

INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATION & 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 2840 Cities in 164 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.	THE LEADERSHIP PRACTICES OF COMBINED ARMY ACADEMY'S DEAN <i>MATEBE TAFERE</i>	1
2.	ELECTRONIC GROCERY SHOPPING: MODELS AND METHODS FOR THE URBAN CONSUMER DELIGHT <i>AMOL RANADIVE & DR. HRUDANAND MISHRA</i>	6
3.	STUDY ON IMPLEMENTING ASSOCIATION RULE MINING IN PARTICLE SWARM OPTIMIZATION <i>T. BHARATHI & DR. P. KRISHNAKUMARI</i>	10
4.	KEY FACTORS TO DEVELOP WOMEN ENTREPRENEURS IN NELLORE (DT), ANDHRA PRADESH <i>A.M.MAHABOOB BASHA, P.SRI SUDHA & V.MADHAVI</i>	18
5.	LAND USE AND LAND COVER DETECTION FOR THREE DECADES USING GIS AND RS -A CASE STUDY OF ERODE DISTRICT <i>C. LALITHA & DR. S. P. RAJAGOPALAN</i>	21
6.	APPRAISAL OF LIQUIDITY PERFORMANCE IN LANCO INDUSTRIES LIMITED: A CASE STUDY <i>N. K. PRADEEP KUMAR & P. MOHAN REDDY</i>	25
7.	ORGANIZATIONAL CULTURE AS A DETERMINANT OF CUSTOMER SERVICE DELIVERY IN LOCAL AUTHORITIES IN KENYA <i>ROBERT K.W. EGESSA, PETER KIBAS & THOMAS CHERUIYOT</i>	30
8.	EMPLOYEE JOB SATISFACTION: A CASE STUDY ON ONGC <i>DR. MEGHA SHARMA</i>	35
9.	SUPPLY CHAIN MANAGEMENT: A STUDY OF PADDY IN ANDHRA PRADESH <i>DR. I. SAI PRASAD</i>	39
10.	PERFORMANCE APPRAISAL PROCESS AT ANDHRA PRADESH STATE ROAD TRANSPORT CORPORATION (APSRTC) <i>RAKHEE MAIRAL RENAPURKAR & DR. SUDHAKAR B INGLE</i>	44
11.	DETECTION OF BRAIN TUMOR USING THRESHOLDING AND MORPHOLOGICAL OPERATIONS <i>SHRIJA MADHU & T.M.SIRISHA</i>	51
12.	ANTECEDENTS OF CUSTOMER RELATIONSHIP MANAGEMENT AND ITS IMPACT ON CUSTOMER LOYALTY IN BANKING SECTOR <i>V.KRISHNAMOORTHY & DR. R. SRINIVASAN</i>	54
13.	ASSESSMENT OF CUSTOMERS' SERVICE EXPECTATIONS AND PERCEPTIONS IN GEM HOSPITAL: GAPS MODEL <i>V. KANIMOZHI & DR. R. ANITHA</i>	60
14.	IMPACT OF CLOUD COMPUTING ON INDIAN SMEs: ADOPTION, BENEFITS AND FUTURE SCOPE <i>NAZIR AHMAD & JAMSHED SIDDIQUI</i>	64
15.	A STUDY ON THE EFFECTIVENESS OF TRAINING AND DEVELOPMENT PRACTICES ON THE BASIS OF LEVEL OF TRUST, COMMUNICATION AND MORALE OF EMPLOYEES AT LIBERTY SHOES LIMITED <i>DR. VANDANA KHETARPAL & REETI ATREJA</i>	67
16.	A SURVEY OF THE DIMENSIONALITY REDUCTION TECHNIQUES IN DATA MINING: A REVIEW PAPER <i>TARANMEET KOUR, AMITPREET KOUR & DR. SANDEEP SHARMA</i>	73
17.	AN IMPERATIVE STUDY ABOUT HUMAN COMPUTER INTERACTION: TRENDS AND TECHNOLOGIES <i>DR. ASHU GUPTA & SAKSHI DUA</i>	76
18.	A REVIEW ON THE COST MANAGEMENT STRATEGIES ADOPTED BY AIRLINES GLOBALLY <i>DR. BINDU NAIR</i>	81
19.	APPLICATION OF ARTIFICIAL BEE COLONY ALGORITHM TO INDEPENDENT COMPONENT ANALYSIS <i>AMRESH KUMAR SINGH</i>	84
20.	ACTIVITY BASED COSTING & TRADITIONAL COST ACCOUNTING SYSTEM: A COMPARATIVE STUDY OF OVERHEAD COST ALLOCATION <i>MONIKA KHEMANI</i>	93
21.	E-MARKETING: CHALLENGES AND OPPORTUNITIES <i>RUCHIKA NACHAAL</i>	97
22.	PERFORMANCE EVALUATION OF TURKISH PENSION FUNDS BY USING ELECTRE METHOD <i>HASAN UYGURTÜRK</i>	100
23.	FROM CHANGE MANAGEMENT TO CHANGE READINESS: KEYS TO SUCCESSFULLY IMPLEMENTING CHANGE <i>AJIT KUMAR KAR & LOPAMUDRA PRAHARAJ</i>	108
24.	A STUDY TO MAXIMIZE INTERPERSONAL EFFECTIVENESS TO OVERCOME GENERATION GAP USING AURA AS A TOOL <i>V. VAIDEHIPRIYAL & DR. N. RAMKUMAR</i>	113
25.	APPLICATION OF ROLE OF PROFESSIONAL MARKETING MANAGERS IN A DYNAMIC BUSINESS ENVIRONMENT <i>DR. ABDULSALAM JIBRIL & DR. MUHAMMAD ISA BAZZA</i>	118
26.	ANALYSIS OF CORPORATE SOCIAL DISCLOSURE PRACTICES IN ANNUAL REPORTS: AN EXPERIENCE WITH THE PRIVATE COMMERCIAL BANKING SECTOR OF BANGLADESH <i>SHARMIN SHABNAM RAHMAN</i>	122
27.	M-LEARNING CONTEXTS COUPLED WITH CONNOTATION OF 4G CONNECTIVITY <i>B.AYSHWARYA & M.DHANAMALAR</i>	130
28.	IMPORTANCE OF OPEN ACCESS IN FLOW OF INFORMATION: WITH SPECIAL EMPHASIS ON RESEARCH <i>A. SIVA KESAVULU & B.DEENADHAYALU</i>	133
29.	VIRTUAL LEARNING ENVIRONMENT: ISSUES AND SUGGESTIONS <i>SUNIL KUMAR SHARMA</i>	136
30.	THE IMPACT OF INTEREST RATES ON THE PERFORMANCE OF BANKS: A CASE STUDY OF CANARA BANK AND HDFC BANK <i>MANASA ELURU, SAHLE YEIBIYO ASGHEDE & SHIFERAW MITIKU TEBEKA</i>	139
	REQUEST FOR FEEDBACK	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

DR. SAMBHAV GARG

Faculty, Shree Ram Institute of Business & Management, Urjani

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. S. L. MAHANDRU

Principal (Retd.), Maharaja Agrasen College, Jagadhri

EDITOR

PROF. R. K. SHARMA

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

EDITORIAL ADVISORY BOARD

DR. RAJESH MODI

Faculty, Yanbu Industrial College, Kingdom of Saudi Arabia

PROF. PARVEEN KUMAR

Director, M.C.A., Meerut Institute of Engineering & Technology, Meerut, U. P.

PROF. H. R. SHARMA

Director, Chhatrapati Shivaji Institute of Technology, Durg, C.G.

PROF. MANOHAR LAL

Director & Chairman, School of Information & Computer Sciences, I.G.N.O.U., New Delhi

PROF. ANIL K. SAINI

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

PROF. R. K. CHOUDHARY

Director, Asia Pacific Institute of Information Technology, Panipat

DR. ASHWANI KUSH

Head, Computer Science, University College, Kurukshetra University, Kurukshetra

DR. BHARAT BHUSHAN

Head, Department of Computer Science & Applications, Guru Nanak Khalsa College, Yamunanagar

DR. VIJAYPAL SINGH DHAKA

Dean (Academics), Rajasthan Institute of Engineering & Technology, Jaipur

DR. SAMBHAVNA

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

DR. MOHINDER CHAND

Associate Professor, Kurukshetra University, Kurukshetra

DR. MOHENDER KUMAR GUPTA

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

DR. SAMBHAV GARG

Faculty, Shree Ram Institute of Business & Management, Urjani

DR. SHIVAKUMAR DEENE

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

DR. BHAVET

Faculty, Shree Ram Institute of Business & Management, Urjani

ASSOCIATE EDITORS

PROF. ABHAY BANSAL

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

PROF. NAWAB ALI KHAN

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

ASHISH CHOPRA

Sr. Lecturer, Doon Valley Institute of Engineering & Technology, Karnal

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 SURVEY OF THE DIMENSIONALITY REDUCTION TECHNIQUES IN DATA MINING: A REVIEW PAPER

TARANMEET KOUR
STUDENT
DEPARTMENT OF COMPUTER SCIENCE
GURU NANAK DEV UNIVERSITY
AMRITSAR

AMITPREET KOUR
STUDENT
DEPARTMENT OF COMPUTER SCIENCE
GURU NANAK DEV UNIVERSITY
AMRITSAR

DR. SANDEEP SHARMA
ASSOCIATE PROFESSOR
GURU NANAK DEV UNIVERSITY
AMRITSAR

ABSTRACT

The objective of data mining is to identify valid novel, potentially useful, and understandable correlations and patterns in existing data [Chung and Gray 1999]. The volume of electronically accessible data in warehouses and on the Internet is growing faster than the speedup in processing times predicted by Moore's Law. Consequently, classical data mining algorithms that require one or more computationally intensive passes over the entire database are becoming prohibitively slow, and this problem will only become worse in the future. The scalability of mining algorithms has therefore become a major research topic. One approach to the scalability problem is to run mining algorithms on a small subset of data. That is Data Reduction is a very important issue. Data reduction techniques can be applied to obtain a reduced representation of the data set that is much smaller in volume, yet closely maintains the integrity of the original data. That is, mining on the reduced data set should be more efficient yet produce the same analytical results. In this paper mainly dimensionality reduction technique to reduce data is been discussed. Under it mainly two techniques i.e. SVD and PCA are been discussed.

KEYWORDS

Data Mining, eigen values, eigen vectors, PCA, SVD.

I. INTRODUCTION

The sheer volume of data today and its expected increase in the next few days are some of the key challenges in data mining and knowledge discovery applications[1]. Besides a huge number of data samples that are collected and processed, the high dimensional nature of data arising in many applications causes a need to develop effective and efficient techniques that are able to deal with massive amount of data[9]. As the dimensionality of data increases, many types of data analysis and classification problems become significantly harder. Dimensionality reduction, feature selection, numerosity reduction, and data compression are some of the techniques to reduce the attribute space.

II. DATA REDUCTION TECHNIQUES

Dimensionality reduction is the process of reducing the number of random variables or attributes under consideration by producing the linear combination of the original data. Dimensionality reduction methods include transforms and principal components analysis which transform or project the original data onto a smaller space. The main advantage of this technique is that the size of attribute space is reduced without losing a lot of information of the original attribute space. The disadvantage is the fact that the linear combinations of the original features are usually not interpretable and the information about how much an original attribute contributes is often lost.

Feature Selection works by removing redundant or irrelevant features from the data set as they can lead to a reduction of the classification accuracy or clustering quality and to unnecessary increase in the computational cost. The advantage of feature selection is that no information about the importance of single feature is lost. On the other hand, if a small set of features is required and the original features are diverse, information may be lost as some of the features must be omitted.

Numerosity reduction techniques replace the original data volume by alternative smaller forms of data representation.

Data compression, transformations are applied so as to obtain a reduced or "compressed" representation of original data.

In this paper we will mainly deal with dimensionality reduction technique of data reduction. Data reduction can be lossless or lossy. In dimensionality reduction, reduction of the dimensionality is done by using low rank approximation techniques to create new attributes that are combinations of the old, original variables. Dimensionality reduction is also referred to as feature transformation, feature extraction or feature construction. Some of the techniques covered under this are Principal Component Analysis(PCA), Singular Value Decomposition(SVD), Pseudoinverse(also called as Moore Penrose Pseudoinverse), etc.

III. PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA)[2] is a standard tool in modern data analysis - in diverse fields from neuroscience to computer graphics - because it is a simple, non-parametric method for extracting relevant information from confusing data sets.

Often, the variables under study are highly correlated and as such they are effectively "saying the same thing". It may be useful to transform the original set of variables to a new set of uncorrelated variables called principal components. These new variables are linear combinations of original variables and are derived in decreasing order of importance so that the first principal component accounts for as much as possible of the variation in the original data. These new attributes (called principle components(PCs)) have to meet the following criteria: The PCs are (i) linear combinations of the original attributes, (ii) orthogonal to each other, and (iii) capture the maximum amount of variation in the data. Also, PCA is a linear dimensionality reduction technique, which identifies orthogonal directions of maximum variance in the original data, and projects the data into a lower-dimensionality space formed of a sub set of the highest variance components.

The goal of PCA is to reduce the dimensionality of the data while retaining as much as possible of the variation present in the original dataset.

A. Advantages and disadvantages of PCA

PCA is a relatively simple, non-parametric, generic method that is useful for finding new, more informative, uncorrelated features and it can be used to reduce dimensionality by rejecting low variance features. Since the principal components are orthogonal[7][8] to each other, every principal component is uncorrelated to every other principal component (i. e., they do not contain any redundant information). The principal components are designed to account for the highest percentage of the variation among the variables with as few PCs as possible. Thus, often the first few PCs account for some large percentage of the total variance, allowing for a compact representation of the data with only low dimensions.

However, PCA is limited to re-expressing the data as combinations of its basis vectors. A main drawback of PCA[5] is that each PC is a linear combination of all original variables, thus leading to a potentially difficult interpretation of the PCs. On the contrary, in a system with many variables PCA may be used to project the dimension down to a reasonable number of plots, and the principal components could be rotated towards a more meaningful representation. Moreover, PCA[6] is sensitive with respect to the units of measurement. If the units and the variances of attributes vary a lot, then variables with high variance tend to dominate the first few principal components. In this case, the data need to be normalized prior to the PCA transformation.

B. Computational complexity

From a computational point of view, the eigen value decomposition for solving the PCA transformation is rather expensive in terms of runtime, especially for a large number of attributes. There are several algorithms for solving symmetric eigen problems, but all of them are of order $O(n^3)$. So the complexity of solving PCA also revolves around the same.

IV. SINGULAR VALUE DECOMPOSITION

Singular Value Decomposition[3] can be viewed from three points of view:

- 1.It is a method for transforming correlated variables into a set of uncorrelated ones that better expose the various relationships among the original data items.
- 2.SVD is a method for identifying and ordering the dimensions along which data points exhibit the most variation.
- 3.Once we have identified where the most variation is, it is possible to find the best approximation of the original data points using fewer dimensions. Hence SVD can be seen as a method of data reduction.

SVD[4] is based on a theorem of linear algebra which says that a rectangular matrix can be broken down into a product of three matrices:

- a) An orthogonal matrix U [i.e. $U^T U = I$] and the columns of U are orthonormal eigen vectors of $A A^T$.
- b) A diagonal matrix S, containing the square roots of eigen values from U or V in descending order.
- c) Transpose of an orthogonal matrix V^T [i.e. $V^T V = I$] and the columns of V are orthonormal eigen vectors of $A^T A$.

i.e. $A_{mn} = U_{mn} S_{mn} V_{nn}^T$

The truncated SVD (or reduced rank SVD) to A can be found by setting all but the first k largest singular values equal to zero and using only the first k columns of U and V . This is usually denoted like

$A_k = U_k S_k V_k^T$

or more explicitly,

$$A_k \approx (u_1, \dots, u_k) \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_k & \\ & & & \ddots \end{pmatrix} \begin{pmatrix} V_1 \\ \vdots \\ V_k \end{pmatrix}$$

A. Computational complexity

In general, the computational complexity of the SVD algorithm for computing all three matrices U, S and V is $O(4m^2n + 8mn^2 + 9n^3)$, for computing only the matrices S and V it is $O(4mn^2 + 8n^3)$, making the decomposition unfeasible for large datasets. The RSVD algorithm has a complexity of $O(4m2n+22n3)$ for computing all three matrices, and a complexity of $O(2mn^2 + 11n^3)$ for computing only S and V.

Other than the data reduction the various other applications of SVD are:

- 1.Numerical Rank: The SVD can be used as a numerically reliable estimate of the effective rank of the matrix. The rank of a matrix is defined as the number of linearly independent columns in a matrix. The way to find these dependencies is to focus on the singular values that are of a larger magnitude than the measurement error. If there are r such singular values, the effective rank of the matrix is found to be r: The SVD is closely related to the idea of selecting the closest rank r approximation to a matrix, which is further considered in the data compression.
- 2.Pseudoinverse: Another application of the SVD is to the computation of what is called the generalized inverse of a matrix. This is very closely related to the linear least squares problem. If A is square and full-rank it has an inverse. But for the tall and fat matrices i.e. when $m > n$ and $m < n$ respectively the inverse cannot be taken out directly. In such cases the pseudoinverse is been taken out. Various methods for calculating the pseudoinverse have been found. The pseudoinverse of a sparse matrix is been found in [12]. Similarly the pseudoinverse can also be taken out by Linear Solver[11] which takes out the inverse by deviding the matrix into two parts.
- 3.Rank Deficient Least Squares: As with most of our other linear algebra tools, SVD provides yet another way to solve linear systems. This is stable when the we deal with square matrices One of the strengths of the SVD is that it works when the matrix is singular. It also deals with the overdetermined and underdetermined cases when the number of equations is not equal to the number of unknowns.

B. Relation to PCA

Calculating the SVD is equivalent of finding the eigen values and eigen vectors of AA^T and $A^T A$. The eigenvectors of AA^T make up the columns of U, the eigenvectors of $A^T A$ make up the columns of V. Moreover, the singular values in S are the square roots of eigen values from AA^T and $A^T A$. The SVD-PCA connection stems from the straightforward linear algebra calculations

$AA^T = (USV^T)(VS^T U^T) = US^2 U^T$, and

$A^T A = (VS^T U^T)(USV^T) = VS^2 V^T$

An often mentioned difference between these two approaches is that PCA removes

the mean of each variable whereas SVD uses the original data (however, the mean could also be removed before computing the SVD). Especially for sparse data it is not always preferable to remove the mean of the data.

V. METHODS

There are various methods to solve matrix equations that are specialised to particular problems. We can select between these based on the type of data. Various methods used for the same are as follows:

- 1.LAPACK: It is the default method for solving dense numerical matrices. When the matrix is square and non-singular the routines dgesv, dlange, and dgecon are used for real matrices and zgesv, zlange, and zgecon for complex matrices. When the matrix is non-square or singular dgelss is used for real matrices and zgelss for complex matrices. If the input matrix uses arbitrary-precision numbers, then LAPACK algorithms extended for arbitrary-precision computation are used.
- 2.Multifrontal: The Multifrontal method is a direct solver used by default if the input matrix is sparse. If the input matrix to the Multifrontal method is dense, it is converted to a sparse matrix. The implementation of the Multifrontal method uses the UMFPACK library.

3. *Krylov*: The Krylov method is an iterative solver that is suitable for large sparse linear systems, such as those arising from numerical solving of PDEs. Basically two Krylov methods are implemented: Conjugate Gradient (for symmetric positive definite matrices) and BiCGSTAB (for non-symmetric systems). The default method for Krylov, BiCGSTAB, is more expensive but more generally applicable. The ConjugateGradient method is suitable for symmetric positive definite systems, always converging to a solution (though the convergence may be slow). If the matrix is not symmetric positive definite the ConjugateGradient may not converge to a solution.

4. *Cholesky*: The Cholesky method is suitable for solving symmetric positive definite systems. This method is far more fast and stable. For dense matrices the Cholesky method uses LAPACK functions such as dpotrf and dpotrs for real matrices and zpotrf and zpotrs for complex matrices. For sparse matrices the Cholesky method uses the TAUCS library.

5. *Symbolic Methods*: There are a number of methods that are specific to symbolic and exact computation: CofactorExpansion, DivisionFreeRowReduction and OneStepRowReduction.

VI. CONCLUSIONS AND FUTURE WORK

In this paper we have discussed about data reduction techniques and from them we conclude various advantages of data reduction. It leads to better understanding of underlying model. It is simply the usage of various visualization techniques(eg looking at peers or triplets of attributes, self organizing maps, etc.). Data reduction significantly reduces the computational cost and memory requirements of the classification algorithm applied to data. Here mainly the dimensionality reduction aspect of data reduction is considered. SVD and PCA can be mainly used for dimensionality reduction. While PCA mainly works with principal components SVD deals with whole data and mainly with sparse matrix. From a theoretical point of view, PCA and SVD have the best theoretically provable properties, and give the best approximation in terms of approximation error. Moreover, both techniques are purely deterministic and always produce reproducible results (for identical input).

The availability of fast and efficient algorithms is also an important aspect of dimensionality reduction methods. The software libraries as described in the method part provide several algorithmic variants for efficiently computing singular value decomposition and eigen value decomposition(to be used in PCA).

The future work for these techniques can be that various applications of these techniques can be explored and the accuracy of data reduction can be increased.

REFERENCES

1. A. Beck and M. Teboulle, "A Fast Iterative Shrinkage Thresholding Algorithm for Linear Inverse Problems", *SIAM J. Imaging Sci.*, vol.2, no. 1, pp. 183-202, 2009.
2. Alter O., Brown P.O., Botstein D. Singular value decomposition for genome-wide expression data processing and modeling. *Proc Natl Acad Sci USA* 2000; 97:10101-06.
3. Fayyad, U., Piatetsky-Shapiro, G., and Smyth "From Data Mining to Knowledge Discovery: An Overview" *Advances in Knowledge Discovery and Data Mining*, MIT Press, 1-36, Cambridge, 1996.
4. Jolliffe, I. T. (2002). *Principal component analysis* (Second ed.): Springer.
5. Larose, D.T., *An Introduction to Data Mining. Discovering Knowledge in Data*. 2005: Willi & Sons, Inc.
6. M.A. Carreira-Perpinan. *A review of dimension reduction techniques*. Technical report CS-96-09, Department of CSE, University of Sheffield, 1997
7. Mark Richardson tutorial on "Principal Component Analysis", 2009.
8. Michael E. Wall "Singular value decomposition and principal component analysis", 2003.
9. Principal components analysis for correlated curves and seasonal commodities: The case of the Petroleum markets (Tolmasky & Hindanov, 2002).
10. Shlens, J., *A Tutorial on Principal Component Analysis*. 2009.
11. Smith, L., *A Tutorial on Principal Component Analysis*. 2002.

REQUEST FOR FEEDBACK

Dear Readers

At the very outset, International Journal of Research in Computer Application and 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

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

