



COLLEGE OF ENGINEERING, PUNE
(An Autonomous Institute of Government of Maharashtra.)
Department of Computer Engineering and Information Technology
PROJECT ABSTRACTS
BTECH COMPUTER ENGINEERING 2017-18

AN UNSUPERVISED APPROACH TO MINE VOCABULARIES AND RULES FROM BUSINESS DOCUMENTS

Enterprises model the behaviour of business to prepare a communication standard for business people and to specify requirements to IT people. The communication gap between IT people and business people because of the different terminologies they use in the same context has led to major software failures. Thus, OMG standard has come up with a new direction – Semantic of Business Vocabulary and Business Rules(SBVR), to reduce the communication gap. Declarative models are provided by SBVR to represent Business Vocabulary and Business rules which can be understood by both IT people and business people. The surplus information provided in enterprises motivated the approach of automatically mining the business rules using text mining. The needs of the business leads to business rules which are constrained by the regulation policy set up by the policy guidelines of the organisation and government regulations set up on the organisation. Business rules are specified in documents like user guides, requirement documents, terms and conditions, dos and don'ts. A business analyst interprets the document and has his/her own perception of the document which leads to potential discrepancies, ambiguities and quality issues in the software system. To minimize such errors, this report presents an unsupervised approach to extract Business Vocabulary and Rules from Business Documents automatically. Business Vocabulary encompasses of the entities and the facts present in the business document belonging to the particular enterprise on the other hand Business rules are the constraints on the atomic business fact.

PACKET LOSS CONCEALMENT USING DEEP LEARNING

Audio Packet loss occurs when a network overload occurs and the information containing packets related to the media on user end are therefore lost and dropped by the network, thus creating bad user experience due to missing elements in the application (here: games). Nvidia hosts gaming as an online platform where players can use the client-server architecture to connect to a remote server and play games without needing to install high performance GPUs and other computation accelerators in their own machines. Owing to network lags there are times when some game-info carrying packets can be lost over the network, due to which the audio may experience glitches and discontinuity, this causes bad user experience To cover up the apparent loss of packets and maintain continuity of the games a lot of mechanisms have been devised, and these are called packet loss correction techniques. However, these packet loss algorithms are simply a repetition of the previous

last intact audio frames and they are repeatedly played with reduced intensity until they die out, after which the glitch cannot be covered anymore. This creates bad and inauthentic user experience. This paper uses Machine Learning to predict the missing audio frames given some audio frames before it.

CLASSIFICATION OF BENIGN AND MALICIOUS PDFS USING STATIC FEATURES AND METADATA

Classification of Benign and Malicious PDFs using Static Features and Metadata Problem Statement: Building an efficient classifier to recognize benign and malware PDFs by using custom features. There's been an unprecedented rise in the potential threats and attacks on the cyber world on a daily basis. PDF is a soft target for many malware writers as it is one of the most used file format. Embedding PDFs with malware easily accomplishes the target, and most of the times by the victim merely opening the file. In some cases, the PDF even requires the user to enter important credentials. This is mostly an illusion as the user is made to think that he is performing some operation but in reality, his credentials are either stolen or the action triggers a URL in the background which downloads malware onto the system. These attacks are well-obfuscated to deceive users. The ways to detect such malicious PDFs can be classified as static and dynamic. Static examination involves making use of the structural features of the PDF source code. Dynamic features involve run-time features of the PDF by executing it in an isolated virtual environment. Some metadata properties play a monumental role in classification of benign and malicious PDFs. In this thesis, a novel combination of static features and metadata has been used to develop feature sets that will enable to identify malicious PDFs with high accuracy. This is highly helpful to the cyber security industry, whereby this project provides new insights into malware classification and guides the developers to come up with a new approach to tackle the problem of malicious PDFs.

PERFORMANCE OPTIMIZATION OF ACTIVE BROWNIAN MOTION CODE ON ACCELERATED ARCHITECTURES

A study in the field of HPC, to survey the various researches done in the field so as to optimize HPC applications using Intel Xeon Phi Processor or Knights Landing which is a many core processor that can be programmed as per the application utilities to improve the efficiency so as to fully exploit the thread and data parallelism capabilities and the high memory bandwidth. Followed by, a stepwise optimization mechanism for a serial code into a highly parallelized one that can exploit available processor capabilities and effectively reduce the time and enhance efficiency of the application. The one under study is a code based on active brownian motion that is highly data dependent with a lot of numeric calculations. We propose a way to optimize the application by parallelizing it through OpenMP and using the available Intel environment variables and hence porting it onto the Xeon

processor and Intel Knights Landing and efficiently using the massive thread and data parallelism available on the same.

OPTIMIZATION OF ACTIVE BROWNIAN MOTION USING GPUs

Classical research into active brownian motion uses stochastic differential equations in order to simulate the motion of these active brownian particles. However, such simulations are very computationally expensive, and have high runtimes which could run in hours or days. The sheer amount of computation in these simulations imposes a limit on the optimization that can be obtained by traditional software-based approaches. This project presents a novel method of profiling and optimizing algorithms used for the simulation of active brownian motion.

A PROTOTYPE SENTIMENT ANALYSIS SYSTEM OF TRANSCRIBED TEXT

The aim of this project is to develop a prototype version of an opinion mining system which will do sentiment analysis on transcribed text. The transcribed text can be text converted from any source. The model uses call centres product feedback as the data source. Call centres are an important intermediary between customers and enterprises. Sentiment Analysis on these calls provide detailed understanding of the products. These combined with Aspect miner helps to find out the topic or the subject of the feedback. The benefits of the proposed model include: Gaining insight into customer attitude on services and products. Improve the brands reputation. Working on and rectifying the negative aspects of the product. Our project incorporates speech to text conversion of the calls, aspect mining and sentiment classification on the converted calls. For speech to text conversion, Google Speech API is an active approach which converts speech to text whilst keeping the grammatical and contextual meaning intact. Sentiment analysis aims to determine the opinion of the customer regarding a certain aspect of the product. Our proposed model uses Naive Bayes classifier for sentiment classification which is an easy and fast algorithm for text classification. A report will be generated in the form of a bar chart wherein every aspect will be represented with positive and negative bar charts. The positive and negative bars represent the percentage of negative and positive sentiments respectively.

DISTRIBUTED COMPUTING FOR AUTONOMOUS DRIVING SYSTEMS

Driver negligence is one the leading causes for road accidents worldwide. More than 12 million people lose their lives every year globally. Light motor vehicles account for 22 percent of the total road accidents in India alone in 2016. With the increasing number of vehicles, the use of efficient computer vision algorithms for safety is the way forward. Autonomous driving systems have made the task of taking smart decisions while driving easier and more efficient. Technologies like cruise control and parking assistance make the driver feel safer and have increased the level of comfort

while driving. These autonomous driving systems often use sensor data to capture images of the road/track on which the vehicle is being driven on. The view spans multiple cameras placed around a car driven through numerous driving Scenarios. The images are captured at a high speed and are sent to the processing unit where they are analyzed using computer vision algorithms. The results are then dispatched to the user interface and thus helping the driver make smarter choices. Using over 100 hours of driving footage as the training set, we propose to make a smart system which helps the driver to make smarter choices while driving. A huge data set requires efficient computation so, an efficient way of handling the image dataset is the Map reduce paradigm. The Hadoop Image Processing Interface (HIPI) is used to implement the paradigm. After analyzing the large volume of images in a distributed environment, through image processing using OpenCV the first series of decisions like lane detection and switching, traffic light and road sign detection will be made.

EMAIL SENTIMENT ANALYSIS

Natural Language Processing (NLP) is a field of computer science concerned with the interaction and programming of computers to fruitfully process natural (human) language. In recent technical advancements, a lot of focus has been on building varied interfaces and platforms that can examine and analyze natural language, with an aim to develop different methods for interacting with computers. Sentiment analysis is a subfield of natural language processing which involves the study of tone and emotion of a natural language data. A typical sentiment analysis focuses on predicting the positive or negative polarity of the given sentence. With the help of machine learning and data mining, tons of data is being obtained from social media websites and e-commerce review sites. This data can be helpful in training algorithms and models to determine the sentiment of a natural language message with appropriate accuracy. The proposed system is implemented by building an application that can carry out the extraction of emails, perform sentiment analysis on them and then send out a text message if a strong negative tone is identified in the email.

ANALYSIS OF TRANSACTIONAL DATA USING DATA MINING TECHNIQUES TO MAXIMIZE THE PROFIT OF COMPANY.

In today's world, economy is growing and the size of data is increasing in all fields. There is a need to analyze the data and draw useful insights from it. Recently e-commerce has seen a huge growth because of its advantages. People prefer to shop on e-commerce websites. An e-commerce company needs to analyze its transactional data to see the new trends, rise or fall in Production etc. Using data analysis, we are able to store, track and analyze information provided by the company. The uses of this information have a huge scope in delivering increased sales and profits for the company. We use a data mining technique which is used for Basket data. This technique is association rule mining. In

transactional data we have the items sold in a set of transactions. Large retailers and e-commerce companies use market basket analysis to find combinations of items that occur frequently together. Using association rule mining we try to find strong rules. The companies can utilize this information to increase profits and target marketing.

IoT BASED AUGMENTED REALITY SYSTEM OF HUMAN HEART: AN ANDROID APPLICATION

Augmented Reality has been defined broadly as combining real and computer generated digital information into the user's view of the physical and interactive real world in such a way that they appear as one environment, thus providing a bridge between digital information and the physical world. Augmented Reality can amplify human perception remarkably. Currently, during a cardiac examination, it becomes extremely difficult for the patient to comprehend the results. As the results are non-intuitive, doctors have a tough time explaining them to patients. Therefore, the patients have to rely on doctor's judgment against his own. The system proposed in this project intends to make use of Augmented Reality (AR) technology for evaluation and visualization of a beating heart to make this doctor-patient interaction more intuitive. The system proposed in this project is an android application which uses device's camera to display the 3D heart model in the augmented view. The real-time pulse data is loaded via a pulse sensor connected with an MCU through the cloud. Hence, the results obtained have shown us that this system has proven to bring more intuitiveness in beating heart visualization and analysis.

SENTIMENT ANALYSIS OF TWITTER DATA

Sentiment Analysis is also called as opinion mining which is used to analyze people's opinions, attitudes, political as well as religious views about specific personality, product, etc and determined whether they are positive or negative. Natural language processing, data mining, web mining and text mining are the areas where sentiment analysis is widely used. The importance of sentiment analysis corresponds with the development of social media such as forum, discussions, micro-blogs and Twitter. Twitter is one of the social media that is achieving Popularity and provides standard to express the opinions and ideas. The results you get from sentiment analysis helps industries to improve their market strategy, customer service and product quality also boost sales revenue. It also aids customers in decision making before any purchase. Positive, negative, neutral sentiment can be determined using scaling system. From the evaluation of this study it can be concluded that the proposed machine learning and natural language processing techniques are an effective and practical methods for sentiment analysis. Our project will involve collection of data from web resources such as e-commerce websites; blogging sites etc by using Hadoop and intend to derive useful inferences and recommendations. As it is textual data it will involve various stages of pre-processing as well as data structurization. State of art machine learning models will be involved in

data analysis and predictions. This will benefit organizations to save on manual effort and give crucial trend of user opinions to aid business discussion.

ANALYSIS AND SUMMARISATION OF TRENDING TOPICS ON TWITTER

Twitter is considered as a rich resource of information about current world actions in all fields. Ample amount of work and efforts are made to find trending topics on Twitter i.e., the latest popular topics of conversation among the users. Analysing these tweets can help in finding trending topics, thus providing important information for real-time decision making in various fields. In this project, we will be focusing on fetching data from Twitter and finding latest popular topics of discussion on Twitter by fetching daily data from Twitter and making use of the Twitter Hash tags and subsequently providing summary by providing related URLs. The important feature of this project is its ability to find the trending topics along with the related URLs, so the user can get complete knowledge of the event. Analysis of this data can help to understand the reaction of masses towards these Trending Topics. The related links can provide complete insight of reaction of people towards current worldwide actions.

COOPERATIVE TRANSPORTATION USING MULTI-AGENT REINFORCEMENT LEARNING

With miniaturization of sensors and faster processors, deploying teams of robots is a viable alternative to human teams in handling dangerous activities after a disaster like search and rescue as well as cleaning up hazardous materials. In our project, we propose a co operative multi-agent system composed of a set of autonomous agents working together in a dynamic environment to accomplish a common goal. Agents are trained using MARL (Multi-Agent Reinforcement Learning), a technique which allows them to explore the environment through trial and error to develop effective co ordination policies using a neural network, gradually improving their Performance through experience. In a real world scenario, we cannot rely on communication infrastructure to facilitate agent co ordination. To accurately model these conditions, each agent only has a local view of its surroundings and no long range communication. Each agent chooses the action it believes will maximize its overall reward according to a reward scheme configured by a human supervisor to indicate high-level goals. To help agents relate new experience with past experience, we incorporate 'Experience Replay' into the regular learning paradigm. Though we do not allow long distance communication, we allow agents who physically meet to exchange the maps they have created of the unknown environment.

ANOMALY DETECTION IN CLOUD INFRASTRUCTURE USING MACHINE LEARNING

Cloud is one of the emerging technologies in the field of computer science and has become exceedingly popular due to its use of flexible resources to provide effective, economical and on-

demand services. As technology starts to grow in scale and complexity, the need for automated anomaly detection and monitoring systems becomes important. Cloud infrastructure is susceptible to various kinds of problems such as CPU over-utilization, RAM overuse, and increase in IO wait time. Inappropriate exploitation of Cloud resources can often lead to faults like crashing of VM, decreased efficiency of cloud system, etc. These faults are often preceded by anomalies which can be used as indicators to prevent such faults thereby reducing the cost spent in the recovery of the Cloud system. It is paramount that the services that are delivered to users by any cloud service provider work in a precise fashion without the occurrence of unexpected detrimental events. Thus anomaly detection holds great significance since abnormalities in data can often translate to critical and significant information that may be actionable. What we present is a comparative study of various machine learning algorithms for detecting anomalies in cloud. These algorithms can be implemented in the system that can be used to detect whether any cloud component is experiencing an anomaly or not and will notify both the system administrator and the user so that appropriate steps can be taken to prevent the system from going into a faulty state.

NOCO: A PROTOCOL FOR EFFICIENT TRANSACTION PROPAGATION IN OPEN BLOCKCHAINS

The emergence of crypto currency like Bit coin has given the world a model of a trust less, secure and a reliable way of transacting wealth. Bit coin is a crypto currency that unlike traditional currencies does not rely on a centralized authority. Instead Bit coin relies on a network of volunteers that collectively implement a replicated ledger and verify transactions. But, the present Bit coin's block chain agreement does not scale well as compared to other centralised payment systems like MasterCard and VISA: At present irrespective of the computational power available Bit coin block chain processes only 7 transactions per second. One of the identified reasons for this is that transactions are mined by all the miners; hence many miners simply waste their computation. Like Bit coin, there are several crypto currencies in the block chain realm, like Ethereum, Litecoins, etc. All of these suffer more or less from the same bottlenecks. There are several techniques proposed for overcoming this: changing consensus protocol (Proof of Stake), increasing the block size and sharing. However, all the techniques do have a trade off. In this report, we propose a new node count based transaction propagation protocol for less intensive mining called as NoCo. NoCo distributes a transaction from a node to the nearby miners based on the node count between them and thereby decreasing the competition, and hence this is beneficial for the miners as well as for the network's scalability. We have developed this over GoEthereum and tested it with various numbers of nodes.

DATA SCIENCE BASED ADVERTISEMENT RECOMMENDATIONS IN INFOTAINMENT

As enormous amount of data is available in various forms on different devices, Internet

advertisement is a challenge for today's world. This paved way for enhancement in advertisement recommendation service. This project mainly focuses on the various techniques used for building the framework which is in turn used to recommend advertisements based on various data sources. The sources of data used are: users personalized information, social Updates and interests, travel details, ongoing video clip etc. The pivot of this project is to extract user's needs and mood from these data sources and recommending advertisements accordingly. Here, we proposed data science based advertisement recommendation system based on user details and video content.

OPEN-DOMAIN QUESTION ANSWERING USING FEATURE

Question Answering is one of the many Natural Language Processing tasks which has recently attracted a lot of attention from the Artificial Intelligence research community. In QA, a computer returns the answer to a users query. Earlier, retrieval systems retrieved related documents based on keyword search. This method however does not capture the meaning of the natural language query. With developments in NLP, understanding the syntax and semantics of natural language has become possible. This has led to development of intelligent Question Answering Systems. Previously, the algorithms used in Question Answering were either statistical and/or rule-based. With recent developments in Deep Learning, neural network methods have shown promise for QA (Question Answering) systems. Researchers have begun to tackle open-domain QA, in which the model is given a question and access to a large, diverse corpus (e.g. Wikipedia). Open Domain QA is complex, as it requires large-scale search for relevant passages by an Information Retrieval component, combined with a Reading Comprehension component that reads the passages to generate an answer to the question. We have designed a novel open-domain system with two components viz., Paragraph Retriever and Paragraph Reader. Traditionally, retrieval systems only retrieve the most relevant documents. Our Retriever component retrieves the most relevant paragraphs with respect to the question. To retrieve the most relevant paragraphs, our Paragraph Retriever system uses a two level search. In the Reader Component, we propose a modification to the DCN model [1] , by augmenting it with other useful features like Parts of Speech of words, Named Entities, Exact Match between question and paragraph, question word identification for extracting the answer from a paragraph.

DATA SHARING AND DISCUSSION PORTAL FOR EDUCATIONAL INSTITUTES

We see students and faculties from small educational institutes facing problems with proper communication with each other. Lack of communication between students and teachers has caused a great impact on the quality of knowledge the students gain in India. The project focuses on the primary information sharing needs of such institutes. The portal provides a platform for students and faculties for discussion and complaint about improvement of the college management. It also avails a personal storage space for storing of academic data, a place for blogging and a mail server

which will further improve the communication standards. The basic requirement of such portal i.e. information management is also a part of this project. A Data Sharing and Discussion Forum equipped with all the necessary facilities can be a great tool to handle the situation very effectively.

CENTRAL BANK INTEREST RATE PREDICTION

Central bank interest rate, also known as the repo rate, is an essential aspect to a large population of any country including industrialists, investors, bankers, etc. Interest rates refer to the general level of interest that a borrower has to pay a lender to borrow a certain amount of money for a certain amount of time. Interest rates directly affect the private banks and the working of stock markets. Their prediction can, thus, help in predicting any crisis and in turn help to take any preventive measures to avoid shocks in an economy. Interest rates are affected by macroeconomic as well as microeconomic factors. We have focused on the macro economic factors in our project such as GDP, CPI, retail sales, etc. We are trying to deeply understand the relationship of each factor with the interest rate using the regression machine learning algorithms. With comparative study between these algorithms, we have built a machine learning model that derives the correlation of all the factors with interest rate. Our goal was to predict the interest rate for an unseen dataset with scope of minimum error using the machine learning model.

POLYPHONIC MUSIC TRANSCRIPTION

Music transcription is usually done manually and requires intuition . It is learnt through experience after listening and categorizing notes in the audio signal. These notes are available online or purchase through sites likemusicnotes.com , sheet music plus etc. and not necessarily available for each instrument and song. There is a need to automate this process so that notes for each instrument in a song can be readily obtained. With the advent of machine learning , it is quite possible to automate this process where the machine will learn from its experience.

Our approach will be to identify and segregate individual notes of each instrument which will be converted to corresponding music sheet. Initially ,human voice and irrelevant noise will be separated from the instrumental part. Each of the audio signals originating from a particular instrument will then be segregated. Next step will be classifying the segregated instruments. The segregated audio signals will be converted to notes corresponding to that instrument. Finally, this would be represented in the form of music sheet for each instrument. Such system has numerous practical applications.

SENTIMENT ANALYSIS ON CUSTOMER REVIEWS

Opinion mining is one of the most important topic in today's world for individuals. Due to development in e-commerce, more data is getting stored in web and also the amount of e-commerce

website users for shopping is increasing. Many of these e-commerce websites have facility to express user opinions such as forums or review sections from which other users can know about the user experience for the specified product. So analysis of these reviews can be beneficiary for users as well as manufacturers. This work concentrates on extracting amazon customer reviews but we can use it for any other e-commerce website. Opinion mining is one of the most important topics in today's world for individuals. Due to development in e-commerce, more data is getting stored in web and also the amount of e-commerce website users for shopping is increasing. Many of these e-commerce websites have facility to express user opinions such as forums or review sections from which other users can know about the user experience for the specified product. So analysis of these reviews can be beneficiary for users as well as manufacturers. The process of seeking the perspective of any user about any issue or object or artifact is called opinion mining. The analysis of the feelings of the extracted opinions is denoted as Analysis of emotions/sentiments. This work concentrates on extracting amazon customer reviews but we can use it for any other e-commerce website.

ULTRASCOPE PLUS

Telescopes are used for gathering and analyzing the radiation released by distant objects. But in today's world people find it difficult to have access to such powerful telescopes and these telescopes cannot be bought by everyone due to their high cost. In addition various other factors such as optimal space, local unavailability, lack of knowledge regarding installation and use ,perfect view further reduce the access of these powerful telescopes to potential amateur astronomers and institutions. Thus we need to bring down the cost of development and provide better access to such telescopes. With recent advancement in technology we can use the technique of additive manufacturing to build telescopes at substantially reduced cost and further via Internet provide remote access of these telescopes to user. Furthermore it's merely impossible to identify, locate and describe stars just by looking at the night sky, which is major cause of disappointment among amateur astronomers who often get confused and discouraged due lack of sufficient knowledge about astronomy.

Thus we need to develop a user friendly interface that can easily locate and describe stars thus reducing the burden of identifying, locating and describing stars, keeping their interest in the field of astronomy alive. In addition to above problems, providing a platform dependent solution, restrict the solution to specific model of telescope which fails to bring out the global impact required in the field of astronomy.

Thus we need to deploy a platform independent solution which can work over a wide variety of telescopes. Able to control your telescope with click of your mouse using one of the biggest open source platform is what's desired by every astronomer. We have come closer to this dream by

providing a hardware independent software solution at one of the most popular open source platform stellarium.

GRAPH BASED QUESTION ANSWERING

In today's digital age, in the dawning era of big data analytics, it is not the information but the linking of information through entities and actions, which defines the discourse. Any textual data either available on the Internet or off-line (like newspaper data, Wikipedia dump, etc) is basically connected information which cannot be treated isolated for its wholesome semantics. There is a need for an automated retrieval process with proper information extraction to structure the data for relevant and fast text analytics. The first big challenge is the conversion of unstructured textual data to structured data. Unlike other databases

Graph databases handle relationships and connections very elegantly. Our project aims at developing a graph based information extraction and retrieval system. The project involves two main phases - firstly, XML and English text to graph database creation and secondly, querying and visualization. In the first phase, we build a graph of entities and relationships from given input XML and unstructured English sentences. We have identified patterns in sentences to extract relationships between entities. In the second phase, we query the graph database with natural language queries to retrieve results. The queries are annotated using different database oriented classifications to be processed for graph creation. The query graph is matched with the document graph, created in the first phase and the matched results are visualized with a graphical depiction of entities and their relations so as to give better and quicker insight into the information.

SENTIMENT ANALYSIS OF CODE MIXED DATA USING NEURAL NETWORKS(CODE - MIXED DATA HINDI/ENGLISH)

An important part of today's information-gathering is to find out what other people think about products, events, famous personalities, movies etc. People express their emotions through tweets and Facebook posts. Often, people mix words of different languages to write on twitter and Facebook. This poses a challenge to judge the sentiments expressed in this data. The aim of this project is to address the problem of sentiment analysis of code - mixed(English-Hindi) tweets and Facebook posts; that is classifying tweets and posts according to the sentiment expressed in them.

MINING AND ANALYZING OPERATIONAL PROCESSES FROM SYSTEM AND APPLICATION LOGS

Process mining is the technique of extracting process models, using logs and traces, from information systems. Obtaining such models can often increase the understanding of the processes in the organization. This is particularly helpful in designing company production policy and market strategies. For businesses that have never ventured into business process management, process

mining is a great place to start. An organization may also use it to keep a check on their current business processes and update them or make changes in the organization as required. Existing mining algorithms produce a process model taking system logs as input. These algorithms are not robust. They can only handle noise in the form of infrequent behaviour. They do not account for many types of discrepancies which occur in real life logs. If real life logs are given as input in their raw form to the existing mining algorithms, they will discover inaccurate models. A number of processing steps are required to clean the real life logs so that they give good results with the existing mining algorithms. Our aim is to identify general methods of pre-processing which will remove some discrepancies, based on the characteristics of the logs. These methods should be applicable on any log in general in order to obtain better results from the mining algorithms. To achieve this, we work on real life logs to identify the steps of pre-processing required on them and then we generalize them so that they could be applied on any logs. In the process, we identify the algorithms that might work best with the given logs and also their configuration settings.

ANALYZER FOR ASSESSING AND VALIDATING EFFICIENCY PARAMETERS IN APACHE STORM

Big data analytics is a term used for processing of data that is too big to be analyzed using conventional processing tools. One such big data analysis tool is Apache Storm. In this project we intend to improve the efficiency of Apache Storm by analyzing the parameters that are of utmost importance when it comes to efficiency. Feature selection algorithms such as Relief along with some cross validations techniques are going to be used to shortlist the most important parameters from a lengthy list. In the end, we intend to shortlist 5-6 parameters and verify the results.

ALGORITHMIC FOREX TRADING

Foreign Currency Exchange market (Forex) is a highly volatile complex time series. Predicting the trends in Foreign Exchange prices is a very challenging task due to the many uncertainties involved and many variables that influence the market value in a particular day such as economic factors, political events, government debt etc which cannot be anticipated just by considering numeric time series data. This project proposes a prediction model that combines artificial neural network (Long Short term memory model) for historical analysis and support vector machine for news analysis which will consider the impact of above mentioned factors. The performance measure for LSTM is quantified in terms of mean absolute error, mean square error and root mean square error. SVM is evaluated using confusion matrix. Next, the predictions from both these models are integrated to design trading rules which eventually maximizes the profit and reduces the risk.

REMOTE DESKTOP INFRASTRUCTURE MANAGEMENT

Many times it is observed that the lab admin or person-in-charge needs to physically visit very

desktop machine in order to check whether it is turned off properly or not. It is very cumbersome for the lab admin to check every machine individually and turn it off. So in this case there is huge need of automating this task with the help of some system management software which would help to ease the task. The purpose of our project is to create centralized system management software using client-server architecture. The Remote Desktop management Software will generate a Dashboard on server machine containing IP-Addresses of all the machines in lab along with their status whether particular machine is on or off. Depending on the status of that machine it will turn off that machine if it is on. The project also helps the administrators for managing power by turning off the remote computers when not required. This will lead to minimization of power consumption and thus saving energy in labs/offices. The practical application of our software will benefit as a solution for Green Backup: Save Power, Save Money. The project also aims at sending message to administrator for the purpose of switching off the main power supply when all the machines will be turned off properly. The purpose of sending the message is to safely turn off the computers without crash or failure as turning off the MCB directly without switching the computers off may damage the systems leading to a major crisis.

FEATURE SELECTION OF PERFORMANCE PARAMETERS FOR APACHE SPARK

Spark is an open-source platform. It provides data processing which is suited for large datasets. Spark provides no file management and hence has to work on Hadoop's cluster management system like yarn. It still provides a standalone mode but this mode as name suggests works on single system only and is not usable for clusters. Spark can perform batch processing but its forte lies in stream workloads, machine based learning and interactive queries. Spark also provides real time data processing capability which gives it upper edge over Hadoop. The parameters of Apache spark are available in unsupervised format and classification algorithm cannot be used on unsupervised data it needs to be converted to supervised format by using clustering algorithm. Then classification algorithms can be applied over the obtained labelled data. Our aim is to get the most important efficiency parameters for Apache Spark working on three system cluster and we do it by applying K-means algorithm to get supervised datasets and then obtaining efficient parameters by applying a feature selection algorithm.

AUGMENTED TRAIL ROOM (ATR)

Now-a-days shopper encounters various problem while shopping, especially during holidays or weekends. Long waiting queue, crowds, various restrictions make it quite an unpleasant and dissatisfying experience during shopping. Minimum numbers of trial rooms and huge number of customers increases waiting time for customers and results in customer dissatisfaction. Also for some security reasons, there is restriction on the number of garments that can be taken, at one

instance of time, for trial. To overcome these problems, we proposed a Augmented Trial Room. In this system, the person will stand in front of monitor, and system will simulate clothes on user body. User will be provided with varieties of clothes, allowing the user to select the best one and making it user friendly. User can change clothes by just moving his hands and the system will recognize the gesture and accordingly impose the other cloth. This will reduce time of waiting queue of trial rooms and apparently will eliminate trial rooms by making it virtually.

APPLICATION OF GENERATIVE ADVERSARIAL NETWORKS IN IMAGE GENERATION

Adversarial training is defined by the process of injecting fake samples in datasets with the aim to fool the model and improve its robustness in the process. One such application of adversarial training is Generative Adversarial Networks (GANs) which are a class of artificial intelligence algorithms used in unsupervised machine learning. In simple terms, GANs can be explained as a system of two neural networks competing with each other, resulting in improved results of both of these neural networks. GANs are based on the game theory approach with the objective to find Nash equilibrium between the two networks, Generator and Discriminator. We have implemented a multi-stage generative adversarial network and tuned its hyper-parameters to obtain quality images from natural language textual description. Initially we convert the input captions to vectors in a multi-dimensional space to obtain text embedding which can be fed into the model. To better map these vectors to the images they represent, a CNN-RNN encoder is used to obtain these vectors. These vectors are then fed into the first stage GAN along with some noise which outputs a low resolution image. These vectors along with the low resolution image are then again fed into the 2nd stage GAN, which then outputs a higher resolution image. To improve training stability and speed, DCGAN (Deep Convolution) approach has been used, i.e. Pooling and Flattening Layers have been eliminated in lieu of Convolutions of same strides. Also, noise vectors are sampled from a Normal Distribution instead of Uniform Distribution.

BANK LOAN DEFAULT PREDICTION

When Common People or Bank lends capital to a business, they believe it will be profitable and give them returns. They are capable to handle Failure and Critical Situation which may face while doing Business. The primary goal of this project is to extract patterns from a common loan approved dataset and then build a model based on these extracted patterns, in order to predict the likely loan defaulters by using classification data mining algorithms. The historical data of the customers like their age, income, loan amount, employment length etc. will be used in order to do the analysis. Later on, some analysis will also be done to and the most relevant attributes, i.e., the factors that the prediction result the most. To be on a Step we are going to add one more layer of validation. This validation is to stop or reduce Bankruptcy which is a big issue of modern India. Such kind of

validation will enhance Banking system and lower the rate of Insolvent of people and move towards our vision 20-20.

INTELLIGENT STOCK PREDICTION SYSTEM

Algorithmic trading is a trading system makes use of highly advanced, robust and scalable complex mathematical models and formulae to take high frequency decisions and stantaneous transactions in the financial market. Several hedge funds depend on such models for trading stocks, mutual funds and other investments. These funds represent hundreds of billions of dollars of capital being handled and manipulated by lines of computer code.

At the same time, an all time high number of hedge funds are also totally controlled and managed by artificial Intelligence-driven trading systems. Today, the results obtained from machine learning represent a substantial in-crease in returns over existing algorithmic trading engines. What we are proposing is an intelligent stock trading system that makes decisions by means of various heuristic strategies, where it truly understands the nature of a company's stock thereby enabling it to make much more informed decisions, and achieving higher returns and better yields. By our logic, we are taking a small, yet significant step towards achieving financial singularity. It is the stage where all market decisions and behaviour is controlled by intelligent machines, and no human quotient is involved. Imagine this, a market, one that works solely on the basis of logic and justified decisions, such it prices all assets in the right manner without any requirement for human validation. The growth of interest in advanced technologies such as artificial intelligence (AI) in the economic sector could spark the advent of machines that are capable of taking decisions in an instant by adapting and reacting to different situations in the market immediately. When the human element is disregarded from markets, then efficient and justfied markets, which have only existed so far theoretically, could be feasible.