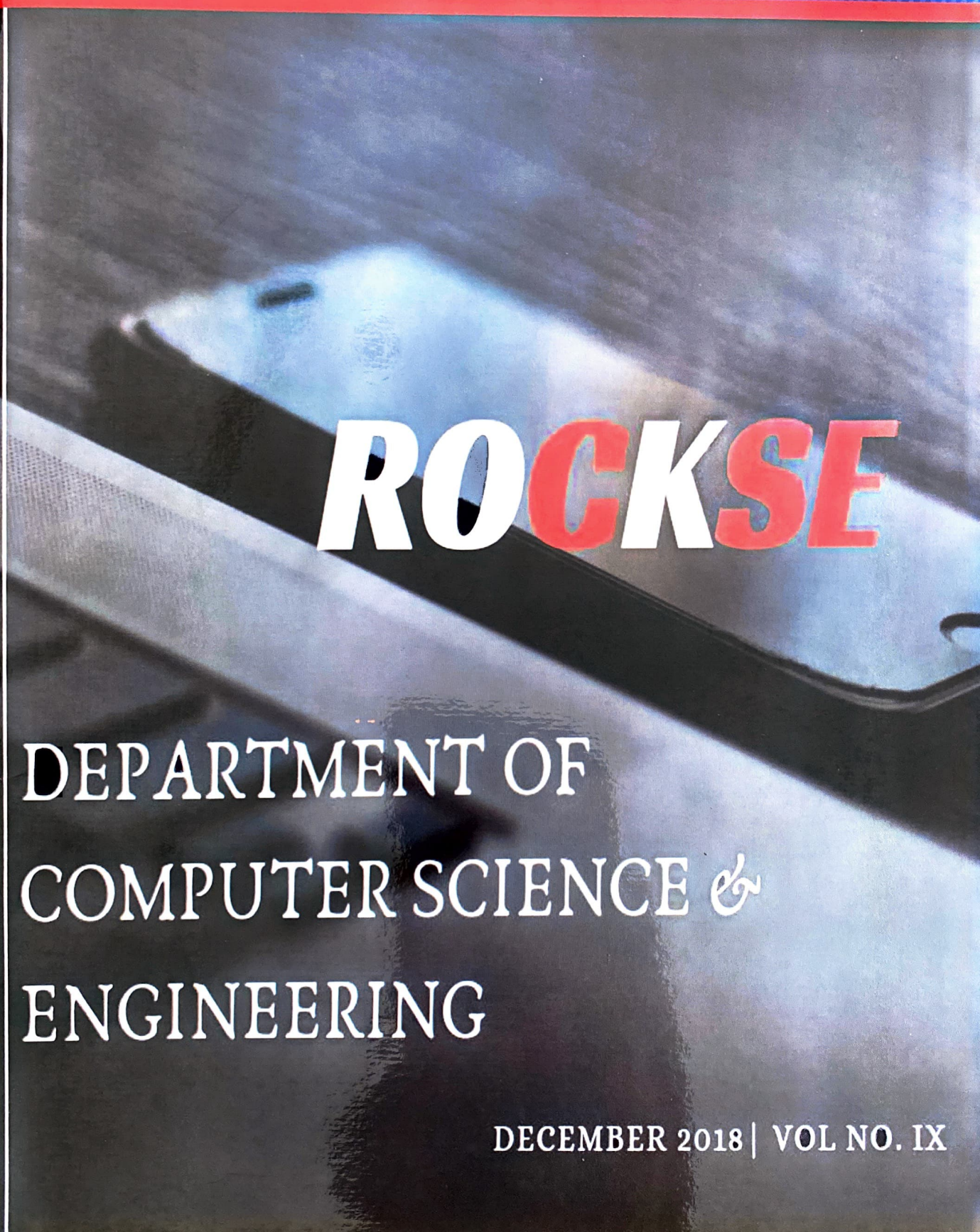
KAKATIYA INSTITUTE OF TECHNOLOGY & SCIENCE
(An Autonomous Institute under Kakatiya University, Warangal)
Opp : Yerragattu Gutta, Hasanparthy (Mandal), WARANGAL - 506 015, Telangana, INDIA.
काकतीय प्रौद्योगिकी एवं विज्ञान संस्थान, वरंगल - ५२९ ०१५
కాకతీయ సాంకేతిక విజ్ఞాన శాస్త్ర విద్యాలయం, వరంగల్ - ౫౦౬ ౦౧౫
Estd : 1980
KITSW



ROCKSE

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DECEMBER 2018 | VOL NO. IX

VIEWS

I am grateful with the idea of this magazine roCkSE. It gives overall views on the current evolving technologies which are trending in computer science field.

I am grateful with the idea of this magazine roCkSE. It gives overall views on the current evolving technologies which are trending in computer science field.

Dr. K. Ashoka Reddy
Principal, KITSW

This magazine summarizes the curent state of CSE, and also information of department. Providing an arena for the student community to showcase their technical talents is a great task.

Dr. P. Niranjan
Professor & Head
Department of CSE

EDITORIAL BOARD

KAKATIYA INSTITUTE OF TECHNOLOGY & SCIENCE
(An Autonomous Institute under Kakatiya University, Warangal)
(Approved by AICTE, New Delhi) (Sponsored by Ekasila Education Society, Warangal)
Opp: Yerragattu Gutta, Hasanparthy (Mandal), WARANGAL - 506 015, Telangana, INDIA.
काकतीय प्रौद्योगिकी एवं विज्ञान संस्थान, वरंगल - ५०६ ०१५
కాకతీయ సాంకేతిక విజ్ఞాన శాస్త్ర విద్యాలయం, వరంగల్ - ౫౦౬ ౦౧౫
Estd 1980
KITSW

Department of
Computer Science and Engineering

Presents...

roCkSE

A Technical Magazine

Volume - IX
DECEMBER - 2018

CONTENTS

Vision

Attaining center of excellence status in various fields of Computer Science and Engineering by offering worthful education, training and research to improve quality of software services for ever growing needs of the industry and society.

Mission

- Practice qualitative approach and standards to provide students better understanding and profound knowledge in the fundamentals and concepts of computer science with its allied disciplines
- Motivate students in continuous learning to enhance their technical, communication, and managerial skills to make them competent and cope with the latest trends, technologies, and improvements in computer science to have a successful career with professional ethics.
- Involve students to analyze, design and experiment with contemporary research problems in computer science to impact socio-economic, political and environmental aspects of the globe.

Program Educational Objectives (PEOs): B.Tech

- Graduates with fundamental knowledge should escalate the technical skills within and across disciplines of Computer Science Engineering for productive career by maintaining professional ethics.
- Graduates should develop and exercise their capabilities to demonstrate their creativity in engineering practice and exhibit leadership with responsibility in teamwork.
- Graduates should refine their knowledge and skills to attain professional competence through life-long learning such as higher education, research and professional activities.

Program Specific Outcomes (PSOs): B.Tech

- **Software Development and Quality assurance:** Transform various legacy or manual systems into computer automated systems using Modern Programming Languages, Integrated Development Environments, and apply Testing Tools for efficient verification and validation of those software systems
- **Maintenance:** Demonstrate knowledge in fixing and updating multidisciplinary software problems working in real time environment.
- **Immediate professional practice:** Work as a software practitioner or continue higher education by adopting advanced technologies in various fields of computer science and Engineering.

Program Outcomes (POs): B.Tech

- Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
- Problem analysis: Identify, formulate, review research literature, and analyse complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

- Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
- Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
- Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modelling to complex engineering activities with an understanding of the limitations.
- The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
- Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
- Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
- Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
- Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
- Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
- Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Program Outcomes (POs): M.Tech

- Engineering Knowledge: An ability to independently carry out research/ investigation and development work to solve practical problems.
- Problem Analysis: An ability to write and present a substantial technical report / document.
- Design/Development of solutions: Students should be able to demonstrate a degree of mastery over as per the Specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program.

Program Specific Outcomes(PSOs): M.Tech

- Software Development and Quality Assurance: Apply the Knowledge and current technologies of software engineering to Pursue Research over complex problems of Computer Science domains.
- Maintenance: Equipped with the Industry Ready, Teaching Skills and Entrepreneurship Capabilities.
- Immediate Professional Practice: plan, manage and assess effectively the software products by using the Software Engineering Concepts and Methodologies.

List of Faculty papers published in Journals:

S.No	Title of paper	Name of the author	Name of journal	ISBN/ISSN number
1	An Overview of Integrating Text Processing Techniques and Geospatial Techniques using SpSJoin for efficiently performing Spatial Similarity Joins	Mr. S. NagaRaju	International Journal of Research and analytical reviews	2348 –1269
2	A study of an enhanced approach towards frequent pattern mining	Mr. V. Chandra Shekar Rao	International Journal of Pure and Applied Mathematics	1314-3395
3	A study of an enhanced approach towards frequent pattern mining	Mr. Siripuri Kiran	International Journal of Pure and Applied Mathematics	1314-3395
4	Review of Feature Extraction Techniques for Character Recognition	Dr S.Narasimha Reddy	IETE Journal of Research	0377-2063
5	A novel local skew correction and segmentation approach for printed multilingual Indian documents	Dr S.Narasimha Reddy	Alexandria Engineering Journal	1110-0168
6	A Review on Big Data Application in Health Care	Mr. G Sridhar	International Journal of Scientific Research in Science and Technology	2395-602X
7	A study on network security administration using the technology of data mining	Mr. Siripuri Kiran	International Journal of Pure and Applied Mathematics	1314-3395
8	Fuzzy Associative Classification Driven MapReduce Computing Solution for Effective Learning from Uncertain and Dynamic BigData	Mr. B. Raghuram	International Journal of Database Theory and Application	2207-9688

9	Multiple objective–based Modified Exponential Gravitational Search Algorithm to VMM strategy for Load balancing in Cloud	Mr. P. Vijay kumar	International Journal of Intelligent Engineering & Systems	2185-3118
10	Exponential Gravitational Search Algorithm based Migration strategy for Load balancing in Cloud Computing	Mr. P. Vijay kumar	International Journal of Modeling, Simulation, and Scientific Computing	1793-9615
11	Decision Tree Analysis Tool with the Design Approach of Probability Density Function towards Uncertain Data Classification	Mr. Siripuri Kiran	International Journal of Scientific Research in Science and Technology	2395-602X
12	Anomaly Detection Using Data Mining Techniques in Social Networking	Mr. Siripuri Kiran	International Journal for Research in Applied Science & Engineering Technology	2321-9653
13	A Case Study on Issues In Privacy Preserving Data Mining	Mr. Siripuri Kiran	International Journal for Science and Advance Research in Technology	2395-1052
14	Real Time NIDS towards Architecture of Data Mining	Mr. Siripuri Kiran	International Journal of Advanced Research in Computer Engineering & Technology	2278 –1323
15	Real Time NIDS towards Architecture of Data Mining	Ms. G. Rekha	International Journal of Advanced Research in Computer Engineering & Technology	2278 –1323

An Overview of Integrating Text Processing Techniques and Geospatial Techniques using SpSJoin for efficiently performing Spatial Similarity Joins

S. Naga Raju,

ABSTRACT *A spatial similarity join of two geospatial datasets discovers sets of records that are at the same time comparable on spatial and textual qualities. Such join is helpful for an assortment of uses, similar to information purifying, record linkage, duplications identification and geocoding upgrade. Proficient techniques exist for the individual joins on either spatial or textual qualities. In any case, the consolidated issue has gotten considerably less re-seek consideration. This paper displays the SpSJoin (Spatial Similarity join) framework to fill in this need. SpSJoin is a stage that unions geospatial and text processing techniques for effectively performing spatial similarity joins. The plat-shape use parallel processing with MapReduce to handle scalability issues in joining huge datasets. The proficiency of the proposed techniques are tentatively approved with a join case for enhancing the geolocation of elements in a genuine geospatial dataset with referential elements of another dataset.*

Key Words: *Spatial Similarity Joins, Text Processing Techniques, Geospatial Techniques.*

1. INTRODUCTION

In current land databases, records contain textual and spatial credits to depict qualities and area of certifiable substances. At the point when the area of the records has low accuracy, e.g. geolocated at the focal point of the city, their area might be upgraded by finding their most comparable records in another database, known to have high area accuracy. For example, Figure 1 demonstrates test records of Physicians database, geolocated at downtown area level exactness and Yellow Pages database with high geolocation accuracy. Instinctively, the most comparable protest of doctor "John F. Smith MD" is "John Smith MD" in Yellow Pages, since the two names are fundamentally the same as and geologically nearer. In this way, finding the most comparable matches between two land databases requires a composite join task that thinks about the two kinds of qualities, textual and spatial. Such kind of join, specifically Spatial Similarity join, has gotten substantially less consideration in the exploration group than singular joins on either textual or spatial qualities. In the textual case, the level of likeness in a similarity join [1] [7] is estimated by a similarity work, e.g. Jaccard coefficient or Levenshtein separation, and sets that fulfill a client characterized similarity edge are incorporated into the yield. As of late, parallel processing with MapReduce, a parallel programming model proposed by Google [3], has been investigated to handle the scalability issue of this kind of joins [6]. In the spatial case, a spatial join [4] between two land datasets matches records in view of their spatial characteristics. The spatial connection might be communicated in a few ways, e.g. separate edge or polygon cover.

Associate Professor,

Computer Science and Engineering

Shivanadhuni Spandana

Student, M. Tech in Software Engineering

A STUDY OF AN ENHANCED APPROACH TOWARDS FREQUENT PATTERN MINING

Abstract

Association rule mining is one of the imperative errands in data mining. The undertaking to locate the frequent patterns is assuming a fundamental part in mining associations and numerous other intriguing highlights among the factors in the transactional database. In any case, this assignment is computationally escalated and utilizes a significant extensive measure of memory. There are numerous components that include the working of a frequent pattern mining algorithm. One of the variables that have a noteworthy impact is the attributes of the database being examined. The well known algorithm works distinctively on inadequate and thick database. Two algorithms are being connected to the database as indicated by the data attributes of the dataset. FEM(FP-Tree and Eclat Method) utilizes a settled edge as an exchanging condition between the two mining techniques while DFEM(Dynamic FP-Tree and Eclat Method) applies an edge dynamically at runtime to efficiently fit the qualities of the database amid the mining procedure. The execution

1 INTRODUCTION

One of the essential undertakings in data mining is association rule mining, which centers around discovering rules that determine the event of the things in the subsets in databases. [1] Frequent pattern mining is an imperative undertaking, used to locate the distinctive sorts of relationship among factors in expansive database. Its fundamental concentration is to look for itemsets, sub successions that co-happen with a base recurrence more noteworthy than the client determined help check. Apriori, FP-development, Eclat are a portion of the generally utilized algorithms. Apriori algorithm utilizes the property that an itemset happen frequently if and just if the majority of its sub-things are frequent. It uses a level-wise or expansiveness first approach of the itemset look space which essentially prunes all the superset. It likewise maintains a strategic distance from the age of any candidate age that has any infrequent subset.FP-development files the database for quick calculation of the help tally by means of the utilization of a data structure called the frequent pattern tree or the FP-tree. Tallying the help check can be enhanced essentially if the database is recorded such that it permits quick recurrence calculations. In level-wise approach, to figure the help tally, it is expected to create subsets of every exchange and check on the off chance that they exist in the prefix tree. Not at all like the level-wise approach Eclat algorithm utilizes the vertical TID rundown to locate the frequent itemset by crossing these TID rundown and then registering their resultant help check [5]. From different trials performed on various databases it is demonstrated that the ARM methods functioned admirably for a curious sort of databases [5], [2], [6], [7]and [4]. The methods either worked for scanty or thick database and inadequately on both. In this paper we talk about two effective algorithms , DFEM (Dynamic FEM) and FEM (FP-development and Eclat Mining) which is a blend of FP-development and Eclat algorithm [8] and [9]. It utilizes FP-tree to store the database minimalistically. The principle highlight of

K. MOUNIKA- STUDENT

V. CHANDRA SHEKHAR RAO-ASSOCIATE PROFESSOR

S. KIRAN – ASSISTANT PROFESSOR

Department of CSE, KITS.

Review of Feature Extraction Techniques for Character Recognition

ABSTRACT
This paper presents a comprehensive review of the feature extraction techniques for character recognition (CR) which will be helpful for the new researchers to understand the insight into the developments of the current research in the field of CR. Feature extraction plays a major role in the performance of the CR. The characteristics of the feature extraction techniques have to be independent of the scalable font characteristics such as type, size, style, tilt, rotation and should be able to describe the complex, distorted, broken characters effectively. A feature vector should be simple, reliable, complete, and compact to recognize any input character with high accuracy similar to human perception. A lot of research has been done on feature extraction techniques for optical CR for the past few decades. Most of the existing CR methods from the literature will work successfully for one or two fonts and they have used the combination of existing features to improve the accuracy. Therefore, we feel there is still scope to work on feature extraction techniques for the recognition of multilingual characters.

KEYWORDS
Character recognition; Feature extraction; Shape-based features; Non-shape-based features; Indian scripts; OCR survey

1. INTRODUCTION

Optical character recognition (OCR) is the process of recognizing the segmented characters from scanned images or standalone characters by the computer by extracting the invariant and discriminative features from the characters [1]. CR is one of the most challenging areas of image processing and pattern recognition with various practical application areas such as automatic license plate (LP) recognition, document analysis and recognition, reading aid for blind, banks, post offices, defense applications, etc. [2]. Initially, the development of OCR has been started as an aid to the visually handicapped during the first decade of the nineteenth century by the Russian scientist Tyurin. The modern OCR development started during the year 1950 because of the advent of digital computers and increased its development pace during mid-1970s due to the emergence of a lot of handwritten applications. During the mid-1980s, the OCR techniques have progressed very rapidly due to intense research and development in the field of OCR [3–6].

Present OCR systems have shown significant progress in the recognition of printed bilingual and multilingual characters extracted from old noisy printed documents with graphics, tables, mathematical symbols present in it and reported more than 95% recognition accuracies [2,7]. Even much work has been done on unconstrained handwritten characters recognition and the recognition accuracies are reported around 96% using single font [8,9]. At present, there are lots of OCR systems available for western countries script, which will work effectively for printed and handwritten characters with limited abilities to recognize unconstrained multilingual characters from non-western countries such as India. Therefore, there is still scope to work on feature extraction techniques for multilingual CR for the countries like India which can lead to the development of Omni-font CR [1].

In India, there are 22 official languages, namely, Asamiya, Bangla, Bodo, Dogri, Gujarati, Hindi, Kannada, Kashmiri, Konkani, Maithili, Malayalam, Manipuri, Marathi, Nepali, Odia, Punjabi, Sanskrit, Santali, Sindhi, Tamil, Telugu, and Urdu. There are around 12 different scripts existing to write these Indian official languages [1]. Developing an unconstrained multilingual OCR for the nations like India is very challenging because of the complex character grapheme of the Indian language scripts and not much effort is put in developing such a system [6,10]. In this paper, we have categorized feature extraction techniques of OCR into shape-based and non-shape-based features as shown in Figure 1. Shape-based features have the capability to describe the shape of the object independently and may get confuse while classifying the objects having similar shapes and which can be discriminated with the help of non-shape-based features.

The remaining sections of this paper are organized as follows. Section 2 elaborates on classification of OCR

Narasimha Reddy Soora -Associate Professor, CSE

Parag S. Deshpande, CSE,,NIT

A novel local skew correction and segmentation approach for printed multilingual Indian documents

KEYWORDS

Character recognition;
Character segmentation;
Document analysis;
Skew correction;
Rough skew;
Text line extraction;
Word segmentation

Abstract Till date, many Indian government organizations do not have robust software to search for words from scanned office documents having complex multilingual Indian scripts. Manual search of such a multilingual Indian document will take few minutes and there will be tens of thousands of documents to be searched for the desired content. Manual search of such a huge number of scanned Indian documents will be tedious, which requires robust automatic searching software. This led us to work toward indexing of aged printed multilingual Indian office documents. This paper presents a novel geometrical technique to group the components which belong to a text line of a document having multi-orientations and a novel approach to find the local skew of Devanagari word. The performance of the proposed technique was evaluated using 280 printed Indian documents with around 6000 text lines having English, Devanagari, and Marathi scripts and achieved 99% success rate for line segmentation indicates the legitimacy of the proposed method. To further assess the performance of the proposed method, we have considered publicly available Tobacco800 document image database and achieved significant performance results as compared with few of the prominent methods from the literature.

1. Introduction

To implement e-governance in an effective way, many government organizations are storing paper documents in digitized form. The aim of digitization was to supply information in less time and also to offer durability because electronic media is more durable than the paper media. To have an effective information access, it is necessary to index the documents [12]. So, the overall process of the task is to scan the documents, indexing the documents and integrating the documents with the e-governance application. The document scanning is done with high quality and high speed scanners and the most of the process is mechanical in nature. Indexing process requires identifying the keywords from the document and storing the words with the document id and word positions from the document. Currently, this process is manual and it requires around 5–6 min for one page document. This process becomes costly and time consuming, when the counts of documents are in millions. In Indian context, the task is more challenging because the keywords are multilingual.

* Corresponding author. Fax: +91 712 2223969.
E-mail addresses: snreddy75@yahoo.co.uk, 12dt04cse005@cse.vnit.ac.in (N.R. Soora), psdeshpande@cse.vnit.ac.in (P.S. Deshpande).

Peer review under responsibility of Faculty of Engineering, Alexandria University.

Narasimha Reddy Soora -Associate Professor, CSE

Parag S. Deshpande

CSE, VISVESVARAYA NATIONAL INSTITUTE OF TECHNOLOGY

A Review on Big Data Application in Health Care

ABSTRACT

Big data technologies are progressively utilized for biomedical and health-care informatics research. A lot of biological and clinical data have been created and gathered at a phenomenal speed and scale. For instance, the new age of sequencing technologies empowers the handling of billions of DNA sequence data every day, and the application of electronic health records (EHRs) is archiving a lot of patient data. The cost of getting and breaking down biomedical data is required to diminish drastically with the assistance of innovation redesigns, for example, the rise of new sequencing machines, the advancement of novel equipment and programming for parallel computing, and the broad extension of EHRs. Big data applications introduce new chances to find new information and make novel strategies to enhance the nature of health care. The application of big data in health care is a quickly developing field, with numerous new disclosures and philosophies distributed over the most recent five years. In this paper, we review and talk about big data application in four noteworthy biomedical sub disciplines: (1) bioinformatics, (2) clinical informatics, (3) imaging informatics, and (4) general health informatics. In particular, in bioinformatics, high-throughput tests encourage the research of new expansive affiliation investigations of diseases, and with clinical informatics, the clinical field benefits from the immense measure of gathered patient data for settling on smart choices. Imaging informatics is presently more quickly incorporated with cloud stages to share medical image data and work processes, and general health informatics use big data methods for foreseeing and observing infectious disease flare-ups, for example, Ebola. In this paper, we review the current advance and achievements of big data applications in these health-care domains and condense the difficulties, holes, and chances to enhance and progress big data applications in health care.

Keywords: Big Data, Literature Review, Health Care, Data-Driven Application

I. INTRODUCTION

In the biomedical informatics domain, big data is another worldview and a biological system that changes case-based investigations to huge scale, data-driven research. It is generally acknowledged that the qualities of big data are characterized by three noteworthy highlights, ordinarily known as the 3Vs: volume, assortment, and speed. To start with and most essentially, the volume of data is developing exponentially in the biomedical informatics fields. For instance, the ProteomicsDB8 covers 92% (18,097 of 19,629) of known human qualities that are explained in the Swiss-Prot database. ProteomicsDB has a data volume of 5.17 TB. In the clinical domain, the advancement of the HITECH Act9 has about tripled the reception rate of electronic health records (EHRs) in doctor's facilities to 44% from 2009 to 2012. Data from a great many patients have just been gathered and put away in an electronic arrangement, and these collected data could possibly improve health-care administrations and increment research opportunities.10,11 what's more, medical imaging (eg, MRI, CT filters) produces tremendous measures of data with considerably more perplexing highlights and more

Sridhar Gujjeti

Assistant Professor, CSE,KITSW

D Dr. Suresh Pabboju

Professor, IT, CBIT

A Study on Network Security Administration using the Technology of Data Mining

ABSTRACT:

A few Network systems are experiencing different security threats including network worms, extensive scale network assaults, and so forth, and network security situation awareness is a compelling route to tackle these issues. The general procedure is to see the network security occasions occurred in a specific day and age and the internet condition, artificial control of security data, investigate the assault practices systems endured, give the worldwide perspective of network security, and survey the entire security situation and foresee the future security patterns of the network.

INTRODUCTION

Security in network has been the endless hot research spot, and it has encountered three phases: obstruction, detection and fault. Not with standing, there are still some security issues left, for example, the confounded structure, and the kittle network assaults, so the network security issues are winding up increasingly stark. The existed proficient network security implies, similar to IDS, Firewall and VDS can't mirror the security status of the network. From that point onward, Tim Bas s, the American exceptional network security master, proposes the idea of network security situational awareness (NSSA), and builds up the system of network situation awareness, which means to take care of the existed network security issues from another purpose of view

Saloni Fathima

Faculty, Department of CSE, KU

Siripuri Kiran

Assistant Professor,Department of CSE, KITSW

Fuzzy Associative Classification Driven MapReduce Computing Solution for Effective Learning from Uncertain and Dynamic Big Data

Abstract

Handling uncertainty and dynamic changes in data sets supposed for analysis is always been challenging task for data analytics community. The same challenges even inherited to the embryonic big data analytics which are generally mentioned as veracity and velocity properties. Indeed, in case of big data, handling uncertainty and dynamism data could be more typical because of the scalability factor which is a result of data storage in distributed file system structure. In order to overcome difficulties of handling uncertainty and dynamic changes in big data analytics and considering efficiency provided by fuzzy associative classification techniques in handling uncertainty of data, we propose a dynamically scalable fuzzy associative classifier extraction model for Map-reduce framework. The important contributions of the paper is that it proposes a data driven fuzzy generalization approach for handling uncertainty, Tid-list based classification approach for easy scalable computation among multiple nodes and data chunks driven updating of classification model in case of dynamic changes to dataset with respect Map-reduce framework. The experimental evaluation results shows that the proposed Map-reduce model for fuzzy associative classification rule extraction can efficiently handle data uncertainty and dynamic changes to data stored in distributed file system, along with satisfying scalability factor.

Keywords: *Fuzzy associative classification, MapReduce, Data Veracity, Dynamic data*

1. Introduction

The digital data floating from various fields including e-commerce, social media, health and economic industry *etc.*, is continuously throwing challenges to knowledge processing community for providing effective storing and analytical techniques. The handling of this immense data resulted from human cognitive process is currently referred as big data analytics [1]. The primary challenge in big data analytics is handling scalability of ever increasing data. This scalability can be offered by distributed file system.

The distributed file system offers scalability by dividing the data into chunks and storing the chunks into disk and whenever the present disk is full the next chunk is stored in another disk. The other advantage of the distributed file system is its reliability which offered by storing same chunk of data into multiple disks by which even if a disk is get failed it can recover data from other disks to ensure high reliability. Out of different open source distributed file systems available Hadoop Distributed File System (HDFS) [2]

Raghuram Bhukya

Assistant Professor, CSE,KITSW

Jayadev Gyani

Assistant Professor, Computer Science and Information Technology College

Majmaah University, Majmaah, Kingdom of Saudi Arabia

Multiple Objective-Based Modified Exponential Gravitational Search Algorithm to VMM Strategy for Load Balancing in Cloud

Abstract: Cloud computing is a trending topic in the field of science and technology since the internet dependent services have been growing rapidly. In this environment, there are a lot of immense infrastructures and resources to satisfy the internet users. When a large number of service requests reach at a particular time, load balancing becomes a necessity. Load balancing involves the effective migration of the resources from the loaded physical machine to the other physical machine. For the effective migration, a method named Modified Exponential Gravitational Search Algorithm based on Virtual Machine Migration strategy (MEGSA-VMM) has been proposed that uses the gravitational concepts for performing the frequency-based velocity computations. MEGSA algorithm is the integration of the gravitational search algorithm and exponential weighted moving average theory. Also, the quality-of-Service (QoS) constraints considered for VM migration are migration cost, migration time, resource usage and energy. Simulation of the proposed method and the comparison of the results obtained, with the traditional methods like Ant-Colony Optimization (ACO), Gravitational Search Algorithm (GSA) and Exponential Gravitational Search Algorithm (EGSA) is performed. The proposed method is found to achieve an optimum migration with a minimum energy at a rate of 0.26 and minimum migration cost at a rate of 0.015.

Keywords: Cloud Computing, Load Balancing, Modified Exponential Gravitational Search Algorithm (MEGSA), VM Migration strategy (VMM), Fitness Function.

1. Introduction

Load balancing [1, 2] is the strategy to solve the uneven distribution of the resources in the network, which is the cause of increased data processing and data storage over the internet. Due to this uneven distribution, the overload and underload problems exist, which paved the way for load balancing [3, 4]. Load balancing [5, 6] is a term that contributes to the proper distribution of the resources among all the virtual machines available in the cloud [7-9]. Also, it does not place any single virtual machine loaded, but it distributes the workload among the multiple virtual machines [10, 11]. This distribution causes the improvement in the performance of the system and results in the proper resource utilization. Thus, load balancing is the trending strategy that contributes to achieving better resource utilization and better response time. Load balancing can be performed in two ways namely, distributed and non-distributed load balancing. In distributed balancing, load from the overloaded virtual machine gets distributed to the other virtual machines, whereas in the non-distributed load balancing, only one virtual machine is involved in load balancing [12-15].

The virtual machines are controlled using the large servers present in the data centers, and they share a lot of resources provided by the servers. These resources may get loaded in certain cases, and a perfect model is required, which computes the exact physical host for load balance [16]. Load balancing with the VMM strategy is the only solution to achieve the perfect load balance in the network. VMM strategy is simply the migration of the virtual machine from the loaded physical machine to the other physical machine. Also, it is essential to determine the physical host to undergo

P. VIJAY KUMAR -ASST.PROFFESOR, KAKATIYA INSTITUTE OF TECHNOLOGY AND SCIENCES

KAILA SHAHU CHATRAPATHI, JNTUH College of Engineering, Manthani, India

Exponential gravitational search algorithm-based VM migration strategy for load balancing in cloud computing

Abstract

With the advancement in the science and technology, cloud computing has become a recent trend in environment with immense requirement of infrastructure and resources. Load balancing of cloud computing environments is an important matter of concern. The migration of the overloaded virtual machines (VMs) to the underloaded VM with optimized resource utilization is the effective way of the load balancing. In this paper, a new VM migration algorithm for the load balancing in the cloud is proposed. The migration algorithm proposed (EGSA-VMM) is based on exponential gravitational search algorithm which is the integration of gravitational search algorithm and exponential weighted moving average theory. In our approach, the migration is done based on the migration cost and QoS. The experimentation of proposed EGSA-based VM migration algorithm is compared with ACO and GSA. The simulation of experiments shows that the proposed EGSA-VMM algorithm achieves load balancing and reasonable resource utilization, which outperforms existing migration strategies in terms of number of VM migrations and number of SLA violations.

Introduction:

In the recent epoch of the computer and information science, cloud computing is one of prominent "state of art" technologies. The cloud computing is a computing paradigm which provides platforms, resources and all the application as services. Cloud computing is defined as "a model for enabling ubiquitous, convenient, on demand network access to a shared pool of configurable computing resources (e.g.,nettworks,servers,storage,applicationsandservices)thatcanberapidlyprovisio ned and released with minimal management effort or service provider interaction".2 Generally, cloud computing consists of a bunch of distributed servers providing demanded services and resources to different clients in network with reliable and scalable data centers. The services in the cloud may be of software resources (Software as a service)3 or physical resources (Platform as a service),4 or hardware or infrastructure resources (Infrastructure as a service).5 Amazon EC2 is one of the examples for the cloud computing servicesvirtualization is a useful concept. Virtual Machine (VM) acts as a basic foundation unit for cloud.

1. P. VIJAY KUMAR & 2. KAILA SHAHU CHATRAPATHI, KITSW & JNTUH College of Engineering, Manthani, India

Decision Tree Analysis Tool with the Design Approach of Probability Density Function towards Uncertain Data Classification

Siripuri Kiran

ABSTRACT

Traditional decision tree classifiers are built utilizing certain or point data as it were. Be that as it may, in numerous genuine applications innately data is constantly uncertain. Quality or esteem uncertainty is characteristically connected with data esteems amid data gathering process. Traits in the preparation data sets are of two kinds – numerical (constant) and clear cut (discrete) characteristics. Data uncertainty exists in both numerical and all out characteristics. Data uncertainty in numerical qualities implies scope of qualities and data uncertainty in all out traits implies set or accumulation of qualities. In this paper we propose a technique for taking care of data uncertainty in numerical properties. One of the least difficult and most straightforward techniques for taking care of data uncertainty in numerical properties is finding the mean or normal or agent estimation of the arrangement of unique estimations of each estimation of a characteristic. With data uncertainty the estimation of a property is generally spoken to by an arrangement of qualities. Decision tree classification precision is tremendously enhanced when property estimations are spoken to by sets of esteems as opposed to one single delegate esteem. Probability density function with equal probabilities is one compelling data uncertainty demonstrating system to speak to each estimation of a property as an arrangement of qualities. Here the principle presumption is that genuine esteems gave in the preparation data sets are found the middle value of or delegate estimations of initially gathered esteems through data accumulation process. For every illustrative estimation of each numerical characteristic in the preparation data set, approximated values relating to the initially gathered esteems are created by utilizing probability density function with equal probabilities and these recently produced sets of qualities are utilized as a part of developing another decision tree classifier.

Keywords : Probability Density Function, Data Mining, Classification, Uncertain Data, Decision Tree, Machine Learning.

I. INTRODUCTION

Classification is a data investigation method. Decision tree is a capable and well known instrument for classification and forecast yet decision trees are predominantly utilized for classification [1]. Primary favorable position of decision tree is its interpretability – the decision tree can without much of a stretch be changed over to an arrangement of IF-THEN decides that are effortlessly justifiable [2].

Cases wellsprings of data uncertainty incorporat estimation/quantization blunders, data staleness, an various rehashed estimations [3]. Data minin applications for uncertain data are – classification o uncertain data, bunching of uncertain data, visi design mining, exception discovery and so forth Cases for specific data are – areas of college structures, schools, universities, eateries, railroa stations and transport stands and so on. Dat uncertainty normally emerges in an extensiv

Assistant Professor, Department of CSE, KITSW

Decision Tree Analysis Tool with the Design Approach of Probability Density Function towards Uncertain Data Classification

Siripuri Kiran

ABSTRACT

Traditional decision tree classifiers are built utilizing certain or point data as it were. Be that as it may, in numerous genuine applications innately data is constantly uncertain. Quality or esteem uncertainty is characteristically connected with data esteems amid data gathering process. Traits in the preparation data sets are of two kinds – numerical (constant) and clear cut (discrete) characteristics. Data uncertainty exists in both numerical and all out characteristics. Data uncertainty in numerical qualities implies scope of qualities and data uncertainty in all out traits implies set or accumulation of qualities. In this paper we propose a technique for taking care of data uncertainty in numerical properties. One of the least difficult and most straightforward techniques for taking care of data uncertainty in numerical properties is finding the mean or normal or agent estimation of the arrangement of unique estimations of each estimation of a characteristic. With data uncertainty the estimation of a property is generally spoken to by an arrangement of qualities. Decision tree classification precision is tremendously enhanced when property estimations are spoken to by sets of esteems as opposed to one single delegate esteem. Probability density function with equal probabilities is one compelling data uncertainty demonstrating system to speak to each estimation of a property as an arrangement of qualities. Here the principle presumption is that genuine esteems gave in the preparation data sets are found the middle value of or delegate estimations of initially gathered esteems through data accumulation process. For every illustrative estimation of each numerical characteristic in the preparation data set, approximated values relating to the initially gathered esteems are created by utilizing probability density function with equal probabilities and these recently produced sets of qualities are utilized as a part of developing another decision tree classifier.

Keywords : Probability Density Function, Data Mining, Classification, Uncertain Data, Decision Tree, Machine Learning.

I. INTRODUCTION

Classification is a data investigation method. Decision tree is a capable and well known instrument for classification and forecast yet decision trees are predominantly utilized for classification [1]. Primary favorable position of decision tree is its interpretability – the decision tree can without much of a stretch be changed over to an arrangement of IF-THEN decides that are effortlessly justifiable [2].

Cases wellsprings of data uncertainty incorporat estimation/quantization blunders, data staleness, an various rehashed estimations [3]. Data minin applications for uncertain data are – classification o uncertain data, bunching of uncertain data, visi design mining, exception discovery and so forth Cases for specific data are – areas of college structures, schools, universities, eateries, railroa stations and transport stands and so on. Dat uncertainty normally emerges in an extensiv

Assistant Professor, Department of CSE, KITSW

Anomaly Detection using Data Mining Techniques in Social Networking

Abstract: Nowadays, there exists a broad development in utilizing Internet in long range internet in social networking communication (e.g., texting, video collections, and so forth.), social insurance, online business, bank exchanges, and numerous different administrations. These Internet applications require a palatable level of security and protection. Then again, our computers are under assaults and defenseless against numerous dangers. There is an expanding accessibility of apparatuses and traps for assaulting and intruding networks. Anomalous exercises in social organizations speak to abnormal and unlawful exercises showing distinctive practices than others exhibit in a similar structure. This paper talks about various sorts of abnormalities and their novel order in view of different qualities. A survey of number of procedures for avoiding and distinguishing anomalies alongside fundamental suppositions and explanations behind the nearness of such inconsistencies is shrouded in this paper. The paper displays an audit of number of data mining approaches used to recognize anomalies.

Keywords: Anomalous activity, anomalies, Data mining techniques, Review analysis, Social Networking

I. INTRODUCTION

Anomaly detection intimates the design recognition in the given collection of information that doesn't fit in with a set up typical conduct. The examples subsequently identified are called anomalies and mean basic and significant data in a few application spaces. Anomalies are likewise alluded to as anomaly, astonishment deviation and so on. Most anomaly recognition calculations require an arrangement of simply typical information to prepare the model and they certainly expect that inconsistencies can be dealt with as examples not seen some time recently. Since an exception might be characterized as an information point which is altogether different from whatever is left of the information, in view of some measure, we utilize a few discovery conspires to perceive how proficiently these plans may manage the issue of outlier recognition. The measurements group has concentrated the idea of exceptions broadly. In these methods, the information focuses are displayed utilizing a stochastic appropriation and focuses are resolved to be anomalies relying on their association with this model. However with expanding dimensionality, it turns out to be progressively troublesome and erroneous to evaluate the multidimensional appropriations of the information focuses [1].

However recent anomaly discovery calculations that we use in this review depend on figuring the full dimensional separations of the focuses from each other and in addition on processing the densities of nearby neighborhoods. The deviation measure is our augmentation of the customary technique for anomalies recognition. As in anomalies identification, correlations are made amongst anticipated and genuine sensor values, and contrasts are deciphered to be signs of anomalies. This crude disparity is gone into a standardization procedure indistinguishable to that utilized for the value change score, and it is this portrayal of relative anomalies which is accounted for [2].

The deviation score for a sensor is least if there are no anomalies and most extreme if the disparity amongst anticipated and real is the best observed to date on that sensor. Deviation requires that a simulation be accessible in any frame for producing sensor value expectations. However the rest of the affectability and falling cautions measures require the capacity to reenact and prevail upon a causal model of the framework being checked. Affectability and falling cautions are an engaging approach to survey whether current conduct is irregular or not is by means of correlation with past conduct. This is the ideal of the unexpected measure. It is

Siripuri Kiran

Assistant Professor,Department of CSE, KITSW

Ajmera Rajesh

Assistant Professor, Department of CSE, KITSW

A Case Study on Issues in Privacy Preserving Data Mining

Abstract- *The development in data mining technology brings serious threat to the individual information. The objective of privacy preserving data mining (PPDM) is to safeguard the sensitive information contained in the data. The unwanted disclosure of the sensitive information may happen during the process of data mining results. In this study we identify four different types of users involved in mining application i.e. data source provider, data receiver, data explorer and determiner decision maker. We would like to provide useful insights into the study of privacy preserving data mining. This paper presents a comprehensive noise addition technique for protecting individual privacy in a data set used for classification, while maintaining the data quality. We add noise to all attributes, both numerical and categorical, and both to class and non-class, in such a way so that the original patterns are preserved in a perturbed data set. Our technique is also capable of incorporating previously proposed noise addition techniques that maintain the statistical parameters of the data set, including correlations among attributes. Thus the perturbed data set may be used not only for classification but also for statistical analysis.*

Keywords- Data Mining, Security, Issues & Remedies, Privacy, Preservation, development, technology, information, process.

I. INTRODUCTION

Data mining is frequently characterized as the way toward finding important, new correlation patterns and trends through non-trifling extraction of certain, already obscure data from extensive measure of data put away in repositories utilizing design acknowledgment and additionally statistical and mathematical techniques.

A Structured Query Language (SQL) is usually stated or written to access a specific data while data miners might not even be exactly sure of what they need. So, the result of a SQL query is usually a part of the database; whereas the result of a data mining query is an analysis of full contents of the database. Data mining tasks can be classified as follows:

2) Classification and prediction
3) Cluster analysis and outlier analysis
4) Web Data mining and search engines'.
5) Evolution analysis

The main focus of this thesis is to obtain secure Clustering results. Achieving accurate clustering results by providing privacy to sensitive data is trivial task. This thesis proposes two approaches for achieving the privacy for sensitive attributes during data mining [1].

Data Mining

Data mining also called as knowledge discovery in databases (KDD). Data Mining is defined as the "process of evaluating interesting, useful and hidden patterns from large volumes of data stores and identifies the relationships among the patterns" [2-4]. Data mining task requires utilities fir statistical data and Artificial Intelligence systems (AI). AI systems includes neural networks and machine learning sometimes one can combine them with database management system for evaluating or analyzing the huge volumes of digital data, which is the derived form of data sets..

Data mining has many applications; those have been listed in the above section. They can broadly categorized in to three area's one is business (insurance company, banking corporation, retail sector), second is science research (astronomy, medicine), and government security (detection of criminals and terrorists).

The large number of organizations, government and private data bases aims to ensure that the individual records are accurate and secure from unauthorized access. The tasks of data mining are targeted towards extracting hidden predictive knowledge about a group rather than the individual.

Figure 1 shows the Data mining process. First, data is collected from various sources in Data selection step. Next, Data will be pre- processed by dealing with null values and

Siripuri Kiran

Assistant Professor,Department of CSE, KITSW

Ajmera Rajesh

Assistant Professor, Department of CSE, KITSW

Real Time NIDS towards Architecture of Data Mining

Abstract

Intrusion detection technology is a successful way to deal with the problems of network security. In this paper, we introduce a data mining-based network intrusion detection framework in real time (NIDS). This framework is a disseminated design consisting of sensor, data preprocessor, extractors of highlights and identifiers. To enhance proficiency, our approach receives a novel FP-tree structure and FP-growth mining technique to extricate highlights in light of FP-tree without hopeful age. FP-growth is simply accord with the arrangement of real-time and updating data regularly as NIDS. We utilize DARPA intrusion detection assessment data set to train and test the attainability of our proposed strategy. Experimental results demonstrate that the execution is effective and palatable. Finally, the advancement pattern of intrusion detection technology and its as of now existing problems are quickly finished up.

Index Terms : FP-growth, Data Mining, Intrusion Detection, Network Intrusion Detection System(NIDS).

1. Introduction

With the advancement of the technology of information, particularly the pervasiveness of the technology of Internet/Intranet, security of more association and individual's PC framework foundation and information asset was debilitated. In this way, the security of information is turned into the a standout amongst the most critical errand in the domain of the technology of information. Conventional model of intrusion detection is been built up inefficient and the cost of research is to such an extent. The technology of data mining goes up against specific predominance in the domain of startling information acquiring. Along these lines Data Mining-based Intrusion Detection is turned out to be pervasive [1], [2], [3], [4]. Fundamentally, Network security is simply network information security. As a rule, all advances and hypotheses about mystery, integrality, ease of use, reality and controllable of network information are the examination domain of network security. Intrusion is an activity that tries to demolish that mystery, integrality and ease of use of network information, which is unlicensed and surpass expert. Intrusion Detection is a decidedly technology of security guard, which gets and examinations review data of PC framework and network from some network point, and to find whether there is the activity of disobeying security system and whether be attacked. Intrusion Detection System is the combination of programming and equipment of Intrusion Detection System. Whatever is left of the paper is sorted out as takes after. Segment 2 depicts how to separate highlights from review data. Area 3 outlines the main segments of our framework. Segment 4 reports the results of our tests. Segment 5 reaches the determination.

Siripuri Kiran

Assistant Professor,Department of CSE, KITSW

G. Rekha

Assistant Professor,Department of CSE, KIT

List of faculty papers published in conference proceedings:

S.No	Name of the Teacher	Title of the paper	Title of the proceedings of the conference	Name of the conference
1	Dr. P.Niranjan	"An Overview towards the Priority of Data mining in IOT systems "	Proceedings on Innovations and discoveries in science engineering and management(ICIDSEM-2018)held on 9-10 April 2018 at Vaagerwari college of engineering	International conference on Innovations and discoveries in science engineering and management(ICIDSEM-2018)held on 9-10 April 2018 at Vaagerwari college of engineering
2	Dr. P.Shireesha	"An Overview towards the Priority of Data mining in IOT systems "	Proceedings on International conference on Innovations and discoveries in science engineering and management(ICIDSEM-2018)held on 9-10 April 2018 at Vaagerwari college of engineering	International conference on Innovations and discoveries in science engineering and management(ICIDSEM-2018)held on 9-10 April 2018 at Vaagerwari college of engineering
3	Mr. S. Kiran	"An Overview towards the Priority of Data mining in IOT systems "	International conference on Innovations and discoveries in science engineering and management(ICIDSEM-2018)held on 9-10 April 2018 at Vaagerwari college of engineering	International conference on Innovations and discoveries in science engineering and management(ICIDSEM-2018)held on 9-10 April 2018 at Vaagerwari college of engineering
4	Mr. S.Venkatramulu	"CSES:cuckoo search based exploratory scale to defend input type validation vulnerabilities of HTTP Requests"co-authored by Dr Guru Rao .C.V	Proceedings on 2nd International Conference on Computational Intelligence and Informatics (ICCII-2017)held during 25-27 September	International Conference on Computational Intelligence and Informatics (ICCII-2017)held during 25-27 September

5	Mr. S. Narasimha Reddy	"Global Skew Detection and Correction using Morphological and statistical Methods"	Proceedings on International Conference on Computational Vision and Bio Inspired computing(ICCVBIC)	International Conference on Computational Vision and Bio Inspired computing(ICCVBIC)
6	Mr. Md Sharfuddin Waseem	"Global Skew Detection and Correction using Morphological and statistical Methods"	Proceedings on International Conference on Computational Vision and Bio Inspired computing(ICCVBIC)	International Conference on Computational Vision and Bio Inspired computing(ICCVBIC)

AN OVERVIEW TOWARDS THE PRIORITY OF DATA MINING IN IOT SYSTEMS

Abstract

Internet of Things is currently a quickening technology in the realm of devices. It encourages us interface every one of the gadgets which we use in our everyday tasks by means of the internet. Beginning from home, office, industry computerization to social insurance and brilliant urban areas internet of things has reformed the world by interconnecting them. Accordingly it produces monstrous volumes of data. For some, this data has huge business esteem and data. This is the place data mining becomes an integral factor which makes such sort of frameworks more sufficiently brilliant for better productivity and more noteworthy openings and administrations. This paper acquaints with the Internet of Things technology and states the need of data mining in our current reality where everything is conveyed over the internet and clarifies the procedure and appropriate calculations required for Internet of things.

Keywords:Data mining, Internet of things, Knowledge Data Discovery.

1

Siripuri Kiran

Assistant professor

Dr P. Niranjan

Professor

Dr .P. Shireesha

Assistant professor,

Department of CSE,

Kakatiya Institute of Technology and Science, Warangal, India

CSES: Cuckoo Search Based Exploratory Scale to Defend Input-Type Validation Vulnerabilities of HTTP Requests

Abstract Web application servers are prone to attacks that are more vulnerable and thousands of security breaches that are taking place everyday. Predominantly, the hackers to breach the web application systems security use the method of SQL injections and XSS models. IDS systems play a pivotal role in identifying the intrusions and alerting about the attacks. Despite that, there are numerous models of IDS systems in place; one of the commonly approached systems is the syntax analyzers. However, the limitations in terms of programming language dependency and the related issues drop the performance levels of syntax analyzer based strategies. To ensure the right kind of http request vulnerabilities, detection methods are in place; the Cuckoo Search based Exploratory Scale (CSES) to defend input-type validation vulnerabilities of HTTP requests is proposed here in this paper. The key objective of CSES is to magnify the speed and accuracy of input-type validation of web applications. The programming language dependency and server level process overhead issues do not impact the performance of CSES. In addition, the other key benefit of CSES model is optimal speed in search related to vulnerability scope detection. The experimental studies that are carried out on a dataset that contains the records prone to cross-site scripting, SQL injection alongside the normal records, depict better performance of the model, when compared to the other benchmarking model of DEKANT. CSES model has delivered improved accuracy levels in identifying the attacks.

Keywords XSS attack · Web application · SQL injection attack
CVE · CSES

S. Venkatramulu

Department of CSE,

K.I.T.S, Warangal, Telangana, India

C. V. Guru Rao

Department of CSE,

SR Engineering College, Warangal, Telangana, India

Global Skew Detection and Correction Using Morphological and Statistical Methods

Abstract. In this paper we have proposed a technique for skew detection and correction for printed documents, and have used an existing Optical Character Recognition (OCR) to recognize characters. The proposed algorithm has the following steps (a) Applying the morphological dilations by defining the various structure elements (SE) (b) extracting the longest connected components (CC) (c) finding the global skew angle by statistical analysis of connected component (d) reference text line estimation and regression line fit to rotate the individual line by estimated angle of rotation. We have conducted experiment using printed images having different languages i.e. English, Devanagari, and Arabic (custom dataset) and have achieved significant performance.

Keywords: Morphological dilations · Statistical analysis · Regression line fit Connected components analysis

1 Introduction

Most of the historical documents are preserved in digital format by scanning the document with proper dpi (dots per inch). To retrieve the information from these digital images, processing needs an optical character recognition to recognize the characters with the help of techniques such as [1, 2].

In document recognition systems, the quality of the input image is essential to the output performance. During the scanning process, adverse effects of tilting the document produces noise or skew, which are unavoidable. Many techniques have been proposed in the literature to overcome the noise. Many character recognition systems need a preprocessing of document to improve the efficiency; it involves noise removal, skew correction.

Narasimha Reddy Soora

Associate Professor

Sharfuddin Waseem Mohammed

Department of Computer Science and Engineering,

Kakatiya Institute of Technology and Sciences, Warangal, Telangana, India

Articles by Students:

S.No	Student Name	Article title
1	M. Sai karthik	Humming on the web
2	Jeevana	Players Unknown Battleground(PUBG) Backend
3	R. Vijaysai	Internet Of Things Core Features
4	P.Satya SaiTeja	Data Security through Cryptography
5	Jeevana	Know many or master One?
6	Maniteja	Dataset shift in Machine Learning
7	K.Ranjith	RAIN: Reliable Array of Independent Nodes

Humming on the web

Hummingbird is the embodiment of the team's experiments to bring Flutter to the web. It is a "web-based implementation of the Flutter runtime that takes advantage of the capability of the Dart platform to compile not just to native ARM code but also to JavaScript. This enables Flutter code to run on the standards-based web without change."

The project is still at the incipient stage and under heavy development but we do have some prototype concepts that we can go through.

Flutter is an open-source mobile application development SDK created by Google. It is used to develop applications for Android and iOS, as well as being the primary method of creating applications for Google Fuchsia.

M. Sai Karthik(B15CS065)

PLAYERS UNKNOWN BATTLE GROUND (PUBG) BACKEND:

A game is never developed in a single language. It has multiple language covering and but at the end it looks like it is one language that makes up as the core of software. Writing code for things where you feel interesting is much better than playing the game once someone has designed would be like time passing thing. So, coming right back to the present scenario people around us are the zombies of the most popular game now a days so called Pubg. People say it very interesting game for those who feel to do the things related to battle-royale, caching, gaming etc. Technically speaking most of them are fascinated about playing that game but I'm sure most of them doesn't concentrate on things running behind the game. It has 2 million users through android, windows, los. The craze for pubg is increasing day by day. Pubg is being played in mobile rather than PC. Pubg is well developed with realistic graphics. Here we will see programming languages used in pubg.

UNREAL Engine 4

This game is developed in special game IDE platform called unreal engine4.This engine is coded in c++ with the features of high portability and with realistic graphics. This mainly targets the eight generation-based PC console and android devices.

JAVA

Pubg is a cross-platform game which is available for PC as well as Mobile. So for developing the cross platform games java is must and this is considered as the side sourcing language. This is particularly declared as the official language for Android game Development. This code running from java is integrated with android studio as well as OpenGL for graphics makes the game more realistic.

PYTHON

The core of the game is developed by the language which most of the companies consider as the greatest, Python. Python makes the core of Pubg. It is Python that makes it so hack-proof. This provided developers with lot of built in functions for gaming and also makes this game more adaptable adn stable in different environments (as this python is portable in nature).This python is good for handling the server side work.

PEARL 6

Perl is a family of two high-level, general-purpose, interpreted, dynamic programming languages. Pearl stands for "Practical Extraction and Reporting Language".pearl has benefits such as powerful object orientations,

short coding, Syntactic simplification, Chained comparison.

JavaScript

JavaScript is an action-oriented language and is very capable language supporting different extra functions. This supports of different API and library.In the making of pubg developers used a minor portion of C#,PHP, MySQL etc. Pubg combat is very much inspired by the movie "Battle Royale". Pubg is considered as a Survival Game.

I hope this article helps you to get a glance about backend of Pubg.

Jeevana(B16CS166)

Internet Of Things core features:

Internet, these days is no longer just a means of communication. Internet of things connects everyday of physical things to the internet. It gives us the information or the details on the environment. In many fields, the Internet Of Things is going to change the way we do things. The IOT means that it doesn't just get the information and store them. It also responds intelligently as per the situation. IOT helps in predicting things and helps in Automation.

The device we use can be tracked all the time and the current information is being uploaded onto the internet. It is the sensors in devices that collect the information. Even if there is a sensor, there should also be a device called actuators that causes a machine or other device to operate. Sensor detects the data which must be analysed and the actuators responds to the collected data through the internet. Microcontroller device reads the input from the sensor and outputs the action to the actuator. We put this whole system in one thing and call it a device.

Both of them directly communicates through the internet and could send the data to the cloud where the data is being stored and is analyzed to give us the response. Every device using internet has its own IP address. So, we have to send a message to particular device we can access it through IP address. There are intermediate devices as well which hosts the network and are network between other devices.

It can solve almost all our issues. However, It also has disadvantages within. There should be frequent data transfer between the devices. If there are many devices that are connected to the internet, then the rate of getting it hacked can also be more. Security factor must be considered in this case. Which is why it takes at least more than a generation to make things fully automated.

There are key concepts such as Networking IOT-defines how IOT devices are connected, Programming IOT- the code we write behind to accept the data and to respond accordingly and securing IOT- where securing the devices is considered.

Most of them lose their jobs but lots of new jobs will be created in this area. We should not see AI and IOT as two different things. IOT can be worked successfully using AI by making things smart and automated.

Cases where IOT is being employed:

1. Automated cars: Consider a case where its fully automated. Automated cars have some IOT devices built-in which is GPS based and they could drive it for us. All the frequencies should be same among other devices. Our car will be communicating through the satellites with other vehicles and picks the fastest possible routes preventing the blockages in traffic. It sends a signal through GPS to other

vehicles which is connected over internet. What actually happens is that it can only sense electronic and internet-based signals but not the road rules which are specified to the driver. So, It takes years to work in fully autonomous mode.

2. It is being used to predict the output in future.
   It is used to control the functionality and notice errors if occurred. It keeps a track on the things.
   The wind Turbines are used to provide the wind energy. We could never guess when the repair may occur. But if we use IOT devices in there, we could predict when the wing of the turbine would worn out and if there are any defects occurred between the blades and the motor, it can be noticed. Of course, someone must monitor it all the time. But it cannot be done manually.

IOT can be used in various other platforms such as in Industries, Agricultural Sector, Power Plants, Education, Medical Field, to determine the condition of a historical monument or any such thing where prediction and responding intelligently is required.

R.VijaySai(B17CS008)

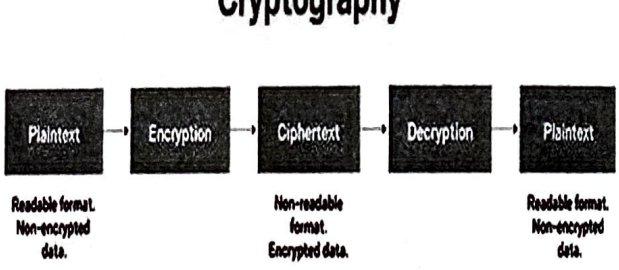
Data Security through Cryptography

Abstract:-

Cryptography is the science of transforming data, using a key, so that the data becomes unintelligible to those who do not have the key (or more precisely, the inverse transformation). Cryptography provides a means to secure data from unauthorized access. Using cryptographic functions, you can perform such operations as "scramble" (*encrypt*) the initially human-readable (*clear text*) contents of a character string so that the contents of the string is no longer humanly readable. If you are authorized, you can later "unscramble" (*decrypt*) such encrypted data so that it is humanly readable again. You can also ensure the integrity of data to verify that it has not been changed in an unauthorized manner. A symmetric cryptosystem is one in which the encryption key and the decryption key are the same. An asymmetric cryptosystem is one in which the decryption key is different from the encryption key. This article describes the various Security features of Some Cryptography Algorithms.

I. Introduction:

Cryptography is the art of secret writing which is used since Roman times to hide information secret or keeping message secure. To keep information secret, a widely used method is an encryption/decryption. Basically, encryption/decryption are the fundamental functions of cryptography. In encryption, a simple message (plain text) is converted into unreadable form called cipher text. While in decryption, a cipher text is converted into the original text (plaintext). Both of these functions are used to secure message against who is not authorized to view the message contents. The

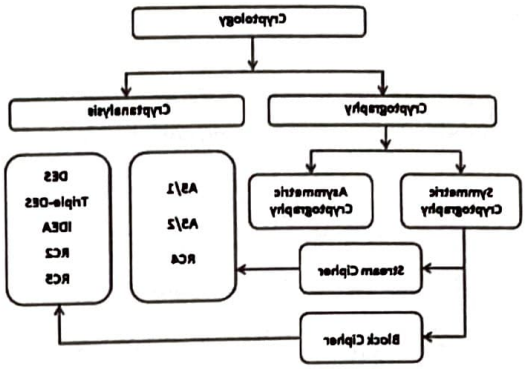


simple working of encryption and decryption functions is shown in Fig. 1.

Cryptography is an important aspect when we deal with network security. 'Crypto' means secret or hidden. Cryptography is the science of secret writing with the intention of keeping the data secret. Cryptanalysis, on the other hand, is the science or sometimes the art of breaking cryptosystems. These both terms are a subset of what is called as Cryptology.

Fig.1 Working of Cryptography

Cryptography is method of design and analysis of mathematical techniques that enable secure communications in the presence of malicious eavesdroppers. It aims to send information to a proper receiver without giving any information to a third party for misuse of data. Cryptography is method of protecting sensitive information as it is stored and transmitted through network communication channel. Cryptography operates by a sending scrambling or encrypting the original data by using different ways that obscures its meaning. The encrypted message is transmitted, and the receiver will recovers the message by using the same way the sender scrambled unscrambling or decrypting the transmission. Today, ciphers are used where the algorithm for encryption and decryption which could be revealed to anybody without compromising the security of a particular message to third parties.So, in such cases ciphers provides a set of specific parameters, called a key, is used along with the data as an input to the encrypting algorithm. This encrypted data along with the key serves as the inputs for the decrypting algorithm. The encrypting and decrypting algorithms mainly depends on the secrecy of the key. To prevent the key from being discovered by accident or systematic search, the key is must be chosen as a very large number. When the connection is established and Once the key is established, at same time secure communication can take place by sending converted crypto text, even over a public channel that is vulnerable to total passive eavesdropping. The main problem in cryptography is the maintain key distribution, for which there are essentially two solutions: one based on classical that is mathematical logics, and one based on Physics (quantum cryptography).But classical cryptography depends on the computational difficulty of factoring large integers, quantum cryptography relies on what we believe to be the universal laws of quantum mechanics. The classical cryptography is divided into two typessymmetric and asymmetric systems. The main aim of the secret key is to protect the data from eavesdroppers. The mathematical methods help the present cryptography methods to stay secure. For instance, an overnight breakthrough in mathematics or logics could make electronic money instantly worthless. To avoid such major economic and



social risks, there is no alternative but to turn to symmetrical cryptosystems.

I. Classification of Cryptography

Cryptanalysis:

Cryptanalysis is the study of ciphertext, ciphers and cryptosystems with the aim of understanding how they work and finding and improving techniques for defeating or weakening them. For example, cryptanalysts seek to decrypt ciphertexts without knowledge of the plaintext source, encryption key or the algorithm used to encrypt it; cryptanalysts also target secure hashing, digital signatures and other cryptographic algorithms.

A. Symmetric Cryptography :

Symmetric cryptography is placed in the category of cryptography schemes in which a shared key is used to convert a plain text into cipher text. A same secret key is shared by both

sender and receiver. Followings are the symmetric cryptography schemes.

- **DES (Data Encryption Standard):**

DES stands for Data Encryption Standard. DES introduced in early 1970 at IBM. The early design of DES is based on Horst Feistel. DES is a symmetric cryptographic algorithm used for encryption and decryption of message. In DES, only one secret key is used for both encryption and decryption. The key size of DES is 56-bit. To perform encryption/decryption, the sender and receiver must have the same key. The DES performs encryption on a block of 64-bit. The DES algorithm is most widely used in many applications and some popular use in military, commercial, and security of communication system , same as DES but key size is different from DES. The key size of 3DES is 168 bit. The 3DES algorithm performs operation three times on each block of data. It is slower than DES .

Algorithm:

```
functior. DES_Encrypt (N, K) where N = (L, R)
        N←IP(N)

        For round ←1 to 16
            do Ki ←SK (K,
            round) L ←L xor
            F(R, Ki) swap(L,
            R)

        Erd

        swap(L, R)
            N←IP¹(N)

        return N

        End
```

- AES (Advanced Encryption Standard):

AES stands for Advanced Encryption Standard which is the advancement of 3DES algorithm. It was introduced in 1997 by the NIST (National Institute of Standards and Technology). Basically, AES is based on the Rijndael cipher developed by two cryptographers, Joan Daemen and Vincent Rijmen. AES is different from DES and 3DES due to variables key sizes such as 128, 192, and 256 bits . Same like DES and 3DES, AES also performs encryption on blocks which are 128-bit . AES algorithm use in small devices for encrypting a message to send over a network. Some other applications are monetary transaction and security applications.

```
Cipher(byte[] output, byte[] input)
{ byte[4,4] State;
copy input[] into State[] AddRndKey
        for (rnd = 1; rnd < Nr-1; ++rnd)
        {
        SubBytes ShiftRows MixColumns AddRndKey
        }
                SubBytes ShiftRows AddRndKey
                        copy State[] to output[]
}
```

- **Blowfish and Twofish**: Both of these block ciphers are free to use and popular among e-commerce platforms for protecting payment information. They were created by the same person and offer symmetric encryption with keys varying in bit length. Twofish is the successor and offers longer encryption keys.

Asymmetric Cryptography

Asymmetric cryptography is also in the category of cryptography schemes. Unlike symmetric cryptography, two keys are used: one is public and second is private. The public key is shared by anyone in the cryptographic system while the private key is kept secret by authenticated user. Followings are the asymmetric cryptography algorithms.

- **RSA (Rivest, Shamir and Adleman):**

RSA stands for Rivest, Shamir and Adleman who introduced the RSA algorithm in 1977. RSA is an asymmetric cryptographic algorithm which is also used for encryption and decryption of the message. RSA is widely used in transferring of keys over an insecure channel. Due to asymmetric nature, there are two keys used in the algorithm. One is public key and second is a private key. The public key is openly accessible to everyone in the cryptosystem and the private key is kept secret by authorized person. RSA provides

confidentiality, integrity, authenticity, and nonrepudiation of data. RSA is more commonly used in electronic industry for online money transfer . In future, RSA can be used in Java cards .

• ElGamal:

ElGamal algorithm was introduced in 1985 by Taher ElGamal . ElGamal is an asymmetric key encryption algorithm that is based on the Diffie-Helman key exchange as an alternative to RSA for public key encryption. ElGamal is also used in digital signature generation algorithm called ElGamal signature scheme. A homomorphic algorithm named paillier used for its semantic security .

• ECC (Elliptic Curve Cryptography):

ECC stands for Elliptic Curve Cryptography. ECC introduced in 1985 by Neal Koblitz and Victor S. Miller. ECC lies in the category of the asymmetric scheme that is based on elliptic curves. The applications of ECC are encryption, digital signatures and pseudo-random generators.

II. Comparisons of Algorthims:

Algorithm	Key Size (Bits)	Block Size (Bits)
DES	64	64
3DES	192	64
Rijndael	256	128
Blowfish	448	64

III. Cryptography Protocols:

A cryptographic protocol is defined as a series of steps and message exchanges between multiple entities in order to achieve a specific security objective ‰ Properties of a protocol (in general): ‰ Everyone involved in the protocol must know the protocol and all of the steps to follow in advance ‰ Everyone involved in the protocol must agree to follow it ‰ The protocol must be unambiguous, that is every step is well defined and there is no chance of misunderstanding ‰ The protocol must be complete, i.e. there is a specified action for every possible situation ‰ Additional property of a cryptographic protocol: ‰ It should not be possible to do or learn more than what is specified in the protocol

IV. Applications of Cryptographic Protocols :

1. Key exchange ‰
2. Authentication ‰
3. Data origin authentication ‰
4. Entity authentication ‰
5. Combined authentication and key exchange
6. Secret splitting ‰
7. Secret sharing ‰
8. Time-stamping ‰
9. Key escrow (ensuring that only an authorized entity can recover keys)
10. ‰ Zero-Knowledge proofs Blind signatures ‰ Secure elections ‰ Electronic money

V. Attacks on Cryptography:

1.Ciphertext-only attack

2.Known-plaintext attack

3.Chosen-plaintext attack

4.Man-in-the-middle attack

5. Brute-force attack

VI. Conclusion:

Cyber Attacks are constantly increasing and evolving in new methods.As to stop them from accessing the private data of user or clients,the cryptography techniques are also evolving in different forms by using the complex mathematical equations and difficult logics to confuse and make the data secure.These are some

of the Cryptography algorithms mainly used in various industries and business.The advancement of the algorithms are testing and work against the cyber attacks that takes place in the present computer era.

P.Satya Sai Teja

KNOW MANY OR MASTER ONE?

No, We do not need so many programming languages. In fact, we need a programming language where it is perfectly fit for our use which would do analysis, perform calculations, alter user experience, sort out performance and accessibility issues and everything needed. This isn't an ideal world and the Technology is varying, so are the requirements. Hence, we need more than a few programming Language to perform all these tasks.

What makes programming language different is that the way they operate and the way they deliver the user experience. we cannot make a programming language and hope it will work for everything under the sun. If you're to develop iOS apps you don't need to dig into java, c, python and c#- all you need is to learn SWIFT Language.

The number of programming languages that a person should learn depends on his/her interests and needs. You really do not have to learn every programming language available to you. Instead you could learn one or two and master them. If all the operations of Facebook could be handled by a CEO itself, Mark Zuckerberg would have never paid millions to hire experts from all around the globe. He needs people that are really good at their chosen language, not someone who knows a little bit about a lot of languages. The point I'm trying to say is that learning a single programming language and mastering that is far better than being jack of all and master of none. So pick the language that suits your interests. There are several platforms out online where you can learn programming and explore the real world problem statements.

I personally suggest hackerrank.com. Google the website and you will find more challenging problems which are to be solved. Find your interests and start practicing it regularly and be eager to learn and evolve as the language evolves.

Jeevana(B16CS166)

Dataset shift in Machine Learning

Dataset shift is a common problem in predictive modeling that occurs when the joint distribution of inputs and outputs differs between training and test stages. Covariate shift, a particular case of dataset shift, occurs when only the input distribution changes. Dataset shift is present in most practical applications, for reasons ranging from the bias introduced by experimental design to the irreproducibility of the testing conditions at training time. (An example is -email spam filtering, which may fail to recognize spam that differs in form from the spam the automatic filter has been built on.) Despite this, and despite the attention given to the apparently similar problems of semi-supervised learning and active learning, dataset shift has received relatively little attention in the machine learning community until recently. This volume offers an overview of current efforts to deal with dataset and covariate shift. The chapters offer a mathematical and philosophical introduction to the problem, place dataset shift in relationship to transfer learning, transduction, local learning, active learning, and semi-supervised learning, provide theoretical views of dataset and covariate shift (including decision theoretic and Bayesian perspectives), and present algorithms for covariate shift.

Maniteja(B15CS063)

RAIN: Reliable Array of Independent Nodes

RAIN (also called channel bonding, redundant array of independent nodes, reliable array of independent nodes, or random array of independent nodes) is a cluster of nodes connected in a network topology with multiple interfaces and redundant storage. RAIN is used to increase fault Tolerance. It is an implementation of RAID across nodes instead of across disk arrays.

RAIN can provide fully automated data recovery in a local area network (LAN) or wide area network (WAN) even if multiple nodes fail. A browser-based, centralized, secure management interface facilitates monitoring and configuration from a single location. There is no limit to the number of nodes that can exist in a RAIN cluster. New nodes can be added, and maintenance conducted, without incurring network downtime.

RAIN originated in a research project for computing in outer space at the California Institute of Technology (Caltech), the Jet Propulsion Laboratory (JPL), and the Defence Advanced Research Projects Agency (DARPA) in the United States. The researchers were looking at distributed computing models for data storage that could be built using off-the-shelf components.

The idea for RAIN came from RAID (redundant array of independent disks) technology. RAID partitions data among a set of hard drives in a single system. RAIN partitions storage space across multiple nodes in a network. Partioning of storage is called disk striping. Several patents have been granted for various proprietary versions of RAIN.

Ranjith Katta (B15CS042)

KAKATIYA INSTITUTE OF TECHNOLOGY & SCIENCE
WARANGAL
ERASILA EDUCATION SOCIETY

Estd : 1980

KITSW

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING