and all these are related with the Information or Computing Sciences; thus it would be better if the School named on 'Information Science and Technology' or 'Information Science and Computing' or simply 'Informatics and Computing'.
- Solving of initial problem on the availability of manpower is positively possible by the adjustment of guest professors, adjunct professors etc.
- Developing countries need to put on information development, information infrastructure building, information systems promotion for solid knowledge economy creation and all these cases information play a healthy role.

10 Conclusion

Information is everything. This is the power; this is the development and reason for the growth of every kind. Today we are living in the Information Technology age where many of us surrounded by the technological devices and products. Behind the development of these products information is there. The promotion of health, education, governance, infrastructure development many ways depends on the information and technologies. And for all these skill manpower is most important and vital. Government of each and every nation moving towards best and healthy interdisciplinary research and in this case *iSchool* should be healthy and positive efforts. It is interesting that many countries already become part of *iSchool* Organization, US such as Uganda, Turkey etc. Even in many developing countries *iSchool* are developed by the individual basis without membership of the US based consortium. But ultimately all these efforts are towards information development and information richness in many prospects. So the rest countries are also need to jump in the academic initiatives for solid and real development.

References

1. Chu H (2012) iSchools and non-iSchools in the USA: an examination of their master's programs. Educ Inf 29(1):1–17
2. https://en.wikipedia.org/wiki/List_of_information_schools. Accessed on 31 July 2016

3. Dillon A (2012) What it means to be an iSchool. J Educ Libr Inf Sci 53(4):267. Martin SB (1998) Information technology, employment, and the information sector: trends in information employment 1970–1995. J Am Soc Inf Sci 49(12):1053–1069
4. Lopatovska I, Pattuelli MC, Bates MJ, Buckland M, Dalbello M, Hastings S, Giannini T (2012) iSchools and l-schools: converging or diverging communities? Proc Am Soc Inf Sci Technol 49(1):1–3
5. Paul PK, Dangwal K LK, Sridevi V (2012) I-programmes: the new interdisciplinary knowledge cluster with ample job opportunities-prospects and problems in Indian scenario. Int J Appl Bus Econ Res 10(2):283–290
6. Hjorland B, Albrechtsen H (1995) Toward a new horizon in information science: domain-analysis. J Am Soc Inf Sci 46(6):400
7. Ibekwe-SanJuan F (2012) The French conception of information science: "Une exception française"? JASIST 63(9):1693–1709
8. Paul PK, Chatterjee D (2013) Information Engineering (IE) and its overview with special reference to MTech—information engineering in Indian perspective. Sci Tech Int J Eng Sci 1(1):71–79
9. Marchionini G, Solomon P, Davis C, Russell T (2007) Information and library science MPACT: a preliminary analysis. Libr Inf Sci Res 28(4):480–500
10. Olson GM, Grudin J (2009) TIMELINES the information school phenomenon. Interactions 16(2):15–19
11. Paul PK (2013) MSc-information science (Geo Informatics): overview emphasizing twoproposed curriculum for sophisticated geo spatial development. Int J Pharm Biol Res 4(5):218–227
12. http://ischools.org/members/directory/. Accessed on 31 July 2016
13. Parameswaran M, Whinston AB (2007) Research issues in social computing. J Assoc Inf Syst 8(6):336
14. Mezick EM, Koenig ME (2008) Education for information science. Annu Rev Inf Sci Technol 42(1):593–624
15. Paul PK, Chatterjee D (2015) I-schools and the present worldwide trend and the Indian scenario. 3rd edn. IGI global-encyclopedia of information science and technology, pp 2525–2534
16. Murray RM, Astrom KJ, Boyd SP, Brockett RW, Stein G (2003) Future directions in control in an information-rich world. IEEE Control Syst Mag 23(2):20–33
17. Patterson L, Martzoukou K (2012) An examination of Canadian information professionals' involvement in the provision of business information synthesis and analysis services. J Libr Inf Sci 44(1):47–64
18. Paul PK (2013) Information systems: from meaning to its changing domain at a glanc. Abhinav Natl J Sci Technol 2(1):1–8
19. Paul PK MSc information science with possible computing specialization: an analytical study. Int J Syst Simul 7(2):60–67
20. Paul PK (2013) MSc-information science: a deep study of the proposed specialized curriculum in Indian scenario. Int J Intell Inf Process 7(2):90–98
21. https://en.wikipedia.org/wiki/Developing_country. Accessed on 31 July 2016
22. https://en.wikipedia.org/wiki/Developed_country. Accessed on 31 July 2016
23. Paul PK, Ghose MK, Ghosh M, Ganguly J (2015) MSc-information science (Quantum Information Science) specialization: its requirement and proposed model for future I-school dedicated to next generation information-technology-community interaction. SRELS J Inf Manag 52(2):93–99
24. Paul PK, Rajesh R (2015) MTech-information science: a new age proposed programme and its requirement with SWOT analysis in Indian educational context. Int J Enterp Netw Manag 6(4):286–298
25. Walters WH, Wilder EI (2015) Disciplinary, national, and departmental contributions to the literature of library and information science, 2007–2012. J Assoc Inf Sci Technol
26. Wallace DP (2009) The iSchools, education for librarianship, and the voice of doom and gloom. J Acad Libr 35(5):405–409

27. Zhang P, Benjamin RI (2007) Understanding information related fields: a conceptual framework. J Am Soc Inform Sci Technol 58(13):1934–1947
28. Wiggins A, Sawyer S (2012) Intellectual diversity and the faculty composition of iSchools. J Am Soc Inform Sci Technol 63(1):8–21
29. Zhu Z (2008) Pondering over the impact of I-Schools movement on China's postgraduate education of information science. Res Libr Sci 8:002
30. Paul PK (2013) Human computer interaction and its emerging affiliation with information science (IS): an overview. J Bus Manag (JBM) Int J 5(2):162–167
31. Paul PK (2013) Web engineering: emerging domain emphasizing its role and possibilities in MSc-information science specialization. Abhinav Natl J Arts Educ 2(10):8–16
32. Wu D, He D, Jiang J, Dong W, Vo KT (2012) The state of iSchools: an analysis of academic research and graduate education. J Inf Sci 38(1):15–36. Buckland M, Liu Z (1995) History of information science. Annu Rev Inf Sci Technol 30:385–416

Information System Used to Study on Effect of Groundwater Quality

Venkateshwarlu Musini, Suresh Merugu and T. Rohini Reddy

Abstract The groundwater high content of fluorine is problem globally. This due to alteration of mineral constituents of the rock such as apatite, fluorite, biotite muscovite and from use of fertilizers in agriculture. Portable drinking water is not available in the region there by using only groundwater. This paper through light on the influence of fertilizer is agriculture add further contamination of ground water made unsafe to drink groundwater using organic forming reduce the contamination levels. Low content of fluorine help in strengthening teeth and effect skeletal and bones of the content is more. In Telangana state India this is one of the major health problem observed in the public. Hydro geochemical studies had been carried out on 25 groundwater samples collected from study area i.e. Wailpalli area. Various Parameters like Electrical conductivity, pH, TDS and major on concentrations such as Ca^{+2}, Mg^{+2}, Na^{+}, K^{+2}, $CO_3{}^{2-}$, $HCO_3{}^{-}$, Cl^{-} and F^{-} were calculated. Analytical result indicates that groundwater is acidic to alkaline and mostly hard in nature.

Keywords Hydro geochemistry · Groundwater quality · Major ions · Fluoride concentration

1 Introduction

One third world population of the world with usage of water for various activities depends on groundwater for drinking [13]. Hence management of water had becomes important. Hamzaoui-Azaza et al. [8], studies the geochemical effect on water quality leads to improve the management of water. Studies done by the geochemical characteristics of groundwater [3, 4, 10, 11, 26, 27]. For groundwater geochemistry hydrochemistry of groundwater made by [7, 14, 18]. Graphical representation and

V. Musini (✉) · T. Rohini Reddy
Department of Civil Engineering, CMR College of Engineering & Technology, Medchal Road, Kandlakoya, Hyderabad, Telangana 501401, India
e-mail: venkatmusini@gmail.com

S. Merugu
Research & Development (R&D) Centre, CMR College of Engineering & Technology, Medchal Road, Kandlakoya, Hyderabad, Telangana 501401, India

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_53

Wilcox diagrams will help to recognize hydro geochemistry of groundwater. To evaluate suitability of the groundwater on agriculture. Increase of unwanted constituents in groundwater is also due to crops paddy and cotton using fertilizer.

Geochemical study of major ions in groundwater of wailpalli area, Nalgonda district of Telangana State had been carried out in addition to fluoride ion distribution. Low Fluoride water (less than 0.6 mg/L) causes dental problems and high fluoride (greater than 1.2 mg/L) results in fluorosis [31, 32]. In India about 62 lakh people are suffering with dental, skeletal and/or non skeletal fluorosis [28]. Present study attempt has been made to investigate the fluoride ion distribution in the groundwater of wailpalli area demarcate safe and unsafe zones based on analysis of 25 groundwater samples fluoride content.

2 Study Area

Study area located in the western part of Nalgonda district and lies between latitudes 17000′–17010′ N and longitudes 78048′–79000′ E covering Survey of India top sheet 56 K/16. Major part of the investigated area falls in Narayanpur mandal while very small part falls in Mungod Mandal in the east and Chandur Mandal in the south mentioned in Fig. 1.

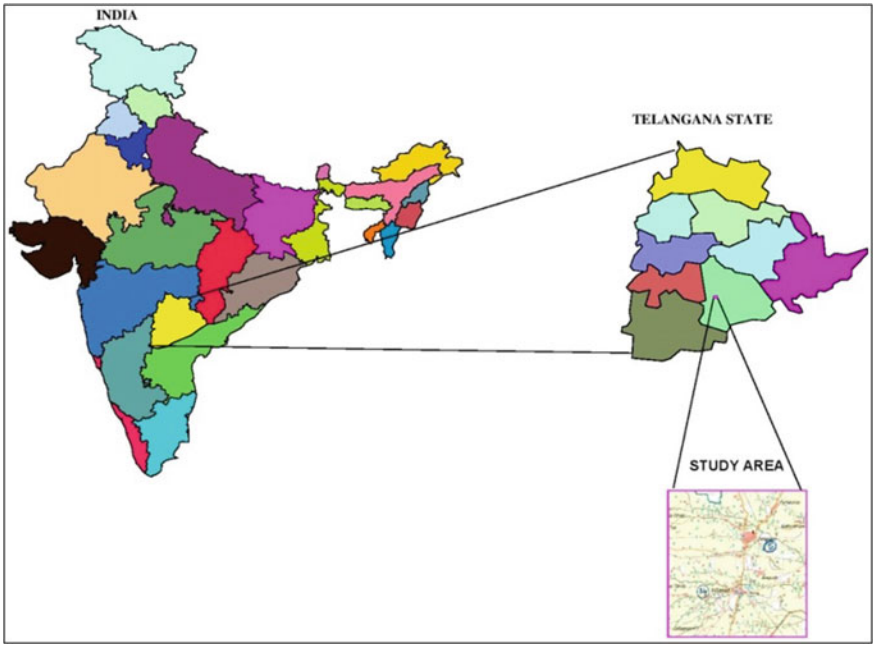


Fig. 1 Location of the area of study

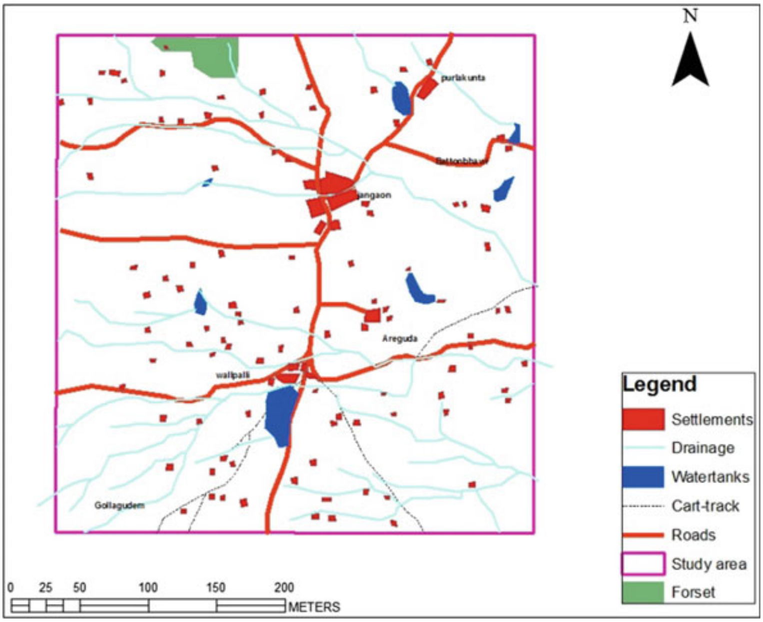


Fig. 2 Base map area

The study area covering of 213 km^2 comprises eleven revenue villages in the three mandals. There about 36 hamlets, which are located in the interior hilly terrain in the west part of the area. These hamlets/tandas are mostly occupied by Lambada tribes. Among eleven villages, five villages viz., Kurmakothaguda, Jangao, Wailapalli, Puttapaka and Chillapur fall in Narayana mandal. Base map area of the area of study given in Fig. 2.

3 Geology of the Study Area

Study area composed of rocks of Archaean group which represent older group and peninsular gnessic complex. Schist of biotite and hornblende belongs to older rocks and porphyritic granite gneisses, pink granite with pegmatite, quartz and epidote as veins injected into these rocks. Among the older rocks, biotite schists and hornblende schists are found to occur as inclusions in the migmatitic gneiss. The common and characteristic feature of these intrusive is there alignment parallel to the gnessocity of enclosing migmatites. Biotite schist inclusions which are fine grained, frequently seen in the grey migmatitic gneisses are in the form of streaks, bands, clots, pods, lenses etc.

Among this group of rocks, grey granite gneiss occupies major part of the area and occurs as sheet like exposures or as gentle demo-like hills. The soil is formed due to weathering of granite gneisses. Seen on plain and valley bottom surrounded by red soils.

4 Structure

Structural changes resulted development of joint fracture in the rocks. Major sets of joint trends are in N 100 W–S 100 E and N 750 E–S 750 W and are vertical in nature. These are tensile, vertical fractures. They occur near the surface and are as a result of decompression of the crystalline rocks. The EW and NS fractures are normally filled with intrusive like dolerite dykes. The drainage of the area mentioned in Figs. 3 and 4 pertain to the road and settlement of the area. The drainage map of the present study region is given in Fig. 3. Evapotranspiration and artificial extraction are the main discharge path way of groundwater.

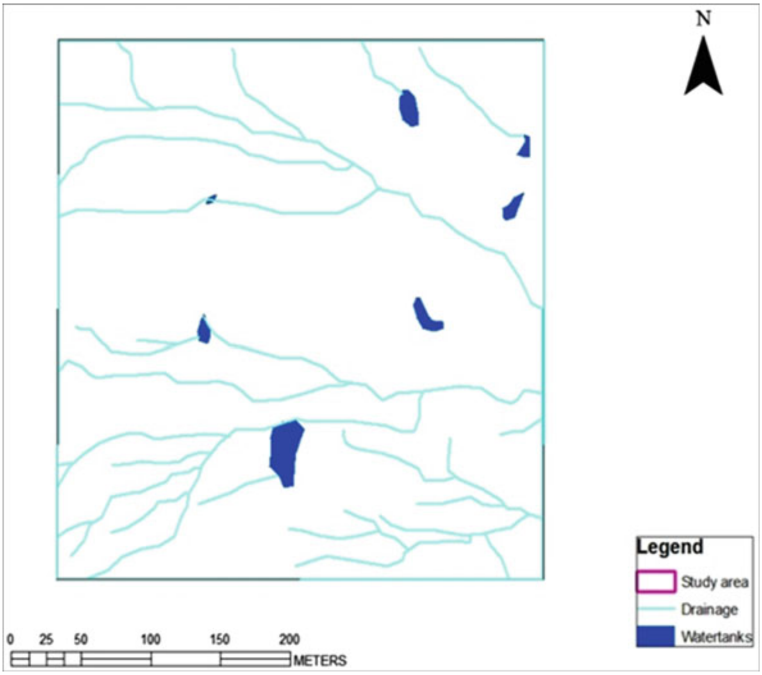


Fig. 3 Drainage of the area

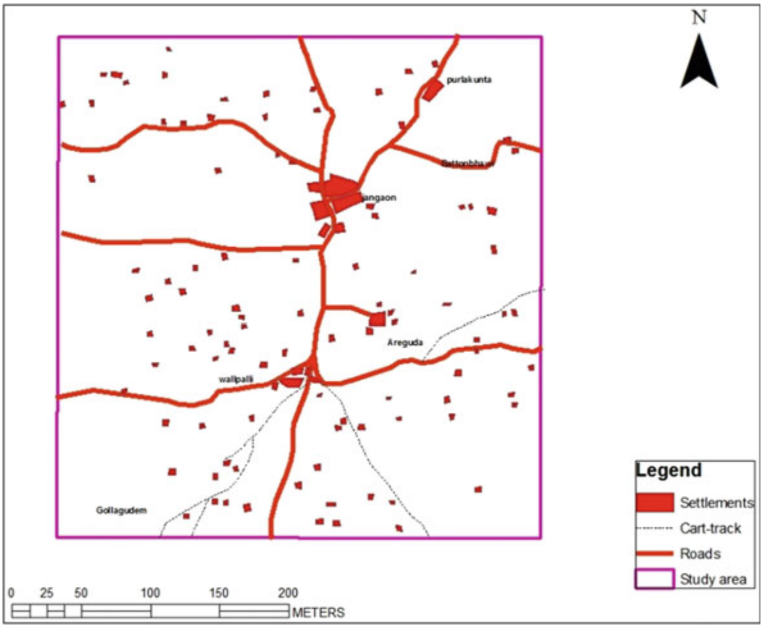


Fig. 4 Road and settlements of the area

5 Materials and Methods

Representative 25 groundwater sample collected from the study area in and around Wailpalli area, Nalgonda District of Telangana, Samples were analyzed for major ions by employing standard methods (APHA 1985). pH and conductivity meter use to measure pH and Electrical conductivity (EC). Total Dissolved Solids (TDS) calculated by multiply with electrical conductivity factor (0.55–0.75) taking into the concentration of ions. Fluoride was analyzed by Spectrophotometer by Zirconium Alizarin Red S method. By using titration method Total alkalinity (TA), Total hardness (TH) as $CaCO_3$ and Calcium (Ca^{+2}) with H_2SO_4 and EDTA respectively, Flame photometer used for estimation of Sodium and Potassium and Chloride estimated by AgNO. All values are expressed in milligrams per liter (mg/L) except pH.

6 Results and Discussions

Result analysis of samples of groundwater mentioned in Table 1 and statistical parameters of groundwater sample mentioned in Table 2. Groundwater has pH between of 7.2 and 8.5 indicate acidic to in TDS varies from 236 to 1322 mg/L the Total hardness varies from 125 to 1690 mg/L. Hardness of these samples were correlated

Table 1 Chemical analysis of grounded samples of water samples in area (Mg/L)

Results of the chemical analysis of grounded samples of water samples in area (Mg/L)

S. No	pH (S/cm)	EC	TH	Ca	Mg	Na	K	CO_3	HCO_3	CL	SO_4	NO_3	F	SAR	RSC
1	7.96	1513	395	52	64	161	0.98	0	537	138	36	92	2.4	3.52	0.9
2	7.96	542	225	42	29	21	0.47	0	275	21	7.7	8	1	0.6	0
3	8.22	526	175	44	16	35	0.27	0	244	18	11	14	2.4	1.13	0.5
4	7.86	670	180	44	17	71	0	0	311	21	12	25	3	2.31	1.5
5	8.11	1150	365	60	52	93	2.1	0	438	117	26	20	1.6	2.11	−0.2
6	8.2	820	225	36	33	63	35	0	317	78	19	14	1.1	1.81	0.7
7	7.92	1400	330	28	63	167	1.8	0	494	50	144	52	2.1	4.01	1.5
8	7.83	2050	350	42	60	279	1.1	0	433	386	96	24	2.2	6.48	0.1
9	8.04	2040	390	34	74	270	0.9	0	653	277	72	19	2.3	5.94	2.9
10	8.12	3650	820	36	177	438	1.6	0	714	666	144	166	2	6.65	−4.7
11	7.54	1150	305	32	55	115	1.4	0	500	64	16	9.4	4.2	2.87	2.1
12	8.34	1680	445	48	79	172	1.1	0	610	167	67	49	3.9	3.54	1.1
13	8.36	1264	370	32	70	114	0.9	0	555	92	26	25	4.2	2.58	1.7
14	7.9	1090	325	48	50	87	0.66	0	488	60	26	43	4.2	2.11	1.5
15	7.59	1130	235	40	33	150	0.39	0	445	64	20	41	4.4	4.25	2.6
16	8.12	677	190	26	30	62	0.2	0	317	25	12	12	3.6	1.96	1.4
17	8.12	300	90	34	1	18	0.2	0	134	14	7.2	16	4.3	0.82	0.4
18	7.59	830	175	44	16	110	0.2	0	409	43	11	11	3.6	3.63	3.2
19	7.25	590	200	40	24	45	0.27	0	281	14	10	13	3.8	1.37	0.6
20	8.25	540	215	32	33	24	0.27	0	262	18	5.3	14	1.8	0.7	0

(continued)

Table 1 (continued)

Results of the chemical analysis of grounded samples of water samples in area (Mg/L)

S. No	pH (S/cm)	EC	TH	Ca	Mg	Na	K	CO_3	HCO_3	CL	SO_4	NO_3	F	SAR	RSC
21	7.4	918	235	60	21	99	0.27	0	427	53	12	13	4	2.81	2.3
22	7.83	820	205	60	13	90	0.27	0	354	60	19	23	4	2.72	1.7
23	7.33	1740	435	50	75	175	0.39	0	738	113	48	79	5.8	3.64	3.4
24	8.14	968	275	56	33	87	0.9	0	445	60	15	7.6	2.8	2.28	1.8
25	8.25	962	235	62	19	106	1.37	0	458	43	14	28	1.2	3	2.8

Table 2 Statistical parameters of ground water samples

Statistical analysis	pH	EC	TH	Ca	Mg	Na	K	CO_3	HCO_3	CL	SO_4	NO_3	F
Mean	7.929	1110.8	295.6	43.28	45.48	119.29	2.12	0	433.56	106.48	35.048	32.72	3.036
Median	7.96	968	235	42	33	99	0.66	0	438	60	19	20	3
Mode	8.12	1150, 820	235	44, 60, 32	33	161, 21, 35	0.27	0	317, 445	60	12, 26	14	4, 2
Standard deviation	0.311	685.66	140.3	10.45	34.75	94.83	6.73	0	147.05	142.03	38.98	34.46	1.23
Skewness	−0.67	1.86	1.91	0.29	1.99	1.6	4.55	0	0.22	2.68	1.79	2.47	0.07
Range	1.11	1524	730	36	176	420	2.1	0	604	642	134	154	4.8
Minimum	7.25	526	90	26	1	18	0	0	134	14	10	8	1
Maximum	8.36	2050	820	62	177	438	2.1	0	738	666	144	166	5.8

with standard values [19]. This results shows that 75.86% of samples belongs to hard category. Carbonate and bicarbonate ions in groundwater of the study area vary from 0 to 45 mg/L. and 27.1 to 225.7 mg/L respectively. It is observed that bicarbonate in all the groundwater samples of the area are in desirable limits of 1000 mg/L. The Concentration of Sodium varies from 21.9 to 151.8 mg/L and in within the prescribed limit of 200 mg/L for drinking water [32], Potassium concentration various from 0.6 to 53.6 mg/L are within the acceptable limit of 10–15 mg/L and the calcium concentration varies from 20 to 514 mg/L which is again within the prescribed limit of calcium in drinking water is 75 mg/L (ISI 1983). It is observed that 79.37% of samples are within prescribed limit. The magnesium varies from 1.2 to 147.1 mg/L. The prescribed limit of magnesium in drinking water is 30 mg/L. (ISI 1983) and it is observe that 56.89% of the samples of the study area exist the desirable limit prescribed for drinking purposes.

As per [31, 33] guidelines the acceptable range of fluoride ion content shall be within 0.6–1.5 mg/L. The optimal range for human consumption is 0.8–1.2 mg/L (Table 1). It's most important to provide a natural protection against dental carries by reducing the solubility of the enamel under acidic condition [33]. People with a drinking water source containing less than the optimal amount of fluoride will therefore be more susceptible for dental problems. Although the advised fluoride contents in drinking water ranges between 0.8 and 1.2 mg/L based on a daily intake of two liters in hot climate when people drinks more than this quantity the optimal fluoride content is lower 0.6–1.0 mg/L [30, 31]. The Investigated area contains 89% of total ground water samples are within optimal ranged less than 0.8 mg/L. The concentration of fluoride varies from 0.5 to 2 mg/L (Table 2). It is observed that fluorides of samples are within the limits of mentioned by standards of BIS 2003 for drinking water. Fluoride concentration in the groundwater of the study area is correlated with the concentration of certain dissolved constituents and other parameters. Leaching of fluoride from minerals in the soil is due to weathering [15, 20, 23, 34]. The distribution of fluoride is high in north eastern parts of area of investigation. In the groundwater high fluoride content is due to climatic change [5, 9, 29, 21], evaporative concentration and hydrology and regional geology of the region influence the concentration of fluorides [1, 12, 16, 23]. Groundwater stastical analysis mentioned in Table 2. The intensity of weathering and leachable fluoride in terrain are the factors fluoride content in water rather than more presence of fluoride bearing minerals in the bulk rocks/soils [16, 17, 24, 25].

The major lithological unit in the study area is extensive dolerite dykes which cut granitic hill ranges and show insignificantly low fluoride ion content but they enhances leaching of fluoride ion by water acting as barriers for groundwater flow there by permitting a longer contact time of water, with fluoride bearing minerals.

7 Sodium Adsorption Ratio (SAR)

Sodium adsorption ratio (SAR) is a tool used towards water suitability in agricultural irrigation, as determined from the concentrations of the main alkaline and earth alkaline cations present in the water. Higher the values of SAR, the cation exchange complex may become saturated with sodium, which can destroy the soil structure owing to dispersion of clay particles which leads to the less suitable the water is for irrigation [22]. It is ratio between concentration of sodium ion and square root half the concentration value of calcium and magnesium ions together.

$$\mathrm{SAR} = \mathrm{Na}/(\surd(\mathrm{Ca} + \mathrm{Mg})/2) \tag{1}$$

The SAR values lies between 0.60 and 6.65 in the water samples collected from the study area (Table 1). Electrical conductivity values and SAR values, High-salinity hazard (C3) and low-sodium hazard noticed in the ground water more than 80% sample (1, 2, 3, 4 and 5) as the fall in the zone of C3-S1. This water can be used for agriculture in all type of soils. Whereas sample (6) fall in the C4-S2 zone which 10% approximately not suitable for agriculture.

7.1 Percent Sodium (NA^+)

Sodium ion concentration (Na^+) is calculated on the basis of ratio between as the given formula.

$$\%\mathrm{Na} = (\mathrm{Na} + \mathrm{K})/(\mathrm{Ca} + \mathrm{Mg} + \mathrm{Na} + \mathrm{K}) * 100 \tag{2}$$

8 Residual Sodium Carbonate (RSC)

Alkalinity of water is effected water Carbonate ions due to precipitation of alkaline earth metals Ca^{2+} and Mg^{2+} [6]. $NaHCO_3$ is formed when carbonates concentration is more than alkaline earth due to presence of Na^{2+} this is known as residual sodium carbonate. Suitability of water for agriculture can be used based on relation between carbonate and alkaline earth concentrations. The RSC is calculated using the formula.

$$\mathrm{RSC} = [(\mathrm{HCO}_3^- + \mathrm{CO}_3^{2-}) - (\mathrm{Ca}^{2+} + \mathrm{Mg}^{2+})] \tag{3}$$

Concentration of ionic expressed in milli equivalents per liter.

RSC values lies between 0 and 3.4 meq/L from the water samples taken (Table 1). In the present study, half of the water samples shows values less than 1.25 meq/L.

Water is categorized as suitable, if the RSC value is less than 1.25 meq/L; marginally suitable, 43% samples of groundwater have RSC 1.25 meq/L indicate it's suitable for agriculture.

9 Conclusion

It is observed that after leaching of fluoride from fluoride rich minerals into groundwater, enriching in hilly terrain itself. Concentration is problematic for drinking in addition to this extensive usage of fertilizer there is abnormal increase in some of the places of the to study area. Giving awareness programs to the farmer's and encouraging the organic farming reduces the concentration of unsafe elements from the groundwater. Groundwater flows down to topographic low (plain) and further gets enriched during the passage. However, its enrichment is neither progressive nor showing any significant trend. The positive relationship of fluoride with bicarbonate and negative relationship with Ca and Mg is the characteristic feature. pH and sodium are also showing positive relationship with fluoride. Fluoride shows highly scattered in all relationships with other constituents where the fluoride concentrations are low (<2 mg/L). Groundwater is having fluoride concentration associated with high sodium and low calcium. Present study on about groundwater quality based on to fluoride ion concentration in Wailpalli area indicates that the groundwater is acidic to alkaline in nature. Increase of fluoride content in groundwater samples of the fluoride content is higher than the optimal limit is few samples.

Acknowledgements We express sincere thanks to Sri Ch. Gopal Reddy, Secretary, CMRGI, Principal and HOD, CMR College of Engineering and Technology Hyderabad, for providing necessary Laboratory facilities, to publish this paper.

References

1. Agarwal V, Vaish AK, Vaish P (1997) Groundwater quality focus on fluoride on fluorosis in Rajasthan. Curro Sci 73:743–746
2. Aghazadeh N, Mogassem A (2010) Geochemical characteristics of groundwater. Environ Monit Assess. https://doi.org/10.1007/S10661-0101574-4
3. Ahmad Z, Qadir A (2011) Geochemical characteristics of groundwater. Environ Monit Assess 175(1–4):9
4. Alexakis D (2011) Geochemical characteristics of groundwater. Environ Monit Assess. https://doi.org/10.1007/s10661-011-1884-2
5. Apambire WM, Boyle DR, Michal FA (1997) Geochemistry genesis and health implications of fluoriferous groundwater in the upper region Ghana. Environ Geol 33:13–24
6. Eaton FM (1950) Significance of carbonates in irrigation waters. Soil Science 69(2):123–134
7. Garrels RM (1967) Genesis of some groundwaters from igneous rocks. In: Abelson PH (ed) Researchers in geochemistry 2. Wiley, New York, pp 405–420
8. Hamzaoui-Azaza F, Ketata M, Bouhlia R, Gueddari M, Ribeiro L (2011) Fluoride containing groundwater in India

9. Handa BK (1975) Geochemistry and genesis of fluoride containing groundwater in India. Groundwater 13:275–281
10. Jeevanandam M, Kannan R, Srinivasalu S, Rammohan V (2006) Hydrogeochemistry and Groundwater Quality Assessment of Lower Part of the Ponnaiyar River Basin, Cuddalore District, South India. Environ Monit Assess 132(1–3):263–274
11. Laluraj CM (2005) Groundwater chemistry of shallow aquifers in the coastal zones of Cochin, India. Appl Ecol Environ Res 3(1):133–139
12. Muralidharan D, Nair AP, Satyanarayan V (2002) Fluoride in shallow aquifer in Rajgarh Tehsil of Churu District Rajasthan; an arid environment. Curr Sci 83:699–702
13. Nickson RT, McArthur JM, Shrestha B, Kyaw-Myint TO, Lowry D (2005) Arsenic and other drinking water quality issues, Muzaffargarh District, Pakistan. Appl Geochem 20(1):55–68
14. Paces T (1973) Steady State Kinetics and Equilibrium between Groundwater and Granite Rock. Geochimica et Cosmochimica Acta. 37:2641–2663
15. Ramamohan Rao NV, Suryaprakash Rao K, schilling RD (1993) Fluorine distribution in water of Nalgonda District, Andhra Pradesh India. Envron Geol 21:84–89
16. Ramesam V, Rajagopalan K (1985) Fluoride ingestion into the natural waters of hard rock areas Peninsular India. J Geol Soc India 26:125–132
17. Sahu NK, Karim MA (1989) Fluoride incidence in natural waters in Ameeli District Gujarat. J Geol soc India 33:450–456
18. Sarin MM, Krishnaswamy S, Dilli K, Somayajulu BLK, Moore WS (1989) Major ion chemistry of the Ganga-Brahmaputra river system: Weathering process and fluxes to the Bay of Bengal. Geochimica et Cosmochimica Acta 53:997–1009 https://doi.org/10.1016/0016-7037(89)90205-6
19. Sawyer CN, Mc Carty PL (1967) Chemistry for pamtary engineers India. McGraw-Hill, New York, 518 p
20. Saxena VK, Ahmad S (2001) Dissolution of fluoride in groundwater, a water-rock interaction study. Environ Geol 40:1084–1087
21. Saxena VK, Ahmed S (2003) Inferring the chemical parameters for the dissolution of fluoride in groundwater. J Environ Geol 43:731–736
22. Singh DK, Singh AK (2002) Groundwater situation in India: Problems and perspective. Int J Water Resour Dev 18(4):563–580
23. Subba Rao N, Krishna Rao G, John Devados D (1998) Variation of fluoride in groundwater of crystalline terrain. J Environ Hydrol 6:1–5
24. Subba Rao N, Rao AT (2003) Fluoride in ground water in a developing area of Guntur District, Andhra Pradesh India. J Appl Geochem 5:94–100
25. Subba Rao N (1998) Groundwater quality in crystalline terrain of Guntur District, Andhra Pradesh, India. Visakha Sci J 2:51–54
26. Subramani T, Elango L, Damodarasamy SR (2005) Groundwater quality and its suitability for drinking and agricultural use in Chithar River Basin, Tamil Nadu, India. Environ Geol 47(8):1099–1110
27. Sujatha D, Reddy BR (2003) Quality characterization of groundwater in the south-eastern part of the Ranga Reddy district, Andhra Pradesh, India. Environ Geol 44(5):579–586
28. Susheela AK (1999) Fluorosis management programmer in India. Curr Sci 77:1250–1256
29. Till Y, Zhu WH (1991) Environmental characteristics of regional groundwater in relation to fluoride poisoning in North China. Environ Geol Water Sci 18:3–10
30. WHO (1993) Guide lines for drinking water quality, vol 1, 2nd edn. Recommendations, Geneva
31. WHO (1996) Guidelines for drinking water quality. 2nd edn. Health criteria and supporting information Geneva, pp 231–237
32. WHO (2004) Guidelines for drinking water quality. 3rd edn. Geneva
33. WHO (1984) Guidelines for drinking water quality, vol 2. Health criteria and supporting information Geneva, pp 100–105
34. Wodeyar BK, Srinivasan G (1996) Occurrence of fluoride in the groundwaters and its impact in Peddavankahalka basin, Bellary District, Karnataka, India—a preliminary study. Curr Sci 70:71–74

A Technical Review on Image Super-Resolution Techniques

Anugya Shukla, Suresh Merugu and Kamal Jain

Abstract Conversion from multiple low resolution (LR) images to high resolution (HR) image is done by using super-resolution techniques. Anyone can achieve more information in detail from high-resolution images, which helps further for many satellite image applications. This growing technology interest in the reconstruction of imagery leads to several methodologies in the field of advanced digital color image processing. The study presents and describes the various conventional algorithms of image SR reconstruction used to date by researchers. In this review paper, we have shown different types of super-resolution techniques starting from image noisy to remotely sensed imagery for super-resolution mapping at a sub-pixel level with their characteristics and major limitations. A precise comparative study was done on different domains like spatial and frequency as shown separately in tabular form.

Keywords Super resolution (SR) · High resolution (HR) · Low resolution (LR) · Image processing · Sub-pixel classification

1 Introduction

In the present scenario, high-resolution images are desired for various requirements for remotely sensed satellite imagery, Medical imagery, Military, Forensic, etc. Basically, high resolution means a number of narrow bands in a dataset that provides complete detail or information. Since 1970, complementary metal-oxide-semiconductor sensors (CMOS) and charge-coupled device (CCD) were used to capture the imagery in digital format. The present resolution of these sensors provide high resolution but not satisfy the consumer price and demand, with the help of SR techniques, the present LR imagery can be converted to high resolution with consideration of present need and demand.

A. Shukla · K. Jain
Indian Institute of Technology Roorkee, Roorkee, Uttarakhand, India
e-mail: anugya.shukla92@gmail.com

S. Merugu (✉)
CMR College of Engineering & Technology, Hyderabad, Telangana, India
e-mail: msuresh@cmrcet.org

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_54

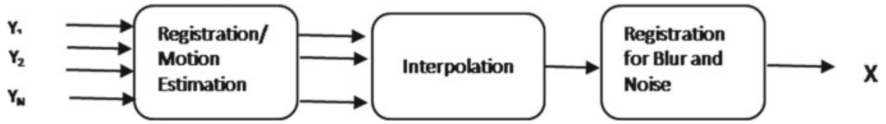


Fig. 1 Basic super-resolution model

The common approach used for increasing the resolution is to reduce the pixel per unit area, which turns into failure as the number of available light decreases and generates shot noise, which depletes the quality of the image. Another approach was adopted in 1980 for refining the spatial resolution by increasing the size of integrated chip (IC), which increased the capacitance too [1] and decrease the charge transfer rate. Various approaches used on spatial and frequency domains to overcome these issues, which are discussed in image processing.

The main advantage of the super-resolution algorithm is that it provides HR images in a cost-effective manner, and the present LR model can still be employed. Figure 1 shows the basic model of super-resolution. The SR image reconstruction has wide application in the field of medical science used for medical image processing (CT and MRI), Video applications, Surveillance, satellite imagery in field of Remote sensing, Forensic science, Digital Video Recorder (DVR). There are five sections in this paper. Section 2 covers the frequency-domain techniques of super-resolution reconstruction. Section 3 explains most of the methods of spatial domain, for example, interpolation briefly, Iterative backpropagation, Set-theoretic and many more. Super-resolution mapping is covered under Sect. 4, where various methods of SR mapping are explained.

2 Frequency Domain Methods

Tsai and Hung in 1984 [2] proposed that HR imagery can be generated from the sequence of multiple LR imagery, with possible sub-pixel motion between frames obtained; thus, the unique direction is required for each frame. This model is based on:

- Fourier Transformation with shift property.
- Image aliasing effect between the Discrete Fourier Transformation (DFT) and Continuous Fourier Transforms (CFT).
- It assumes that information is band-limited in the original scene.

Tasi and Huang's work was extended by Kim and Bose in 1990, where they considered noise and blur in the image and formulated with weight least square method, and the assumption was made that all LR images have the same noise and blur.

Rhee and Khang in 1999 [3], proposed a discrete cosine transformation (DCT) based model, which has reduced memory and computational cost compared to DFT.

The advantage of the frequency domain is computationally easier due to frequency domain hardware implementation complexity with less cost [4].

Let a scene denoted by $f(x, y)$, which is a continuous high-resolution image. $F_r(u, v)$ be its continuous Fourier transformation and then the global translation that produce rth shift as

$$F_r(x, y) = f(x + \Delta x_r, y + \Delta y_r) \tag{1}$$

where $r = 1, 2, 3, \ldots, R$.

$$CFT = F_r(u, v) \tag{2}$$

$$F_r(u, v) = F(u, v) \tag{3}$$

If $F_r(x, y)$ is the shifted image sampled at T_1 and T_2 the time period to generate LR image

$$Y_r[t_1, t_2] = f(mT_1 + \Delta x_r, nT_2 + \Delta y_r) \tag{4}$$

where, $m = 0, 1, 2, \ldots, M - 1$, and $n = 0, 1, \ldots, N - 1$.

Now DFT of these LR images can be represented by $Kr[m, n]$

$$\mathrm{K_r[m, n]} = \frac{1}{\mathrm{T_x T_y}} \sum_{\mathrm{n_1 - 1}}^{\mathrm{L_1 - 1}} \mathrm{F} * \left(\frac{2\pi}{\mathrm{T_x}} \left(\frac{\Omega_1}{\mathrm{N_1}} + \mathrm{n_1} \right), \frac{2\pi}{\mathrm{T_y}} \left(\frac{\Omega_2}{\mathrm{N_2}} + \mathrm{n_2} \right) \right)$$

Assuming $F(x, y)$ band limited above equation can be written as matrix-vector as

$$Y = \Phi F \tag{5}$$

where Y is represented as a column vector of a shifted image including DFT and F is a column matrix of unknown continuous Fourier transformation of $f(x, y)$.

Y is a matrix showing the relationship between Y and F, i.e., discrete Fourier transformation of the observed LR imagery are sampled to continuous HR images.

3 Spatial Domain Methods

In the spatial domain method, the spatial property is for the reconstruction of images. The effect of a blur, i.e. motion and optical blur, non-ideal sampling, Point Spread Function and more, are used for reconstruction of images [5]. A glance at comparative analysis of all conventional techniques of super-resolution is mentioned in Table 1.

Table 1 Comparative analysis of image reconstruction algorithms

S. No.	Domain	Literature	Method	Blur and noise	Registration parameter	Prior constrain	Advantage	Disadvantage
1	Frequency	Tsai and Huang [2]	Relative motion	Not considered	Not considered	Not considered	The earliest method to eliminate aliasing artifact due to ideal under sampling	Considered hypothetical (impulse) sampling
2		Kim, Bose and Valenzuela [46]	Recursive total least square	Not considered	Considered	Included	Improvement of SNR ratio	No registration parameter
3		Kim and Su [47]	Weighted least square	Diff for each image	Not considered	Not included	Updated regularization parameter, Incorporation of the new frame without repletion of the computation process	Prior constraints and registration parameter are not considered

(continued)

Table 1 (continued)

S. No.	Domain	Literature	Method	Blur and noise	Registration parameter	Prior constrain	Advantage	Disadvantage
4		Rhee and Khang [3]	Discrete cosine transformation	Reduced	Not considered	Included	Reduction in distortion caused by in-accurate motion in reconstructed image and memory requirement, and the computational cost was also reduced	Due to no data correlation in the frequency domain, the spatial domain prior knowledge was difficult to apply
5	Spatial	Ur and Gross [6]	Non-linear interpolation	Same for all images	Considered	Included	Less complex and low computational time	The degradation model is limited
6		Irani and Peleg [7]	IBP		Considered	Not included	Easy in understanding and simple approach	Approach don't provide a unique solution
7		N. Kbose, S. Lertrattanapanich and J. Koo [39]	Deterministic	Reduced	Considered	Included	Provide a unique solution, and computational cost is less	A large number of Low-Resolution images are required for a better solution

(continued)

Table 1 (continued)

S. No.	Domain	Literature	Method	Blur and noise	Registration parameter	Prior constrain	Advantage	Disadvantage
		InYadav et al. [53]		Reduced	Considered	Included	Less complex reduces the noise, preserve edges, and less computational time	It still suffers from artifacts
		X. Yang, T. Liu and D. Zhou (2015)		Reduced	Considered	Included	Better edge preservation.	It still suffered from the noise.
		Shen et al. [8]		Reduced	Considered	Included	Balanced result for edge preservation and smoothing	It uses the L2 norm that has an optimum solution for white Gaussian distribution
		Kiani and Drummond [9]		Reduced	Considered	Included	Less computational cost and better results.	Not suitable for most practical situations.

(continued)

Table 1 (continued)

S. No.	Domain	Literature	Method	Blur and noise	Registration parameter	Prior constrain	Advantage	Disadvantage
8		Richar R Schultz and Robert L. Stevenson [23]	Stochastic (i) Bayesian	Assumed independent (noise-free)	Considered	Included	It is very simple to understand, and it reduces ambiguity and noise	Not suitable for nonlinear functions
		Tom and Katsaggelos [10]	(ii) Maximum likelihood	Reduced and included for each frame	Considered	Not included	Estimated the sub-pixel shift, noise variance, and high-resolution image together	Not able to define the position of low-resolution pixel w.r.t. to the HR grid
		Hardie et al. [11]	(iii) MAP joint	Reduced and block matching technique	Considered	Included	Applicable to sophisticated observation and degradation model	Conjugate gradient technique not applicable and with a change in registration technique cost increases in the present algorithm
		Zeng and Lu [12]	(iv) Tikhonov	Reduced	Considered	Included	Reduces noise.	Loss of information from image regions
		Yuan et al. [13]	(v) SWTV	Considered	Considered	Included	Better in artifacts reduction and edge preservation	It causes an exchange among edge details during preservation

(continued)

Table 1 (continued)

S. No.	Domain	Literature	Method	Blur and noise	Registration parameter	Prior constrain	Advantage	Disadvantage
		Chen et al. [14]	(vi) Lmix	Considered	Considered	Included	Better noise removal and edge preservation	The shortcomings of the global TV model
		Zhang et al. [15]	(vii) MAP-MRF	Considered	Considered	Included	Better in blur image case	Produce insufficient results
		Abedi and Kabir [16]	(viii) SWDTV	Considered	Considered	Included	Smmothifies the image.	Unable to reconstruct edges of text details and has a high computation cost.
		Huang et al. [17]	(ix) GVFHF-ADSF	Considered	Considered	Included	Better in edge preservation	It has a high computational cost
9			Set-theoretic	Not considered	Considered	Included	Simplicity and Flexible in the spatial domain	Slow convergence and high computational cost
		Stark and Oskaui [18]	POC					
		Tekalp and Oskoui (1992)		Included	Considered	Included		
10		Zeng et al. [19]	Hybrid	Included	Considered	Included	Better in edge preservation and fine details.	Unable to differentiate between the edges and noises.

(continued)

Table 1 (continued)

S. No.	Domain	Literature	Method	Blur and noise	Registration parameter	Prior constrain	Advantage	Disadvantage
		Ghosh et al. [20]		Included	Considered	Included	It is simple and robust.	The computational resource is limited
		Long et al. [21]		Included	Considered	Included	Poor edge preservation. Loss of details	It reduces noise
11		Elad and Feuer [22]	Motionless SR reconstruction	Included	Considered	Included	Combination of ML, MAP, and POCS, Less number of LR images required for reconstruction	The generalized method does not work for special cases
		M. V. Joshi and S. Chaudhri (2007)		Included	Considered	Included	Used zoom factor for increasing resolution	Used convectional technique Bilinear interpolation technique

Let x_r be the multiple LR images where $r \in \{1, 2, \ldots, R\}$ (Images are written in lexicographically order vector) z is the original image.

Then, x_r maybe written as

$$x_r = H_r z \tag{6}$$

where H_r is a matrix containing motion compensation, degradation effect, and subsampling.

$$X = H_y z + N \tag{7}$$

X is an observation equation $\left[x_1^T \ldots x_R^T\right]^T$ and $H = \left[H_1^T \ldots H_R^T\right]^T$.

N represents the observation noise.

3.1 Non-linear Interpolation Method

Non-linear interpolation comprises of three phases.

- Registration
- Interpolation
- Restoration for blur and noise

Restoration performed by a deconvolution method that includes the noise. Low complexity and computational time are one of the major advantages of non-uniform interpolation, but this approach is only applicable to uniform blur and noise characteristics for all LR images.

Ur and Gross in 1992 [6] reduced the blur and noise effect, increased the spatial resolution with multiple shifted LR images spatially, and applied the sampling theorem with multichannel of Pauli's method, which further modified by Brown. A real-time image registration technique for infrared images and SR reconstruction was proposed by Hardie.

3.2 Nearest Neighbor

This technique assigns a gray value (DN) to a new output pixel location (i, j) according to the DN value of the new output pixel (i, j) DN value of a spatially nearest pixel. This method preserves original values of the pixel, and this doesn't alter the thematic information. Due to its simplicity, it doesn't create an anti-aliasing effect. The result of nearest-neighbor doesn't have sub-pixel accuracy and result in discontinuities when arbitrary rotation and scale is done.

3.3 Bicubic Convolution Method

The result of non-linear interpolation lacks high-frequency information; it was improved in bilinear interpolation. It uses information from the original pixel and 16 surrounding pixels to determine the DN value of the new pixel created from an original pixel. Bi-cubic is the enhancement of nearest-neighbor interpolation. The resultant image is much smoother than of bilinear interpolation method. It also alters the DN value and requires more computational time, and it provides improved image quality.

3.4 Iterative Back Propagation Method

In 1993 Irani proposed a different approach for super-resolution restoration problems based on IBP, which was suitable for computer-supported tomography where 2D objects are reconstructed from its 1D projection. In this method, the difference is treated as a residual error and emulated in the next iteration for HR image calculation. In this algorithm uniform translation and rotation of pixel, displacements are assumed for the entire image. The decrease in de-blur was observed when algorithm tested on a single image without increasing the sampling rate. This process is repeated until some convergence is achieved.

The contribution of the error in the reconstructed image at each iteration was determined by the backpropagation kernel. The major drawback of this approach is that it does not provide any unique solution, and no prior constraints are present.

IBP produce HR image by the back projection of error between ith LR image $X(i)$ and observed LR image by BP operation H^{BP}.

$$\widehat{Z}(i+1) = \widehat{Z}(i) + H^{BP}\Big(X - \widehat{X}(i)\Big) \tag{8}$$

3.5 Regularized Method

3.5.1 Deterministic

The deterministic approach solves the inverse problem by using prior information about the solution to a problem [1].

$$\Big[||X_K - W_{Ky}||^2 + \alpha||C_y||^2\Big],$$

where C is represented as a high pass filter, α represents langrage's multiplex on increasing the value of α solution becomes smoother. This method is fissile with less number of LR images while increasing the number of images amount of noise decreases with small α and lead to a good solution.

$$\left[\sum_{k=1}^{p}\left(W_K^T W_K + \alpha C^T C\right)\right]\hat{y} = \left[\sum_{k=1}^{p}\left(W_K^T X_K\right)\right] \tag{9}$$

$$\hat{y}_{n+1} = \hat{y}_n + \beta\left[\sum_{k=1}^{p}\left(W_K^T\left(X_K - W_K\hat{y}_n\right) - \alpha C^T C\hat{y}_n\right)\right] \tag{10}$$

where β represents the convergence parameter. For a better optimum value of regularization parameter using L-curve, Bose pointed out the registration parameter and proper constrained least square (CLS) reconstruction.

3.5.2 Stochastic

(a) ***Bayesians Method***

The Bayesian approach uses Probability Density Function (PDF) of original imagery. The MAP estimation of x maximizes the posterior Probability Density Function $P(Y|X_k)$ with respect to y [11]

$$x = argmax\, P(y|x_1, x_2, \ldots, x_n) \tag{11}$$

Applying Bayes theorem to condition probability then MAP optimization

$$x = argmax\ P\{y|x_1, x_2, \ldots, x_n|y| + \ln P(x)\} \tag{12}$$

$P(y) = priori\ image\ model$
$P\{y|x_1, x_2, \ldots, x_n|y| \ln P(x)\} = conditional\ density$

.

In order to compute the MAP estimate, the conditional density $P(z)$ must be defined

$$P(Y = y) = 1/z\ \exp\{-1/\lambda \Sigma_{c\epsilon s} V_C(y)\} \tag{13}$$

MAP estimate contains prior knowledge $P(y)$
Z is the normalizing condition
λ temperature parameter of density
$Vc(y)$ function derivative of an image.

The main advantages of the Bayesians algorithm provide edge-preserving and image expansion prior to the algorithm. The Bayesian approach is not suitable for nonlinear function because the result becomes smoothed and edge preservation is not possible [23].

$$y = \text{argmin}\left\{ ||X - W_k||2 + \frac{1}{\lambda}\lambda \Sigma_{c\epsilon s} V_C(Y) \right\} \tag{14}$$

When error and noise assume to be independently estimated, it becomes equal to MAP.

(b) ***Maximum Likelihood***
In 1997 Maximum Likelihood technique developed to estimate transitional motion and High-resolution image, i.e., HR images are obtained from multiple noisy, blurred and un-sampled images with registration parameters associated with each frame in which Point Spread Function was assumed to be unknown, but this was not successful as it fails to locate the position of every obtained LR pixel with respect to HR grid.

(c) ***MAP Joint Method***
MAP uses Huber-Markov random field for image prior model and their motion calculation using the block-matching technique. Here SR image is derived from three phases' registration, restoration and image interpolation [11].

$$Z, s = \arg\min_{z,s}\{-\log[Pr(x|z,s)] - \log[Pr(z)] - \log[Pr(s)]\} \tag{15}$$

Here, we have to specify $Pr(z)$= prior image density
$Pr(s)$ = prior motion density
$Pr(x|z,s)$ = conditional density.

(d) ***Maximum Posterior Method***

In 1996 Bayesian developed the maximum posterior algorithm in which a single image frame and image expansion were performed. The maximum posterior is also known as Tikhonov-Arsenium Regularization. It considered the special case of MRF priors in the Bayesian framework [5].

3.6 Set-Theoretic Method

3.6.1 Projection into Convex (POCS)

In set-theoretic methods, the POCS algorithm is the most popular method because of its simplicity and powerful usability in spatial domain observation. The prior constrains about the solution to the reconstruction process. POCS method was first suggested by Stark and Oskaui [18] which has further extended by Tekalp et al. [5]. The prior knowledge, flexibility in spatial domain and simplicity add its advantage, but due to non-uniqueness of solution and dependence on initial values, degrade this method. Additionally, slow convergence and high computational cost make it less fissile.

3.6.2 Bounding Ellipsoid

Tom and Kalsaggelous [10] found a similar technique to POCS in which they modify the constraint set by using an ellipsoid to bound constrain set in which centroid is taken as a super solution estimate. As direct implementation is infeasible, so the iterative procedure was used [22].

3.6.3 Hybrid Methods

In order to reduce the ill effect of various methods ML, MAP and POCS, the combination among them is taken for SR reconstruction. For example, the ML-POCS method in which the simplicity of ML and non-ellipsoid constrain used in POCS are utilized to form a new hybrid method. The advantage of this hybrid was that prior knowledge provides a single solution in contrast to POCS [24].

3.7 Filtering Method

The adaptive filter approach was used in 1999 by M. Elad. The observation model was modified to remove the time dependence. In this approach least square (LS) estimator was based on the pseudo-PLS algorithm. This technique lacked the inclusion of prior constraints as compared to POCS and Bayesian.

3.8 Motionless SR Reconstruction Method

All SR reconstruction techniques mentioned above uses relative sub-pixel motion between observed images, but the motionless reconstruction technique uses different blurred images without motion. In 2001 similar technique was proposed using intensity and depth map using the MRF model; other techniques uses zoom was proposed by Joshi and Chaudhari.

3.9 Other Techniques

There are many more techniques for SR reconstruction; for example, a single frame image SR reconstruction was done using single LR observation. Freeman in 2002 [25], proposed an example-based SR technique. Kwok-Wai Hung prepares interpolation between HR image and LR image. Van-Cutterd, Richardsu Lucy, scale space, blind de-convolution were developed.

4 SR Mapping in Classification

In order to generate HR images from LR images in the SR reconstruction technique, many classification methods are used. Wavelet-based classification is based on wavelet theory, which was developed from signal processing, comprises of wave and decaying amplitude in both negative and positive direction [26]. In this approach, a signal passed through the number of low pass and high pass filters to analyze in the frequency domain.

The image restoration process is affected by filtering technique, and then by applying down-sampling and up-sampling, this process was developed by Mallat in 1989 [27]. Pixel-based fusion was another classification technique in which registration and conventional transformation are performed on a multi-spectral LR image then replacing the first component of transformed LR image by HR panchromatic image. IHS, Borvey transformation, SWT, Gram Schmidt and more are some image fusion methods.

4.1 Neural Network

According to Chua and Tay [28] convolution neural network had better performance than other reconstruction methods, here training data was used to train the neural network, which makes it more robust for motion blur and noise. Convolution neural network consists of a hidden layer, feature extraction and output layer. The performance of the convolution neural network depends upon the process of rectifying the number of hidden layers and neurons.

Taten A. J., Lewis G. H., Aktinsion P. M. and Nixon M.S. in 2001 developed Super Resolution Hopfield neural network for remotely sensed images, they proposed solutions for network neurons considered to be arranged in a regular grid and referred the neurons by coordinate notation. Consider an example where example neuron (i, j), row i and Colum j of the grid. Figure 2 shows an example where it contains $2 * 2$ pixel image, p and q are image dimensions, x and y are image pixel, coordinates, i, and j represent neuron coordinate.

Fig. 2 Relation between (x, y) with network (i, j)

4.2 Linear Reconstruction Method for Super-Resolution

In super-resolution with a variable pixel, linear reconstruction combines the LR image in a linear non-uniform way. The algorithm defines a variable pixel linear reconstruction method proposed by Fruchter and Hook [29] called drizzle detailed. This technique weights input according to the statistical significance of each pixel and it also works on removing geometrical detector on both images. It has the advantage of observation effects. It has a huge memory as compared to handing noise, multi-temporal and others to others and capable of handling large LR images [30].

The major drawback of this approach is that it does not provide any unique solution, and no prior constraints are present.

4.3 Application of SR Mapping in Subpixel Classification

A pixel represents more than one class. The convectional image classification method (Hard classification) classifies only one class and not able to distinguish mixed pixels from a pure pixel.

4.3.1 Land Cover Mapping Using Multiple Sub-pixel Shifts

The solution of a sub-pixel problem was considered as soft classification for remotely sensed images [31, 32]) as soft classification do not assign single class feature but also predicts fractional correlation of each ground cover class falls in every pixel also include linear spectral un-mixing modeling, Artificial neural network, Support vector machine and regression tree [4].

Fen and Ling in 2010 proposed SR mapping method with numerous sub-pixel shift method in which mixed pixel problem reduced in classification process where LR imagery are classified with various soft classification in order to obtain proportional images and these proportional imageries are to be used as inputs for further processing called super-resolution mapping to extract high resolution land cover class. In this research high overall accuracy and kappa coefficients are observed.

4.3.2 Sub-pixel Mapping Using Radial Basis Function Interpolation

This method was proposed by Qunming Wang in 2014 in which the RBF interpolation method was introduced in subpixel mapping to embark ground cover features for each and every sub-pixel [33].

In this paper, they determine the spatial relationship between a coarser pixel and mixed pixel, and coefficients of a coarser pixel are observed. To calculate mixed pixel fraction abundance value, they used weighted basis function value. This methodology then tested on three various types of remotely sensed data and compared with existing SPM methods.

4.3.3 Sub-pixel Mapping of Landcover Using Interpolation-Based SR

In 2015 Feng Ling proposed this method in which they classified the remote sensing images using soft classification for a generation of fine indicator maps of each land cover and combines the fine resolution indicator maps of all land cover classes to generate High-Resolution land cover. Consider an example under the land cover fraction of the raster image of 3 * 3 LR i = if shift in both images is by 0.5 [34], then the sub-pixel shift proves more useful Land cover information as shown in Fig. 3. In this paper, he used spline, IDW, and kriging interpolation methods. The work concluded that increasing the zoom factor performance SRM approach decreases, whereas the spline and kriging method generates high accuracy.

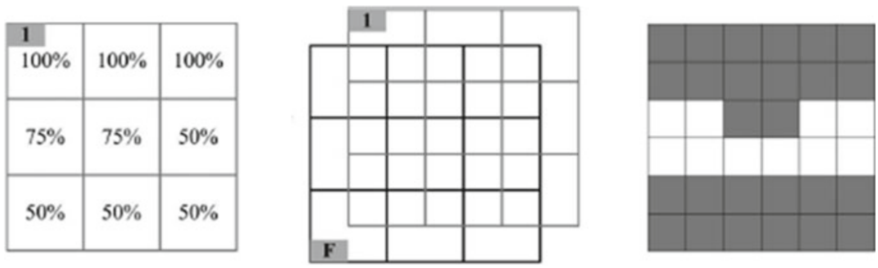


Fig. 3 Fraction of ground surface raster grid of 3 * 3 LR image mapping

4.3.4 Colorimetry Based Method for Sub-pixel Level Arrangements for Super-Resolution Mapping

Suresh and Jain [36], in remotely sensed imagery, each pixel records the detailed information of color spectra. The major contribution of this work is an allocation algorithm that allocates classes to subpixel in a non-random manner and developed an optimized algorithm for solving the problem of multiple and non-allocated sub-pixels. As the execution time of the super-resolution algorithm depends on the initial allocation of classes to subpixel. Depending upon sub-pixel level arrangements and which is to analyze the super-resolution mapping algorithm on different scale factors. Experimental results show better information extraction with sub-pixel level arrangements of high-resolution imagery, as shown in Table 2.

5 Advances and Future Perspective

There are various issues in SR reconstruction one of the major issue is registration. Therefore, an accurate registration method should be adopted based on robust motion models. Cannel adaptive regularization is used to lower down the registration error in this method more number of registration errors should contribute less in estimated HR image. Kang further expands this approach in terms of set-theoretic. Boese estimated error generated by inaccurate registration and found the complete least square approach to reduce the errors.

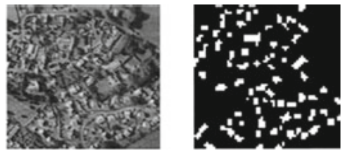
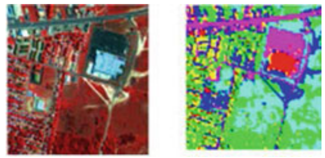
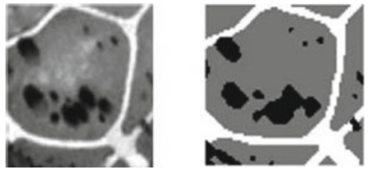
The blurring process is assumed to be known in many SR reconstruction techniques, but in some cases, it is unknown; this is called blind SR reconstruction. A blind multichannel method was developed by Wirawan using multiple-finite-impulse-filters.

The efficient and robust algorithm proposed by Nuygen for solving the Ticknov-Regularize SR problem with the accelerated CG (Conjugate Gradient) is proposed as circulant type pre-conditioners. These pre-conditioners can be easily regularized, and the operation can be performed easily by using FFT, and the number of CG iteration is dynamically reduced. Michael Elad, in 2001, proposed a method in which deblurring can be separated from the fusion process. They considered the blur the same for all measurements and invariant space, and additive noise is white. The method proved to be efficient and low computational cost.

6 Conclusion

This paper provides a brief review of different types of methodology used in super-resolution reconstruction in the field of digital image processing. Table 1 describes the various algorithms with their characteristics in spatial and frequency domain, and Table 2 explains the recent approaches of classification in super-resolution. The

Table 2 Super-resolution mapping using different classification algorithms

S. No.	Method	Literature	Algorithm	Features	Results
1	Super-resolution mapping using classification	A. J. Lewis, G. H. Atkinson and Nixon MS, 2002	Neural network	Robust, Efficient, and simple technique. Maps at any scale	
2		Xu et al. [35]	Active substitute wavelet	Works on Inherent structures of subpixel accuracy	
3		Feng Ling, Yun Du, Fei, Xiao, Huaiping Xue and Shengiun Wu, 2009	Multiple sub-pixel shifts	The sequence of sub-pixel shifted imagery is used for high-resolution accuracy	

(continued)

Table 2 (continued)

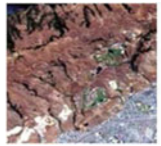
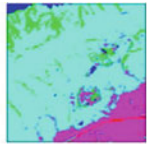
S. No.	Method	Literature	Algorithm	Features	Results
4		Qunming Wang, Wenzhong and Peter M. Atkinson [33]	Radial basis function Interpolation	Use RBF interpolation, No need prior information, no iteration required	 
5		Suresh and Jain [36]	Colorimetry	Chromaticity mapping of a multispectral imagery	 

Table 3 Comparative analysis between spatial and frequency domain

S. No.	Parameters	Frequency based SR	Spatial based SR
1	Motion	Global translation	Global and non-Global translation
2	Degradation model	LSI limited	Unlimited LSI
3	Accountability	Low	High
4	Implacability cost	Less	High
5	A priors info	Bandlimited	Unlimited
6	Hardware complexity	Less	Low
7	Application	Limited	Unlimited
8	Disadvantage	Lack of data correlation	Registration and blur error

overall conclusion can be drawn from Table 3, where the spatial domain has higher overall computational flexibility and improved results as compared to the frequency domain. The applicability of the spatial domain in classification mapping provides a robust solution for the super-resolution mapping of satellite imagery. The much work has been done in this field, but further improvement is required for extraction finer details from a low-resolution image.

References

1. Park SC, Park MK, Kang MG (2003) Super-resolution image reconstruction: a technical overview. Signal Process Mag 20(3):21–36
2. Tsai RY, Huang TS (1984) Multi-frame image restoration and registration. Adv Comput Vis Image Process 1(2):317–339
3. Rhee S, Kang MG (1999) Discrete cosine transform based regularized high-resolution image reconstruction algorithm. Opt Eng 38(8):1348–1356
4. Ling F et al (2010) Super-resolution land-cover mapping using multiple sub-pixel shifted remotely sensed images. Int J Remote Sens 31(19):5023–5040
5. Tekalp AM, Ozkan MK, Sezan MI (1992) High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration. In: 1992 IEEE international conference on acoustics, speech, and signal processing, 1992. ICASSP-92, vol 3. IEEE
6. Ur H, Gross D (1992) Improved resolution from sub-pixel shifted pictures. CVGIP Graph Models Image Process 54(2):181–186
7. Irani M, Peleg S (1993) Motion analysis forimage enhancement: resolution, occlusion, and transparency. J Vis Commun Image Represent 4(4):324–335
8. Shen H, Peng L, Yue L, Yuan Q, Zhang L (2016) Adaptive norm selection for regularized image restoration and super-resolution. IEEE Trans Cybern 46:1388–1399
9. Kiani KA, Drummond T (2017) Solving robust regularization problems using iteratively re-weighted least squares. In: 2017 IEEE winter conference on applications of computer vision (WACV). IEEE, pp 483–492
10. Tom BC, Katsaggelos AK (1995) Reconstruction of a high-resolution image by simultaneous registration, restoration, and interpolation of low-resolution images. In: Proceedings of 1995 IEEE international conference on image processing, vol 2, Washington, DC, Oct 1995, pp 539–542

11. Hardie RC, Barnard KJ, Armstrong EE (1997) Joint MAP registration and high-resolution image estimation using a sequence of under-sampled images. IEEE Trans Image Process 6(12):1621–1633
12. Zeng WL, Lu XB (2013) A robust variational approach to super-resolution with nonlocal TV regularisation term. Imag Sci J 61(2):268–278
13. Yuan Q, Zhang L, Shen H (2012) Hyperspectral image denoising employing a spectral–spatial adaptive total variation model. IEEE Trans Geosci Remote Sens 50(10):3660–3677
14. Chen C, Liang H, Zhao S, Lyu Z, Sarem M (2014) A novel reconstruction model for multi-frame super-resolution image based on lmix prior. Comput Electr Eng 40:142–153
15. Zhang D, Jodoin P-M, Li C, Wu Y, Cai G (2015) Novel graph cuts method for multi-frame super-resolution. IEEE Signal Process Lett 22:2279–2283
16. Abedi A, Kabir E (2016) Stroke width-based directional total variation regularisation for document image super-resolution. IET Image Proc 10:158–166
17. Huang S, Sun J, Yang Y, Fang Y, Lin P (2017) Multi-frame super-resolution reconstruction based on gradient vector flow hybrid field. IEEE Access 5:21669–21683
18. Stark H, Oskoui P (1989) High-resolution image recovery from image-plane arrays, usingconvex projections. JOSA A 6(11):1715–1726
19. Zeng W, Lu X, Fei S (2015) Image super-resolution employing a spatial adaptive prior model. Neurocomputing 162:218–233
20. Ghosh D, Kaabouch N, Hu W-C (2016) A robust iterative super-resolution mosaicking algorithm using an adaptive and directional Huber-Markov regularization. J Vis Commun Image Represent 40:98–110
21. Long W, Lu Y, Shen L, Xu Y (2017) High-resolution image reconstruction: an env‘1/TV model and a fixed-point proximity algorithm. Int J Numer Anal, Model, p 14
22. Elad M, Feuer A (1997) Restoration of a singlesuper-resolution image from several blurred, noisy, and under-sampled measured images. IEEE Trans Image Process 6(12):1646–1658
23. Schultz RR, Stevenson RL (1994) A Bayesian approach to image expansion for improved definition. IEEE Trans Image Process 3(3):233–242
24. Nguyen N, Milanfar P, Golub G (2001) A computationally efficient super-resolution image reconstruction algorithm. IEEE Trans Image Process 10(4):573–583
25. Freeman WT, Jones TR, Pasztor EC (2002) Example-based super-resolution. Comput Graph Appl IEEE 22(2):56–65
26. Young RK (1993) Wavelet theory and applications
27. Mallat SG (1989) A theory for multi-resolution signal decomposition: the wavelet representation. IEEE Trans Pattern Anal Mach Intell 11(7):674–693
28. Chua KK, Tay YH (2013) Enhanced image super-resolution technique using convolutional neural network. In: Advances in visual informatics. Springer International Publishing, 157–164
29. Fruchter AS, Hook RN (2002) Drizzle: a method for the linear reconstruction of under-sampled images. Publ Astron Soc Pac 114(792):144
30. Merino MT, Nunez J (2007) Super-resolution of remotely sensed images with variable-pixel linear reconstruction. IEEE Trans Geosci Remote Sens 45(5):1446–1457
31. Foody GM (1996) Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data. Int J Remote Sens 17(7):1317–1340
32. Liu J et al (2005) Spatial and temporal patterns of China’s cropland during 1990–2000: an analysis based on Landsat TM data. Remote Sens Environ 98(4):442–456
33. Wang Q, Shi W, Atkinson PM (2014) Sub-pixel mapping of remote sensing images based on radial basis function interpolation. ISPRS J Photogramm Remote Sens 92:1–15
34. Ling F et al (2013) Interpolation-based super-resolution land cover mapping. Remote Sens Lett 4(7):629–638
35. Xu X, Zhong Y, Zhang L (2014) Adaptive subpixel mapping based on a multiagent system for remote-sensing imagery. IEEE Trans Geosci Remote Sens 52(2):787–804
36. Suresh M, Jain K (2017) Subpixel level mapping of remotely sensed imagery to extract fractional abundances using colorimetry. J Remote Sens Spat Sci. https://doi.org/10.1016/j.ejrs.2017.02.004

Bibliography

37. Bose NK, Kim HC, Valenzuela HM (1993) Recursive total least squares algorithm for image reconstruction from noisy, under-sampled frames. Multidimens Syst Signal Process 4(3):253–268
38. Bose NK, Kim HC, Zhou B (1994) Performance analysis of the TLS algorithm for image reconstruction from a sequence of under-sampled noisy and blurred frames. In: Image processing, 1994. Proceedings. ICIP-94. IEEE international conference, vol 3. IEEE
39. Bose NK, Lertrattanapanich S, Koo J (2001) Advances in superresolution using L-curve. In: The 2001 IEEE international symposium on circuits and systems, ISCAS 2001, vol 2. IEEE (2001)
40. Elad M, Yacov H-O (2001) A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur. IEEE Trans Image Process 10(8):1187–1193
41. Glasner D, Bagon S, Irani M (2009) Super-resolution from a single image. In: 2009 IEEE 12th international conference on computer vision. IEEE
42. Hung K-W, Siu W-C (2012) Single image super-resolution using iterative Wiener filter. In: 2012 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE
43. Irani M, Peleg S (1991) Improving resolution by image registration. CVGIP Graph Models Image Process 53(3):231–239
44. Joshi MV, Chaudhuri S, Panuganti R (2004) Super-resolution imaging: use of zoom as a cue. Image Vis Comput 22(14):1185–1196
45. Malczewski K (2009) Super resolution for multimedia, image, and video processing applications. In: Studies in computational intelligence
46. Kim SP, Bose NK, Valenzuela HM (1990) Recursive reconstruction of high resolution image from noisy undersampled multiframes. IEEE Trans Acoust Speech Signal Process 38(6):1013–1027
47. Kim SP, Su W-Y (1993) Recursive high-resolution reconstruction of blurred multi-frame images. IEEE Trans Image Process 2(4):534–539
48. Merugu S, Rai AK, Jain K (2015) Sub pixel level arrangement of spatial dependences to improve classification accuracy. In: 2015 international conference on advances in computing, communications and informatics (ICACCI). IEEE
49. Tatem AJ et al (2002) Super-resolution land cover pattern prediction using a Hopfield neural network. Remote Sens Environ 79(1):1–14
50. Tekalp AM (1995) Digital video processing. Prentice-Hall, Inc.
51. Tom BC, Katsaggelos AK (1995) Reconstruction of a high-resolution image by simultaneous registration, restoration, and interpolation of low-resolution images. In: Proceedings of international conference on image processing, vol 2. IEEE
52. Verhoeye J, De Wulf R (2002) Land cover mapping at sub-pixel scales using linear optimization techniques. Remote Sens Environ 79(1):96–104
53. Yadav M, Garg R, Bhavsar A (2014) Better guiding the guided range image filter for range-image super-resolution. In: Proceedings of the 2014 Indian conference on computer vision graphics and image processing. ACM, p. 62
54. Yang J, Huang T (2010) Image super-resolution: historical overview and future challenges. Super-Resolut Imag, 20–34
55. Duhamel P, Maitre H (1999) Multi-channel high resolution blind image restoration. In: Proceedings, 1999 IEEE international conference on acoustics, speech, and signal processing, 1999, vol 6. IEEE
56. Yang S et al (2002) The design of scaffolds for use in tissue engineering. Part II. Rapid prototyping techniques. Tissue Eng 8(1):1–11
57. Zhong Y et al (2015) An adaptive subpixel mapping method based on MAP model and class determination strategy for hyperspectral remote sensing imagery. IEEE Trans Geosci Remote Sens 53(3):1411–1426

Augmented Reality on Sudoku Puzzle Using Computer Vision and Deep Learning

Azhar Talha Syed, Suresh Merugu and Vijaya Kumar

Abstract Many people try solving Sudoku puzzles every day. These puzzles are usually found in newspapers, magazines and so on. Whenever a person is unable to solve a puzzle or is running short on time to solve the puzzle, it will be very convenient to show the solved puzzle as an augmented reality. Objectives: In this paper, proposed an optimal way of recognizing a Sudoku puzzle using computer vision and Deep Learning, and solve the puzzle using constraint programming and backtracking algorithm to display the solved puzzle as augmented reality. Also, a comparative performance analysis with the previous work is shown and provided at the end of this paper. Methods: In order to implement augmented reality on to the Sudoku puzzle, image classification itself won't be sufficient as the solved puzzle has to be shown on top of the area of the unsolved puzzle in the original image. So puzzle detection has to be performed and for doing so proposed work used CNN and Object Localization algorithms. After the detection this should store the values detected in each 9×9 cells and ran a constraint programming and backtracking algorithm to solve the puzzle and finally filled the detected empty cells with correct values of the solved puzzle. Applications/Improvements: Usually the Sudoku puzzles that will find in newspapers and magazines are surrounded by a lot of noise such as text (characters) irrelevant to the puzzle and borders of the newspaper which could be similar to a Sudoku puzzle structure. In this paper it emphasize on how to handle such disturbances and improve the performance.

Keywords Object localization · Sudoku · CNN · Augmented reality · Computer vision

A. T. Syed · S. Merugu (✉) · V. Kumar
Department of Computer Science and Engineering, CMR College of Engineering & Technology, Hyderabad, Telangana, India
e-mail: msuresh@cmrcet.org

A. T. Syed
e-mail: syedazhartalha@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_55

1 Introduction

Ever since the breakthrough of AlexNet in the imagenet challenge, CNN's have become one of the most prominent algorithms in image classification problems. However, just classifying whether or not the image contains a Sudoku puzzle is not enough. The bounding box of the Sudoku puzzle has to be identified and in order to do so, we used Object Localization [1]. The reason for not using other algorithms such as Selective Search [2], YOLO [3] or Region Proposals [4] for finding the bounding box is, all these algorithms are computationally expensive when compared to Object Localization. This is another important performance criteria as the number of frames per second has to be taken into consideration for a better AR experience, as the computation increases the number of frames per second decreases. Moreover, these algorithms are very effective when detecting multiple objects in a frame and undergo additional computation, such as in the case of selective search a support vector machine is trained to find the regions of interest [2].

Where as in our problem we are only trying to detect one class which is the Sudoku board so Object Localization is a good solution for this problem as the cost of finding the bounding box is very less and it works well for detecting a single object in a frame [1]. After finding the bounding box of the Sudoku board we iterate over each of the 9×9 cells, on each cell classification using CNN is performed to find the digits in the cell. For training this CNN we used the MNIST dataset [5] and achieved 99.67% accuracy on the test data. The CNN used for performing Object Localization has architecture similar to AlexNet [6]. The architectures and the training methodologies used for training Object Localization for detecting the Sudoku board and CNN for classifying digits will be further illustrated later in Sect. 3.

This paper is organized as follows. In Sect. 2, it describes the previous work done on solving Sudoku using augmented reality. In Sect. 3, an illustration of steps involved to solve the problem is provided along with a description of methodologies used in training Deep Neural Networks used in the solution. In Sect. 4, an analysis of solution is provided and results. Finally, in Sect. 5, a conclusion on the work done is provided with a suggested solution.

2 Related Work

In the year 2012, Pramod J Simha, Suraj K v, and Tejas Ahobala, proposed Recognition of Numbers and Position using Image Processing Techniques for Solving Sudoku Puzzles [7]. For detecting the Sudoku Puzzle they made an assumption that the largest contour in the frame is the Sudoku Puzzle. However, this could not be true every time as there could be larger contours in a frame other than the Sudoku Puzzle. After finding the largest contour they proposed to split the contour into 81 parts, each part representing a cell as a Sudoku contains 9×9 cells. On each cell a template matching algorithm was used using template data to find the digits in each

Fig. 1 A typical Sudoku Puzzle found in daily newspaper

cell. After finding all the values in each the Sudoku was solved using backtracking (Fig. 1).

Snigdha Kamal, Simarpreet Singh Chawla, Nidhi Goel proposed, "Detection of Sudoku Puzzle using Image Processing and Solving by Backtracking, Simulated Annealing and Genetic Algorithms: A Comparative Analysis" [8] in 2015. For detecting the Sudoku Puzzle these authors used an algorithm similar to the one proposed by Pramod J Simha, Suraj K v, and Tejas Ahobala, which is assuming that the largest contour in the frame is the sudoku puzzle. After finding the puzzle, OCR was performed on each cell and the values corresponding to each cell were stored. Three algorithms were used to solve the Sudoku which were Backtracking, Simulated Annealing and Genetic Algorithm. The results of these algorithms were compared with different difficulties of the Sudoku problems and their performance analysis was made based upon the execution time that each problem took to be solved.

In 2005, Helmut Simonis, proposed Sudoku as a Constraint Problem [9].

A common way of solving Sudoku is by using Backtracking algorithm, which is basically a blind search algorithm similar to Depth First Search Algorithm (DFS). As these blind searches can take exponentially large time in some cases, Helmut proposed a better algorithm which doesn't involve search. He viewed Sudoku as a constraint problem which means choosing an answer based on some set of constraints. This will really boost the speed of solving the problem. But however, the solution will not work for all the problems and in such cases Backtracking has to be used again. For detecting the Sudoku puzzle, A. Van Horn proposed the use of Hough transformation to detect the puzzle in his literature, "Extraction of sudoku puzzles using the Hough transform" [10]. The same algorithm was used by Baptiste Wicht and Jean Hennebert in their work, "Mixed handwritten and printed digit recognition in Sudoku with Convolutional Deep Belief Network" [11]. After detecting the Sudoku Puzzle, Baptiste Wicht and Jean Hennebert used Convolutional Deep Belief Network to classify the digits. Keeping the previous works in mind we have worked on improvising the existing techniques especially in the case of Sudoku Puzzle detection and also provided a full solution for the problem.

3 Methodology

For solving this problem we have divided it into four sub-problems.

They are:

- Detecting the Sudoku Puzzle.
- Classifying and storing the values in each cell.
- Detecting the Sudoku Puzzle.
- Classifying and storing the values in each cell.
- Detection of Sudoku Puzzle (Fig. 2).

The objective here is to find the bounding box of the Sudoku Puzzle in a frame. One way of doing this is finding the largest contour in the frame and assuming it to be the puzzle, calculating the areas of all the contours and treating the contour with the largest area as the puzzle. However, in real life there are a lot of disturbances in a frame, it is also possible that Sudoku puzzle is not even present in the frame. For

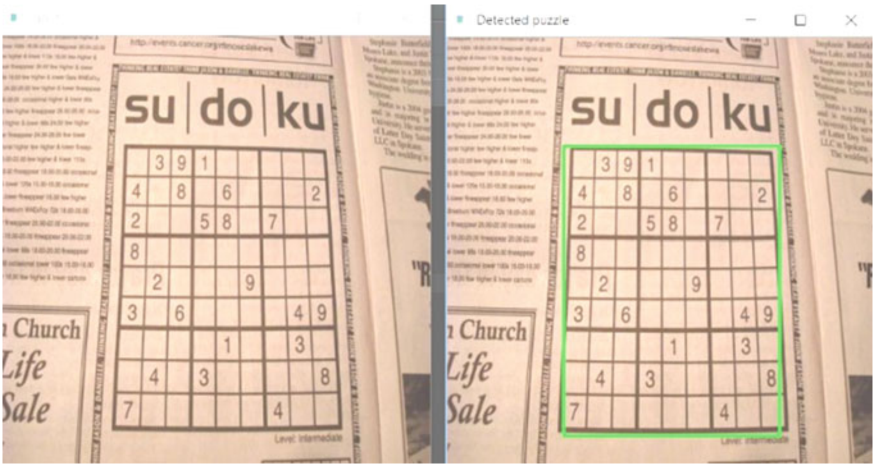


Fig. 2 Objective of the problem, (left) input frame, (right) detected Sudoku Puzzle

example in Fig. 1, the puzzle is surrounded by another contour, which is the border of the paper, and if we use the same detection algorithm the border of the paper will be detected as the puzzle because it has a larger area compared to the puzzle. So this particular approach is not very effective and a better algorithm is needed for detecting the puzzle.

Sliding Window, in this algorithm a fixed set of sizes of windows are chosen and starting from the top left corner of the frame, part of the frame is chosen along with the size equal to the size of the window. This part of the frame is given as input to a classifier to know whether or not it is a Sudoku. Same is done for the whole frame with a fixed stride. Different window sizes are applied and the process is done again [12]. However, the algorithm is not very effective for this problem as it is hard to predict the exact window size which is equal to the size of the Sudoku Puzzle because precision is very important for implementing Augmented Reality on the same detected area. Other detection algorithms such as Selective Search combined with CNN, YOLO, and Region Proposals combined with

CNN are good detection algorithms but involve heavy computation which will in turn lead to reduction in number of frames per second. In Fig. 3 have plotted the average frames per second in a one minute window for detecting the puzzle (not solving). Object localization has outperformed the other algorithms as it has only one CNN which is the same for classifying and finding the bounding box as well [1]. Selective Search, YOLO, Region Proposals were proposed for detecting multiple objects in a frame [2–4], but in our problem we are only trying to detect one object. So, there is no need of extra computation which is required to detect multiple

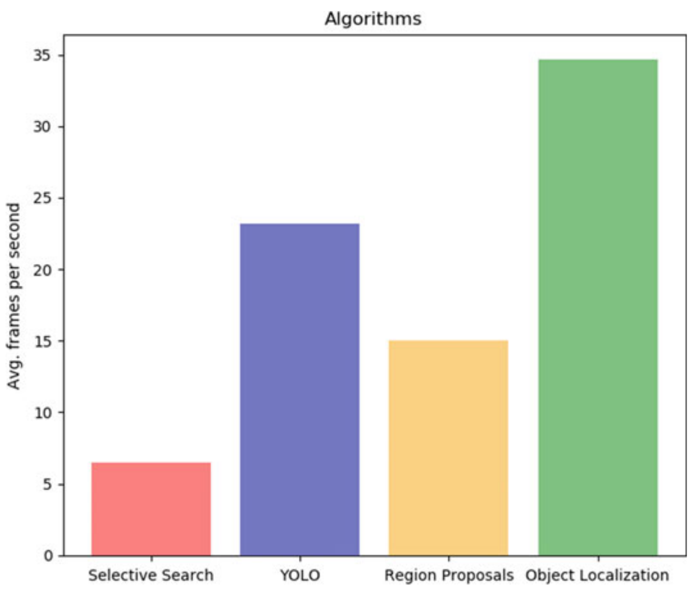


Fig. 3 Average frames per second for different detection algorithms

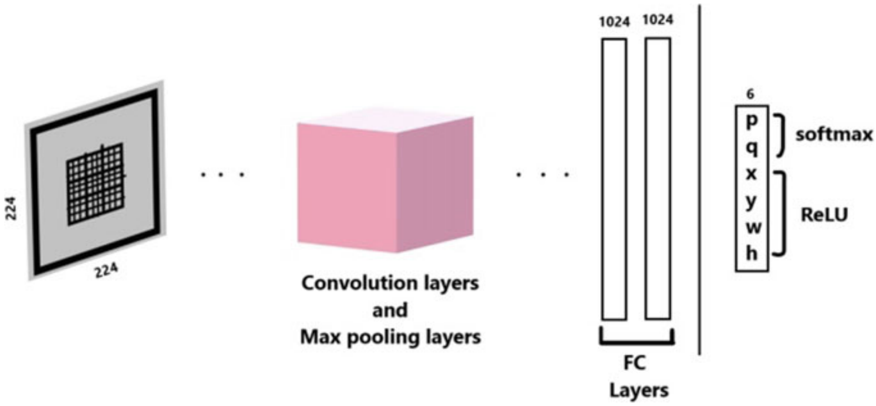


Fig. 4 Architecture of CNN used for Object Localization

objects. We will further describe the architecture and the training methodologies used for Object Localization as we notice it performing well over other algorithms in our empirical data for this problem [13–15] (Fig. 4).

First the frame is resized to 224 × 224 pixels. Note that a pixel represents a real value x_{ij} where x_{ij}s [0, 1]. As the image is converted to grayscale and the values between [0, 255] are normalized to real values between [0, 1].

This resized frame is passed as input to a CNN with following architecture.

Convolution Layer: Input—a tensor of dimensions (224 × 224 × 1); number of filters—96; kernel dimensions—(11 × 11 × 1); Padding—0; stride—4;
Max Pooling Layer: Input—a tensor of dimensions (54 × 54 × 96); Pool size—(2 × 2 × 1); strides—(2 × 2 × 1);
Convolution Layer: Input—a tensor of dimensions (27 × 27 × 96); number of filters—256; kernel dimensions—(5 × 5 × 96); Padding—0; stride—1;
Max Pooling Layer: Input—a tensor of dimensions (23 × 23 × 256); Pool size—(3 × 3 × 1); strides—(2 × 2 × 1);
Convolution Layer: Input—a tensor of dimensions (11 × 11 × 256); number of filters—384; kernel dimensions—(3 × 3 × 256); Padding—0; stride—1;
Convolution Layer: Input—a tensor of dimensions (9 × 9 × 384); number of filters—384; kernel dimensions—(3 × 3 × 384); Padding—0; stride—1;
Convolution Layer: Input—a tensor of dimensions (7 × 7 × 384); number of filters—256; kernel dimensions—(3 × 3 × 384); Padding—0; stride—1;
Max Pooling Layer: Input—a tensor of dimensions (5 × 5 × 256); Pool size—(2 × 2 × 1); strides—(2 × 2 × 1);
Fully Connected Layer: Input—a vector of length 1024; output neurons—1024;
Fully Connected Layer: Input—a vector of length 1024; output neurons—1024;
Output Layer: input—a vector of length 1024; softmax neurons—2; ReLU neurons—4; output neurons—6;

The architecture is similar to AlexNet [6]. However, there are some major changes in the fully connected layers and output layer. AlexNet contains 4096 neurons in fully connected layers, whereas, we used 1024 neurons in the fully connected layers and by doing so there is a total reduction of 18 million parameters, which is a great performance boost with respect to computation. And also AlexNet was designed for 1000 class classification problem but we are concerned with only a single class which is the Sudoku Puzzle along with four other regression outputs. We noticed an accuracy of 98.3 on the test data when using 4096 neurons in the fully connected layers and an accuracy of 98.2 with 1024 neurons in the output layer. The reason for having such a less difference in the test data accuracy could be over fitting. As the number of parameters increase there is a chance that the model is over fitting [16], however, weight decay (L2) regularization was used in both cases.

The difference in the accuracies is very less but the differences in the computation are extremely large so we decided to use 1024 neurons in the fully connected layers. Speaking of output layer, it outputs 6 values. p, q, x, y, w, h. Where p is the probability that the Sudoku Puzzle is in the frame and q is the probability that the sudoku puzzle is not in the frame. So, in an ideal case p should be 1 and q should be 0 if the puzzle is in the frame, if not p should be 0 and q should be 1. The values x, y, w, h represent the bounding box. Where (x, y) are coordinates of the top left corner of the bounding box, values w and h represent width and height of it [17, 18]. For all the convolution layers and fully connected layers ReLU is used as the activation function. In the output layer, values p and q have Softmax activation and values x, y, w, h have ReLU activation. This is because for values p and q we expect a probability where the value of $p + q$ is always equal to 1, it can be achieved using Softmax activation over p and Values x, y, w, h are always non negative integers. ReLU always outputs a non negative real number, so ReLU is a good activation function for predicting these values. Some other hyper parameters used while training:

Initialization [19] was used for all the kernels.

Weight Decay (L2 regularization) with lambda value 1e−5.

Adam [9] was used as the optimizer with a learning rate of 0.001.

Batch Normalization after every convolution layer.

Digit Recognition on Each Cell

Now that we have the bounding box of the Sudoku Puzzle, we cropped the puzzle from the original frame and resized it to 252×252 images. Before we went further some image processing had to be performed. Which is removing the margins of the cells, as the margins of the cells are straight lines we used Hough Line detection to detect the margins and remove them. By doing the same frame will be similar to left image as shown in Fig. 5.

There are a total of 9×9 cells in Sudoku, as we are solving 9×9 standard size Sudoku puzzles. So after resizing the image each cell will be of 28×28 pixels. These 28×28 pixels will be given as an input to a CNN which will classify if the cell is empty or which of the 9 digits (sudoku doesn't contain does the cell contain. The classified values are stored for further solving the problem. The CNN used for this sub-problem has the following structure.

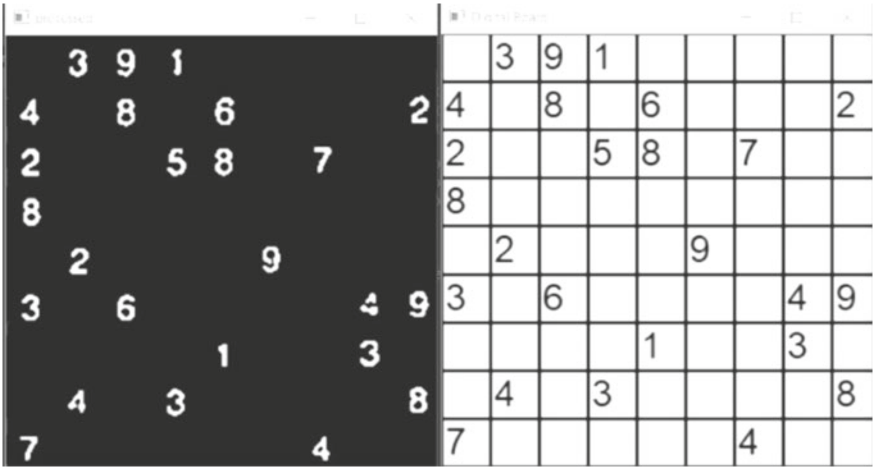


Fig. 5 (left) Processed frame, (right) classified cells

Convolution Layer: Input—a tensor of dimensions (28 × 28 × 1); number of filters—64; kernel dimensions—(3 × 3 × 1); Padding—1; stride—1;
Max Pooling Layer: Input—a tensor of dimensions (28 × 28 × 64); Pool size—(2 × 2 × 1); strides—(2 × 2 × 1);
Convolution Layer: Input—a tensor of dimensions (14 × 14 × 64); number of filters—128; kernel dimensions—(3 × 3 × 64); Padding—1; stride—1;
Max Pooling Layer: Input—a tensor of dimensions (14 × 14 × 128); Pool size—(2 × 2 × 1); strides—(2 × 2 × 1);
Convolution Layer: Input—a tensor of dimensions (7 × 7 × 128); number of filters—256; kernel dimensions—(3 × 3 × 128); Padding—1; stride—1;
Convolution Layer: Input—a tensor of dimensions (7 × 7 × 256); number of filters—256; kernel dimensions—(3 × 3 × 256); Padding—1; stride—1;
Max Pooling Layer: Input—a tensor of dimensions (7 × 7 × 256); Pool size—(2 × 2 × 1); strides—(2 × 2 × 1);
Fully Connected Layer: Input—a vector of length 4096; output neurons—1024;
Fully Connected Layer: Input—a vector of length 1024; output neurons—1024;
Softmax Layer: input—a vector of length 1024; output neurons—10;
Training strategies used:
ReLU activation for all the layers except pooling and Softmax Layer.

Dropout with rate = 0.5. Batch normalization after last fully connected layer. Initialization for all the kernels. Adam optimizer with learning rate 0.001. Solving the puzzle (Fig. 6).

By the time we reach this step, the complete Sudoku puzzle is detected and values in each cell are loaded into the memory. Many solutions have been proposed to solve Sudoku the simplest and the most popular one being the backtracking. In backtracking all the possible values are tested in each empty cell until there are no

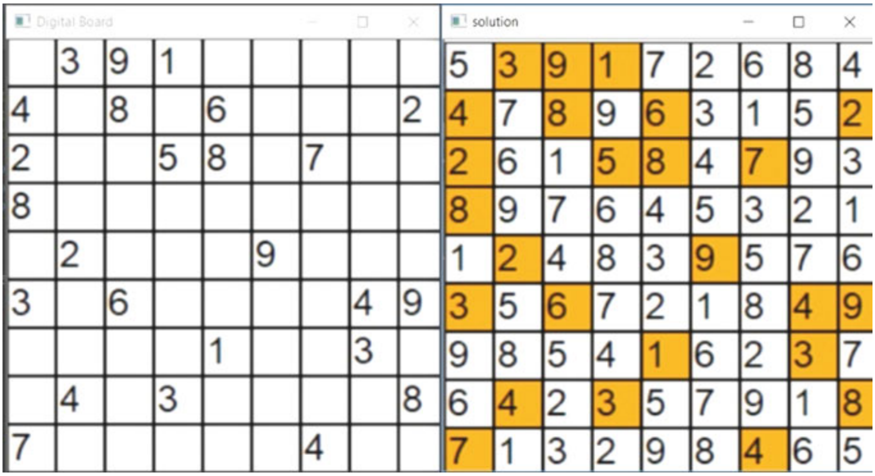


Fig. 6 (left) Classified cells, (right) solved puzzle

other possibilities left, if the solution is not yet reached then the next possibility of the previous empty cell is checked. Another algorithm for solving Sudoku is Constraint programming. This method was proposed by Helmut Simonis where he proposed solving Sudoku as a constraint problem [9]. However, the algorithm might not work for some difficult problems so backtracking has to be used in such cases. So a combination of both algorithms will result in a faster solution. An implementation of such algorithm is provided by Rhollor [20]. Which is the same algorithm used for solving this problem?

Drawing the Solution onto the Original Frame

At this point we have the solution for the puzzle and in order to fill the solution in the empty cells and give an AR experience. The x, y, w, h values obtained during Object Localization will be used again. If a digit belonging to cell c_{ij} is to be drawn onto the frame, then the pixel on which we need to start drawing is

$$(x + ((x + w)/9) * i + 1, y + ((y + h)/9) * j + 1)$$

The same is done for all the empty cells and finally our goal is achieved. By the end of all the four steps the frame will look as the output in Fig. 7.

Training Data

The data set used for training CNN for object localization was created using 200 images of Sudoku Puzzle from news- papers and magazines and 1000 images which do not contain the Puzzle. For each image 6 labels are attached p, q, x, y, w, h which will be used by loss function during the training process. Generating this data was very expensive. We have written a program which will provide the x, y, w, h values when a box is drawn around the puzzle using the mouse. The same had to be done

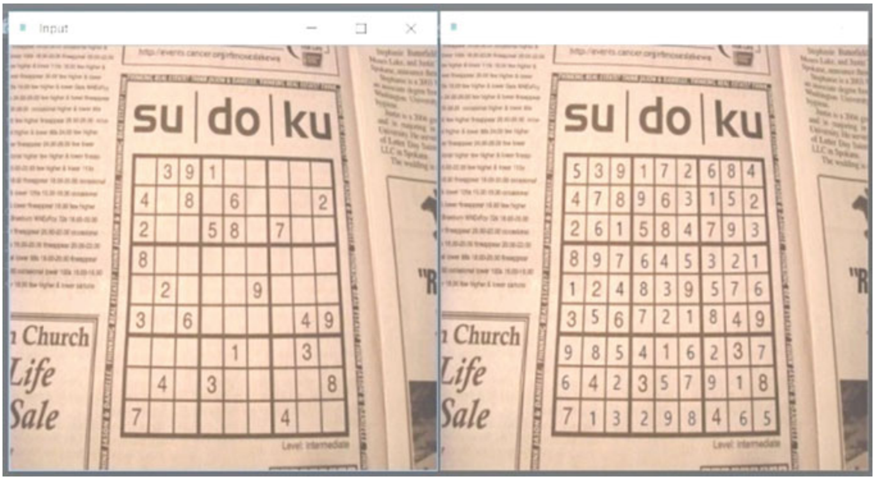


Fig. 7 (left) Original frame, (right) solved puzzle as Augmented Reality

for every image. And also we wanted to use the data for training, so we opted for Data augmentation with the following techniques.

Rotation—we rotated the y—axis between [−15, 15] degrees arbitrary and generated 100 more images.

Translation—By translation of x and y axis another 400 images were generated.

Gaussian noise—By adding Gaussian noise with a standard deviation 1 another 100 images were added.

The CNN used for classification of digits is trained using MNIST dataset, as it does not contain empty cells all the zero labeled images are made as blank images. Because the digit '0' is not used in Sudoku but in there are very high chances that the cell could be empty. So these '0' digit images were replaced with blank images in the MNIST dataset for training. No data augmentation was performed here as the MNIST dataset is a very rich dataset [5] and we achieved 99.67% accuracy on the test data set.

4 Results

The CNN used for Object Localization has 99.33 accuracy in correctly classifying the test dataset and an average of 0.87 Intersection over Union (IoU) for finding the bounding box. Whereas, CNN used for classifying digits has an accuracy of 99.67 in correctly classifying the digits. Overall the performance of our solution in terms of frames per second is as following. For an easy level of difficulty it was approximately 26 frames per second, for intermediate level of difficulty it was 20–22 frames per second and for hard level of difficulty 13–16 frames per second were noticed. Also an analysis of each step involved in solving the problem was provided along with exploring the other possible solutions and discussing their drawbacks.

5 Conclusion

A full fledged system for showing the solution of a Sudoku puzzle as Augmented Reality was proposed in the paper. Using image processing, object localization, image classification, constraint programming and backtracking. Figure 7. Shows how the solution is shown on top of the original frame by the end of all the steps. Addressed some substantial improvements to the previous methodologies used for the problem set. Such as, using Object Localization for puzzle detection and the Convolutional Neural Network to be used for implementing it, classifying digits from the puzzle, solving Sudoku and finally filling the solution in the empty cells precisely.

Compliance with Ethical Standards Both the authors declares that he/she has no conflict of interest and have written this after reviewing various research papers and then given a easy solution to solve the Puzzles.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

Appendix

This Paper was implemented by Azhar Talha Syed under the supervision and guidance of Dr. Suresh Merugu. The implementation of this work was done in Python using Tensorflow and OpenCV libraries at AI&DL Centre of Excellence Centre, CMR College of Engineering & Technology, Hyderabad, Telangana, India.

References

1. Andrew Ng. Lecture on "Object Localization" in the course "Convolutional Neural Networks". [Online]. Available: https://www.youtube.com/watch?v=GSwYGkTfOKk
2. Uijlings JRR, van de Sande KEA, Gevers T, Smeulders AWM (2013 Sept) Selective search for object recognition. Int J Comput Vis 104(2):154–171
3. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: The IEEE conference on computer vision and pattern recognition (CVPR), pp 779–788
4. Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in neural information processing systems 28 (NIPS 2015)
5. The MNIST dataset. By LeCun Y, Cortes C, Burges CJC [Online]. Available: http://yann.lecun.com/exdb/mnist/

6. Krizhevsky A, Sutskever I, Hinton GE, ImageNet classification with deep convolutional neural networks. In: Advances in neural information processing systems, vol 25(2). [Online]. Available: https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf
7. Simha PJ, Suraj KV, Ahobala T (2012) Recognition of numbers and position using image processing techniques for solving Sudoku Puzzles. In: IEEE-international conference on advances in engineering, science and management (ICAESM-2012)
8. Kamal S, Chawla SS, Goel N, Detection of sudoku puzzle using image processing and solving by back-tracking, simulated annealing and genetic algorithms: a comparative analysis. In: 2015 third international conference on image information processing (ICIIP)
9. Simonis H, Sudoku as constraint problem. Paper presented at the Eleventh international conference on principles and practice of constraint programming
10. Van Horn (2012) Extraction of sudoku puzzles using the hough transform. University of Kansas, Department of Electrical Engineering and Computer Science, Tech. Rep.
11. Wicht B, Hennebert J, Mixed handwritten and printed digit recognition in sudoku with convolutional deep belief network. In: 2015 13th international conference on document analysis and recognition (ICDAR)
12. Andrew Ng. Lecture on "Object detection" in the course "Convolutional Neural Networks". [Online]. Available: https://www.youtube.com/watch?v=5e5pjeojznk
13. Merugu S, Reddy MCS, Goyal E, Piplani L (2019) Text message classification using supervised machine learning algorithms. In: Kumar A, Mozar S (eds) ICCCE 2018. Lecture Notes in Electrical Engineering, vol 500. Springer, Singapore. ISSN 1876-1100
14. Yadav BC, Merugu S, Jain K (2019) Error assessment of fundamental matrix parameters. In: Kumar A, Mozar S (eds) ICCCE 2018. Lecture Notes in Electrical Engineering, vol 500. Springer, Singapore. ISSN 1876-1100
15. Suresh M, Jain K (2015) Semantic driven automated image processing using the concept of colorimetry. In: Second international symposium on computer vision and the internet (VisionNet'15), Procedia Computer Science, vol 58, pp 453–460
16. Nielsen M, in his book Neural networks and deep learning—Chapter 3: Overfitting and regularization
17. Sharma SK, Jain K, Suresh M (2019) Quantitative evaluation of panorama softwares. In: Kumar A, Mozar S (eds) ICCCE 2018. Lecture Notes in Electrical Engineering, vol 500. Springer, Singapore, ISSN 1876-1100
18. Suresh M, Jain K, Colorimetry-based edge preservation approach for color image enhancement. J Appl Remote Sens (SPIE)
19. He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: The IEEE international conference on computer vision (ICCV), pp 1026–1034
20. Rhollor, Sudokusolver. GitHub. Rhollor. Retrieved 8 Dec 2016

Fuzzy Inference System Based Assessment of Pollution Aspects in Hyderabad

Musini Venkateshwarlu and Suresh Merugu

Abstract Urbanization and industrialization rapid growth of Hyderabad and surrounding area led to air pollution. The air pollution is mainly from industries and emission from transportation segment like suspended particulate matter (SPM), monoxide and dioxides of C, S, N and fuel burning. The air pollution is high effect the health of individuals such as lung diseases and respiratory problem. This paper deals with study on our pollution and health hazards and remedial action to minimize this. To understand the concentration of pollutants three different zones residential, industrial and commercial areas sample are collected. To control pollution methods to be adopted and also mentioned.

Keywords Air pollution · Health damages · Mitigating activities · Air quality parameters

1 Introduction

The world pollution is an adaptation of the Latin word 'Pollutionene', meaning defilement, from pollute, to soil or defile. In the later part of the 19th century the word pollution has been used to signify contamination of the environments [1]. According to Edwards [2] pollution is the release of substances or energy to the environment by man and industry in quantities that damage either his health or resources. Globally environmental deterioration is the major issue due to green house effect reduction in ozone layer CO_2 increase fluorocarbon into atmosphere further deteoried environment growth in population and urbanization followed by industrialization added pollutants to the echo system lead to air pollution due to use of hydrocarbons in transportation and industries.

M. Venkateshwarlu (✉)
Department of Civil Engineering, CMR College of Engineering & Technology, Kandlakoya (V), Medchal Road, Medchal District, Hyderabad 501401, Telangana, India
e-mail: venkatmusini@gmail.com

S. Merugu
Research &Development (R&D) Centre, CMR College of Engineering & Technology, Kandlakoya (V), Medchal Road, Medchal District, Hyderabad 501401, Telangana, India

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_56

Hyderabad, sixth largest city in India and capital of Telangana state with an ancient civilization on and culture with Hussain Sagar Lake is an artificial lake connecting Hyderabad and Secunderabad part of Deccan Plateau 1776 ft above sea level spread over 100 Sq.Miles.

The new city growth is mainly due to industrialization and commercialization based on modern technical know-how and a good transport network system. Hyderabad city has become strategically very significant after the war with China in 1962, because of its central location. The Government of India has established a huge military base at Dindigal (Hakimpet) with the establishments of defence organization like DRDL, DRO, Aircrafts Pilot Training Centre, etc., A large number of government, public and private sector have come up like BHEL, NFC, ECIL, HMT, PRAGA TOOLS, BHARATH DYNAMICS, Republic Forge, Allwyn, SETWIN and a number of pharmaceutical, etc., are established for the promotion of commercialization of business activities for its economic growth. This rapid urbanization and increased industrialization has attracted a large chunk of rural population for employment opportunities and better standard of living. Social and Cultural changes, are attributed to distinct changes of socio-economic conditions of the people and increased land values resulted in the form of morphological structural changes of the city.

Due to presence of undesirable elements in the air due to pollution lead to health problem such as asthma, lung problems. The main causes of the air pollution are industrialization without any control and treatment, heavy increase in the number of automobiles and their emissions, increase use of chemicals and petrochemicals, rapid growth of population, fast rate of deforestation and led to environmental pollution. It has been reported in most industrial towns and metropolitans of India and abroad such as Delhi, Mumbai, Chennai, Kolkata, Hyderabad, Kanpur, Firozabad, London, New York, Tokyo, Pittsburg etc.

2 Sources of Air Pollution

The sources of air pollution are of two types.

(a) Natural sources.
(b) Anthropogenic sources.

Natural sources include windblown dust, ash and gases, smoke and trace gases from forest fires. Anthropogenic sources over agricultural operations, transportation systems, combustion of fuels, process industries etc.

3 Classification of Air Pollutants

Air pollutants are classified into two groups

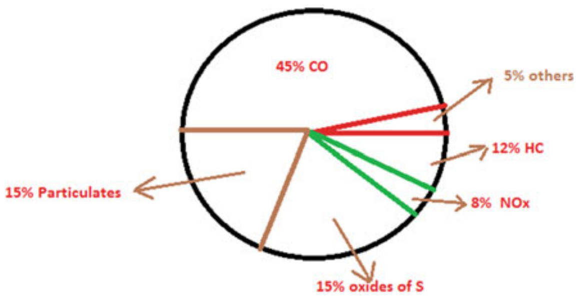


Fig. 1 Pollutants in the atmosphere

1. Primary air pollutants and
2. Secondary air pollutants.

4 Primary Pollutants

Pollutants that come directly into the atmosphere from the source of generation are called primary pollutants.

The major primary pollutants are

(a) CO
(b) Particulates
(c) Oxides of sulphur
(d) Hydrocarbons and
(e) Oxides of nitrogen.

In Fig. 1 distribution of various pollutants are given.

5 Secondary Pollutants

Secondary pollutants are generated in atmosphere due to chemical interaction of primary contaminants with oxygen, nitrogen, CO, etc. In the air. Secondary pollutants constitute ozone, aldehydes, PAN (Peroxy acetyl nitrate), photochemical smog, acid mists, sulphur trioxide, and nitrogen dioxide.

6 Sampling and Analysis of Air Pollutants

Twin cities (Hyderabad and Secunderabad) of Telangana are growing at a faster due to increased industrial activity, vehicular usage and urbanization. Over last 20 years air pollution has worsened due to NO_2, CO emissions, SPM, and SO_2.

Air pollution is due to usage of carbon compounds for cooking heating and transportation in cities. Intensity of air pollution increased due to concentration of population due to urbanization and industrialization concentrating and around cities going to deteriorate further in next 20 years. Therefore the need is felt to assess the ambient air quality and its seasonal variation in different zones, so that based on the knowledge careful planning of the future industrial and residential areas can be envisaged.

There sampling stations are selected in twin cities in such a way that they are located in there different zones namely residential, industrial and commercial to cover spatial variations various types of industries are located in the industrial estate of Uppal and is also densely populated Sanjeeva Reddy Nagar is one of the residential areas in twin cities having many colleges and tutorials with moderate traffic and homes. Abids is a busy commercial are with busy traffic and frequent idle-runs of vehicles at traffic signals.

7 Air Quality Parameters

For ascertaining the quality of ambient air, three major air quality parameters are selected.

I. SPM
II. SO_2
III. NO_X

And monitoring is carried out for a period of three months during summer, monsoon and winter covering all seasons of 1997 to observe the effects of seasonal changes. The frequency of sampling for SPM, SO_2 and NO_X is four times a month in each season for each sampling station and the sampling is carried for a period of 24 h during each sampling.

To collect samples for SO_2, SPM and NO_X envirotech model with gaseous sampling attachment is used and samples are collected round the clock during each sampling. SO_2 and NO_X are sampled on eight hourly basis to get three samples in a day in sequential order by drawing air at a flow rate of one litre per minute (LPM) through Sodium tetra chloro mercurate and sodium hydroxide absorbing solution respectively.

Spume is collected on eight hourly basis for 24 h in order by drawing air at a flow rate of 1.1 m^3 in through glass micro fibber filter paper and estimated by gravimetric method using the formula

$$\begin{aligned}\mathrm{SPM}\left(\mu g/m^3\right) &= \frac{\text{Wt diff in grams} \times 10^6}{\text{Avg air flow rate in } m^3/\text{min} \times \text{time in Min}} \\ &= \frac{W_2 - W_1}{V} \times 10^6\end{aligned}$$

where

W_1 = Initial weight of filter in grams.
W_2 = Final weight of filter in grams.
V = Total volume of air in m^3.
10^6 = Conversion factor from grams to μg.

The meteorological date for Hyderabad and Secunderabad is collected from IMD (India meteorological Department) for the above seasons.

8 Effects of Air Pollution

It has been found that air pollutants effecting human health, vegetation, animals and property. Some of the effects are physico-chemical and physiological.

Visibility is effected corrosion in metals change on material characters in turn effected life system and extent of damage depends upon the time of exposure and nature and concentration of contaminant [3].

Most of the effects of polluted air on man are related to ailments of respiratory tracts causing irritation [4]. Asthma, bronchitis, lung cancer etc. In every major air pollution episode, the combination of sulphur oxides and particulates has been implicated as cause of the excess mortality observed. Industry, cement dusts, photochemical smog cause injury, partially or complete destruction of vegetation, reduction of growth and ultimately crop yield, SO_2, NO_2 ammonia, fluorides, O_3Cl_2, PAN are responsible for destruction, dying, burning of leaf tissues (NECROSIS) and bleaching of green chlorophyll (CHLOROSIS) [5].

9 Control of Environmental Air Pollution

Some of the methods are suggested to control various types of pollutions.

- Controlling or confirming the pollutants at source and methods are suggested to do the above.
- Dilution of the pollutants in the atmosphere to permissible levels before they can reach the receptor.
- To mitigate the industrial pollution "green belt" is effective and economical method. It remains effective for considerable period of time without much recurring expenditure as compared to other pollution control measure [6]. Therefore in every industrial area provision should be made for allocating space for developing green belt by selecting proper trees namely *Albizzia lebbek*, *Azadirachta*, *Azadirachta indica*, *Tectona grandis*, *Terminalia arjuna*, *Cassia fistula*, *Polylthia longfolia* etc. Having high AIR pollution tolerance index (APTI) value for minimizing environmental pollution caused by gaseous pollutants [7].
- Environmental pollutions caused by industrial effluents can be reduced in two ways: (a) Raw materials and process must be selected carefully so that the quantity

of effluents should be harmless. (b) The effluents must be analyzed with the max. Allowable conc. (MAC) led by "Pollution Control Board". If the conc. given to the effluents before being disposed off.

- It has been found that by starting a common effluent treatment plants in industrial estate, pollution can be reduced to a greater extent. "The Jeedimetla common effluent treatment plant" in Hyderabad is the best example.
- Environmental education at all levels of non formal and formal education should be giving very much importance.
- Catalytic convertors should be made compulsory for all petro driven vehicles and dieselized vehicles must have a particle trap.
- To encourage the introduction of modern technology in vehicles, the price differential between diesel and petrol should be minimized as diesel is greater sources of pollutant releases. As both petrol and diesel are used mainly for public use (to the extent of 70–80%), it is not desirable to allow one to be more expensive than the other: this only prevents modern technology in the I.C. engines from coming to India [8].
- Enforcing check on heavy vehicle's emission by respective authorities and up gradation of fuel quality.

Details of sampling stations:

S. No.	Name of the sampling station	Name of the activity
1	S.R. Nagar	Residential
2	Uppal	Industrial
3	Abids	Commercial cum traffic

Effect of Materials:

Pollutant	Types of effect	Materials effected
Particulates	Abrasion and corrosion caused when combined with gaseous pollutants	Metals, paints and textiles
Sulphur oxides	Corrosion, fading of colors	Steed, zinc, electrical equipment, lime stone, statues, textiles, leather
	Electrochemical deterioration	Iron, Al, Cu, silver, building materials, leather, paper textiles
Nitrogen oxides	Cause dyes to fade and whites to turn yellow	Textiles

10 Statistical Analysis of Air Pollutants: SPM, SO_2, NO_X

Sampling station: Uppal (Industrial Estate)

Parameter: SPM ($\mu g/m^3$)

Season	No. of samples	Minimum	Concentration maximum	Mean
Summer	12	510	710	605
Monsoon	12	315	630	460
Winter	12	300	680	490

Parameter: SO_2 ($\mu g/m^3$)

Season	No. of samples	Minimum	Concentration maximum	Mean
Summer	12	25	60	37.9
Monsoon	12	20	60.8	40.4
Winter	12	35.72	60.75	48.3

Parameter: NO_X ($\mu g/m^3$)

Season	No. of samples	Minimum	Concentration maximum	Mean
Summer	12	45.2	200.5	122.84
Monsoon	12	40	180.2	110.1
Winter	12	74.25	218.11	146.1

Sampling Station: S.R. Nagar (Residential)

Parameter: SPM

Season	No. of samples	Minimum	Concentration maximum	Mean
Summer	12	190	470	330
Monsoon	12	150	505	328
Winter	12	170	530	350

Parameter: SO_2

Season	No. of samples	Minimum	Concentration maximum	Mean
Summer	12	18.5	30.5	24.5
Monsoon	12	15.0	29.5	26.6
Winter	12	20	37.2	29.1

Parameter: NO_X

Season	No. of samples	Minimum	Concentration maximum	Mean
Summer	12	24	80.2	50.2
Monsoon	12	31.4	84.5	55.1
Winter	12	35.0	74.2	54.5

Sampling Station: Koti (Commercial cum Traffic)

Parameter: SPM

Season	No. of samples	Minimum	Concentration maximum	Mean
Summer	12	290	610	450
Monsoon	12	280	615	448
Winter	12	370	510	440

Parameter: SO_2

Season	No. of samples	Minimum	Concentration maximum	Mean
Summer	12	28.98	45.96	39.0
Monsoon	12	20.43	37.23	29.88
Winter	12	30.35	45.14	38.01

Parameter: NO_X

Season	No. of samples	Minimum	Concentration maximum	Mean
Summer	12	25.07	186.35	95.5
Monsoon	12	40.74	146.93	89.64
Winter	12	85.71	221.87	142.01

11 Conclusion and Results

From the observation it is found that all the monitoring stations are affected by one or more poll-ting activities. Some of the results are

(1) Study zones, indicate air borne dust pollution.
(2) Winter is critical for maximum account of gaseous pollution and summer for air borne dust pollution.
(3) High incidence of NO_X at high volume traffic points contrary signal.

Acknowledgements We express our thanks to Ch. Gopal Reddy, Secretary, CMRGI, Principal and HOD, CMR College of Engineering and Technology Hyderabad, for help and encouragement to publish this paper.

References

1. Warren Charles E (1971) Biology and water pollution control
2. Edwards Peter (1972) Cultured red alga to measure pollution. Marine Pollution Bulletin 3(12):184–188
3. T.T.T.I Courses material—environmental pollution
4. FARMCO (Agriculture based magazine), Dept. of Farm Science & Rural Development, Loyola Academy, SEC-BAD, vol 12 (Aug 1995)
5. Indian J Environ Pollut 17(10) (1991 Oct)
6. Murty MN, Kumar Surender (2003) Environmental and economic accounting for industry. Oxford University Press, New Delhi
7. Ostro B et al (1995) Air pollution and mortality: results from Santiago, Chile. Policy Research Department, Working Paper 1453. World Bank, Washington, DC
8. Murty MN, James AJ, Misra S (1999) Economics of water pollution: the indian experience. Oxford University Press, New Delhi

Support Vector Machines: Introduction and the Dual Formulation

Rashmi Pathak

Abstract Machine Learning means "learning through Machines". We make machine learn and predict the behavior in order to find a solution to a problem. Machine is set to make various predictions based on the learning mechanisms that have been incorporated in them. There are various techniques through which machines can learn. Learning can be supervised, Unsupervised or Semi-supervised. Under these learning schemes we have various classifications. Under Supervised learning we have Classification and Regression. Classification works on continuous values and Regression works on discrete values. Support Vector Machine is an efficient classifier which are mostly sort of linear and comes under supervised method of learning. SVM also find its application in real life for Face Detection, Bioinformatics, Handwriting recognition, image classification and many others.

Keywords SVM · Dual formulation · Support vector machines · Classifiers · Machine learning

1 Introduction

Support Vector Machines is one of the most effective classifiers among those which are sort of linear [1]. We are also able to handle particular cases of non-linearity using non-linear basis function. Support Vector Machines have been one of the most popular classifier in recent times. This is because SVMs have clever way to prevent overfitting and we can work with large number of features at a time without getting involved in any complications.

Before getting started on SVM, basic hands on in logistic regression is required. Below are the basics for the logistic regression before proceeding with SVMs.

- $P(y = 1 \mid x) = h(x) = \partial(\beta^T x)$.
- We predict 1 when $h(x) > 0.5$ and 0 when $h(x) < 0.5$.

R. Pathak (✉)
Reckitt Benckiser Healthcare India Private Limited, Hyderabad, India
e-mail: Rashmi.pathak3012@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_57

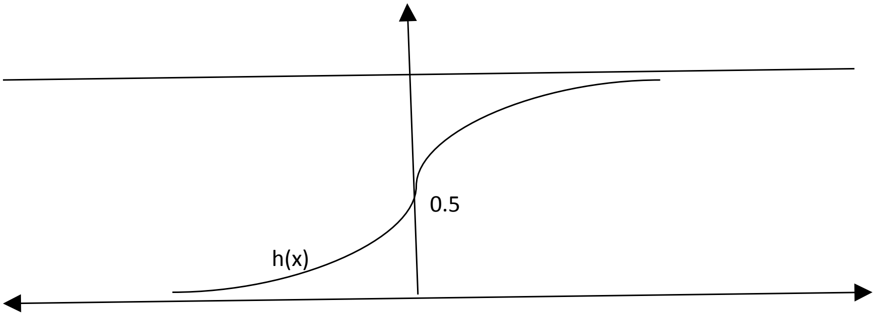


Fig. 1 Sigmoid function h(x)

- h(x) mentioned here is sigmoid function, meaning that whenever value h(x) increases or decreases than 0.5 confidence level increases. In other words, we can also say that as the value of h(x) increases or decreases from 0.5, the probability of the confidence level increases (Fig. 1).
- We can say that liner regression gives us linear classifier. For example, if we have 2 classes positives and negatives and if we try to draw a decision surface, then the points that are closer to the decision surface will have less confidence level as compared to the point that far off. This means that as the distance for the classes increases with the decision surface the confidence level increases and vice versa (Fig. 2).

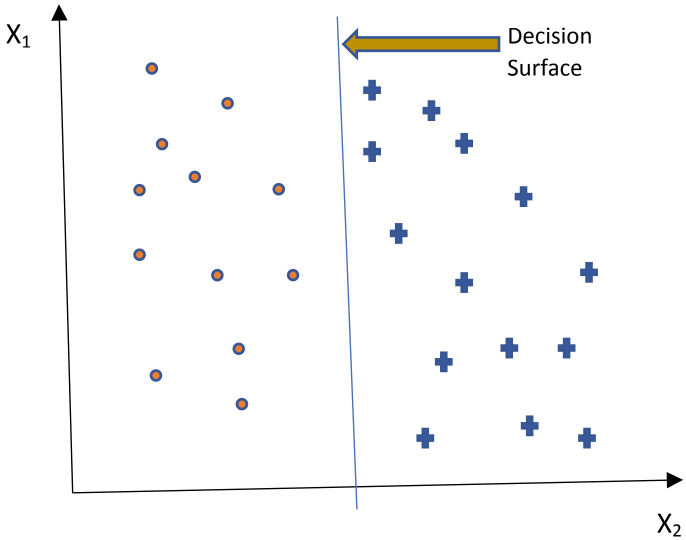


Fig. 2 Decision surface as linear classifier

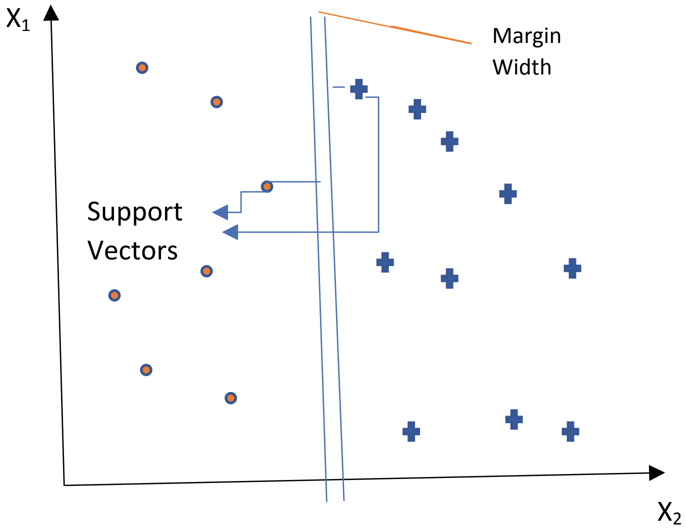


Fig. 3 Margin and margin width

In Support Vector Machine, we want a classifier so that it maximizes the distance between the points and the decision surface. Hence, we move forward to define a few terms related to SVM.

Margin

The closest distance from a point (training instance) to the decision surface is known as Margin. We can also have margin width defined such that the distance of the points from both and are same. We need to choose that decision surface for which margin width is highest. These conditions are under assumption that the 2 points or the 2 classes are linearly separable by a linear decision surface (Fig. 3).

Support Vectors

If we consider a decision surface and then consider the point closer to decision surface, and also equidistant from the decision surface. These points are called as support vectors [2].

There is can be two or more than two support vectors but in totality the number of support vectors are extremely small. The support vectors are the one that determine the equation of the line.

The decision surface determines which point given a test point lies on what side of decision surface and based on this we can classify the points as or .

Functional Margin

The distance of (x_i, y_i) from the decision boundary (w, b).

Suppose that the equation of the decision surface is given by $wx + b = C$ (Fig. 4).

Fig. 4 Functional margin

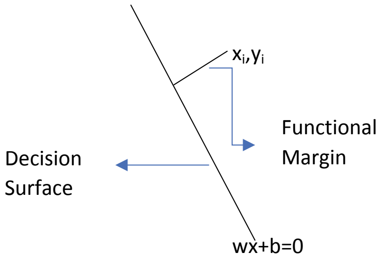
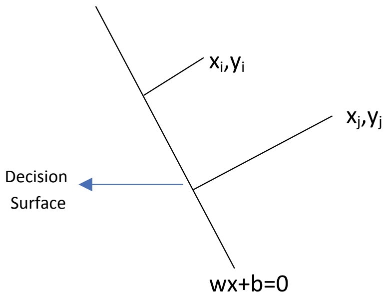


Fig. 5 Functional margin comparison



Perpendicular distance from points x_i, y_i is given by:

$$\boxed{\gamma^i = y_i\left(w^T x_T + b\right)}$$

Suppose we have another point in the line $wx + b = 0$ called as x_j, y_j such that functional margin of x_j, y_j > functional margin of x_i, y_i (Fig. 5).

In case of points x_j, y_j, equation of line is still the same and since its functional margin is higher, it has more confidence level than points x_i, y_i. Distance from the functional margin to the decision surface not only depends on the equation of line but also on the coefficients. The coefficients represents the functional margin. As the coefficient changes so does the functional margin. So in order to normalize the same so that we arrive at normalized distance, we derive below formulae:

Suppose we have set of training points:

$$S = \left\{x_1, y_1, x_2, y_2, \ldots, x_m, y_m\right\}$$

Then the normalized distance from the training point to the decision surface is given by:

$$\gamma = \min_{i=1}^{m} y^i$$

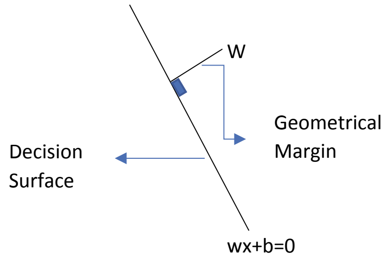


Fig. 6 Geometrical margin

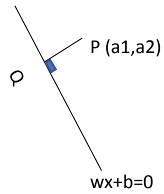


Fig. 7 Geometrical margin distance from decision surface

So the distance of set of training points from the decision surface is the minimum distance among the training points to the decision surface. To overcome this drawback, we define another term known as geometrical margin.

Geometrical Margin

This is invariant to the scaling of the equation. Geometrical margin is defined by vector W perpendicular to the decision surface [3]. Value if this unit vector is given by w/||w||. For example: if the vector W is given by (2, 3) then the value of unit vector will be $2/\sqrt{3}$ and $3/\sqrt{13}$ (Fig. 6).

Geometric Margin distance from (P, Q) from (w, b) (Fig. 7):

$$\mathrm{P} = \mathrm{Q} + \gamma \cdot \frac{\mathrm{w}}{\|\mathrm{w}\|}$$

- $(\mathrm{a1}, \mathrm{a2}) = (\mathrm{b1}, \mathrm{b2}) + \gamma \cdot \frac{\mathrm{w}}{\|\mathrm{w}\|}$
- $\mathrm{W}^{\mathrm{T}}\left((\mathrm{a1}, \mathrm{a2}) - \gamma \cdot \frac{\mathrm{w}}{\|\mathrm{w}\|}\right) + \mathrm{b} = 0$
- $\gamma = \frac{\mathrm{W}^{\mathrm{T}}(\mathrm{a1},\mathrm{a2}) + \mathrm{b}}{\|\mathrm{w}\|}$
- $\gamma = \frac{\mathrm{W}^{\mathrm{T}}(\mathrm{a1},\mathrm{a2})}{\|\mathrm{w}\|} + \frac{\mathrm{b}}{\|\mathrm{w}\|}$
- $\boxed{\begin{array}{l}\gamma = \mathrm{y} \cdot \left(\frac{\mathrm{W}^{\mathrm{T}}(\mathrm{a1},\mathrm{a2})}{\|\mathrm{w}\|} + \frac{\mathrm{b}}{\|\mathrm{w}\|}\right) \\ \text{If } \|\mathrm{w}\| = 1, \text{ then } \gamma = \mathrm{y} \cdot \left(\mathrm{W}^{\mathrm{T}}(\mathrm{a1}, \mathrm{a2}) + \mathrm{b}\right)\end{array}}$

2 Problem of Optimization: Maximization of Margin Width

If (w, b) characterizes the decision surface and $\frac{\gamma}{\|w\|}$ is the geometrical margin and we want to learn the values of (w, b) such that the value of $\frac{\gamma}{\|w\|}$ is the largest subject to the constraints [4] (Fig. 8)

$$\begin{aligned} WX_i + b &\geq \gamma \text{ for all +ve points} \\ WX_i + b &\leq \gamma \text{ for all −ve points} \end{aligned}$$

If $\gamma = 1$, then the problem becomes maximize $\frac{1}{\|w\|}$, which is equivalent to minimize $||w|| = W \cdot W^T$.

Our constraints can be re-written as:

$$\boxed{y \cdot (WX_i + b) \geq \gamma[= 1] \quad \text{for all } \ i = 1, 2, 3, \ldots, m}$$

For normalizing, we scale w such that geometric margin is $\frac{1}{\|w\|}$.

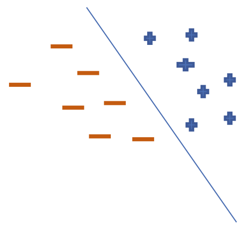
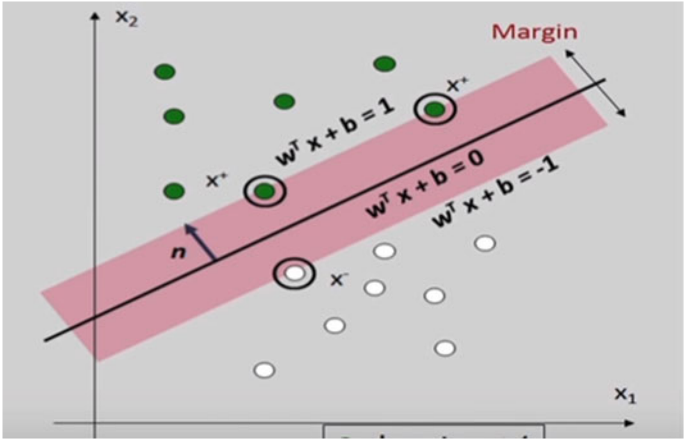


Fig. 8 Decision surface dividing positive and negative points

3 Formulation for Large Margin Linear Classifier



 denotes +1

 denotes −1

Minimize $\frac{1}{2} \cdot \|w\|^2$
Such that

$$\boxed{y \cdot (WX_i + b) \geq 1 \quad \text{for all} \;\; i = 1, 2, 3, \ldots, m}$$

Solving the Optimization Problem

$$\boxed{\begin{array}{l} \text{Minimize } \frac{1}{2} \cdot \|w\|^2 \\ \text{Such that} \\ y \cdot (wx_i + b) \geq 1 \quad \text{for all}\, i = 1, 2, 3, \ldots, m \end{array}}$$

This is the optimization problem with the convex quadratic objectives and the linear constraints. We use Lagrange Duality to get the optimization problem in dual form [5]. This has the following below advantages:

- This allows us to make use of kernels so as to get the optimal margin classifiers in order to work efficiently in high dimensional spaces.
- This allows us to make derivation of very effective algorithm to solve the above optimization problem that performs well as compared to any other software.

Before we take up any solution for above optimization, we define Lagrangian duality as below:

The Primal Problem

$$\boxed{\begin{array}{l} \mathrm{Min_w f(w)} \\ \text{Subjected to constraints:} \\ \mathrm{g_i(w) \le 0, \quad i = 1, 2, \ldots, k} \\ \mathrm{h_i(w) = 0, \quad i = 1, 2, \ldots, 1} \end{array}}$$

The generalized Lagrangian:

$$\boxed{\begin{array}{l} \mathrm{L(w, \alpha, \beta) = f(w) + \sum_{i=1}^{k} \cdot \alpha_i g_i(w) + \sum_{i=1}^{k} \cdot \beta_i h_i(w)} \\ \alpha \text{ and } \beta \text{ are known as Lagrangian Multipliers and are } \ge 0 \end{array}}$$

Lemma

$$\left\{ \mathrm{Max_{\alpha,\beta,\alpha i} L(w, \alpha, \beta)} = \begin{array}{ll} \mathrm{f(w)}, & \text{if w satisfies primal constraints} \\ \infty, & \text{otherwise} \end{array} \right.$$

$$\boxed{\begin{array}{l} \text{Re-writing Primal:} \\ \mathrm{Min_w Max_{\alpha,\beta,\alpha i \ge 0} L(w, \alpha, \beta)} \end{array}}$$

So, our Primal problem is $\mathrm{p^* = Min_W Max_{\alpha,\beta,\alpha i \ge 0} L(w, \alpha, \beta)}$.
The dual problem is $\mathrm{d^* = Max_{\alpha,\beta,\alpha i \ge 0} Min_W L(w, \alpha, \beta)}$.

Theorem

Case 1: Weak Duality

$$\mathrm{d^* = Max_{\alpha,\beta,\alpha i \ge 0} Min_W\, L(w, \alpha, \beta) \le Min_W Max_{\alpha,\beta,\alpha i \ge 0}\, L(w, \alpha, \beta) = p^*}$$

Case 2: Strong Duality

If there exists a saddle point of $\mathrm{L(w, \alpha, \beta)}$ then we have $\mathrm{d^* = p^*}$.

KKT Conditions

If we have a saddle point of L, then it should satisfy the following KKT "Karush-Kuhn-Tucker" conditions:

$$\boxed{\begin{array}{ll} \frac{\partial L(w,\alpha,\beta)}{\partial w_i} = 0, & i = 1, 2, \ldots, k \\ \frac{\partial L(w,\alpha,\beta)}{\partial \beta_i} = 0, & i = 1, 2, \ldots, l \\ \alpha_i g_i(w) = 0, & i = 1, \ldots, m \\ g_i(w) \le 0, & i = 1, \ldots, m \\ \alpha_i \ge 0, & i = 1, \ldots, m \end{array}}$$

Theorem If w^*, a^*, b^* satisfy KKT condition then it is also a solution to the primal and dual problems.

Support Vectors

- Few α_i's can only be non-zero.
- Training Data points whose α_i is non-zero are known as support vectors.

$$\boxed{\alpha_i g_i(w) = 0 \quad i = 1, 2, \ldots, m}$$

if $\alpha > 0$ then $g(w) = 0$.

Now we try to solve the optimization problem we discussed before,

$$\begin{array}{l} \text{Minimize } \frac{1}{2} \cdot \|w\|^2 \\ \text{Such that } y \cdot (WX_i + b) \ge 1 \quad \text{for all } i = 1, 2, 3 \ldots, m \end{array}$$

When comparing above problem with Lagrangian Duality, only have f(w) and 1st condition $g_i(w) \le 0$, i = 1, 2, …, k in place.

Taking Lagrangian Function of the above expression we get:

$$\boxed{\begin{array}{l} \text{minimize} \quad L_p(w, b, \alpha_i) = \frac{1}{2} \cdot \|w\|^2 - \sum_{i=1}^{n} \cdot \alpha_i (y \cdot (wx_i + b) - 1) \\ \text{s.t.} \quad \alpha_i \ge 0, \text{ p stands for "primal"} \end{array}}$$

We will minimize w and b for fix α.

$$\frac{\partial L_p}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{n} \cdot \alpha_i \cdot y_i x_i$$

$$\frac{\partial L_p}{\partial b} = 0 \Rightarrow \sum_{i=1}^{n} \cdot \alpha_i \cdot y_i = 0$$

$$L_p(w, b, \alpha_i) = \sum_{i=1}^{m} \cdot \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \cdot \alpha_i \alpha_j y_i y_j \left(x_i^T x_j\right) - b \sum_{i=1}^{m} \cdot \alpha_i y_i$$

$$L_p(w, b, \alpha_i) = \sum_{i=1}^{m} \cdot \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \cdot \alpha_i \alpha_j y_i y_j \left(x_i^T x_j\right)$$

Now the dual optimization problem reduces as below:

$$\boxed{\begin{array}{l} \max\limits_{\alpha} J(\alpha) = \sum\limits_{i=1}^{m} \cdot \alpha_i - \frac{1}{2} \sum\limits_{i,j=1}^{m} \cdot \alpha_i \alpha_j y_i y_j \left(x_i^T x_j\right) \\ \text{s.t.} \quad \alpha_i \geq 0 \quad i = 1, 2, \ldots, k \\ \sum\limits_{i=1}^{m} \cdot \alpha_i \cdot y_i = 0 \end{array}}$$

This is a quadratic programming problem, there always exist a global maximum of α_i.

4 Support Vector Machine Formulation

Once we have the Lagrangian multiplier $\{\alpha_j\}$, weighted parameter 'w' can be reconstructed as weighted combination of training example:

$$w = \sum_{i=1}^{m} \cdot \alpha_i \cdot y_i x_j$$

$$w = \sum_{i(-SV} \cdot \alpha_i \cdot y_i x_i$$

If we have new data set 'z', then we compute:

$$w^T z + b = \sum_{i(-SV} \cdot \alpha_i \cdot y_i \left(x^T z\right) + b$$

Based on the above computation, we classify 'z' as class 1 when the Sum gives positive and class 2, when the sum gives negative. 'w' value need not be formed explicitly.

Solving the Optimization Problem

The discriminant function reduces to be:

$$\boxed{g(x) = w^T x + b = \sum_{i(-SV} \cdot \alpha_i \cdot y_i \left(x_j^T x\right) + b}$$

The above expression relies on the dot product between support vectors x_i and the test point x. So, in order to solve the optimization problem, we need to compute the dot products between x_i^T and x_j. x_i can be a high dimensional vectors but the dot product results are scalar.

'w' is optimal and forms the linear combination of for all data points that are small.

5 SVM: Maximum Margin with Noise

There are few data sets which are incorporated with noise, in these cases it becomes difficult to make the data linearly separable. From linear SVM formulation, we find w and b such that the Margin width is maximized i.e., $\frac{2}{\|w\|}$ or $\frac{\|w\|}{2}$ is maximized or we can say $\|w\|^2 = w \cdot w$ is minimized [6].

For each training point, i = 1, …, n (x_i, y_i), we have:

$$Y_1(wx_i + b) \geq 1$$

These help separating those instances that are linearly separable. We cannot use linear SVMs for the noisy data or data that are non-linear. In order to overcome this drawback, we extend the definition of maximum margin to allow non-separating planes.

Now we redefine our decision surface in order to fit the noisy data. For this either we can increase the decision surface, or we try to minimize the misclassification of data.

Now the objective function changes to:

$$w \cdot w + c * \text{number of training errors}$$

where "c" is used for controlling the importance that we want to give for number of training errors. The problem now is that we cannot use Quadratic Programming to solve this objective function. Hence, we try to look at some alternative methods to find the solution of this optimization problem.

We assume that, given some points, we are able to define some margin. Corresponding to each point that are not correctly classified, assuming a particular decision surface, and the corresponding margin, for each such point, we can give a "penalty". Penalty can be given based on how far the misclassified points are from the correct position so that they are classified correctly [7] (Fig. 9).

We draw a perpendicular from the misclassified points to the margin and the perpendicular distance between the misclassified points and the margin are given by introducing a slack variable €.

Modified Objective Function

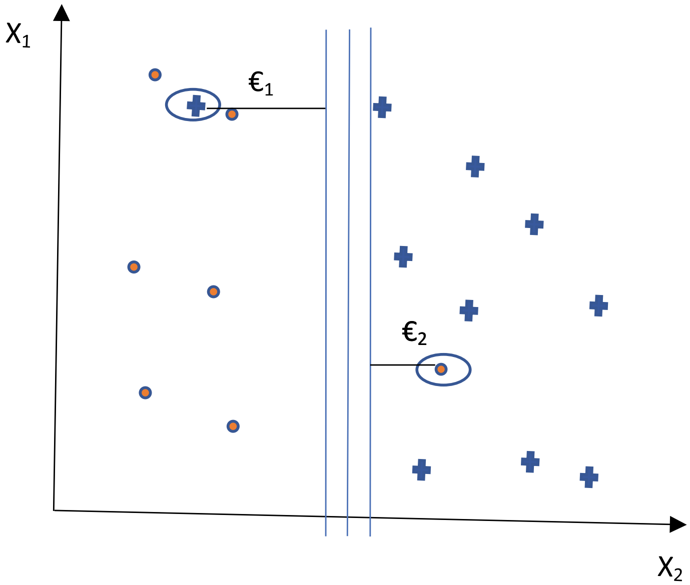


Fig. 9 Decision surface with misclassification of training data

$$\mathrm{w} \cdot \mathrm{w} + \mathrm{c} \cdot \sum_{k=1}^{m} €_{\mathrm{k}}$$

where “m” is the number of training examples.

The final equation becomes:
Minimize:
w ·w + C · (distance of error points from their correct zones)
And then add the slack variable $€_1$.
“C” controls the relative importance of maximizing the margin and fitting the training data and control overfitting (Fig. 10).
So, for each training sets we have:

$$\mathrm{Y_i}(\mathrm{wx_i} + \mathrm{b}) \geq (1 - €_{\mathrm{k}}) \quad \text{for } \mathrm{i} = 1, 2, \ldots, \mathrm{m}$$

Lagrangian:

$$\mathrm{L}(\mathrm{w}, \mathrm{b}, €, \alpha, \beta) = \frac{1}{2}\mathrm{w} \cdot \mathrm{w} + \mathrm{c} \cdot \sum_{k=1}^{m} €_{\mathrm{k}} + \sum_{i=1}^{m} \cdot\alpha_{\mathrm{i}}\left[\mathrm{y_i}(\mathrm{x} \cdot \mathrm{w} + \mathrm{b}) - 1 + €_{\mathrm{l}}\right] - \sum_{i=1}^{m} \cdot\beta_{\mathrm{i}}€_{\mathrm{i}}$$

α_i’s and β_i’s are the Lagrange multipliers ($\geq$0).
Dual Formulation:
For $\alpha_{1,}$ α_2, …, α_m such that

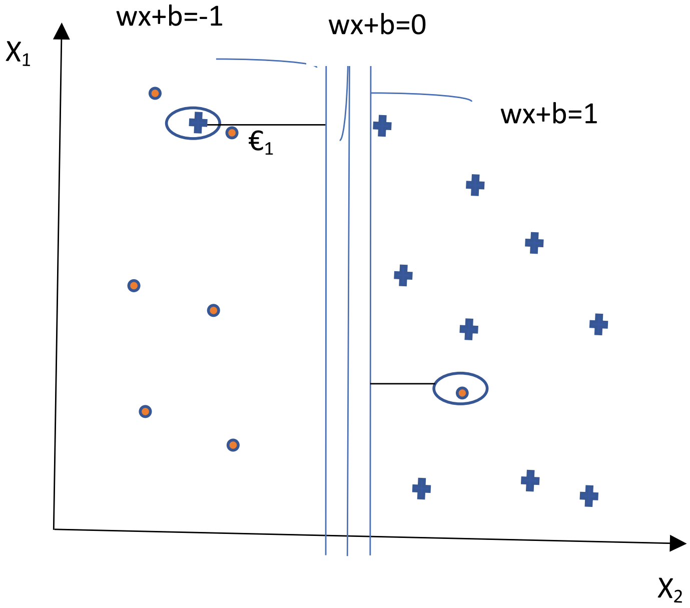


Fig. 10 Decision surface with misclassification of training data

$$\max_{\alpha} J(\alpha) = +\sum_{i=1}^{m} \cdot \alpha_i - 1 \sum_{i,j=1}^{m} \cdot \alpha_i \alpha_j y_i y_j \left(X_i^T X_j\right)$$

For Linear SVM:	**Noise Accounted**:
s.t. $\alpha_i \geq 0$,	s.t. $0 \leq \alpha_i \leq c$
$\sum_{i=1}^{m} \cdot \alpha_i y_j = 0$	$\sum_{i=1}^{m} \cdot \alpha_i y_j = 0$

Soft SVM:

In soft SVM, we do not have hard decision surface that clearly separates the positive and negative points [8].

- $\mathbf{x_i}$ with non-zero α_i's will be support vectors
- Solution to dual problem is:

$$w = \sum_{i=1}^{m} \cdot \alpha_i y_j x_i$$

$$b = y_k(1 - €_k) - \sum_{i=1}^{m} \cdot \alpha_i y_j x_i x_k$$

- For any k s.t. $\alpha_k > 0$

For classification,

$$\mathrm{F(x)} = \sum_{i=1}^{m} \cdot\alpha_{\mathrm{i}} \mathrm{y_j x_i} \cdot \mathrm{x} + \mathrm{b} \quad \text{(we need not compute w in this case explicitly)}$$

6 Conclusion

The above formulation and SVM knowledge opens us the door for various real world applications. As we know SVM depends on the supervised learning, the need of SVM is to perform classification of the unseen data. Some of the common applications we can say are:

Face Detection:

SVMs performs classification of a face based on non-face and face and a square shaped boundary is formed around the face.

Categorization of Hypertext and Text:

SVMs uses training data for the classification for documents into various documents. Categories gets a score and the values are compared with the threshold value.

Classification of Images:

SVMs are used in image classification as they provide better accuracy for search as compared to any of the traditional methods.

Bioinformatics:

SVMs use classification for cancer and different kinds of proteins. This helps in gene and patient classification based on genes or any other problems related to some biology. SVM algorithms are often used for detection of protein folding.

Handwriting Recognition:

We use SVMs widely for recognizing the handwriting based on the training data available.

Wide variety of applications depicts the popularity of the Support Vector Machine. SVMs have been found to perform very well with the data in limited amount. Algorithm used for SVM are simpler and when we try to apply these into natural language processing, we do not require complex formulations and are found dependable in these cases.

There are many terms that are dig out when we talk about SVMs such as kernel functions, non-liner SVMs. Naïve Bayes is also used for similar sets of applications but once you are confident on your training data Support Vector Machines proves to be most efficient. There are many discoveries going on in these fields and we are looking forward for finding out the solution for more complex problems.

References

1. Hsu C-W, Lin C-J (2002) A comparison of methods for multiclass support vector machines. IEEE Trans Neural Netw 13(2):415–425
2. Zhang X (2000) Introduction to statistical learning theory and support vector machines. Acta Autom Sin 26(1):32–42
3. Dibike YB et al (2001) Model induction with support vector machines: introduction and applications. J Comput Civ Eng 15(3):208–216
4. Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, Cambridge
5. Chapelle O, Vapnik V (2000) Model selection for support vector machines. In: Advances in neural information processing systems
6. Osuna E, Freund R, Girosit F (1997) Training support vector machines: an application to face detection. In: Proceedings of IEEE computer society conference on computer vision and pattern recognition. IEEE
7. Tay FEH, Cao L (2001) Application of support vector machines in financial time series forecasting. Omega 29(4):309–317
8. Tong S, Koller D (Nov 2001) Support vector machine active learning with applications to text classification. J Mach Learn Res 2: 5–66

Towards Designing a Computationally Efficient Neural Network Model for Improved Financial Time Series Forecasting

Sarat Chandra Nayak, Koppula Vijaya Kumar and B. Satyanarayana

Abstract Designing computationally efficient forecasting models with improved accuracy is always a challenging task and attracting the attention of researchers. This paper develops a computationally efficient hybrid forecasting model which uses a Pi-Sigma neural network (PSNN) as the base architecture and Jaya algorithm to search the optimal model parameters. The PSNN has only one layer trainable weights thus offering less computational complexity. Jaya is a simple but powerful optimization algorithm that necessitates only some control parameters rather any parameters specific to algorithm. Hence the hybrid model synergies the advantages of PSNN and Jaya algorithm to form a quite less complex and efficient model. We termed the model as J-PSNN and evaluated it on predicting the closing indices of few real stock markets. The performance of the proposed J-PSNN model is compared to that of few state-of-the-art prediction models similarly trained and established its superiority.

Keywords Financial time series forecasting · Pi-Sigma neural network · Stock market forecasting · Java algorithm · Multilayer perceptron

1 Introduction

Stock market performance is extremely erratic because of the influence of uncertainties, nonlinearities, high volatility, and discontinuities associated with it. Also, it is affected by instant movement happening in other stocks markets, political influences,

S. C. Nayak · K. V. Kumar
Department of Computer Science and Engineering, CMR College of Engineering & Technology, Hyderabad, India
e-mail: saratnayak234@gmail.com

K. V. Kumar
e-mail: vijaykoppula@gmail.com

B. Satyanarayana (✉)
Department of Computer Science and Engineering, CMR Institute of Technology, Hyderabad, India
e-mail: bsat777@gmail.com

V. K. Gunjan et al. (eds.), *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, Lecture Notes in Electrical Engineering 643,
https://doi.org/10.1007/978-981-15-3125-5_58

various macro-economical factors, psychology of common investors, and sometimes rumor. Prediction of such irregular and highly fluctuating stock data is very much complex and typically subject to large error. Hence, robust forecasting methods are intensely desirable for decision making of speculators, financial managers, naive investors etc. A small improvement in prediction may probable leads to elevated financial pay back. Over the last few decades a large number of statistical [1–3] as well as machine learning approaches [4–11] have been formulated and suggested in the literature of stock forecasting. On the other hand, the accuracy of a model is problem specific and varies among datasets. Though few methods claimed superior over others, achieving improved forecasting accuracy is still an issue. The selection of most promising forecasting model is itself a tricky job. With the swell in the number of available forecasting techniques, the studies interrelated to their comparative assessment are also growing incessantly.

Artificial Neural Network (ANN) has the resemblance with the judgment power of human brain and thus mimicking it [12, 13]. The ANN can follow the practice of human brain to determine nonlinear multifaceted problems and got wide popularity. By means of little a priori assumptions about the generating procedure, they are very capable to model non-linear processes. ANNs are considered to be a powerful modeling practice for the cases where input-output mapping has regularities and exceptions together. These advantages tempted researchers to pave the path toward developing ANN based models to study the chaotic actions of stock market and forecast its trend. Dealing with ambiguity and nonlinearity associated with stock market data using ANN based forecasting methods mainly involve two steps, i.e. recognition of patterns in the data and predict future events using such patterns.

Among the ANN based forecasting models, the multilayer perceptron (MLP) is probably the most widespread model adopted by the researchers. However, MLP has certain issues like adding more computational complexity, leads to lower learning rate and accuracy. It has been characterized with structural complexity, computational overhead and has black box visualization. In contrast to MLP, higher order neural networks (HONN) are characterized with stronger approximation capability, fast learning properties, larger storage and higher fault tolerance capability etc. Also, it has a prevailing mapping single layer tunable weight set which overcome the black box nature of MLP as suggested by authors in [14]. Use of higher order terms can amplify the information capacity and help solving intricate nonlinear task even with a petite network without compromising the convergence capability. In the other hand, with the swell in order of network there is intensification in tunable weights, which may lead to additional computation cost. To alleviate these drawbacks, a special kind of HONN called as Pi-Sigma neural network (PSNN) has been introduced by Shin and Ghosh [15] using less number of connection weights. Several applications are found in the literature using HONN for prediction of stock market future [4–11].

The population based optimization algorithms are broadly categories into two classes. The first category includes the evolutionary algorithms inspired from the theory of natural evolution such as genetic algorithm (GA), differential evolution (DE), evolution programming (EP), etc. The second category is based on swarm intelligence such as particle swarm optimization (PSO), ant colony optimization

(ACO), bee colony optimization (BCO), and so on. Apart from these two classes, there are some other optimization techniques, which are inspired from various natural phenomena such as gravitational search algorithm (GSA), harmony search algorithm (HSA), fireworks algorithm (FWA), etc. The above mentioned optimization techniques are applied to explain various real life multifaceted problems and their superiority have been established. Some applications of GA in the stock market prediction can be found in [4, 6, 10]. However, their performance essentially depends upon fine-tuning several algorithm specific control parameters. Choosing appropriate control parameters is a difficult task and requires intensive human interventions. Inappropriate selection of algorithm parameters may add computational burden or lands the model in local optima. Therefore, adopting an optimization technique that requires very few controlling parameters without losing the performance can be the best choice for solving real world problems.

Jaya is a simple and robust optimization algorithm recently proposed by Rao [16]. It does not entail any algorithm specific parameters rather few general control parameters only. The algorithm focuses on moving the current solution toward finest solution simultaneously ignoring most awful solution. The capability of Jaya has been experimented in several works [16, 17] and found successful.

The objective of this work is to develop an efficient forecasting model with lower computational complexity. In order to achieve this we used a higher order neural network, i.e. a PSNN which contains only one layer tunable synaptic weights, thus offering lower computation overhead. To search the model parameters the efficient Jaya algorithm has been employed. The model combines the advantages of both PSNN and Jaya algorithm to form a hybrid model J-PSNN. The model is evaluated on predicting the next days' closing prices of DJIA, KOSPI, HIS, and BSE. The input signals are selected by using a sliding window technique. The model uses only few input closing prices, trained adaptively hence provides lower computational cost.

The rest of the paper is organized as follows. The introduction about the stock market forecasting and earlier research is discussed in Sect. 1. The methods and methodologies are discussed in Sect. 2. The proposed J-PSNN forecasting is explained in Sect. 3. Section 4 presents the experimental results and analysis. Section 5 gives the concluding remarks and followed by a list of references.

2 Materials and Methods

This section shortly describes about the base neural model, i.e. the PSNN and the learning algorithm, i.e. the Jaya algorithm

2.1 Pi-Sigma Neural Network

The PSNN belongs to the class of HONN and having architecture of double layered. The network is feed forward and fully connected [15]. The first layer is a composed of summing units and the second layer is composed of product units. The network input are feed to input layer consist of summation units and the consequent outputs are feed to the output layer consist of a product unit. The synaptic input-summing layer weights are trainable where as the connection summing-product layer weights are non-tunable and set to a value of unity. Since only one tunable weight set is required, the network achieved a significant decline in training time. Linear activation functions are used at summing layer units and nonlinear activation functions are used by the product units. The order of the network increases by one in addition of each extra summing unit. The product unit offers higher order capabilities to the network by intensifying the input space from lower to higher dimension. In this way, the network avoids being suffered from exponential swell in weights and simultaneously offering better nonlinear separable capacity to the network.

2.2 Jaya Algorithm

Jaya is a simple and dominant optimization method recently proposed by Rao [16]. It only requires few common control parameters rather any algorithm specific parameters to reach the global optima. The algorithm focuses on moving the current solution toward preeminent solution simultaneously ignoring the bad solution. In the simplest form the working principle of the algorithm can be described as follows:

Let the objective function to be optimized = $f(x)$.

Let the number of design variable = m and number of candidate solution = n.

Let the fitness of best and worst candidate in the entire candidate solution are represented as $f(x)_{best}$ and $f(x)_{worst}$ respectively. Let $C_{j,k,i}$ is the value of the j-th variable ($j = 1, 2, \ldots, m$) for the k-th ($k = 1, 2, \ldots, n$) candidate during the i-th iteration. Then the modified candidate $\hat{C}_{j,k,i}$ can be computed as follows:

$$\hat{C}_{j,k,i} = C_{j,k,i} + rand_{1,j,i}\left(C_{j,best,i} - \left|C_{j,k,i}\right|\right) - rand_{2,j,i}\left(C_{j,worst,i} - \left|C_{j,k,i}\right|\right), \quad (1)$$

where, $C_{j,best,i}$ and $C_{j,worst,i}$ are the value of j for best and worst contestant respectively. $rand_{1,j,i}$ and $rand_{2,j,i}$ are random values in the range [0,1] for the j-th variable in the i-th iteration. The term "$+rand_{1,j,i}\left(C_{j,best,i} - \left|C_{j,k,i}\right|\right)$" represents the inclination of the solution nearer to the best solution. Similarly, the term "$-rand_{2,j,i}\left(C_{j,worst,i} - \left|C_{j,k,i}\right|\right)$" represents the propensity of the solution to shift away from the worst solution. In this way the algorithm tries to avoid the failure and get closer to the global optimum. Based on the above concept the Jaya algorithm steps can be presented as in Fig. 1.

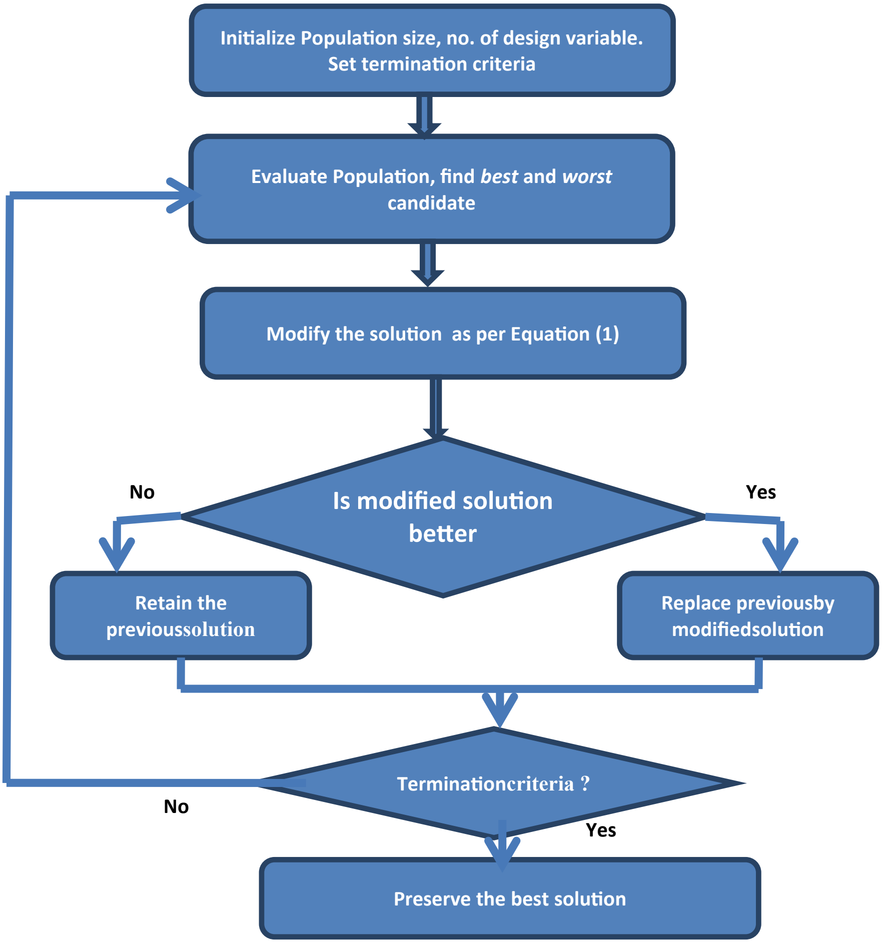


Fig. 1 Process of Jaya algorithm

3 Proposed J-PSNN Based Forecasting

The proposed J-PSNN based model is shown in Fig. 2. The input-hidden layer weights are trainable while that of hidden-output layer is fixed. The output of J-PSNN is calculated as follows. Let the output at j-th summing component of hidden layer is computed by summing up the products from each input x_i and corresponding weight w_{ij} as in Eq. (2).

$$y_j = \sum_{i=1}^{n} w_{ij} * x_i \tag{2}$$

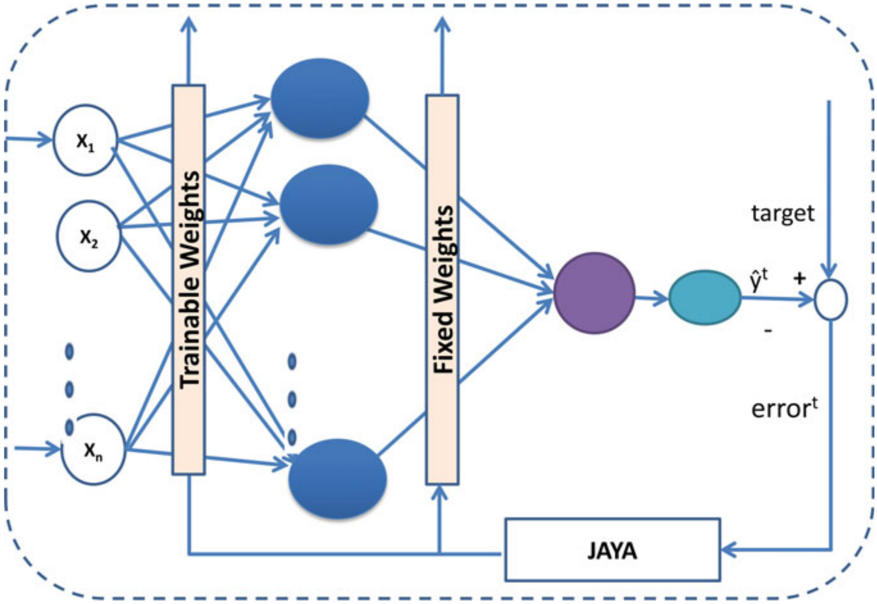


Fig. 2 J-PSNN based forecasting

where n is the total volume of input signals fed to the input layer. The output unit computes the product of the outputs from the summing units of hidden layer and forwarding it to a nonlinear sigmoid activation as in Eq. (3).

$$\hat{y} = \sigma\left(\prod_{j=1}^{m} y_j\right) \tag{3}$$

where m is the order of the network and symbol σ is the product operator. Now the target is supplied to the output neuron, compared with model estimation and the error value is calculated. Each candidate solution represents one potential weight set for the model. The candidate with least error value is considered as the best and that one having maximum error value is treated as worst candidate solution. Now the weight set is optimized as explained in Sect. 2.2 by Jaya algorithm.

4 Experimental Results and Analysis

This section describes about the data for experiment, statistical description of datasets, data normalization, input selection for the model and experimental results and analysis of the results.

4.1 Data Collection and Description

For experiment work, historical closing indices are collected for each transactional day form https://in.finance.yahoo.com/quote/. The lists of financial time series are summarized in Table 1. The statistical descriptions of the datasets are summarized in Table 2.

The daily closing prices for KOSPI, HSI, BSE, and DJIA are plotted in the Figs. 3, 4, 5 and 6 respectively. The histograms of daily returns of the KOSIP, HIS, BSE, and DJIA stock datasets against the theoretical normal distribution are drawn and presented by Figs. 7, 8, 9 and 10 respectively. For all the financial time series the skewness values are found to be negative. This negative value is an implication that all the datasets are spread out more toward left. The kurtosis values for all datasets are more outlier prone. Similarly, the Jarque-Bera test statistic values show that all datasets are non normal distributed.

The data for training and test were generated from original financial time series by using a sliding window technique. A window of predetermined range is moving on the whole series by one step each time. Each move forms a new pattern which is feed to the model as an input. The size of window is determined experimentally. The original input measures are normalized by sigmoid method [18] as follows:

Table 1 Information about the stock datasets

S. No.	Stock data	Stock exchange	Period	No. of data
1	HSI	Hang Seng Index, Hong Kong stock market	02.01.2015–20.09.2018	921
2	DJIA	Dow Jones Industrial Average's, US market Index	31.12.2014–19.09.2018	938
3	BSE	Bombay stock exchange S&P 500	01.01.2015–19.09.2018	923
4	KOSPI	Korean stock exchange	02.01.2015–20.09.2018	916

Table 2 Descriptive statistics of closing prices for different stocks

Stock Index	Descriptive statistics						
	Min.	Max.	Mean	Std dev.	Skewness	Kurtosis	Jarque-Bera test statistics
DJIA	1.5660e+04	2.6617e+04	2.0232e+04	3.0450e+03	−0.6524	6.9293	668.5418 (h = 1)
BSE	9.1880e+03	1.5846e+04	1.2290e+04	1.7848e+03	−0.9764	8.5211	1.3161e+03 (h = 1)
KOSPI	1.8298e+03	2.5982e+03	2.1600e+03	203.8555	−0.4836	4.8327	163.5340 (h = 1)
HSI	1.8320e+04	3.3154e+04	2.5138e+04	3.4698e+03	−0.4274	5.5983	286.4947 (h = 1)

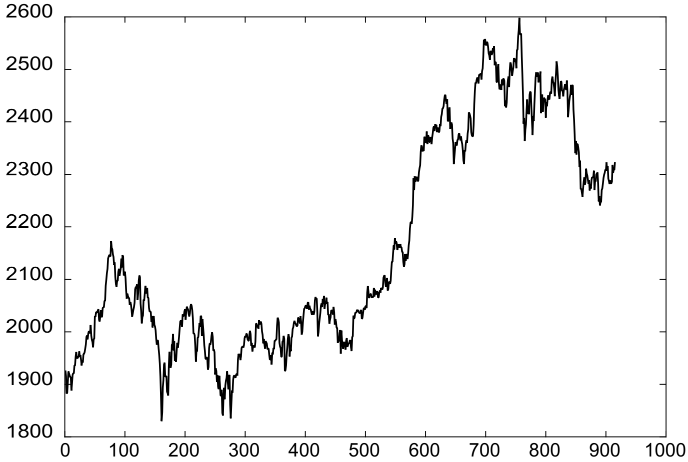


Fig. 3 Daily observations of KOSPI for the period 1st January 2015 to 20th September 2018

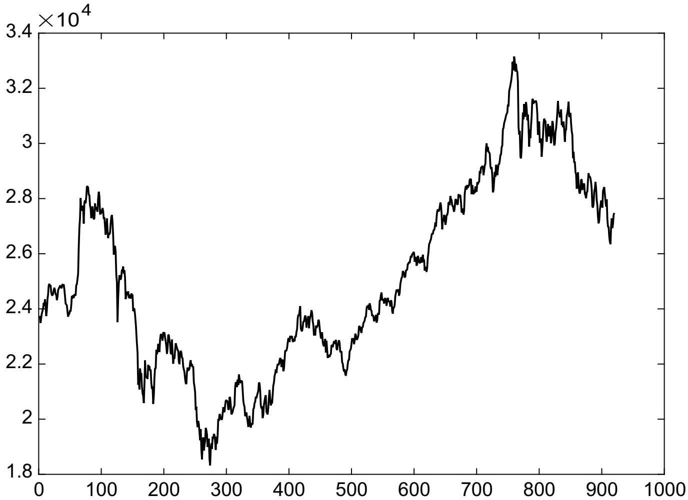


Fig. 4 Daily observations of HSI for the period 1st January 2015 to 20th September 2018

$$x_{norm} = \frac{1}{1 + e^{-\left(\frac{x_i - x_{min}}{x_{max} - x_{min}}\right)}}, \tag{4}$$

where x_{norm} and x_i are the normalized and original price respectively. x_{max} and x_{min} are the maximum and minimum price of the current input pattern respectively.

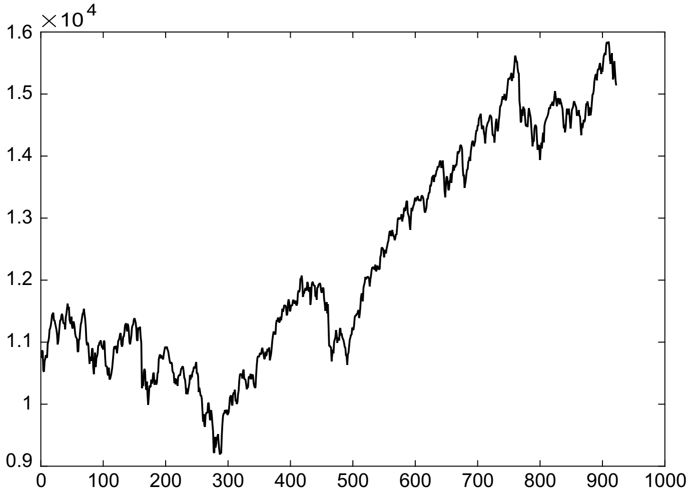


Fig. 5 Daily observations of BSE for period 1st January 2015 to 20th September 2018

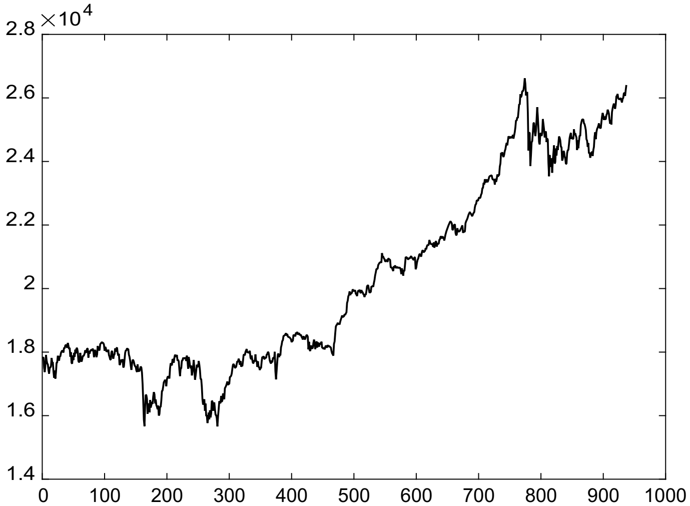


Fig. 6 Daily observations of DJIA for period 1st January 2015 to 20th September 2018

4.2 Experimental Results and Analysis

The same training and test patters are fed to all the models considered for forecasting. In order to avoid the stochastic nature of the forecasting model, we simulated each model 10 times and considered the average error statistics for performance

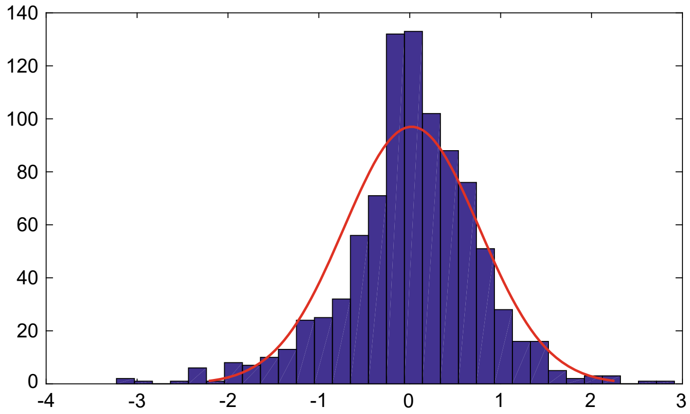


Fig. 7 Histogram of daily returns of the KOSIP against the theoretical normal distribution

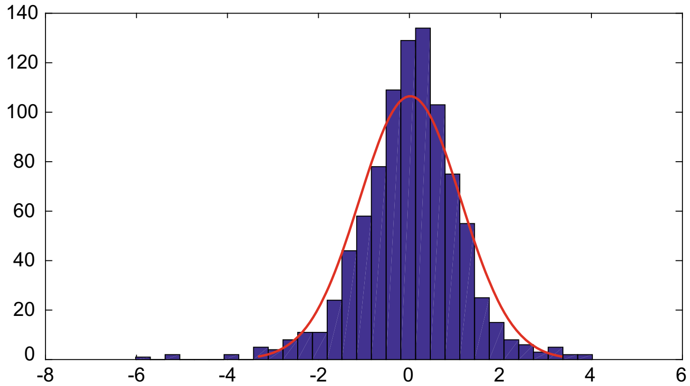


Fig. 8 Histogram of daily returns of the HSI against the theoretical normal distribution

comparison. The performance of J-PSNN model is compared with that of a multilayer perceptron (MLP), gradient descent PSNN (GD-PSNN), GA based PSNN (GA-PSNN), and PSO based PSNN (PSO-PSNN).

It is observed from Table 3 that, the average error values generated by the proposed J-PSNN forecasting model are lower than that of other models. The better average error values are highlighted in boldface letter. For an example, the average error value generated by the J-PSNN from the KOSPI financial time series is 0.0049 and that from HIS, BSE, and DJIA are 0.0045, 0.0039, and 0.0032 respectively. For more clarity on performance, we calculated the gain in average error of J-PSNN over other models as per the following formula and the results are shown in Fig. 11.

$$\text{Average gain} = \frac{(\text{Average of existing model} - \text{Average of proposed model})}{\text{Average of existing model}} \times 100\% \quad (5)$$

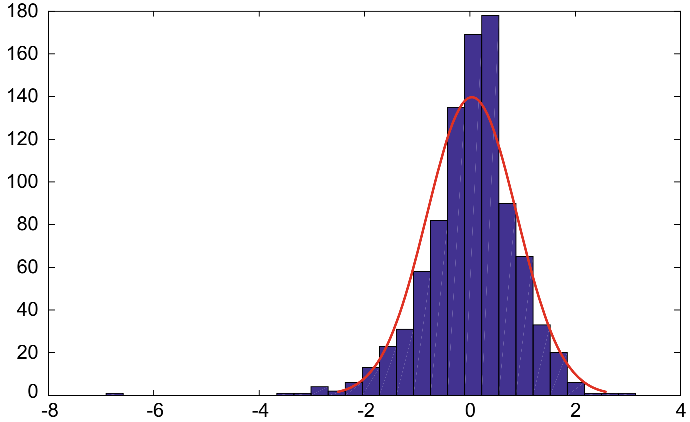


Fig. 9 Histogram of daily returns of the BSE against the theoretical normal distribution

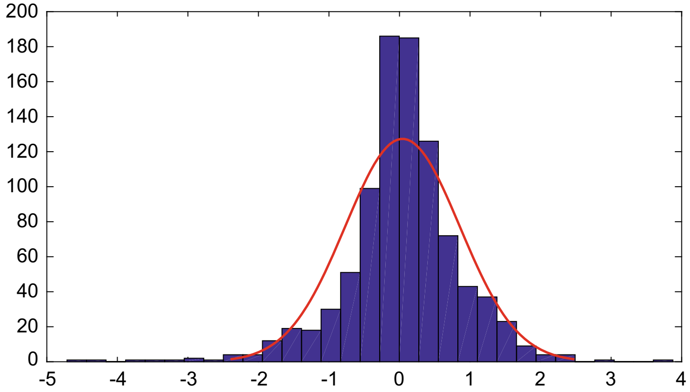


Fig. 10 Histogram of daily returns of the DJIA against the theoretical normal distribution

It is also observed that, standard error gains by the proposed model over other models are quite better. Considering four datasets, the average gain % of J-PSNN over MLP was found 65.37%, over GD-PSNN it was 63.96%, over GA-PSNN it was 55.35%, and over PSO-PSNN it was 55.04%. These observations are in favor of the superiority of J-PSNN over other models. Further, in order to establish the superiority of the proposed model, actual versus estimated prices are plotted and shown in Figs. 12, 13, 14 and 15. The plots are showing the closeness of the model estimations to the actual closing prices.

Table 3 Error statistics from five forecasting models considering four stock indices

Stock index	Error statistic	Forecasting models				
		MLP	GD-PSNN	GA-PSNN	PSO-PSNN	J-PSNN
KOSPI	Minimum	9.0384e−05	8.6386e−05	9.6336e−06	9.6336e−05	9.6265e−06
	Maximum	0.0502	0.0335	0.0322	0.0376	0.0303
	Average	0.00995	0.00975	0.0088	0.0107	**0.0049**
	Std. deviation	0.00874	0.00813	0.0075	0.0075	0.0045
HSI	Minimum	6.9508e−05	5.9528e−05	5.0527e−06	6.0507e−05	5.6507e−06
	Maximum	0.0382	0.0340	0.0312	0.0308	0.0299
	Average	0.0128	0.0122	0.0096	0.0094	**0.0045**
	Std. deviation	0.0166	0.0159	0.0055	0.0057	0.0040
BSE	Minimum	4.8234e−06	4.5224e−06	4.1226e−06	4.1760e−06	3.0763e−06
	Maximum	0.0467	0.0455	0.0358	0.04552	0.0312
	Average	0.0180	0.0159	0.0138	0.0098	**0.0039**
	Std. deviation	0.0133	0.0102	0.0082	0.0074	0.0032
DJIA	Minimum	6.84822e−05	5.84879e−05	5.74875e−05	6.44875e−05	5.8975e−06
	Maximum	0.0473	0.0453	0.0393	0.0399	0.0362
	Average	0.00987	0.00985	0.0067	0.0069	**0.0032**
	Std. deviation	0.00469	0.00475	0.0042	0.0048	0.0033

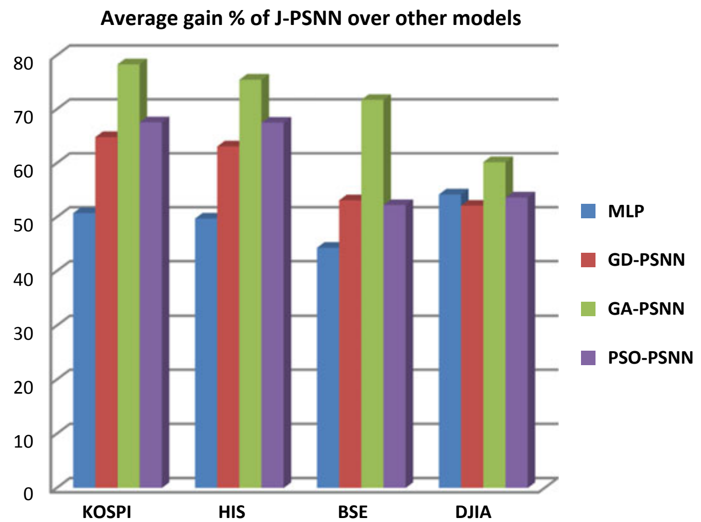


Fig. 11 Average gain percentage of J-PSNN over MLP, GD-PSNN, GA-PSNN, and PSO-PSNN forecasting model considering four financial time series

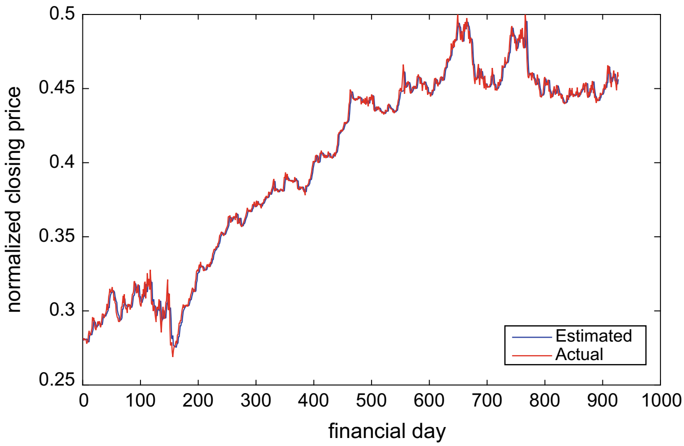


Fig. 12 Actual versus estimated daily closing prices by J-PSNN from DJIA dataset

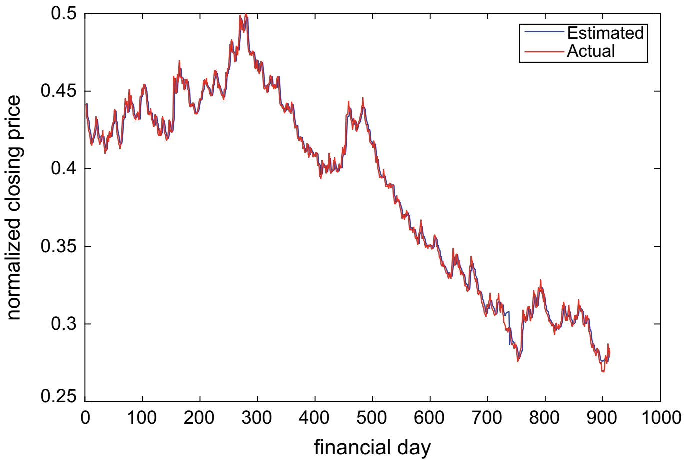


Fig. 13 Actual versus estimated daily closing prices by J-PSNN from BSE dataset

5 Conclusions

To capture the uncertainties allied with the stock data efficiently, this article presents a computationally efficient neural network based model. The model uses a higher order neural network, particularly a PSNN as the base architecture. To search the model parameters an optimization technique called Jaya algorithm is employed which requires only few common parameters. Hence, the model synergies the strengths of PSNN and Jaya algorithm forming a hybrid model, i.e. J-PSNN forecasting model. The model was engaged to predict the next day's closing indices of Korean stock

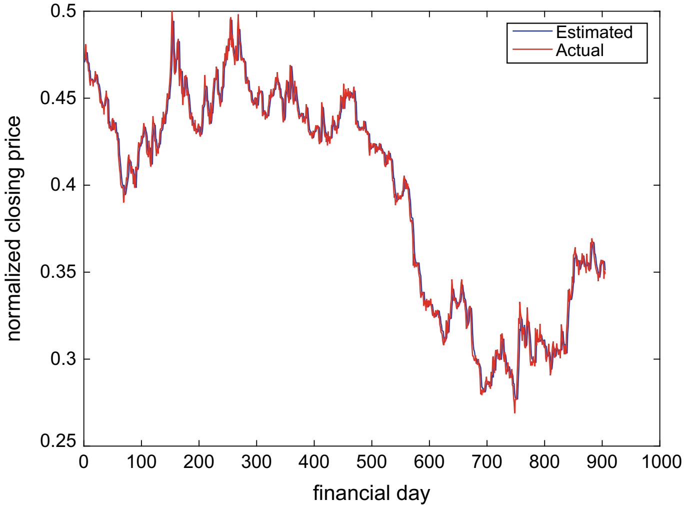


Fig. 14 Actual versus estimated daily closing prices by J-PSNN from KOSPI dataset

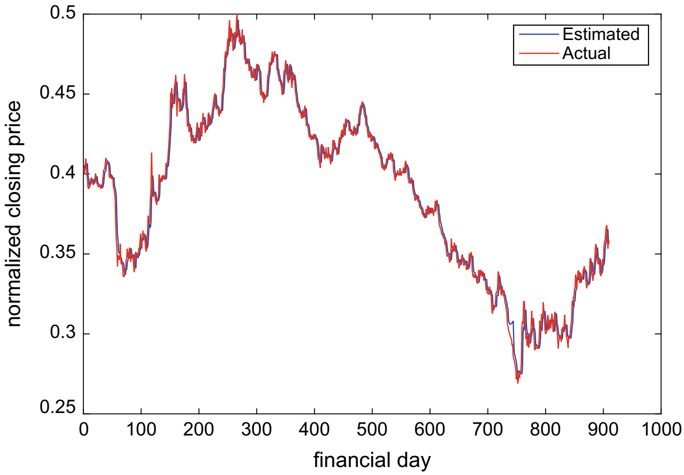


Fig. 15 Actual versus estimated daily closing prices by J-PSNN from HSI dataset

exchange, Hong Kong stock, Bombay stock, and Dow Jones Industrial Average. The performance of the model was compared with that of four state-of-the-art models, i.e. MLP, GD-PSNN, GA-PSNN, and PSO-PSNN and found quite satisfactory. Therefore the proposed forecast can be adopted as an efficient tool for financial instrumentation. This work further can be extended using other sophisticated neural networks and competent optimization techniques.

References

1. Liu H-C, Lee Y-H, Lee M-C (2009) Forecasting China stock markets volatility via GARCH models under skewed-GED distribution. J Money Invest Bank, 5–14
2. Zhang GP (2003) Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50(2003):159–175
3. Ghazali R, Hussain AJ, Nawi NM, Mohamad B (2009) Nonstationary and stationary prediction of financial time series using dynamic ridge polynomial neural network. J Neurocomput 72:2359–2367
4. Nayak SC, Misra BB, Behera HS (2016) Fluctuation prediction of stock market index by adaptive evolutionary higher order neural networks. Int J Swarm Intell 2(2–4):229–253
5. Nayak SC (2017) Development and performance evaluation of adaptive hybrid higher order neural networks for exchange rate prediction. Int J Intell Syst Appl 9(8):71
6. Nayak SC, Misra BB, Behera HS (2016) An adaptive second order neural network with genetic-algorithm-based training (ASONN-GA) to forecast the closing prices of the stock market. Int J Appl Metaheuristic Comput (IJAMC) 7(2):39–57
7. Nayak SC, Misra BB, Behera HS (2015) A pi-sigma higher order neural network for stock index forecasting. In: Computational intelligence in data mining, vol, 2. Springer, New Delhi, pp 311–319
8. Nayak SK, Nayak SC, Behera HS (2016) Evolving low complex higher order neural network based classifiers for medical data classification. In: Computational intelligence in data mining, vol 2. Springer, New Delhi, pp 415–425
9. Nayak SC, Misra BB, Behera HS (2017) Adaptive hybrid higher order neural networks for prediction of stock market behavior. In: Nature-inspired computing: concepts, methodologies, tools, and applications. IGI Global, pp 553–570
10. Nayak SC, Misra BB, Behera HS (2018) On developing and performance evaluation of adaptive second order neural network with GA-based training (ASONN-GA) for financial time series prediction. In: Advancements in applied metaheuristic computing. IGI Global, pp 231–263
11. Nayak SC, Misra BB, Behera HS (2017) Exploration and incorporation of virtual data positions for efficient forecasting of financial time series. Int J Ind Syst Eng 26(1):42–62
12. Haykin SS, Haykin SS, Haykin SS, Haykin SS (2009) Neural networks and learning machines, vol 3. Pearson, Upper Saddle River, NJ, USA
13. Rajasekaran S, Pai GV (2003) Neural networks, fuzzy logic and genetic algorithm: synthesis and applications (with cd). PHI Learning Pvt. Ltd
14. Wang Z, Fang JA, Liu X (2008) Global stability of stochastic high-order neural networks with discrete and distributed delays. Chaos, Solitons Fractals 36(2):388–396
15. Ghosh J, Shin Y (1992) Efficient higher-order neural networks for classification and function approximation. Int J Neural Syst 3(04):323–350
16. Rao R (2016) Jaya: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems. Int J Ind Eng Comput 7(1):19–34
17. Rao RV, More K, Taler J, Ocłoń P (2016) Dimensional optimization of a micro-channel heat sink using Jaya algorithm. Appl Therm Eng 103:572–582
18. Nayak SC, Misra BB, Behera HS (2014) Impact of data normalization on stock index forecasting. Int J Comp Inf Syst Ind Manag Appl 6:357–369