

INTERNATIONAL JOURNAL OF RESEARCH IN COMMERCE, IT & MANAGEMENT

I
J
R
C
M



A Monthly Double-Blind Peer Reviewed (Refereed/Juried) Open Access International e-Journal - Included in the International Serial Directories

Indexed & Listed at:

Ulrich's Periodicals Directory ©, ProQuest, U.S.A., EBSCO Publishing, U.S.A., Cabell's Directories of Publishing Opportunities, U.S.A.

Open J-Gate, India [link of the same is duly available at Inlibnet of University Grants Commission (U.G.C.)],
Index Copernicus Publishers Panel, Poland with IC Value of 5.09 & number of libraries all around the world.

Circulated all over the world & Google has verified that scholars of more than 2980 Cities in 165 countries/territories are visiting our journal on regular basis.

Ground Floor, Building No. 1041-C-1, Devi Bhawan Bazar, JAGADHRI – 135 003, Yamunanagar, Haryana, INDIA

<http://ijrcm.org.in/>

CONTENTS

Sr. No.	TITLE & NAME OF THE AUTHOR (S)	Page No.
1.	EFFICIENT MARKET HYPOTHESIS: A PRIMARY EXPLORATORY STUDY ON RELEVANCE OF INFORMATION BIAS <i>PRADEEPA.M, VIDYA.R & DR. R.HIREMANINAIK</i>	1
2.	COMPARATIVE ANALYSIS OF THE PARAMETERS OF DYNAMIC CHANNEL ALLOCATION FOR COLLISION LESS DATA TRANSMISSION <i>SUSHANT JHINGRAN & SAOUD SARWAR</i>	5
3.	A STUDY OF OVERDUES IN SELECTED PACS'S: WITH SPECIAL REFERENCE TO MANGALAGIRI BRANCH OF GDCCB LTD., TENALI, DURING 2008-'12 <i>DR. MADDALI ARAVIND & PALUTLA NAGAMANI</i>	8
4.	FDI IN AVIATION: WILL IT SERVE AS A GAME CHANGER FOR INDIAN AIRLINES INDUSTRY? <i>P.K. KOTIA & MEENAL LODHA</i>	14
5.	FACTORS INFLUENCING THE EMPLOYEES' INTENTION TO CHANGE JOB FROM PUBLIC TO PRIVATE SECTOR BANKS & VICE VERSA: AN EMPIRICAL STUDY OF BANKING SECTOR EMPLOYEES IN INDIA <i>DR. RENU SHARMA</i>	19
6.	HIGHER EDUCATION IN INDIA: CONFRONTING THE CHALLENGE OF CHANGE <i>DR. PAWAN KUMAR SHARMA</i>	24
7.	A TECHNIQUE WITH DIFFERENTIATED SAMPLING IN ANOMALY DETECTION SYSTEM FOR OUTLIER IDENTIFICATION <i>SARANYA.C & VEENA.S</i>	27
8.	AN IMPROVED APPROACH OF RISK ANALYSIS FOR IT & ITES ORGANIZATIONS <i>CHELLAM SHENBAGAM</i>	31
9.	DERIVATIVES MARKET IN INDIA <i>GHANATHE RAMESH, CHEGU JYOTHI, KONDA SANDEEP KUMAR & GOWLIKAR VINESH KUMAR</i>	37
10.	CORPORATE GOVERNANCE IN BRICS: A COMPARATIVE STUDY <i>MINNY NARANG, DEEPALI MALHOTRA & SWATI SETH</i>	41
11.	EFFECT OF INSTITUTIONAL PRESSURES ON THE RELATION BETWEEN FINANCIAL AND SUSTAINABLE PERFORMANCE OF FIRMS <i>AMOGH TALAN, PRIYANKA PANDEY & GAURAV TALAN</i>	48
12.	FOREIGN DIRECT INVESTMENT: ECONOMIC GROWTH AND ISLAMIC BANKING INDUSTRY <i>MEHDI BEHNAME & MAHDI MOSTAFAVI</i>	52
13.	THE EFFECT OF CAPITAL STRUCTURE ON PROFITABILITY: EVIDENCE FROM THE PETROCHEMICAL COMPANIES IN THE KINGDOM OF SAUDI ARABIA <i>AHMED AL AJLOUNI & MUNIR SHAWER</i>	56
14.	ERA OF KNOWLEDGE MANAGEMENT IN INDUSTRY AND INFORMATION RESEARCH WORLD <i>G.SANTOSH KUMAR & P.SHIRISHA</i>	63
15.	AN APPROACH INTO COMMERCE EDUCATION AFTER GLOBALIZATION-CHALLENGES AND OPPORTUNITIES <i>RAVINDER KAUR & MANMEET KAUR</i>	66
16.	A STUDY ON STRESS LEVEL OF EMPLOYEES OF INFORMATION TECHNOLOGY COMPANIES IN CHENNAI, TAMILNADU <i>DR. RETHINA BAI.R</i>	70
17.	INNOVATIVE FINANCIAL PRODUCTS: A STUDY OF CHALLENGES AND OPPORTUNITIES AT UDAIPUR, INDIA <i>DR. YOGESH JAIN</i>	73
18.	FINANCIAL MEASURES USING Z- SCORES WITH SPECIAL REFERENCE TO BAJAJ AUTO LIMITED <i>KOKILA H S</i>	84
19.	CONSUMER BEHAVIOUR TOWARDS REFRIGERATOR IN MYSORE CITY <i>ALUREGOWDA</i>	88
20.	THE STUDY OF FACTORS RELATED TO VOCATIONAL WEAR IN MUNICIPALITY'S EMPLOYEES <i>MAJID NILI AHMADABADI & HAMID DEGHANI</i>	94
21.	UTILIZING INTERNET AS ON-LINE SALES TOOL FOR EMPOWERMENT OF BUSINESS EDUCATION GRADUATES IN NIGERIA <i>TITUS AMODU UMORU</i>	97
22.	CONSUMERS' ATTITUDES TOWARDS THE DAIRY PRODUCT IN THE ETHIOPIAN MARKET: CASE OF ADDIS ABABA <i>SARFARAZ KARIM, SRAVAN KUMAR REDDY & ELIAS GIZACHEW</i>	100
23.	IMPLEMENTATION OF ABC IN BANGLADESH: REQUIRED PREREQUISITES AND THEIR AVAILABILITY <i>TANZILA AHMED & TAHMINA AHMED</i>	105
24.	WIDENING REGIONAL ECONOMIC DISPARITIES IN INDIA <i>SUSANTA KR. SUR & DR. TOWHID E AMAN</i>	109
25.	MODELLING A STACKELBERG GAME IN A TWO STAGE SUPPLY CHAIN UNDER RETURN POLICY CONTRACTS: SOLVING A DECISION PROBLEM FOR A CAPACITY CONSTRAINED MANUFACTURER <i>SHIRSENDU NANDI</i>	113
26.	JOB SATISFACTION IN BANKING: A COMPARATIVE STUDY BETWEEN PUBLIC AND PRIVATE SECTOR BANKS IN DEHRADUN, UTTARAKHAND <i>RATNAMANI</i>	118
27.	PERFORMANCE APPRAISAL SYSTEM FOR SALES FORCE IN FASTENER INDUSTRY: STUDY OF LPS ROHTAK <i>HARDEEP</i>	124
28.	IMPACT OF GLOBAL RECESSION ON INDIAN FINANCIAL MARKET <i>SHIKHA MAKKAR</i>	129
29.	IMPACT OF PRIVATIZATION ON INDIAN BANKING SECTOR IN THE GLOBALIZATION ERA <i>PRIYANKA RANI, NANCY ARORA & RENU BALA</i>	134
30.	POST IMPACT ANALYSIS OF GLOBALIZATION ON TOURISM SERVICES <i>BIVEK DATTA</i>	139
	REQUEST FOR FEEDBACK & DISCLAIMER	142

CHIEF PATRON

PROF. K. K. AGGARWAL

Chairman, Malaviya National Institute of Technology, Jaipur

(An institute of National Importance & fully funded by Ministry of Human Resource Development, Government of India)

Chancellor, K. R. Mangalam University, Gurgaon

Chancellor, Lingaya's University, Faridabad

Founder Vice-Chancellor (1998-2008), Guru Gobind Singh Indraprastha University, Delhi

Ex. Pro Vice-Chancellor, Guru Jambheshwar University, Hisar

FOUNDER PATRON

LATE SH. RAM BHAJAN AGGARWAL

Former State Minister for Home & Tourism, Government of Haryana

Former Vice-President, Dadri Education Society, Charkhi Dadri

Former President, Chinar Syntex Ltd. (Textile Mills), Bhiwani

CO-ORDINATOR

AMITA

Faculty, Government M. S., Mohali

ADVISORS

DR. PRIYA RANJAN TRIVEDI

Chancellor, The Global Open University, Nagaland

PROF. M. S. SENAM RAJU

Director A. C. D., School of Management Studies, I.G.N.O.U., New Delhi

PROF. M. N. SHARMA

Chairman, M.B.A., Haryana College of Technology & Management, Kaithal

PROF. S. L. MAHANDRU

Principal (Retd.), Maharaja Agrasen College, Jagadhri

EDITOR

PROF. R. K. SHARMA

Professor, Bharti Vidyapeeth University Institute of Management & Research, New Delhi

CO-EDITOR

DR. BHAVET

Faculty, Shree Ram Institute of Business & Management, Urjani

EDITORIAL ADVISORY BOARD

DR. RAJESH MODI

Faculty, Yanbu Industrial College, Kingdom of Saudi Arabia

PROF. SANJIV MITTAL

University School of Management Studies, Guru Gobind Singh I. P. University, Delhi

PROF. ANIL K. SAINI

Chairperson (CRC), Guru Gobind Singh I. P. University, Delhi

DR. SAMBHAVNA

Faculty, I.I.T.M., Delhi

DR. MOHENDER KUMAR GUPTA

Associate Professor, P. J. L. N. Government College, Faridabad

DR. SHIVAKUMAR DEENE

Asst. Professor, Dept. of Commerce, School of Business Studies, Central University of Karnataka, Gulbarga

ASSOCIATE EDITORS

PROF. NAWAB ALI KHAN

Department of Commerce, Aligarh Muslim University, Aligarh, U.P.

PROF. ABHAY BANSAL

Head, Department of Information Technology, Amity School of Engineering & Technology, Amity University, Noida

PROF. A. SURYANARAYANA

Department of Business Management, Osmania University, Hyderabad

DR. SAMBHAV GARG

Faculty, Shree Ram Institute of Business & Management, Urjani

PROF. V. SELVAM

SSL, VIT University, Vellore

DR. PARDEEP AHLAWAT

Associate Professor, Institute of Management Studies & Research, Maharshi Dayanand University, Rohtak

DR. S. TABASSUM SULTANA

Associate Professor, Department of Business Management, Matrusri Institute of P.G. Studies, Hyderabad

SURJEET SINGH

Asst. Professor, Department of Computer Science, G. M. N. (P.G.) College, Ambala Cantt.

TECHNICAL ADVISOR

AMITA

Faculty, Government M. S., Mohali

FINANCIAL ADVISORS

DICKIN GOYAL

Advocate & Tax Adviser, Panchkula

NEENA

Investment Consultant, Chambaghat, Solan, Himachal Pradesh

LEGAL ADVISORS

JITENDER S. CHAHAL

Advocate, Punjab & Haryana High Court, Chandigarh U.T.

CHANDER BHUSHAN SHARMA

Advocate & Consultant, District Courts, Yamunanagar at Jagadhri

SUPERINTENDENT

SURENDER KUMAR POONIA

CALL FOR MANUSCRIPTS

We invite unpublished novel, original, empirical and high quality research work pertaining to recent developments & practices in the areas of Computer Science & Applications; Commerce; Business; Finance; Marketing; Human Resource Management; General Management; Banking; Economics; Tourism Administration & Management; Education; Law; Library & Information Science; Defence & Strategic Studies; Electronic Science; Corporate Governance; Industrial Relations; and emerging paradigms in allied subjects like Accounting; Accounting Information Systems; Accounting Theory & Practice; Auditing; Behavioral Accounting; Behavioral Economics; Corporate Finance; Cost Accounting; Econometrics; Economic Development; Economic History; Financial Institutions & Markets; Financial Services; Fiscal Policy; Government & Non Profit Accounting; Industrial Organization; International Economics & Trade; International Finance; Macro Economics; Micro Economics; Rural Economics; Co-operation; Demography; Development Planning; Development Studies; Applied Economics; Development Economics; Business Economics; Monetary Policy; Public Policy Economics; Real Estate; Regional Economics; Political Science; Continuing Education; Labour Welfare; Philosophy; Psychology; Sociology; Tax Accounting; Advertising & Promotion Management; Management Information Systems (MIS); Business Law; Public Responsibility & Ethics; Communication; Direct Marketing; E-Commerce; Global Business; Health Care Administration; Labour Relations & Human Resource Management; Marketing Research; Marketing Theory & Applications; Non-Profit Organizations; Office Administration/Management; Operations Research/Statistics; Organizational Behavior & Theory; Organizational Development; Production/Operations; International Relations; Human Rights & Duties; Public Administration; Population Studies; Purchasing/Materials Management; Retailing; Sales/Selling; Services; Small Business Entrepreneurship; Strategic Management Policy; Technology/Innovation; Tourism & Hospitality; Transportation Distribution; Algorithms; Artificial Intelligence; Compilers & Translation; Computer Aided Design (CAD); Computer Aided Manufacturing; Computer Graphics; Computer Organization & Architecture; Database Structures & Systems; Discrete Structures; Internet; Management Information Systems; Modeling & Simulation; Neural Systems/Neural Networks; Numerical Analysis/Scientific Computing; Object Oriented Programming; Operating Systems; Programming Languages; Robotics; Symbolic & Formal Logic; Web Design and emerging paradigms in allied subjects.

Anybody can submit the **soft copy** of unpublished novel; original; empirical and high quality **research work/manuscript anytime** in **M.S. Word format** after preparing the same as per our **GUIDELINES FOR SUBMISSION**; at our email address i.e. infoijrcm@gmail.com or online by clicking the link **online submission** as given on our website ([FOR ONLINE SUBMISSION, CLICK HERE](#)).

GUIDELINES FOR SUBMISSION OF MANUSCRIPT

1. **COVERING LETTER FOR SUBMISSION:**

DATED: _____

THE EDITOR
IJRCM

Subject: SUBMISSION OF MANUSCRIPT IN THE AREA OF.

(e.g. Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/Engineering/Mathematics/other, please specify)

DEAR SIR/MADAM

Please find my submission of manuscript entitled ' _____ ' for possible publication in your journals.

I hereby affirm that the contents of this manuscript are original. Furthermore, it has neither been published elsewhere in any language fully or partly, nor is it under review for publication elsewhere.

I affirm that all the author (s) have seen and agreed to the submitted version of the manuscript and their inclusion of name (s) as co-author (s).

Also, if my/our manuscript is accepted, I/We agree to comply with the formalities as given on the website of the journal & you are free to publish our contribution in any of your journals.

NAME OF CORRESPONDING AUTHOR:

Designation:
Affiliation with full address, contact numbers & Pin Code:
Residential address with Pin Code:
Mobile Number (s):
Landline Number (s):
E-mail Address:
Alternate E-mail Address:

NOTES:

- a) The whole manuscript is required to be in **ONE MS WORD FILE** only (pdf. version is liable to be rejected without any consideration), which will start from the covering letter, inside the manuscript.
- b) The sender is required to mention the following in the **SUBJECT COLUMN** of the mail:
New Manuscript for Review in the area of (Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/Engineering/Mathematics/other, please specify)
- c) There is no need to give any text in the body of mail, except the cases where the author wishes to give any specific message w.r.t. to the manuscript.
- d) The total size of the file containing the manuscript is required to be below **500 KB**.
- e) Abstract alone will not be considered for review, and the author is required to submit the complete manuscript in the first instance.
- f) The journal gives acknowledgement w.r.t. the receipt of every email and in case of non-receipt of acknowledgment from the journal, w.r.t. the submission of manuscript, within two days of submission, the corresponding author is required to demand for the same by sending separate mail to the journal.

2. **MANUSCRIPT TITLE:** The title of the paper should be in a 12 point Calibri Font. It should be bold typed, centered and fully capitalised.

3. **AUTHOR NAME (S) & AFFILIATIONS:** The author (s) **full name, designation, affiliation (s), address, mobile/landline numbers, and email/alternate email address** should be in italic & 11-point Calibri Font. It must be centered underneath the title.

4. **ABSTRACT:** Abstract should be in fully italicized text, not exceeding 250 words. The abstract must be informative and explain the background, aims, methods, results & conclusion in a single para. Abbreviations must be mentioned in full.

5. **KEYWORDS:** Abstract must be followed by a list of keywords, subject to the maximum of five. These should be arranged in alphabetic order separated by commas and full stops at the end.
6. **MANUSCRIPT:** Manuscript must be in **BRITISH ENGLISH** prepared on a standard A4 size **PORTRAIT SETTING PAPER**. It must be prepared on a single space and single column with 1" margin set for top, bottom, left and right. It should be typed in 8 point Calibri Font with page numbers at the bottom and centre of every page. It should be free from grammatical, spelling and punctuation errors and must be thoroughly edited.
7. **HEADINGS:** All the headings should be in a 10 point Calibri Font. These must be bold-faced, aligned left and fully capitalised. Leave a blank line before each heading.
8. **SUB-HEADINGS:** All the sub-headings should be in a 8 point Calibri Font. These must be bold-faced, aligned left and fully capitalised.
9. **MAIN TEXT:** The main text should follow the following sequence:

INTRODUCTION**REVIEW OF LITERATURE****NEED/IMPORTANCE OF THE STUDY****STATEMENT OF THE PROBLEM****OBJECTIVES****HYPOTHESES****RESEARCH METHODOLOGY****RESULTS & DISCUSSION****FINDINGS****RECOMMENDATIONS/SUGGESTIONS****CONCLUSIONS****SCOPE FOR FURTHER RESEARCH****ACKNOWLEDGMENTS****REFERENCES****APPENDIX/ANNEXURE**

It should be in a 8 point Calibri Font, single spaced and justified. The manuscript should preferably not exceed **5000 WORDS**.

10. **FIGURES & TABLES:** These should be simple, crystal clear, centered, separately numbered & self explained, and **titles must be above the table/figure. Sources of data should be mentioned below the table/figure.** It should be ensured that the tables/figures are referred to from the main text.
11. **EQUATIONS:** These should be consecutively numbered in parentheses, horizontally centered with equation number placed at the right.
12. **REFERENCES:** The list of all references should be alphabetically arranged. The author (s) should mention only the actually utilised references in the preparation of manuscript and they are supposed to follow **Harvard Style of Referencing**. The author (s) are supposed to follow the references as per the following:
 - All works cited in the text (including sources for tables and figures) should be listed alphabetically.
 - Use (ed.) for one editor, and (ed.s) for multiple editors.
 - When listing two or more works by one author, use --- (20xx), such as after Kohl (1997), use --- (2001), etc, in chronologically ascending order.
 - Indicate (opening and closing) page numbers for articles in journals and for chapters in books.
 - The title of books and journals should be in italics. Double quotation marks are used for titles of journal articles, book chapters, dissertations, reports, working papers, unpublished material, etc.
 - For titles in a language other than English, provide an English translation in parentheses.
 - The location of endnotes within the text should be indicated by superscript numbers.

PLEASE USE THE FOLLOWING FOR STYLE AND PUNCTUATION IN REFERENCES:**BOOKS**

- Bowersox, Donald J., Closs, David J., (1996), "Logistical Management." Tata McGraw, Hill, New Delhi.
- Hunker, H.L. and A.J. Wright (1963), "Factors of Industrial Location in Ohio" Ohio State University, Nigeria.

CONTRIBUTIONS TO BOOKS

- Sharma T., Kwatra, G. (2008) Effectiveness of Social Advertising: A Study of Selected Campaigns, Corporate Social Responsibility, Edited by David Crowther & Nicholas Capaldi, Ashgate Research Companion to Corporate Social Responsibility, Chapter 15, pp 287-303.

JOURNAL AND OTHER ARTICLES

- Schemenner, R.W., Huber, J.C. and Cook, R.L. (1987), "Geographic Differences and the Location of New Manufacturing Facilities," Journal of Urban Economics, Vol. 21, No. 1, pp. 83-104.

CONFERENCE PAPERS

- Garg, Sambhav (2011): "Business Ethics" Paper presented at the Annual International Conference for the All India Management Association, New Delhi, India, 19-22 June.

UNPUBLISHED DISSERTATIONS AND THESES

- Kumar S. (2011): "Customer Value: A Comparative Study of Rural and Urban Customers," Thesis, Kurukshetra University, Kurukshetra.

ONLINE RESOURCES

- Always indicate the date that the source was accessed, as online resources are frequently updated or removed.

WEBSITES

- Garg, Bhavet (2011): Towards a New Natural Gas Policy, Political Weekly, Viewed on January 01, 2012 <http://epw.in/user/viewabstract.jsp>

A TECHNIQUE WITH DIFFERENTIATED SAMPLING IN ANOMALY DETECTION SYSTEM FOR OUTLIER IDENTIFICATION

SARANYA.C

STUDENT

S.A ENGINEERING COLLEGE

VEERARAGHAVAPURAM

VEENA.S

ASSOCIATE PROFESSOR

S.A ENGINEERING COLLEGE

VEERARAGHAVAPURAM

ABSTRACT

Anomaly detection is the process of identifying unusual behavior and also a small group of instances that deviate remarkably from the existing data. The real world application of anomaly detection includes intrusion or credit card fraud detection that requires a most efficient framework for identifying the deviated data instances. The technique called principal component analysis (PCA) which require large amount of computation memory requirements and therefore it is not suitable for large scale data like online applications. Therefore a new technique called online oversampling principal component analysis (osPCA) algorithm along with online updating technique is used for detecting the existence of outliers from a large number of data. When oversampling a data instance the online updating technique enables the osPCA to update the principle direction effectively without solving the eigenvalue decomposition. The feasibility of osPCA provides more efficient and accurate results. The work extends by detecting outliers from high dimensional dataset using some clustering techniques with lesser time consumption.

KEYWORDS

Anomaly detection, online updating, oversampling principal component analysis, least square, KDD dataset.

1. INTRODUCTION

Anomaly detection also called as outlier detection is a process of searching of an item that does not conform to an expected pattern. These patterns that are detected are generally called as anomalies. These kinds of anomalies are translated to critical and actionable information in several application domains. Practically anomalies are patterns in data that do not match a well defined notion of normal behavior. The major task in data mining is the difficulty of detecting the outliers from a large set of data objects and this motivates multiple number of anomaly detection techniques [1], [2]. Anomaly detection finds its use in wide variety of applications that lead to illegal or abnormal behavior, such as credit card fraud detection, homeland security, and network intrusion detection, and insurance fraud, medical diagnosis, marketing, or customer segmentation, intrusion detection for cyber-security [4]. These real world applications contain only a limited amount of labeled data, how to detect anomaly of unseen events draws attention from the researchers in machine learning and data mining areas.

The existences of anomalies are often determined from the amount of deviated data. The presence of such kind of deviated data immensely affects the distribution of data. The calculation of least squares and data mean associated with the linear regression model is both sensitive to outliers. In that event, anomaly detection needs to solve an unsupervised yet unbalanced data learning problem. The most frequently cited sentence for the meaning of outliers is "one person's noise is another person's signal". Many approaches are framed for detecting outliers from a large scale of data as well as high dimensional data yet most of the known methods are least theoretically applicable to high dimensional data.

A well known framework called Principal component analysis (PCA) along with LOO (leave one out) strategy calculates effect of outlierness from the derived principal direction of the data set without the presence of target instance and that of the original data set. However in the real-world anomaly detection problems dealing with large amount of data, adding or removing a target data instance produces only a negligible difference in the resulting eigenvectors and it is not elementary to apply the (decremental PCA (dPCA) technique for detecting the anomalies [3], [5]. The LOO anomaly detection procedure with an oversampling strategy will strikingly increase the computational load as well as the storage capacity. Henceforth, the problem of detecting outliers in machine learning, data mining literature lose their algorithmic effectiveness for high dimensional data. Eventhough a well known power method is able to produce approximated PCA solutions; it cannot be extended easily to applications with streaming of data. For this reason an online updating technique along with osPCA is framed, which allows to efficiently calculate the approximated dominant eigen-vector without storing the entire covariance matrix. In order to increase the speed of the system a clustering technique called hierarchical clustering is used in accordance with osPCA online updating technique. Correlated to other anomaly detection methods like ABOD, PCA with LOO strategy, osPCA and power methods the proposed framework incomparably reduces the imperative computational costs and memory requirements. And thence the proposed method is eminently desirable in online, streaming data, or large-scale problems. The proposed clustering technique along with the online updating osPCA is computationally preferable to precedent anomaly detection methodologies.

2. RELATED WORKS

Previously many number of outlier detection techniques have been proposed. One among those is the well known PCA method. This PCA is a renowned unsupervised dimension reduction method to determine the principal directions of the data distribution. In LOO strategy the principle direction of the data set can be calculated without the presence of target instance and that of the original data set by adding or removing the abnormal (or normal) data instance. It is examined that removing (or adding) an abnormal data will produce a greater difference in the principal direction when compared to same operation performed with the normal data [3], [5]. In such a way, the outlierness can be discovered by the variation of the resulting principal directions.

Though PCA is a well known dimension reduction method, it is not suitable for real-world an anomaly detection problem which deals with a large amount of data. Therefore, adding or removing a target instance will create only a negligible difference in the resulting eigenvectors. The PCA method retains characteristics of the data set which contributes most of its variance by handling with the lower-order principal components. This method is sensitive to outliers, and the main data structure is represented with the help of few principal components [3]. In these techniques, the presence of an outlier data instance is determined by many processes, but one of those processes is to determine the corresponding principal direction of the data set. Thus, whenever an outlier data instance is introduced, accordingly the principal direction will create corresponding angle changes. That is, adding or removing an outlier data or a deviated data instance will cause larger effect on these principal directions.

Consequently, the variation of principle directions while removing or adding an instance is discovered. Accordingly, the first principal direction is greatly affected whenever an outlier or a deviated data instance is added or removed from the existing data set. In such a case the first principal direction is affected badly. Therefore, the initial principal direction is varied and that forms a large angle between itself and the old one. Therefore, in such kind of situations the initial

principal direction will not get affected and it only form an extremely smaller angle between the old initial principal direction and the new one if the normal data instance is removed from its original data set. By means of this observation the LOO procedure is used in each and every individual point by adding or removing (with or without) effect. The same procedure is used with the LOO technique but with the incremental strategy. This methodology enables adding a data instance to determine the possible variations of the principal directions from the above mentioned strategies it was found that the principal directions are significantly affected with the removal of an outlier data instance while this variation of the principal direction will be much smaller with the removal of a normal data instance. Conversely adding an outlier data instance will also cause significantly larger influence on the principal directions while the variation of the principal direction will be notably smaller with adding a normal data instance. However, it is not applicable for detecting the presence of outliers from a large scale data, since in case of a large scale data or streaming data removing a single outlier will not create such a difference in the resulting principal direction.

To address the above problem another technique called oversampling PCA (osPCA) is proposed [3]. This is because, the effect of "with or without" a specified data instance might be diminished when the size of the data is large. Thence to overcome the problem, this oversampling strategy employs duplicating the target instance on such an oversampled data, and this makes detecting of outliers to be much easier.

In the PCA scheme for anomaly detection, n PCA analysis for a data set with n data instances in a p-dimensional space has to be performed, despite it is computationally infeasible to compute. In the practical anomaly detection problems like real world anomaly detection problems this may not be so easy to discover and monitor the variation of the principal directions caused by the presence of a single outlier. Since in a real-world data set, the size of the data is typically large and thence detecting the presence of outlier from such a large amount of data is really a tedious task.

The power method for osPCA is a simple iterative algorithm which does not need to compute matrix decomposition. This method requires only the matrix multiplications and it does not bother about decomposition of matrix that is formed. Due to these reasons the computation cost can be mitigated in calculating the computational cost. But this method is computationally expensive and also it cannot be applied for large-scale anomaly detection techniques. In order to reduce the computational cost and the memory requirements of the anomaly detection system a new methodology called osPCA with online updating technique was proposed. Though the cost of computation is not so expensive as well it does not require large amount of memory the time taken to train the entire set of real world data is tremendous. This training delays the other activities of the system and therefore a new framework has to be proposed in order to reduce the system response time and the performance time. However the computational complexity and the memory requirements for the online osPCA method is very low when compared to other methods including osPCA power method.

It is undesirable that the above described methods are typically implemented in batch mode. Therefore it is not so elementary to perform the anomaly detection problems with online data (large-scale data) or streaming data.

3. PROPOSED WORK

With regards to the above mentioned methodologies some of the techniques for enforcing anomaly detection is mentioned below

A. DATA TRAINING

The data training set is a set of data used in different areas of information science. The main objective of training the data set is to conceive a potentially predicated relationship. While considering a real time KDD (knowledge discovery in databases) which consists of a large amount of data instances. This data training methodology includes training the undefined and unstructured data set in such a way that these data sets are arranged accordingly in the database. The KDD data set or any real time can be used for this purpose. The most dominant application which requires detection in online updating strategies is credit card fraud detection, fault detection, detection in cyber security. Many of the anomaly detection techniques need a set of completely pure normal data set to train the model. However, these kind of anomaly detection techniques implicitly assumes that these anomalies can also be treated as patterns that are not observed before.

By this reason, an outlier may be defined as the data point which is highly unique from the rest of the data instances, based on some predefined measure. And therefore, several outlier detection schemes are proposed in order to verify how efficiently the problem of anomaly detection can be pledged.

B. CLEANING DATA

The training data set can be selected based on the user assumption in case of handling a real-world data instances. The main goal of data cleaning phase is to identify the suspicious outliers. A set of data instances from the original data set after the data training process is taken as the predefined input. These data sets may be contaminated due to noise, incorrect data labeling etc., but they become an error free data due to training of data. Using LOO strategy the absolute value of cosine similarity is determined to measure the difference of the principle direction and obtain the suspicious outlier scores. When the suspicious outlier score is higher, it implies the higher probability of being an outlier data instance. The ranking for all the data instances can be made after calculating the suspicious outlier scores for each data instances. The over-sampling principal component analysis outlier detection algorithm is used for cleaning the data. The data cleaning process aims to filter out the most deviated data using osPCA. This process is done offline before performing the online anomaly detection process. The percentage of training normal data instance to be disregarded can be determined by the user. While performing the data cleaning process the smallest score of outlierness is used which of the remaining training data instances as the threshold for outlier detection. The trained data extracted from the data cleaning process is completely recycled by the online detection process in order to detect each arriving target instance.

C. CLUSTERING THE DATA

The process of training the data is done based on the user assumption and hence there may be a problem that even an outlier data can be considered as a normal data instance. This is because of training the data based on the assumption made by the user.

Clustering method is the best solution for the above mentioned problem. The type of clustering used here is the hierarchical clustering. This kind of clustering technique creates an easily understandable hierarchical cluster model for the data instance. Thereby it tend to increase the speed of the anomaly detecting system and also the amount of time consumption of the remaining processes tend to get reduced accordingly. The clusters are usually formed for the input data instances, where the outlier calculation is applied to each cluster for detecting the exact outlier data instance. The score of outlierness is updated accordingly from the process of data cleaning.

D. ONLINE DETECTION

As soon as the data cleaning process is completed, the suspicious points are filtered. As a result a pure normal data is obtained to which the online anomaly technique is to be applied. The main objective of the online anomaly detection method is to identify the newly arriving abnormal data instance. Similarly the oversampling process is also applied for the newly arriving instance. The idea behind this method is to determine the mean and standard scores that are computed from all normal data points. Once when the mean and the standard deviation is calculated, a newly arriving data instances will be marked as an outlier if it's suspicious score than the previously calculated values.

For a KDD dataset there are four categories of attacks which are to be considered as an outlier data instance.

- DoS (Denial of Service), example, ping-of-death, teardrop, etc.,
- U2R, it is the unauthorized access to local super user privileges by a local privileged user, example, and various buffer overflow attacks.
- PROBING, surveillance and probing, example, port-scan, ping-sweep, etc.,
- R2L, unauthorized access from a remote machine, example, guessing password.

FIG. 1 ARCHITECTURE

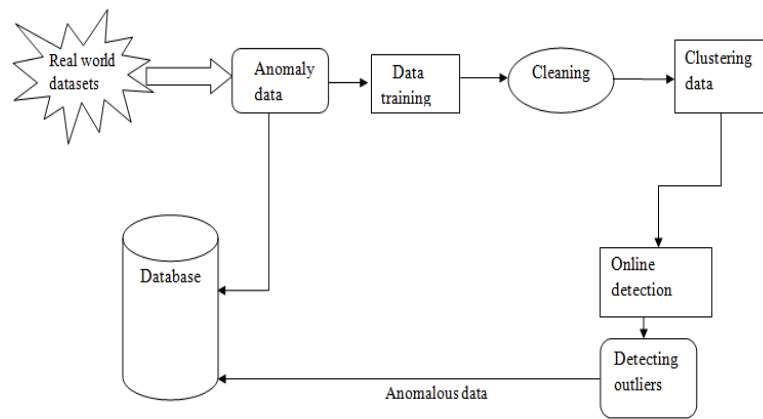


Fig.1 depicts the architecture of the proposed anomaly detection system. The input is the KDD data set which is obtained from the KDD website. It consists of thousands of data sets including both the normal data instance and the outlier data instances. This KDD data set is considered as the anomaly data set and is processed based on the proposed methodology.

Any real world data sets excluding KDD data set can also be given as the input anomaly data. These data sets are trained in the data training process. After storing these data sets into the database it is cleaned in the data cleaning process. These processes include filtering of the most deviated or outlier data. The cleaning block includes extracting the normal data from the contaminated data set. The data cleaning process calculates the score of outlierness of each receiving test input data instance. The data cleaning determines the threshold and report the correct and incorrect rates over the receiving data instances.

The clustering of data instance includes priority wise or range wise clustering of data which makes searching and cleaning of data faster which in turn makes the implementation becomes much easier. The result of clustering or the data cleaning is fed as input to the online detection process. It aims at detecting the newly arriving abnormal instance.

As a result of the online anomaly detection process, the possible suspicious data instances, say abnormal data instances are detected and these sets of data are stored separately stored in the database and therefore whenever a new data set is taken for processing the suspicious data sets can be detected easily and ignored in order to obtain the system security. Therefore any unauthorized access from a remote machine, unauthorized access to a local super user machine or any probing attacks can be easily detected. Eventhough these kinds of data instances resembles same as the normal data instances with a background noise that characterize as a normal data. The attacked data files can be listed out as a suspicious data instances by comparing its features with the some predefined list files. Many of the connections are being used by the DoS and probing attacks and therefore the proposed framework must able to capture even those kinds of data that represents the same characteristics. However the connection feature of the data content was the important characteristics for detecting R2L (the unauthorized access from a remote machine) and U2R (the unauthorized access to a local super user privileges) attack types, while the time-based and the connection-based features were the most important features for detecting and determining the DoS (denial-of-service) attack types.

E. OUTLIER DETECTION

Since it impossible for a human analyst to investigate practically all the outlier detection scenarios, the clustering technique along with the online osPCA is proposed which involves faster detection of outlier data instances and storing it in a separate field on the database. So, whenever a real time data like credit card data has to be detected through online then the previously investigated results can be used in order to compare it with the newly arrived data and henceforth the suspicious data instances can be detected and blocked. These kinds of anomaly detection techniques can be most probably used in many application domains which consist of real world data sets. The online osPCA with clustering technique achieves slightly a larger false positive rate than the other anomaly detection methods. It also achieves a comparable performance significantly less computation time along with less memory requirement; this is because the proposed framework does not require storing the entire data matrix during the online detection process.

As soon as the anomalous data sets are detected they are separately stored in the database and whenever a newly arriving data instance that possess the same characteristics of the already detected anomalous data then the system is given an alert message or an acknowledgement that the newly arrived data is the suspicious data, this helps the system user to protect the original data sets from the attacker sets of data and also ignore these kinds of data automatically. Thus the proposed work involves obtaining much faster and more efficient results when compared to other methods.

4. RESULTS

Initially the input to the system is given as a text file which is then trained by the anomaly detection system in order to convert it into the datafields which are then stored onto the database for further calculations and proceedings. The results generated for the above methods is given as below:

The screenshot shows a software application window titled "TRAIN KDD DATA SET". At the top right, there is a "NEXT" button. Below it, the text "Browse File" is followed by a colon and a file path: "E:\saranyal3rd semDM\dataset\Copy of set.txt". There are two buttons, "Browse" and "Train", below the file path. The main area of the window is divided into two panes. The left pane contains a list of network traffic records, each with a long string of binary digits (0s and 1s) followed by protocol and port information. The right pane contains a table with columns for protocol, port, and source/destination IP addresses. A "Message" dialog box is overlaid in the center, displaying a warning icon and the text "KDD trained successfully" with an "OK" button.

5. CONCLUSION

An anomaly detection methodology for detecting intrusions and suspicious data sets are proposed in this paper. To support the applicability of anomaly detection schemes, several clustering techniques are used accordingly along with the online oSPCA methodology by determining the outlier scores for each data instances. This outlier scores are framed in order to rank the outliers in an efficient manner. Moreover the proposed method does not require storing the entire data matrix during the online detection process. Therefore, when compared to other anomaly detection methodologies the proposed framework produces a better result. On the other hand, it can be used for detecting anomalies in large scale data including online data stream or any unbalanced data distribution (includes network security problems). The future work can be extended in order to have a high detection rate and also to increase the speed of the anomaly detection system which automatically deletes the anomalous data from the original data set.

REFERENCES

1. M. Breunig, H-P. Kriegel, R.T. Ng, and J. Sander, "LOF: Identifying Density-Based Local Outliers," Proc. ACM SIGMOD Int'l Conf. Management of Data, 2000.
2. Management of Data, 2000.H-P. Kriegel, M. Schubert, and A. Zimek, "Angle-Based Outlier Detection in High-Dimensional Data," Proc. 14th ACM SIGKDD Int'l Conf. Knowledge Discovery and data mining, 2008.
3. Y-R Yeh, Z-Y. Lee, and Y-J. Lee, "Anomaly Detection via Oversampling principal component Analysis," Proc. First KES Int'l Symp. Intelligent Decision Technologies, pp. 449-458, 2009.
4. W. Wang, X. Guan, and X. Zhang, "A Novel intrusion detection method based on principal component analysis in computer security," Proc. Int'l Symp. Neural Networks, 2004.
5. Lee, Yuh-Jye; Yeh, Yi-Ren; Wang, Yu-Chiang Frank, "Anomaly Detection via Online Oversampling Principal Component Analysis," Knowledge and Data Engineering, IEEE Transactions on , vol.25, no.7, pp.1460,1470, July 2013.
6. Pokrajac, A. Lazarevic, and L. Latecki, "Incremental Local Outlier Detection for data streams," Proc. IEEE Symp. Computational Intelligence and Data Mining, 2007.

REQUEST FOR FEEDBACK

Dear Readers

At the very outset, International Journal of Research in Commerce, IT & Management (IJRCM) acknowledges & appreciates your efforts in showing interest in our present issue under your kind perusal.

I would like to request you to supply your critical comments and suggestions about the material published in this issue as well as on the journal as a whole, on our E-mail infoijrcm@gmail.com for further improvements in the interest of research.

If you have any queries please feel free to contact us on our E-mail infoijrcm@gmail.com.

I am sure that your feedback and deliberations would make future issues better – a result of our joint effort.

Looking forward an appropriate consideration.

With sincere regards

Thanking you profoundly

Academically yours

Sd/-

Co-ordinator

DISCLAIMER

The information and opinions presented in the Journal reflect the views of the authors and not of the Journal or its Editorial Board or the Publishers/Editors. Publication does not constitute endorsement by the journal. Neither the Journal nor its publishers/Editors/Editorial Board nor anyone else involved in creating, producing or delivering the journal or the materials contained therein, assumes any liability or responsibility for the accuracy, completeness, or usefulness of any information provided in the journal, nor shall they be liable for any direct, indirect, incidental, special, consequential or punitive damages arising out of the use of information/material contained in the journal. The journal, nor its publishers/Editors/Editorial Board, nor any other party involved in the preparation of material contained in the journal represents or warrants that the information contained herein is in every respect accurate or complete, and they are not responsible for any errors or omissions or for the results obtained from the use of such material. Readers are encouraged to confirm the information contained herein with other sources. The responsibility of the contents and the opinions expressed in this journal is exclusively of the author (s) concerned.

ABOUT THE JOURNAL

In this age of Commerce, Economics, Computer, I.T. & Management and cut throat competition, a group of intellectuals felt the need to have some platform, where young and budding managers and academicians could express their views and discuss the problems among their peers. This journal was conceived with this noble intention in view. This journal has been introduced to give an opportunity for expressing refined and innovative ideas in this field. It is our humble endeavour to provide a springboard to the upcoming specialists and give a chance to know about the latest in the sphere of research and knowledge. We have taken a small step and we hope that with the active co-operation of like-minded scholars, we shall be able to serve the society with our humble efforts.

Our Other Journals

