VOLUME NO. X // AUGUST 2019

roCkSE

KAKATIYA INSTITUTE OF TECHNOLOGY & SCIENCE

(An Autonomous Institute under Kakatiya University, Warangal)

DEPARTMENT OF CSE

VIEWS

DR. K.ASHOKA REDDY

PRINCIPAL, KITSW

"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. P. NIRANJAN

PROF & HEAD
DEPARTMENT OF CSE

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

KAKATIYA INSTITUTE OF TECHNOLOGY & SCIENCE

(An Autonomous Institute under Kakatiya University, Warangal)
Warangal, Telangana -506015

Department of
Computer Science and Engineering

Presents….

roCkSE

A Technical magazine

Volume -X

AUGUST 2019

CONTENTS

TECHNICAL MAGAZINE 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.

FACULTY PAPERS IN JOURNALS:

S.No	Name of the Author	Title of the paper	Name of journal	Page no.
1	**Dr P. Niranjan**	A Study on the Limitations of Blockchain and the usage of Consensus Models in Blockchain Technology	International Journal of Research	5
		Cloud Computing Product Information Retrieval by Using Secure and Efficient Methods		6
		Efficiency-Based Analysis of Homomorphic Encryption Implications, Advances in Intelligent Systems and Computing		7
		An Overview towards the Priority of Data Mining in IOT Systems	International Journal of Pure and Applied Mathematics	8
		Quantifying Location Privacy, Global Journal of Engineering Science and Researches		9
2	**Dr. V Shankar**	Enhanced Disease Prediction and Notification Mechanism over Medical Big Data	IJR	10
		Automatic testing of Modern Web Applications	GJESR	11
3	**V. Swathy**	A retrieval technique for software reusable components using static and dynamic keyword ontology	IJSRCSAM	12
		A discretization technique based on behaviour for normalizing diversified code reuses repositories	IJRECE	13

4	**C. Srinivas**	Feature Vector based Clustering for Software Component Reuse	ICEMIS	14
		Big Data analytics Transformations and Tools for Bigdata mining and analysis	International Journal of Research	15
5	**S. Venkatramulu**	EC-CSES: Ensemble Classification using Cuckoo Search based Explorative Scale for Input Parameter Vulnerabilities of HTTP Requests	Journal of Advanced research in Dynamical and Control Systems	16
6	**V. Chandra Shekhar Rao**	A Novel approach for Iceberg Query Evaluation on Multiple Attributes Using Set Representation	Periodicals of Engineering and Natural Sciences (PENS), vol 6, issue no 2,2018, page no 302 - 315	17
		A Glimpse on Iceberg Query Evaluation Techniques	International Journal of Engineering & Technology (IJET), vol 7, 2018,997 - 1000	18
7	**Dr P. Shireesha**	An Overview towards the priority of data mining in IOT systems	International journal of pure and applied mathematics (IJPM), vol 8, issue 3, issn no 2236-6124, march 2019, 459 - 465	8
8	**B. Srinivas**	Hadoop Map reduce for Tactical Clouds	IOSR Journal of Computer Engineering (IOSR-JCE), vol 20, issue 4, ISSN no 2278-8727, 2018, 61-68	19
		Predicting Diseases-Emergency Department Visits using Big Data with Privacy Preservation	International Journal of Engineering Science and Computing, February 2019, 2018	20
9	**B. Sridhara Murthy**	Efficient Data Storage Mechanism in Cloud Computing	Global Journal of Engineering Science and Researches, June 2018, 16 - 21	21
		An Overview of IOT Healthcare and Applications	Research Review Journal, vol 4, issue no 1, Jan 2019, 645 - 649	22

10	**P. Vijay Kumar**	Degsa-VMM: dragonfly-based exponential gravitational search alg. To VMM strategy for load balancing in Cloud Computing	Kyberbetes (SCI), 2018	23
11	**N. C. Santosh Kumar**	Optimized Principal Curvatures Based Efficient Segmentation of Blood Vessels from Retinal Fundus Images	Biomedical Research, 2018	24
12	**K. Vinay Kumar**	A Similarity Function for Feature Pattern Clustering and High Dimensional Text Document Classification	Springer, 2018	25
		A Similarity Function for Feature Pattern Clustering and High Dimensional Text Document Classification	Springer, 2018	25
13	**V. Gowthami**	A Comprehensive study on Machine Learning	JETIR, vol 6, issue 3, Issn no 2349-5162, march 2019, 559-565	26
14	**S. Kiran**	Decision Tree Analysis Tool with the Design Approach of Probability Density Function towards Uncertain Data Classification	Journal of Scientific Research in Science and Technology, 2018	27
		Anomaly Detection using Data Mining Techniques in Social Networking	International Journal for Research in Applied Science & Engineering Technology, 2018	28
		A Case Study on Issues in Privacy Preserving Data Mining	International Journal for Science and Advance Research in Technology, 2018	29
		A Study on Network Security Administration using the Technology of Data Mining	International Journal of Pure and Applied Mathematics, 2018	30
		An Overview towards the Priority of Data Mining in IoT Systems	International Journal of Pure and Applied Mathematics, 2018	8
		A Study on Applications of IoT	Indian Journal of Public Health Research & Development, 2018	27
		A Study on the limitations of Blockchain and the usage of consensus models in Blockchain Technology	International Journal of Research, 2018	5

		Providing Location Proofs for Mobile Users with Privacy Guarantees	Providing Location Proofs for Mobile Users with Privacy Guarantees	30
15	**A. Praveen**	A Review towards the Development of Efficient Skyline Query	Journal of Emerging Technologies and Innovative Research (JETIR) , vol-5,issue-9,OCT- 2018, ISSN 2349-5162.	31
16	**N. Gayatri**	Incremental feature selection method for software defect Prevention	IJRTE, vol 8, issue 2S3, Issn no 277-3878,July 2019, 1345-1353	32
		Index verification based on MMB Cloud service	IJR, vol 8, issue 5, Issn no 2236-6124, may 2019, 1181-1198.	33

A STUDY ON THE LIMITATIONS OF BLOCKCHAIN AND THE USAGE OF CONSENSUS MODELS IN BLOCKCHAIN TECHNOLOGY

SIRIPURI KIRAN[1], P. NIRANJAN[2], P. SHIREESHA[3]

[1,3]Assistant Professor Department of CSE, Kakatiya Institute of Technology and Science, India

[2]Professor, Department of CSE, Kakatiya Institute of Technology and Science, India

Abstract

Threats to data ethics are consequently paramount relevance, as tampering with data can maliciously affect crucial business decisions. This issue is especially true in cloud computing surroundings , where data owners can't get a handle on fundamental info elements, such as the physical storage of data as well as also the control of its own Travels. Blockchain has just emerged as a fascinating technology which, among others, provides persuasive qualities about data integrity. Using the blockchain to manage data integrity threats is apparently a natural choice, however, its existing limitations of non throughput, higher latency, and weak stability interfere with the practical feasibility of virtually any blockchain-based solutions. At their basic levelthey empower a network of users to record trades in a shared ledger within that community, for example under normal operation of this blockchain system no trade may be changed once released. This record provides a high level technical overview of blockchain technology. The purpose is to help readers comprehend just how blockchain technology works. **Index Terms:** consensus model, blockchain, cryptocurrency

I. Introduction

Blockchains are evident and tamper immune electronic ledgers employed in a dispersed manner (i.e.(with no fundamental repository) and commonly without a fundamental power (i.e., a banking, organization, or authorities). In their essential amount they empower an network of customers to directly list trades within an shared ledger inside of that area, for example under ordinary performance of their blockchain system no trade might be shifted once released. Back in 2008the blockchain notion was united with a lot of different engineering and calculating theories to generate modern crypto currencies: digital dollars shielded as a result of cryptographic mechanics rather than a fundamental repository or jurisdiction. The very first these blockchain established crypto currency has been Bit coin.The many elements of blockchain technology together side its dependence on cryptographic primitives and dispersed technologies may create it hard to comprehend. Nevertheless, every component might be explained only and employed like a construction block to know the bigger

International Journal of Research ISSN NO:2236-6124

Cloud Computing Product Information Retrieval by Using Secure and Efficient Methods

K. NAVYA[1], Dr P. NIRANJAN REDDY[2]

[1]Student [2]Professor, Head of CSE Dept. KITS Warangal, India.

Abstract: In Cloud registering, cloud clients can remotely store their information into the cloud in order to appreciate the on-request top notch applications and administrations from a mutual pool of configurable processing assets. The way that information proprietors and cloud server are no longer in the equivalent believed space may put the redistributed decoded information in danger. It pursues that touchy information must be scrambled preceding re-appropriating for information security. Nonetheless, information encryption makes successful information use an exceptionally testing undertaking given that there could be a lot of re-appropriated information records. Additionally, redistributing information to the cloud causes loss of authority over information on an information proprietor's part. This loss of command over information is additionally heightened with the absence of dealing with clients' entrance to the information from down to earth distributed computing points of view. We address these difficult issues utilizing Ranked Searchable Symmetric Encryption Scheme.

Keywords: Product information retrieval, cloud computing, secure data access, keyword search.

I. Introduction: Cloud storage administrations enable the clients to re-appropriate their information in the distributed storage servers and recover them at whatever point and wherever required. This keeps away from the expense of structure and keeping up their information store. Be that as it may, the clients need to give protection to the information and furthermore to have the capacity to seek it without losing security. The clients dependably look through their reports through catchphrase in plaintext, which may spill security of clients in distributed storage condition. So, permitting a cloud specialist organization (CSP), whose design is predominantly for making a benefit, to take the authority of delicate information, raises fundamental security and protection issues.

Efficiency-Based Analysis of Homomorphic Encryption Implications

P. NIRANJAN[1], V. GOWTHAMI[2]

[1]Professor, Department of CSE, Kakatiya Institute of Technology and Science, India

[2]Assistant Professor Department of CSE, Kakatiya Institute of Technology and Science, India

ABSTRACT:

Homomorphic-based encryption incorporates authenticated encryption techniques which permits for computational principles generated on encrypted data with exclusive of necessitate execution of the decryption-based key. The traditional encryption schemes are used to provide the private environment for outsourcing the data storage to the third party end users, which will not work out for proving high-level security for avoiding data leakage-based uncertainties while the data is in different modes and also especially when the data without initially decrypting it. Huge data loss can occur with the latest TLS and SSH based attacks with lack of proper implementation of encryption methods with concern precautions. There are some other complex problems by using the traditional computational techniques such as privacy is concern when processing the confidential data through the third-party providers; integrity of confidential data to facilitate the secure access and association from the illegal execution of entire data and secreted unofficial accessibility. Homomorphic-based hybrid encryption will overcome these complexities and perform the encryption-based derivations by exclusive of initial decrypting the data. The numerous dedicated environments such as searchable-based encryption, deterministic-based encryption, order-preserving-based encryption, partial and fully homomorphic-based encryptions consent to precise group of computations to be derived on encrypted data for implications in real life circumstances for real-life circumstances. This paper describes the hybrid-based encrypted approaches, methodology implications, applications and lagging issues and focuses on the efficiency analysis of homomorphism encryption implications along with how homomorphic encryption-based algorithms have composed for realistic complex problems and described the analysis of past designed homomorphic-based encryption approaches with their implications and advances.

International Journal of Pure and Applied
Mathematics Volume 120 No. 6 2018, 7393-7406
ISSN: 1314-3395 (on-line version)
url: http://www.acadpubl.eu/hub/
Special Issue

AN OVERVIEW TOWARDS THE PRIORITY OF DATA MINING IN IOT SYSTEMS

Siripuri Kiran[1], Dr P. Niranjan[2],
Dr .P. Shireesha[3],

[1,3]Assistant professor, [2]Professor,
Department of CSE,
Kakatiya Institute of Technology and Science,
Warangal, India

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.

[Manasa, 5(6): June 2018] ISSN 2348 – 8034
DOI- 10.5281/zenodo.1262227 Impact Factor-5.070

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

QUANTIFYING LOCATION PRIVACY

M. Manasa*[1], Dr P. Niranjan[2]& Dr P. Shireesha[3] *[1]M.Tech Student, Department of CSE, Kakatiya Institution of Technology and Science, Warangal District, Telangana, India

[2,3]Professor & HOD, Department of CSE, Kakatiya Institution of Technology and Science, Warangal District, Telangana, India

[3]Professor, Department of CSE, Kakatiya Institution of Technology and Science, Warangal District, Telangana, India

ABSTRACT

The speedy progress in human-genome sequencing is resulting in a high convenience of genomic information. This information is notoriously terribly sensitive and stable in time. it's conjointly extremely related to among relatives. A growing variety of genomes are getting accessible on-line (e.g., as a result of escape, or once their posting on genome-sharing websites). What are then the implications for kin genomic privacy? We have a tendency to formalize the matter and detail economic reconstruction attacks supported graphical models and belief propagation. With our approach, Associate in nursing aggressor will infer the orderings of the relatives of a personal whose genome or phenotype(s) are discovered, by notably wishing on Mendel's Laws, applied mathematics relationships between the genomic variants, and between the phenotypes and therefore the variants. We have a tendency to evaluate the result of those applied mathematics relationships on privacy with relation to the number of discovered relatives and variants. We have a tendency to conjointly study however the recursive performance evolves after we take these varied relationships under consideration. What is more, to quantify the extent of genomic privacy as a result of the projected abstract thought attack, we have a tendency to discuss doable definitions of genomic privacy metrics, and compare their values and evolution. Genomic information reveals botanist disorders and therefore the chance of developing severe diseases like Alzheimer's. We have a tendency to evaluate our approach on actual genomic information from a pedigree and show the threat extent by combining information gathered from a genome-sharing web site and from a web social network. Finally, we have a tendency to show however further data of constitution info will improve the abstract thought attack's success.

8. INTRODUCTION

MOBILE devices, starting from good phones to connected vehicles, supply a large vary of location based mostly services (LBS) like navigation, ride-sharing, eating recommendations, and automobile collision warnings. LBS applications square measure exploding in quality, e.g., Uber, Google Maps, Yelp, and connected vehicles serve tens to many scores of users per day. However, these widespread, necessary services impose important privacy threats to their users as a result of they need access to the placement info of mobile devices. mass with different collected personal information, this info permits adversaries to infer sensitive info that goes so much on the far side user location: their habits, relationships, employments, hobbies, etc. Such privacy compromises may be launched by varied forms of adversaries: the LBS system might compromise users' privacy by mercantilism non-public location info to advertisers; malevolent employees of LBS systems will access users' info for fun or profit (as exemplified during a recent Uber scandal [4], [5]); And cybercriminals might forced an entry the placement info of an LBS system [6] or launch Sybil attacks [7], [8]. This work was supported by National Science Foundation underneath grants CCF 0844725 and CCF 1421957 attributable to the importance of privacy in LBS systems, researchers have devised location privacy protection mechanisms (LPPMs) [9]–[15]. Existing LPPMs square measure typically tailored to specific LBS systems and may be classified into 2 main classes: identity perturbation LPPMs [12], [14],

Enhanced Disease Prediction and Notification Mechanism over Medical Big Data

MOUNIKA KALLURI
M. Tech Student

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
KAKATIYA INSTITUTE OF TECHNOLOGY AND SCIENCE
WARANGAL, TELANGANA- 506 015

Dr V. SHANKAR
Associate Professor

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
KAKATIYA INSTITUTE OF TECHNOLOGY AND SCIENCE
WARANGAL, TELANGANA- 506 01

By rapid phase of development, the occurrences of rare diseases increased and also lot many treatments to treat them are developed, by big data advancement and healthcare communities accurate analysis of patient medical records aids prior disease detection, patient care and community services, Electronic Health Records store data of patient on their every visit. But when the quality of medical data is incomplete analysis accuracy is reduced. Also, the diseases and outbreaks depend upon demographics, cultural, life style and other different characteristics which differ from region to region. In this paper we used machine learning algorithms for effective prediction of diseases in disease frequent countries and a notification mechanism to eliminate possibility of wrong diagnosis which can also make the Electronic Health Record data free from misleading diagnosis. We experimented on regional chronic disease cerebral infarction. We put forward new convolutional neural network based multimodal disease risk prediction (CNN-MDRP) algorithm using structured and unstructured patient medical data retrieved from hospital after following a notification mechanism layer and the prediction accuracy of our proposed algorithm reaches 95% with a convergence speed which is faster than that of CNN-based unimodal disease risk prediction (CNN-UDRP) algorithm and it also stacks the percentage of unpredictable condition patients differing from the good condition and bad condition patients.

.

Keywords: *Big data analytics (BDA), Healthcare, Machine Learning (ML), Notification Mechanism*

1. Introduction

A well-documented medical data serves as road map for patient treatment and diagnosis so many technologies are emerging to store the medical data of patients Electronic Health Records (EHR) are one of those technologies being used to document patient medical history, lab reports, medication and other details and technologies like Machine learning, Health chatbots, AI induced applications are making their way in patient diagnosis slowly. ML facilitates quick analysis of data and delivers results, using historical and real-time data. With the help of ML algorithms, healthcare service providers can fabricate better decisions on patient's diagnoses and treatment options faster and at good accuracy rate, leading to the betterment of healthcare services and a better diagnosis. Earlier, carrying out tasks of collection and analysis of the huge volume of data

[Lavanya, 5(6): June 2018]
DOI- 10.5281/zenodo.1262195

ISSN 2348 – 8034
Impact Factor- 5.070

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

AUTOMATIC TESTING OF MODERN WEB APPLICATIONS

K. Lavanya*[1] & Dr V. Shankar[2]

*[1]M.Tech Student, Department of CSE, Kakatiya Institution of Technology and Science, Warangal District, Telangana, India

[2]Assoc.Professor, Department of CSE, Kakatiya Institution of Technology and Science, Warangal District, Telangana, India

ABSTRACT

AJAX-based internet a pair of applications considers state full asynchronous client/server communication, and client-side run-time manipulation of the DOM tree. This not solely makes them basically totally different from ancient internet applications, however conjointly a lot of erring and tougher to check. We tend to suggest a technique for testing Ajax applications mechanically, supported a crawler to infer a state-flow graph for all (client-side) program states. We tend to determine AJAX-specific faults which will occur in such states (related to e.g., DOM validity, error messages, discoverability, and back-button compatibility) additionally as DOM-tree invariants which will function oracles to observe such faults. Our method, raised to as ATUSA, is required in an exceptionally tool giving generic invariant examination elements, a plugin-mechanism to feature application-specific state validators, and generation of a check suite covering the methods obtained throughout crawl. we tend to describe 3 case studies, consisting of six subjects, evaluating the kind of invariants which will be obtained for Ajax applications additionally because the fault revealing capabilities, quantifiability, needed manual effort, and level of automation of our testing approach.

Keywords*: automated testing, web applications, Ajax*.

21. INTRODUCTION

Web applications are becoming increasingly important in recent years as people more and more tend to interact with Internet to carry out their everyday activities, e.g. browsing, shopping, gaming, working and socializing. Compared to traditional desktop applications, the bar of developing and deploying web applications is relatively lower. To some extent, anyone could rent some space from an Internet Service Provider and post his/her own web applications. On one hand, this substantially facilitates the prosperity of Internet and web applications. On the other hand, it welcomes many web applications that are not sufficiently validated or tested. These applications may cause all sorts of problems, with some of them having financial impact. Web application testing is hence a very important challenge for the health of web application engineering. Despite its importance, testing web applications poses many new challenges. First, web applications are often highly dynamic. A server-side script could generate a huge no. of client-side pages that may seem different and carry out different functionalities, depending on the client-side user inputs and the internal states on the server-side. AJAX (Asynchronous JavaScript and XML), a popular client-side technique, allows on-the-fly page updates by interacting with server scripts. Such page updates do not need loading a fresh page, but rather self-modify part of the current page (e.g. resizing a frame, repositioning a button, replacing part of the page with an image). Such updates also change both the interface and the functionality of the current client page. These active features make web application testing hard. Recently, researchers have developed various web application testing techniques that aim to address the challenges. Particularly, S. Artzi et al. proposed a test generation technique for dynamic web applications by combining concrete and symbolic executions [8]. Their tool Apollo can automatically generate test inputs to expose faults in web applications, guided by similarity of code artifacts. G. Wassermann et al. developed a test generation technique to handle dynamic features in web applications
[29]. Their technique also controls console execution to maximize test coverage of certain code artefacts in server scripts, such as string values, objects, and arrays. While the aforementioned techniques have substantially advanced

Volume 8, Issue 2 (March 2019) ISSN 2319 – 1953

International Journal of Scientific Research in Computer Science Applications and Management Studies

A Retrieval Technique for Software Reusable Components using Static and Dynamic Keyword Ontology

Swathy Vodithala[1]
Dept. of CSE, KITS, Warangal, Telangana, India
swathyvodithala@gmail.com

***Abstract*- Software engineering is a domain that has a scope for emerging new technical research work with integration of existing methods. In software engineering, Component Based Software Development (CBSD) is one of the methodologies for developing the software. The main objective of CBSD is Software Reuse. In particular, proposed work focuses on code reuse means using the existing components rather than doing the code from scratch. This paper explains the retrieval of software components using keyword search along with the ontology. The proposed work explains two different ways of using the keyword ontology, one is static keyword ontology and the other is dynamic keyword ontology.**

***Keywords* – CBSD, Software reuse, ontology, keyword, retrieval algorithm, static, dynamic, semantics**.

I. INTRODUCTION

The software engineering is an application of engineering to the software development. Component based software engineering (CBSE) or CBSD is a specialized process methodology used to develop software in software engineering. Software reuse is a major objective of CBSE, which decreases the cost and time to develop the software. A software component is a software element that conforms to a component model and can be independently deployed and composed without modification according to a composition standard. Software element contains a sequence of abstract program statements that describe the computation to be performed by a machine [1] [7]. The component used in software reuse may be code, design pattern or even documentation. The reuse of the code i.e., code reuse is gaining major attention according to the survey when compared with design or documentation [8] [9][11].

The four major areas in software reuse that should be considered are [12]

1.Identifing the components required.

2. Describing the components.

3. Organizing the components

4. Retrieving the appropriate components.

Finding the component from the software repository includes a variety of search techniques and retrieval techniques. It also includes querying [10] and indexing of the components in software repository.

The structure of the paper is organized as follows: The next section describes the related work, which has two techniques keyword search and ontology with its own advantages and disadvantages. The section III, explains the proposed algorithm based on static and dynamic keyword ontology. In section IV, we have the experimentation results and later sections have conclusion and future scope.

II. RELATED WORK

In order to attain efficiency in retrieving the software components, the software repository is to be well defined, structured and organized. Even though there are many retrieval techniques explained in the literature, only some of the techniques are explained in this section which are relevant to the proposed work.

A. keyword based search

The keyword search works with a principle of assigning some keywords which are relevant to the software component stored in the repository. The Table-1 shows an example of components along with the keywords stored in the database.
Example:

TABLE 1: KEYWORDS OF COMPONENTS

S.No	Name of the component	Keywords
1	Component 1	Kw1 ,Kw2,KW3
2	Component 2	Kw1 ,Kw2,KW4
3	Component 3	Kw3 , Kw4,KW5
4	Component 4	Kw6 ,Kw7,KW8

If the user wants to search a component based on the keyword search the query nothing but the keyword is to be given as the input say Kw3, based on the Keyword given by the user the components will be matched accordingly and retrieved say component1 and component3 are retrieved for Kw3. The disadvantage of keyword search is that the resultant may have many irrelevant components. [4][5]

INTERNATIONAL JOURNAL OF RESEARCH IN ELECTRONICS AND COMPUTER ENGINEERING

A Discretization Technique Based on Behavior for Normalizing Diversified Code Reuse Repositories

Swathy Vodithala
Assistant Professor, Dept. of CSE, KITS, Warangal
Suresh Pabboju
Professor, Dept. of IT, CBIT, Hyderabad

ABSTRACT: The major motivation of Component Based Software Development (CBSD) or Component Based Software Engineering (CBSE) is Software reuse. The importance of software reuse has been magnified as mostly the component reused refers to source code rather than documentation, tools and design patterns. The component in the proposed work is source code taken from the reuse repository. The proposed work explains a technique that can extract the behavior of software components taken from diversified code reuse repositories. The behavior of components taken from different code reuse repositories is normalized to a single dataset which can further be used by any retrieval algorithm. Apart from the behavior other relevant facets can be identified to describe the software component.

KEYWORDS: *Behavior; Component Description; Discretization; Diversified dataset; Code reuse.*

1. INTRODUCTION

Software engineering is the application of engineering to software. Component-Based Software Development (CBSD) or Component Based Software Engineering (CBSE) focuses on the development of applications based on existing software components rather than doing it from scratch.[2][3][4]. CBSE is having a higher level of importance as it is the key technology followed by many industries and resulting in high-quality software systems that are developed on time. The main aim of CBSE is to minimize the cost and time, consequently gives profitable results. The components are the pieces of code written by different programmers belong to different companies who follow different standards. The internal assumptions of code reuse repository differ because of different standards followed by the companies, so the substitutability and compatibility of software components plays a vital challenge in CBSD. The adaptability of a component is to be verified, because in reuse a component must successfully replace another in a particular application.

Re-use may be on design pattern, program elements or tools. But the widely reused software component is source code. There are three major areas in software engineering which has to be focused when considering the components for software reuse [5]. These are described as

a) Classifying/Clustering the components needed.

b) Describing the components.

c) Finding the appropriate component.

The structure of the paper is organized as follows. The section II, describes the related work which explains the existing approaches that forms the basis for proposed work. The section III explains the proposed algorithm along with the architecture .In the section IV, we have the experimentation results. The paper is concluded by conclusion, future scope and references.

2. RELATED WORK

There are many ways by which the behavior of a software component can be explained. The word specification can also be interchangeably used to the term behaviour [1]. The behavior of software is normally explained by one of the techniques given below:

- Informal specifications
  - a) comments embedded in code
  - b) informal metaphors
- Formal specifications
  - a) formal mathematics.
    - i) algebraic specifications
    - ii) model based specifications
  - b) predicate calculus

2.1. Specifications of components using Larch

There are many ideas proposed to retrieve a behavior of a software component. The specifications of software components have been compared by Zamarski both for functions (e.g., C routines, Ada procedures, ML functions) and modules (roughly speaking, sets of functions) written in some programming language [8]. These components might typically be stored in a program library, shared directory of files, or software repository. Associated with each component is a signature and a specification of its behavior.

Whereas signatures describe a component's type information (which is usually statically-checkable),

FEATURE VECTOR BASED COMPONENT CLUSTERING FOR SOFTWARE REUSE

Chintakindi Srinivas
Dept. of Computer Science & Egg
Kakatiya Institute of Technology & Science, Warangal, INDIA
csrinivas067@gmail.com

C.V.Guru Rao
Dept. of Computer Science & Egg
S.R.Engineering College
Warangal, India
guru_cv_rao@hotmail.com

V.Radhakrishna
Information Technology Dept
VNR VJIET
HYDERABAD, INDIA
vrkrishna2014@gmail.com

ABSTRACT
Software reuse is concerned about the possibility of reusability of software components. It is important to think about ways, methods and approaches for extracting knowledge from software components. Knowledge mining from software components has various practical applications that range from software reuse to financial applications. This paper proposes a similarity measure for similarity computation between software components by extending our previous research. The proposed measure is used to perform component clustering. Software component clustering facilitates software reusability and software segmentation.

Categories and Subject Descriptors
D.2 **[Software Engineering]; D.2.13 [Reusable Software]; H.3 [Information Storage and Retrieval]; H.3.3 [Information Search and Retrieval]:** Clustering

ICEMIS'18, June 19–21, 2018, Istanbul, Turkey.

ACM 978-1-4503-6392-1/xx/xxxx …$15.00.
DOI: https://doi.org/10.1145/3234698.3234737

Keywords
Software Clustering, Feature Vector, Components, Classification

General Terms:
Software Components, Reuse, Similarity

1. INTRODUCTION
Software component reuse is an interesting area of research that has received wide attention from academia. Unsupervised learning of software components further received significant importance from researchers. R. Ibba et.al [35] discusses clustering software components by considering component structure. In terms of unsupervised learning of software components, we can view a similarity function being reused for computing similarity between software components. Concept analysis-based method is proposed by P. Tonella et.al [36] for restructuring modules of software components. We can view concept as objects with shared attributes or features that are grouped. Grouping of objects in general requires commonly shared attributes and features. This means structure of components or software objects are functions of attributes. If chosen set of attributes best represents the component (or object) then, the grouping achieved shall be cohesive in nature.

This property is expected when we perform clustering of software components. Detailed information on approach for module identification is discussed [36]. Y. Chiricota et.al [37] applies the quality metric (MQ) for achieving software component clusters. Clustering performed is graph based clustering. The approach [37] depends on identification of strong and weak edges.

Fuzzy approach for component clustering [38] is proposed by applying principles of fuzzy relations. Disease prediction and analysis requires analysis of respective medical dataset. [39] Applies principal of software reuse for medical domain. Statistical techniques are applied for dataset analysis. For evaluating the statistical model developed [39] applies Fenton and Melton metric. System usability is one of the attributes that is related to both application design and also user interface. The principle of software reuse [40] to web-based applications is addressed.

A novel approach for identifying software components, CCIC is proposed. Given a set of analysis classes and set of software architect preferences, CCIC [41] identifies logical software components and applies a customized hierarchical evolutionary algorithm. CCIC facilitates to decide constraints or rules w.r.t implementation and deployment structure. [42] Carried a recent survey on software reuse and its implications. Marlon Monçores [43] applies large neighbourhood search to clustering software modules.

RESEARCH REVIEW JOURNALS
Volume-04
Issue-06
June-2019

ISSN: 2455-3085 (Online)
RESEARCH REVIEW International Journal of Multidisciplinary
www.rrjournals.com[UGC Listed Journal]

Big Data Analytics Transformation and Tools for Big Data Mining and Analysis

[1]Akinapalli Sanjeev Raju, [2]Ravi Kumar Akinapally & [3]Chintakindi Srinivas

[1]*Senior Assistant Professor, Department of CSE, Kamala Institute of Technology & Science (India)*
[2]*Assistant Professor, Department of IT, Sreenidhi Institute of Science & Technology (India)*
[3]*Associate Professor, Department of CSE, Kakatiya Institute of Technology & Science (India)*

ARTICLE DETAILS

Article History
Published Online: 12 June 2019

Keywords
Big data, Big data analytics, Hadoop.

***Corresponding Author**
Email: asanjeevaraju@gmail.com

ABSTRACT

In this paper we will certainly take a look at future patterns of in development excavating that are made use of for evaluation and also the assumption of big data. Big data analytics describes the technique of assessing massive quantities of data, or big data. The big data is gathered from a big array of resources, such as social media networks, video clips, electronic photos, as well as sensing units. The significant purpose of Big Data Analytics is to uncover brand-new patterns and also partnerships which may be unseen, and also it can supply brand-new understandings concerning the individuals that developed it. There are a variety of devices readily available for mining of Big Data and also Analysis of Big Data, both specialist as well as non-professional.

1. Introduction

Big data suggests the datasets which cannot be identified, acquired, took care of, assessed, as well as refined by existing devices. Various meanings of big data have actually been offered by various customers of Big Data as well as various experts of Big Data like research study scholars, data experts, as well as technical experts. According to Apache Hadoop "Big data is a dataset which might not be caught, handled, as well as refined by basic computer systems within an appropriate extent.

In nowadays the amount of data produced per secondly is huge. Data stream live analysis is called for to handle this big data, with the correct analysis we can obtain critical data, via this, we can forecast network website traffic, invasion relevant task, climate. log documents or click-streams in internet checking out, making procedures, call information documents, e-mail, blog writing, twitter articles as well as others [1] data created from the stream simply picture of stream data. Photo is based upon time period. Stream data mining formula handling based time and also room restraint. The main of formulas is the use of sources (sources can be Memory and also time). In-stream data source, to carry out stream mining we need to think about, precision, the quantity of room, the time needed to find out from training instances for obtaining forecast. Data is big and also expanding, there are necessary patterns as well as patterns in the data. We do not totally recognize where to look or exactly how to locate them. Big data analysis is crucial since the data is continually transforming based upon the period of time to keep big data a lot of firms are making use of cloud arrangement.

Really big data was specified in 2001 for the very first time. Doug Laney, specified the 3Vs version, i.e., Quantity, Range as well as Rate 2. Even with the reality that the 3Vs version was not utilized to specify big data, Gartner as well as lots of various other companies, like IBM3 and also Microsoft4 still utilizes the "3Vs" design to specify big data5. In the "3Vs" design, Quantity implies, the dataset is so big as well as big that it is really challenging to examine; Rate implies the data gathered as well as collected so quickly to use it to the optimum; Selection reveals various sorts of data like structured, semi-structured as well as disorganized data i.e. sound, video clip, website, and also message. IDC (Worldwide data Firm), among one of the most leading leaders in the study areas of Big Data, is of various sight regarding Big Data. According to an IDC record of 2011 "Big Data modern technologies define a brand-new generation of modern technologies as well as styles, made to financially remove worth from large quantities of a variety of data, by allowing the high-velocity capture, exploration, and/or analysis" 6. According to this meaning, big data features can be Quantity (big quantity), Selection (numerous kinds as well as the framework of data), Rate (fast development), and also Worth (excellent worth yet extremely reduced resemblance).

2. What is big data analytics?

As one of the most "hyped" terms on the market today, there is no agreement regarding just how to specify big data. The term is frequently made use of synonymously with a relevant idea such as Company Knowledge (BI) and also data mining[5]. It holds true that all 3 terms have to do with examining data as well as in a lot of cases progressed analytics. Yet big data principle is various from both others when data quantities, variety of deals and also the variety of data resources are so big and also complicated that they call for unique techniques and also modern technologies in order to attract understanding out of data (as an example, standard data stockroom services might fail when handling big data).

This additionally creates the basis for the most second-hand meaning of big data, the 3V: Volume, Velocity and also Variety as displayed in Number 1.

Springer

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

S. Venkatramulu and C. V. Guru Rao

Abstract Web application servers are prone to attacks that are more vulnerable and thousands of security breaches that are taking place every day. 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 analysers. However, the limitations in terms of programming language dependency and the related issues drop the performance levels of syntax analyser 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 com-pared 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 (Associate Professor)

Department of CSE, K.I.T.S, Warangal, Telangana, India e-mail: venkatramulu10@gmail.com

C. V. Guru Rao
SR Engineering College, Department of CSE, Warangal, Telangana, India e-mail: guru_cv_rao@hotmail.com

V. Bhateja et al. (eds.), *Proceedings of the Second International Conference on Computational Intelligence and Informatics*, Advances in Intelligent Systems and Computing 712, https://doi.org/10.1007/978-981-10-8228-3_23

Periodicals of Engineering and Natural Sciences ISSN 2303-4521
Vol.6, No.2, December 2018, pp. 302~315
Available online at: *http://pen.ius.edu.ba*

A Novel Approach for Iceberg Query Evaluation on Multiple Attributes Using Set Representation

V. Chandra Shekhar Rao [1], Dr P. Sammulal [2],
[1]Research Scalar, Department of CSE, JNTU, Hyderabad, vcsrao.kitswgl@gmail.com
[2]Associate Professor, Department of CSE, JNTU, Hyderabad, sammulalporika@gmail.com

ABSTRACT

Iceberg query (IBQ) can be an really identifying kind of aggregation question that calculate aggregations up-on user given threshold (T). In data mining field, effective investigation of compounding queries was because of by the majority of investigators because the tremendous generation of information outside of industrial and businesses industries. Conclusion assist database and discovery of the majority of information connected systems largely calculate the worthiness of most fascinating features having an critical level of information from data foundations that may be tremendous. By means of the paper, we propose that an initial Manner of calculating IBQ, which builds a choice for every attribute nicely value, but additionally includes a One of a Kind events Inside the attribute column also plays specify operations for creating closing Outcomes. We formulated highly effective GUI software for just 2 characteristics, numerous traits employing egotistical prepare and several features utilizing lively plan. If data collection comprises two traits, then it truly is substantially more advanced than apply just two traits. In the event of information collection comprises multiple traits, predicated up on anyone choice suitable module could potentially be decided on. If characteristic uniqueness changes from characteristic in to the following characteristic, then vibrant variety approach is very powerful. This strategy somewhat reduces performance memory and time space contrast with additional processes.

I. Introduction

Industry awareness and Recognizing Detection [1] from Dealing databases/warehouses always hardy fire-arms, So that you may acquire competitive wages in the current industry community. An Iceberg Query(IBQ) could be an excellent Form of a aggregation question that divides values to a Individual Outlined threshold (T) meanings. It truly is of this differentiating comprehension of most those end users in bringing advanced level degree worth which frequently take more considerable in manufacturing companies. The Syntax with the Iceberg query as to a romantic relationship REL (C1, C2. . .Cn) is revealed beneath:

SELECT Ci, Cj, …, Cm, AGG(*) FROM R GROUP BY Ti,Tj…, Tm HAVING AGG (*) > = T

This aggregation performs,"at which by Ci, Cj,...,Tm" suggests a sub set pair of capacities in well-known mix faculties. Aggregation functions just like COUNT(), COUNT(*),MIN,MAX,SUM and AVG. The more expensive in comparison with equal to >= could possibly be considered a index used like a contrast predicate.

Article Info

Article history:

Received Jun 12th, 2018
Revised Aug 20th, 2018
Accepted Aug 26th, 2018

International Journal of Engineering & Technology, 7 (4.39) (2018) 997-1000

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Glimpse on Iceberg Query Evaluation Techniques

V. Chandra Shekhar Rao*[1], Dr. P. Sammulal[2]

[1]*Research Scalar, Department of CSE, JNTU, Hyderabad,*
[2]*Associate Professor, Department of CSE, JNTU, Hyderabad*
**Corresponding author E-mail: vcsrao.kitswgl@gmail.com*

Abstract

Queries are required to determine distinct quality worth's and also their accumulation that is over a predefined limit from this significant variety of documents. The information keeping as well as getting are playing a significant function in the information clustering and also information warehousing methods. The performance of an information retrieving approach relies on the information details queries for getting the information from the data source. Iceberg query is an one-of-a-kind course of gathering query, which calculates accumulation worth's over a provided limit. The queries which are having this sort of nature are called as Iceberg(IB) queries.
Existing data source system perform it similar to typical query so it take even more time to implement. It is difficult job to essence fascinating as well as crucial details promptly from big data source. Great deals of research study has actually been done to raise the rate of IB query. Initially scientists makes use of tuple check strategy to perform IB query which is time consuming and also need even more memory. To get rid of these troubles scientists recommended IB query examination utilizing Bitmap Indexing strategies. This method prevent total table check so time called for to carry out IB query is decreased as well as memory demand is likewise lowered.

Index Terms : *Bitmap Index , Aggregation functions, Iceberg Queries, Counting co-occurrence*

1. Introduction

Data mining software application evaluates connections as well as patterns in saved purchase information based on flexible individu-al queries. A number of kinds of a logical software program are readily available: analytical, artificial intelligence, and also se-mantic networks
Many data mining queries are specially iceberg queries. For ex-ample, market experts per kind of market basket query on huge information stockrooms that keep consumer sales deals. These queries find customer acquiring patterns, by discovering thing sets that are combined with lots of consumers. Target collections are items required to sustain the thing set. Given that these queries operate huge datasets, addressing such iceberg queries successful-ly is a vital issue.

The customers of the DW are the choice manufacturers, service expert as well as expertise employees company. They use infor-mation from an information storage facility to anticipate some concerns regarding their organization. As necessary they take choices concerning their company. Once they have actually taken choice they focus on execution of the very same. The info which they accumulate from an information storage facility is tiny info from a massive dataset. To eliminate such a type of details the question executed is of the nature event of some worth on some specified issue or limitation. This kind of question is called as IB inquiry.

IB questions were really initial investigated by a scientist called Minutes Fang et.al [1] According to him, an iceberg inquiry has a lot of application in data-warehousing, information mining along with information gain access to systems. Iceberg Inquiry defined as the type of question which do collecting attribute on some col-lection of particular adhered to by having a terms on some issue or on restriction well worth. As a result of having stipulation the accumulated feature which does not please limit problem will certainly obtain remove from the outcome.

2. Review of related works

A handful of looks into are readily available in literary works for iceberg queries. Below, we evaluate the current jobs offered in the literary works for assessment of iceberg queries. Scientists were constantly extremely eager to figure out the effective means to carry out the iceberg queries as a result of the restricted computer sources. There are outcomes which are revealing that implement-ing iceberg queries on information takes even more time than discovering the organization regulation from the information col-lections.

IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 20, Issue 4, Ver. II (Jul - Aug 2018), PP 61-68 www.iosrjournals.org

Hadoop Map reduce for Tactical Clouds

[#1]S. Swapna, ,[2]**B**. Srinivas,
Dept Of Cse, Kakatiya Institute Of Technology And Science, Warangal, Ts, India.
Corresponding Author: S. Swapna

***Abstract**: We envision a future where real-time computation on the battlefield provides the tactical advantage to an Army over its adversary. The ability to collect and process large amounts of data to provide actionable information to soldiers will greatly enhance their situational awareness. Our vision is based on the observation that the U.S. Military is attempting to equip soldiers with smartphones. While individual phones may not be sufficiently powerful for processing large amount of data, using the mobile devices carried by a squad or platoon of Soldiers as a single distributed computing platform, a Tactical Cloud, would enable large-scale data processing to be conducted in battlefields. In order for this vision to be realized, two issues have to be addressed. The first is the complexity of writing applications for distributed computing environments, and the second is the vulnerability of data on mobile devices. In this paper, we propose combining two existing technologies to address these issues. The first is Hadoop MapReduce, a scalable platform that provides distributed storage and computational capabilities on clusters of commodity hardware, and the second is the Mobile Distributed File System (MDFS) which allows distributed data storage with built-in reliability and security. By making the MDFS file system work with Hadoop on mobile devices, we hope to enable big data applications on tactical clouds.*

***Keywords**—Mobile cloud, Hadoop, Map-reduce.*

I. Introduction

With advances in technology, mobile devices are becoming capable computing platforms. The new generations of mobile devices are relatively powerful with gigabytes of memory and multi-core processors. These devices have sophisticated applications and sensors capable of generating and collecting hundreds of megabytes of data. This data can range from raw application data to images, audio, video, or text files. With these enhancements in mobile device capabilities, big data processing in environments such as disaster recover sites and battlefields is becoming a reality [1]. There is currently an effort by the military to equip Soldiers with smartphones [2]. We propose utilizing these mobile devices to collect and process data in order to provide Soldiers with enhanced situational awareness. Current mobile applications that perform massive computing tasks, such as big data processing, offload data and tasks to data centers or powerful servers in the cloud [3]. Hadoop MapReduce [4] is one of the frameworks that exist to make such computation easier. It splits user jobs into smaller tasks and runs them in parallel on different nodes, reducing the overall execution time. In extreme environments, access to the traditional cloud may not be available.

Thus, the ability to carry out computation across a group of mobile devices, a Tactical Cloud carried by a squad of Soldiers or a team of first responders, is essential. This requires a Hadoop-like framework that is resilient to network failures and can operate across wireless mobile ad-hoc networks [5] typical of such scenarios. A concern that has to be addressed to enable distributed computation across mobile devices is data security, due to the envisioned applications for such systems involving sensitive information [6], [7].

Traditional security mechanisms tailored for static networks are inadequate for tactical clouds (i.e., tactical grade security) due to the ease with which mobile devices can be lost or captured (and data could be compromised, even if encrypted). One approach proposed to address this security vulnerability is the k-out-of-n computing framework [8] which distributes data across n nodes with the property that the data from at least k nodes is necessary to reconstruct the original information. In this paper, we replace Hadoop‟s native distributed file system, HDFS [9], with the Mobile Distributed File System (MDFS) [8], [10] that uses the k-out-of-n principle in order to provide the security necessary for the application domain.

In addition to the lack of tactical-grade security, a main drawback of HDFS in mobile environments is its inefficient use of resources. HDFS does not consider device energy and relies on low latency and high availability networks to replicate file blocks across multiple devices to increase reliability. Interestingly, the aforementioned k-out-of-n-enabled MDFS [8], [10] also ensures high energy efficiency. Replacing HDFS with MDFS mitigates these drawbacks while allowing Hadoop MapReduce to be used as a framework for distributed computing on mobile devices

ISSN XXXX XXXX © 2019 IJESC

Research Article Volume 9 Issue No.2

Predicting Diseases-Emergency Department Visits using Big Data with Privacy Preservation

B.Srinivas[1], Yerram Sneha[2]
Associate Professor[1], M.Tech Scalar[2]
Department of Computer Science and Engineering
Kakatiya Institute of Technology & Science, Warangal, Telangana, India

Abstract:
The idea that the purely knowledge that we can extract by analyzing large amounts of data can be useful in healthcare. This paper provides an overview of recent developments in big data in the context of biomedical and health informatics. Diseases are the most important and Costly, which can/cannot be cured. However, accurate and timely surveillance data could allow for timely and targeted interventions at the community or individual level. We introduce a novel method of using multiple data sources for predicting the number of diseases-related Information visits in specific Purposes. Twitter data, Google search interests, environmental sensor data and Lab Diagnostics centre's were collected for this purpose. Our preliminary findings show that our model can predict the number of Diseases Related ED visits based on near-real-time environmental and social media data with approximately 70% pre- cision. The results can be helpful for public health surveillance, ED preparedness, and targeted patient interventions.

Keywords: Big data, Emergency Department (ED) Visits, Environmental Sensors, Predictive Modelling, Social Media.

1. INTRODUCTION

A Top-k Dominating Query binds the benefits of both top –k queries and Skyline queries. In a top –k query a user provides a ranking function to order the objects by their scores and thus retrieve top-k best objects, while A Skyline is defined as those points which are not dominated by any other point. A point dominates another point if it is as good or better in all dimensions and better in at least one dimension. In Top- k output size can be controlled by using parameter –k, and In skyline query you need not use any ranking function.
Let us consider a Dataset D let $A(d_1,d_2,d_3….d_n)$,$B(d_1,d_2,d_3….d_n)$ be the entities in a dataset D, where $d_1,d_2,d_3….d_n$ indicates dimensions of an entities.

Entity A is set to be dominate object B when it is not bad in all dimensions or at least better one dimension while compared to B.Incomplete data indicates not having appropriate or necessary information. Let us take an entity$X(d_1,d_2,_,…,_,.d_n)$ where some of the dimensions are missing, so this is an incomplete object, where "_" indicates a missing dimension , a sample object X(3,_,2) which has three dimensions where one of it's the dimension is missing.

Moreover, we apply dominance relationship only when object and comparing entities have common dimensions. For instance take two entities. A(1,_,4) and B(_,_,3) having three dimensions this can be comparable. But in case if A(1,_4) and B(_,4,_) ,there are no common dimension hence they cannot be comparable i.e; we cannot apply dominance relationship. In a Real-time system, whenever people visit e-Commerce website, people only tend to rate a product which he/she buys, in this case missing data occurs as whenever a new user visits a e-commerce website he is enable to decide which product is good. So Tkd queries helps in finding a solution to the given problem by making use of Tkd algorithms. Let us consider an example of e-commerce website. where there are few products and ratings are given by website users. The below Table shows the ratings from various users for various products.
Product p1 is better than or dominates other products i.e; for P_1
$P_1.[1]=5 > P_2.[1]=3$ and $P_1.[1]=5 > P_3.[1]=2$, $P_1.[2]=4 > P_3.[2]=3$, $P_1.[4]=4 > P_2.[4]=3$ and for product P_2
$P_2.[1]=3 > P_3.[1]=2$, P_2 dominates product P_3 .Score is calculated based on the number of objects it dominates product P_1 score[P_1]=2 as it dominates two products, product P_2 score[P_2]=1 as it dominates one product P_2 .A Tkd query ,given K=1 it returns product P_1 as best product.

[Varshini, 5(6): June 2018]
DOI- 10.5281/zenodo.1262287

ISSN 2348 – 8034
Impact Factor- 5.070

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

EFFICIENT DATA STORAGE MECHANISM IN CLOUD COMPUTING

Vasam Varshini*[1] & B.Sridhara Murthy[2]

*[1]M.Tech Student, Department of CSE, Kakatiya Institution Of Technology and Science, Warangal District, Telangana, India

[2]Assistant Professor, Department of CSE, Kakatiya Institution Of Technology and Science, Warangal District, Telangana, India

ABSTRACT

As data gradually grows within data storage areas, the cloud storage systems nonstop face challenges in saving storage capacity and providing capabilities necessary to move big data within an acceptable time frame. In this paper, we propose the Boafft, a cloud storage system with distributed deduplication. The Boafft achieves scalable throughput and capacity using multiple data servers to de- duplicate data in parallel, with a minimal loss of deduplication ratio. Initially, the Boafft uses an effective data routing procedure based on data similarity that decrease the network burdenby quickly identifying the storage location. Secondly, the Boafft maintains an in-memory similarity indexing in each data server that helps avoid a large number of random disk reads and writes, which in turn accelerates local data deduplication. Thirdly, the Boafft build hot fingerprint cache in each data server based on access occurrence, so as to improve the data deduplication ratio. Our comparative analysis with EMC's stateful routing algorithm reveals that the Boafft can provide a comparatively high deduplication ratio with a low network bandwidth overhead. Moreover, the Boafft makes best usage of the storage area, with higher read/write bandwidth and good load balance.

Keywords: *Big data, cloud storage, data deduplication, data routing, file system.*

17. INTRODUCTION

As data is at a riotous growth, the redundancy in data is also escalating and with this comes apre- condition for anorderly method to classify and organize such a huge amount of data. Big data is an embryonic term for datasets that are extremely huge which are not capable to manage by conventional techniques. Since data is being outsourced to cloud storage, the effectual management of storage space asks for more attention. Deduplication [1]turns up to be asuitableway out for data detonation in big data epoch by decelerating the data expansion speed by wiping out the redundant data. The olddeduplication methods work on primary storage only. The management of such massive data turns out to be very intricate. Data deduplication is a lossless compression technique that averts the replicated data from being stored into the storage devices. Hasty increase in data is momentous challenge to be conquered. Data deduplication fundamentallyorientationsthe data previouslystored on the disk by a phenomenon that supersedes the indistinguishable data in a file or identical regions of file (similar data).

RESEARCH REVIEW JOURNALS
Volume-04
Issue-01
January -2019

ISSN: 2455-3085 (Online)
RESEARCH REVIEW International Journal of Multidisciplinary
www.rrjournals.com [UGC Listed Journal]

An Overview of IOT Healthcare and Applications

Bejugama Sridhara Murthy

Assistant Professor, Department of CSE, Kakatiya Institute of Technology & Science (India)

ARTICLE DETAILS

Article History
Published Online: 10 January 2019

Keywords
health care, Internet of things, security

ABSTRACT

Even the IoT has a number of application domains, for example healthcare. The IoT revolution is currently organised contemporary healthcare together with promising technological, economical, and social prospects. This paper studies advances in IoT-based healthcare technology and testimonials the advanced network architectures/platforms, programs, and industrial tendencies in IoT-based healthcare solutions. Additionally, this paper assesses distinct IoT security and privacy features, such as security conditions, hazard models, and assault taxonomies in the healthcare perspective. Further, that this newspaper polls smart collaborative security version to minimise security threat. Rather than the traditional Web, along with people, an IoT joins a High Number of machines cellular, tabletcomputer, computer, resource- restricted devices and detectors employing heterogeneous wired and wireless networks.

1. Introduction

Definition of IoT and Its Role in the Healthcare Industry

According to[1], the Internet of Things will be a international network infrastructure, connecting virtual and physical objects throughout the manipulation of information capture and communication capacities. This infrastructure comprises existing and between Web and network improvements. It will provide specific thing - identification, detector and link capacity as the foundation for the evolution of independent joint applications and services. These can be characterized with a high level of autonomous information capture, event transport, network connectivity and interoperability. "

These days, Internet of Things (IoT) joins the Web with detectors and a great number of devices, largely utilizing IP-based communications. In healthcare business, IoT offers alternatives to remote observation, early prevention, and clinical therapy for institutionalized handicapped. Such tags and devices ease access by individuals' health professionals. By way of instance, RFIDs labels of individuals or patients' private devices (like medical devices) are more readable, familiar, locatable, and dialing through IoT programs [2].

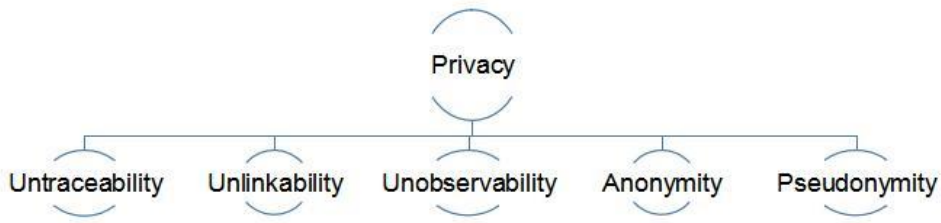
IoT enables a vast selection of smart services and applications to deal with challenges which people or healthcare industry faces [3]. This intelligently connects people, machines, intelligent devices, and energetic methods so as to guarantee an effective healthcare system [4].

2. Definitions of Privacy and Security

Definition of Privacy. Ensuring privacy necessitates making sure that individuals keep the right to control exactly what information is collected about them, who keeps it, who uses it, how it's utilized, and what point it's used for.

Figure 1 shows the main privacy services1 and properties as described by [5]:

1) *Untraceability*: Making it difficult for an adversary to identify that the same subject performed a given set ofactions.
2) *Unlinkability*: Hiding information about the relationship between any items, such as subjects, messages, actions, etc.
3) *Unobservability*: Hiding the fact that a message was sent (as opposed to hiding the identity of the sender of message).
4) *Anonymity*: Hiding information who performed a given action or who is described by a given dataset.
5) *Pseudonymise*: Using pseudonyms instead of using real identifiers.



DEGSA-VMM: Dragonfly-based exponential gravitational search algorithm to VMM strategy for load balancing in cloud computing

Vijaykumar Polepally, K. Shahu Chatrapati ,Kybernetes

ISSN: 0368-492X

Publication date: 4 June 2018 Reprints & Permissions

Abstract:

Purpose

This paper aims to develop the Dragonfly-based exponential gravitational search algorithm to VMM strategy for effective load balancing in cloud computing. Due to widespread growth of cloud users, load balancing is the essential criterion to deal with the overload and underload problems of the physical servers. DEGSA-VMM is introduced, which calculates the optimized position to perform the virtual machine migration (VMM).

Design/methodology/approach

This paper presents an algorithm Dragonfly-based exponential gravitational search algorithm (DEGSA) that is based on the VMM strategy to migrate the virtual machines of the overloaded physical machine to the other physical machine keeping in mind the energy, migration cost, load and quality of service (QoS) constraints. For effective migration, a fitness function is provided, which selects the best fit that possess minimum energy, cost, load and maximum QoS contributing toward the maximum energy utilization.

Findings

For the performance analysis, the experimentation is performed with three setups, with Setup 1 composed of three physical machines with 12 virtual machines, Setup 2 composed of five physical machines and 19 virtual machines and Setup 3 composed of ten physical machines and 28 virtual machines. The performance parameters, namely, QoS, migration cost, load and energy, of the proposed work are compared over the other existing works. The proposed algorithm obtained maximum resource utilization with a good QoS at a rate of 0.19, and minimal migration cost at a rate of 0.015, and minimal energy at a rate of 0.26 with a minimal load at a rate of 0.1551, whereas with the existing methods like ant colony optimization (ACO) gravitational search algorithm (GSA) and exponential gravitational search algorithm, the values of QoS, load, migration cost and energy are 0.16, 0.1863, 0.023 and 0.29; 0.16, 0.1863, 0.023 and 0.28 and 0.18, 0.1657, 0.016 and 0.27, respectively.

Originality/value

This paper presents an algorithm named DEGSA based on VMM strategy to determine the optimum position to perform the VMM to achieve a better load balancing.

ISSN 0970-938X
www.biomedres.info

Optimized maximum principal curvatures-based segmentation of blood vessels from retinal images.

Santosh Kumar NC*, Radhika Y

Department of CSE, Githam University, Vizag, Andhra Pradesh, India

Abstract

In retinal image of the human eye, extracting tree shaped retinal vasculature is an important feature which helps eye care specialists or ophthalmologists to pursue proper diagnostic procedures. In this paper, an approach named Optimized Maximum Principal Curvatures Based (OPCB) segmentation is been proposed for efficient extraction of blood vessels from retinal fundus images. This algorithm proceeds into two stages. Firstly, pre-processing on input retinal images is done by Particle Swarm Optimization (PSO) technique which is an automatic process for computing the global optimum pixels of the image in order to avoid working with all or random pixels. Later, these optimal pixels are made to undergo further processing with Gaussian Filter to remove the noisy pixels among them. Secondly, the post-processing is carried out in four steps: (i)Maximum Principal Curvatures (maximum-eigenvalues) of the second order derivative matrix (Hessian) quantity of the pre-processed PSO image are computed by using 'Lambda Function', which then does region growing of the tree-shaped blood vessels by convolving Maximum Principal Curvatures with the mathematical erosion structuring element of. (ii)After extraction of blood vessels, section-wise contrast enhancement is performed by using Adaptive Histogram Equalization that work on 8 × 8 tiles of image being segmented for smoothing artificially introduced boundaries if any, and also for eliminating over amplified noise. (iii) ISODATA (Iterative Self-Organizing Data Analysis Technique) thresholding is used to classify the image globally where the image's foreground vascular structure is segmented from the background. (iv) A 'morphologically opened' operation is performed to prune falsely segmented isolated regions, to achieve very accurate segmentation. This proposed technique tested online available colored retinal images of STARE and DRIVE databases. As an outcome, the proposed approach achieves the superior segmentation accuracy of 96% which outperformed many empirically proven segmentation methods which were proposed in the past.

Keywords: Retinal fundus image, Gaussian filter, Particle swarm optimization, Maximum principal curvature, Adaptive histogram equalization, ISO data thresholding, STARE and DRIVE databases.

Introduction

Frightening disease diabetes is affecting and deteriorating health of humans around the globe. Regardless of speedy attainments in medical field, this disease surely will get amplified in the future. According to Verma et al. [1], many patients losing their eye sight due to spread of diabetes on retina, which is well known as 'diabetic retinopathy' and this, is because of disparities in retina vascular arrangement. "Prevention is better than cure", it is better to discover and manage the irregularities in the retina vessels that can reduce suffering of patients from vision failure. Currently figure of patients affected with diabetes is terrifying; research is proving that above 85% cases of the patients with complications of diabetic can be condensed, if the diagnosis for diabetic retinopathy is done early. The blood vessels segmentation outcome is a verdict of cardiovascular and ophthalmologic disorders including diabetic retinopathy, glaucoma, hypertension etc. Computerization of analysis of retinal image gives greater insights to ophthalmologists who can make an efficient assessment of the state of the patients by managing, analysing, and storing retinal images. In mechanizing the discovery of vessels -vasculature for a better analysis, the substance that is at the background of the image is an important measure and fetching of that background is a difficult job since the vessels are indifferent in terms of diameter and width [2]. These measuring parameters of the varied blood vessels give an elegant specification of ophthalmologic disease type and its intensity. The segmentation of vessels which can be either supervised or unsupervised in nature is a precursor in the development of automated diagnosis system [3] that concludes an ophthalmic disease.

A Similarity Function for Feature Pattern Clustering and High Dimensional Text Document Classification

Abstract:

Text document classification and clustering is an important learning task which fits to both data mining and machine learning areas. The learning task throws several challenges when it is required to process high dimensional text documents. Word distribution in text documents plays a very key role in learning process. Research related to high dimensional text document classification and clustering is usually limited to application of traditional distance functions and most of the research contributions in the existing literature did not consider the word distribution in documents. In this research, we propose a novel similarity function for feature pattern clustering and high dimensional text classification. The similarity function proposed is used to carry supervised learning-based dimensionality reduction. The important feature of this work is that the word distribution before and after dimensionality reduction is the same. Experiment results prove the proposed approach achieves dimensionality reduction, retains the word distribution and obtained better classification accuracies compared to other measures.

An Efficient Approach for Dimensionality Reduction and Classification of High Dimensional Text Documents

Abstract:

Feature representation and dimensionality reduction techniques are two important tasks in text clustering and classification. In this paper, an approach for feature representation and dimensionality reduction of text documents is described. The feature representation and dimensionality reduction approaches that are introduced retain the original distribution of features. Output of feature representation is a hard representation matrix. The hard matrix is used to obtain the low dimensionality document matrix. The input for clustering is the low dimensional matrix. The working of proposed approach is explained using a case study that supports the importance of the approach and advantage of dimensionality reduction. Results prove that the proposed approach has better dimensionality reduction achieved and is also better suited for classification.

K. Vinay Kumar

Assistant Professor, Department of CSE, Kakatiya Institute of Technology & Science (India)

© 2019 JETIR March 2019, Volume 6, Issue 3 www.jetir.org
(ISSN-2349-5162)

A Comprehensive Study on Machine Learning

Pranavi Singarapu[1], Gouthami Velakanti[2]
[1,2]Assistant Professor, Department of CSE, Kakatiya Institute of Technology and Science, India

Abstract

Over the previous 20 years, Machine Learning has actually turned into one of the primaries- keeps of infotech and also keeping that, an instead main, albeit normally concealed, component of our life. With the ever before raising quantities of information appearing there is a great factor to think that wise information evaluation will certainly end up being a lot more prevalent as a required component for technical progression. The function of this paper is to supply the viewers with a review over the large series of applications which contend their heart a machine learning issue as well as to bring some level of order to the zoo of troubles.

Index Terms : Machine Learning, Applications, Regression

I. Introduction to Machine Learning

Machine learning can show up in numerous semblances. We currently review a variety of applications, the kinds of information they take care of, as well as lastly, we define the issues in a rather even more elegant style. The last is crucial if we intend to stay clear of transforming the wheel for each brand-new application. Rather, a lot of the art of machine learning is to lower a series of rather inconsonant issues to a collection of relatively slim models. Much of the scientific research of machine learning is after that to fix those troubles as well as give excellent assurances for the remedies.

Understanding, like knowledge, covers such a wide variety of procedures that it is difficult to specify specifically. A thesaurus meaning consists of expressions such as "to get understanding, or understanding of, or ability in, by the research study, direction, or experience,"[1] as well as "adjustment of a behavior propensity by experience." Zoologists as well as psychotherapists research understanding in pets as well as human beings. In this publication, we focus on finding out in devices. There are numerous parallels between pet and also machine learning. Definitely, several methods in machine learning stem from the initiatives of psychotherapists to make even more specific their concepts of the pet as well as human knowing with computational designs. It promises likewise that the principles and also methods being discovered by scientists in machine learning might brighten specific elements of organic discovering.

As concerns devices, we could claim, extremely generally, that a device discovers whenever it alters its framework, program, or information (based upon its inputs or inaction to exterior info) in such a fashion that its predicted future efficiency enhances. Several of these modifications, such as the enhancement of a document to an information base, autumn conveniently within the district of various other self-control as well as are not always far better recognized for being called discovering [2]. However, as an example, when the efficiency of a speech-recognition maker enhances after listening to a number of examples of an individual's speech, we really feel fairly warranted because instance to claim that the maker has actually discovered.

II. Applications

A lot of visitors will certainly know with the idea of website position. That is the procedure of sending a question to an online search engine, which after that locates websites pertinent to the inquiry as well as which returns them in their order of importance. See e.g. Figure 1 for an instance of the inquiry results for "machine discovering". That is, the internet search engine returns an arranged checklist of websites provided an inquiry To accomplish this objective, an internet search engine requires to 'understand' which web pages matter and also which web pages match the question.

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

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

A Study on Applications of IoT

Abstract:

The government has adopted the field of "Internet of Things" as a national strategic project, announcing the Internet of Things master plan to achieve a leading country of hyper-connected digital revolution last May. The government has the promotional strategy of reinforcing the competitiveness in software (SW) sensor components devices, training specialists that will lead Internet of Things (IoT) services and products, and internalizing security for Internet of Things products and services. Internet of Things, thus expected to grow from W2 trillion and 300 billion last year to W30 trillion by 2020 in the market, refers to the Internet environment where people, things around, data, etc. are all connected to the wired and wireless network to mutually create, collect, share, and utilize information. Internet of Things is the technology and service that includes generating information (sensor) - acquisition (parts, devices) - sharing (Clyde) - utilization (Big Data) - application software. Internet of Things is getting a lot of attention from the public due to the effect of increasing export businesses and jobs and has become a buzzword among businesses. The Internet of Things market has a tendency to grow in the future as it is extended from social infrastructure (utilities, transportation, automation, etc.) and safety management to the consumer sector centered on life services. Sophisticated wireless communication technology is expected to form a huge network connected to all object units as a communication function.

Siripuri Kiran

Assistant Professor, Department of CSE, Kakatiya Institute of Technology & Science (India)

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue II, February 2018- Available at www.ijraset.com

Anomaly Detection Using Data Mining Techniques in Social Networking

Ajmera Rajesh[1], Siripuri Kiran[2]
[1]*Academic Consultant, Ku College Of Engineering And Technology, Kakatiya University, Warangal. India.*
[2]*Assistant Professor, Kakatiya Institute Of Technology and Sciences, Warangal, India.*

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

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.

A Case Study on Issues in Privacy Preserving Data Mining

Siripuri Kiran[1], Ajmera Rajesh[2]
Assistant Professor[1], Academic Consultant[2 1, 2] Kakatiya Institute Of Technology and Sciences, Warangal, India. [2]Ku College Of Engineering And Technology, Kakatiya University, Warangal. India.

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:

1) Association rule mining or market basket analysis Page

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.

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. Data mining task requires utilities fir statistical data and Artificial Intelligence systems (AI). AI systems includes neural networks and machine learning

ISSN: 1314-3395 (on-line version)
url: http://www.acadpubl.eu/hub/
Special Issue
Abs

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.

Key Words: Intrusion Detection System, Data Mining,
IDS, Network Security

Providing location proofs for mobile users with privacy guarantees

Abstract:

Location based services became popular in the real world where location of query is very important to provide services. The accuracy of user location information plays vital role in rendering needed services. However applications that provide location based services may be subjected to malicious attacks unless they are designed with end-to-end security. In this paper, we propose and implement a protocol that provides secure location-based services. Using this protocol, adhoc mobile users can generate location proofs. This will help trusted mobile users and wireless access points to have genuine communication as part of location-based service applications. Non-transferability of location proofs and integrity achieved with this protocol can have high level of security to location-based information.

Siripuri Kiran

Assistant Professor, Department of CSE, Kakatiya Institute of Technology & Science (India)

A Review towards the Development of Efficient Skyline Query

Ankam Praveen[1], Punnamchandar Pulyala[2]
[1]Assistant Professor, Department of CSE, KAKATIYA INSTITUTE OF TECHNOLOGY AND SCIENCE, India
[2]Assistant Professor, Department of CSE, Ganapathy Engineering College,
India

ABSTRACT

There has been an expanded development in various applications that normally create huge volumes of dubious information. By the coming of such applications, the help of cutting edge investigation questions, for example , the skyline and its variation administrators for enormous indeterminate information has turned out to be essential. Skyline questions intend to prune an inquiry space of extensive quantities of multidimensional information things to a little arrangement of fascinating things by wiping out things that are ruled by others. Skyline query preparing has gotten significant consideration from database and data recovery look into networks. This paper tends to the issue of noting skyline query preparing in versatile condition productively.

Index Terms: Skyline Query processing, wireless sensor networks, SKYLINE EVALUATION ALGORITHM

INTRODUCTION

As one of the mainstream queries in present day databases, the skyline query has gotten much consideration as of late by the database network because of its wide applications in multi-criteria basic leadership. Most past examinations [4, 3] for the skyline issue depend on concentrated databases and distributed figuring condition. As of late some exertion has been taken for skyline query assessment and support in remote sensor systems (WSNs, for example, [2, 1, 5]. In this paper we center around skyline query assessment and maintenance in WSNs by formulating vitality proficient assessment and support calculations for skyline queries to amplify the system lifetime. To accomplish that, we concoct another idea - the neighborhood skyline declaration, which will be utilized to locate a worldwide channel. In light of the worldwide channel, we at that point devise channel based, disseminated assessment and support calculations for skyline queries.

II. Skyline Query Processing

A skyline query is an intense apparatus for multi criteria information investigation, information mining, and basic leadership. Given an arrangement of information tuples with different properties, a skyline query recovers an arrangement of information tuples, called skyline tuples, to frame a skyline. These skyline tuples are not ruled by some other tuples. Here a tuple p is said to command another tuple q if p isn't more awful than q on all traits and p is entirely superior to q on somewhere around one property.

These days, cell phones, for example, PDAs, advanced mobile phones, tablets, and so forth are quickly spreading and turning into a vital piece of human life. In this way, new rising skyline query preparing for portable origination is a promising answer for successful and helpful multicriteria information investigation, information mining and basic leadership in versatile condition. The quick advancement of Skyline queries in portable condition turns into a great pattern in the improvement of multicriteria information examination and additionally information mining and enterprises. In any case, there are still a considerable measure of difficulties for cell phones execution, for example, battery life, stockpiling limit, handling power and secure correspondence ability. Nature of administration essentially constrained by absence of accessible assets.

International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8, Issue-2S3, July 2019

Incremental Feature Selection Method for Software Defect Prediction

N. Gayatri, S. Nickolas, A. Subbarao

Abstract— Software defect prediction models are essential for understanding quality attributes relevant for software organization to deliver better software reliability. This paper focuses mainly based on the selection of attributes in the perspective of software quality estimation for incremental database. A new dimensionality reduction method Wilk's Lambda Average Threshold (WLAT) is presented for selection of optimal features which are used for classifying modules as fault prone or not. This paper uses software metrics and defect data collected from benchmark data sets. The comparative results confirm that the statistical search algorithm (WLAT) outperforms the other relevant feature selection methods for most classifiers. The main advantage of the proposed WLAT method is: The selected features can be reused when there is increase or decrease in database size, without the need of extracting features afresh. In addition, performances of the defect prediction models either remains unchanged or improved even after eliminating 85% of the software metrics.

Key Words: Software defect prediction, ANOVA, Wilk's Lambda, incremental feature selection.

1. INTRODUCTION

Defect prediction has lot of importance in software engineering process. An established approach for this task is building software quality prediction model that estimates program module's quality in terms of defect prone and not defect prone [1-5]. Practitioner can apply such models towards implementing a software quality improvement activity for more cost effective usage of the limited project resources. Software qualities of products or process are characterized by software attributes for software development [6]. From the literature, it is understood that focus is given to software metrics like code level metrics and defect data for building these models as it is assumed that these software metrics will confirm the quality of end product. One can build effective defect prediction model by exploring the knowledge from the historical data. Generally defect prediction models formed from previous data available and after validating the model; it is ready to predict the quality in terms of the fault proneness of modules which are under current development. The goal is to obtain high software reliability and quality with effective use of resources. Since quality prediction models are built using the available software metrics and knowledge stored in them, the selection of relevant metrics data becomes an integral and important part of model building in order to achieve high classification accuracies [6]. In other words, the selection of appropriate quality measurement data is necessary for the model to achieve high predictive accuracy. The number of features will also be reduced to lower the classifier's complexity and computational time. Furthermore, as the days go on there, may be a chance of incrementing or updating the database as some changes may occur in the source code in the maintenance phase. In this case when the database gets incremented, accordingly there is a chance that some metrics may lose their strength and some may gain strength for contributing to the final accuracy of the classifier to predict the class. So there is a need of reselecting the important attributes for better accuracy. Hence the process needs to be run from the scratch which is a time-consuming procedure. To overcome this disadvantage there is a need for finding the feature selection model which works for incremental databases also.

The paper mainly focuses on

i. Feature selection process is based on the statistical measure (WLAT) and thus improving the quality of software defect prediction models and

5. Re usage of the selected features for incremental database without any time overhead with enhanced performance.

There are two types of feature selection methods, namely: Wrappers and Filters. Wrapper method chooses the relevant features based on the predictive accuracy of the model [8]. Though the wrapper methods give better performances, they are comparatively computationally expensive. Filter methods select the features without constructing the predictive accuracy of the model, but by heuristically determined relevant knowledge [7] and wrapper method will be too expensive also for incremental database computing.

This paper focuses on the statistical method based on Analysis of Variance Discriminant Analysis (ANOVA DA) Wilk's lambda to select features which can be used when the database is updated. The comparative analysis is done using three feature ranking techniques and validated using six classifiers. For most of the classifiers, features selected through the proposed method gives better performance and reusability of the selected metrics is also shown.

Revised Version Manuscript Received on July 10, 2019.

N. Gayatri, Assistant Professor, Kakatiya Institute of Technology and Science, Warangal, Telangana, India E-mail: gayatriakkala@gmail.com

J. Nickolas, Professor, National Institute of Technology, Tiruchirappalli, Tamil Nadu, India E- Mail: nickolas@nitt.edu

A. Subbarao, Associate Professor, S R Engineering College, Warangal, Telangana, India E-mail: subbarao_ka@yahoo.com

INDEX VERIFICATION BASED ON MMB CLOUD SERVICE

A.NAVYA SREE
M.Tech Student
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
KAKATIYA INSTITUTE OF TECHNOLOGY AND SCIENCE
WARANGAL, TELANGANA- 506 015

Dr.N.GAYATRI
Assistant Professor
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
KAKATIYA INSTITUTE OF TECHNOLOGY AND SCIENCE
WARANGAL TELANGANA -506 015

Abstract

Cloud brokers have been lately brought as an additional computational layer to facilitate cloud selection and provider control tasks for cloud customers. But, present brokerage schemes on cloud carrier selection typically anticipate that agents are completely trusted, and do no longer offer any assurances over the correctness of the carrier guidelines. It's far then feasible for a compromised or dishonest broking to effortlessly take benefit of the limited competencies of the customers and provide incorrect or incomplete responses. To address this trouble, we endorse a modern Cloud carrier choice Verification (CSSV) scheme and index structures (MMB cloud-tree) to enable cloud customers to hit upon misbehavior of the cloud brokers at some point of the carrier choice system. We show correctness and efficiency of our tactics each theoretically and empirically.

Index terms: Brokerage System, Cloud Service Selection, Merkle hash Tree, Verification

INTRODUCTION

Cloud offerings offer a scalable type of storage space and computing talents, which might be widely employed with the aid of increasingly more business proprietors. This has resulted in a huge range of cloud service carriers (CSPs), offering an extensive range of sources. the provision of various, possibly complicated alternatives, but, makes it difficult for capacity cloud customers to weigh and decide which options fit their necessities the pleasant. The demanding situations are twofold: 1) its miles difficult for cloud customers to collect records about all of the CSPs to be had for his or her choices; 2) its also computationally high priced to pick out an appropriate CSP from a probably massive CSP pool. In mild of these difficulties, each enterprise and academia cautioned introducing an additional computing layer (referred to as cloud brokerage systems) on the pinnacle of the bottom carrier provisioning to enable tasks inclusive of discovery, mediation, and tracking.

FACULTY PAPERS IN CONFERENCES:

S.No	Name of the Faculty	Details of conference where papers were presented	Dates of the conference	Title of the paper presented, including page numbers	PageNo
1	Dr P. Niranjan	International Conference on Inventive Systems and Control (ICISC 2018) IEEE Xplore Compliant - Part Number: CFP18J06-ART	2018	New Approach for Securing Cloud Applications	35
2	K. Vinay Kumar		2018	An Efficient Approach for Dimensionality Reduction and Classification of High Dimensional Text Documents	35
3	Dr V. Shankar	International Conference on Data Science, Machine learning & Applications	Mar 2019	Spatial data management system for Dengue and Swine Flu Epidemiology	36
4	C MADAN KUMAR	ICCIIOT2018	Dec 2018	Text Extraction from Business Cards and Classification of Extracted Text into Predefined Classes	37
5	Siripuri Kiran	International Conference on Research Trends in Science, Technology, Engineering & Management (ICRSTEM-2019)	2019	A Study on Wired Connections with TCP Congestion Control Techniques	38
6	MSB Phridviraj		Jun 2019	A Novel approach for Unsupervised Learning of Transaction Data	38
7	Md. Waseem	2nd International conference on Cognitive Information & Soft Computing (CISC-2019)	April 2019	A Neural Network Model for Attacker Detection using GRU and Modified Kernel of SVM	39

New Approach for Securing Cloud Applications

Abstract:

Cloud computing is a set of IT services such as networks, software systems, storage, hardware, software and resources that are available to customers over the network. IT services for cloud computing are provided by third-party providers that have an infrastructure. The benefits of cloud storage are easy to access means your knowledge, scalability, resiliency, cost efficiency and high reliability of data access anytime, anywhere. Because of these benefits, each organization moves its data to the cloud, which means it uses storage services provided by cloud providers. Therefore, it is necessary to protect these data from unauthorized access, modification or denial of service. Protecting the Cloud means securing the processing (computing) and storage (the database hosted by the cloud provider). In this research paper, the proposed work plan is to use encryption algorithms to eliminate concerns about data privacy in order to enhance security in the cloud from different perspectives of cloud clients.

N Gowtham Kumar
Computer Science & Engineering, KL University, Vijayawada, AP, India

Dr P. Niranjan
Computer Science & Engineering, KITS, Warangal, TS, India

D. Aruna Kumari
Department of Electronics & Computer Engineering, KLEF, Vijayawada, AP, India

A Similarity Function for Feature Pattern Clustering and High Dimensional Text Document Classification

Abstract:

Text document classification and clustering is an important learning task which fits to both data mining and machine learning areas. The learning task throws several challenges when it is required to process high dimensional text documents. Word distribution in text documents plays a very key role in learning process. Research related to high dimensional text document classification and clustering is usually limited to application of traditional distance functions and most of the research contributions in the existing literature did not consider the word distribution in documents. In this research, we propose a novel similarity function for feature pattern clustering and high dimensional text classification. The similarity function proposed is used to carry supervised learning based dimensionality reduction. The important feature of this work is that the word distribution before and after dimensionality reduction is the same. Experiment results prove the proposed approach achieves dimensionality reduction, retains the word distribution and obtained better classification accuracies compared to other measures.

Vinay Kumar Kotte
Department of CSE, Kakatiya Institute of Technology and Science, Warangal, Telangana, India

Srinivasan Rajavelu & Elijah Blessing Rajsingh
Karunya Institute of Technology and Sciences (Deemed to be university), Coimbatore, Tamilnadu, India

SPATIAL DATA MANAGEMENT SYSTEM FOR DENGUE AND SWINEFLU EPIDEMIOLOGY

V. SHANKAR

Associate Professor, Dept. of Computer Science & Engineering, Kakatiya Institute of Technology & Science -Warangal, Telangana, India-506 015
E-mail: vuppu.shankar@gmail.com

Abstract - Recent developments in information technology have enabled collection and processing of vast amounts of personal data, business data and spatial data. It has been widely recognized that spatial data analysis capabilities have not kept up with the need for analyzing the increasingly large volumes of geographic data of various themes that are currently being collected and archived. Our study is carried out on the way to provide the mission-goal strategy (requirements) to predict the epidemic. The co-location rules of spatial data mining are proved to be appropriate to design nuggets for disaster identification and the state-of-the-art and emerging scientific applications require fast access of large quantities of data. Large databases are routinely being collected in science, business and medicine. It has been widely recognized that spatial data analysis capabilities have not kept up with the need for analyzing the increasingly large volumes of geographic data of various themes that are currently being collected and archived. This management system is to obtain and process the data, to interpret the data, and to use the designed algorithms for decision makers (Health Companion) as a basis for action. Our contribution in this paper is to design Algorithms to identify spreading of the dengue from huge datasets

Keywords - Spatial Data Mining; Data Sets; Algorithms; Dengue; Knowledge; Collocation Pattern.

I. INTRODUCTION

During the last decade computer-assisted applications have proven to be of value for the diagnosis of various forms of viruses like Cholera, Dengue Fever, Malaria, and Polio. The increasingly large amount of geographical data available to physicians calls for computer-assisted methods to extract information and knowledge from the available data. Health maps have become available as the use of geographical information systems in health-related contexts increased. Many literary research works has been taken place such as [1] [4]. A formula implemented as Hazard science to Risk Science, towards understanding the hazards and their consequences (risks).

II. APPLYING SPATIAL DATA MINING

Spatial data mining becomes more interesting and important as more spatial data have been accumulated in spatial databases [9]. Spatial Statistics Using spatial statistics measures, dedicated techniques such as cross k-functions with Monte following a probabilistic approach using spatial data mining [1]. Due to larger heterogeneity of spatial data, the providers of geographic data specify different models for same spatial objects. Unknown and unexpected patterns, trends or relationships can hide deep in a huge feature space and make it very hard for analytical methods or visual approaches to find (Miller and Han 2000).

2.1 Mining Collocation Patterns

Mining collocation patterns gives the standard of observing the generic characteristics of a given spatial zone with more relevant Boolean features with their s %(support) and c (confidence) [6] [7]. The spatial Collocation pattern mining framework presented in the erstwhile works has bias on popular events. It may miss some highly confident but "infrequent" Collocation rules by using only "support"-based pruning. In a spatial database S, let F = {f_1… fk} be a set of Boolean spatial features.

Let I = {i_1,…,i_n} be a set of n instances in the spatial database S, where each instance is a vector consisting of [instance-id, location, spatial features]. ~Neighbourhood relation R over pair wise locations in S exists ~ is assumed. The object of this collocation rule mining is to find rules in the form of **A B**, where **A** and **B** are subsets of spatial features. **A** determines the set of spatial features that form the antecedent part of the rule and **B** defines the action and its consequential parts the support and the confidence. The rule indicates the coincidence of the spatial collocation rule absorbs the action of the rule in the "nearby" regions of the spatial objects that comply with the collocation rule. A collocation
pattern C is a set of spatial features, i.e., C F. A neighbor-set L is said to be a row instance of collocation pattern C if every feature in C appears in an instance of L, and there exists no proper subset of L does so. We denote all row instances of a collocation pattern C as row set (C). In other words, row set (C) is the set of neighbor-sets where spatial features in C collocate.

The conditional probability is the probability that a neighbor-set in row set (A) is a part of a neighbor-set in row set (B). Intuitively, the conditional probability p indicates that, whenever we observe the occurrences of the spatial features in A, the probability to find the occurrence of B in a nearby region is p [5].

Text Extraction from Business Cards and Classification of Extracted Text into Predefined Classes

C Madan Kumar[a] and M Brindha[b]

[a]*Assistant Professor of CSE Department, KITS, Warangal, Telangana, India.*
[b]*Assistant Professor, CSE Department, NIT Trichy.*

Abstract: Optical Character Recognition (OCR) is the technology for identification of characters with utmost accuracy possible by employing suitable pre-processing, processing and post processing refinements. Practical application of OCR is very wide ranging from day-to-day need to scientific research purposes. One very crucial application is to Automate Digitalisation of Business cards. Since Business cards comes in different fonts and sizes, and most importantly with different lighting conditions, applying OCR can be done after careful processing. Avoiding noise from source image is one of the most crucial step in any image-processing process and it has major weightage in the accuracy of further step and thus indirectly has a huge contribution for our final outcome. Further proper noise cancellation in our source image can reduce number of future steps required to attain good accuracy and also avoid problem of iterating our sample over cycles to avoid better contrast or to distinguish text in our source lying in a noisy matrix. Digitalising business cards aims at classification of the text extracted from our source image in the hard copy of the business card directly into respected following classified fields so that it becomes a lot easier to proceed any desired function ones aims to do with that business card in this online era.

Keywords: Optical Character Recognition (OCR), Text Extraction, Biggest Contour Method, Houglines Method, Canny Edge Detection, Image Segmentation

1. Introduction

Digital image processing refers to use of computerized algorithms to perform image processing on digital images. Optical character recognition or OCR(Tesseract algorithm) is a form of information entry for business cards, e-mails, pan cards, Id cards, which scans a document in written form or printed form and retrieving the text out of it. The idea of Business card Digitalisation has been evolved from Automatic License Plate Recognition.

Digitizing the Business Cards is a real challenge. They come in different formats and fonts. Text extraction in different lighting conditions is very difficult. A formal structure for the business card reader which was innovated is reported. A Boundary Detection method is proposed called Biggest Contour method and Houghlines Transformation method, image extraction and segmentation technique based on a statistical method called Connected Component Method.

The task is to detect the boundary of the card by eliminating the background. The image was subdivided into an array of smaller blocks, over which gray thresholding i**s** used to compute local thresholds. These thresholds are then stored in another array. This method of dividing the image into smaller images and then applying threshold, instead of applying threshold on the image as a whole is known as adaptive threshold.

A Study on Wired Connections with TCP Congestion Control Techniques

Abstract:

There are numerous sorts of TCP Congestion Control Techniques consisting of Tahoe, Reno, New Reno, SACK, Las vega and also these techniques are various from each other. Whenever for any type of package the timeout happens, these techniques enter into duty as well as what is the result on throughput, performance, efficiency when compared to TCP Las vega. When congestion takes place at any kind of router their function is to adhere to a particular protocol, comply with some algorithm for complying with one more course and also info regarding that package is saved in the storage space media as well as the simulator which is utilized for executing this is ns2 set up in Ubuntu variation.

Siripuri Kiran

Assistant Professor, Department of CSE, Kakatiya Institute of Technology & Science (India)

M. Anusha

Software Programmer, United States of America

A Novel approach for Unsupervised Learning of Transaction Data

ABSTRACT:

Incremental clustering is a technique which can be applied when the dataset is not constant and keeps updating. Normally when kmeans clustering is applied and if the dataset is modified then the clustering must be done from start. Similarly, for maximum capture procedure proposed in our previous research the clustering task must be carried from the start. In this paper, we propose an incremental approach for clustering transaction data which can be used for customer segmentation and other related applications. Experiments are conducted and three approaches are compared in terms of CPU utilization. It is observed that incremental approach required less CPU utilization.

MSB Phridviraj

Assistant Professor, Department of CSE, Kakatiya Institute of Technology & Science (India)

A Neural Network Model for Attacker Detection using GRU and Modified Kernel of SVM

Abstract:

Over past few decades neural network changed the way of traditional computing many different models has proposed depending upon data intensity, predictions, and recognition and so on. Among which Gated Recurrent Unit (GRU) is created for variety of long short-term memory (LSTM) unit, which is part of recurrent neural network (RNN). These models proved to be dominant for range of machine learning job such as predictions, speech recognition, sentiment analysis and natural language processing. In this proposed model, a support vector machine (SVM) with modified kernel as final output layer for prediction is used instead of traditional approach of softmax and log loss function is used to calculate the loss. Proposed technique is applied for binary classification for intrusion detection using honeypot dataset (2013) network traffic sequence of Kyoto University. Results shows a prominent change in training efficiency of ≈89.45% and testing efficiency of ≈88.15% when compared with softmax output layer. We can conclude that linear SVM with modified kernel as output layer outperform compared with softmax in prediction time

Dr Narsimha reddy Soora

Associate Professor, Department of CSE, Kakatiya Institute of Technology & Science (India)

MD. SHARFUDDIN WASEEM

Assistant Professor, Department of CSE, Kakatiya Institute of Technology & Science (India)

STUDENT PAPERS IN JOURNALS

KAKATIYA INSTITUTE OF TECHNOLOGY AND SCIENCES

Contents

S.No	Name of the Student	Title of the paper	Name of the journal	Page No.
1	Mohammed Ehsan Ur Rahman	"Image retrieval by Integrating Global Correlation of Colour and Histograms with Local Texture Features "	Springer	41
2	P. Satya Sai Teja	"A Secure Communication through Quantum Key Distribution Protocols"	IRJECE	42

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-019-08029-7

Image Retrieval by Integrating Global Correlation of Color and Intensity Histograms with Local Texture Features

Suresh Kumar Kanaparthi[1] & U. S. N. Raju[1] & P. Shanmukhi[2] & G. Khyathi Aneesha[2] & Mohammed Ehsan Ur Rahman[3]

Received: 9 September 2018 / Revised: 8 June 2019 / Accepted: 19 July 2019
Published online: 20 August 2019

Abstract:

Research on Content-Based Image Retrieval is being done to improvise existing methods. Most of the techniques that were proposed use color and texture features independently. In this paper, to get the correspondence between color and texture, we use congruence on Hue, Saturation, and Intensity by using inter-channel voting. Gray Level Co-occurrence Matrix (GLCM) on Diagonally Symmetric Pattern is computed to get texture features of an image. The similarity metrics between two images is computed using congruence and GLCM. To measure the performance; Average Precision Rate (APR), Average Recall Rate (ARR), F-measure, Average Normalized Modified Retrieval Rank (ANMRR) are calculated. In addition to these parameters, one more parameter has been proposed: Total Minimum Retrieval Epoch (TMRE) to calculate the average number of images to be traversed for each query image to get all the images of that category. To validate the performance of the proposed method, it has been applied to six image databases: Corel-1 K, Corel-5 K, Corel-10 K, VisTex, STex, and Color Brodatz. The results of most of the databases show significant improvement.

Keywords CBIR · Inter-Channel Voting · Total Minimum Retrieval Epoch · Diagonally Symmetric Pattern · Color Auto Correlogram

Introduction

The exponential growth of digital libraries due to mobile phones and digital cameras has resulted in enormous size of image database. Maintaining this huge database is an extremely monotonous and unmanageable task. So, an effective technique needs to be developed, which can automatically retrieve the required image from the database. Content-based image retrieval (CBIR) is one of the widely accepted solutions to accomplish this task [43]. “Content-Based”

DOI: 10.24178/irjece.2018.4.3.14 IRJECE Vol 4(3)
Sep 2018

A Secure Communication through Quantum Key Distribution Protocols

P.Satya Sai Teja
Kakatiya Institute of Technology and Science, Hanamkonda, Warangal, India
Satyateja401@gmail.com

P.Mounika Lakshmi
TechMahindra
Hyderabad, Telangana, India
pmounikalakshmi401@gmail.com

K.Vinay Kumar
Kakatiya Institute of Technology and Science, Hanamkonda, Warangal, India
kotte.vinaykumar@gmail.com

Abstract - Quantum cryptography is a new method of communication offering the security of the inviolability by using Law of Nature.Quantum Cryptography uses different secure communication by applying the phenomena of quantum physics. Unlike traditional classical cryptography, which uses mathematical techniques to restrict eavesdroppers, quantum cryptography is focused on the properties of physics of light for information. Quantum cryptography depends only on the validity of quantum theory, i.e., it is guarantied directly by the laws of physics. This is a different from any classical cryptographic techniques. This paper summarizes the current state of quantum cryptography and provides potential extensions of its feasibility as a mechanism for securing existing communication systems.

Keywords: Cryptography, Classical Cryptography, Quantum Cryptography, Qubits, Quantum Key Distribution, QKD Protocols

I. Introduction

The physics of quantum cryptography helps in achieving the possibilities for strong cryptography. The characteristics of quantum mechanics include the existence of indivisible quanta and of entangled systems, both of which are present at the root of quantum cryptography (Q). Q is one of the few best commercial applications of quantum physics at the single quantum level for achieving the secure communication.

Other applications of quantum mechanics to cryptography, which tend to come in three flavours:

- ✓ Quantum mechanism helps in breaking classical cryptography protocols.
- ✓ Quantum states are undetective which helps them to stay secure when classical cryptography methods are breakable y eavesdroppers.
- ✓ Cryptography is used to protect the information stored in quantum instead of classical information.

II. Classical 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 types-symmetric 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.

ARTICLES CONTRIBUTED BY STUDENTS

KAKATIYA INSTITUTE OF TECHNOLOGY AND SCIENCES

Contents

Hacking the human brain

One of the greatest challenges facing artificial intelligence development is understanding the human brain and figuring out how to mimic it. Now, one group reports in *ACS Nano* that they have developed an artificial synapse capable of simulating a fundamental function of our nervous system -- the release of inhibitory and stimulatory signals from the same "pre-synaptic" terminal.

The human nervous system is made up of over 100 trillion synapses, structures that allow neurons to pass electrical and chemical signals to one another. In mammals, these synapses can initiate and inhibit biological messages. Many synapses just relay one type of signal, whereas others can convey both types simultaneously or can switch between the two. To develop artificial intelligence systems that better mimic human learning, cognition and image recognition, researchers are imitating synapses in the lab with electronic components. Most current artificial synapses, however, are only capable of delivering one type of signal. colleagues sought to create an artificial synapse that can reconfigurably send stimulatory and inhibitory signals.

The researchers developed a synaptic device that can reconfigure itself based on voltages applied at the input terminal of the device. A junction made of black phosphorus and tin selenide enables switching between the excitatory and inhibitory signals. This new device is flexible and versatile, which is highly desirable in artificial neural networks. In addition, the artificial synapses may simplify the design and functions of nervous system simulations.

J. Sahithi **(B16CS022)**

Artificial Intelligence for obtaining chemical fingerprints

Drastic advances in research of artificial intelligence have led to a wide range of fascinating developments in this area over the last decade. Autonomously driven cars, but also everyday applications such as search engines and spam filters illustrate the versatility of methods from the field of artificial intelligence.

Infrared spectroscopy is one of the most valuable experimental methods to gain insight into the world of molecules. Infrared spectra are chemical fingerprints that provide information on the composition and properties of substances and materials. In many cases, these spectra are very complex -- a detailed analysis makes computer-aided simulations indispensable. While quantum chemical calculations in principle enable extremely precise prediction of infrared spectra, their applicability in practice is made difficult by the high computational effort associated with them. For this reason, reliable infrared spectra can only be calculated for relatively small chemical systems.

An international group of researchers led by Philipp Marquetand from the Faculty of Chemistry at the University of Vienna has now found a way to accelerate these simulations using artificial intelligence. For this purpose, so-called artificial neural networks are used, mathematical models of the human brain. These are able to learn the complex quantum mechanical relationships that are necessary for the modelling of infrared spectra by using only a few examples. In this way, the scientists can carry out simulations within a few minutes, which would otherwise take thousands of years even with modern supercomputers -- without sacrificing reliability. "We can now finally simulate chemical problems that could not be overcome with the simulation techniques used up to now," says Michael Gast egger, the first author of the study.

Based on the results of this study, the researchers are confident that their method of spectra prediction will be widely used in the analysis of experimental infrared spectra in the future.

D. Sheethal **(B16CS036)**

Deep learning enables real-time imaging around corners

The new imaging system uses a commercially available camera sensor and a powerful, but otherwise standard, laser source that is similar to the one found in a laser pointer. The laser beam bounces off a visible wall onto the hidden object and then back onto the wall, creating an interference pattern known as a speckle pattern that encodes the shape of the hidden object.

Reconstructing the hidden object from the speckle pattern requires solving a challenging computational problem. Short exposure times are necessary for real-time imaging but produce too much noise for existing algorithms to work. To solve this problem, the researchers turned to deep learning.

"Compared to other approaches for non-line-of-sight imaging, our deep learning algorithm is far more robust to noise and thus can operate with much shorter exposure times," said co-author Prasanna Rangarajan from Southern Methodist University. "By accurately characterizing the noise, we were able to synthesize data to train the algorithm to solve the reconstruction problem using deep learning without having to capture costly experimental training data."

Seeing around corners

The researchers tested the new technique by reconstructing images of 1-centimeter-tall letters and numbers hidden behind a corner using an imaging setup about 1 meter from the wall. Using an exposure length of a quarter of a second, the approach produced reconstructions with a resolution of 300 microns.

The research is part of DARPA's Revolutionary Enhancement of Visibility by Exploiting Active Light-fields (REVEAL) program, which is developing a variety of different techniques to image hidden objects around corners. The researchers are now working to make the system practical for more applications by extending the field of view so that it can reconstruct larger objects.

K. Prashanth Reddy **(B17CS139)**

Bitcoin causing carbon dioxide emissions

Although Bitcoin is a virtual currency, the energy consumption associated with its use is very real. For a Bitcoin transfer to be executed and validated, a mathematical puzzle must be solved by an arbitrary computer in the global Bitcoin network. The network, which anyone can join, rewards the puzzle solvers in Bitcoin. The computing capacity used in this process -- known as Bitcoin mining -- has increased rapidly in recent years. Statistics show that it quadrupled in 2018 alone.

Consequently, the Bitcoin boom raises the question of whether the cryptocurrency is imposing an additional burden on the climate. Several studies have attempted to quantify the CO_2 emissions caused by Bitcoin mining. "These studies are based on a number of approximations, however," says Christian Stoll, who conducts research at the Technical University of Munich (TUM) and the Massachusetts Institute of Technology (MIT).

"Detective work" to track down the power consumption

Therefore, a team of management sciences and informatics researchers at TUM has carried out the most detailed calculation of the carbon footprint of the Bitcoin system

to date. Working like detectives, they proceeded step by step to gather conclusive data.

The team began by calculating the power consumption of the network. This depends primarily on the hardware used for Bitcoin mining. "Today special systems are used, known as ASIC-based miners," explains Stoll. In 2018 the three manufacturers who control the ASIC miner market planned IPOs. The team used the mandatory IPO filings to calculate the market shares of the companies' respective products. The study also had to consider whether the mining was being done by someone running just one miner at home or in one of the large-scale "farms" set up in recent years by professional operators. "In those operations, extra energy is needed just for the cooling of the data centre," says Stoll. To investigate the orders of magnitude involved, the team used statistics released by a public pool of different miners showing the computing power of its members.

68 percent of computing power located in Asia

The researchers determined the annual electricity consumption by Bitcoin, as of November 2018, to be about 46 TWh. And how much CO_2 is emitted when this energy is generated? Here, too, the research team wanted to go beyond mere estimates. The key question, therefore: Where are the miners located?

Once again, live tracking data from the mining pools provided the decisive information. "In these groups, miners combine their computing power in order to get a quicker turn in the reward for solving puzzles -- similar to people in lottery pools," explains Stoll. The IP addresses in the statistics published by the two biggest pools showed that miners tend to join pools in or near their home countries. Based on these data, the team was able to localize 68 percent of the Bitcoin network computing power in Asian countries, 17 percent in European countries, and 15 percent in North America. The researchers cross-checked this conclusion against the results of another method by localizing the IP addresses of individual miners using an internet of things search engine. They then combined their results with statistics on the carbon intensity of power generation in the various countries.

"Linking large-scale mining operations to renewable energy production"

The conclusion of the study: The Bitcoin system has a carbon footprint of between 22 and 22.9 megatons per year. That is comparable to the footprint of such cities as Hamburg, Vienna or Las Vegas.

"Naturally there are bigger factors contributing to climate change. However, the carbon footprint is big enough to make it worth discussing the possibility of regulating cryptocurrency mining in regions where power generation is especially carbon-intensive," says Christian Stoll. "To improve the ecological balance, one possibility might be to link more mining farms to additional renewable generating capacity."

K. Raj Kamal **(B19CS042)**

What Is the Internet of Bodies? And How Is It Changing Our World?

When the Internet of Things (IoT) connects with your body, the result is the Internet of Bodies (IoB). The Internet of Bodies (IoB) is an extension of the IoT and basically connects the human body to a network through devices that are ingested, implanted, or connected to the body in some way. Once connected, data can be exchanged, and the body and device can be remotely monitored and controlled.

There are three generations of Internet of Bodies that include:

· Body external: These are wearable devices such as Apple Watches or Fitbits that can monitor our health.

· Body internal: These include pacemakers, cochlear implants, and digital pills that go inside our bodies to monitor or control various aspects of our health.

· Body embedded: The third generation of the Internet of Bodies is embedded technology where technology and the human body are melded together and have a real-time connection to a remote machine.

Progress in wireless connectivity, materials, and tech innovation is allowing implantable medical devices (IMD) to scale and be viable in many applications.

The most recognized example of Internet of Bodies is a defibrillator or pacemaker, a small device placed in the abdomen or chest to help patients with heart conditions control abnormal heart rhythms with electrical impulses. In 2013, former United States Vice President Dick Cheney got his Wi-Fi-connected defibrillator replaced with one without Wi-Fi capacity. It was feared that he could be assassinated by electric shock if a rogue agent hacked the device.

A "smart pill" is another IoB device. These pills have edible electronic sensors and computer chips in them. Once swallowed, these digital pills can collect data from our organs and then send it to a remote device connected to the internet. The first digital chemotherapy pill is now in use that combines chemotherapy drugs with a sensor that captures, records, and shares information with healthcare providers (with the patient's consent) regarding the drug dosage and time, plus other data on rest and activity, heart rate and more.

"Smart contact lenses" are being developed that integrate sensors and chips that can monitor health diagnostics based on information from the eye and eye fluid. One smart contact lens in development aims to monitor glucose levels that will hopefully allow diabetics to monitor their glucose levels without repeated pinpricks throughout the day.

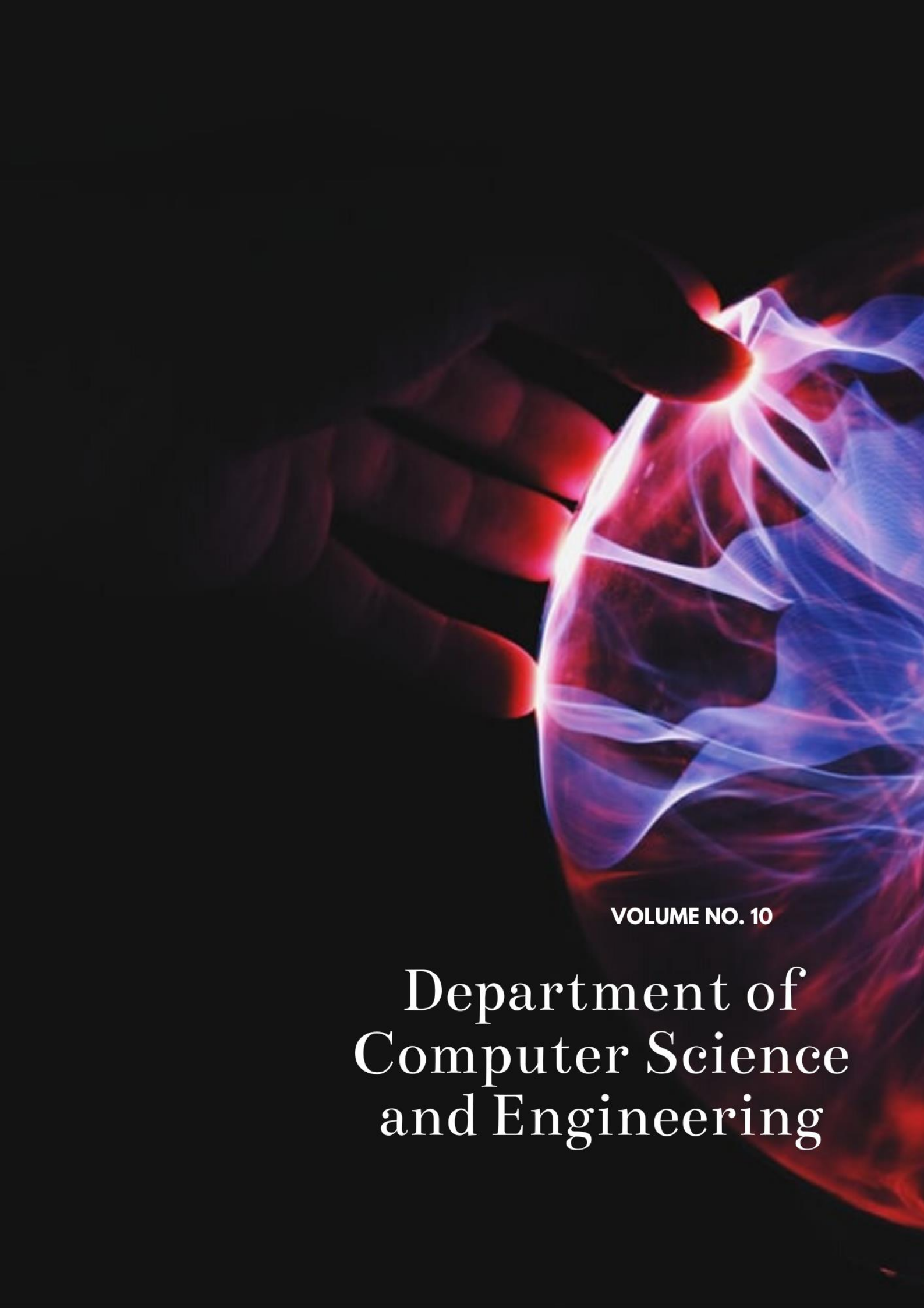
Taking it up a notch is the Brain Computer Interface (BCI), where a person's brain is actually merged with an external device for monitoring and controlling in real-time. The ultimate goal is to help restore function to individuals with disabilities by using brain signals rather than conventional neuromuscular pathways.

The security challenges faced by Internet of Bodies tech are similar to what plagues Internet of Things generally, but there can be life and death consequences when IoB devices are involved. Additionally, IoB devices create another cyber security challenge that will need to be safeguarded from hackers.

Privacy is also of paramount concern. Questions about who can access the data and for what purpose need answers. For example, a device that monitors health diagnostics could also track unhealthy behaviours. Will health insurance companies be able to deny coverage when a customer's IoB device reports their behaviour? A cochlear implant could restore hearing, but it might also record all audio in a person's environment. Will that data remain private?

As Internet of Bodies tech continues to grow, regulatory and legal issues will have to be resolved and policies built around the proper use of the technology.

S. Amogh Madhav **(B17CS001)**



VOLUME NO. 10

Department of Computer Science and Engineering