Anurag Sharma

4th year student
5 year Dual Degree B.Tech IT and M.Tech Software Engineering
IIIT Allahabad

ism2013018@iiita.ac.in
+91-7704914402
anurageldorado (Skype)
aedorado (Linkedin)
aedorado (Github)

Summary

To augment my existing theoretical knowledge and gain practical insights to provide a base for my future research endeavours and channelize my time into research activities.

Education

5 year Dual Degree B.Tech Information Technology and M.Tech Software Engineering (2013 - 2018) CGPA: 9.00 (7 SEMESTERS)
Indian Institute of Information Technology, Allahabad

ISC (2010 - 2012) 95.25%
St.Joseph's Academy

ICSE (2008 - 2010) 95%
St.Joseph's Academy

Academic Projects

P2P File Sharing GUIDES: PROF. S. SANYAL, DR. BIBHAS GHOSHAL
DISTRIBUTED SYSTEMS Jan '17 – ongoing

The project envisions a P2P file sharing system, where a super-peer indexes the shared files while peers can search the index and transfer files amongst themselves. Some features envisioned are replication, caching and incentives/penalties.
The project is in nascent stage of development.

Citation Network Analysis GUIDES: DR. JAGPREET SINGH, DR. BIBHAS GHOSHAL
BIG DATA, SOCIAL NETWORK ANALYSIS, Hadoop, HBase, Java Jan '16 – Dec '16

The project involved 100GB of data of the Microsoft Academic Database containing data about authors, publications and citations of publications.
The data was first filtered to include only those papers related to the Computer Science fields lending a sub-network of 5 million nodes and 15 billion edges. Each of these publication was classified into 24 research communities(Eg. Algorithms, Networks etc.). The complex network of publication citations was analyzed to calculate various metrics for the citation network which included: Expansion, Cut Ratio, Fraction over Median Degree, Conductance, Flake Odf and Inwardness.
Also the metrics were calculated in a temporal (by taking subsets which contained paper of a particular time period) and a geographic (by geotagging a paper's author) manner.
C^3-index for the authors was also calculated by using 3 networks: publication citation network, author-author citation network and the co-authorship network.
Hadoop was used for all processing tasks and HBase was used as database. The hadoop cluster had 350GB of RAM and 40 Vcores.

Dynamic Aspect-based Semantic Analysis on Review Sets Guide: Dr. Ranjana Vyas

NATURAL LANGUAGE PROCESSING, NLTK, Python3, Java8 Aug '15 – Nov '15

This is a system that can automatically summarize opinions from a set of reviews and present them in an easy to process manner.

For this projects, a large amount of reviews were extracted(scraped) from websites like amazon, flipkart etc. to objectively provide ratings to the concerning products and provide summaries for the same based on aspects of the product/service.

A lexicon for the analyzer is constructed using seed words and wordnet. This lexicon is appropriately scored using an iterative algorithm.

The reviews were first tagged into parts of speech using the stanford pos-tagger and the most frequently occurring noun and noun compounds are extracted as dynamic features of the product/service. The related adjectives and adverbs the describe these aspects are then scored using the aforesaid lexicon.

The most semantically high scoring sentences about an aspect are included in the summary for the aspect.

Shelter Animal Outcome Guide: Dr. K.P. Singh

MACHINE LEARNING, Scikit-learn, Python3 Mar '16 – Apr '16

The project involved a data set related to an animal shelter of stray cats and dogs containing features such as age, gender, breed, etc. along with 5 outcomes such as return to owner, transfer etc.

Various visualizations were performed on the dataset using seaborn and matplotlib.

After data cleaning and pre-processing, various algorithms were used for classification: Naive Bayes, Decision Tree, Random Forest, KNN, SVM and Neural Network.

Techniques like cross-validation, pipe-lining, gird-search, etc. were used to enhance the accuracy. The average accuracy obtained was about 62% and the results were satisfactory.

License Plate Detection Guide: Prof. Anupam Agarwal

IMAGE PROCESSING, Matlab, OpenCV, Python3 Mar '16

License Plate Detection was performed using various algorithms in MATLAB and a few in OpenCV. The algorithms used were: Mathematical morphology, Sobel operation, Histogram Processing, Top-hat transformation and Canny-edge detection. Bounding box was used to segment the characters on the number plate.

The dataset used was obtained from NTUA Media Lab website. An average accuracy of 71% was obtained. The highest accuracy of 74% was obtained by the Sobel operator.

Sybil Attack Detection Guide: Prof. Jagpreet Singh

WIRELESS SENSORS, NS2, C++, TCL Nov '15

Sybil attack is an attack where the attacker (sybil node) tries to forge multiple identification in a certain region. This project is a Received Signal Strength Indicator (RSSI) based scheme for Sybil Attack detection in wireless sensor networks with a hundred percent detection rate and only a few false positives, based on empirical evidence. The technique emphasises the use of 2 instead of 4 detectors for detection of Sybil Attack in Wireless Sensor Networks and uses ration of RSSI to detect Sybil Attack.

The algorithm was implemented in C++and a demonstration presented in NS2.

Software Development Projects

Examcell @ IIITA (Guided by: Dr. B Sanjeev, Dr. S Venkateshan) May '16 – Sept '16

Developed Exam Cell Software which is currently installed and working in IIIT, Allahabad in a team of 5. The software basically autmates the manual tasks of the employees of the exam cell in IIIT Allahabad such as generation of transcripts, grade cards, declaration of results,

admitting new students etc. The software comprised 40 tables and more than 10K SLOC. PHP was used for backend.

C-PROXY in C++. Nov '15
A proxy server built in C++.

Chat Application in Java Oct '15
Multi-client, encrypted group chat application build in Java.

Productivity Chrome Extensions Oct '16, Oct '15
Exty-Switchy - Easy control panel for chrome extensions.
Redirecto - Redirect one url to another.

Skills

Programming & scripting languages: *C/C++, Java, Python, HTML/CSS, Javascript, PHP, SQL, Bash.*

Frameworks: *Hadoop, OpenCV, Django, Reactjs, jQuery.*

Tools: *HBase, Sqlite3, MySQL, Git, Sublime Text, Linux, AWS, Apache, Latex.*

Natural languages: *Hindi* (mother tongue), *English* (full professional proficiency) *and Spanish* (beginner).

Coursework

Introduction to Programming	*Basic Electronics*	*Digital Electronics*
Data Structures	*Discrete Maths*	*Computer Organization & Architecture*
Operating Systems	*Microprocessor*	*Object Oriented Methodologies*
Theory of Computation	*Computer Graphics*	*Design & Analysis of Algorithms*
Probability & Statistics	*Software Engineering*	*Principles of Programming Languages*
Artificial Intelligence	*Computer Networks*	*Database Management Systems*
Compiler Design	*Image & Vision Processing*	*Natural Language Programming*
Simulation & Modelling	*Software Metrics*	*Machine Learning Tools & Techniques*
Maths for IT	*Data mining & Warehousing*	*Software Requirements & Estimation*
Distributed Systems	*Analysis of Software Systems*	*Software Testing & Quality Management*

Positions of Responsibility

Prayaas (Co-ordinator), 2014-2016: *Coordinator at a local NGO, Prayaas, dedicated to educating children of poor slums. Responsibilities involved managing volunteers, organizing classes for underprivileged children.*

Hack in the North (Lead Organizer), 2015 - present: *Lead organizer at Hack in the North, the largest student held hackathon in India.*

Web Development Wing, Geek Haven Technical Society (Co-ordinator), 2014-2015: *Responsibilities involved conducting lectures and tutorials and development of major portals of the institute.*

Interests

Non-exhaustive and in alphabetical order: *Automation, Economics, Gaming, History, Music, Open Source, Philanthropy, Philosophy, Psychology, Reading, Spirituality and Travel*