

**DEPARTMENT OF COMPUTER SCIENCE
IGDTUW**



**MAJOR PROJECT ABSTRACTS
BTech (8TH SEM)**

Compiled by

**Zeenat(PhD Scholar), Manasi, Amrita Srivasatava, Amrita Joshi, Jithu,
Nicy, Kirti (MTech-MPC)**

Index: Abstracts of B.Tech Final Year

| S.No | Name of the Project | Student Name | Guide Name | Page No |
|------|--|--|---------------------------------------|---------|
| 1 | Distributed Denial Of Service Attack | Priyanka Choudhary, Damini Sinha, Preeti | Monika Choudhary | 5 |
| 2 | Intrusion detection system | Kirti Dhir, Ishmeet Kaur | Indra Thanaya | 6 |
| 3 | Educated youth unemployment analysis using social media | Jyoti saini Aayushi Khandelwal | Dr.Devendra Tayal | 7 |
| 4 | Sentiment analysis and opinion mining on government policies | Himani and Jaya Sharma | Ms. Monica Choudhary | 8 |
| 5 | A new method for answer-script evaluation based on type-2 fuzzy sets | Anshul Mehra, Astha Singh, Leena Varghese | Dr.Devendra Tayal | 9 |
| 6 | A novel secure edge based image Steganography | | | 10 |
| 7 | Text Summarization Tanvi Tanwar, Shilpi | Tanvi Tanwar, Shilpi | Dr.Seeja | 11 |
| 8 | Data/File Transfer between two devices through Light(Li-Fi Technology) | Meenu, SadhanaYadav, Megha Sagar | Mrs.vibha | 12 |
| 9 | User preference analysis of online blogging site (mashable.com) using big data | Neha Chaurasia,sakshi Singhal, Divyasha | Mr. Sumit Yadav/ Mrs. Kavita Sachdeva | 13 |

| | | | | |
|----|---|---|---|----|
| 10 | Perplexed Bayes Classification of tweets in context to Terrorism | Shubhita Garg, Neha Kakkar | Mr. Sumit Yadav | 14 |
| 11 | Customization of Open Source UI Framework (PubNub EON) for Sensor Interface in Smart Cities | Neerali ,Nikita Mohan , Surabhi Santra | Dr. S.R. N. Reddy | 15 |
| 12 | Solution for the problems in face detection techniques | Himani Garg & Ravita Kumari | NAJME ZEHRA (Assistant professor) | 16 |
| 13 | Emotion re recognition using facial expressions | Shivangi shori,sweta dhar,megha | Dr. S.R. N. Reddy/ Ms. Monica Choudhary | 17 |
| 14 | Energy efficient protocol for wireless sensor network | Ankita rustagi,vinny rana,yashika kashyap | Dr.S.R.N.Reddy/Mr.vivekanand jha | 18 |
| 15 | Analysis of fake accounts on facebook | Anjali,zulfitaha | Ms.vibha | 19 |
| 16 | Currency recognition using image processing | Ila amar,priyanka kumari | Ms. Monica Choudhary | 20 |
| 17 | Privacy preserving data mining with 3-d rotation transformation | Somya upadhaya,chetna Sharma,pravishti Sharma,prachi bhardwaj | Mrs. Seeja | 21 |
| 18 | Detection of wormholes in wireless sensor networks | Srishti marwah,shivani tripathi,somya goyal | Mr.vivekanand jha | 22 |
| 19 | Building feature rich parts of speech tagger for a morphologically complex macronic language:hinglish | Richa malhotra,mansi mittal | Mr. Sumit yadav | 23 |
| 20 | Dimension reduction of multivariate non-linear data | Nisha sinha,chitra baghel,manisha Sharma,manju | NAJME ZEHRA (Assistant professor) | 24 |
| 21 | Home alert system | Pooja dahiya,neha | Dr.S.R.N.Reddy | 25 |
| 22 | Probability distribution function to assist compressive sensing in | Mansi tripati,shubhangi mahajan | Mr.vivekanand jha | 26 |

| | | | | |
|----|--|---|-----------------------------------|----|
| | wsn | | | |
| 23 | Election opinion analysis using a hadoop system. | Prachi mahajan,prachi aggarwal | Prof. Ela kumar | 27 |
| 24 | Blackhole detection in manet | Bhawana khadgi,silvy aggarwal,surbhi bhalla | Najme zehra (assistant professor) | 28 |
| 25 | Diagonis system for thyroid in Parkinson disease | Rohini parle,akanksha,puja,avnee | Ms. Monica Choudhary | 29 |
| 26 | Automated tool for teacher feedback using opininon mining and fuzzy logic | Leena mangla,amitoj kaur | Dr.Devendra Tayal | 30 |
| 27 | Internship at expedia ,inc.- code automation tool for mvc layers of spring and design pattern implementation for restful web service calls | Himani bajaj | Ms. vibha | 31 |
| 28 | Predictive analysis in healthcare | Gursift kaur,taniya,simrati | Prof. Ela kumar | 32 |
| 29 | Hybrid approaches to software testing and its practical applications | Mansi singhal | Mr Indra Thanya | 33 |
| 30 | Environment allocater | Aditi goel | Mr.Sumit yadav | 34 |

Distributed Denial Of Service Attack

Submitted By: Priyanka Choudhary, Damini Sinha, Preeti

Under the guidance of

Monika Choudhary

ABSTRACT

Denial-of-service (DoS) attacks occur when the attacks are from a single source, whereas Distributed Denial-of-service (DDoS) attacks occur when many compromised systems flood the resources or bandwidth of a target system. We have shown HTTP flooding by making a service server and service client class. The client are the infected bots which will affect further systems. One command and controller is used as a botnet master. After starting rmi registry, connection is established with client and make it service bound. Infecting bots start flooding the target with http requests. For detection, state analyzer is used for analysing the system, Syn requests are continuously sent and If acknowledgment is returned, then the system is active else down. Through this, monitoring of network is done.

KEYWORDS: DoS attack, DDoS attack, HTTP flooding, state analyzer, botnet, RMI registry, infecting bots.

INTRUSION DETECTION SYSTEM

Submitted By: Kirti Dhir, Ishmeet Kaur

Under the guidance of

Indra Thanaya

ABSTRACT

Providing security in an exceedingly distributed system needs over user authentication with passwords or digital certificates and confidentiality in information transmission. Distributed model of cloud makes it vulnerable and at risk of subtle distributed intrusion attacks like Distributed Denial of Service (DDOS) and Cross website Scripting (XSS). To handle massive scale network access traffic and body management of information and application in cloud, a replacement multi-threaded distributed cloud IDS model has been planned. Our planned cloud IDS handles massive flow of information packets, analyze them and generate reports with efficiency by integration data and behaviour analysis to sight intrusions.

Today, several organizations are moving their computing services towards the Cloud. This makes their pc process out there rather more handy to users. However, it additionally brings new security threats and challenges regarding safety and dependableness. In fact, Cloud Computing is a lovely and cost-saving service for consumers because it provides accessibility and dependableness choices for users and ascendable sales for suppliers. In spite of being engaging, Cloud feature poses numerous new security threats and challenges once it involves deploying Intrusion Detection System (IDS) in Cloud environments. Most Intrusion Detection Systems (IDSs) are designed to handle specific forms of attacks. It's evident that no single technique will guarantee protection against future attacks. Hence, there's a necessity for an integrated theme which may offer strong protection against a whole spectrum of threats.

Cloud computing refers to the availability of machine resources on demand via a electronic network. Users or shoppers will submit a task, like data processing, to the service supplier, like Google, while not really possessing the specified computer code or hardware. The consumer's pc could contain little or no computer code or information (perhaps a minimal software package and browser only), serving as very little over a show terminal connected to the net. Since the Cloud is that the underlying delivery mechanism, Cloud primarily based applications and services could

support any form of computer code application or service in use nowadays. The essential characteristics of Cloud Computing embrace On-demand self-service that allows users to consume computing capabilities (e.g., applications, server time, and network storage) as and once needed. Resource pooling that permits combining computing resources (e.g., hardware, software, processing, network bandwidth) to serve multiple customers - such resources being dynamically allotted. Measured provision to optimize resource allocation and to supply a metering capability to work out usage for charge functions Extension to existing hardware and application resources, thus, reducing the price of extra resource provisioning. The cloud isn't merely the newest trendy term for the net.

Keywords:DDOS, IDS, Cloud Computing, server

EDUCATED YOUTH UNEMPLOYMENT ANALYSIS USING SOCIAL MEDIA

Submitted By: Jyoti saini (0041350712), Aayushi Khandelwal (04313502712)

Under the guidance of

Dr.Devendra Tayal

Associate Professor ,CSE

ABSTRACT

In India, there has been an alarming phase of mass-unemployment among the educated youth. Post-graduates and graduates are walking pillar to post in search of employment. The number of unemployment youths in our country is increasing every year.

Many researchers have revealed facts about the E-Recruitment and Social Media trends in India with increasing internet usage among youth and what benefits are experienced by the organisations using these online-recruitment methods.

There are thousands of resumes available on the job portals of IT professionals. Over 120 million professionals use social network LinkedIn to exchange information, ideas and opportunities .This tells us, how the Internet is bringing radical change to corporate recruiting.

This project addresses the problem of educated unemployment in youth and analyses it on the basis of parameters such as gender, location, industry, education, age groups using fuzzy logic and various data mining techniques. It will help Govt. Of India as well as different public and private sectors to increase the employment opportunities.

Key words: LinkedIn, Social media, Unemployment, Big Data Analysis, Fuzzy logic, Data mining ,naukri.com

SENTIMENT ANALYSIS AND OPINION MINING ON GOVERNMENT POLICIES

Submitted By: Himani and Jaya Sharma

Under the guidance of

Ms. Monica Choudhary

ABSTRACT

Online Social Networks (OSNs) are deemed to be the most sought-after societal tool used by the masses world over to communicate and transmit information. Our dependence on these platforms for seeking opinions, news, updates, etc. is increasing. While it is true that OSNs have become a new medium for dissemination of information, at the same time, they are also fast becoming a playground for the spread of misinformation, propaganda, fake news, rumors, unsolicited messages, etc. The trustworthiness of online opinions has been neglected most often. There is no reported study on assessing the trustworthiness of reviews, which is crucial for all opinion based applications. Consequently, we can say that an OSN platform comprises of two kinds of users namely, Spammers and Non-Spammers. Spammers, out of malicious intent, post either unwanted (or irrelevant) information or spread misinformation on OSN platforms. As part of our work, we propose mechanisms to detect such users (Spammers) in Twitter social network. The trustworthiness of the reviews is assessed as spam or a non spam review which includes both duplicate and near duplicate reviews classified as spam reviews, and unique reviews classified as non spam reviews. We propose a novel and effective technique, namely, Conceptual level similarity measure used for detecting spam reviews based on the policy features that have been commented in the reviews. The efficiency of the task of web based customer review spam detection can be enhanced by identifying and eliminating duplicate and near duplicate spam reviews, thereby providing a summary of the trusted reviews for customers to make buying decisions.

KEYWORDS: Customer reviews, Conceptual level similarity measure, Feature extraction, Web mining.

A new method for answer-script evaluation based on type-2 fuzzy sets

Submitted By: Anshul Mehra, Astha Singh, Leena Varghese

Under the guidance of

Dr. Devendra Tayal

ABSTRACT

In this project, a new method for evaluating students' answer scripts based on fuzzy grade sheets is presented. The evaluator assigns scores to each answer using three criteria, namely, 'complexity', 'importance', 'accuracy rate'. Fuzzy marks, represented by type-2 fuzzy sets, are generated from these assigned values. The degree of similarity between the graded type-2 fuzzy mark and a set of standard type-2 fuzzy criteria is calculated using an equality operator. The degree of optimism of the evaluator is indicated by the index of optimism, λ , where $\lambda \in [0,1]$. The normalized grade point is calculated for each answer. The total percentage is calculated using the derived grade point and the maximum marks for each answer. A web application implementing this evaluation system has also been created. Experimental results show that the proposed method is more accurate and precise than recent methods for students' answer scripts evaluation.

KEYWORDS: Fuzzy grading system, Index of Optimism, Equality Operator, Students' Evaluation

A novel secure edge based image Steganography

ABSTRACT

Security of information is one of the important issues of digital communication. Apart from encryption, steganography has been one of the solutions to ensure secure data transmission over the network. Steganography is the transmission of messages from a sender to a receiver with the intention that third party cannot reliably detect the secret communication between sender and receiver. A novel secure edge based image steganography technique for 24 bit RGB images has been proposed. The proposed method first encrypts the secret data using encryption algorithm and then the encrypted data is embedded in the edge pixels of the cover image in the order of strongest to weakest edge pixels depending on message length. Edges are used as an embedding position as they possess sharp change in visual and statistical properties. To make interception of secret message difficult, XOR between the green component of pixel and mutually decided secret key has been used as location deciding factor among red and blue components of pixel. Eyes being more sensitive to green, data has been stored in red and blue components of pixel. With the two level security, the proposed method aims to protect data from unauthorized access and making its interception by any intruder more difficult.

Text Summarization

Submitted By: Tanvi Tanwar, Shilpi

Under the guidance of

Dr. Seeja

ABSTRACT

Automatic summarization is the process of reducing the text document with the computer program in order to create a summary that retains the most important points of the original document. There are two approaches to automatic summarization: Extraction and Abstraction. Extractive methods work by selecting a subset of existing sentences in the original text to form the summary. In contrast, abstractive methods build an internal semantic representation and then use natural language generation techniques to create a summary. We have implemented an extraction based text summarization approach on the basis of term ranking. Ranking is done through different algorithms - position based algorithm, numerical based algorithm, keyword extraction and calculation of Tf-Idf. Keyword extraction is done using the software weka. Tf-idf is a numerical statistic that is intended to reflect how important a word is to a text. After applying all the above mentioned algorithms and approaches, we are getting the most important sentences of the text document as our result.

KEYWORDS: Text Summarization, Tf-Idf

Data/File Transfer between two devices through Light(Li-Fi Technology)

Submitted By: Meenu, SadhanaYadav, Megha Sagar

Under the guidance of

Mrs. Vibha

ABSTRACT

Li-Fi is a VLC (Visible Light Communication) technology. It will show how li-fi takes VLC further by using light emitting diode (LEDs) to realise fully network wireless systems. Synergies are harnessed as luminaries become Li-Fi autocells resulting in enhance wireless capacity providing the necessary connectivity to realise the Internet of things, controlling to for the fifth generation of cellular systems(5G) and beyond. It covers all of the key research areas from Li-Fi /wireless fidelity (WiFi) networks to illustrate that Li-Fi attocells are not a theoretical Concept anymore, but at the point of real world deployment.

Keywords: Communication systems, Communication network, Light emitting diodes(LEDs), Wi-Fi (wireless Fidelity), VLC (Visible Light Communication).

USER PREFERENCE ANALYSIS OF ONLINE BLOGGING SITE (mashable.com) USING BIG DATA

Submitted By: Neha Chaurasia(00813502712), Sakshi Singhal(03613502712) , Divyasha
(08613502712)

Under the guidance of
Mr. Sumit Yadav/ Mrs. Kavita Sachdeva

ABSTRACT

Big data is a term for data sets that are so large or complex that traditional data processing applications are inadequate. To determine user preference information from big data recorded by blogging sites, this project analyzed big data sourced from Mashable.com (a leading source for news & information for the Connected Generation). To understand the project further, we reviewed research papers on Big Data, Hadoop, Hive, Hue, Map-Reduce, Cloudera Platform, Blogging. Sites and their social impacts. For implementation, a range of values was extracted and an algorithm was devised (using MapReduce technology) which was run onto the database. Through the results obtained, we observed that majority of viewers who follow these posts prefer more text over media, the optimum number of words in the post title should be in the range of 7-13, and links, 0-4. The small figures (100-299 words per article) show that readers prefer news in a concentrated format. We also concluded that readers read news more during mid-weekdays, and go for 'World News' genre over others. An overwhelming number of successful posts (83%) were positive in nature.

KEYWORDS: Mashable.com, Social Media Newsfeed, Hadoop, Big Data

Perplexed Bayes Classification of tweets in context to Terrorism

Submitted By: Shubhita Garg, Neha Kakkar

Under the guidance of

Mr Sumit Kumar Yadav

ABSTRACT:

Social Media has become a tool in the hands of Terror Groups. A large repository of information is posted on such platforms on daily basis. Our approach analyse a subset of this data, viz the Tweets posted on Twitter. Our method uses two approaches, namely, Naives Bayes Classifier and Perplexed Bayes Classifier, to categorize tweets into three broad categories of: Towards, Against and Neutral in the context of terrorism. The analysis is done based on a training dataset that contains tweets which are manually categorized in these three categories. The knowledge from the training dataset is used to find the inclination of the tweet towards terrorism. A comparative study of both the approaches is also shown. The results observed shows that for same training set and **same set of tweets, Perplexed Bayes classifier performed better than Naive bayes classifier.**

KEYWORDS: Social Media, Terrorism, Twitter, Tweets, Naives Bayes Classifier, Perplexed Bayes Classifier.

Customization of Open Source UI Framework (PubNub EON) for Sensor Interface in Smart Cities

Submitted By: Neerali (06613502712), Nikita Mohan (06713502712), Surabhi Santra (07513502712)

Under the guidance of

Dr. S.R. N. Reddy

ABSTRACT:

PubNub EON is an open-source graphical user interface framework (employing charts and maps) for the embedded world, from Raspberry Pi, to Arduino, to Atmel MCUs, enabling users to collect sensor readings and visualize the data readings on a live-updating UI. Live visualization of Realtime data has been made easy by creating a better user experience for research based experiments in IGDTUW embedded labs. With application-specific customization of PubNub EON framework, it expands the control capability of Raspberry Pi interfaced home-automated sensors , digital maker projects, music machines, parent detectors to easily create a custom user interface program running on an attached Raspberry Pi, with the help of simple html codes and C3js charting scripts.. During customization, mapbox widgets and other map themes based on geographical locations have been reduced from the package.

Keywords : Open-source, Graphical User Interface, Raspberry Pi, Temperature & Humidity sensor interfacing, realtime dashboard, framework, customization

SOLUTION FOR THE PROBLEMS IN FACE DETECTION TECHNIQUES

Submitted By: Himani Garg & Ravita Kumari

Under the guidance of

NAJME ZEHRA (Assistant professor)

ABSTRACT

In past few years, Face detection is one of the most studied, basic, important and challenging topic in computer vision literature. There have been very considerable developments in the field of detection of faces in various directions since its inception early in the area of image processing and computer vision. The previous algorithms have many problems like: - detection of black person images, lighting condition, brightness, problem with overlapping sub window, problem with variation with angle and poses. This paper proposed a new algorithm, which minimize problems with previous algorithm and both the false negative, false positive rates in order to achieve an acceptable performance.

KEYWORDS: Adaboost algorithm, Hausdorff distance, Cascade classifier.

EMOTION RECOGNITION USING FACIAL EXPRESSIONS

Submitted By: Shivangi Shory, Shweta Dhar, Megha

Under Guidance of

Dr. S.R.N. Reddy & Ms. Monica Choudhary

ABSTRACT

Automatic recognition of spontaneous facial expressions is a major challenge in the field of affective computing. Head rotation, face pose, illumination variation, occlusion etc. are the attributes that increase the complexity of recognition of spontaneous expressions in practical applications.

Here, we have described emotion recognition system based on facial expressions. A fully automatic facial expression recognition system is based on three steps: face detection, feature extraction and facial expression classification. Viola - Jones algorithm has been used to detect the face in the image which is further pre-processed to remove inconsistencies. Now feature extraction is carried out using HOG (Histogram of Gradients) and PCA (Principle Component Analysis) techniques. Furthermore classification of images is carried out using K-Nearest Neighbor and Neural Network techniques.

The four possible combinations of techniques are carried out and the best possible combination is then selected based on the degree of accuracy of the combinations. This emotion recognition system is then embedded in an application which uses the emotion of the image to execute its functionality.

KEYWORDS: Viola-Jones, Feature Extraction, HOG, PCA, K-Nearest Neighbor, Neural Network, Emotion Recognition

Energy Efficient Protocol For Wireless Sensor Networks

Submitted By: Ankita Rustagi, Vinny Rana, Yashika Kashyap

Under the guidance of

Dr. S.R.N. Reddy, Mr. Vivekanand Jha

ABSTRACT

A Wireless Sensor Network consists of a sink and randomly distributed sensor nodes over an area of interest. The sensor nodes generate the data packet periodically which must be delivered to the sink within a specific period of time. The sensor nodes near the sink deplete their energy quicker than other nodes in the network leading to an energy hole problem. The sensor nodes collect the data from neighbours and aggregate it to the sink. An inefficient data aggregation mechanism leads to data delivery latencies and data redundancy. In this protocol, we propose a virtual ring for sink mobility pattern and cluster heads selection on the basis of cost factor. The sensor nodes send their data to the associated cluster head. During Data Aggregation among the cluster heads, a bi-parted tree is constructed and cluster heads are scheduled.

KEYWORDS: Wireless Sensor Networks, energy hole problem

Analysis of fake accounts on facebook

Submitted By: Anjali(06813502712)

Zulfitaha(07613602712)

**Under the guidance of
Ms. Vibha**

ABSTRACT

In present time, social life of everyone has become linked with the online social networking sites. These sites have made a severe change in the way we pursue our social life. Making friends and keeping in contact with them and their updates has become easier. But with their rapid growth, many problems like fake profiles, online impersonation have also grown. There are no feasible solution exist to control these problems. In this project, we came up with a framework with which automatic detection of fake profiles is possible and is efficient. This framework uses classification techniques like Support Vector Machine, Naive Bayes and Decision trees to classify the profiles into fake or genuine classes. As, this is an automatic detection method, it can be applied easily by online social networks which has millions of profile whose profiles can not be examined manually.

KEYWORDS: Online Social Networks (OSN)

Currency Recognition Using Image Processing

Submitted By: Ila Amar ,Priyanka Kumari
Under the guidance of

Monika Choudhary

ABSTRACT

Currency has great importance in day to day life and may be because the currency recognition is a great area of interest for researchers. Image processing is the most popular and effective method of currency recognition. Image processing based currency recognition technique consists of few basic steps like image acquisition, its preprocessing and finally recognition of the currency. The software we are using is MATLAB. The phases are as follows: 1. Read image, the format of the image is JPEG or PNG. 2. Pre-processing, removing noise, smoothing image. 3. Image process, edge detection, segmentation, pattern matching. 4. Results printing. We are giving our output in the form of voice or speech, which can be an effective tool for visually challenged people.

KEYWORDS: Currency Recognition, Image Processing, Image Acquisition, MATLAB, Pre-processing, Smoothing Image, Edge Detection, Segmentation, Pattern Matching.

Privacy Preserving Data Mining with 3-D Rotation Transformation

Submitted By: SomyaUpadhyay ,Chetana Sharma, Pravishti Sharma, Prachi Bharadwaj

Under the guidance of

Seeja K R

Abstract

Data perturbation is one of the popular privacy preserving data mining techniques. A major challenge in data perturbation is to balance privacy protection and data utility, which are normally considered as a pair of conflicting factors. This paper proposes a Geometric Data Perturbation (GDP) method using data partitioning and three dimensional rotation. In this method, attributes are divided into groups of three and each group of attributes are rotated about different pair of axis. The rotation angle is selected such that the variance based privacy metric is high which makes the original data reconstruction difficult. As many data mining algorithms like classification and clustering are invariant to geometric perturbation, the data utility is preserved in the proposed method. The experimental evaluation shows that the proposed method provides good privacy preservation results and data utility compared to the state-of-the-art techniques.

Keywords: Data perturbation, variance, three dimensional rotation, privacy preserving, data mining

TITLE: DETECTION OF WORMHOLES IN WIRELESS SENSOR NETWORKS

Submitted By: Srishti Marwah (06213502712), Shivani Tripathi (08413502712), Soumya Goyal (08513502712)

Under the guidance of

Mr. Vivekanand Jha

ABSTRACT:

Wireless Sensor Network (WSN) is an evolved technology that shows great promise for various futuristic applications both for military and mass public. The sensing technology combined with processing power and wireless communication makes it lucrative for being exploited in abundance in future. The inclusion of wireless communication technology also incurs various types of security threats like wormhole, sinkhole, black hole, Sybil, selective forwarding attacks.

We present a secure and efficient scheme to overcome one of the attacks; wormhole attack in WSN. It aims at maximizing wireless sensor security from wormhole attacks, by blacklisting the malicious nodes (wormholes). Various types of wormhole attacks are hidden wormholes, exposed wormholes, out of band wormholes, wormholes using packet encapsulation.

We have proposed an algorithm that focusses OUT OF BAND wormholes. It includes the concept of distance based neighbor discovery and a counter based scheme to detect wormholes in the network

KEYWORDS: Sensors, wormholes, wireless

Building Feature-Rich Parts-of-Speech Tagger for a Morphologically Complex Macaronic Language: HINGLISH

Student Name: Richa Malhotra (02113502712), Mansi Mittal (02913502712)

Guide Name: Sumit Yadav

Abstract

The word Hinglish refers to a combination of English and Hindi which is becoming increasingly common in India and Indian communities throughout the world. Hence the growing influence of Hinglish and taking into the account for the future of the language, we aimed to build a POS tagger for Hinglish which encapsulated the properties of wildly spoken languages namely Hindi and English. We have used the Maximum Entropy based methods, which can deal with diverse, overlapping features. Maximum Entropy is a very flexible method of statistical modeling which handles the sparse data problem. Under this model, a natural combination of several features can be easily incorporated. The accuracy of our approach is 98.75% for known words and 95.63% for unknown words. The model described is very simple and is effective for automatic tagging even when the amount of available labeled text is small.

Keywords: POS tagging, Hinglish, ME, Supervised machine learning

DIMENSION REDUCTION ON MULTIVARIATE NON-LINEAR DATA

**Submitted By: NISHA SINHA, CHITRA BAGHEL, MANISHA SHARMA & MANJU
LATA SORENG**

**Under the guidance of
NAJME ZEHRA NAQVI**

ABSTRACT

Real life datasets have large number of dimensions dependent on each other representing different features of the dataset. 'The curse of dimensionality' is a major factor affecting performance of any big data analysis algorithms. Therefore, before applying any prediction or computation algorithm, it is necessary to project it to a new feature space with independent and reduced number of dimensions. The Principal Component analysis (PCA) is one of the basic techniques to reduce dimensions based on Pearson's Product Moment Correlation. This method works fine for linearly related features but it cannot capture non-linear relationships among them. To address this drawback, we propose a modification in the correlation coefficient calculation of PCA, such that it is capable of finding dependence of any order among the dimensions and hence allows PCA to reduce more dimensions as compared to standard PCA. **KEYWORDS:** Multivariate; PCA; Pearson's Product moment correlation; Spearman's rank order correlation; Distance correlation; Eigen value decomposition

HOME ALERT SYSTEM

Submitted By:Pooja Dahiya, Neha

Under the guidance of

Dr. SRN Reddy

ABSTRACT

We are living in an hi-tech era where crime is increasing day by day. We want to be safe and at the same time we also want to protect our home assets from any kind of hazards. Knowing our home is protected provides peace of mind when we are away. The customers require simple, reliable and high performance core system that can be easily implemented.

The major concern of this project is to design a budget Home Alert System based on wireless sensor network using wi-fi and cloud technology. It can detect the theft, fire, leakage of gas, smoke and send alarm message remotely. This is done by uploading all the sensor's data to cloud so that it can be further accessed from an android application. In this project, we survey the current work on security system and applications. We examine the existing work, which is held by using different sensors and contributes to better understanding of the challenges in existing work on security system and further research direction. In this project we have taken an overview on how to alert home's owner remotely in case of accidental miss happenings like fire, theft and safety issues.

Key words: Home Security, Microcontroller, wifi, cloud, Android Application, sensors.

Probability Distribution Function to Assist Compressive Sensing in Wireless Sensor Networks

Submitted By: Mansi Tripathi (03813502712), Shubhaangi Mahajan (08713502712)

**Under the guidance of
Mr. Vivekanand Jha**

Abstract:

Wireless Sensor Networks are self-organizing, large-scale and high dimensional networks that are a wireless association of small, low-power and inexpensive battery operated embedded devices called sensor nodes. One way of increasing the network lifetime, as inferred from the literature, is by using an efficient data acquisition technique. Compressive Sensing is a data acquisition technique that samples the original signal into, where $m \ll n$, and reconstructs the desired signal from the few samples. To the best of our knowledge there exists no node deployment strategy that completely supports the technique of compressive sensing in Wireless Sensor Networks to achieve an energy efficient environment. The node deployment is based upon a Probability Distribution Function which shall exploit the benefits of Compressive Sensing in Wireless Sensor Networks. The project aims at designing a novel. Probability Distribution Function that assists Compressive Sensing in increasing the network lifetime while providing full coverage of the sensing field.

Keywords: Probability Distribution Function, Compressive Sensing, Sparsity, Incoherence, Wireless Sensor Networks

ELECTION OPINION ANALYSIS

USING A HADOOP BASED SYSTEM

Submitted By:Prachi Mahajan(04913502712),Prachi Aggarwal(03213502712)

**Under the guidance of
Prof. Ela Kumar**

ABSTRACT

In recent times, social media has become a popular medium for many election campaigns. It not only allows candidates to reach out to a large section of the electorate, it is also a potent medium for people to express their opinion on the proposed policies and promises of candidates. Analyzing social media data is challenging as the Data can be noisy, sparse and even multilingual. This generates a huge unstructured data for data mining. In the midst of this big data rush, Hadoop, has been heavily promoted as the one-size fits of all solution for big data problems. Hadoop has efficient architectures, is massively parallel, highly scalable and available to handle very large structured/unstructured data volumes up to several petabytes. This Project aims to analyze the public opinion regarding the US Presidential Candidate Nomination 2016 using Twitter Data. It not only focuses on the sentiment regarding the candidates but aims to perform various other analysis regarding the election on the basis of popularity, positive public support, reaction to various policies and trends.

KEYWORDS: Elections, opinion poll, Hadoop, BigData

BLACKHOLE DETECTION IN MANET

Submitted By: Bhawna Khadgi, Silvy Aggarwal, Surbhi Bhalla

**Under the guidance of
Ms. NAJME ZEHRA NAQVI**

ABSTRACT:

Mobile Ad hoc Network (MANET) is a self configuring network of mobile nodes connected by wireless links and considered as network without infrastructure. Securing MANETs is an important part of deploying and utilizing them, since they are often used in critical applications where data and communications integrity is important. In the network, some malicious nodes act as blackhole pretending to be intermediate nodes of a route to some given destinations, drop any packet that subsequently goes through it, is one of the major types of attack. Here, we propose an improved Watchdog Protocol using Sequence Number based approach that provides efficient and secure routing with prevention of Blackhole Attack. This protocol aims at delivering the packets accurately at their appropriate destination and minimizing packet loss by detecting blackhole nodes in the network.

KEYWORDS: MANET, Blackhole, Watchdog, Sequence Number

Diagnosis System for Thyroid and Parkinson's disease

Submitted By: Rohini Parle(07213502712), Akanksha Tickoo(09313502712), Puja Gupta(09513502712), Avnee Goyal(00313507213)

**Under the guidance of
Monika Choudhary**

ABSTRACT:

Diagnosis systems are the most common type of system in routine clinical use. They contain medical knowledge, usually about a very specifically defined task, and are able to reason with data from individual patients to come up with reasoned conclusions. Although there are many variations, the knowledge within a diagnosis system is typically represented in the form of a set of rules along with validation scheme to increase the accuracy and prevent from over-fitting and for correct data prediction. The thyroid gland is one of the most important organs in the body as thyroid hormones are responsible for controlling metabolism. Thyroid disease can be difficult to diagnose because symptoms are easily confused with other conditions. Parkinson's disease (PD) is the second most common neurodegenerative disorder and the most common movement disorder. It is characterized by progressive loss of muscle control, which leads to trembling of the limbs and head while at rest, stiffness, slowness, and impaired balance.

KEYWORDS: Validation schemes, data prediction

Automated Tool for Teacher Feedback Using Opinion Mining and Fuzzy Logic

Submitted By: Leena Mangla (02713502712), Amitoj Kaur Chawla(03713502712)

**Under the guidance of
Dr. Devendra K. Tayal**

ABSTRACT:

This project aims at developing an automated tool for teacher feedback. The students can enter feedback for a particular teacher and subject in natural language. This feedback will then be preprocessed using NLP techniques like tokenizing, tagging and dependency tree extraction. After preprocessing of input, aspects are identified using a structural SVM classifier. We adopt a vector based approach here, i.e., vector representations of words are used as features in the SVM. Identification of aspect is followed by aspect categorization after which the sentiment score is calculated using the cosine similarity distance. After sentiment analysis of feedback, the final performance score of the teacher is calculated by using Fuzzy Logic. Using Fuzzy logic to calculate the cumulative score provides the opportunity to incorporate the uncertainties associated with the real world, which cannot be modeled otherwise.

KEYWORDS: Teacher Feedback, Aspect based Sentiment Analysis, Vector Space Model, Fuzzy Logic

**INTERNSHIP AT EXPEDIA ,INC.- CODE AUTOMATION TOOL FOR MVC LAYERS
OF SPRING AND DESIGN PATTERN IMPLEMENTATION FOR RESTFUL WEB
SERVICE CALLS**

Submitted By:Himali Bajaj

**Under the guidance of
Ms. Vibha (Assistant Professor ,CSE)**

ABSTRACT:

Code Automation tool for MVC Layers of Spring aims to build a tool that automates the MVC layers of Spring to generate REST end points as well the unit tests and Integration tests to test these. Once built , this tool would be used by all the teams of the company to generate REST endpoints. This would greatly reduce the time and effort required to do the coding of Spring MVC Layers by all teams. Design Pattern Implementation for RESTful Web Service calls aims to implement a design pattern in java to build an abstraction layer for the various clients-Spring's REST Template ,Jersey implementing JAX-RS being used in the company project. Factory pattern was developed to deal with the same so that we can interchange from one client to another in the company project without any code changes in the project .This also should prove to be very beneficial in the future.

KEYWORDS:Spring,MVC,REST,WebServices,Design Pattern,Automation

PREDICTIVE ANALYSIS IN HEALTHCARE

Submitted By:Gursift Kaur,Taniya Chaudhary, Simrati

**Under guidance of
Prof. Ela Kumar**

ABSTRACT:

With the increasing volume of data in healthcare , advantages of Map Reduce programming model can be leveraged to reduce the time taken to predict certain diseases. Our project compares two clustering techniques and thereafter finds which location is prone to which disease out of the two similar diseases using naive bayes theorem. Cure clustering does random sampling which might not represent the whole class. Cure clustering enhanced with Mapreduce techniques consumes less time in clustering than K Means clustering.

KEYWORDS: MapReduce, K Means, clustering, healthcare,CURE Algorithm, Naive Bayes, disease

Hybrid Approaches to Software Testing and its practical applications

Submitted By:Mansi Singhal

Under the guidance of

Mr Indra Thanaya

Abstract

Management System (IUMS) is a business intelligence enabled web based ERP product for effectively managing complete university functions Testing is context-based and risk-driven. As against the older perception it actually begins even before the first line of code can be written. Software Testing Life Cycle consists of phases like Gathering Requirements(Requirement phase),Analysis, Test case planning, Test case I am doing my internship from expedien solutions. It uses Integrated University development, Review and correction, Test case execution & bug reporting &Test cycle closure. Bugzilla is a "Defect Tracking System" or "Bug-Tracking System". After finding bugs in the product, we report it using bugzilla tool. Quality Center is a comprehensive test management tool. It is a web-based tool and supports high level of communication and association among various stakeholders. **Selenium** is a portable software testing framework for web applications. Selenium provides a record/playback tool for authoring tests without learning a test scripting language. It is used for automating websites. I have made use of these tools plus have done manual testing too for testing modules of websites and other things.

ENVIRONMENT ALLOCATOR

Submitted By: ADITI GOEL

Under Guidance of

Mr. SUMIT YADAV

ABSTRACT:

Environment Allocator Tool is used to share the environments within the orgs in Amazon. As a developer in Amazon when someone take up a task which requires code change, before making a code change the developer has to setup the environment on his box. A significant Developer bandwidth is being spent in creating, debugging and maintaining environments. Thus shared environment and environment allocation can solve this problem. Developer places a request for the environment he needs using Environment Allocator tool. Environment allocator looks for an available environment and locks the environment against the userId of the developer. The tool then syncs the environment from parent so that the developer starts working on the latest changes. Developer releases the environment via the Environment Allocator after his work is done. The Tool then puts this environment back into the pool by assigning it to NULL again and making it available for others to use.

KEYWORDS: Environment Allocator, shared environments, Amazon