INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATION & MANAGEMENT

IJRCM

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-Gage, India [link of the same is duly available at Inflibnet of University Grants Commission (U.G.C.)],

Index Copernicus Publishers Panel, Polandwith 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 2401 Cities in 155 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

CHIEF PATRON

PROF. K. K. AGGARWAL
Chancellor, Lingaya's University, Delhi
Founder Vice-Chancellor, GuruGobindSinghIndraprasthaUniversity, Delhi
Ex. Pro Vice-Chancellor, GuruJambheshwarUniversity, 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.), MaharajaAgrasenCollege, Jagadhri

EDITOR

PROF. R. K. SHARMA
Professor, Bharti Vidyapeeth University Institute of Management & Research, New Delhi

EDITORIAL ADVISORY BOARD

DR. RAJESH MODI
Faculty, YanbuIndustrialCollege, Kingdom of Saudi Arabia
PROF. PARVEEN KUMAR
Director, M.C.A., Meerut Institute of Engineering & Technology, Meerut, U. P.
PROF. H. R. SHARMA
Director, Chhatarpati 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), GuruGobindSinghI. P. University, Delhi
PROF. R. K. CHOUDHARY
Director, Asia Pacific Institute of Information Technology, Panipat
DR. ASHWANI KUSH
Head, Computer Science, UniversityCollege, KurukshetraUniversity, Kurukshetra

DR. BHARAT BHUSHAN
Head, Department of Computer Science & Applications, GuruNanakKhalsaCollege, 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, KurukshetraUniversity, Kurukshetra
DR. MOHENDER KUMAR GUPTA
Associate Professor, P.J.L.N.GovernmentCollege, 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, AligarhMuslimUniversity, 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

Weinvite unpublished novel, original, empirical and high quality research work pertaining to recent developments & practices in the area of Computer, Business, Finance, Marketing, Human Resource Management, General Management, Banking, Education, Insurance, Corporate Governance and emerging paradigms in allied subjects like Accounting Education; 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; Monetary Policy; Portfolio & Security Analysis; Public Policy Economics; Real Estate; Regional Economics; Tax Accounting; Advertising & Promotion Management; Business Education; Management Information Systems (MIS); Business Law, Public Responsibility & Ethics; Communication; Direct Marketing; E-Commerce; Global Business; Health Care Administration; Labor 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; Public Administration; Purchasing/Materials Management; Retailing; Sales/Selling; Services; Small Business Entrepreneurship; Strategic Management Policy; Technology/Innovation; Tourism, Hospitality & Leisure; Transportation/Physical Distribution; Algorithms; Artificial Intelligence; Compilers & Translation; Computer Aided Design (CAD); Computer Aided Manufacturing; Computer Graphics; Computer Organization & Architecture; Database Structures & Systems; Digital Logic; Discrete Structures; Internet; Management Information Systems; Modeling & Simulation; Multimedia; Neural Systems/Neural Networks; Numerical Analysis/Scientific Computing; Object Oriented Programming; Operating Systems; Programming Languages; Robotics; Symbolic & Formal Logic and Web Design. The above mentioned tracks are only indicative, and not exhaustive.

Anybody can submit the soft copy of his/her manuscript **anytime** in M.S. Word format after preparing the same as per our submission guidelines duly available on our website under the heading guidelines for submission, at the email address: infoijrcm@gmail.com.

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 mentionthe 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, KurukshetraUniversity, 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

DOCUMENT CLUSTERING BASED ON CORRELATION PRESERVING INDEXING IN SIMILARITY MEASURE SPACE

D. JENCY
INSTRUCTOR
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
SRI MUTHUKUMARAN INSTITUTE OF TECHNOLOGY
CHIKKARAYAPURAM

ABSTRACT

Document Clustering Based on Correlation Preserving Indexing is a new spectral clustering method, which is performed in the correlation similarity measure space. In this framework, latent semantic indexing is an indexing and retrieval method that uses a mathematical technique called singular value decomposition (SVD) to identify patterns in the relationships between the terms and concepts contained in an unstructured collection of text. The documents are projected into a low-dimensional semantic space that can be solved by generalized eigenvalve problem. Consequently, the proposed CPI method can effectively discover the intrinsic structures embedded in high-dimensional document space.

KEYWORDS

Document clustering, correlation measure, correlation latent semantic indexing, dimensionality reduction, singular value decomposition.

1 INTRODUCTION

Document clustering aims to automatically group related documents into clusters. The k-means method is one of the methods that use the euclidean distance,

$$dist = \sqrt{\sum_{k=1}^{n}(p_k - q_k)^2}$$

Where n is the number of dimensions (attributes) and p_k and q_k are, respectively, the k^{th} attributes (components) or data objects p and q. Which minimizes the sum of the squared euclidean distance between the data points and their corresponding cluster centres. Since the document space is always of high dimensionality, it is preferable to find a low-dimensional representation of the documents to reduce computation complexity.

Low computation cost is achieved in spectral clustering methods, in which the documents are first projected into a low-dimensional semantic space and then a traditional clustering algorithm is applied to finding document clusters. An effective document clustering method must be able to find a low-dimensional representation of the documents that can best preserve the similarities between the data points. Locality preserving indexing (LPI) method is a different spectral clustering method based on graph partitioning theory.

In this paper, we propose a new document clustering method based on correlation preserving indexing (CPI), which explicitly considers the manifold structure embedded in the similarities between the documents. It aims to find an optimal semantic subspace by simultaneously maximizing the correlations between the documents in the local patches and minimizing the correlations between the documents outside these patches.

1.1 SPECTRAL CLUSTERING

A special case of graph-based clustering, that has enjoyed much recent interest, constructs a bipartite graph of the rows and columns of the input data. The underlying assumption behind co clustering is that words which occur together are associated with similar concepts, and so it not just groups of similar documents that are important, but also groups of similar words.

Cuts in this bipartite graph produce co clusters of words (rows) and documents (columns). It has been shown that optimizing these cuts is an equivalent problem to computing the singular value decomposition of the original matrix.

1.2 VECTOR SPACE MODEL

The vector model was originally developed for automatic indexing. Under the vector model, a collection of *n* documents with *m* unique terms is represented as an *mxn* term-document matrix where each document is a vector of m dimensions. Several terms weighing schemes have been used, including binary term frequency and simple term frequency (how many times the words occur in the document). The document vectors are composed of weights reacting the frequency of the terms in the document multiplied by the inverse of their frequency in the entire collection (tf x idf). The assumption is that words which occur frequently in a document but rarely in the entire collection are of highly discriminative power. Under all these schemes, it is typical to normalize document vectors to unit length.

Two important properties should be stressed. First, in a collection of heterogeneous, the number of unique terms will be quite large. This results in document vectors of high dimensionality.

2 DOCUMENT CLUSTERING BASED ON CORRELATION PRESERVING INDEXING

Correlation as a similarity measure is suitable for capturing the manifold structure embedded in the high-dimensional document space. Mathematically, the correlation between two vectors (column vectors) u and v is defined as

$$Corr(u,v) = \frac{u^T v}{\sqrt{u^T u}\sqrt{v^T v}} = \left\langle \frac{u}{\|u\|}, \frac{v}{\|v\|} \right\rangle.$$

$$\theta$$

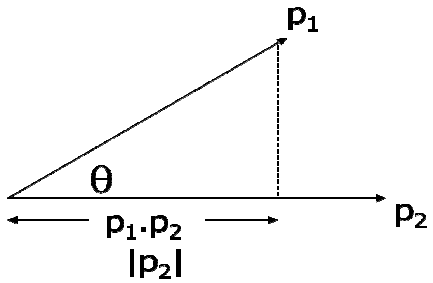
$$cos\theta = Corr(u,v).$$

Correlation corresponds to an angle Θ such that *cos Θ = Corr(u,v)*, the stronger the association between the two vectors *u* and *v*.

x = (−2.8, −1.8, −0.8, 1.2, 4.2) and y = (−0.028, −0.018, −0.008, 0.012, 0.042), from which

$$\cos\theta = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|} = \frac{0.308}{\sqrt{30.8}\sqrt{0.00308}} = 1 = \rho_{xy},$$

COSINE MEASURE DIAGRAM



$$dist(p_1, p_2) = \theta = \arccos(p_1.p_2/|p_2||p_1|)$$

2.1 CLASSIFICATION OF DOCUMENTS INTO CLUSTERS

A1. If two documents are close to each other in the original document space, then they tend to be grouped into the same cluster.

A2. If two documents are far away from each other in the original document space, they tend to be grouped into different clusters.

Based on these assumptions, we can propose a spectral clustering in the correlation similarity measure space through the nearest neighbours graph learning.

2.2 EXTERNAL EVALUATION OF CLUSTER QUALITY

Typical objective functions in clustering formalize the goal of attaining high intra-cluster similarity (documents within a cluster are similar) and low inter-cluster similarity (documents from different clusters are dissimilar).

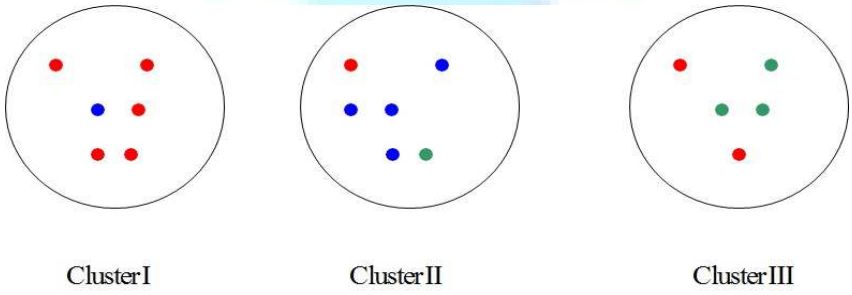
- Simple measure: purity, the ratio between the dominant class in the cluster π_i and the size of cluster ω_i

$$Purity(\omega_i) = \frac{1}{n_i}\max_j(n_{ij}) \quad j \in C$$

- Biased because having *n* clusters maximizes purity

Others are entropy of classes in clusters (or mutual information between classes and clusters)

CLUSTER PURITY



Cluster I Purity = 1/6(max(5,1,0))=5/6

Cluster II Purity = 1/6(max(1,4,1))=4/6

Cluster III Purity = 1/5(max(2,0,3))=3/5

The Rand index penalizes both false positive and false negative decisions during clustering. The F measure in addition supports differential weighting of these two types of errors.We can use the F measures to penalize false negatives more strongly than false positives by selecting a value β >1.

$$P = \frac{TP}{TP+FP} \qquad R = \frac{TP}{TP+FN} \qquad F_\beta = \frac{(\beta^2+1)PR}{\beta^2 P + R}$$

$$RI = \frac{TP+TN}{TP+FP+FN+TN}$$

Where, TP as True Positive, FP as False Positive, TN as True Negative and FN as false negative.

2.3 CLUSTERING ALGORITHM BASED ON CPI

The algorithm for document clustering based on CPI can be summarized as follows:

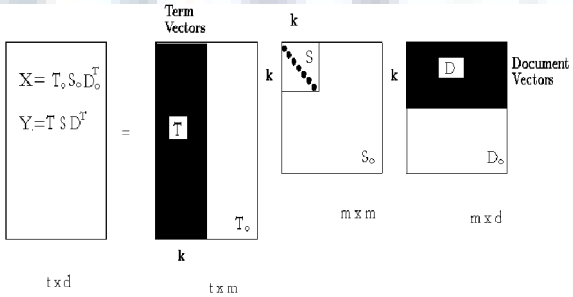
1. Construct the local neighbour patch, and compute the matrices MS and MT .

2. Project the document vectors into the singular value decomposition SVD subspace by throwing away the zero singular values. To establish a mapping from the term space to the topic space using latent semantic analysis, we use the SVD of the document-term index. The SVD is of the form:

$$X = U\Sigma V^T$$

Where X is our document by term matrix with elements representing the frequency of each term in each document, U and V are the set of left and right singular vectors respectively. The SVD has the properties that if we keep the greatest s singular values and remove the rest. This is a direct consequence of the Eckart-Young theorem.

FIGURE 1: SINGULAR VALUE DECOMPOSITION- Y IS THE APPROXIMATED X



3. Compute CPI Projection. Compute CPI Projection. Based on the multipliers $\lambda_0, \lambda_1, \ldots . \lambda_n$ the eigenvalue equation for D is Differential equation $Df = \lambda f$. Let W_{CPI}be the solution of the generalized eigenvalue problem $M_S W = \lambda M W$. Then the low dimensional representation of the document can be computed by $Y = W^T X$.

4. Cluster the documents in the CPI semantic subspace. Since the documents were projected on the unit hyper sphere, the inner product is a natural measure of similarity.

3. DOCUMENT REPRESENTATION

Step 1: Stemming - The process of reducing words to their base form, or stem. Porter's algorithm is the standard stemming algorithm.
Step 2: Stop word removal - A stop word which is not thought to convey any meaning as a dimension in the vector space.
Step 3: Pruning - Removes words that appear with very low frequency throughout the corpus. The underlying assumption is that these words, even if they had any discriminating power, would form too small clusters to be useful. Some words which occur too frequently are also removed.
Step 4: Technique – compute the term frequency Vector. Well known tf x idf (term frequency times inverted document frequency) weighting system defined as:

$$\mathrm{tf}(t,d) = \frac{\mathrm{f}(t,d)}{\max\{\mathrm{f}(w,d) : w \in d\}}$$

$$\mathrm{idf}(t,D) = \log \frac{|D|}{|\{d \in D : t \in d\}|}$$

$$\mathrm{tfidf}(t,d,D) = \mathrm{tf}(t,d) \times \mathrm{idf}(t,D)$$

Where D is the total number of documents in the corpus, $|\{d \in D : t \in d\}|$ is total number of documents where the term *t* appears.

TABLE 1: EXAMPLE OF IDF VALUES. HERE WE GIVE IDF'S OF TERM WITH VARIOUS FREQUENCIES IN THE REUTERS COLLECTION OF 806,791 DOCUMENTS

Term	df_t	idf_t
Car	18,165	1.65
Auto	6,723	2.08
Insurance	19,241	1.64
Best	25,235	1.5

```
//Calculates TF-IDF weight for each term t in //document d
private static float FindTFIDF(string document, string term)
{
float tf = FindTermFrequency(document, term);
float idf = FindInverseDocumentFrequency(term);
return tf * idf;
}
private static float FindTermFrequency(string document, string term)
{
int count = r.Split(document).Where(s => s.ToUpper() == term.ToUpper()).Count();
//ratio of no of occurance of term t in document d //to the total no of terms in the document
return (float)((float)count / (float)(r.Split(document).Count()));
}
private static float FindInverseDocumentFrequency(string term)
{
//find the  no. of document that contains the term //in whole document collection
int count = documentCollection.ToArray().Where(s => r.Split(
s.ToUpper()).ToArray().Contains(term.ToUpper())).Count();
/*
* log of the ratio of  total no of document in the collection to the no. of document containing the term
* we can also use Math.Log(count/(1+documentCollection.Count)) to deal with divide by zero case;
*/
return (float)Math.Log((float)documentCollection.Count() / (float)count);
}
}
```

4. CONCLUSION

In this paper, we present a new document clustering method based on correlation preserving indexing. It simultaneously maximizes the correlation between the documents in the local patches and minimizes the correlation between the documents outside these patches. Consequently, a low dimensional semantic subspace is derived where the documents corresponding to the same semantics are close to each other.

5. REFERENCES

1. D. Cai, X. He, and J. Han, "Document Clustering Using Locality Preserving Indexing," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 12, pp. 1624-163
2. D.M. Blei, A.Y. Ng, and M.I. Jordan, "Latent Dirichlet Allocation,"J. Machine Learning Research, vol. 3, pp. 993-1022, 2003.
3. Md Maruf, HASAN Yuji MATSUMOTO, "Document Clustering: Before and After the Singular Value Decomposition"
4. S.C. Deerwester, S.T. Dumais, T.K. Landauer, G.W. Furnas, and R.A. Harshman, "Indexing by Latent Semantic Analysis" J. Am. Soc. Information Science, vol. 41, no. 6, pp. 391-407, 1990.
5. Taiping Zhang, Yuan Yan Tang,Bin Fang and Yong Xiang "Document Clustering in Correlation Similarity Measure Space" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 6, JUNE 2012.
6. The Measure of Correlation, Macintosh HD:DA:DA IX:Volume II stuff:02 Correlation March 19, 1997
7. Thomas K Landauer, Peter W. Foltz, Darrell Laham "An Introduction to Latent Semantic Analysis",
8. Uppe Nanaji, Majji Nagaraju " Statistical Measurement Space for Document Clustering Based on Correlation Preserving Indexing"International Journal of Engineering Research & Technology (IJERT), Vol. 1 Issue 7, September – 2012

WEBSITES

9. en.wikipedia.org/wiki/Eigendecomposition_of_a_matrix
10. en.wikipedia.org/wiki/Latent_semantic_indexing.

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 tosupply 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 youhave 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