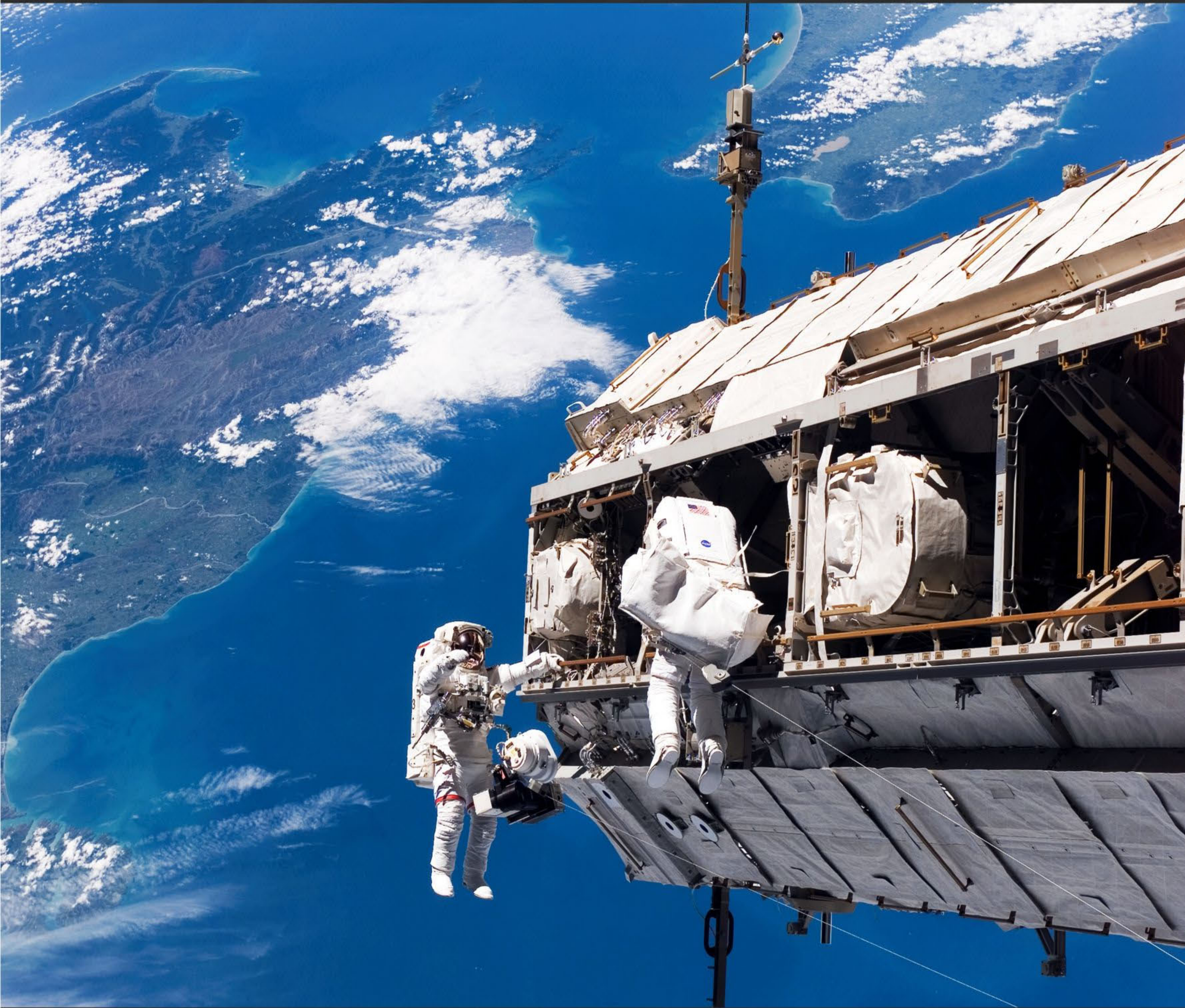
Online ISSN: 2249-4596
Print ISSN: 0975-5861

Global Journal of Researches in Engineering: I Numerical Methods

DISCOVERING THOUGHTS AND INVENTING FUTURE



Highlights

- Bayesian Spam Filtering
- Robust Multiple Watermarking
- Decision Tree Construction

Volume 11
Issue 7
version 1.0

Global Journal of Research in Engineering :I

Numerical Methods

Global Journal of Research in Engineering :I

Numerical Methods

Volume 11 Issue 7 (Ver. 1.0)

Open Association of Research Society

Incorporation No.: 0423089
License No.: 42125/022010/1186
Registration No.: 430374
Import-Export Code: 1109007027
Employer Identification Number (EIN):
USA Tax ID: 98-0673427

Global Journals Inc.

*(A Delaware USA Incorporation with "Good Standing"; **Reg. Number: 0423089**)*
Sponsors: Open Association of Research Society
Open Scientific Standards

Publisher's Headquarters office

Global Journals Inc., Headquarters Corporate Office, Cambridge Office Center, II Canal Park, Floor No. 5th, ***Cambridge (Massachusetts)***, Pin: MA 02141
United States
USA Toll Free: +001-888-839-7392
USA Toll Free Fax: +001-888-839-7392

Offset Typesetting

Open Association of Research Society, Marsh Road, Rainham, Essex, London RM13 8EU
United Kingdom.

Packaging & Continental Dispatching

Global Journals, India

Find a correspondence nodal officer near you

To find nodal officer of your country, please email us at *local@globaljournals.org*

eContacts

Press Inquiries: *press@globaljournals.org*
Investor Inquiries: *investers@globaljournals.org*
Technical Support: *technology@globaljournals.org*
Media & Releases: *media@globaljournals.org*

Pricing (Including by Air Parcel Charges):

For Authors:
22 USD (B/W) & 50 USD (Color)
Yearly Subscription (Personal & Institutional):
200 USD (B/W) & 250 USD (Color)

Contents of the Volume

Global Journal of researches in engineering
Numerical Methods
Volume 11 Issue 7 Version 1.0 December 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 2249-4596 & Print ISSN: 0975-5861

A New Approach for Calculating Average Value (Including Null) Without Aggregate Function

By Mridul Kanti Das, Goutam Biswas, Md. Masudur Rahman, Dr. Md. Nasim Akhtar

Dhaka University of Engineering & Technology (DUET), Gazipur, Bangladesh

Abstract - An evaluation of aggregate functions in relational database system considerable impact on performance in many application areas like geographic information systems and statistical and scientific databases. The problem with existing systems is inefficient execution of aggregate functions with large data volumes and lack of flexibility. It is not possible to extend the systems with new aggregate (average) functions. We show how this could be implemented into a database. We also describe how support for special kinds of aggregate queries and data structures can help in designing future high performance systems.

Keywords : Aggregate Function, NULL, SQL, Boolean Algebra, Sub Queries.

GJRE-I Classification : FOR Code: 010303

A NEW APPROACH FOR CALCULATING AVERAGE VALUE INCLUDING NULLWITHOUT AGGREGATE FUNCTION

Strictly as per the compliance and regulations of :

A New Approach for Calculating Average Value (Including Null) Without Aggregate Function

Mridul Kanti Das[α], Goutam Biswas[Ω], Md. Masudur Rahman[β], Dr. Md. Nasim Akhtar[ψ]

Abstract - An evaluation of aggregate functions in relational database system considerable impact on performance in many application areas like geographic information systems and statistical and scientific databases. The problem with existing systems is inefficient execution of aggregate functions with large data volumes and lack of flexibility. It is not possible to extend the systems with new aggregate (average) functions. We show how this could be implemented into a database. We also describe how support for special kinds of aggregate queries and data structures can help in designing future high performance systems.

Keywords : Aggregate Function, NULL, SQL, Boolean Algebra, Sub Queries.

I. Introduction

Aggregate functions perform a calculation on a set of values and return a single value. Aggregate function avg( ) only calculate average without null values. It provides average result, eliminating null values. Null does not have a value (and is not a member of any data domain) but it is a placeholder or "mark" for missing information. Comparisons with Null can never result in either True or False but always in the third logical result is Unknown. So comparing two null is difficult. We discus about

(1) review of the research for handling null values in database system using aggregate function (2) problem structure with null value with respect to database (3) describes existing solution and proposed solution and its algorithm as well as how it works (4) details the experimental work that has been carried out. The experimental evaluation has been performed using a large amount of datasets.

Application:

Database System & its related Application software.

Author α : Bachelor of Science in Engineering degree from Department of Computer Science and Engineering (CSE), Dhaka University of Engineering & Technology (DUET), Gazipur, Bangladesh. PH-+8801715543754. E-mail : mridul.duet@yahoo.com

Author Ω : Bachelor of Science in Engineering degree from Department of Computer Science and Engineering (CSE), Dhaka University of Engineering & Technology (DUET), Gazipur, Bangladesh. PH:+8801710741874. E-mail : welovegoutam@yahoo.com

Author β : Bachelor of Science in Engineering degree from Department of Computer Science and Engineering (CSE), Dhaka University of Engineering & Technology (DUET), Gazipur, Bangladesh. PH-+8801710370054. Email : masudcse21@gmail.com

Author ψ : Asst. Prof. of Department of CSE, DUET and he achieved his B.Sc & M.Sc from NTUU, Ukraine and Ph. D. from MITXT, Russia . PH-+8801711395537. Email : nasim_duet@yahoo.com

Related woks:

Baekgraard and Mark [1] on this work, developing a nested algebra for describing sub-queries, but without considering aggregates, duplicates, or null values,

Jeff Smith [2] is software developer, he compare all columns in database, and also handles comparing NULL values to other NULLs, Using UNION operator.

- Aggregate function avg (X) returns the average value of all non-NULL X within a group ignoring null values .

So, we do not get appropriate result.

Aggregate avg ( ) function with Null value problem solution:

- *SQL>select avg (Amount) from bank;*

Table bank

name	*Address*	*account_no*	*amount*
Mridul	*Sylhet*	*49501*	*500*
outam	*Gopalganj*	*49502*	*NULL*
Masud	*Narial*	*49503*	*600*
Ashim	*Khulna*	*49504*	*600*
Swapan	*Dinajpur*	*49505*	*800*
Kishor	*Kushtia*	*49506*	*600*
Shahid	*Khulna*	*49507*	*1100*

Normal Avg Query Result is

Avg=(4200)/6

=700

II. Existing Solution

- Aggregate avg ( ) function with Null value problem solution:
- SQL>select sum (amount)/count(*) from bank;

Global Journal of Researches in Engineering (I) Volume XI Issue VII Version I December 2011

Table bank			
name	*Address*	*account_no*	*amount*
Mridul	*Sylhet*	*49501*	*500*
outam	*Gopalganj*	*49502*	*NULL*
Masud	*Narial*	*49503*	*600*
Ashim	*Khulna*	*49504*	*600*
Swapan	*Dinajpur*	*49505*	*800*
Kishor	*Kushtia*	*49506*	*600*
Shahid	*Khulna*	*49507*	*1100*

Normal Avg Query Result is

Avg=(4200)/7

=600

III. Performance Measure Table

Table bank		
Amount of Data	*Amount of NULL*	*Existing solution Execution Time (sec)*
40000	1000	0.2840
80000	4000	0.5720
120000	6000	0.8440
160000	7000	1.1841
180000	7000	1.3301
200000	7000	1.5511
220000	7000	1.7511
240000	7000	1.9371
260000	8000	2.0601
280000	8000	2.1851

IV. Proposed Algorithm

```
Algorithm  Average_WN (d, Avg, size )
// d is a data table which like two dimensional
//array,   size is maximum data row. r and c //mean
row and column of data table //respectively.
// By this algorithm find the average value
//considering NULL
{
sum:=0.0, Avg:=0.0;
size:=row size of data table;
For r:=0 to size-2  step 2
{
If( d[r][c]=NULL and d[r+1][c] = NULL ) then ;
else if( d[r][c]=NULL ) then
sum:=sum+ d[r+1][c] ;
else if( d[r+1][c]=NULL ) then
sum:=sum+ d[r][c] ;
else
sum:=sum+ d[r][c] + d[r+1][c] ;
}
If( size mod 2=1)
{
If(size=1) then
sum := sum + d[r][c];
else
sum := sum + d[r-1][c];
}
Avg:=sum/size ;
Print AVG ;
}
```

V. Performance Measure Table of Proposed Solution

Table bank		
Amount of Data	*Amount of NULL*	*Proposed Solution Execution Time (sec)*
40000	1000	0.2460
80000	4000	0.5040
120000	6000	0.8250
160000	7000	1.1591
180000	7000	1.3011
200000	7000	1.5011
220000	7000	1.5781
240000	7000	1.7641
260000	8000	1.9671
280000	8000	2.0762

In this solution we see that if the number of data in database gradually increased then the execution time is increased. Built-in function sum() and count() scan record individually from top to bottom in a table. Therefore Existing query are not efficient to calculating average value in large amount of data with Null values.

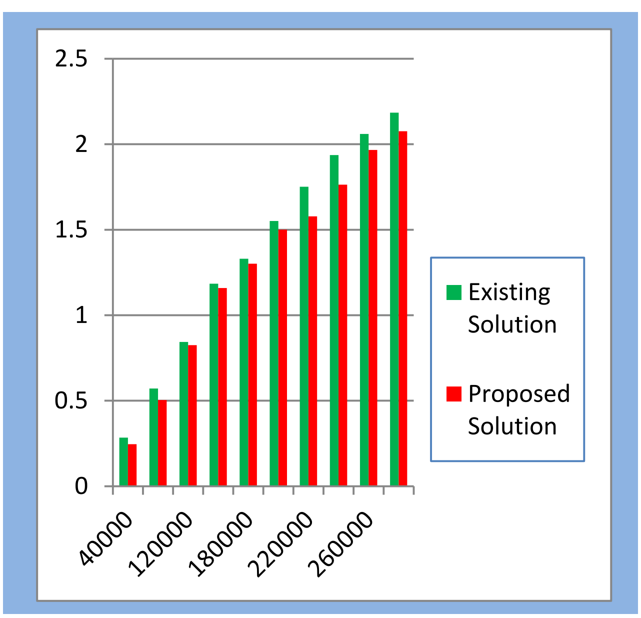
VI. Comparison Table Between Existing and Proposed Solution

Table bank		
Amount of Data	*Existing solution Execution Time (sec)*	*Proposed Solution Execution Time (sec)*
40000	0.2840	0.2460
80000	0.5720	0.5040
120000	0.8440	0.8250
160000	1.1841	1.1591
180000	1.3301	1.3011
200000	1.5511	1.5011
220000	1.7511	1.5781
240000	1.9371	1.7641
260000	2.0601	1.9671
280000	2.1851	2.0762

From comparison table we see that our propose system takes less times than existing system. By proposed system can reduce time and reduce the problem of existing system. To understand easily a graph chart is given below.

VII. Graph of Existing Solution VS. Proposed Solution

Our propose solution is efficient to calculate average value with Null values from large amount of data.



From the above graph Green bar indicates Existing solution time and Red Bar indicates proposed solution time. We see that in proposed system needs execution time less than existing system.

VIII. Conclusion

At the age of globalization most of all bank already has been computerized. They store their customer information, balance, transaction etc. in database. And they need to calculate average number of transaction after a certain period of time. Even stock exchange Ltd. Hospital, Airlines etc. need to calculate average number of transaction frequently. So our proposed system will be best for them which can save their times.

References Références Referencias

1. "A Relational Model of Data for Large Shared Data Banks" E. F. Codd, Communications of the ACM, Vol. 13, No. 6, June 1970, pp. 377–387.
2. Domain (Data Type) and null. Posted June 17,2009 http://dbtips.wordpress.com/2009/05/31/some-notes-about-null/
3. Jeff Smith is software developer, he using UNION operator comparing NULL values to other NULLs. www.weblogs.sqlteam.com.
4. Analytic Functions from Oracle® Database SQL Reference 10g Release 1 (10.1) Part Number B10759-01
5. L. Baekgraard and L. Mark. Incremental computation of nested relational query expression. ACM TODS, 20(2):111-148, June 1995.
6. Comparing Tables By Bill Graziano on 07 January 2002.
7. SQL Functions from Oracle Database Globalization Support Guide
8. "Structured Query Language (SQL)". International Business Machines. October 27, 2006. Retrieved 2007-06-10.

This page is intentionally left blank

Global Journal of researches in engineering
Numerical Methods
Volume 11 Issue 7 Version 1.0 December 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 2249-4596 & Print ISSN: 0975-5861

Bayesian Spam Filtering Using Statistical Data Compression

By Gumpina V V Satya Prasad, Satya P Kumar Somayajula
Avanthi Institute of Engineering and Technology, Makavarapalem, Visakhapatnam

Abstract - The Spam e-mail has become a major problem for companies and private users. This paper associated with spam and some different approaches attempting to deal with it. The most appealing methods are those that are easy to maintain and prove to have a satisfactory performance. Statistical classifiers are such a group of methods as their ability to filter spam is based upon the previous knowledge gathered through collected and classified e-mails. A learning algorithm which uses the Naive Bayesian classifier has shown promising results in separating spam from legitimate mail.

GJRE-I Classification : FOR Code: 010303

BAYESIAN SPAM FILTERING USING STATISTICAL DATA COMPRESSION

Strictly as per the compliance and regulations of :

Bayesian Spam Filtering Using Statistical Data Compression

Gumpina V V Satya Prasad[α], Satya P Kumar Somayajula[Ω]

Abstract - The Spam e-mail has become a major problem for companies and private users. This paper associated with spam and some different approaches attempting to deal with it. The most appealing methods are those that are easy to maintain and prove to have a satisfactory performance. Statistical classifiers are such a group of methods as their ability to filter spam is based upon the previous knowledge gathered through collected and classified e-mails. A learning algorithm which uses the Naive Bayesian classifier has shown promising results in separating spam from legitimate mail.

I. Introduction

Spam has become a serious problem because in the short term it is usually economically beneficial to the sender. The low cost of e-mail as a communication medium virtually guaranties profits. Even if a very small percentage of people respond to the spam advertising message by buying the product, this can be worth the money and the time spent for sending bulk e-mails. Commercial spammers are often represented by people or companies that have no reputation to lose. Because of technological obstacles with e-mail infrastructure, it is difficult and time-consuming to trace the individual or the group responsible for sending spam. Spammers make it even more difficult by hiding or forging the origin of their messages. Even if they are traced, the decentralized architecture of the Internet with no central authority makes it hard to take legal actions against spammers. The statistical filtering (especially Bayesian filtering) has long been a popular anti-spam approach, but spam continues to be a serious problem to the Internet society. Recent spam attacks expose strong challenges to the statistical filters, which highlights the need for a new anti-spam approach. The economics of spam dictates that the spammer has to target several recipients with identical or similar e-mail messages. This makes collaborative spam filtering a natural defense paradigm, wherein a set of e-mail clients share their knowledge about recently receivedspame-mails, providing a highly effective defense against a substantial fraction of spam attacks. Also, knowledge sharing can significantly alleviate the burdens of frequent training stand-alone spam filters. However, any large-scale collaborative anti-spam approach is faced with a fundamental and important challenge, namely ensuring the privacy of the e-mails among un trusted e-mail entities. Different from the e-mail service providers such as Gmail or Yahoo mail, which utilizes spam or ham(non-spam) classifications from all its users to classify new messages, privacy is a major concern for cross-enterprise collaboration, especially in a large scale. The idea of collaboration implies that the participating users and e-mail servers have to share and exchange information about the e-mails (including the classification result). However, e-mails are generally considered as private communication between the senders and the recipients, and they often contain personal and confidential information. Therefore, users and organizations are not comfortable sharing information about their e-mails until and unless they are assured that no one else (human or machine) would become aware of the actual contents of their e-mails. This genuine concern for privacy has deterred users and organizations from participating in any large-scale collaborative spam filtering effort. To protect e-mail privacy, digest approach has been proposed in the collaborative anti-spam systems to both provide encryption for the e-mail messages and obtain useful information (fingerprint) from spam e-mail. Ideally, the digest calculation has to be a one-way function such that it should be computationally hard to generate the corresponding e-mail message. It should embody the textual features of the e-mail message such that if two e-mails have similar syntactic structure, then their fingerprints should also be similar.Afew distributed spam identification schemes, such as Distributed Checksum Clearinghouse (DCC) [2] and Vipul's Razor [3] have different ways to generate fingerprints. However, these systems are not sufficient to handle two security threats: 1) Privacy breach as discussed in detail in Section 2 and 2) Camouflage attacks, such as character replacement and good word appendant, make it hard to generate the same e-mail fingerprints for highly similar spam e-mails.

II. Statistical Data Compression

Probability plays a central role in data compression: Knowing the exact probability distribution governing an information source allows us to construct optimal or near-optimal codes for messages produced by the source. A statistical data compression algorithm

Author α : M.Tech, Asst. Professor, IT Dept, Sir. C. R. Reddy College of Engg, Eluru, A.P. India. E-mail : prasad17gumpina@gmail.com
Author Ω : M.Tech, Asst. Professor, CSE Dept, Avanthi Institute of Engineering and Technology, Makavarapalem, Visakhapatnam. E-mail : balasriram1982@gmail.com

exploits this relationship by building a statistical model of the information source, which can be used to estimate the probability of each possible message. This model is coupled with an encoder that uses these probability estimates to construct the final binary representation. For our purposes, the encoding problem is irrelevant. We therefore focus on the source modeling task

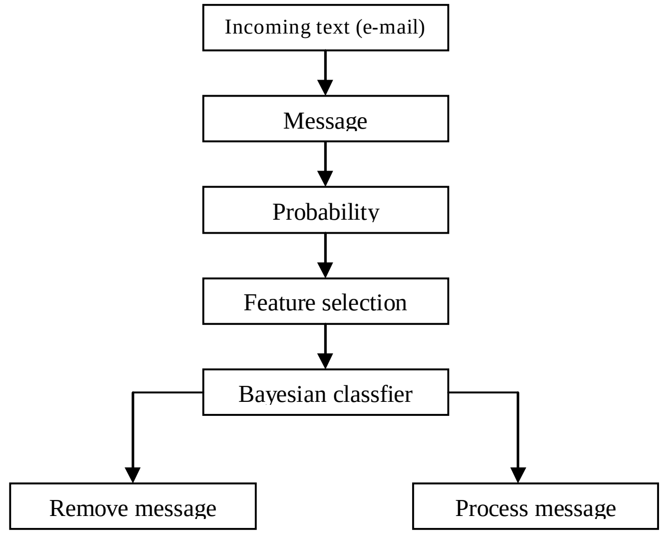
III. Preliminaries

We denote by X the random variable associated with the source, which may take the value of any message the source is capable of producing, and by P the probability distribution over the values of X with the corresponding probability mass function p. We are particularly interested in modeling of text generating sources. Each message **x** produced by such a source is naturally represented as a sequence $\mathbf{X}=x1^{n} = x1\ldots xn$ $\boldsymbol{\Sigma}$*of symbols over the source alphabet $\boldsymbol{\Sigma}$. The length of a sequence can be arbitrary. For text generating sources, it is common to interpret a symbol as a single character, but other schemes are possible, such as binary (bitwise) or word-level models. The entropy $H(X)$ of a source X gives a lower bound on the average per-symbol code length required to encode a message without loss of information: $H(x) = E_{x\sim P}(-\frac{1}{|x|}\log p(x))$

This bound is achievable *only* when the true probability distribution P governing the source is known. In this case, an average message could be encoded using no less than $H(X)$ bits per symbol. However, the true distribution over all possible messages is typically unknown. The goal of any statistical data compression algorithm is then to infer a probability mass function over sequences $f:\Sigma^* \rightarrow [0,1]$, which matches the true distribution of the source as accurately as possible. Ideally2, a sequence **x** is then encoded with $L(\mathbf{x})$ bits, where $L(\mathbf{x}) = -\log f(\mathbf{x})$. The compression algorithm must therefore *learn* an approximation of P in order to encode messages efficiently. A better approximation will, on average, lead to shorter code lengths. This simple observation alone gives compelling motivation for the use of compression algorithms in text categorization.

IV. Bayesian Spam Filtering

Bayesian spam filtering can be conceptualized into the model presented in Figure 1. It consists of four major modules, each responsible for four different processes: message tokenization, probability estimation, feature selection and Naive Bayesian classification.



When a message arrives, it is firstly tokenized into a set of features (tokens), F . Every feature is assigned an estimated probability that indicates its spaminess. To reduce the dimensionality of the feature vector, a feature selection algorithm is applied to output a subset of the features. The Naive Bayesian classifier combines the probabilities of every feature in 1 F , and estimates the probability of the message being spam. In the following text, the process of Naive Bayesian classification is described, followed by details concerning the measuring performance. This order of explanation is necessary because the sections concerned with the first three modules require understanding of the classification process and the parameters used to evaluate its improvement.

V. Performance Evolution

Precision and recall a well employed metric for performance measurement in information retrieval is precision and recall. These measures have been diligently used in the context of spam classification (Sahami et al.1998). Recall is the proportion of relevant items that are retrieved, which in this case is the proportion of spam messages that are actually recognized. For example if 9 out of 10 spam messages are correctly identified as spam, the recall rate is 0.9. Precision is defined as the proportion of items retrieved that are relevant. In the spam classification context, precision is the proportion of the spam messages classified as spam over the total number of messages classified as spam. Thus if only spam messages are classified as spam then the precision is 1. As soon as a good legitimate message is classified as spam, the precision will drop below 1. Formally: Let gg n be the number of good messages classified as good (also known as false negatives). Let gs n be the number of good messages classified as spam (also known as false positives).(9). Let ss n be the number of spam messages classified as spam (also known as true positives). Let sg n be the number of spam messages classified as good (also known as true negatives). The precision calculates the occurrence of false positives which are good messages classified as spam. When this happens p drops below 1. Such misclassification could be a disaster for the user whereas the only impact of a low recall rate is to receive spam messages in the inbox. Hence it is more important for the precision to be at a high level than the recall rate. The precision and recall reveal little unless used together. Commercial spam filters sometimes claim that they have an incredibly high precision value of 0.9999% without mentioning the related recall rate. This can appear to be very good to the untrained eye. A reasonably good spam classifier should have precision very close to 1 and a recall rate > 0.8. A problem when evaluating classifiers is to find a good balance between the precision and recall rates. Therefore it is necessary to use a strategy to obtain a combined score. One way to achieve this is to use weighted accuracy.

VI. Cross Validation

There are several means of estimating how well the classifier works after training. The easiest and most straightforward means is by splitting the corpus into two parts and using one part for training and the other for testing. This is called the holdout method. The disadvantage is that the evaluation depends heavily on which samples end up in which set. Another method that reduces the variance of the holdout method is k -fold cross-validation. In k -fold cross-validation (Kohavi 1995) the corpus, M , is split into k mutually exclusive parts, k M ,M ,...M 1 2 . The inducer is trained on i M \ M and tested against i M . This is repeated k times with different i such that iÎ{1,2,...,k}. Finally the performance is estimated as the mean of the total number of tests.

VII. Conclusion

Optimal search algorithm called SFFS was applied to find a subset of delimiters for the tokenizer. Then a filter and a wrapper algorithm were proposed to determine how beneficial a group of delimiters is to the classification task. The filter approach ran about ten times faster than the wrapper, but did not produce significantly better subsets than the base-lines. The wrapper did improve the performance on all corpuses by finding small subsets of delimiters. This suggested an idea concerning how to select delimiters for a near-optimal solution, namely to start with space and then add a few more. Since the wrapper generated subsets had nothing in common apart from space, the recommendation is to only use space as a delimiter. The wrapper was far too slow to use in spam filter.

References Références Referencias

1. Almuallim, H. and T. Dietterich. (1991), Learning with many irrelevant features. In Proceedings of the Ninth National Conference on Artificial Intelligence, pp. 547-552. Menlo Park, CA: AAAI Press/The MIT Press.
2. Androutsopoulos I., Paliouras G., Karkaletsis V., Sakkis G., Spyropoulos C. and Stamatopoulos, P. (2000a) Learning to filter spam email: A comparison of a naive bayesian and a memory-based approach. In Workshop on Machine Learning and Textual Information Access, 4th European
3. Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD 2000). Androutsopoulos, I., Koutsias, J., Chandrinos, K.V., Paliouras, George and Spyropoulos, C.D. (2000b),
4. An Evaluation of Naive Bayesian Anti-Spam Filtering. In Potamias, G., Moustakis, V. and van Someren, M. (Eds.), Proceedings of the Workshop on Machine Learning in the New
5. Information Age, 11th European Conference on Machine Learning (ECML 2000), Barcelona, Spain, pp. 9-17.
6. Androutsopoulos, I., Paliouras, G., Michelakis, E. (2004), Learning to Filter Unsolicited Commercial E-Mail. Athens University of Economics and Business and National Centre for Scientific Research “Demokritos” Bevilacqua-Linn M. (2003),
7. Machine Learning for Naive Bayesian Spam Filter Tokenization Breiman, L., and Spector, P. (1992), Submodel selection and evaluation in regression: The Xrandom case. International Statistical Review, 60, 291-319.
8. Androutsopoulos, G. Paliouras, and E. Michelakis. Learning to filter unsolicited commercial e-mail. Technical Report 2004/2, NCSR “Demokritos”, October 2004.
9. F. Assis, W. Yerazunis, C. Siefkes, and S. Chhabra. CRM114 versus Mr. X: CRM114 notes for the TREC 2005 spam track. In *Proc. 14th Text REtrieval Conference (TREC 2005)*, Gaithersburg, MD, November 2005.
10. A.R. Barron, J. Rissanen, and B. Yu. The minimum description length principle in coding and modeling. *IEEE Transactions on Information Theory*, 44(6):2743–2760, 1998. D. Benedetto, E. Caglioti, and Loreto V. Language trees and zipping. *Physical Review Letters*, 88 (4), 2002.
11. A.Bratko and B. Filipi Spam filtering using character-level markov models: Experiments for the TREC 2005 Spam Track.

This page is intentionally left blank

Global Journal of researches in engineering
Numerical Methods
Volume 11 Issue 7 Version 1.0 December 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 2249-4596 & Print ISSN: 0975-5861

Three Dimensional Discrete Wavelet Transform Based Robust Multiple Watermarking Algorithm for Spectral Images

By S.Maheswari, K.Rameshwaran
Kongu Engineering College, Perundurai, Tamil Nadu, India

Abstract - Three Dimensional (3-D) imaging is a fast growing technology in the areas of computer graphics, robotics, metrology, medicine and cultural heritage etc. With the development of internet, 3-D contents are easily manipulated, transmitted and widely distributed. Watermarking has been proved to be an efficient technology in copyright protection of multimedia information. In this paper, a novel blind multiple watermarking algorithm for copyright protection of spectral images is presented. Watermarking is based on 3-D Discrete Wavelet Transform (DWT). Dual binary watermarks are proposed to embed in the mid frequency bands of cover image. One level 3-D DWT is applied to the original spectral image and one level 2-D DWT is applied on two binary watermarks. Eigen values of selected subbands of spectral image are modified by the eigen values of selected subbands of two binary watermarks. Inverse SVD (Singular Value Decomposition) and Inverse 3-D DWT is taken on the modified subbands to obtain the watermarked image. During extraction process, two binary watermarks are recovered from the test image with the help of a secret key. An experimental result shows that, the proposed algorithm produces very high imperceptibility and robustness against various types of attacks like low pass filtering, median Filtering, JPEG compression, histogram equalization, cropping, various angles of rotation and addition of noise like Gaussian, Salt &Pepper, Speckle, Poisson etc.

Keywords : *DWT, 3-D DWT, SVD and Multiple Watermarking.*

GJRE-I Classification : *FOR Code: 010399*

THREE DIMENSIONAL DISCRETE WAVELET TRANSFORM BASED ROBUST MULTIPLE WATERMARKING ALGORITHM FOR SPECTRAL IMAGES

Strictly as per the compliance and regulations of :

Three Dimensional Discrete Wavelet Transform Based Robust Multiple Watermarking Algorithm for Spectral Images

S.Maheswari[α], K.Rameshwaran[Ω]

Abstract - Three Dimensional (3-D) imaging is a fast growing technology in the areas of computer graphics, robotics, metrology, medicine and cultural heritage etc. With the development of internet, 3-D contents are easily manipulated, transmitted and widely distributed. Watermarking has been proved to be an efficient technology in copyright protection of multimedia information. In this paper, a novel blind multiple watermarking algorithm for copyright protection of spectral images is presented. Watermarking is based on 3-D Discrete Wavelet Transform (DWT). Dual binary watermarks are proposed to embed in the mid frequency bands of cover image. One level 3-D DWT is applied to the original spectral image and one level 2-D DWT is applied on two binary watermarks. Eigen values of selected subbands of spectral image are modified by the eigen values of selected subbands of two binary watermarks. Inverse SVD (Singular Value Decomposition) and Inverse 3-D DWT is taken on the modified subbands to obtain the watermarked image. During extraction process, two binary watermarks are recovered from the test image with the help of a secret key. An experimental result shows that, the proposed algorithm produces very high imperceptibility and robustness against various types of attacks like low pass filtering, median Filtering, JPEG compression, histogram equalization, cropping, various angles of rotation and addition of noise like Gaussian, Salt &Pepper, Speckle, Poisson etc.

Keywords : DWT, 3-D DWT, SVD and Multiple Watermarking.

I. Introduction

Digital watermarking and steganography techniques are used to address digital rights management, protect information, and conceal secrets. During the last decade, 3-D meshes have been widely used in industrial, medical and entertainment applications. Since 3-D watermarking has also become an active research topic to protect 3-D information, we present a new frame work for copyright protection of 3-D spectral images in this paper. Spectral imaging is a practical tool for various applications like medical imaging, industrial quality control, digital commerce and maintenance of cultural heritage in digital museums etc [1]. In spectral color imaging, color of an object can be represented more accurately as compared to traditional three channel RGB images.

Three conflicting requirements of watermarking systems are robustness, imperceptibility and capacity [2]. Any watermarking system should allow embedding large number of secret information. Embedding of secret message should not degrade the quality of image; it should withstand in the image; for all kinds of attacks like JPEG compression, low pass filtering, median filtering, noise addition, histogram equalization, rotation, scaling, cropping etc. Embedding of multiple watermarks enhances the robustness property. Embedding of multiple watermarks, instead of a single one into one cover image improves the robustness of watermarking scheme [3].

Transform domain technique offers very high robustness as compared to spatial domain technique; but it needs more computational complexity because input images are converted into transform coefficients by using various image transforms like DCT [4-5], DFT and DWT [6-9] etc. 3-D watermarking techniques are classified into two types [10], similar to that of 2-D watermarking, spatial domain technique [10] and spectral domain technique [4-9]. Spatial domain techniques are further classified into two types. They are Geometry modification technique and Topology modification technique. In spectral method; information can be embedded in one of the mesh transform domains like mesh spectral decomposition, wavelet transform or spherical transform.

The use of wavelets in image and video coding has increased significantly over the years, mainly due to the superior energy compaction property of wavelets compared with the traditional transforms like DCT [11]. In this paper, we have proposed a method to embed the watermark in wavelet transform domain. The 3-D object is processed for obtaining 2-D slices, so that each 3-D objects are represented by a set of 2-D slices. A 3-D DWT domain, obtained by performing a 2-D spatial wavelet transform and then a temporal 1-D wavelet transform [12]. The Haar Wavelet Transform consistently outperforms the more complex ones when using non-coloured watermark [13]. Therefore, we propose 3-D Haar Wavelet Transform based watermarking scheme. Eigen values of mid frequency subbands of a cover image and two binary watermarks are obtained by

Author α : Kongu Engineering College, Perundurai, Erode-638052, Tamil Nadu, India. (E-mail : maheswari_bsb@yahoo.com).
Author Ω : MAM Group of Institutions, Triuchirapalli-620009, Tamil Nadu, India. (E-mail : krameshwaran @gmail.com).

Singular Value Decomposition (SVD)[1, 8]. Eigen values of mid frequency subbands are modified by the Eigen values of dual binary watermarks. Section (2) deals with 3-D DWT. Section (3) discusses about the proposed embedding and extraction algorithm, Section (4) gives the experimental results of proposed scheme, followed by conclusion in section (5).

II. Three Dimensional DWT

A 3-D DWT performs wavelet transform in the three directions x, y and z on the image. A 3-D image is an extension of 2-D image along with time axis. To use the wavelet transform for 3-D images, we must implement a 3-D version of analysis and synthesis filter banks [12]. Figure 1 shows the decomposition structure of one level 3D-DWT.

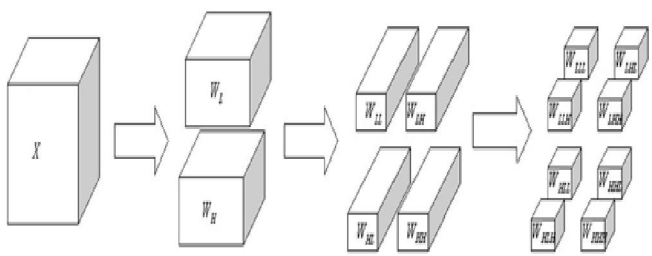


Fig. 1: One Level 3-D DWT

If the data is of the size of npixels. The steps of the 3-D Discrete Wavelet Transform are defined as follows. n_1, n_2 and n_3. After applying DWT in one dimension, we obtain two subbands of the size of $n_1/2$, n_2 and n_3. After applying DWT in second dimension, we obtain four subbands of the size of $n_1/2$, $n_2/2$ and n_3. After applying third dimension, we obtain the eight subbands of the size of $n_1/2$, $n_2/2$ and $n_3/2$ [11]. Let X is an image of size $m \times n \times l$ pixels. The steps of the 3-D Discrete Wavelet Transform are defined as follows.

Step 1 : In the horizontal direction, the original image x (m,n,l) is filtered by the filters $H_0(n)$ and $H_1(n)$respectively. $Two images\ x_0(m,n,l)\ and\ x_1(m,n,l)$ are produced.

Step 2 : In the vertical direction, the two images $x_0(m,n,l)$ and $x_1(m,n,l)$ are filtered by the filters $H_0(n)$ and $H_1(n)$ respectively. This gives four images $x_{00}(m,n,l), x_{01}(m,n,l), x_{10}(m,n,l) and\ x_{11}(m,n,l)$.

Step 3 : In the temporal direction, the four images, $x_{ij}(m,n,l)$, $0 \le i \le 2$,$0 \le j \le 2$ are filtered by the filters $H_0(n)$ and $H_1(n)$ respectively. This gives eight images $x_{000}(m,n,l), x_{001}(m,n,l), x_{010}(m,n,l), \ldots x_{ijk}(m,n,l)$ where $0 \le i \le 2, 0 \le j \le 2\ and\ 0 \le k \le 2$.

Step 4 : Steps 1 to 3 can be repeated on the subimage $x_{ijk}(m,n,l)$ so as to get the other subimages for the next scale.

Figures 2 and 3 show the analysis and synthesis filter banks of 3-D DWT.

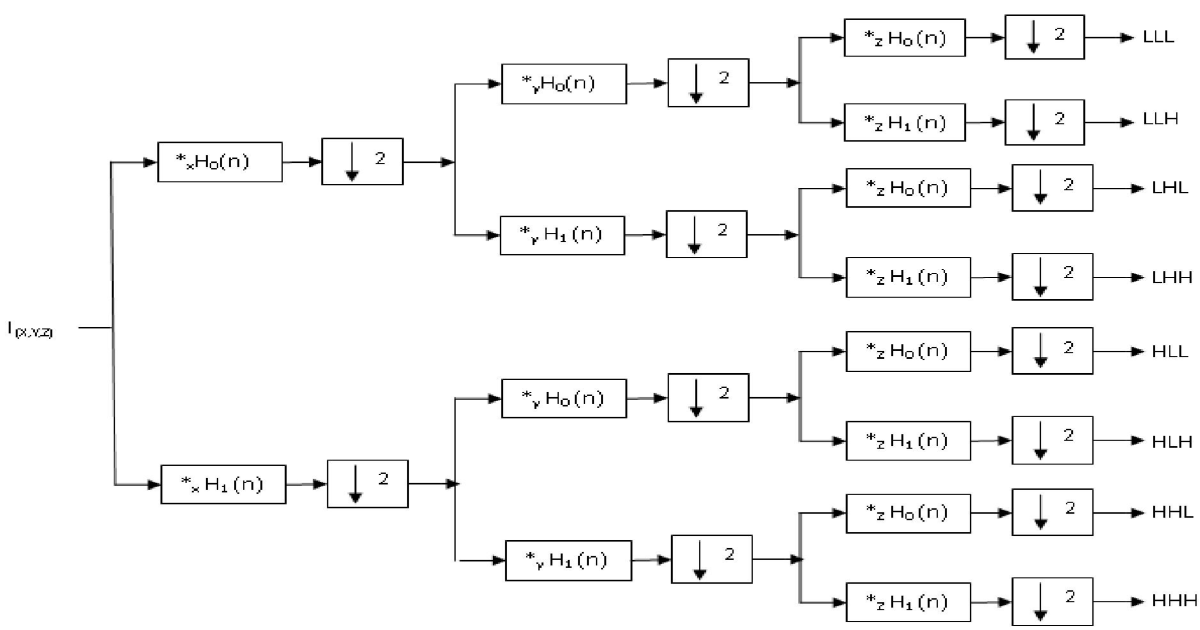


Fig. 2 : Analysis Filter Bank of 3-D DWTFig.3 Synthesis Filter Bank of 3-D DWT

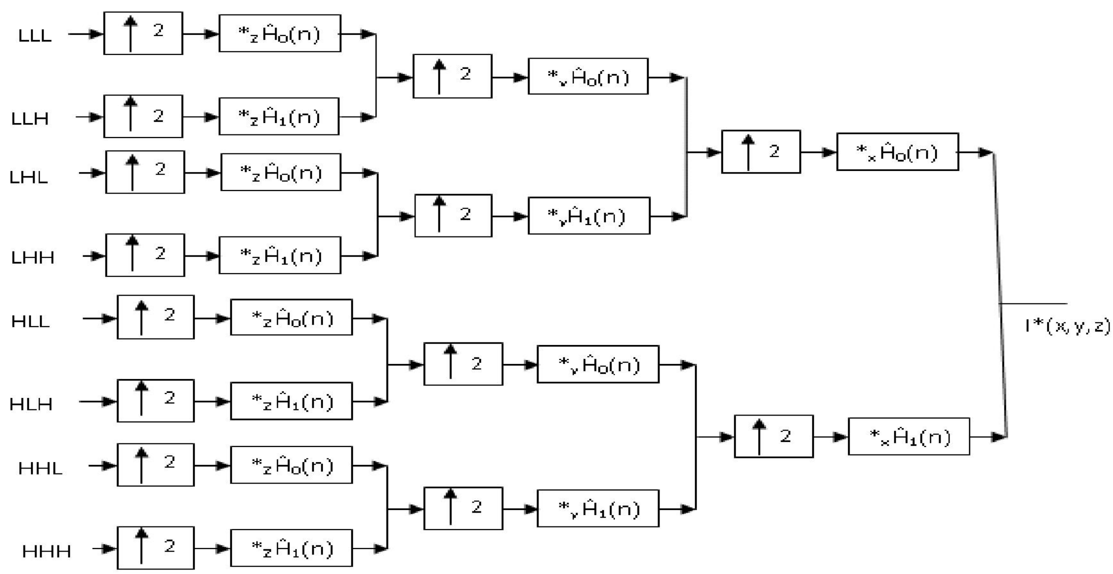


Fig.3 : Synthesis Filter Bank of 3-D DWT

III. Proposed Method

In recent years, several 3-D watermarking techniques have been proposed for gray scale and colour images for authentication, copyright protection, finger printing and ownership assertion. Here, we propse a novel blind watermarking scheme based on wavelet transform for spectral images. Designing of blind watermarking i.e. extracting the watermark without original image and original watermark is a very difficult task [9]. In this paper, blind watermarking scheme has been employed. Watermark is extracted with the help of a secret key.Mid frequency bands are selected to embed the secret message in order to compromise between imperceptibility and robustness. Watermark embedding and extraction processes are explained in the following sections.

a) Embedding Algorithm

One level 3-D DWT is applied on the spectral image (X). It decomposes the cover image into eight subbands ($X_{lll,}$ $X_{llh,}$ $X_{lhl,}$ $X_{lhh,}$ $X_{hll,}$ $X_{hlh,}$ $X_{hhl,}$ X_{hhh}). One level 2-DDWT is applied on the dual binary watermark (W_1 & W_2). It decomposes each watermark into four subbands($W_{ll,}$ W_{lh} W_{hl} W_{hh}).Singular Value Decomposition (SVD) is applied on the selected mid frequency subbandsof a cover image ($X_{llh,}$ $X_{lhl,}$ $X_{hlh,}$ X_{hhl}) and watermark image ($W_{lh1,}$ $W_{hl1,}$ $W_{lh2,}$ W_{hl2}). Eigen valuesof the selectedsubband (σ_{llh}, σ_{lhl}, σ_{hlh}, σ_{hhl}) is modified with the Eigen values of mid frequency subbands of binary watermark ($\sigma_{lhw1,}$ $\sigma_{hlw1,}$ $\sigma_{lhw2,}$ σ_{hlw2}) which is multiplied with appropriate strength factor (α). Inverse SVD is applied on the new Eigen values (σ^*_{llh}, σ^*_{lhl}, σ^*_{hlh}, σ^*_{hhl}) to obtain the modified subbands(X^*_{llh}, X^*_{lhl}, X^*_{hlh}, X^*_{hhl}) and inverse 3D-DWT is applied on the modified subbands to obtain the watermarked image (X^*). PSNR value is calculated for the watermarked image. Watermark embedding scheme has been shown in Figure 4.

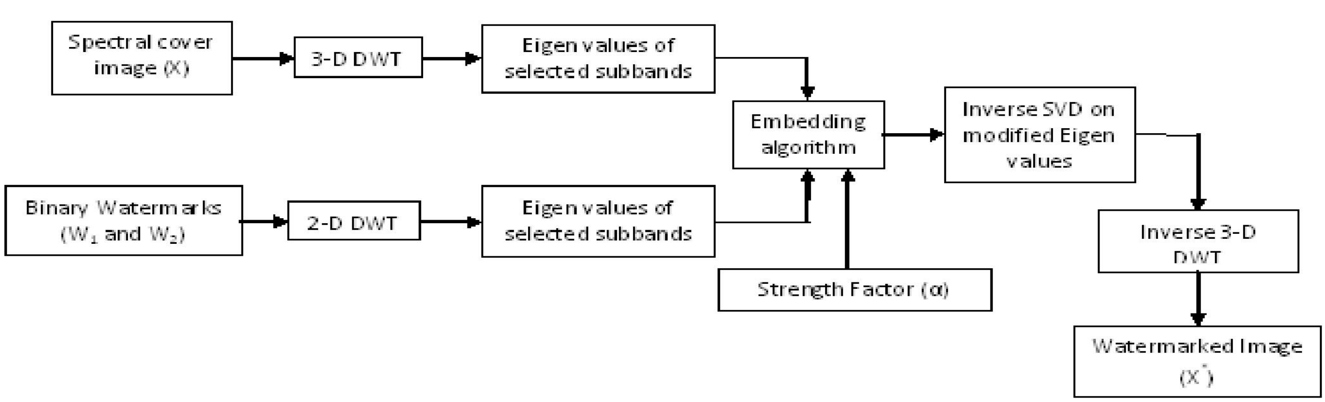


Fig.4 : Watermark Embedding Algorithm

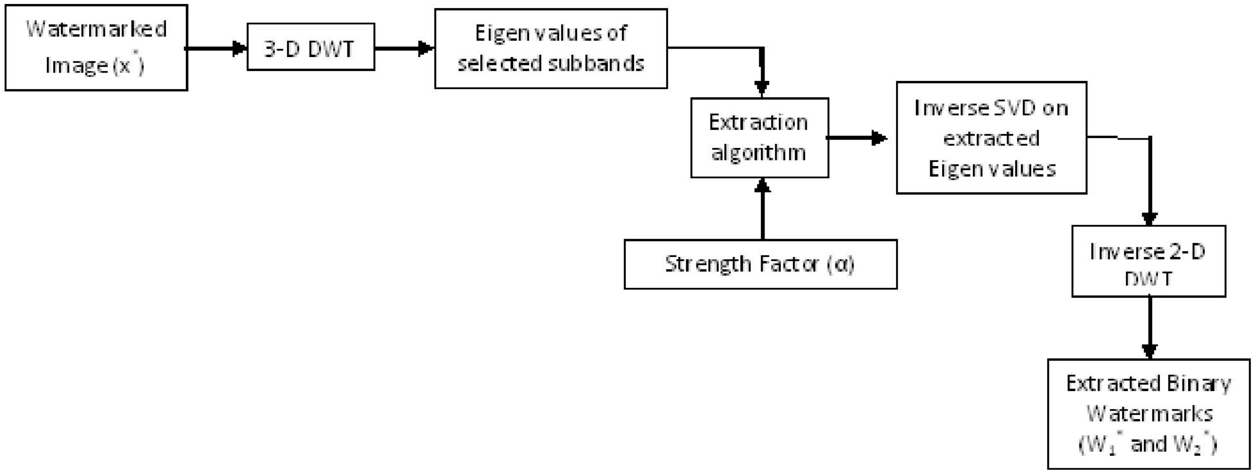


Fig.5 : Watermark Extraction Algorithm

b) Extraction Algorithm

One level 3-D DWT is applied on the Test image (X^*). It decomposes the cover image into eight subbands ($X^*_{lll}, X^*_{llh}, X^*_{lhl}, X^*_{lhh}, X^*_{hll}, X^*_{hlh}, X^*_{hhl}, X^*_{hhh}$). Singular Value Decomposition (SVD) is applied on the selected mid frequency subbands of a test image ($X^*_{llh}, X^*_{lhl}, X^*_{hlh}, X^*_{hhl}$). Eigen values of mid frequency subbands of binary watermark ($\sigma^*_{lhw1}, \sigma^*_{hlw1}, \sigma^*_{lhw2}, \sigma^*_{hlw2}$) is recovered from the Eigen values ($\sigma^*_{llh}, \sigma^*_{lhl}, \sigma^*_{hlh}, \sigma^*_{hhl}$) of selected subband($X^*_{llh}, X^*_{lhl}, X^*_{hlh}, X^*_{hhl}$) of test image with the help of strength factor (α). Inverse SVD is applied on the obtained Eigen values ($\sigma^*_{lhw1}, \sigma^*_{hlw1}, \sigma^*_{lhw2}, \sigma^*_{hlw2}$) to Recover the subbands and inverse 2D-DWT is applied to reconstruct the watermark (W^*). Normalised Correlation of recovered watermark is calculated on comparing with original watermark. Watermark extraction scheme has been shown in Figure 5.

IV. Experimental Results

The experiments were performed on two different hyperspectral natural images [14]. Natural image of the size 256x256x30 and two binary watermarks of size 32x32 were taken to evaluate the proposed algorithm. A single level 3-D DWT was taken on the host image and single level 2-D DWT was taken on the two binary watermarks. Eigen values of selected mid frequency subbands ($\sigma_{llh}, \sigma_{lhl}, \sigma_{hlh}, \sigma_{hhl}$) of the original image and two binary watermarks ($\sigma_{lhw1}, \sigma_{hlw1}, \sigma_{lhw2}, \sigma_{hlw2}$) were obtained by applying SVD. Eigen values of binary watermarks were multiplied with the strength factor (α=10 for scene2) and modified with the Eigen values of selected subbands. Watermarked image was obtained by applying inverse SVD and inverse 3-D DWT to the modified sub bands. Figure 6 illustrate the Host image, Original watermarks w_1 and w_2, Watermarked image and recovered watermarks w^*_1, w^*_2.

Figure 6 : (a) Host image (b) Original Watermark (c) Watermarked Image (d) Extracted Watermark

PSNR value of the watermarked image is calculated by using the equations (1) and (2).

$$PSNR = 10\log\frac{255^2}{MSE} \qquad (1)$$

$$MSE = \frac{1}{m*n*l}\sum_{x=1}^{m}\sum_{y=1}^{n}\sum_{z=1}^{l}\left(f(x,y) - \hat{f}(x,y)\right)^2 \qquad (2)$$

where m, n, and l are dimensions of the image $f(x,y)$ and $\hat{f}(x,y)$ are the pixel values of the original and watermarked image. Normalized Correlation (NC) between original watermark and extracted watermark from the test image is calculated by using the equation (3).

$$NC = \frac{\sum_i\sum_j\sum_k p_{ijk}\, p_{ijk}^*}{\sum_i\sum_j\sum_k (p_{ijk})^2} \qquad (3)$$

PSNR value of the watermarked image and Normalised Correlation of recovered watermark are shown in Table 1.

Table 1: PSNR Value of Watermarked Image and NC

	PSNR Value in dB	Normalised correlation	
		W_1^*	W_2^*
Natural Image I	69.7494	1	1
Natural Image II	57.7162	1	1

Any watermarking system should be robust against various image processing attacks. It should not be removable by unauthorized users and it should not degrade the quality of the images. There are many attacks against which image watermarking system could be judged. The attacks include JPEG compression, Histogram equalization, various filtering operations like, average filtering, median filtering, addition of noise like Salt and Pepper noise, Gaussian noise, speckle noise and Poisson noise, cropping, various angles of rotation and so on. These attacks are applied to the watermarked images to evaluate recovery process. Table 2, 3 and 4 shows the PSNR value and NC under various attacks. Table 5 shows the extracted watermark image under various attacks from Natural Image I.

Table 2 : PSNR Value of Watermarked Image Under various Attacks.

Image Type	PSNR in db of Watermarked image under various Attacks					
Histogram equalization						
Natural Image I	50.247					
Natural Image II	51.778					
JPEG Compression						
Natural Image I	36.0904					
Natural Image II	36.0912					
Low pass filtering						
	3x3	5x5	7x7	9x9	11x11	15x15
Natural Image I	72.0312	72.3203	72.4340	72.5012	72.5499	72.6231
Natural Image II	58.4857	58.5827	58.6308	58.6664	58.6968	58.75
Median filtering						
	3x3	5x5	7x7	9x9	11x11	15x15
Natural Image I	71.61	72.2512	72.4937	72.6184	72.9642	72.7934
Natural Image II	58.3489	58.5241	58.6099	58.6576	58.6923	58.7413
Angle of rotation						
	15	30	45	60	75	90
Natural Image I	71.2821	71.6741	71.7716	71.7003	71.2946	69.7760
Natural Image II	58.4666	58.8021	58.9701	58.8654	58.6201	57.7558
Addition of noise						
	Gaussian	Salt&Pepper	Speckle		Poisson	
Natural Image I	69.5155	67.5837	75.1603		75.1603	
Natural Image II	64.2688	63.0395	63.9917		64.0872	

Table 3 : Normalised Correlation of Recovered Watermark (W_1) under various Attacks.

Image Type	Normalised Correlation of Recovered Watermark (W_1) under various Attacks					
Histogram equalization						
Natural Image I	0.9864					
Natural Image II	0.9653					
Cropping						
Natural Image I	0.9855					
Natural Image II	0.9716					
JPEG Compression						
Natural Image I	0.9307					
Natural Image II	0.9307					
Low pass filtering						
	3x3	5x5	7x7	9x9	11x11	15x15
Natural Image I	0.9582	0.9473	0.9429	0.9399	0.9378	0.9355
Natural Image II	0.9598	0.9486	0.9416	0.9389	0.9374	0.9359
Median filtering						
	3x3	5x5	7x7	9x9	11x11	15x15
Natural Image I	0.9697	0.9548	0.9471	0.9418	0.9382	0.9349
Natural Image II	0.9706	0.9579	0.9470	0.9420	0.9393	0.9365
Angle of rotation						
	15°	30°	45°	60°	75°	90°
Natural Image I	0.9531	0.9473	0.9471	0.9417	0.9382	0.9443
Natural Image II	0.9594	0.9546	0.9481	0.9428	0.9406	0.9443
Addition of noise						
	Gaussian	Salt&Pepper	Speckle	Poisson		
Natural Image I	0.9757	0.9771	0.9782	0.9782		
Natural Image II	0.9658	0.9672	0.9664	0.9680		

Table 4 : Normalised Correlation of Recovered Watermark (W_2) under various Attacks.

Image Type	Normalised Correlation of Recovered Watermark (W_2) under various Attacks					
Histogram equalization						
Natural Image I	0.9864					
Natural Image II	0.9653					
Cropping						
Natural Image I	0.9843					
Natural Image II	0.9698					
JPEG Compression						
Natural Image I	0.9307					
Natural Image II	0.9307					
Low pass filtering						
	3x3	5x5	7x7	9x9	11x11	15x15
Natural Image I	0.9576	0.9477	0.9429	0.9401	0.9378	0.9355
Natural Image II	0.9590	0.9493	0.9413	0.9394	0.9374	0.9360
Median filtering						
	3x3	5x5	7x7	9x9	11x11	15x15
Natural Image I	0.9697	0.9550	0.9472	0.9418	0.9383	0.9350
Natural Image II	0.9704	0.9583	0.9469	0.9423	0.9393	0.9365
Angle of rotation						

	15°	30°	45°	60°	75°	90°
Natural Image I	0.9531	0.9473	0.9467	0.9415	0.9383	0.9443
Natural Image II	0.9594	0.9544	0.9479	0.9428	0.9404	0.9443
Addition of noise						
	Gaussian	Salt&Pepper	Speckle	Poisson		
Natural Image I	0.9759	0.9772	0.9782	0.9782		
Natural Image II	0.9659	0.9672	0.9664	0.9680		

Table 5 : Image Results of Recovered watermarks from Natural Image I

Addition of Noise				JPEG Compression	Histogram Equalization
Gaussian	Salt and Pepper	Poisson	Speckle	QF=10	
0.9757,0.9759	0.9771,0.9772	0.9795,0.9795	0.9782,0.9782	NC=0.9307	0.9864,0.9864
Low Pass Filtering					
3x3	5x5	7x7	9x9	11x11	15x15
0.9582,0.9576	0.9473,0.9477	0.9429,0.9429	0.9399,0.9401	0.9378,0.9378	0.9355,0.9355
Median Filtering					
3x3	5x5	7x7	9x9	11x11	15x15
0.9697,0.9697	0.9548,0.9550	0.9471,0.9472	0.9418,0.9418	0.9382,0.9383	0.9349,0.9350
Rotation					
α=0.25	α=0.5	α=0.75	α=1	α=15	α=30
0.9843,0.9845	0.979,0.9782	0.9711,0.9709	0.9857,0.9857	0.9531,0.9531	0.9473,0.9473
α=45	α=60	α=90	α=180	α=270	Cropping
0.9471,0.9461	0.9417,0.9415	0.9443,0.9443	1,1	0.9443,0.9443	0.9855,0.9843

The proposed watermarking scheme is compared with existing recently published papers by Arto Kaarna *et al* [15], Long Mal et[16],the results are shown in Tables 6 and 7.

Table 6 : Comparison of PSNR value of Watermarked Image in Proposed method and existing methods

Methods	PSNR in dB
Arto Kaarna et al	38.95
Long Mal et al	48.74
Proposed method	69.74

Table 7 : Comparison of Proposed Method and Existing Methods under Image Processing Attacks

LPF	Arto Kaarna et al	Long Mal et al	Proposed method	
			W_1	W_2
3 * 3	0.66	0.73	0.9582	0.9576
5 * 5	0.43	0.56	0.9473	0.9477
7 * 7	0.30	0.43	0.9429	0.9429

Global Journal of Researches in Engineering (I) Volume XI Issue VII Version I December 2011

V. Conclusion

Three Dimensional mesh watermarking is an interesting and promising research area. In this paper, a novel blind multiple watermarking algorithm based on 3-D DWT for copyright protection of spectral images has been proposed. Dual binary watermarks have been proposed to embed in mid frequency bands to increase the robustness against attacks and to improve imperceptibility. Extracting the secret message without the help of original image and original watermark is a very difficult task. We extracted dual binary watermarks with the help of details of subband selection and strength factor only. Experimental result shows that the proposed scheme achieves very high imperceptibility and robustness against various image processing attacks like LPF, Median Filtering, JPEG Compression, histogram equalization, cropping, various angle of rotation and addition of noise like Gaussian, Salt & Pepper, Speckle, Poisson etc.

References Références Referencias

1. Long Ma, Changjun Li and Shuni Song, "Digital Watermarking of Spectral Images Using DWT-SVD", *Proceedings ofInternational Conference on Communications, Circuits and Systems,*2006. vol. 1, pp.15-18,June
2. ChirawatTemi, SomsakChoomchuay, and AttasitLasakul "A Robust Image Watermarking Using Multiresolution Analysis of Wavelet," *Proc.of International symposium on communication and information technologies*, vol. 2, pp. 623-62, 2005.
3. S.RavalMehul "Discrete Wavelet Transform Based Multiple Watermarking Scheme", proc. *IEEE Conference on Convergent Technologies for Asia-Pacific Region*,vol.3, pp. 935 – 938, 2003.
4. XiaoYuan Yang, KeNiu, YiJie Wu, Wei Zhang, "Robust Watermarking for Video Based on Quasi-3D DWT/DCT" *Proc. of The IET International Conference on Wireless Mobile and Multimedia Networks (ICWMMN 2006),*pp. 1-4,November 2006, Hangzhou, China.
5. Yonggang Fu, "Robust image watermarking scheme based on 3D-DCT", *Proc. of Sixth International Conference on Fuzzy Systems and Knowledge Discovery,*vol. no. 5, pp. 437-44, 2009.
6. Hongmeiliee, NuoeChen,Hwang, " A Robust 3D Wavelet Video Watermarking", *IEEE Trans. on Image Processing* ,vol. 32, pp. 208–212, October 2003.
7. Dumitru Dan Burdescu, Liana Stanescu, Anca Ion and RazvanTanasie, "A New 3d Watermarking Algorithm", *Proc.of IEEE Technical Conference (TENCON* 2009), pp.381,May 2008.
8. Gaurav Bhatnagar and Balasubramanian Raman, "A new robust reference watermarking scheme based DWT-SVD," *Computer Standards and Interfaces*, vol. 31, no. 5, pp. 1002-1013, September 2009.
9. *Ce Wang, Chao Zhang, and PengweiHao*, "A Blind Video Watermark Detection Method Based on 3D-DWT Transform" *Proc.of IEEE 17th International Conference on Image Processing,*pp. 3693-3696, September, 2010, Hong Kong.
10. Saoussen Ben Jabra and EzzeddineZagrouba, "A New Approach of 3D Watermarking Based on Image Segmentation", *Proc. of IEEE symposium on Computers and Communications*, pp. 994-999, July 2008.
11. S. Mallat, "A Wavelet Tour of Signal Processing" *Academic Press*, New York, 1999.
12. *http://taco.poly.edu/WaveletSoftware/standard3D.html*
13. Evelyn Brannock, Michael Weeks, Robert Harrison "The Effect of Wavelet Families on Watermarking" *Journal of Computers*, vol. 4, no. 6, pp. 554-566,June 2009.
14. Nascimento, S.M.C., Ferreira, F., and Foster, D.H. (2002). Statistics of spatial cone-excitation ratios in natural scenes. *Journal of the Optical Society of America A*, 19, 1484-1490.
15. Arto Kaarna, Pekka Toivanen, "Digital watermarking of spectral images in PCA/Wavelet-transform domain", *proc. of IEEE symposium on Geoscience and Rem ote Sensing,* vol.6, pp no. 3564 – 3567, 2003.
16. Long Ma, Changjun Li, and Shuni Song, "Digital watermarking of spectral images using SVD in PCA-transform domain", proc. of IEEE International Symposium on Communications and Information Technology, pp no.1489 – 1492, 2005.

Global Journal of researches in engineering
Numerical Methods
Volume 11 Issue 7 Version 1.0 December 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 2249-4596 & Print ISSN: 0975-5861

Cohesion Metric and Its Relation with Coupling: A Class Level Variable Assessment Approach

By Dr.M.V.Vijaya Saradhi, Dr.B.V.Ramana Murthy
Aurora's Engineering College (AEC), Bhongiri, Nalgonda (Dist), Andhra Pradesh, India

Abstract - Cohesion Metrics is an important technical development which helps for the better Assessment Of class cohesion that is the measurement of relatedness among members In a class. The higher this relatedness is the best the performance will be. Hence this is an important feature of Object oriented systems. Present paper presents an advanced metrics. The pervious metrics have got few limitations. Whereas this advanced metrics considers few more characteristics of class Cohesion. This is based on common object parameters. Moreover this metric is statistically advanced and measures cohesion by observing the relationship between cohesion Metric and Coupling. This is cohesion measurement tool for Java and it is tested on several systems. These systems which are used for the experiment are deferent in size and domain. This test proved that this extended metrics captures other pairs of relatedness among members in a class and also the correlation between cohesion Metric and coupling.

Keywords : Metrics, Cohesion, LCOM, CBO, MIC, DCD, DCDE, DCI, TCC, LCC,MIC,DCIE, Coupling, Class level variables, Spearman correlation,MPC,CLC,OLC.

GJRE-I Classification : FOR Code: 080309

COHESION METRICAND ITS RELATION WITH COUPLINGA CLASS LEVEL VARIABLE ASSESSMENT APPROACH

Strictly as per the compliance and regulations of :

Cohesion Metric and Its Relation with Coupling: A Class Level Variable Assessment Approach

Dr.M.V.Vijaya Saradhi[α], Dr.B.V.Ramana Murthy[Ω]

Abstract - Cohesion Metrics is an important technical development which helps for the better Assessment Of class cohesion that is the measurement of relatedness among members In a class. The higher this relatedness is the best the performance will be. Hence this is an important feature of Object oriented systems. Present paper presents an advanced metrics. The pervious metrics have got few limitations. Whereas this advanced metrics considers few more characteristics of class Cohesion. This is based on common object parameters. Moreover this metric is statistically advanced and measures cohesion by observing the relationship between cohesion Metric and Coupling. This is cohesion measurement tool for Java and it is tested on several systems. These systems which are used for the experiment are deferent in size and domain. This test proved that this extended metrics captures other pairs of relatedness among members in a class and also the correlation between cohesion Metric and coupling.

Keywords : Metrics, Cohesion, LCOM, CBO, MIC, DCD, DCDE, DCI, TCC, LCC,MIC,DCIE, Coupling, Class level variables, Spearman correlation,MPC,CLC,OLC.

I. Introduction

In Software Engineering there is a raising importance to Metrics and cohesion which are interlinked [28, 27]. Metrics are very much useful in assessing several software characteristics such as complexity, cohesion, coupling and size. Cohesion helps for the better performance and metrics to decide the level of cohesion. So both are equally important in the field of software development. Hence research is going on to improve the cohesion as well as metrics which is a tool to measure it. Cohesion can be defined as the degree of relatedness among elements in a component. It was first introduced and utilized by Yourdon and Constantine in the context of traditional applications. They used it as a tool to estimate level of functional relationships of the elements in a module [30]. These modules are structured for different uses. Class cohesion is a very important feature of object oriented software. This feature of software is of great help to software developers and managers to improve the software quality during the development process.

Author [α] : Professor& HoD in the Dept. of IT at Aurora's Engineering College (AEC), Bhongiri, Nalgonda (Dist), Andhra Pradesh, India. (Telephone : +91-9849275094, E-mail : meduri_vsd@yahoo.co.in),
Author [Ω] : Professor in the Dept.of CSE at Jyothismathi College of Eng. & Tech, Shamirpet, R.R.Dist, Andhra Pradesh, India. (Telephone : +91-9247331980, E-mail : drbvrm@gmail.com)

There are three different types of cohesions. Functional, sequential and coincidental etc [19] is few among them. This feature of software is important because of its wide range uses. If the component has good range of cohesiveness among its members, it can be reused and maintained well [25,7,9,13]. With the help of cohesion one can evaluate the structure quality of class. A class which has good cohesiveness can not be broken easily. It can be used to find out the poorly structured classes. In addition while designing a class cohesion is of great assistance to bring out the best [24]. As a reference we can observe Grady Booch views. He describes that high functional cohesion can be achieved it the elements of a component has good cohesiveness among them which provides some well bounded behavior [8]. This can be brought with single logical function. If all the parts of a class contribute to this single logical function high degree of cohesiveness can be achieved. In contrast if the members are desperate and non related then the coherence will be very low. Since cohesion has such a large scale importance, much number of metrics was proposed to estimate it in object oriented systems. Most of these metrics have been experimented and widely discussed in literature [19,12, 16, 18, 11, 5]. From the above paragraph one can understand that class cohesion is a very important feature of object oriented systems. Many successful metrics have been elaborated and categorized in [10]. These metrics estimate the cohesiveness according to them relatedness among the elements of the class. They count two features; first one is the number of instance variable used by methods and second is the number of method pairs that share instance variables. Though many metrics were proposed in literature they were not total successful in finding out the cohesiveness of classes [22, 13, and 3]. There are some basic reasons for this failure. Some of them are that they do not undertake few features of classes that are the size of cohesive components and relatedness among elements as said in [13].

Few other serious drawbacks which above stated metrics are that they are based only on few categories like instance variable and number of method pairs as stated in [23]. This often leads to wrong estimation of the cohesiveness among members in a component. So the previous metrics face a serious problem when the systems work in functional relationship. In this category, cohesiveness can not be

Global Journal of Researches in Engineering (I) Volume XI Issue VII Version I December 2011

decided by the above said connections but has to be done with the help of the relationships that may exist among methods. If the same old metrics is followed many features of class cohesion will not represented. Hence we believe that class cohesion will not be exact if it does not go beyond above cited categories. Research on source code on several systems tells us that several methods functionally attached even without sharing any instance variable and these can't be divided into different classes. As such the focus is to be extended by taking into account different ways of estimating class cohesion and should not be restricted to any two. First development of the metrics is that connections among class methods will be considered [5, 6]. These systems help to find out much number of pairs of related methods which are not found by previous cohesion metrics. This criterion proved successful when it was tested on several Java systems. In these experiments they gave correct statistical information. In these last years many such developments were brought in. of them [17], concentrates on tradition of maintainability [Zho 03] and [1] on the intimation of mistakes, and [23] on examine the relationship between cohesion and coupling in the one hand and the relationship between cohesion and changeability in the other hand. The area has a raising importance in these last years.

Till now cohesion metrics were limited to object oriented systems but we used this information for other systems also [5, 6]. The previous metrics were based only on instant variables and number of methods pairs. But this information is very primitive to depend on. With this one cannot give good cohesion results. As such research was undertaken to find out other categories which are more authentic and will be helpful to give exact cohesive results. This paper gives an extension of two methods that is DCD and DCI which was proposed in [5, 6]. A new criterion of common object parameters was introduced to calculate cohesion levels. This criterion tells us that two methods of a given class can very well share same object passage in a parameter without being correlated. It does not need to share a method or an instance variable to get connected. So depending on instance variable take us wrong, where as this object passage can present authentic results. Further more it was discovered that in the object context objects themselves collaborate to accomplish a given task. As such certain design principles [24] like design patterns among others and classes play an important role in successful completion of a given job. This collaboration can be located at two levels. One at the group of objects belonging to different classes, second at the collaboration between groups of methods with in a unique given class. This last kind of collaboration can be observed among other things also. These are used in the form of instance variable or passes as arguments at the method level, public in particular. In this kind of collaboration cohesion helps to assign responsibilities to classes [24] in a cohesive manner. Form the above conclusions and according to the experiments done since 2003, this new category to estimate cohesion levels is more dependable. This is proved after conducting many experiments on systems. These clearly show that the extended cohesion metrics based on the addition of the proposed category captured more pairs connected methods than the old metrics DCD and DCI did. These experiments that were done on several systems gave correct, authentic, statistical results.

Software Engineering developers state that there is a correlation between cohesion and coupling. They state that if the cohesion is high, coupling will be low and vice versa [24, 28, 27]. But this notion was not proved by any empirical work. Many experiments were done to bring out the realities and they could only present the necessity for a refined cohesion metrics. They failed because of the limitations of the previous metrics [23]. This paper presents such an extended cohesion metrics. This technique is tested to find out the truth in the relationship between extended cohesion and coupling. Here it was decided to use both the old and new cohesion metrics. The experiment shows that there is a significant correlation between our cohesion metrics and the considered coupling metrics [CBO – Coupling between objects] of Chidamber et al [14, 15]. But the correlation degrees between them varied much when they are presented by the considered cohesion metrics. The empirical experiments as well as the obtained results are discussed in section [7]. Our final goal is to validate the new cohesion metrics as a good indicator for changeability, testability and for many more things. This will be dealt seriously in the future research.

The following paper is arranged in the following way: Section 2 presents an overview of major class cohesion metrics. Section 3 gives the idea of coupling between objects and few important coupling metrics. Section 4 gives some related work which focuses the relatedness of object-oriented metrics and few quality characteristics. Section 5 redefines class cohesion that was proposed depending on the new criterion that we introduced in this paper. Section 6 gives the first step of the experiment that was done (statistic test). Section 7 gives the empirical investigation that we have done to find out the relationship between cohesion and coupling. Finally, conclusions and future work ideas are presented in section 8.

II. Class Cohesion Metrics

Classes are that primary units of object oriented software. In these classes we find cohesion. It is an important feature in software design. The higher the cohesion better the performance will be. A class will have best cohesiveness as stated in [13] if large number of its instance variables are used by a method (LCOM5 [19], Coh [10]), or a larger number of methods pairs share instance variables (LCOM1 [14], LCOM2 [15], LCOM3 [25], LCOM4 [21], Co [21], TCC and LCC [7], DC [4]). Other than the above two, sometimes these metrics also observe the relatedness between the methods to assess cohesion. To get such good cohesion software developers struggle hard at the design phase of classes. If the modeling of these classes is not at its best then cohesion will automatically go down. Hence after designing the classes one would like to assess class cohesion. To assess them many metrics have been proposed in literature. Different authors have defined class cohesion by proposing their cohesion metrics. These cohesion metrics have been presented in detail and are categorized in [10]. One such cohesion metric is LCOM. It is lack of cohesion in methods. It is a metric which is defined by Chidambar and Kemeter [14,15,16]. This metric stands as a role model for many other proposed cohesion metrics. In addition to other proposed metrics, few others tried to redefine LCOM itself. All this cohesion metrics have come into literature to find out the class cohesion in objected oriented systems.

a) Coupling Between Classes

Cohesion metrics measure cohesion between members. Whereas coupling measures the strength of a connection between two modules. Stevens & al [29] explain coupling as one which assesses the strength of the association which is formed by the relatedness between two modules. Coupling between classes helps to assess in which proportion an entity uses other entities. There are both positive and negative effects of the coupling. To speak of the positive low coupling between components helps to minimize interdependencies and gives a chance for evolution [24, 28, 27]. If the modules are structured with low coupling then the complexity of a system can also be minimized. On the negative side because of high coupling module becomes complex. Because of this complexity module will be tough to understand, tough to detect and correct errors, and to change that module. After analyzing these effects one can understand that low coupling is preferable. So it is always encouraged by software engineers. After much research it was discover that with the help of these coupling metrics the maintainability of the oo systems can be easily imagine. Moreover there are few empirical insufficiencies due which there is lot of importance to coupling metrics for the prevision in maintainability [17]. Among coupling metrics, we cite CBO (Coupling between Objects) of Chidamber and Kemerer [15], MPC (Message-Passing Coupling) and DAC (Data Abstraction Coupling) of Li and Henry [25, 26] or OLC (Object Level Coupling) and CLC (Class Level Coupling) of Hitz and Montazeri [21]. Brian & al. counted 23 coupling metrics [9]. For this research work CBO was used which is proposed by [15]. Largely known as a good coupling metric between classes. In our future work, we plan on extending our study to integrate other coupling metrics.

III. Related Work

During the research of the last three decades many software metrics like coupling, cohesion, and complexity were proposed to calculate certain aspects like maintainability, testability, changeability and many more things. But in finding out the quality these were not totally successful. Few reasons are that they are to some extent based on the little understanding of the empirical hypothesis. In addition to it all these proposed metrics can be used to determine any aspect of quality. There is no particular division that this metrics is for the assessment of this quality. As such there is no information about which metric is most suitable to assess a quality. This is the second major problem. Through the research it was also found that of all the metrics only six are efficient and can present sufficient information to depend on. Whereas all the remaining metrics can't give any extra information and they just correspond to subsets of the retained metrics. These drawbacks were discussed by many engineers. Dag &al opines in [17] about the prediction of maintainability. He argues that because of the empirical insufficiencies that these metrics have got the assessment will not be very much dependable. Of all the research papers Dagpinar & al [17] paper presents in new development. In this paper they opines that inheritance cohesion and indirect exportation coupling are not the right factors by which maintainability can be measured well. They advise taking to consideration metrics of size and direct coupling importation which can give good results. A lot of research was undertaken to find out the correlation among the proposed coupling metrics and how much they are prone to fault results. One such research was done by Aggar & al [1]. In prediction model of [2] it was clearly proved that these metrics are very much prone to fault reasons. Zhou & al [31] also undertake similar work to find out the relationship among design metrics (CBO ,WMC, RFC, LCOM etc) and fault proneness when taking fault severity into account. This was understood after conducting a thorough research on many number of oo coupling metrics [1]. This study focuses to find out the best methods according to the given data. After all the research about the relationship between coupling and cohesion leads us to confusion that their may exist a connection between cohesion and for example maintainability, testability as well as fault proneness. Any how a lot more is necessary to find out the direct relationship that may exist between cohesion and above said attributes. This final aspect will be the subject of further research and is out of discussion of the present paper.

Global Journal of Researches in Engineering (I) Volume XI Issue VII Version I December 2011

IV. Class Cohesion Assesments : a New Measure

Class cohesion at the beginning of this paper is defined as the relative number of related members in a class. This definition is redefined twice in this paper so as to get best methodology which can give authentic assessments. As a first step, two more strategies were added to the relative number. First one is the extension of the methods invocation criteria and indirect utilization of the characteristics explained by Bienan & al in [17]. This idea was extended to the methods invocation criterion as well. After defining the new methodology was tested on several systems [5,6]. This shows a lot of improvement in assessment than the first noted procedure. From the results one can observe that new criteria and the extension of the original criteria are capable of finding out more pairs of connected methods which were not found by the old methods. To come to this conclusion the procedure was tested on several systems. These experiments gave a chance to observe the code of some program. With this code observation and from the obtained results it was found that methods of a class may be functionally, related in other ways. In addition some facts about attributes were also found. From the experiment it was known that attributes, on which first development is dependent on, are not unique to any method. These attributes in reality are reference attribute. Such a one which is not unique but shared is used to assess class cohesion. Many systems were analyzed to come to the above conclusion and observations say that more than 20% of the attributes were reference attributes. This is very much possible oo systems because classes collaborate in accordance with the respective responsibilities so as to finish a given task. Reference attributes are utilized to confirm the needed visibility between objects [24]. Because of these drawbacks we tend to improve this second definition also. We tried to bringing in criteria which will not be primitive, shared but will be more authentic. For this a new criteria that is common objects parameter is introduced. In the following page we explain this and the metrics which work with this new methodology. Our first and second procedures to assess cohesion were already talked about and also tested in the previous papers [5, 6]. This newly introduced criterion is the one which will be prominently discussed in this paper. Both these ways have got many similarities. This new way to assess class cohesion is very much dependent on different connections that are present between its methods. All the three proposed criterions: Attributes Usage Criterion, Methods Invocation Criterion, and Common Objects Parameters will be utilized to find out the functional cohesion in a class. Class cohesion can be said as the connectedness of public methods of a class, with the help of functionalities utilized by its clients. The others methods of the class are included indirectly through the public methods.

a) Attributes Usage Criterion (UC)

Let us take a class C. Let A = {A1, A2,…, Aa} be the group of its characteristics and SPM ={M_1, M_2, …, M_n} be the group of its public methods. Let UCMi be the group of all the characteristics used directly or indirectly by the public method Mi. A characteristic is used directly by a method M_i, if the characteristic shown in the body of the method Mi. The characteristic is indirectly used by the method M_i, if it is used directly by another method of the class that is implored directly or indirectly by Mi. There are n sets UC_{M1}, UC_{M2}, …, UC_{Mn}. Two public methods M_i and M_j are directly related by the UC relation if $UC_{Mi} \cap UC_{Mj} \# \Phi$. It shows that there is at least one characteristic shared (directly or indirectly) by the two methods.

b) Methods Invocation Criterion (MIC)

Let us take a class C. Let SPM = {M1, M2, …, Mn} be the group of its public methods and PRM = {I1, I2, …, Ik} be the group of its other (private and protected) methods. Let SPMMi be the group of all the public methods of the class C, which are implored directly or indirectly by the public method Mi. A public method Mj is called directly by a public method Mi, if Mj is seen in the body of Mi. A public method Mj is indirectly called by a public method Mi, if it is called directly by another method of the class C that is implored directly or indirectly by Mi. There are n sets SPM_{M1}, SPM_{M2}, …, SPM_{Mn}. Let PRM_{Mi} be the group of all the other methods (private and protected) of the class C, which are implored directly or indirectly by the public method Mi. There are n sets PRM_{M1}, PRM_{M2}, …, PRM_{Mn}. Let $MIC_{Mi} = PRM_{Mi} \cup SPM_{Mi}$ be the group of all the methods of the class C, which are implored by the public method Mi. There are n sets MIC_{M1}, MIC_{M2}, …, MIC_{Mn}. Two public methods Mi and Mj are directly connected by the MIC relation if $MIC_{Mi} \cap MIC_{Mj} \# \Phi$. We also take it into account that Mi and Mj are directly related if $Mj \in MIC_{Mi}$ or $Mi \in MIC_{Mj}$.

c) Class level variables (CO)

Let us consider a class C. Let SPM = {M1, M2, …, Mn} be the group of its public methods. Let UCOMi be the group of all the parameters (of object type) of the method Mi. There are n sets UCOM1, UCOM2, …, UCOMn. Two public methods Mi and Mj are directly related by the UCO relation if $UCOMi \cap UCOMj \# \Phi$. From the above we understand that there is at least one parameter of object type that is utilized by the two methods.

d) Cohesion based on the direct relation

Two public methods Mi and Mj may be directly interlinked in different ways: they share at least one instance variable in common (UC relation), or get connected at least with another method of the same class (MIC relation), or share at least one object passed as argument (CO relation). In this context, the two methods may be directly interlinked by one or more criteria. It shows that the two methods are directly interlinked if: $UC_{Mi} \cap UC_{Mj} \# \Phi$ or $MIC_{Mi} \cap IM_{Mj} \# \Phi$ or $UCO_{Mi} \cap UCO_{Mj} \# \Phi$.Let us consider a class C with SPM = {M_1, M_2, …, M_n in character are directly connected. Let ED be the number of edges in the graph GD. The level of cohesion in the class C is dependent on the direct connection between its public methods is explained as: DC} the group of its public methods. The highest number of public methods pairs, is n * (n – 1) / 2. Let us take an undirected graph GD, in which vertices are the public methods of the class C, and there is an edge between two vertices if the methods which are equal DC_{DE} = |ED| / [n * (n – 1) / 2] [0,1]. DC_{DE} (as an extension of DC_D [5, 6]) presents the percentage of public methods pairs, which are directly (as defined below) connected. The Lack of Cohesion in the Class (LCC_{DE}) is than given by : LCC_{DE} = 1- DC_{DE} € [0,1].

e) Cohesion based on the indirect relation

Two public methods Mi and Mj can be indirectly connected if they are directly or indirectly related to a method Mk. The indirect relation, brought in by Bieman and Kang in [7], is the transitive closure of the direct relation. We use this idea in our method to mark the indirect related methods. Let us take now an undirected graph GI, where the vertices are the public methods of the class C, and there is an edge between two vertices if the methods are directly or indirectly connected (transitive closure of the graph GD). Let EI be the number of edges in the graph GI. The degree of cohesion in the class C in this case (direct and indirect relations) is said as: DCIE = |EI| / [n * (n – 1) / 2] Є [0,1]. DCIE (as an extension of DCI [5, 6]) presents the percentage of public methods pairs, which are directly or indirectly related. The Lack of Cohesion in the Class (LCCIE) is than given by: LCCIE = 1- DCIE Є [0, 1].

V. Experimental Study

Several systems were downloaded from the web to experiment on the new criterion. The goal was to achieve significant and general results. To collect the significant data was the main goal of these experiments. Hence many number of Java classes from different systems are taken. Through this experiment it was explored if the proposed criterion is statistically significant before more investigation. We extended the cohesion measurement tool (in Java) for Java programs, that we developed for [6], to automate the computation of our metrics (DCD, DCDE, DCI and DCIE). Many classes in the chosen systems have only one method or do not have any methods. These classes were taken as special classes and have not used for our measurements. All abstract classes are also not used. Overloaded methods within the same class were taken as one method. In addition to it, all special methods (constructors, destructors) were not used. We gathered the values for all the selected metrics from the test systems. For each metric, we calculated some descriptive statistics (minimum, maximum, mean, median, and standard deviation).

VI. Selected Systems

The experiment concerned more than 800 classes. The followed methodology and the obtained results are presented in the following sections. The selected systems are:

- System1 : JIU0.10 (Java Imaging Utilities) is a library in Java for the change, the edition, the analysis and the backup of pixels of image files (http://sourceforge.net/projects/jiu). This system consists of 180 classes.
- System2 : JIU0.11 (Java Imaging Utilities) is an improvement of the first system (http://sourceforge.net/projects/jiu) and consists of 191 classes.
- System3 : FujabaUML is a software development tool which helps for the easy betterment of UML and the progress with Java by adding plug-ins (http://www.fujaba.de). This system consists of 186 classes.
- System4 : Wbemservices is a Java open source implementation of Web Based Enterprise Management (WBEM) for commercial and non commercial applications. It is compiled of API, of servers, client applications and tools http://wbemservices.sourceforge.net/). It contains 463 classes.

Systems	Des. Stat	DC_D	DC_{DE}	DC_I	DC_{IE}
Jiu1	Mean	0.16027	0.17384	0.1922	0.2178
	Sdt.dev	0.13686	0.1378	0.1638	0.2178
Jiu2	Mean	0.2497	0.2635	0.3102	0.3350
	Sdt.dev	0.16466	0.1714	0.2292	0.2246
Fujaba	Mean	0.01597	0.05244	0.0207	0.0656
	Sdt.dev	0.01479	0.05861	0.0201	0.0739
WBEM	Mean	0.08138	0.2286	0.1013	0.2747
	Sdt.dev	0.14164	0.2051	0.1678	0.2332

Table 1: Average values of cohesion.

VII. Results

We assessed cohesion values for the 4 chosen systems. Table 1 gives the mean values of the metric for the chosen system. The results that we have got for DCDE et DCIE show clearly that they find more pairs of connected methods than DCD et DCI did.

Global Journal of Researches in Engineering (I) Volume XI Issue VII Version I December 2011

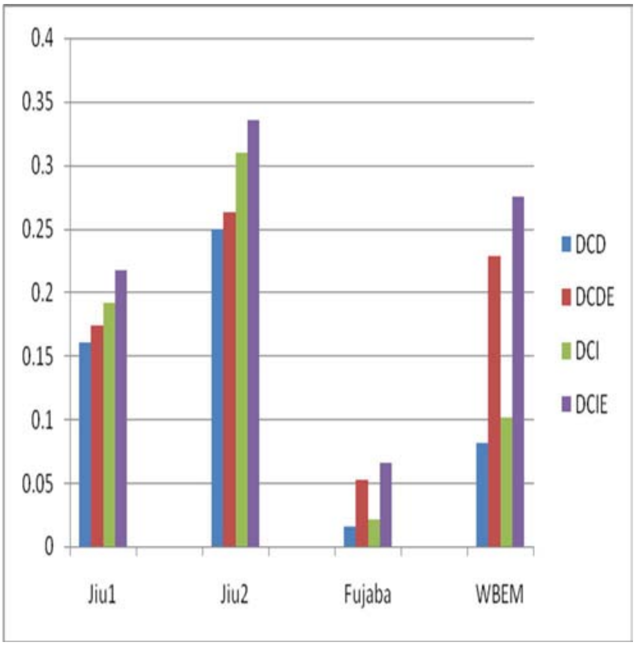


Fig 1: Representation and comparison of the average values of selected systems

From the above figures we can come to a conclusion that DCDE and DCIE are able to find out many other details of attributes of classes which are not found by other metrics. Through this research we would like to prove the importance of new criteria. Hence we are not going to evaluate the cohesion values of the selected systems. From the above given statistics in table 1, it can be said that these systems don't have cohesiveness.

VIII. Validation of the New Criterion

The goal of this chapter is to check the effects of DCD and DCDE on one side and the effects of DCI and DCIE on the other. The goal of this comparison is to find out if there is any difference brought by the introduced new criteria. Through this we would like to prove that DCDE and DCIE are much preferable to DCD and DCI. Because DCDE and DCIE help to find out more pairs of related methods. To prove our above said assumption we have under taken one statistical test: the PAIRED t-TEST [20]:

Let μ1 be the mean value of DCDE (or DCIE) and μ2 be the mean value of DCD (or DCI).

Below we give two statistical hypotheses :

- H0 : μ1= μ2 The metrics are equivalent.
- H1 : μ1> μ2 DCDE (or DCIE) is more significant than DCD (or DCI).

Let Diff be the value of (μ1- μ2). The above test is equivalent to:

- H0 : Diff = 0.
- H1 : Diff >0.

The test statistic is: Z =d/ [Sd / sqrt(N)]

With d : the mean value of sample Diff
Sd : the standard deviation of sample Diff and
N : the number of classes in sample Diff.

Tables 2 and 3 present respectively the comparison between DCD and DCDE on one side and DCI and DCIE on the other.

Systems	Des. Stat	DCD	DCDE	Diff	Z	Zα
Jiu1	Mean	0.16027	0.17384	0.01356	1.799	1.645
	Sdt .dev	0.13686	0.1378	0.01685		
Jiu2	Mean	0.2497	0.2635	0.0228	2.4635	1.645
	Sdt .dev	0.16466	0.1714	0.0207		
Fujaba	Mean	0.01597	0.05244	0.03646	2.6547	1.645
	Sdt .dev	0.01479	0.05861	0.05663		
WBEM	Mean	0.08138	0.2286	0.1472	4.7917	1.645
	Sdt .dev	0.14164	0.2051	0.1869		

Table 2 : Comparison between DCD and DCDE

The methodology consists on comparing Z, for each system, to a value Zα (the value of α is 0.05). If the value of Z is higher than Zα, we do not agree with hypothesis H0 : Diff = 0 and accept H1 : Diff > 0. In this case, the statistical test is significant and we can conclude that metric DCDE (or DCIE) is preferable than metric DCD (or DCI). This means that the added criterion is significant and allows capturing an additional aspect of classes' properties. We have taken data on the metrics from the selected systems and estimated Diff and Z for these systems. These observations are given in tables 2 and 3.

Systems	Des. Stat	DCI	DCIE	Diff	Z	Zα
Jiu1	Mean	0.1922	0.2178	0.0255	1.620	1.645
	Sdt .dev	0.1638	0.2178	0.0352		
Jiu2	Mean	0.3102	0.3350	0.02485	1.5498	1.645
	Sdt .dev	0.2292	0.2246	0.0358		
Fujaba	Mean	0.0207	0.0656	0.0448	2.6494	1.645
	Sdt .dev	0.0201	0.0739	0.0697		
WBEM	Mean	0.1013	0.2747	0.1734	5.7969	1.645
	Sdt .dev	0.1678	0.2332	0.1819		

Table 3 : Comparison between DCI and DCIE

They clearly show that, for the many tested systems Zα lower than Z. The systems for which Zα is higher than Z are the systems for which N is low. Through out the world, the results show that DCDE (or

DCIE) is preferable than DCD (or DCI). This statistical validation shows the applicability of the new cohesion criterion for finding new pairs of connected methods. The results that we have got prove that the extended cohesion metrics, based on the newly introduced criteria, find more pairs of connected methods than metrics DCD and DCI.

IX. The Relationship Between Extended Cohesion Assessment and Coupling

To validate our metrics we went for further experimentation. Software developers believe that cohesion and coupling are correlated. Though not proved it is said that coupling will be low when cohesion is high and vice versa [24, 28, 27]. Using our new criteria of cohesion we tried to know the facts about this belief. If the facts about this can be brought out, we can prove that new metrics on the new criteria is most successful way of assessment. Through these experiments we can bring out the relationship that the metrics can directly have with high level quality attributes like testability, changeability and maintainability. But this is a very beginning stage and no conclusions can be brought. We need more research to confirm the above relationship.

An empirical study

The experiment we performed considered six systems that vary in size (number of classes) and domain. The selected systems are (more than 500 classes):

- **System 1** : Gnujsp 1.0.1, GNUJSP is a free implementation of Java Server Pages of Sun (http://klomp.org/gnujsp). This system contains 56 classes.
- **System 2** : JIU 0.12, JIU (Java Imaging Utilities) is a library in Java for loading, editing, analyzing and saving pixels in image files (http://sourceforge.net/projects/jiu). This system has 77 classes.
- **System 3** : fujabaUml.4, FujabaUML is a software development tool allowing the easy extension of UML and Java development with the use of plug-ins (http://www.fujaba.de). This system contains 60 classes.
- **System 4** : jexcelapi 2.6, JExcelApi is a Java library that grants the possibility of reading, writing and modifying Microsoft Excel Worksheets (http://sourceforge.net/projects/jexcelapi). It contains 110 classes.
- **System 5** : moneyjar 0.8, Moneyjar is a Java library for financial applications. It simplifies treasury management, currency exchange, tax calculations and invoice management (http://sourceforge.net/projects/moneyjar). It contains 20 classes.
- **System 6** : wbemservices 1.0.0, Wbemservices is an open source Java implementation of Web Based Enterprise Management (WBEM) for commercial and non commercial applications. It is a project composed of APIs, of servers, of client applications and of tools (http://wbemservices.sourceforge.net/). This system contains 180 classes.

Experimental Process: First phase

We started our experiment to find out the relationship between cohesion and coupling. Of all the selective systems six were chosen and taken for the experiments. From these data is collected about the four cohesion metrics and also about the CBO metrics. The study conducted with this data proves that there is a definite correlation between cohesion and coupling. Here after it is not a belief but fact that when cohesion is high coupling will be low and inverse is also true.

Experimental Process: Second phase

In this second step we would like to explain how we have come to the above conclusion. From the following results we can prove the hypothesis that there exists a relationship between coupling and cohesion. To prove the above four cohesion metrics: DCI, DCD, DCDE and DCIE and for coupling CBO metric are undertaken. As a first step data on the selected metrics from each of the considered systems is collected. Later to find out the relationship Spearman coefficient was used. This experiment is important because it proves the above said belief. This test is well suited since the dependence seems to be non linear as stated to the previous graphs. Studies of the data sets are done by calculating the Spearman dependence coefficients for each pair of metrics (a metric of cohesion, CBO). The Spearman statistic is based on ranks of the observations. The value of the Spearman statistic is a number between -1 and 1, -1 being a perfect negative dependence and +1 a perfect positive dependence.

Results

X. Regression Study

Main aim of this study is to find out if there is any linear connection between cohesion metrics and coupling. For this a regression study was done between coupling and under different cohesion metrics. As a first step cohesion metric connected to the retained coupling.

Later a regression evaluation between two variables was done. Below are some terms utilized in this section of the paper.

- Regression model: It is the regression model used. DCDE, DCIE, DCD, DCI are the variables which are not dependent and coupling metric CBO is not independent
- Dependant variable: A random variable to predict;
- Independent variable: A predictive variable;
- R^2(r-square): The percentage of change in the dependent variable described by the independent variables in the regression model for the given example of the population.

- Population : The group of classes that are chosen into consideration at the experiment level;
- Adjusted R-square: The percentage of change in the dependent variable described by the independent variables in a regression model of the population;
- Sum of squares of regression: The variance of the dependent variable described by the regression model;
- Sum of squares of residual: The change is not described by the dependent variable;
- Mean squares of residual: Total of squared residues divised by the number of freedom degrees of the residues;

To analyze some other variant of this relatedness between the metrics of cohesion and the coupling metric, the logarithm of the coupling value was explained. To get this value a regression between this logarithm and the cohesion is undertaken. The out come of the above is given in table 5.

For this first experiment we have utilized R2 statistic through this we have tried to find out the areas that connect coupling and cohesion. To find out this, values of system JIU are used. For this DCDE and DCIE values are 0.0228 and 0.0267 respectively. These values present the variance of coupling brought in by the cohesion metrics, which can be said as 2.28% and 2.67% in percentages.

System	Cohesion Metric	R^2 vs Coupling
FujabaUml	DC_{DE}	0.0118
	DC_{IE}	0.0081
	DC_D	0.0081
	DC_I	0.0054
Gnujsp	DC_{DE}	0.2835
	DC_{IE}	0.2676
	DC_D	0.4657
	DC_I	0.4506
JIU	DC_{DE}	0.0228
	DC_{IE}	0.0267
	DC_D	0.0186
	DC_I	0.0221
Moneyjar	DC_{DE}	0.0226
	DC_{IE}	0.0237
	DC_D	0.032
	DC_I	0.0331

Table 4 : Values of R2 in the different systems.

Concerning table 5, for the above said JIU system, values 0.0341 and 0.0430, respectively for cohesion metrics DCDE and DCIE, present the percentages of the logarithm of the variance described by the cohesion metrics. Hence, 3.41% and 4.3% of the logarithm of variance is described respectively by cohesion metrics DCDE and DCIE. Given the obtained values in this test and observing the noted observations in previous part (the relationship seems to be non linear), we undertook a second experiment utilizing the Spearman correlation.

System	Cohesion Metric	R^2 vs logCoupling
FujabaUml	DC_{DE}	0.0118
	DC_{IE}	0.0081
	DC_D	0.0081
	DC_I	0.0054
Gnujsp	DC_{DE}	0.2835
	DC_{IE}	0.2676
	DC_D	0.4657
	DC_I	0.4506
JIU	DC_{DE}	0.0228
	DC_{IE}	0.0267
	DC_D	0.0186
	DC_I	0.0221
Moneyjar	DC_{DE}	0.0226
	DC_{IE}	0.0237
	DC_D	0.032
	DC_I	0.0331

Table 5 : R2 obtained with the log of a coupling value

Spearman Correlation study (rank statistic)

Further, we calculated the correlation degree (according to Spearman) between the cohesion metrics and coupling in the chosen systems. Table 6 gives the results that we have got.

The aim of this research was to identify a correlation (negative) between the cohesion metrics and coupling metric we have chosen. This tests main goal is to find out if the connectedness is significatively lower than 0 (in the statistical sense) for a negative dependence. A statistical research was conducted. The statistical research must then be correlated to a Student variable computed with n-2 freedom degrees, and where n is the size of the example. The P-value shows the probability of getting such a value under the null hypothesis of absence of dependence. In general, if P-value < 0.05 (error margin), we come to a conclusion that a negative dependence is significant. Hence, for the group of tested systems and from the values of table 6, only the moneyjar system has P-values > 0.05 for all combinations (cohesion metric – coupling metric). We examine values of 0.48996, 0.46740, 0.4649, and 0.442451 for, respectively, cohesion metrics DCIE, DCDE, DCI, DCD correlated to the coupling metric CBO.

For the remaining chosen systems, the P-values are all < 0.05 for the whole group of combinations (cohesion metric – coupling metric). As per table 6, all systems show a significant negative dependence between cohesion and coupling. But this is not possible with moneyjar system. The reason behind this exception is that this particular system has got less number of classes [20] than the other systems. Hence we can come to a conclusion that to observe significant negative dependency is better to select systems with high number of classes. From the above results, it can be said that there is a correlation between cohesion metric and coupling metric.

Other than this it is also observed that if the number of classes are high in a system then the dependency level between cohesion and coupling (Non linear dependency relations) can be confirmed easily. Different kind of systems were selected to prove the correlation between cohesion and coupling and is proved. But there are many other kind of systems on which it is not tested. Hence it is better to prove the same on other systems also before we give any global declaration.

System	Statistic	DCIE-CBO	DCDE-CBO	DCI-CBO	DCD-CBO
Gmujsp	S .Coeff.	0.354545	-0.35892	-0.35455	-0.35892
	Test statistic	-2.786373	-2.8258	-2.78637	-2.8258
	P-value	0.0036697	0.003299	0.00367	0.003299
Jiu	S.Coeff.	0.50857	-0.47888	-0.50337	-0.47584
	Test statistic	-5.11527	-4.72409	-5.04502	-4.68533
	P-value	1.17E-06	5.27E-06	1.53E-06	6.11E-06
fujabaUml	S.Coeff.	0.425590442	-0.43809	-0.29225	-0.31195
	Test statistic	-3.5817696	-3.71155	-2.3273	-2.5005
	P-value	0.000349459	0.000232	0.011731	0.007625
Jexcelapi	S.Coeff.	0.18723	-0.22039	-0.19318	-0.21987
	Test statistic	-1.98076	-2.34805	-2.04612	-2.3423
	P-value	0.02508	0.010346	0.021587	0.010499
Moneyjar	S.Coeff.	0.00602	-0.01955	-0.02105	-0.03459
	Test statistic	-0.02552 -	0.08295 -	0.08934	-0.14683
	P-value	0.48996	0.467402	0.4649	0.442451
WBEM	S.Coeff.	0.242732	-0.295322901	-0.26708	-0.3055
	Test statistic	-3.338282	-4.124041493	-3.69767	-4.28049
	P-value	0.0005133	2.85E-05	0.000145	1.52E-05

Table 6 : Results of the Spearman rank statistic method

XI. Conclusion

With the help of this research we introduced a new criterion and gave a better, revised definition of class cohesion [5, 6]. Common objects parameters are the new criteria which is introduced and also validated in this paper. This becomes the new way to measure class cohesion. We enhanced a cohesion measurement tool for Java programs to automate the calculation of the class cohesion metrics that we propose. Different kinds of systems were taken for the experiment to prove that the new criterion and the proposed metrics for class cohesion give the best assessment. These systems are very much different in size and domain. In this test many number of classes were analyzed. After all the experiments it was understood that the extended metrics with the help of new criterion is capable of finding out more pairs of connected methods. In addition to the above experiment one more was also conducted. It helped to validate our new metrics. This was helpful to prove the correlation between cohesion and coupling. To the second step we got a chance to observe several hundreds of classes. From this second test it was found in the selected systems that there exists a negative correlation between cohesion and coupling. More over through the results we could see that if the number of classes in a system is high then the dependency relation between cohesion and coupling can be confirmed easily.

References Références Referencias

1. K.K.Aggarwal, Yogesh Singh, Arvinder Kaur, RuchikaMalhotra, Empirical study of object-oriented metrics, In Journal of Object Technology, vol. 5. no. 8, November-December 2006, pp. 149-173.
2. K.K.Aggarwal, Yogesh Singh, Arvinder Kaur, RuchikaMalhotra, Investigating the effect of coupling metrics on fault proneness in objectoriented systems, SQP, vol. 8 no. 4, 2006.
3. H. Aman, K. Yamasaki, H. Yamada and MT. Noda, A proposal of class cohesion metrics using sizes of cohesive parts, Knowledge-Based Software Engineering, T. Welzer et al. (Eds), pp. 102-107, IOS Press, September 2002.
4. L. Badri, M. Badri and S. Ferdenache, Towards Quality Control Metrics for Object-Oriented Systems Analysis, Proceedings of TOOLS (Technology of Object-Oriented Languages and Systems) Europe'95, Versailles, France, Prentice-Hall, March 1995.
5. L. Badri and M. Badri, A New Class Cohesion Criterion: An empirical study on several systems, Proceedings of QAOOSE'03, 2003.
6. L. Badri and M. Badri, A proposal of a new class cohesion criterion: An empirical Study, In Journal of Object Technology, 2004. [Bas 96] V.R. Basili, L.C. Briand and W. Melo, A validation of object-oriented design metrics as quality indicators, IEEE

Transactions on Software Engineering, 22 (10), pp. 751-761, October 1996.
7. J.M. Bieman and B.K. Kang, Cohesion and reuse in an object-oriented system, Proceedings of the Symposium on Software Reusability (SSR'95), Seattle, WA, pp. 259-262, April 1995.
8. G. Booch, Object-Oriented Analysis and Design With Applications, Second edition, Benjamin/Cummings, 1994.
9. L. C. Briand, J. Daly, V. Porter, and J. Wuest, The Dimensions of Coupling in Object-Oriented Design, OOPSLA'97,1997.
10. L.C. Briand, J. Daly and J. Wusr, A unified framework for cohesion measurement in object-oriented systems, Empirical Software Engineering, 3 (1), pp. 67-117, 1998.
11. L. Briand, J. Wuest, J. Daly and V. Porter, Exploring the relationships between Design Measures and software quality in object-oriented Systems, Journal of Systems and Software, No. 51, pp. 245-273, 2000.
12. H. S. Chae and Y.R. Kwon, A cohesion measure for classes in objectoriented systems, Proceedings of the fifth International Software Metrics Symposium, Bethesda, MD, pp. 158-166, November 1998.
13. H. S. Chae, Y. R. Kwon and D H. Bae, A cohesion measure for objectoriented classes, Software Practice and Experience, No. 30, pp. 1405- 1431, 2000.
14. S.R. Chidamber and C.F. Kemerer, Towards a Metrics Suite for Object-Oriented Design, Object-Oriented Programming Systems, Languages and Applications (OOPSLA), Special Issue of SIGPLAN Notices, Vol. 26, No.10, pp. 197-211, October 1991.
15. S.R. Chidamber and C.F. Kemerer, A Metrics suite for object Oriented Design, IEEE Transactions on Software Engineering, Vol. 20, No. 6, pp. 476-493, June 1994.
16. S.R. Chidamber, David P. Darcy, and C.F. Kemerer, Mangerial use of metrics for object-oriented sofytware : An exploratory analysis, IEEE Transactions on Software Engineering, Vol. 24, No. 8, pp. 629-639, August 1998.
17. Melis Dagpinar and Jens H. Jahnke, Predicting maintenability with objectoriented metrics – An empirical comparaison, Proceedings of the 10th working conference on reverse engineering (WCRE'03), IEEE computer society, 2003.
18. K. El Emam and W. Melo, The prediction of faulty class using objectoriented design metrics, National Research Council of Canada NRC/ERB 1064, 1999.
19. B. Henderson-sellers, Object-Oriented Metrics Measures of Complexity, Prentice-Hall, 1996.
20. W. W. Hines, D. C. Montgomery, D. M. Goldsman and C. M. Borror, Probability and statistics in engineering, Fourth edition, John Wiley & Sons, Inc., 2003.
21. M. Hitz and B. Montazeri, Measuring coupling and cohesion in object oriented systems, Proceedings of the Int. Symposium on Applied Corporate Computing, pp. 25-27, October 1995.
22. H. Kabaili, R.K. Keller, F. Lustman and G. Saint-Denis, Class Cohesion Revisited: An Empirical Study on Industrial Systems, Proceeding of the Workshop on Quantitative Approaches Object-Oriented Software Engineering, QAOOSE'2000, France, June 2000.
23. H. Kabaili, R.K. Keller and F. Lustman, Cohesion as Changeability Indicator in Object-Oriented Systems, Proceedings of the Fifth European Conference on Software Maintenance and Reengineering (CSMR 2001), Estoril Coast (Lisbon), Portugal, March 2001.
24. G. Larman, Applying UML and Design Patterns, An introduction to object-oriented analysis and design and the unified process, Second edition, Prentice Hall, 2003.
25. W. Li and S. Henry, Object oriented metrics that predict maintainability, Journal of Systems and Software, Vol. 23, pp. 111-122, February 1993.
26. W. Li, S. Henry, D. Kafura and R. Schulman. Measuring Object-Oriented Design. In Journal of Object-Oriented Programming, Vol. 8, No. 4, pages 48-55, July/August 1995.
27. R. S. Pressman, Software Engineering, A practitioner's approach, Fifth edition, Mc Graw Hill, 2005.
28. I. Sommervile, Software Engineering, 2004.
29. W.P. Stevens, G.J. Myers and L.L. Constantine. Structured Design. In IBM Systems Journal, Vol. 13, No. 2, pages 115-139, May 1974.
30. E. Yourdon and L. Constantine, Structured Design, Prentice Hall, Englewood Cliffs, N.J., 1979.
31. Yuming Zhou, Hareton Leung, Empirical analysis of object-oriented design metrics for predicting high and low severity faults, IEEE Transactions on software engineering, vol. 32, no. 10, October 2006.

Global Journal of researches in engineering
Numerical Methods
Volume 11 Issue 7 Version 1.0 December 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 2249-4596 & Print ISSN: 0975-5861

Decision Tree Construction: A Continues Label Support Degree Based Approach

By N.Madhuri, T.Nagalakshmi, D.Sujatha
ATRI, Parvathapur, Uppal, Hyderabad, India

Abstract - Data mining and classification systems utilize decision tree algorithms since they proffer rapid speediness, advanced exactness and also simple organization of those algorithms. An ideal decision can be built only when the appropriate attributes are chosen. This paper focuses on throwing light on choosing characteristics based on the theory of attribute support degree on account of which a unique decision tree construction algorithm is proposed on the basis of rough set and granular computing theory. It is henceforth proved that the decision tree proposed by the new approach yields far more better results in terms of precision and consistency as compared to the decision trees yielded by ID3, C4.5 and DTBAS.

Keywords : rough set: decision tree: granular computing: attribute support degree: attribute selection.

GJRE-I Classification : FOR Code: 080299

DECISION TREE CONSTRUCTIONA CONTINUES LABEL SUPPORT DEGREE BASED APPROACH

Strictly as per the compliance and regulations of :

Open Association of Research, Soceity (Uinted States)

Decision Tree Construction: A Continues Label Support Degree Based Approach

N.Madhuri[α], T.Nagalakshmi[Ω], D.Sujatha[β]

Abstract - Data mining and classification systems utilize decision tree algorithms since they proffer rapid speediness, advanced exactness and also simple organization of those algorithms. An ideal decision can be built only when the appropriate attributes are chosen. This paper focuses on throwing light on choosing characteristics based on the theory of attribute support degree on account of which a unique decision tree construction algorithm is proposed on the basis of rough set and granular computing theory. It is henceforth proved that the decision tree proposed by the new approach yields far more better results in terms of precision and consistency as compared to the decision trees yielded by ID3, C4.5 and DTBAS.

Keywords : rough set: decision tree: granular computing: attribute support degree: attribute selection

I. INTRODUCTION

Decision sets can be denoted using tree structures with the help of decision tree which is a unique, spontaneous, data illustration scheme and also a competent classifier. Quinlan et al[1] proposed ID3, decision tree algorithm and hence has been persistently augmented which have been advanced to C4.5 [2]. The preeminent attribute is chosen as the existing attribute which is then recursively inflates the decision tree branches unless and until the conditional statement is achieved, which ultimately makes use of top-down greedy algorithms. There are different classification schemes that can be achieved concerning different solutions which poses two issues [3] in decision tree construction. Choosing characteristics for crafting new branches in a tree is one issue while the other one is pruning which is all about omitting and decreasing the tree. DTBAS[10] considered the Assortment of attribute as main concern, which was refined and improved by considering assortment of continuous labels also that is discussed in this paper.

Z. Pawlak et al[4] recommended rough set theory which is an expansion of set theory for studying intelligent systems which is followed up by inadequate and partial data information. There is a thriving submission of the rough set theory in the disciplines of data mining, pattern recognition, machine learning, decision analysis etc in recent times. Models are categorized into various resembling classes that houses imperceptible objects in terms of few attributes. Issues pertaining to feature selection, data reduction and pattern extraction can be amicably taken care of such that it can liberate the system of redundant data in systems containing null values or missing data.

Lin et al[5] proposed the expression of Granular Computing which spans itself covering all aspects of concerning theories, tactics, practices and means essential in solving a problem that makes use of granules. Granular Computing has witnessed vast inputs from different practices such as fuzzy sets, rough sets, shadowed sets, probabilistic sets etc.

A crucial step that needs to be taken care of while building a decision tree is choosing characteristics of nodes of a tree that houses minimum number of branches. Decision tree based on continuous label support degree (DTBLSD) algorithm is introduced which is considered as a splitting criterion on account of rough set theory and granular computing. Trial results have approved the usage of DTBLSD algorithm that assures and provides uncomplicated structures and superior categorization accuracy.

The rest of the paper is organized as follows: Section 2 discusses concepts relevant to rough set theory and granular computing. Section 3 gives a basic introduction to our new method and presents a simple example. Experimental comparison of the proposed method with 103 and C4.S is given in section 4. The final section concludes the research work of this paper.

II. BASIC CONCEPT

Few fundamental concepts of rough set theory [6, 7] and granular computing[8] are first initiated for ease of demonstration.

Definition 1 (Information System) : An information system can be labeled as $S = (U, A, V, f)$ wherein U is a finite set of object known as the universe; $A = C \cup D$, which is a non-vacant finite group of attributes; C and D depict set of condition and decision attributes respectively as also $v = \cup v_a, \forall a \in A$, which says that v_a is a value set of the attribute a and

Author α : M.Tech, Department of CSE, ATRI, Parvathapur, Uppal, Hyderabad, India. E-mail : madhuri.neerubavi@gmail.
Author Ω : Sr,Asst.prof, Department of CSE, ATRI, Parvathapur, Uppal, Hyderabad, India. E-mai : nlakshmi.t@gmail.com
Author β : Assoc.Prof and HOD, Department of CSE, ATRI, Parvathapur, Uppal, Hyderabad, India.
E-mail : sujatha.dandu@gmail.com

Global Journal of Researches in Engineering (I) Volume XI Issue VII Version I December 2011

$f = U \times A \rightarrow V$ is known as the information function also known as total decision function, such that $f(u,a) \in V_a$, for every $a \in A$, $u \in U$.

Definition 2 (Lower and Upper Approximation of Sets, Boundary of $X(BND(X))$ Let $S = (U, A, V, f)$ be an information system, and let $R \subseteq A$ and $X \subseteq U$. X can be estimated using information present in R by building lower and upper estimate $\underline{R}(X)$ and $\overline{R}(X)$ and $BND(X)$ known as boundary of X. We can now designate $\underline{R}(X)$ Ji(x), $\overline{R}(X)$, and $BND(X)$ as follows.

$$\underline{R}(X) = \{x \in U \mid [x]_R \subseteq X\} ....(1)$$

$$\overline{R}(X) = \{x \in U \mid [x]_R \cap X \neq \phi\} ...(2)$$

$$BND(X) = \overline{R}(X) - \underline{R}(X) ....(3)$$

Definition 3 (Indiscernibility Relation) Let $S = (U, A, V, f)$ be an information system and subset $P \subseteq A$ be known as the indiscernibility relation, indicated by IND(P), that can be termed as

$$IND(P) = \{(x, y) \in U \times U \mid \forall_a \in P, f(x,a) = f(y.a) ....(4)$$

Where *IND(P)* is an equivalence relationship that separates U into equivalence classes labeled as $\frac{U}{IND(P)}$, which contains group of objects that is unobvious concerning A.

Definition 4 (Granular Degree) suppose $K = (U, R)$ is a repository. $R \subseteq U \times U$ is equivalence relations in the universe U known as knowledge. $GD(R)$ depicts Granular degree of knowledge R. Its definition is as follows:

$$GD(R) = \frac{|R|}{|U \times U|} = \frac{|R|}{|U|^2} ...(5)$$

Where $|R|$ is the cardinal number of $R \subset U \times U$.

When in an equivalence relation R, the granular degree of R reaches the minimum size $|U| / |U|^2 = 1/|U|$; When R is a domain relation, the granular degree of R attains the maximum size $|U|^2 / |U|^2 = 1$

Definition 5 : Assume R is knowledge of repository $K = (U, R)$, $U / R = \{X_1, X_2, ..., X_n\}$, the granular degree of basic knowledge is defined as

$$GD(X_i) = \frac{\sum_{i=1}^{n} |x_i|^2}{|U|^2} ...(6)$$

III. Proposed Algorithm

This section aims at familiarizing the algorithm of building a decision tree on the basis of attribute support degree.

a) *The Principle of Label and Attribute Selection*

The label that represents least average uncertainty is supposed to be chosen as the test label and then choose attribute with less uncertainty as test attribute from the class represented by the selected label, which because it makes apt decisions when compared to existing test attribute selection in different decision tree algorithms. $\{a_1, a_2, ...a_{ls}\}$

Definition 6 : (Label support Degree) let Label l representing the group of attributes here is total number of attributes grouped under label .Then average uncertainty represented by label can measure as

$$avg_{uc}(l) = \frac{\sum_{i=1}^{|l|} uc(a_i)}{|l|} ....(7)$$

Here $avg_{uc}(l)$ represents average uncertainty of label l

$uc(a_i)$ Represents uncertainty of attribute a_i of label l

Definition 7 (Attribute Support Degree) Let $S = (U, A, V, f)$, $A = C \cup D$ be an information system l is a label contains subset of attributes represented as $Q \subseteq C$. Attribute support degree can be denoted as follows based on the definitions mentioned above.

$$S(Q, D) = \frac{GD(Q \cup D)}{GD(Q)} = \frac{|IND(QUD)|}{IND(Q)} ...(8)$$

Where $|IND(Q)|$ denotes the cardinal number of $IND(Q) \subseteq U \times U$.

D using Q can be estimated with the help of a measure S (Q,D) . Definition 5 states that whenever we get the relations among them, namely, when GD(R) is smaller, the distinguishable degree is stronger and S(Q,D) is greater, thereby Q is better sets of test attribute of D. On the contrary, the smaller S(Q,D) is, the worse we get Q as sets of test attribute of D.

b) *The Description of DTBLSD*

The basic notion of DTBLSD expresses the point that whenever label support degree with association of label level attribute support degree is made use of as a customary for choosing a test attribute concerning every node in the decision tree. The attribute reduction set assists in selecting a condition attribute that possesses the highest degree of label level attribute which can be put to use at the root of the decision tree. There will be a testing of the remaining condition attributes on each and every branch of the root node and so, the algorithm persists in a recursive manner by addition of new sub-trees to every division until the leaf is reached.

According to the above idea, using the S(Q,D) as the splitting criterion, we propose our algorithm DTBLSD. Current sample set is depicted by T, set of labels depicted by L, condition attribute set of a label is depicted by the l_{al} . $|l_{al}|$ depicts the number of attributes in the condition attribute set of label l . All attributes of the condition attribute set are discrete and continuous values are discretized by continuous labeling. Following are the specific steps of the algorithm.

Algorithm : A decision tree is created by DTBLSD (T, attribute list) that using the given training data.

Input : The training set samples, represented by discrete valued attributes; the set of condition attribute, attribute list.

Output : A decision tree.

Step1 : create a node N;

Step2 : if samples are all of the same class C, then return N as a leaf node labeled with the class C;

Step3 : if attribute list is empty, then return N as a leaf node labeled with the most common class in the samples;

Step 4 : Select a label l that represents average uncertainty is low.

Step5 : select test attribute in l_{al} with the highest degree of attribute support;

Step6 : label node N with test attribute;

Step7 : for each known value aj of test attribute, grow a branch from node N according to the condition test attribute= aj;

Step8 : let Sj be the set of samples in samples for which test attribute= aj;

Step9 : if Sj is empty, then attach a leaf labeled with the most common class in samples;

Step10 : else attach the node returned by . $DTBLSD(s_i, l_{al}, test\_attribute)$.

The top-down recursive divide and conquer approach for construction of a decision tree wherein the recursion related division takes place only when any one criterion mentioned below is gratified. A common class contains :

1. All specimens for a specific branch which restores a leaf that is termed with the concerned class. Here, a large case of voting is provisioned to change the present working node into a leaf that is termed with the concerned class that in in demand from amongst various specimens.
2. In addition, there are no more specimen test attributes and the class division specimens can be placed wherein a leaf is generated and termed with the most featured class in specimens.

IV. EXPERIMENTS

a) Example Analysis

Table 1 showcases a data tuple training group originated from All Electronics customer records that are implemented using polic mentioned in reference (6). The first step is to estimate the degree of attribute support for each situational attribute or characteristic.

$$\frac{U}{IND(al)} = \{ \{l,2,8,9, 11\},\{3,7,12,13\},\{4, 5, 6, 10, 14\}\};$$

$$\frac{U}{IND(az)} = \{ \{1, 2, 3, 13\},\{4, 8, 10, 11, 12, 14\}, \{5, 6, 7, 9\}\};$$

$$\frac{U}{IND(a3)} = \{\{1, 2, 3, 4, 8, 12, 14\}, \{5, 6, 7, 9, 10, 11, 13\}\};$$

$$\frac{U}{IND(a4)} = \{ \{l,3,4,5,8,9,10,13\},\{2,6,7,11,12,14\}\};$$

$$\frac{U}{IND(d)} = \{ \{1 ,2,6,8,14\},\{3,4,5,7,9, 10, 11,12, 13\}\};$$

$$\frac{U}{IND(al, D)} = \{\{I, 2, 8\},\{3, 7, 12, 13\},\{4,S, 1 0\},\{6, 14\},\{9, II\}\};$$

$$\frac{U}{IND(az, d)} = \{\{1, 2\},\{3, 13\},\{4, 10, 11, 12\},\{5, 7, 9\},\{6\},\{8, 14\}\};$$

$$\frac{U}{IND(az, d)} = \{\{l, 2, 8, 14\}, \{3, 4, 12\}, \{5, 7, 9, 10, 11, 13\}, \{6\}\};$$

$$\frac{U}{IND(a4, d)} = \{\{1, 8\}, \{2, 6, 14\}, \{3, 4, 5, 9, 10, 13\}, \{7, 11, 12\}\}.$$

$$S(a_1, d) = \frac{|IND(a_1 \cup d)|}{IND(a_1)} = 0.636$$

$$S(a_2, d) = \frac{|IND(a_2 \cup d)|}{IND(a_2)} = 0.559$$

$$S(a_3, d) = \frac{|IND(a_3 \cup d)|}{IND(a_3)} = 0.633$$

$$S(a_4, d) = \frac{|IND(a_4 \cup d)|}{IND(a_4)} = 0.580$$

Table I : Training data tuple from the AllElectronics customer database

U	age	income	student	credit	buy
1	<=30	high	no	fair	No
2	<=30	high	no	excellent	No
3	31<=age<=40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	No
7	31<=age<=40	low	yes	excellent	Yes
8	<=30	medium	no	fair	No
9	<=30	low	yes	fair	Yes
10	>40	medium	yes	fair	Yes
11	<=30	medium	yes	excellent	Yes
12	31<=age<=40	medium	no	excellent	Yes
13	31<=age<=40	high	yes	fair	Yes
14	>40	medium	no	excellent	No

Notations used in example descriptions:

Age➔ , Income➔ , Student➔ , Credit➔ , Buy➔

a_1 is selected as the min root of the decision tree and is tagged with age since $S(a_1,d)$ is the maximum extent of a degree from amongst all the condition attributes as also various number of divisions which are branched in reference to a range of different attributes. In case where age=1, all the specimens that are grouped into this should belong to the same class and hence a leaf should be generated at the end of every division and should be tagged with d=yes. The figure above depicts the final decision tree that is built by DTBLSD.

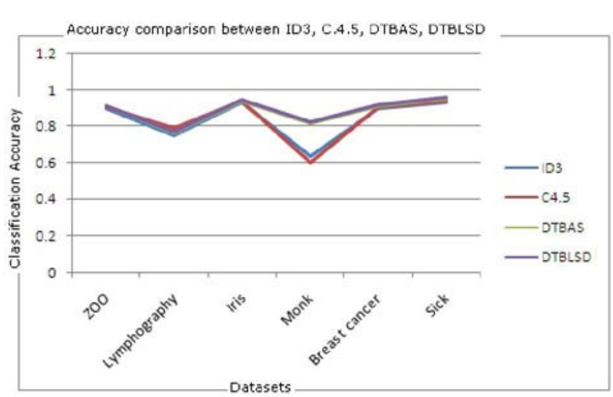
b) Experimentl Comparison

Experimental comparison of DTBLSD with respect to ID3[1], C4.5[2] and DTBAS[10] is discussed in this section. The real datasets that are used in this are approved from University of California, Irvine (UCI), and is known as the machine learning database repository where C++ design language is implemented to form the requisite algorithm. WEKA 3.7 is used for successful accomplishment of ID3[1] and C4.5[2] which is a compilation of machine learning algorithms used for data mining generated and procured by Frank that involved the 10 fold cross estimation to calculate classification authenticity. All experiments were performed on a PC, Intel(R) Pentium(R) 4 CPU, 2.93GHz, 512MB RAM.

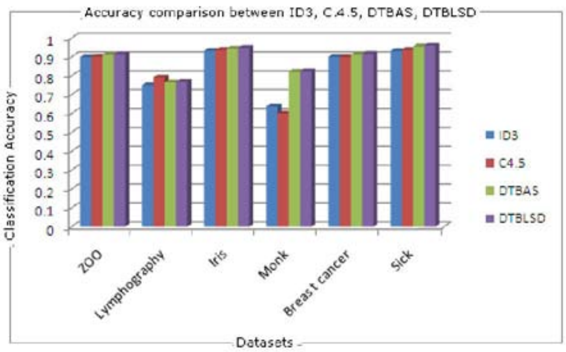
It has been observed from the analytical results that as compared to ID3, C 4.5 and DTBAS, our specified algorithm has better accuracy and lower computation cost. Listing of comparison report follows:

Table II : Tabular representation of accuracy comparison between ID3, C4.5, DTBAS, DTBLSD

Dataset	ID3	C4.5	DTBAS	DTBLSD
ZOO	0.899	0.901	0.911	0.914
Lymphography	0.75	0.791	0.765	0.767
Iris	0.932	0.936	0.943	0.947
Monk	0.637	0.6	0.821	0.824
Breas t cancer	0.9	0.9	0.912	0.916
Sick	0.931	0.937	0.955	0.959



(a) Line Chart of Comparison report



(b) Bar chart of comparison report

Fig 1: Accuracy comparison graphs

V. Conclusion

The paper first focuses on explaining the basic notion of label support degree and attribute support degree [10] and selecting it as a basic decisive factor on the basis of degree of involvement between condition attribute and decision attribute accordingly where a unique decision algorithm tree based on continuous label support degree and label level attribute support degree (DTBLSD) is recommended. Accordingly a suitable methodology is devised which is flexible enough to accommodate and provides lower complexity and high level of accuracy as compared to other algorithm generating methods. A disadvantage identified in [10] is issues pertaining to adjustment with adaptability of samples, which has been overcome successfully in our model.

References References Referencias

1. J. R. Quinlan, "Induction of Decision Trees", Machine Learning, voLI, 1986, pp. 81-106.
2. J. R. Quinlan, "Improved Use of Continuous Attributes in C4.5", Journal of Artificial Intelligence Research, volA, 1996, pp.77-90.
3. F. Seifi, H. Ahmadi, M. Kangavari, "Twins Decision Tree Classification: A Sophisticated Approach to Decision Tree
4. Z. Pawlak, "Rough Sets," International Journal of Information and Computer Science, voLII, 1982, pp. 341 -356.
5. T. Y. Lin, "Granular Computing II Infrastructure for AI-Engineering Examples, Intuitions and Modeling," Proc. 2006 IEEE International Conference on Granular Computing(GrC 2006), 2006, pp. 2-7.
6. J. W. Han, M. Kamber, Data Mining Concepts and Techniques[M). Morgan Kaufmann Publishers, 2001.
7. B. S. Ding, Y. Q .. Zheng, S.Y. Zang, "A New Decision Tree Algorithm Based on Rough Set Theory," Proc. Asia-Pacific Conference on Information Processing(APCIP 2009), 2009, pp. 326-329.
8. D. Q. Miao, G. Y. Wang, Granular Computing: Past, Now, Future[M), Science publishing house, 2007.
9. E. Frank, M. Hall, WEKA, http://www. cs. waikato.ac. nz/ml/weka
10. Qing Lin; Zongzhuan Ding; Jianping Yong; Jun Zhou; , "An algorithm of decision tree construction based on attribute support degree," Educational and Information Technology (ICEIT), 2010 International Conference on , vol.2, no., pp.V2-516-V2-519, 17-19 Sept. 2010.
11. Y.L.Chen,C.L. Hsu, and S.C. Chou, “Constructing a Multi-Valued and Multi-Labeled Decision Tree”,Expert Systems with Apllications,vol.25,pp. 199-209, 2003.

This page is intentionally left blank

Global Journal of researches in engineering
Numerical Methods
Volume 11 Issue 7 Version 1.0 December 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 2249-4596 & Print ISSN: 0975-5861

C-Proeq Algorithm for Subcarrier Allocation for Wimax

By V. Pradeep Sunder, M. Arun
Sri Krishna College of Engineering and Technology, Coimbatore

Abstract - WiMAX is one of the latest techniques for providing Broadband Wireless Access in metropolitan area. Phenomenal growth in the demand for higher bandwidth increases the importance of subcarrier allocation. WiMAX uses OFDMA techniques based on multiuser diversity such as Adaptive Modulation Coding. Resource Allocation Algorithm improves performance based on user channel condition. Fair Allocation Algorithm maximizes the minimum data rate and allocates equal bandwidth to all users. However this allocation reduces the bandwidth efficiency of the system. In order to overcome this drawback Heuristics allocation such as Proportional Allocation and Equal Capacity Increment are proposed. To improve their performance, both allocations are combined and C-Proeq (Combination of Proportional and Equal capacity increment algorithms) is proposed. The other objective is to share better trade off between bandwidth efficiency and fairness. Performance has been compared on the basis of bit error rate with respect to signal to noise ratio and various user distances from base station.

Keywords : Adaptive Modulation Coding, Subcarrier allocation, WiMAX. OFDMA.

GJRE-I Classification : FOR Code: 080699

C-PROEQ ALGORITHM FOR SUBCARRIER ALLOCATION FOR WIMAX

C-Proeq Algorithm for Subcarrier Allocation for Wimax

V. Pradeep Sunder[α], M. Arun[Ω]

Abstract - WiMAX is one of the latest techniques for providing Broadband Wireless Access in metropolitan area. Phenomenal growth in the demand for higher bandwidth increases the importance of subcarrier allocation. WiMAX uses OFDMA techniques based on multiuser diversity such as Adaptive Modulation Coding. Resource Allocation Algorithm improves performance based on user channel condition. Fair Allocation Algorithm maximizes the minimum data rate and allocates equal bandwidth to all users. However this allocation reduces the bandwidth efficiency of the system. In order to overcome this drawback Heuristics allocation such as Proportional Allocation and Equal Capacity Increment are proposed. To improve their performance, both allocations are combined and C-Proeq (Combination of Proportional and Equal capacity increment algorithms) is proposed. The other objective is to share better trade off between bandwidth efficiency and fairness. Performance has been compared on the basis of bit error rate with respect to signal to noise ratio and various user distances from base station.

Keywords : Adaptive Modulation Coding, Subcarrier allocation, WiMAX. OFDMA.

I. Introduction

The consumer demand in multimedia applications for high data rate services, has led to grow the expansion of wireless communication systems. Major companies such as Intel and Motorola are promoting the WiMAX interfaces. Deregulation of the telecom industry and rapid growth of the internet, evaluated in favour of WiMAX. According to WiMAX forum, Physical represents Broadband Wireless Access Techniques. WiMAX operates in both LOS and NLOS operation such as 10-66 GHz and 2-11 GHz. Multipath fading occurs mostly in NLOS operation region, since each frequency arrives at the receiving point via a different radio path. This can be reduced by increasing delay spread. Frequency selective fading can also be equalized or reduced at the subcarrier level.

Inter Symbol Interference is reduced by cyclic prefix and longer symbol period. Fading can be completely removed using diversity techniques where multiple antennas are connected at transmitter and receiver ends. This paper deals with Multiuser Diversity such as Adaptive Modulation Coding for OFDMA (Wiesbeck, 2004), where different constellations are being used.

According to the IEEE 802.16 standard OFDM is to divide the available higher data rate into several low parallel data rate and these low data rate are mapped to individual subcarrier and being modulated. Orthogonal Frequency Division Multiplexing Access (OFDMA) uses multiuser diversity for adjacent subcarriers. Symbols are allocated to different users based on different data rate and channel condition. In OFDMA (Pinola, 2005), these subcarriers are divided to smaller group of subcarrier called sub channel.

Scheduling (Marco Cecchi, Romano Fantacci and Daniele Tarchi, 2008) of the multimedia traffic has become recent trend for research community. In this paper various adaptive modulation schemes are being identified and adapted to improve the bandwidth efficiency of the system according to the various channel condition.

To have more flexibility and higher efficiency, adaptive OFDM schemes are adopted to maximize the system capacity and maintain the desired system performance in the inherent multi-carrier nature of OFDM allows the use of link adaptation techniques according to the narrow-band channel gains. Performance improvement (Alessandro Biagioni and Romano Fantacci, 2009) in Bit Error Rate and users various distances from base station has been simulated in this paper. Techniques for transmission which does not adapt for the transmission parameters to the fading channel are designed to maintain at acceptable performance under the worst-case channel conditions with a consequent of insufficient utilization of the available resources. Hence if the channels fade level is known at the transmitter then the Shannon capacity can be approached by matching transmission parameters to time-varying channel.

To improve the performance of existing algorithms (Alessandro Biagioni and Romano Fantacci, 2009), a combined proportional and equal capacity increment algorithm, **C-Proeq,** is proposed in this work.

II. System Model

Our system uses Mobile WiMAX OFDMA-PHY where K users communicate with the base station using a group of subcarrier called as sub channel or slot. Each subcarrier as in (Marabissi, Tarchi, Genovese, and Fantacci,2007)is assigned power, user bit

Author[α] : Assistant Professor/ ECE,
Author[Ω] : Associate Professor / ECE,
Sri Krishna College of Engineering and Technology.
Coimbatore-641 008.
(Ph.No : 9487974733. E-mail : pradeepvsunder@gmail.com)

modulation to QAM symbols. These symbols are combined using IFFT, then they are transmitted to time varying, frequency selective channel, with each user experiencing an independent channel. Important to know that Channel State Information (CSI) are known at the transmitter level so that transmitter can adapt to the time varying channel condition and attempts to use available resource in efficient way(Ian Wong and Brain Evans,2008).

According to Mobile WiMAX, slot or sub channel is formed by group of adjacent subcarrier which is subdivided into 48 data subcarrier and 6 pilot subcarrier for a total of 54 sub carriers. This structure uses TDD hence uplink and downlink is same. In this paper a bin is formed which has 6 subcarrier in which 5 data subcarrier and 1 pilot subcarrier. Total of 54 bits per channel transmitted, hence 6x54=324 bits are being transmitted. Fig 1 represents OFDMA sub channels.

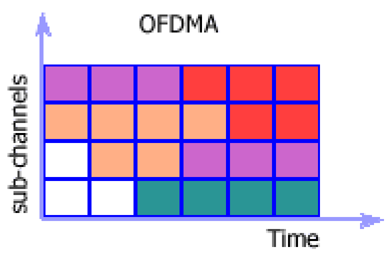


Fig.1 : Subchannel Allocation in OFDMA

Mobile WiMAX uses SOFDMA which scales FFT size to the channel bandwidth and also keeping frequency spacing constant 10.94 kHz. Sub carrier spacing is obtained by channel bandwidth multiplied by sampling rate where sampling rate is 28/25 i.e. 10*(28/25)/1024=10.94 kHz. Frame duration is 8ms and 1024FFT for 10 MHz channel bandwidth. Total OFDMA Symbol time is 102.8 ms, i.e. symbol time 91.4 + guard time.11.4.

TDD structure has guard time 5 μs for uplink and 5 μs for downlink. This structure is solution for data traffic such as IP based multi rate and multi-QoS services. 40 OFDM symbols for downlink sub frame and 39 OFDM symbol for uplink .Total of 79 OFDMA symbols per frame.

a) Adaptive Modulation Coding

Adaptive Modulation Coding chooses best modulation and channel coding schemes according to the propagation condition of radio link channel known at the transmitter end. In our paper channel degradation is due to path loss and multipath fading. Throughput maximization, Fig 2 is without any constraint of data reliability.

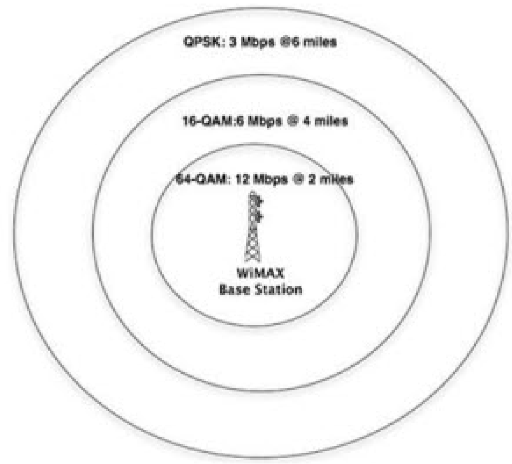


Fig.2 : Example of Adaptive Modulation Coding

III. RESOURCE ALLOCATION

In OFDMA system, resources are considered both in the low frequency and time dimension where different user can be assign sub carriers on a slot basis belonging to different frequencies or to different OFDM symbol times within the same frame. Hence, the available resources within the OFDMA frame form a grid of slots that are assigned to users according to suitable criteria. Subcarrier is divided into group of subcarrier called slots known as grid of frames.

Resource Allocation is usually formulated by two ways 1.minimization of total transmitted power under the constraint of user data rate. 2. Maximization of total data rate under the constraint of total transmitted power. Earlier Technique of Max Sum Rate or Greedy Optimization Algorithm is used to maximize the sum rate of the users with given power constraint.

Pk,l denotes transmitted power in the subcarrier *l*. Signal to interference plus noise ratio for user k in subcarrier *l* denotes as SINR k, l , can be expressed as in eq(1)

$$SINR_{k,l} = \frac{P_{k,l}\ h_{k,l}^2}{\sum_{J=1\ J=K}^{K} P_{j,l}\ h_{k,l}^2 + \sigma^2 \frac{B}{L}} \qquad (1)$$

Using Shannon capacity formula as the throughput measure, MSR algorithm maximizes the quantity by eq(2)

$$maxP_{k,l} \sum_{k=1}^{K} \sum_{l=1}^{L} {}^{B}/_{L} \log(1 + SINR_{k,l}) \qquad (2)$$

With Total Power constraint is given by eq(3)

$$\sum_{k=1}^{K} \sum_{l=1}^{L} P_{k,l}^2 \le P_{tot} \qquad (3)$$

The sum capacity is maximized if total throughput is maximized. Hence maximum sum capacity optimization problem can be decoupled into L simpler problem, one for each sub carrier. Further, capacity in subcarrier *l* is denoted as Cl written as in eq(4)

$$C_l = \sum_{k=1}^{K} log \left| 1 + \frac{P_{k,l}}{P_{tot,l-}P_{k,l} + \frac{B}{L}} \right| \quad (4)$$

a) Fair Allocation

In Fair allocation a slot is assigned the user in order to maximize the minimum number of users with user capacity Ck. This can be obtained by searching for each slot allocation matrix Xk that guarantees the minimum of Ck for k=1,2,....K. For example if 100 user are available then all available resources are allocated to them equally .The main drawback is more bandwidth is wasted for poor channel condition and Complexity. The fair allocation problem is formulated as follows as in eq (5),(6),(7)

$$\text{argmax}(\text{min}C_{SLOT}^{k}) \quad (5)$$

Subjected to

$$\sum x_{i,j}^{k} = 1 \quad (6)$$

$$\sum_{i=0}^{A-1} \sum_{j=0}^{B-1} x_{i,j}^{k} \geq 1 \quad (7)$$

b) Proportional Allocation

User with the best channel condition obtains lower number of resource with respect to the user with worst channel condition and hence channel capacity condition is not fully exploited. This is one of the drawbacks of Fair allocation this can be overcome by different allocation strategy by assuming the user with the best channel condition obtains larger amount of capacity. Being Maximum capacity is allocated to the each user given by (Keller and Hanzo, 2000), the slot allocation for the algorithm consideration searches for the capacity values for each user

$$C_1 : C_2 : \ldots\ldots : C_K = C_{max,1} : C_{max,1} : C_{max,2} : \ldots. : C_{max,K}$$

Problem formulation i.e. maximization of ergodic rates is not instantaneous rate. This allocation is based on priority based allocation algorithm expressed as in eq (8),(9)&(10)

$$\text{argmax}(\text{min}\frac{C_k}{C_{max,k}}) \quad (8)$$

Subjected to

$$\sum_{k=1} x_{i,j}^{k} = 1 \quad (9)$$

$$\sum_{i=0}^{A-1} \sum_{j=0}^{B-1} x_{i,j}^{k} \geq 1 \quad (10)$$

The main difference between fair and this allocation algorithm is that in this case the selected user is the one who has the minimum ratio between the actual capacity value and maximum obtainable capacity value. For the fair allocation case suboptimal allocation achieves result close to optimal solution with low complexity.

c) Equal Capacity Increment Allocation

The equal increment of capacity for the entire user with respect to non adaptive strategy. For example 100 user are active in a network under some circumstances 10 user became inactive then for these user power is divided equally among them self. Thus bandwidth is not wasted. Slot is belonging to the different user can be distinguished only by the position in the time domain different channel condition has different frequency are averaged and the amount of capacity assigned to the Kth user results to be as in eq(11)

$$C_K' = \frac{1}{k} C_{max,k} = \frac{1}{k} \sum_{i=0}^{A-1} \sum_{j=0}^{B-1} C_{SLOT_{i,j}^{k}} \quad (11)$$

This value is used to be estimation of capacity of non adaptive algorithm. Gk is the difference between assigned capacity assigned in eq (12) to the k-th user and C'K

$$G_k = R_k - C_K' \quad (12)$$

Equal capacity increment allocation scheme is consider in maximizing the minimum Gk for all the users. It follows optimal slot allocation problem accordingly in eq(13),(14),(15)

$$\text{argmax}(\text{min}G_k) = \text{argmax}(\text{min}(R_k - C_k')) \quad (13)$$

Subjected to

$$\sum_{k=1} x_{i,j}^{k} = 1 \quad (14)$$

$$\sum_{i=0}^{A-1} \sum_{j=0}^{B-1} x_{i,j}^{k} \geq 1 \quad (15)$$

*d) Combined Proportional And Equal Capacity Increment Algorithm (**C-Proeq**)*

In this paper we combine proportional allocation algorithm with equal capacity increment allocation algorithm by allocating priority in adaptive manner. This allocation is done by BPSK modulation schemes. The computational complexity is low and suboptimal solution is obtained. The same equations are being carried for proportional and equal capacity increment.

IV. Simulation Result and Disscusion

In this section result of MATLAB simulation is presented and the system performance is assumed in carrying out our analysis.

Simulation results are based on

1. OFDMA system based on 1024 FFT
2. Channel bandwidth of 10Mz
3. Rayleigh channel fading is assumed
4. No of active user is 20
5. Radio carrier frequency Fc =3.5Ghz
6. Guard band is 1/8 of OFDMA symbols duration

Neglecting path loss on the performance of consideration of allocation schemes when users are at the same distance from base station the total number of active user is 20 and maximum throughput is achieved by adaptive modulation and coding. The path loss formula (Hata,1980) and (Okumura, Ohmori, and Fukua, 1968) is given by eq(16)

$$PL = L + 10 . \gamma . log_{10}\left(\frac{D}{D_0}\right), D > D_0 \qquad (16)$$

Where D is the distance between user and base station, Do is the reference base station (100 meters) γ is the path loss exponent (set to 4.375 accordingly to an urban environment) and L equal to 20 log10 ($4\pi D0$) is the path loss value at the reference distance (with respect to the wavelength λ) that, in this case, is equal to 83.32 dB. Cell radius of 10 km has been considered and the system provides SNR 7dB at the edge of the cell.

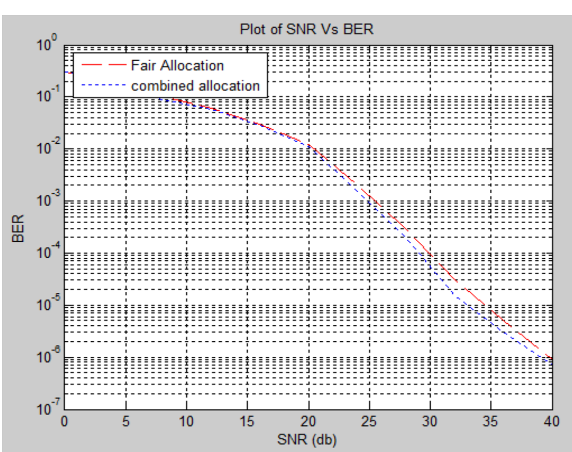


Fig.3 : Simulation of SNR vs BER

In figure3 the performance of the bit error rate vs. signal to noise ratio has almost same optimal solution when the path loss effect is removed.

Figure 4 shows the best performance of throughput up to 7 km for combination allocation .This shows more spectral efficient. Fair allocation produces constant throughput at any distance, resulting disadvantages for the user near to the base station (because it has lower throughput than combined allocation) and profitable for the user far from the Base station

Station.

Figure 5 represents users are at the same distance to the base station, fixed QAM modulation measures throughput with respect to distance. Hence adaptive allocation allows improving performance with respect to non adaptive allocation schemes.

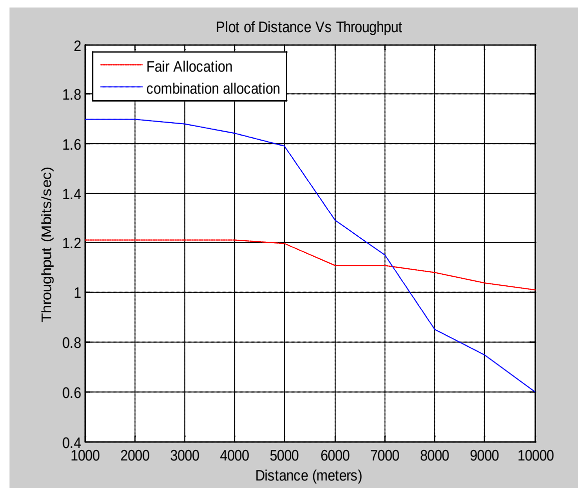


Fig.4 : Simulation of Distance vs Throughput

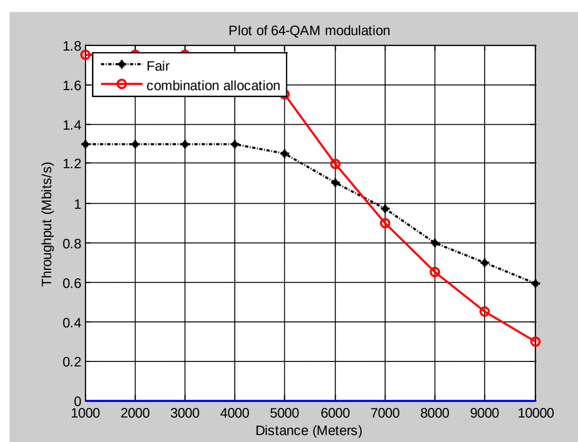


Fig.5 : Simulation of 64 QAM Throughput vs Distance

Figure 7 shows when the user are far from the base station throughput is high at the near end in combination allocation case as distance increases fair allocation regains throughput this is mainly due to the fact that fair allocation algorithm allocates subcarrier in the fair manner while the other combination allocation allocates in order to maximize the cell capacity by damaging far user.

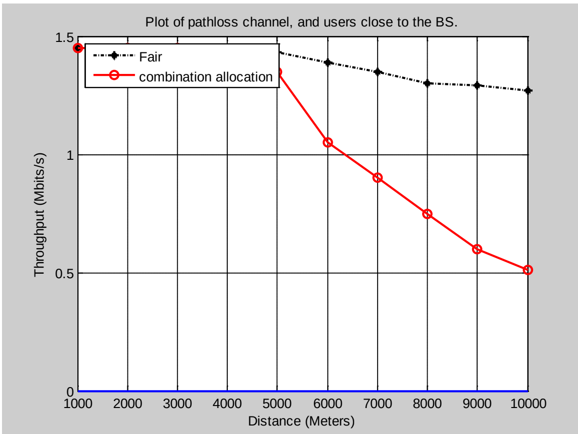


Fig.6 : Simulation of User Near to the BS

Figure 6 shows when user close to the base station the throughput is same (high) up to 3km after that throughput reduces as the distance proceeds.

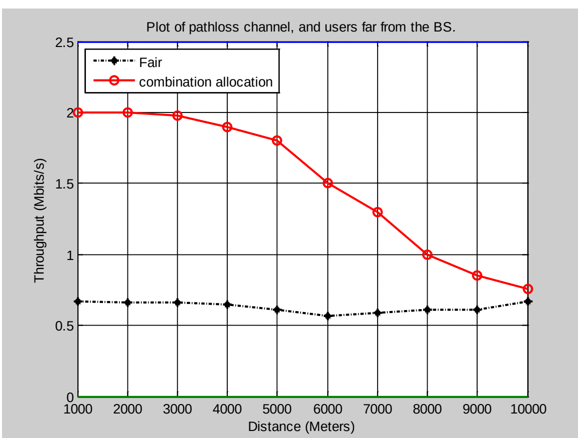


Fig.7 : Simulation of User Far to the Base Station

V. Conclusion

The proposed C-Proeq resource allocation algorithm is proposed in adaptive manner for the different user based on the different channel condition. By combining the equal capacity and proportional fair allocation algorithms, better bandwidth efficiency of C-Proeq algorithm is obtained and better trade-off between fair and bandwith efficiency is achieved. Thus the performance is compared with respect to bit error rate vs signal to noise ratio & user with different distance from base station.

References Références Referencias

1. Wiesbeck, W., (2004) IEEE Standard for local and metropolitan area network- part 16.
2. Pinola, J., (2005), Amendment to IEEE Standard for Combined Fixed Broadband Wireless Access system-physical and Mac layer for combined Fixed and Mobile operation in Licensed Band, 10-30.
3. Marco Cecchi, Romano Fantacci and Daniele Tarchi (2008) "Adaptive Scheduling Algorithms for Multimedia Traffic in wireless OFDMA system " IEEE communication GLOBECOM" 45-51.
4. Alessandro Biagioni and Romano Fantacci (2009) "Adaptive subcarrier allocation schemes For wireless OFDMA system in WiMAX Networks" IEEE Journal on selected area in communications vol 27, no 2, 217-225.
5. Marabissi, D., Tarchi, D., Genovese, F and R. Fantacci (2007) "Adaptive Modulation in wireless OFDMA systems with finite state modelling," In Proc. IEEE Globecom ,USA,40-54
6. Ian Wong and Brain Evans "Resource Allocation in Multiuser Multicarrier Wireless Sytems"acid free paper "Springer"2008
7. Keller, T., and Hanzo L., (2000) "Adaptive modulation techniques for duplex OFDM transmission," IEEE Trans. Veh. Technol., vol. 49, no. 5, pp.1893–1906.
8. Hata, O., (1980) "Empirical Formula for Propagation Loss in Land Mobile Radio Services," IEEE Trans. Vehicular Technology, vol. 29, pp. 317–325.
9. Okumura, T.,Y., Ohmori,E., and Fukua k., (1968) "Field strength and its Variability in UHF and VHF land-mobile radio service,"Rev. Electronic Communication Lab., vol. 16, no. 9,.

This page is intentionally left blank

Global Journal of researches in engineering
Numerical Methods
Volume 11 Issue 7 Version 1.0 December 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 2249-4596 & Print ISSN: 0975-5861

A Unified Field Approach on Fractional-Ordered Micropolar Thermoelasticity with Diffusion

By Soumen Shaw, Basudeb Mukhopadhyay
Bengal Engineering and Science University, Shibpur, India

Abstract - The present paper is completely devoted on derivation of some basic fundamental relations in generalized thermodiffusive micropolar elasticity with fractional-ordered derivatives. The generalized heat conduction and mass diffusion equations have been modified by using fractional calculus. A variational principle is obtained and hence the uniqueness theorem for those equations has been proved.

Keywords : Fractional calculus, Micropolar elasticity, Thermodiffusion, Variational principle, Uniqueness theorem.

GJRE-I Classification : FOR Code: 020304

A UNIFIED FIELD APPROACH ON FRACTIONAL-ORDERED MICROPOLAR THERMOELASTICITY WITH DIFFUSION

Strictly as per the compliance and regulations of :

A Unified Field Approach on Fractional-Ordered Micropolar Thermoelasticity with Diffusion

Soumen Shaw[α], Basudeb Mukhopadhyay[Ω]

Abstract - The present paper is completely devoted on derivation of some basic fundamental relations in generalized thermodiffusive micropolar elasticity with fractional-ordered derivatives. The generalized heat conduction and mass diffusion equations have been modified by using fractional calculus. A variational principle is obtained and hence the uniqueness theorem for those equations has been proved.

Keywords : Fractional calculus, Micropolar elasticity, Thermodiffusion, Variational principle, Uniqueness theorem.

I. Introduction

It is well established that the thermoelasticity theory is a fusion of the theory of heat conduction and the theory of elasticity. In classical theory of thermoelasticity there was a diffusive phenomenon on the heat propagation and thermal signals propagate with infinite speed. This physically unacceptable drawback of infinite speed of heat propagation was inherent in that theory. Modifying the Fourier's law of heat conduction, Lord-Shulman [1] introduced one non-classical theory of heat propagation with one relaxation time which can avoid that paradox. Green-Lindsay [2], in the year 1972, Proposed another one with two relaxation times. These non-classical theories are referred as generalized theory of thermoelasticity. Dhaliwal and Sherief [3] extended that generalized theory for anisotropic media. Later on, during the year 1991-1993 Green and Naghdi [4, 5, 6] introduced a new theory of thermoelasticity and divide their theory into three parts, referred as types I, II and III. In an extensive review work on the development of generalized/ hyperbolic thermoelasticity till 1998 is available in the review article of Chandrasekharaiah [7].

Diffusion can be defined as the random movement of the particles from the higher concentrated regions to the lower concentrated regions because of the non-zero concentration gradient which can be expressed in terms of changes of the concentration at that position. In recent past it has been observed that there are so many researchers are interested to study on this aspects due to a great application in geophysics and in industry e.g. so many oil companies are interested in the thermodiffusion process for more efficient extraction of oil from the oil deposits.Diffusion is used to form base and emitter in bipolar transistors, form integrated resistors and used to introduce 'dopants' in controlled amounts into the semiconductor substrate. The thermodiffusion in elastic solids is due to the coupling among the temperature, elastic strain and mass diffusion in addition with the exchange between heat and mass in the nature.

In 1974, Nowacki [8-11] developed the theory of coupled thermoelastic diffusion. The generalized theory in thermoelastic diffusion was introduced by Sherief et al. [12] in 2004. Again, in the year 2005, Sherief et al. [13] studied a half space problem in the theory of generalized thermoelastic diffusion. The influence of diffusion on generalized thermoelastic problems of infinite body with a cylindrical cavity studied by Rong-hou et al. [14]. Singh [15, 16] in his couple of papers discussed the reflection of waves from the free surface in generalized thermoelastic diffusion. In recent times Kumar and Kansal [17, 18] studied about the Rayleigh and Lamb wave propagation on free surface in transversely isotropic thermoelastic diffusion. Sharma et al. [19-20] studied on thermodiffusive surface wave propagation in heat conducting materials and Kumar et al. [21] discussed on the plane strain deformation in generalized thermoelastic diffusion in 2007-2008.

The linear theory of micropolar thermoelasticity has been developed by extending the theory of micropolar elasticity including thermal effect by Eringen [22, 23] and Nowacki [24, 25]. Minagawa et al. [26] discussed the propagation of plane harmonic waves in a cubic micropolar medium. Kumar and Rani [27] studied time harmonic sources in a thermally conducting cubic crystal and Mechanical/ thermal sources in a micropolar thermoelastic medium with cubic symmetry by Kumar and Aliawalia [28]. In the year 2006, Kumar and Aliawalia [29] studied on deformation due to time harmonic sources in micropolar thermoelastic medium with two relaxation times.

Fractional calculus has been used successfully to modify many existing model of physical process. In the formulation of tautochrone problem, Abel applied fractional calculus to solve integral equation and that was first application of fractional derivatives. Using fractional derivatives, for the description of viscoelastic materials, Caputo [30], Caputo and Mainardi [31] found an agreement between the experimental results with theoretical one. Recently, Sherief et al. [32] introduced

Author αΩ : Department of Mathematics, Bengal Engineering and Science University, Shibpur, India, 711103.
E-mail α : shaw_soumen@rediffmail.com,
E-mail Ω : bmukharjee2006@yahoo.co.in

Global Journal of Researches in Engineering (I) Volume XI Issue VII Version I December 2011

the fractional ordered thermoelasticity by using Caputo's [30] Phenomenon.

In the present investigation we are concerned about the interaction between thermoelastic diffusion with micropolar-elasticity by using fractional derivatives.

Nomenclature:

U : Internal energy per unit mass, K : Kinetic energy per unit mass, L : Power of external force, Q: Heat absorbed by the material body, W: Quantity of heat generated in unit time in unit volume, $\vec{q}$: Heat flux vector, $\vec{F}$: External body force, $\vec{l}$: Body Couple, σ_l : Microstretch rotatory inertia, ρ : Constant mass density of the medium, $\vec{u}$: Displacement vector, $\vec{\xi}$: Micro-displacement vector, $\vec{\phi}$: Microrotation vector, j : Microinertia, $\vec{n}$: Outward drawn normal vector, t_{ij} : Stress tensor, m_{ij} : Coupled stress tensor, ε_{ij} : Micropolar strain tensor, $\upsilon_i : \dot{\phi}_i$, time rate change of microrotation component, $\gamma_{ij} : \phi_{i,j}$, microrotation tensor, $\vec{\eta}$: Flow of diffusion mass vector, T : Absolute temperature, S : Entropy per unit mass, P : Chemical potential per unit mass, C : Concentration, λ, μ : Lame' constants, $\alpha, \beta, \gamma, \kappa$: Micropolar elastic constants, $\beta_1 = (3\lambda + 2\mu + \kappa)\alpha_t$, $\beta_2 = (3\lambda + 2\mu + \kappa)\alpha_c$, α_t : Coefficient of linear thermal expansion, α_c : Coefficient of linear diffusion expansion, C_E : Specific heat at constant strain, a : Measure of thermodiffusion effects, b : Measure of thermodiffusive effects, υ : Thermal relaxation time, τ : Diffusion relaxation time, D : Thermoelastic diffusion constant, K : Coefficient of thermal conductivity, ε_{ijk} : Permutation tensor, δ_{ij} : Kronecker delta.

II. Fundamental Equations

The Riemann-Liouville fractional integral is introduced as a natural generalization of the convolution type integral [33, 34, 35, 36, 37],

$$I^{\alpha} f(t) = \frac{1}{\Gamma(\alpha)} \int_0^t (t-\tau)^{\alpha-1} f(\tau)\, d\tau, \quad (\alpha > 0) \tag{2.1}$$

The Laplace transform for this fractional integral is defined by,

$$L[I^{\alpha} f(t)] = \frac{1}{s^{\alpha}} L[f(t)]. \tag{2.2}$$

The Riemann-Liouville derivative of fractional of fractional order α is defined as the left-inverse of the fractional integral I^{α} as

$$D_{RL}^{\alpha} f(t) = D^n I^{n-\alpha} f(t) = \frac{1}{\Gamma(n-\alpha)} \frac{d^n}{dt^n} \int_0^t (t-\tau)^{n-\alpha-1} f(\tau)\, d\tau, \quad n-1 < \alpha < n \tag{2.3}$$

and for Laplace transform, the initial values of the fractional integral $I^{n-\alpha} f(t)$ and its derivatives of order $k = 1,2,3,\ldots,n-1$ are required, where

$$L[D_{RL}^{\alpha} f(t)] = s^{\alpha} L[f(t)] - \sum_{k=0}^{n-1} s^{n-1-k} D^k I^{n-\alpha} f(0), \quad n-1 < \alpha < n. \tag{2.4}$$

An alternative definition of fractional derivative was proposed by Caputo [30] as,

$$D_C^{\alpha} f(t) = \frac{1}{\Gamma(n-\alpha)} \int_0^t (t-\tau)^{n-\alpha-1} \frac{d^n}{dt^n} f(\tau)\, d\tau, \quad n-1 < \alpha < n \tag{2.5}$$

and for the Laplace transform, this definition has an advantage, the initial values of $f(t)$ and its integer derivatives of order $k = 1,2,3,\ldots,n-1$ are required unlike the fractional ordered derivatives given by the equation (2.3), so that

$$L[D_C^{\alpha} f(t)] = s^{\alpha} L[f(t)] - \sum_{k=0}^{n-1} s^{\alpha-1-k} f^{(k)}(0), \quad n-1 < \alpha < n \tag{2.6}$$

Here we shall use the fractional derivatives of order $\alpha \in (0,1]$, according to Caputo [30].

Let V be any arbitrary volume element of a material body bounded by the closed surface A. Now the first law of thermoelasticity, the law of conservation of energy for the volume element V can be written in the following form:

$$\frac{d}{dt}\int_V \rho(\mathrm{U}+\mathrm{K}) = \mathrm{L} + \frac{d\mathrm{Q}}{dt} \tag{2.7}$$

Here,

$$\mathrm{L} = \int_V \rho(F_i\,\dot{u}_i + l_i\,\upsilon_i)dV + \int_A (t_{ji}\,\dot{u}_i + m_{ji}\upsilon_i)n_j dA \tag{2.8}$$

$$\dot{\mathrm{Q}} = -\int_A \mathrm{q}_i n_i dA + \int_V W dV \tag{2.9}$$

Now the equations of the motion are as follows:

$$t_{ji,j} + \rho F_i = \rho \ddot{u}_i \tag{2.10}$$

$$m_{kl,k} + \varepsilon_{lmn} t_{mn} + \rho l_l = \rho \sigma_l \tag{2.11}$$

where $\sigma_k = j_{kl}\dot{\upsilon}_l + \varepsilon_{klm} j_{mn}\upsilon_l \upsilon_n$

After using divergence theorem and the equations of motion (2.10), (2.11) we obtain from the equation (2.7),

$$\frac{d}{dt}\int_V \rho(\mathrm{U}+\mathrm{K}) = \int_V \left[\frac{d}{dt}(\rho\mathrm{K}) + t_{ji}(u_{i,j} + \varepsilon_{ijk}\upsilon_k) + m_{ji}\upsilon_{i,j} + \rho\varepsilon_{ijk} j_{kl}\,\upsilon_i\upsilon_j\upsilon_l - q_{j,j}\right]dV \tag{2.12}$$

Now for micropolar thermoelastic solids, the deformations, micro-rotations and the changes of temperature are very small in natural state of the body, so a linear approximation is possible. For linear approximation we introduce, as was done by Eringen [23], the norm function of the array

$$W = (\vec{u}_{,i}, \vec{\phi}, \vec{\phi}_{,i}, \theta, \theta_{,i})$$

by $\varepsilon = \|W\| = (W \cdot W)^{1/2}$

and for some neighborhood of $\varepsilon = 0$, there exist positive constants k and n such that

$$\|\mathrm{O}(\varepsilon^n)\| \le k\,\varepsilon^n \quad , k > 0 , n > 0.$$

Now, for linear (1st order) approximation equation (2.12) can be written in the following form:

$$\rho\dot{\mathrm{U}} = \mathrm{t}_{ji}\dot{\varepsilon}_{ji} + m_{ji}\dot{\gamma}_{ij} - q_{i,i} + W \tag{2.13}$$

Where,

$$\varepsilon_{ji} = u_{i,j} + \varepsilon_{ijk}\phi_k \tag{2.14}$$

Now we consider the entropy balance law,

$$\rho T\dot{S} = -q_{i,i} + P\eta_{i,i} + W \tag{2.15}$$

The equation of conservation of mass:

$$\eta_{i,i} = -\dot{C} \tag{2.16}$$

where C is the concentration.

Now using the equations (2.15) and (2.16), from the equation (2.13) we obtain,

$$\rho\, dU = t_{ji} d\varepsilon_{ji} + m_{ji} d\gamma_{ij} + \rho T dS + P dC \tag{2.17}$$

For further proceedings, we introduce the helmholtz free energy function Ψ given by,

$$\Psi = U - TS \tag{2.18}$$

Now using the equation (2.17), the total differential of the function Ψ can be written as,

$$\rho\, d\Psi = t_{ji} d\varepsilon_{ji} + m_{ji} d\gamma_{ij} + P dC - \rho S dT \tag{2.19}$$

Here the function Ψ and all other functions under consideration can be expressed in terms of the independent variables $\varepsilon_{kl}, \gamma_{kl}, C, T$. Now by chain rule we get,

$$\rho\, d\Psi = \rho \frac{\partial \Psi}{\partial \varepsilon_{ji}} d\varepsilon_{ji} + \rho \frac{\partial \Psi}{\partial \gamma_{ij}} d\gamma_{ij} + \rho \frac{\partial \Psi}{\partial T} dT + \rho \frac{\partial \Psi}{\partial C} dC \tag{2.20}$$

Comparing the equations (2.19) and (2.20) we obtain,

$$t_{ji} = \rho \frac{\partial \Psi}{\partial \varepsilon_{ji}} \tag{2.21}$$

$$m_{ji} = \rho \frac{\partial \Psi}{\partial \gamma_{ij}} \tag{2.22}$$

$$P = \rho \frac{\partial \Psi}{\partial C} \tag{2.23}$$

$$S = -\frac{\partial \Psi}{\partial T} = -\frac{\partial \Psi}{\partial \theta} \tag{2.24}$$

Now expanding the function Ψ into the Taylor series in terms of independent variables about its natural state, we obtain,

$$\rho\Psi = \rho\Psi_0 + a_0\theta + A_{ji}\varepsilon_{ji} + b_0 C + B_{ij}\gamma_{ij} - \frac{\rho C_E}{2T_0}\theta^2 + \frac{1}{2}bC^2 + \frac{1}{2}A_{jilk}\varepsilon_{ji}\varepsilon_{lk} + a_{ji}\varepsilon_{ji}\theta + b_{ji}\varepsilon_{ji}C - a\theta C$$

$$+\frac{1}{2}B_{ijkl}\gamma_{ij}\gamma_{kl} + c_{ij}\gamma_{ij}\theta + d_{ij}\gamma_{ij}C + C_{jikl}\varepsilon_{ji}\gamma_{kl} + \dots\dots\dots \tag{2.25}$$

where $\theta = T - T_0$, T_0 is the temperature of the medium in natural state such that $\left|\frac{\theta}{T_0}\right| << 1$.

In the natural state of the material body, we consider,

$$\Psi = 0,\ \ \theta = 0,\ \ t_{ij} = 0,\ \ C = 0,\ \ \varepsilon_{ij} = 0,\ \ \gamma_{ij} = 0.$$

and we obtain,

$$\Psi_0 = 0,\ a_0 = 0,\ A_{ji} = 0,\ b_0 = 0,\ B_{ij} = 0.$$

Now equation (2.25) can be recast in following form keeping only second order terms,

$$\rho\Psi = -\frac{\rho C_E}{2T_0}\theta^2 + \frac{1}{2}bC^2 + \frac{1}{2}A_{jilk}\varepsilon_{ji}\varepsilon_{lk} + a_{ji}\varepsilon_{ji}\theta + b_{ji}\varepsilon_{ji}C - a\theta C + \frac{1}{2}B_{ijkl}\gamma_{ij}\gamma_{kl} + c_{ij}\gamma_{ij}\theta + d_{ij}\gamma_{ij}C$$

$$+ C_{jikl}\varepsilon_{ji}\gamma_{kl}\,. \tag{2.26}$$

Now from equations (2.21) – (2.24) and equation (2.26) we obtain,

$$t_{ij} = A_{ijkl}\varepsilon_{kl} + a_{ij}\theta + b_{ij}C + C_{ijlk}\gamma_{lk} \tag{2.27}$$

$$m_{ji} = B_{ijkl}\gamma_{kl} + c_{ij}\theta + d_{ij}C + C_{jikl}\varepsilon_{lk} \tag{2.28}$$

$$P = bC + b_{ji}\varepsilon_{ji} - a\theta + d_{ij}\gamma_{ij} \tag{2.29}$$

$$\rho S = \frac{\rho C_E}{T_0}\theta - a_{ji}\varepsilon_{ji} + aC - c_{ij}\gamma_{ij} \tag{2.30}$$

Now, in isotropic solids, from the equation (2.26) it has been seen that, since free energy is a scalar quantity, each terms of right hand side of equation (2.26) are scalars and hence,

$$A_{ijkl} = \lambda\delta_{ij}\delta_{kl} + (\mu+\kappa)\delta_{ik}\delta_{jl} + \mu\delta_{il}\delta_{jk}\ ,\ B_{ijkl} = \alpha\delta_{ij}\delta_{kl} + \beta\delta_{il}\delta_{jk} + \gamma\delta_{ik}\delta_{jl}\ ,\ C_{ijkl} = 0\ ,$$
$$a_{ij} = -\beta_1\delta_{ij}\ ,\ b_{ij} = -\beta_2\delta_{ij}\ ,\ c_{ij} = 0\ .$$

Diffusion is a random walk of the particles of the material body from higher concentrated regions to lower concentrated regions. Here d_{ij} represents the coupling coefficients between γ_{ij} and the concentration C of the body, in an isotropic solid the gradient of microrotation and the concentration are independent. So we take $d_{ij} = 0$ in isotropic body. In addition, the material is called spin-isotropic if $j_{\alpha\beta} = j\delta_{\alpha\beta}$.

Hence, in isotropic solids constitutive equations are,

$$t_{ij} = \lambda\varepsilon_{kk}\delta_{ij} + (\mu+\kappa)\varepsilon_{ij} + \mu\varepsilon_{ji} - \beta_1\theta\delta_{ij} - \beta_2 C\delta_{ij} \tag{2.31}$$

$$m_{ij} = \alpha\gamma_{kk}\delta_{ij} + \beta\gamma_{ij} + \gamma\gamma_{ji} \tag{2.32}$$

$$P = bC - \beta_2\varepsilon_{kk} - a\theta \tag{2.33}$$

$$\rho T_0 S = \rho C_E\theta + \beta_1 T_0\varepsilon_{kk} + aT_0 C \tag{2.34}$$

Now, the linear equations of balance law are given by,

$$t_{\beta\alpha,\beta} + \rho F_\alpha = \rho\ddot{u}_\alpha \tag{2.35}$$

$$m_{\beta\alpha,\beta} + \varepsilon_{\alpha mn}t_{mn} + \rho l_\alpha = \rho j\ddot{\phi}_\alpha \tag{2.36}$$

The linearized form of heat conduction is,

$$\rho T\dot{S} = -q_{i,i} + W \tag{2.37}$$

and without contradiction with the second law of thermodynamics, we assume a generalized form of Fourier's law of heat conduction equation as,

$$\left(1+\tau_0\frac{\partial^\alpha}{\partial t^\alpha}\right)q_i = -K\theta_{,i} \tag{2.38}$$

where α is a constant such that $0 \le \alpha \le 1$.

Now using the equations (2.34) and (2.37), from the equation (2.38) we obtain,

$$K\theta_{,ii} = \rho C_E\left(1+\tau_0\frac{\partial^\alpha}{\partial t^\alpha}\right)\frac{\partial\theta}{\partial t} + \beta_1 T_0\left(1+\tau_0\frac{\partial^\alpha}{\partial t^\alpha}\right)\frac{\partial\varepsilon_{kk}}{\partial t}\varepsilon_{kk} + aT_0\left(1+\tau_0\frac{\partial^\alpha}{\partial t^\alpha}\right)\frac{\partial C}{\partial t} - \left(1+\tau_0\frac{\partial^\alpha}{\partial t^\alpha}\right)W \tag{2.39}$$

We may consider the equation (2.39) is an extended version of fractional ordered heat conduction equation in isotropic, micropolar, thermodiffusive elastic solids.
Now we consider the equation of mass flux vector,

$$\left(1+\tau_1\frac{\partial^\alpha}{\partial t^\alpha}\right)\eta_i = -DP_{,i} \tag{2.40}$$

Now using the equation (2.16) and (2.33), from the equation (2.40) we obtain,

$$DbC_{,ii} = D\beta_2\nabla^2\varepsilon_{kk} + Da\theta_{,ii} + \left(1+\tau_1\frac{\partial^\alpha}{\partial t^\alpha}\right)\frac{\partial C}{\partial t} \tag{2.41}$$

This equation may be considered as the fractional ordered generalized mass diffusion equation in isotropic, micropolar elastic solids.

Special cases :
When $\alpha \to 1$, equations (2.39) and (2.41) reduces to

$$K\theta_{,ii} = \rho C_E\left(\dot\theta + \upsilon\ddot\theta\right) + \beta_1 T_0\left(\frac{\partial}{\partial t} + \upsilon\frac{\partial^2}{\partial t^2}\right)\varepsilon_{kk} + aT_0\left(\dot C + \upsilon\ddot C\right) - \left(1+\upsilon\frac{\partial}{\partial t}\right)W \tag{2.42}$$

$$DbC_{,ii} = D\beta_2\nabla^2\varepsilon_{kk} + Da\theta_{,ii} + \left(\frac{\partial}{\partial t} + \tau\frac{\partial^2}{\partial t^2}\right)C \tag{2.43}$$

These are the generalized heat conduction and mass diffusion equations in isotropic, micropolar elastic solids.
Again, when $\alpha \to 1,\ \kappa \to 0$ equations (2.39), (2.41) transform to

$$K\theta_{,ii} = \rho C_E\left(1+\upsilon\frac{\partial}{\partial t}\right)\frac{\partial\theta}{\partial t} + \beta_1 T_0\left(1+\upsilon\frac{\partial}{\partial t}\right)\left(\vec\nabla\cdot\dot{\vec u}\right) + aT_0\left(1+\upsilon\frac{\partial}{\partial t}\right)\frac{\partial C}{\partial t} - \left(1+\upsilon\frac{\partial}{\partial t}\right)W \tag{2.44}$$

$$DbC_{,ii} = D\beta_2\nabla^2\left(\vec\nabla\cdot\vec u\right) + Da\theta_{,ii} + \left(1+\tau\frac{\partial}{\partial t}\right)\frac{\partial C}{\partial t} \tag{2.45}$$

Equations (2.44) and (2.45) represent the heat conduction and mass diffusion equations for isotropic elastic solids, as was done by Sherief et al. [12].

III. A Variational Principle

The variational theorem in classical thermoelasticity first derived by Biot [38] and explained their applications by means of several examples. The variational theorem on classical thermoelastic diffusion was done by Sherief et al. [12]. In micropolar thermoelasticity the variational principle and uniqueness theorem was done by Eringen [22, 23]. Recently, a variational principle of fractional order generalized thermoelasticity was done by Youssef and Al-Lehaibi [39]. Now we shall present a compact derivation of a variational principle on fractional-ordered micropolar thermoelastic diffusion.
We consider,

$$w = \frac{1}{2}\int_V \left[A_{klmn}\varepsilon_{kl}\varepsilon_{mn} + B_{klmn}\gamma_{kl}\gamma_{mn}\right] dV\,, \tag{3.1}$$

where the integrand is homogeneous quadratic form of strain tensor and microrotation tensors.
Now we consider a virtual displacement i.e. for a neighboring state in which the displacement, strain tensor, microrotation tensors are changed by the quantities δu_i , $\delta\varepsilon_{ij}$, $\delta\gamma_{ij}$ respectively, we obtain,

$$\delta w = \frac{1}{2}\int_V \left[A_{klmn}\left(\delta\varepsilon_{kl}\right)\varepsilon_{mn} + A_{klmn}\varepsilon_{kl}\left(\delta\varepsilon_{mn}\right) + B_{klmn}\left(\delta\gamma_{kl}\right)\gamma_{mn} + B_{klmn}\gamma_{kl}\left(\gamma_{mn}\right)\right]dV \tag{3.2}$$

Now using the constitutive equations;

$$t_{ij} = \lambda\varepsilon_{kk}\delta_{ij} + (\mu+\kappa)\varepsilon_{ij} + \mu\varepsilon_{ji} - \beta_1\theta\delta_{ij} - \beta_2 C\delta_{ij} \tag{3.3}$$

$$m_{ij} = \alpha\gamma_{kk}\delta_{ij} + \beta\gamma_{ij} + \gamma\gamma_{ji} \tag{3.4}$$

Taking into account the equations of motion;

$$t_{\beta\alpha,\beta} + \rho F_\alpha = \rho\ddot{u}_\alpha \tag{3.5}$$

$$m_{\beta\alpha,\beta} + \varepsilon_{\alpha mn}t_{mn} + \rho l_\alpha = \rho j\ddot{\phi}_\alpha \tag{3.6}$$

Corresponding boundary conditions;

$$p_i = t_{ji}n_j \tag{3.7}$$

$$m_i = m_{ji}n_j \tag{3.8}$$

and using divergence theorem, we obtain,

$$\int_V \rho F_i\,\delta u_i\,dV + \int_A p_i\,\delta u_i\,dA + \int_V \rho l_i\,\delta\phi_i\,dV + \int_A m_i\,\delta\phi_i\,dA - \int_V \rho\ddot{u}_i\,\delta u_i\,dV - \rho j\int_V \ddot{\phi}_i\,\delta\phi_i\,dV$$
$$= \delta W - \beta_1\int_V \theta\,\delta\varepsilon_{kk}\,dV - \beta_2\int_V C\,\delta\varepsilon_{kk}\,dV \tag{3.9}$$

This is the first variational equation and it would be complete for uncoupled thermoelasticity if the temperature θ and the concentration C in last two integrations of the right hand side of the equation (3.9) were known. Taking into account the coupling between the strain field, temperature and concentration it is observed that θ and C are unknown. Hence it is necessary to introduce other relations considering the phenomena of heat conduction as well as mass diffusion.

Now we introduce one vector $\vec{H}$ as was done by Biot [40], related with entropy by,

$$\rho S = -div\vec{H} = -H_{i,i} \tag{3.10}$$

Again we know the relations,

$$\rho T_0 S = \rho C_E\theta + \beta_1 T_0\varepsilon_{kk} + aT_0 C \tag{3.11}$$

$$q_{i,i} = -\rho T_0\dot{S} \tag{3.12}$$

$$\left(1+\tau_0\frac{\partial^\alpha}{\partial t^\alpha}\right)q_i = -K\theta_{,i} \tag{3.13}$$

Now using the equation (3.10), from the relations (3.11), (3.12) and (3.13) we obtain,

$$-H_{i,i} = \frac{\rho C_E}{T_0}\theta + \beta_1\varepsilon_{kk} + aC \tag{3.14}$$

$$q_i = T_0\dot{H}_i \tag{3.15}$$

$$T_0\left(\frac{\partial}{\partial t} + \tau_0\frac{\partial^{1+\alpha}}{\partial t^{1+\alpha}}\right)H_i = -K\theta_{,i} \tag{3.16}$$

Now multiplying both sides of equation (3.16) by δH_i and then integrating over the region V of the body we obtain,

$$\int_V \left[\theta_{,i} + \frac{T_0}{K}\left(\frac{\partial}{\partial t} + \tau_0 \frac{\partial^{1+\alpha}}{\partial t^{1+\alpha}} \right) H_i \right] \delta H_i \, dV = 0 \tag{3.17}$$

Now,

$$\int_V \theta_{,i}\, \delta H_i \, dV = \int_A \theta n_i \,\delta H_i \, dA + \frac{\rho C_E}{T_0}\int_V \theta\,\delta\theta \, dV + \beta_1 \int_V \theta\,\delta\varepsilon_{kk} \, dV + a\int_V \theta\,\delta C \, dV \tag{3.18}$$

Using the equation (3.18), from the equation (3.17) we obtain,

$$\int_A \theta n_i \,\delta H_i \, dA + \frac{\rho C_E}{T_0}\int_V \theta\,\delta\theta \, dV + \beta_1 \int_V \theta\,\delta\varepsilon_{kk} \, dV + a\int_V \theta\,\delta C \, dV + \frac{T_0}{K}\int_V \left(\frac{\partial}{\partial t} + \tau_0 \frac{\partial^{1+\alpha}}{\partial t^{1+\alpha}} \right) H_i \delta H_i \, dV = 0 \tag{3.19}$$

Here we introduce heat potential function,

$$IP = \frac{\rho C_E}{2T_0}\int_V \theta^2 \, dV \tag{3.20}$$

And

$$\delta IP = \frac{\rho C_E}{T_0}\int_V \theta\,\delta\theta \, dV \tag{3.21}$$

Again, the heat dissipation function ID, where

$$\delta ID = \frac{T_0}{K}\int_V \left(\frac{\partial}{\partial t} + \tau_0 \frac{\partial^{1+\alpha}}{\partial t^{1+\alpha}} \right) H_i \,\delta H_i \, dV \tag{3.22}$$

Therefore, from equation (3.19) we obtain,

$$\int_A \theta n_i \,\delta H_i \, dA + \delta(IP + ID) + \beta_1 \int_V \theta\,\delta\varepsilon_{kk} \, dV + a\int_V \theta\,\delta C \, dV = 0 \tag{3.23}$$

This is the second variational equation connected with heat conduction.

In order to obtain the variational equation connected with the process of thermodiffusion we introduce another vector as was done by Sherief et al [12] related with mass concentration by,

$$C = -div\vec{G} = -G_{i,i} \tag{3.24}$$

Now from the relations,

$$P = bC - \beta_2 \varepsilon_{kk} - a\theta \ , \tag{3.25}$$

$$\eta_{i,i} = -\dot{C} \tag{3.26}$$

$$\left(1 + \tau_1 \frac{\partial^{\alpha}}{\partial t^{\alpha}} \right)\eta_i = -DP_{,i} \tag{3.27}$$

and using the equation (3.24) we obtain,

$$\eta_i = \dot{G}_i \tag{3.28}$$

$$\left(\frac{\partial}{\partial t} + \tau_1 \frac{\partial^{1+\alpha}}{\partial t^{1+\alpha}} \right) G_i = -DP_{,i} \tag{3.29}$$

Now multiplying both sides of equation (3.29) by δG_i and then integrating over the region V we obtain,

$$\int_V \left[P_{,i} + \frac{1}{D}\left(\frac{\partial}{\partial t} + \tau_1 \frac{\partial^{1+\alpha}}{\partial t^{1+\alpha}}\right) G_i \right] \delta G_i \, dV = 0 \tag{3.30}$$

Now,

$$\int_V P_{,i} \, \delta G_i \, dV = \int_A P \, n_i \, \delta G_i \, dA + b \int_V C \, \delta C \, dV - \beta_2 \int_V \varepsilon_{kk} \, \delta C \, dV - a \int_V \theta \, \delta C \, dV \tag{3.31}$$

Using the equation (3.31), from the equation (3.30) we obtain,

$$\int_A P \, n_i \, \delta G_i \, dA - \beta_2 \int_V \varepsilon_{kk} \, \delta C \, dV + b \int_V C \, \delta C \, dV - a \int_V \theta \, \delta C \, dV + \frac{1}{D}\int_V \left(\frac{\partial}{\partial t} + \tau_1 \frac{\partial^{1+\alpha}}{\partial t^{1+\alpha}}\right) G_i \, \delta G_i \, dV = 0 \tag{3.32}$$

Now we define, the diffusion potential function,

$$IP' = \frac{b}{2}\int_V C^2 \, dV \tag{3.33}$$

And

$$\delta IP' = b \int_V C \, \delta C \, dV \tag{3.34}$$

Again, the diffusive dissipation function ID', where

$$\delta ID' = \frac{1}{D}\int_V \left(\frac{\partial}{\partial t} + \tau_1 \frac{\partial^{1+\alpha}}{\partial t^{1+\alpha}}\right) G_i \, \delta G_i \, dV \tag{3.35}$$

Therefore, from the equation (3.32) we obtain,

$$\int_A P \, n_i \, \delta G_i \, dA - \beta_2 \int_V \varepsilon_{kk} \, \delta C \, dV - a \int_V \theta \, \delta C \, dV + \delta\left(IP' + ID'\right) = 0 \tag{3.36}$$

Equation (3.36) represents the third variational equation.

Now from the equations (3.9), (3.23) and (3.36) we obtain,

$$\delta\left[\ \mathcal{W} + IP + IP' + ID + ID' - \beta_2 \int_V C \, \varepsilon_{kk} \, dV \ \right] = -\int_A \theta \, n_i \, \delta H_i \, dA - \int_A P \, n_i \, \delta G_i \, dA + \int_V \rho F_i \, \delta u_i \, dV$$
$$+ \int_A p_i \, \delta u_i \, dA + \int_V \rho \, l_i \, \delta \phi_i \, dV + \int_A m_i \, \delta \phi_i \, dA - \rho \int_V \ddot{u}_i \, \delta u_i dV - \rho j \int_V \ddot{\phi}_i \, \delta \phi_i \, dV \tag{3.37}$$

Now, $\delta u_i = \dot{u}_i \, dt$, $\delta \theta = \dot{\theta} \, dt$, $\delta H_i = \dot{H}_i \, dt$, $\delta G_i = \dot{G}_i \, dt$ etc.

Therefore, equation (3.37) reduces to,

$$\frac{d}{dt}\left[\ \mathcal{W} + \mathrm{K} + IP + IP' + ID + ID' + \frac{a}{b}\int_V P \theta \, dV \ \right] = \int_V \rho F_i \, \dot{u}_i \, dV + \int_V \rho \, l_i \, \dot{\phi}_i \, dV + \int_A p_i \, \dot{u}_i \, dA + \int_A m_i \, \dot{\phi}_i \, dA$$
$$+ \frac{K}{T_0}\int_A \theta \frac{\partial \theta}{\partial n} dA + \int_A P \frac{\partial \eta_i}{\partial n} dA \tag{3.38}$$

where the mass concentration C is written in terms of the chemical potential P.

IV. Uniqueness Theorem

Uniqueness theorem states that there is only one solution of the equations (2.25), (2.36) (2.39) and (2.41) subject to the boundary conditions;

$$p_i = t_{ji}\,n_j\,,\; m_i = m_{ji}\,n_j\,,\; \theta = \theta_1(x,t)\,,\; P = P_1(x,t) \quad , \quad x\in A,\; t>0$$

and initial conditions;

$$u_i(x,0)=u_{i0}(x),\; \dot{u}_i(x,0)=\dot{u}_{i0}(x),\; \phi_i(x,0)=\phi_{i0}(x),\; \dot{\phi}_i(x,0)=\dot{\phi}_{i0}(x),\; \theta(x,0)=\theta_0(x)$$

$$P(x,0)=P_0(x)\,,\; \dot{P}(x,0)=\dot{P}_0(x)\,,\; x\in V,\; t=0\,,$$

where the body occupying the region V bounded by the surface A.

Proof : We consider, if possible, there exist two sets of solutions (u_i,ϕ_i,θ,P) and $(u_i',\phi_i',\theta',P')$.

We take, $\hat{u}_i = u_i - u_i'$, $\hat{\phi}_i = \phi_i - \phi_i'$, $\hat{\theta}_i = \theta_i - \theta_i'$, $\hat{P} = P - P'$

Hence the solution $(\hat{u}_i,\hat{\phi}_i,\hat{\theta},\hat{P})$ must satisfy the equations of motion, heat conduction equation and equation of mass diffusion with no body forces, body couples and without heat sources term i.e.,

$$t_{\beta\alpha,\beta} = \rho\ddot{u}_\alpha \tag{4.1}$$

$$m_{\beta\alpha,\beta} + \varepsilon_{\alpha mn}t_{mn} = \rho j\ddot{\phi}_\alpha \tag{4.2}$$

$$K\theta_{,ii} = \rho C_E\left(1+\tau_0\frac{\partial^\alpha}{\partial t^\alpha}\right)\frac{\partial\theta}{\partial t} + \beta_1 T_0\left(1+\tau_0\frac{\partial^\alpha}{\partial t^\alpha}\right)\frac{\partial\varepsilon_{kk}}{\partial t}\varepsilon_{kk} + aT_0\left(1+\tau_0\frac{\partial^\alpha}{\partial t^\alpha}\right)\frac{\partial C}{\partial t} \tag{4.3}$$

$$DbC_{,ii} = D\beta_2\nabla^2\varepsilon_{kk} + Da\theta_{,ii} + \left(1+\tau_1\frac{\partial^\alpha}{\partial t^\alpha}\right)\frac{\partial C}{\partial t} \tag{4.4}$$

with homogeneous boundary and initial conditions.

Thus we arrive with a system for which the displacement $\hat{u}_i$, microrotation $\hat{\phi}_i$, temperature $\hat{\theta}$ and the chemical potential $\hat{P}$ are vanishes inside the body initially and the surface traction $\hat{p}_i$, surface couple $\hat{m}_i$, temperature $\hat{\theta}$ and the chemical potential $\hat{P}$ are vanishes on surface A.

Hence, it is enough to prove that the measure of strain tensor, microrotation tensor, temperature and the chemical potential are vanishes inside the body.
Now from the equation (3.38) we obtain,

$$\frac{d}{dt}\left[\ \mathrm{W} + \mathrm{K} + IP + IP' + ID + ID' + \frac{a}{b}\int_V P\theta\, dV\ \right] = 0 \quad \text{or,}$$

$$\frac{d}{dt}\left[\ \mathrm{W} + \mathrm{K} + IP + IP'\ \right] = -\frac{a}{b}\int_V P\theta\, dV - \frac{T_0}{K}\int_V \dot{H}_i\left(1+\tau_0\frac{\partial^\alpha}{\partial t^\alpha}\right)\dot{H}_i\, dV - \frac{1}{D}\int_V \dot{G}_i\left(1+\tau_1\frac{\partial^\alpha}{\partial t^\alpha}\right)\dot{G}_i\, dV \le 0 \tag{4.5}$$

The integral in the left hand side of equation (4.5) is zero initially. On the other hand the inequality proves that the left hand side of the equation is either zero or decreases taking negative values. Since the integrand is a sum of squares and vanishes at $t=0$, therefore, first possibility holds.

Hence, $\mathcal{W} + \mathrm{K} + IP + IP' = 0$, and $\hat{u}_i = 0$, $\hat{\hat{\phi}}_i = 0$, $\hat{\theta} = 0$, $\hat{P} = 0$, $\hat{\varepsilon}_{ij} = 0$, $\hat{\gamma}_{ij} = 0$ for $t \geq 0$.

Since $\hat{t}_{ij}$ and $\hat{m}_{ij}$ are the linear functions of $\hat{\varepsilon}_{ij}$, $\hat{\gamma}_{ij}$, $\hat{\theta}$ and $\hat{P}$ which are zero for $t \geq 0$,

therefore, $\hat{t}_{ij} = 0 = \hat{m}_{ij}$, $t \geq 0$.

Hence, $\hat{u}_i = \hat{\phi}_i = \hat{\theta} = \hat{P} = 0$ for $t \geq 0$.

This completes the proof of uniqueness theorem.

References Références Referencias

1. H. Lord, Y. Shulman (1967). A generalized dynamical theory of thermoelasticity. J. Mech. Phys. Solid, 15, 299-309.
2. A.E. Green, K.A. Lindsay (1972). Thermoelasticity. J. Elast., 2, 1-7.
3. R.S. Dhaliwal, H. Sherief (1980). Generalized thermoelasticity for anisotropic media. Quarterly Appl. Math.,33, 1-8.
4. A.E. Green, P.M. Naghdi (1991). A re-examination of the basic postulates of thermomechanics. Proc. Roy. Soc. London, A-432, 171-194.
5. A.E. Green, P.M. Naghdi (1992). On undamped heat waves in an elastic solid. J. Thermal Stresses, 15, 253-264.
6. A.E. Green, P.M. Naghdi (1993). Thermoelasticity without energy dissipation. J. Elast., 31, 189-209.
7. D.S. Chandrasekharaiah (1998). Hyperbolic thermoelasticity: a review of recent literature. J. Appl. Mech. Rev., 51, 705-729.
8. W. Nowacki (1974). Dynamical problems of thermodiffusion in solids I. Bull. Acad. Pol. Sci. Ser. Tech., 22, 55-64.
9. W. Nowacki (1974). Dynamical problems of thermodiffusion in solids II. Bull. Acad. Pol. Sci. Ser. Tech., 22, 129-135.
10. W. Nowacki (1974). Dynamical problems of thermodiffusion in solids III. Bull. Acad. Pol. Sci. Ser. Tech., 22, 257-266.
11. W. Nowacki (1974). Dynamical problems of thermodiffusion in solids. Proc. Vib. Prob., 15, 105-128.
12. H.H. Sherief, F.A. Hamza, H. A. Saleh (2004). The theory of generalized thermoelastic diffusion. Int. J. Eng. Sci., 42, 591-608.
13. H.H. Sherief, H.A. Saleh (2005). A half-space problem in theory of generalized thermoelastic diffusion. Int. J. solids structures, 42, 4484-4493.
14. Xia Rong-hou, Tian Xiao-geng, Shen Ya-peng (2009). The influence of diffusion on generalized thermoelastic problems of infinite body with a cylindrical cavity. Int. J. Eng. Sci., 47, 669-679.
15. B. Singh (2005). Reflection of P and SV waves from free surface of an elastic solid with generalized thermodiffusion. J. Earth Syst. Sci., 114(2), 159-168.
16. B. Singh (2006). Reflection of SV waves from free surface of an elastic solid in generalized thermodiffusion. J. Sound Vib., 291, 764-778.
17. R. Kumar, T. Kansal (2008). Propagation of Rayleigh wave on free surface of transversely isotropic generalized thermoelastic diffusion. Appl. Math. Mech. 29(11), 1451-1462.
18. R. Kumar, T. Kansal (2008). Propagation of Lamb waves in transversely isotropic thermoelastic diffusion plate. Int. J. solids structures, 45, 5890-5913.
19. J. N. Sharma (2007). Generalized thermoelastic diffusive waves in heat conducting materials. J. Sound Vib., 301, 979-993.
20. J. N. Sharma, Y. D. Sharma, P. K. Sharma (2008). On the propagation of elasto-thermodiffusive surface waves in heat conducting materials, J. Sound Vib., 351(4), 927-938.
21. N. Sharma, R. Kumar (2008). Paras Ram, Plane strain deformation in generalized thermoelastic diffusion. Int. J. Thermophys., 29, 1503-1522.
22. A.C. Eringen (1966). Linear theory of micropolar elasticity. J. Math. Mech., 15, 909-923.
23. A.C. Eringen (1999). Microcontinuum field theories, I. Foundations and solids. Springer-Verlag, New York.
24. W. Nowacki (1970). Theory of micropolar elasticity, Udline course and lecture no.25. Springer-Verlag, Wien.
25. W. Nowacki (1977). Theory of disclinations elastic cosserat media. Arch. Mech., 29, 531-547.
26. S. Minagawa, K. Arakawa, M. Yamada (1981). Dispersion curves for waves in a cubic micropolar medium with reference to estimations of the material constants for diamond. Bull. JSME, 24, 22-28.
27. R. Kumar, L. Rani (2003). Elastodynamics of time harmonic sources in a thermally conducting cubic crystal. Int. J. Appl. Mech. Engg. 8, 637-650.
28. R. Kumar, P. Ailawalia (2007). Mechanical/ Thermal sources in a micropolar thermoelastic medium possessing cubic symmetry without energy dissipation. Int. J. Thermophys., 28, 342- 367.
29. R. Kumar, P. Ailawalia (2006). Deformation due to time harmonic sources in micropolar thermoelastic medium possessing cubic symmetry with two

relaxation times. Appl. Math. Mech. (English Edition), 27, 781- 792.

30. M. Caputo (1967). Linear model of dissipation whose Q is almost frequency independent-II. Geophysics J. Roy. Astro. Soc., 13, 529-539.
31. M. Caputo, F. Mainardi (1971). A new dissipation model based on memory mechanism. Pure Appl. Geophysics, 91(1),
32. H.H. Sherief, A.M.A. El-Sayed, A.M.A. El-Latief (2010). Fractional order theory of thermoelasticity. Int. J. Solids Struc., 47, 269-275.
33. K.S. Miller, B. Ross (1993). An introduction to the fractional calculus and fractional differential equations. John Wiley and Sons, Inc., New York.
34. I.Podlubny (1999). Fractional differential equations. Academic Press, New York.
35. Y.Z. Povstenko (2009). Thermoelasticity that uses fractional heat conduction equation. J. Appl. Math. Sci., 162(2), 296-305.
36. Y.Z. Povstenko (2005). Fractional heat conduction equation and associated thermal stresses. J. Thermal Stresses, 28(1), 83-105.
37. Y.Z. Povstenko (2005). Thermoelasticity based on fractional heat conduction equation. Proc. 6th int. Cong. Thermal Stresses, Vienna, 501-504.
38. M. A. Biot (1959). New thermomechanical reciprocity relations with application to thermal stresses. J. Aero-Space science, 26, 7, 401-408.
39. H.M. Youssef, E.A. Al-Lehaibi (1010). Variational principle of fractional order generalized thermoelasticity. Appl. Math. Lett., 23, 1183-1187.
40. M. A. Biot (1956). Thermoelasticity and irreversible thermo-dynamics. J. Appl. Phys., 27, 240-253.

Global Journal of researches in engineering
Numerical Methods
Volume 11 Issue 7 Version 1.0 December 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 2249-4596 & Print ISSN: 0975-5861

Attribute Relational Analysis (ARA) For Coherent Association Rules: A Post Mining Process For Parallel Edge Projection And Pruning (PEPP) Based Sequence Graph Protrude Approach For Closed Itemset

By Kalli Srinivasa Nageswara Prasad, Prof. S. Ramakrishna
Sri Venkateswara University, Tirupati, Andhra Pradesh, India

Abstract - Association rules present one of the most impressive techniques for the analysis of attribute associations in a given dataset related to applications related to retail, bioinformatics, and sociology. In the area of data mining, the importance of the rule management in associating rule mining is rapidly growing. Usually, If datasets are large, the induced rules are large in volume. The density of the rule volume leads to the obtained knowledge hard to be understood and analyze. One better way of minimizing the rule set size is eliminating redundant rules from rule base. Many efforts have been made and various competent and excellent algorithms have been proposed. But all of these models relying either on closed itemset mining or expert's evaluation. None of these models are proven best in all data set contexts. Closed itemset model is missing adaptability and expert's evaluation process is resulting different significance for same rule under different expert's view. To overcome these limits here we proposed a post mining process called ARA as an extension to our earlier proposed closed itemset mining algorithm called PEPP.

Keywords : post mining, association rule mining, closed itemset, PEPP, Inference analysis, rule pruning.

GJRE-I Classification : FOR Code: 080109

Attribute Relational Analysis (ARA) For Coherent Association Rules: A Post Mining Process For Parallel Edge Projection And Pruning (PEPP) Based Sequence Graph Protrude Approach For Closed Itemset

Kalli Srinivasa Nageswara Prasad[α], Prof. S. Ramakrishna[Ω]

Abstract - Association rules present one of the most impressive techniques for the analysis of attribute associations in a given dataset related to applications related to retail, bioinformatics, and sociology. In the area of data mining, the importance of the rule management in associating rule mining is rapidly growing. Usually, If datasets are large, the induced rules are large in volume. The density of the rule volume leads to the obtained knowledge hard to be understood and analyze. One better way of minimizing the rule set size is eliminating redundant rules from rule base. Many efforts have been made and various competent and excellent algorithms have been proposed. But all of these models relying either on closed itemset mining or expert's evaluation. None of these models are proven best in all data set contexts. Closed itemset model is missing adaptability and expert's evaluation process is resulting different significance for same rule under different expert's view. To overcome these limits here we proposed a post mining process called ARA as an extension to our earlier proposed closed itemset mining algorithm called PEPP.

Keywords : post mining, association rule mining, closed itemset, PEPP, Inference analysis, rule pruning.

I. Introduction

In general, association rules tend to deliver an efficient method of analysing binary or discretized data sets that are large in volume. One common practice is to determine relationships between binary variables in transaction databases, which is known as 'market basket analyses. In the case of non-binary data, initially data being coded as binary and then association rules will be used to analyse. Association rules having their impact on analysing large binary datasets and considered as versatile approach for modern applications such as detection of bio-terrorist attacks [1] and the analysis of gene expression data [2], to the analysis of Irish third level education applications [4].

Author α : Research Scholar in Computer Science ,Sri Venkateswara University, Tirupati, Andhra Pradesh, India, Mobile: 9490171769
E-mail : kallisnprasad@gmail.com
Author Ω : Department of Computer Science, Sri Venkateswara University, Tirupati, Andhra Pradesh, India, Mobile: 944149572
E-mail : profsramakrishnasvu@gmail.com

The steps involved in a typical association rule analysis are "Coding of data as binary if data is not binary" -> "Rule generation" -> "Post-mining". This survey focused on post mining. It was a century after the introduction of association rules (associations initially discussed in 1902), it is still continuing that the absence of items from transactions is often ignored in analyses.

a) Association rule mining

Given a set I that is non-empty, a rule of association is a statement of the form $X \Rightarrow Y$, where $X, Y \subset I$ such that $X \neq \emptyset$, $Y \neq \emptyset$, and $X \cap Y = \emptyset$. The dataset X is called the antecedent of the rule, the set Y is called the consequent of the rule, and we shall call I the master itemset. Association rules are generated over a large set of transactions, denoted by T. Xn association rule can be deemed interesting if the items involved occur together often and there are suggestions that one of the sets might in some cases lead to the presence of the other set. Xn association rules are characterised as interesting, or not, based on mathematical notions called 'support', 'confidence' and 'lift'. Although there are now a multitude of measures of interestingness available to the analyst, many of them are still based on these three functions.

In many applications, it is not only the presence of items in the antecedent and the consequent parts of an association rule that may be of interest. Consideration, in many cases, should be given to the relationship between the absence of items from the antecedent part and the presence or absence of items from the consequent part. Further, the presence of items in the antecedent part can be related to the absence of items from the consequent part; for example, a rule such as {margarine} ⇒ {not butter}, which might be referred to as a 'replacement rule'. One way to incorporate the absence of items into the association rule mining paradigm is to consider rules of the form $X \Rightarrow / Y$ [20]. Another is to think in terms of negations. Suppose $X \subset I$, then write $\neg X$ to denote the absence or negation of the item, or items, in X from a transaction. Considering X as a binary {0, 1} variable, the presence

December 2011 51 Global Journal of Researches in Engineering (I) Volume XI Issue VII Version I

of the items in X is equivalent to X = 1, while the negation ¬X is equivalent to XX`€=X`€0. The concept of considering association rules involving negations, or "negative implications", is due to Silverstein et al [20].

b) Post Mining

Pruning rules and detection of rule interestingness are employed in the post-mining stage of the association rule mining paradigm. However, there are a host of other techniques used in the post-mining stage that do not naturally fall under either of these headings. Some such techniques are in the area of redundancy- removal. There are often a huge number of association rules to contend with in the post-mining stage and it can be very difficult for the user to identify which ones are interesting. Therefore, it is important to remove insignificant rules, prune redundancy and do further post-mining on the discovered rules [18, 11]. Liu et al [11] proposed a technique for pruning and summarising discovered association rules by first removing those insignificant associations and then forming direction-setting rules, which provide a sort of summary of the discovered association rules. Lent et al.[19] proposed clustering the discovered association rules.

c) Influences of Input formats

The input formats that influence the post mining methodologies are binary data, text data and streaming data. In association rule mining, much of recent research work has focused on the difficult problem of mining data streams such as click stream analysis, intrusion detection, and web-purchase recommendation systems. In the case of streaming data, it is not possible to perform mining on cached and fixed data records. The attempt of caching data leads to memory usage issues, and the attempt of mining static data leads to worst time complexity since continuous dataset update leads to continuous passes through dataset. From the data mining point of view, texts are complex data giving raise to interesting challenges. First, texts may be considered as weakly structured, compared with databases that rely on a predefined schema. Moreover, texts are written in natural language, carrying out implicit knowledge, and ambiguities. Hence, the representation of the content of a text is often only partial and possibly noisy. One solution for handling a text or a collection of texts in a satisfying way is to take advantage of a knowledge model of the domain of the texts, for guiding the extraction of knowledge units from the texts. One of the obvious hot topics of data mining research in the last few years has been rule discovery from binary data. It concerns the discovery of set of attributes from large binary records such that these attributes are true within a same line often enough. It is then easy to derive rules that describe the data, the popular association rules though the interest of frequent sets goes further.

Based on the proposals [14, 12, 10, 3] recently cited in literature and their motivations, it is observable that the process of rule pruning will opt to one of the two models.

1) Rule pruning under post mining process that demands domain experts observation
2) Rule pruning under post mining process that aims to avoid domain experts role in pruning process.

As in the first case rule pruning accuracy depends on domain expert's awareness on attribute relations. In this case it is always obvious to prune the rules under reliable domain expert's observation. In second case the models prune the rules based on dynamically determined attribute relations. This limits the solution to specific data models. Hence it is not adaptable for all contexts.

The Rest of the paper organized as; in section II we discussed the most frequently cited post mining models to improve rule accuracy. Section III briefs the post mining process [] that we opted. Section IV briefs the approach of closed itemset mining, and inference approach for itemset pruning. Section V explores the process of Attribute Relational weights analysis for rule pruning. Results discussion and comparative study will be in Section VI that fallowed by conclusion and references.

II. Related Work

Huawen Liu, et al[14] proposed post processing approach for rule reduction using closed set to filter superfluous rules from knowledge base in a post-processing manner that can be well discovered by a closed set mining technique. Most of these methods are based on the rule structure analysis where the relations between the rules have been analysed using corresponding problem. This procedure claims to eliminate noise and redundant rules in order to provide users with compact and precise knowledge derived from databases by data mining method. Further, a fast rule reduction algorithm using closed set is introduced. Other endeavours have been attempted to prune the rule bases directly. The typical cases have been elaborated and illustrated by eminent people from all over.

Modest number of proposals is addressed on pre-pruning and post-pruning. In a line, the pruning operation occurs at the phase of generation of significant rules. To add to this, the post pruning technique mainly concerns primarily emphasises that pruning operation occurs after the rule generation, among which the rule cover is a representative case. To extract interesting rules, approri knowledge has also been taken into account in literatures and a template denotes which attributes should occur in the antecedent or the consequent part of the rule.

An association rule is an implication expression illustrates a kind of association relation. A rule is said to be interesting or valid if its support and confidence are user specified minimum support and confidence thresholds respectively. The association rule primarily comprises two phases based on the identification of frequent item sets from the given data mining contexts. However, the problem of massive real world data transactions can be rounded off by adopting other alternatives, which in turn benefits in lossless representation of data. Theoretically, transaction database and relation database are two different inter-transformable representations of data.

The production of association mining is a rule base with hundreds to millions of rules. In order to highlight the important and key ones, certain other rules are proposed which are Second –order rule which states that if the cover of the item set is known, then the corresponding relation can be easily derived. All the technical definitions given hence forth deal with the transaction of the data through item-sets in association mining. The equivalent property significantly states that rules and classes of the same hierarchical database support the power in the content. Traditional data mining techniques are implemented in order to justify the property and its corresponding definition in the specified context. The thus identified second order rules can be used to filter out useless rules out of the priority rule-set.

The effectiveness and efficiency of the classical methods in plating the rules is thoroughly verified under the 2.8 GHZ Pentium PC. Two group experiments were conducted to prune the insignificant association rules and to remove useless association classification rules. Removal of non-predictive rules by virtue of information gain metric is much similar to CHARM and CBA which also work on the same track. To generate association and classification rules by pruning method of Apriori software, some external tools are essentially required. The effectiveness of the pruning algorithm can be inversely related to the number of rules. This along with the computational time consumed, determine an efficient criterion of pruning.

Efficient post processing methods are hence proposed to remove pointless rules from rule-ways by eliminating redundancy among rules. The dependent relation, exploitation, makes this method a self manageable knowledge. The pruning procedure has been sliced into three stages starting from derivation to pruning operation on rule-set by the use of close rule-sets. It is cost-effective and consumes very little time for the transaction. Hence forth, it can be applied to exploit sampling techniques and data structures, thereby increasing the efficiency.

Huawen Liu, et al[14] presented a technique on post-processing for rule reduction using closed set that was targeting to filter the otiose rules in a post-processing of rule mining. The empirical study proved that the discovery of dependent relations from closed set helps to eliminate redundant rules. Hetal Thakkar et al [12]. In the case of stream data, the post-mining of association is more challenging and continuous post mining of association rules is an unavoidable requirement, which is discussed by this author. He presented a technique for continuous post-mining of association rules in a data stream management system. He described the architecture and techniques used to achieve this advanced functionality in the Stream Mill Miner (SMM) prototype, an SQL-based DSMS designed to support continuous mining queries, which is impressive. Hacene Cherfi et al, [10] discussed a post association rule mining approach for text mining that combines data mining, semantic techniques for post-mining and selection of association rules. To focus on the result analysis and to find new knowledge units, classification of association rules according to qualitative criteria using domain model as background knowledge has been introduced. The authors carried out an empirical study on molecular biology dataset that proved the benefits of taking into account a knowledge domain model of the data. Ronaldo Cristiano Prati [3]. The Receiver Operating Characteristics (ROC) graph is a popular way of assessing the performance of classification rules, but they are inappropriate to evaluate the quality of association rules, as there is no class in association rule mining and the consequent part of different association rules might not have any correlation at all. Chapter VIII presents a novel technique of QROC, a variation of ROC space to analyze itemset costs/benefits in association rules. It can be used to help analysts to evaluate the relative interestingness among different association rules in different cost scenarios.

III. Attribute Relational Analysis (ARA) Framework for Coherent association Rules

The approach Attribute Relational Analysis in short can refer as ARA is post mining process to prune the rules based on attribute relational relevancy. The process of ARA Framework can be classified as

- Closed itemset mining
- Describing item class descriptor

The input for ARA Framework is

1. A set of rules
2. An XML descriptor describes attributes, classes, class properties and class relations.

Here in this proposal we considered our earlier work [] to find closed itemsets.

Global Journal of Researches in Engineering (I) Volume XI Issue VII Version I December 2011

The process steps involved in ARA framework are

1. Initially ARA measures the property support degree for each attribute involved in given rule.
2. By using the property support degree of the attributes, Attribute Relation support of attribute pairs of the given rule will be measured.
3. With the help of Attribute Relation supports of all attribute pairs of an itemset that belongs to a given rule, Attribute relation support degree of that itemset will be measured.
4. Using these Attribute relation support degrees of Left Hand Side and Right Hand Side itemsets of the given rule, relation confidence of the rule will be determined.
5. Prunes the rules based on their attribute relation support degree.

Detailed explanation of each step can be found in Section V.

IV. Closed Itemset Mining Using PEPP [34] and Inference Relations [35]

a) Dataset adoption and formulation

Item Sets I : A set of diverse elements by which the sequences generate.

$$I=\bigcup_{k=1}^{n} i_k$$

Note: 'I' is set of diverse elements

Sequence set 'S': A set of sequences, where each sequence contains elements each element 'e' belongs to 'I' and true for a function p(e). Sequence set can formulate as

$$s=\bigcup_{i=1}^{m} <e_i|(p(e_i),e_i\in I)>$$

Represents a sequence's' of items those belongs to set of distinct items 'I'.

'm': total ordered items.

$P(e_i)$: a transaction, where e_i usage is true for that transaction.

$$S=\bigcup_{j=1}^{t} s_j$$

S: represents set of sequences

't': represents total number of sequences and its value is volatile

s_j: is a sequence that belongs to S

Subsequence : a sequence s_p of sequence set 'S' is considered as subsequence of another sequence s_q of Sequence Set 'S' if all items in sequence S_p is belongs to s_q as an ordered list. This can be formulated as

If $$(\bigcup_{i=1}^{n} s_{pi}\in s_q)\Rightarrow(s_p\subseteq s_q)$$

Then $$\bigcup_{i=1}^{n} s_{pi} <: \bigcup_{j=1}^{m} s_{qj}$$ where $s_p\in S \text{ and } s_q\in S$

Total Support 'ts' : occurrence count of a sequence as an ordered list in all sequences in sequence set 'S' can adopt as total support 'ts' of that sequence. Total support 'ts' of a sequence can determine by fallowing formulation.

$$f_{ts}(s_t)=|s_t <: s_p \ (\text{for each } p=1..|DB_S|)|$$

DB_S Is set of sequences

$f_{ts}(s_t)$: Represents the total support 'ts' of sequence s_t is the number of super sequences of s_t

Qualified support 'q_s': The resultant coefficient of total support divides by size of sequence database adopt as qualified support 'qs'. Qualified support can be found by using fallowing formulation.

$$f_{qs}(s_t)=\frac{f_{ts}(s_t)}{|DB_S|}$$

Sub-sequence and Super-sequence: A sequence is sub sequence for its next projected sequence if both sequences having same total support.

Super-sequence: A sequence is a super sequence for a sequence from which that projected, if both having same total support.

Sub-sequence and super-sequence can be formulated as

If $f_{ts}(s_t) \geq rs$ where 'rs' is required support threshold given by user

And $s_t <: s_p$ *for any p value* where $f_{ts}(s_t)\cong f_{ts}(s_p)$

b) Closed Itemset discovery using PEPP: Parallel Edge Projection and Pruning Based Sequence Graph protrude[34]

As a first stage of the proposal we perform dataset pre-processing and itemsets Database initialization. We find itemsets with single element, in parallel prunes itemsets with single element those contains total support less than required support.

Forward Edge Projection

In this phase, we select all itemsets from given itemset database as input in parallel. Then we start projecting edges from each selected itemset to all possible elements. The first iteration includes the pruning process in parallel, from second iteration onwards this pruning is not required, which we claimed as an efficient process compared to other similar techniques like BIDE. In first iteration, we project an itemset s_p that spawned from selected itemset s_i from

DB_S and an element e_i considered from 'I'. If the $f_{ts}(s_p)$ is greater or equal to rs, then an edge will be defined between s_i and e_i. If $f_{ts}(s_i) \cong f_{ts}(s_p)$ then we prune s_i from DB_S. This pruning process required and limited to first iteration only.

From second iteration onwards project the itemset s_p that spawned from $S_{p'}$ to each element e_i of 'I'. An edge can be defined between $S_{p'}$ and e_i if $f_{ts}(s_p)$ is greater or equal to rs. In this description $S_{p'}$ is a projected itemset in previous iteration and eligible as a sequence. Then apply the fallowing validation to find closed sequence.

Edge pruning

If any of $f_{ts}(s_{p'}) \cong f_{ts}(s_p)$ that edge will be pruned and all disjoint graphs except s_p will be considered as closed sequence and moves it into DB_S and remove all disjoint graphs from memory.

The above process continues till the elements available in memory those are connected through direct or transitive edges and projecting itemsets i.e., till graph become empty

c) Inference Analysis [35]

Inferences:

Pattern positive score is sum of no of transactions in which all items in the pattern exist, no of transactions in which all items in the pattern does not exist

Pattern negative score is no of transactions in which only few items of the pattern exist

Pattern actual coverage is pattern positive score-pattern negative score

Interest gain: Actual coverage of the pattern involved in association rule

Coherent rule Actual coverage of the rule's left side pattern must be greater than or equal to actual coverage of the right side pattern

Inference Support ia_s: refers actual coverage of the pattern

$f_{ia}(s_t)$: Represents the inference support of the sequence s_t

Approach:

For each pattern s_p of the pattern dataset, If $f_{ia}(s_t) < ia_s$ then we prune that pattern

Detailed explanation of the PEPP Algorithm can find at []:

d) Description of Inference Analysis

Set I = {i1, i2… im} be the universe of items composed of m different attributes, ik(k=1,2,...,m) is item. Transaction database D is a collection of transaction T, A transaction t = (tid, X) is a tuple where tid is a unique transaction ID and X is an itemset. The count of an itemset X in D, denoted by count(X), is the number of transactions in D containing X. The support of an itemset X in D, denoted by supp(X), is the proportion of transactions in D that contain X. The negative rule X⇒¬Y holds in the transaction set D with confidence conf (X⇒¬Y) = supp $(X \cup \neg Y)$ /supp (X).

In Transaction database, each transaction is a collection of items involved sequences. The issue of mining association rules is to get all association rules that its support and confidence is respectively greater than the minimum threshold given by the user.

The issues of mining association rules can be divide into two sub-issues as follows:

· Find frequent itemsets, Generate all itemsets that support is greater than the minimum support;

· Generate association rules from frequent itemsets.

In logical analysis, the direct calculation of support logical analysis is not convenient, To calculate the support and confidence of negative associations using the support and confidence of positive association that is known: set A, B⊂I, A∩B=Φ, then:

$$\sup(\neg A) = 1 - \sup(A);$$
$$\sup(A \cup \neg B) = \sup(A) - Sup(A \cup B);$$
$$\sup(\neg A \cup B) = \sup(B) - \sup(A \cup B)$$
$$\sup(\neg A \cup \neg B) = 1 - \sup(A) - \sup(B) + \sup(A \cup B);$$

Based on the above formulas we perform the logical analysis to derive the actual support of the patterns that improves the rule coherency.

Inference analysis by example:

Let A, B∈I where I is itemset generated with the association of A,B are individual items or subsets.

Under logical analysis we determine $f_{ts}(\neg A \cup \neg B)$, $f_{ts}(A \cup \neg B)$ and $f_{ts}(\neg A \cup B)$.

The support $f_{ts}(I)$, $f_{ts}(\neg A \cup \neg B)$ we consider as positive support and $f_{ts}(A \cup \neg B)$, $f_{ts}(\neg A \cup B)$ we consider as negative support.

Finally we determine $f_{ia}(I) = f_{ts}(I) + f_{ts}(\neg A \cup \neg B)$ - $f_{ts}(A \cup \neg B)$ - $f_{ts}(\neg A \cup B)$;

Global Journal of Researches in Engineering (I) Volume XI Issue VII Version I December 2011

V. Attribute Relation Analysis Framework

The proposed post mining process Attribute Relation Analysis described in detailed here. Table 1 represents the notations used in ARA framework.

Class Descriptor: The domain expert classifies the attributes involved in transactions will be classified in to different categories. The process of classification as follow

1. Initially classes will be derived based on the properties; hence each class contains set of properties. These classes can be recursive i.e., a class may refer one or more other classes as sub classes.
2. Based on attribute properties, attributes will be categorized into a class.
   a) Ex: if most of the attribute 'a' properties matched to class 'c' then $a \in c$
3. The domain expert also initiates to derive the relation between classes. The relation can be between any two classes, such as
   b) Relation between class and sub-class of other class
   c) Relation between two direct classes
   d) Relation between two sub classes

Note : All related classes of a sub class also related to it's parent class

An xml based attribute class descriptor will be prepared. Fig1 shows an example descriptor. Notations description equations can found in table 1 that fallows.

```
<class-descriptor>
    <properties>
        <property id=1>
            <name>prop1</name>
        </property>
        <property id=2>
            <name>prop2</name>
        </property>
        <property id=3>
            <name>prop3</name>
        </property>
    </properties>
    <attributes>
        <attribute name="item-1" id="1" properties="{list of property ids}" />
        <attribute name="item-2" id="2" properties="{list of property ids}" />
        <attribute name="item-3" id="3" properties="{list of property ids}" />
        .
        .
        <attribute name="item-n" id="n" properties="{list of property ids}" />
    </attributes>
    <classes>
        <class name="class-1" id="1" properties="{list of property ids}" />
        <class name="class-2" id="2" properties="{list of property ids}" />
        <class name="class-3" id="3" properties="{list of property ids}" />
        .
        .
        <class name="class-n" id="n" properties="{list of property ids}" />
    </classes>
    <child-classes>
        <!--parent value must be unique -->
        <child-class parent="class-id" child="{list of class ids}"/>
        .
        .
        <child-class parent="class-id" child="{list of class ids}"/>
    </child-classes>
    <relations>
        <!--lhs value must be unique -->
        <!--
        classes that related to a child-class also related to it's parent class
        -->
        <relation lhs="class-id" rhs="{list of class ids}" />
        .
        .
        <relation lhs="class-id" rhs="{list of class ids}" />
    </relations>
</class-descriptor>
```

Fig 1 : Class-descriptor

Global Journal of Researches in Engineering (I) Volume XI Issue VII Version I December 2011

Table1 : Notations used in Attribute Relational Analysis

1	r_{lhs}	Left side itemset of the Rule r
2	r_{rhs}	Right side itemset of the rule r
3	RS	Rule set
4	cpc_c	Class properties count of class c
5	apc_a	Attribute property count of attribute a
6	psd	Property support degree
7	tp_c	Total properties of class c
8	ps_a $ps_a = \frac{apc_a}{cpc_c}$	Property support of attribute a of class c
9	$psd_a = \frac{ps_a}{tp_c}$	where $a \in c$
10	rs_c	Relation support of class c is max threshold value 1.
11	ARS	Attribute Relation support
12	$ARSD_i$	Attribute Relation support Degree of itemset i.
13	If attributes a_i and a_j belongs to same class c	$ARS(a_i, a_j) \cong 1$, *where* $\{a_i, a_j\} \in c$
14	If attribute a_i belongs to class c_i and attribute a_j belongs to class c_j, c_i and c_j relation is true then	$ARS(a_i, a_j) = \frac{psd_{a_i} + psd_{a_j}}{rs_{c_i} + rs_{c_j}}$
15	If c_i and c_j relation is false	$ARS(a_i, a_j) = 0$,
16	$ARS(a_i, a_j) \cong ARS(a_j, a_i)$	Applicable in all cases such as both belongs to same class, both belongs to different classes that are not having relation and both belongs to different classes that are having relation
17	No of attribute pairs in an itemset	pc
18	$pc = 0;$ $\sum_{i=1}^{n-1} pc + i \quad (or) \quad pc = \frac{n-1}{2} \times n$	pc : is total number of pairs in a given itemset. n : is total number of attributes in the given itemset
19	$pc_{r_{lhs}}$	Pair-count of itemset, which is lhs of rule r .
20	$pc_{r_{rhs}}$	Pair-count of itemset, which is rhs of rule r .
21	$pc_{r_{lhs} \cup r_{rhs}}$	Pair-count of itemset that generated from $r_{lhs} \cup r_{rhs}$
22	$PS_i = \{p_1, p_2, ..., p_m\}$	Pair set that generated from itemset i

23	$ARS(p_i)$	Attribute relation support of pair p_i.
24	$ARSD_i = \frac{\sum_{k=1}^{\lvert ps_i \rvert} ARS(p_k)}{\lvert ps_i \rvert}$	Attribute relation support degree of itemset i And $p_k \in ps_i$, here p_k is k[th] pair of pair-set ps of itemset i
25	RC_r	Relation confidence of rule r.
26	$ARSD_{lhs} = \frac{\sum_{k=1}^{\lvert PS_{lhs} \rvert} p_k}{\lvert PS_{lhs} \rvert}$	Attribute relation support degree of itemset, which is lhs of rule r And $p_k \in ps_{r_{lhs}}$, here p_k is k[th] pair of pair-set ps of itemset lhs of rule r
27	$ARSD_{rhs} = \frac{\sum_{k=1}^{\lvert PS_{rhs} \rvert} p_k}{\lvert PS_{rhs} \rvert}$	Attribute relation support degree of itemset, which is rhs of rule r And $p_k \in ps_{r_{rhs}}$, here p_k is k[th] pair of pair-set ps of itemset rhs of rule r
28	$ARSD_{lhs \cup rhs} = \frac{\sum_{k=1}^{\lvert PS_{lhs \cup rhs} \rvert} p_k}{\lvert PS_{lhs \cup rhs} \rvert}$	Attribute relation support degree of itemset that generated from $lhs \cup rhs$ of rule r And $p_k \in PS_{lhs \cup rhs}$, here p_k is k[th] pair of pair-set ps of itemset that generated from $lhs \cup rhs$ of rule r
29	$rc_r = \frac{ARSD_{lhs \cup rhs}}{ARSD_{lhs}}$	Relation confidence of r is the coefficient emerged as result when 30attribute support degree of all attribute involved in rule r is divided by attribute support degree of rule r's lhs

Property Support : No of attribute properties are matched to number class properties to which that attribute belongs to [Table 1 row: 8].

Property Support degree : indicates the ratio of properties matched to class level properties [table 1 row: 9].

Ex: $psd_a = \frac{ps_a}{cpc_c}$; here a is an attribute of class c $[a \in c]$

Attribute Relation support : Indicates the strength of the relation between two attributes of an itemset that are considered as pair for equation [see table 1 row: 11, 13].

Pair Count : Total number of two attributes sets; here these attribute sets must be unique [see table 1, row 17, 18]

Attribute Relation support degree : is an itemset level measurement representing average relation strength of the attributes those belongs to an itemset [see table 1 row: 24]

Relation confidence : is a rule level measurement concludes the relation strength between left hand side itemset and right hand side itemset of a given rule[see table 1 row: 29]

ARA algorithm :

Input: Rule set RS , Class Descriptor CD and relation confidence threshold rct

Output: Significant Rule set RS' which is subset of RS

$RS' \subseteq RS$

Begin:

While RS is not empty

Begin:

Read a rule r from RS

Find property support ps_a and property support degree psd_a of each attribute a of r_{lhs} [Table 1 row: 8, 9]

Find property support ps_a and property support degree psd_a of each attribute a of r_{rhs} [Table 1 row: 8, 9]

Find unique two attribute pair set $PS_{r_{lhs}}$ from r_{lhs} [Table 1 row: 22]

Find Attribute relation support ARS_p for each pair p, where $p \in PS_{r_{lhs}}$ [Table 1; row: 13, 15, 15 and 16]

Find Attribute relation support degree $ARSD_{r_{lhs}}$ of r_{lhs} [Table 1; row: 24, 26].

Find unique two attribute pair set $PS_{r_{lhs}}$ from r_{lhs} [Table 1 row: 22]

Find Attribute relation support ARS_p for each pair p, where $p \in PS_{r_{rhs}}$ [Table 1; row: 13, 15, 15 and 16]

Find Attribute relation support degree $ARSD_{r_{rhs}}$ of r_{lhs} [Table 1; row: 24, 27].

Find unique two attribute pair set $PS_{r_{lhs} \cup r_{rhs}}$ from $r_{lhs} \cup r_{rhs}$ [Table 1 row: 22]

Find Attribute relation support ARS_p for each pair p, where $p \in PS_{r_{lhs} \cup r_{rhs}}$ [Table 1; row: 13, 15, 15 and 16]

Find Attribute relation support degree $ARSD_{r_{lhs} \cup r_{rhs}}$ of $r_{lhs} \cup r_{rhs}$ [Table 1; row: 24, 28].

Find unique two attribute pair set $PS_{r_{lhs} \cup r_{rhs}}$ from $r_{lhs} \cup r_{rhs}$ [Table 1 row: 22]

Find Attribute relation support ARS_p for each pair p, where $p \in PS_{r_{rhs} \cup r_{lhs}}$ [Table 1; row: 13, 15, 15 and 16]

Find Attribute relation support degree $ARSD_{r_{rhs} \cup r_{lhs}}$ of $r_{lhs} \cup r_{rhs}$ [Table 1; row: 24, 28].

Find Relation confidence rc_r of rule r

If $rc_r \geq rct$ then add rule r to resultant rule set RS'

End

End

Fig 2 : Attribute Relation Analysis algorithm

This segment focuses mainly on providing evidence on asserting the claimed assumptions that 1) The post mining framework ARA is competent enough to momentously surpass results when evaluated against other post mining techniques [14, 10, 12. 2) Utilization of memory and computational complexity is less when compared to other post mining techniques. 3) There is the involvement of an enhanced occurrence and a probability reduction in the memory exploitation rate with the aid of the trait equivalent prognosis and also rim snipping of the PEPP with inference analysis and ARA. This is on the basis of the surveillance done which concludes that ARA implementation is far more noteworthy and important in contrast with the likes of other notable models [10, 12, 14].

JAVA 1.6_ 20th build was employed for accomplishment of the ARA along with PEPP under inference analysis. A workstation equipped with core2duo processor, 2GB RAM and Windows XP installation was made use of for investigation of the algorithms. The parallel replica was deployed to attain the thread concept in JAVA.

Dataset Characteristics [34, 35]:

We used same experiments platform described in our earlier work [34, 35].Hence the dataset that we opted is Pi and it's characteristics as described in our earlier work [34, 35]:

Pi is supposedly found to be a very opaque dataset, which assists in excavating enormous quantity of recurring clogged series with a profitably high threshold somewhere close to 90%. It also has a distinct element of being enclosed with 190 protein series and 21 divergent objects. Reviewing of serviceable legacy's consistency has been made use of by this dataset. Fig. 3 portrays an image depicting dataset series extent status.

In assessment with all the other regularly quoted forms like [14,12,10], Post-Processing for Rule Reduction Using Closed Set[14] has made its mark as a most preferable, superior and sealed example of post mining copy, taking in view the detailed study of the factors mainly, experts involvement, memory consumption and runtime.

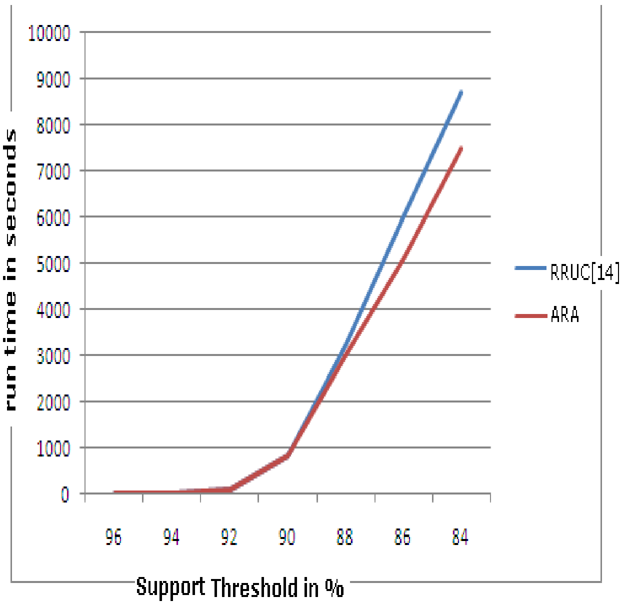


Fig 3 : A comparison report for Runtime[34, 35]

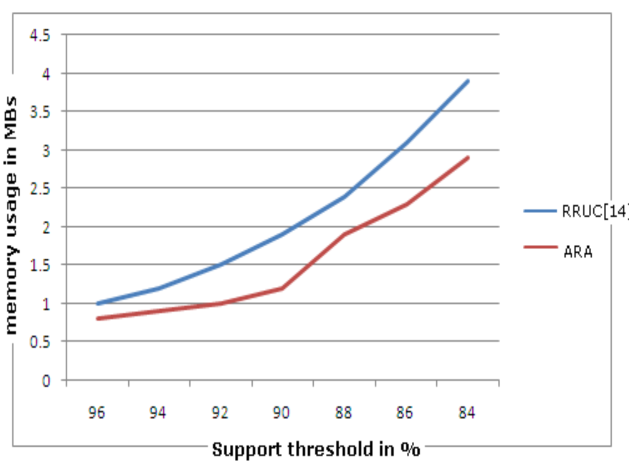


Fig: 4 : A comparison report for memory usage [34, 35]

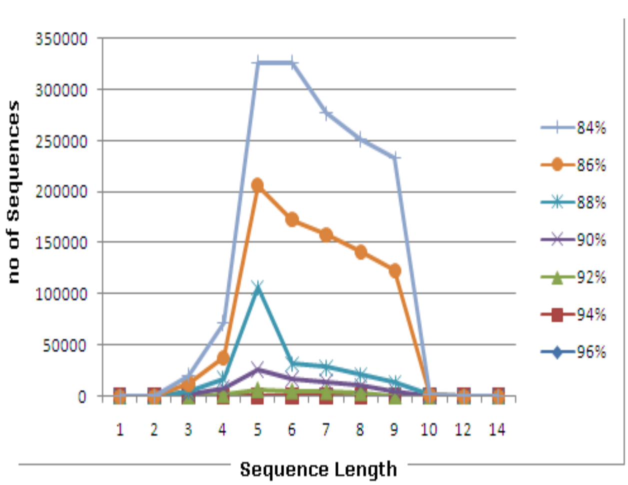


Fig 5 : Sequence length and number of sequences at different thresholds in Pi dataset[34, 35]

In contrast to ARA and RRUC [14], a very intense dataset Pi is used which has petite recurrent closed series whose end to end distance is less than 10, even in the instance of high support amounting to around 90%. The diagrammatic representation displayed in Fig 3 explains that the above mentioned two algorithms execute in a similar fashion in case of rules that are generated at support 90% and above. But in situations when the support case is 88% and less, then the act of ARA surpasses RRUC [14]'s routine. The disparity in memory exploitation of ARA and RRUC [14] can be clearly observed because of the consumption level of ARA being low than that of RRUC. Fig 5 indicates that rules that are contextually irrelevant have been pruned by ARA in high probable rate and stable. Apart from the benefits observed, the rules identified by ARA are more relevant to transaction consequences. It becomes possible since there is no task of experts evaluation in post mining process. Due to the concept of attribute class relation descriptor, the relation between attributes involved in rule is stable.

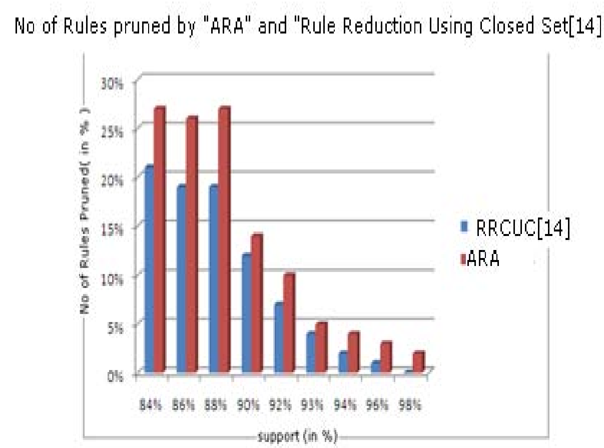


Fig 6 : Rules pruned in % by ARA and RRUC [14]

VI. Conclusion

We proposed a post mining process called Attribute relation analysis framework (ARA) for pruning association rules that are contextually irrelevant. In earlier works [10, 12, 14] the contextual irrelevancy was identified in various ways such as (1) rule evaluation by domain expert, (2) rule evaluation by itemset closeness. We argued that none of these two models is significant in all contexts. In second case adaptability to various data contexts is missing. In the first case, rule selection highly influenced by the experts view, that is when expert changes then rule significance might be rated differently. To defend these limits here we proposed a post mining process as an extension to our earlier proposed closed itemset mining algorithm PEPP with inference analysis [34, 35]. Here in this proposed post mining process ARA, the experts view is not defending one to other, rather it extends or refines. This become possible in ARA because of the proposed concept called attribute class relation descriptor. In this work we consider relation confidence as bench mark for rule pruning; in future this work can be extended to prune the rules by opting relation confidence threshold under inference analysis.

References References Referencias

1. Fienberg, S. E. & Shmueli, G: "Statistical issues and challenges associated with rapid detection of bio-terrorist attacks. Statistics in Medicine, 2005
2. Carmona-Saez, P., Chagoyen, M., Rodriguez, A., Trelles, O., Carazo, J. M., & Pascual-Montano, A: :Integrated analysis of gene expression by association rules discovery". BMC Bioinformatics, 2006
3. Ronaldo Cristiano Prati, "QROC: A Variation of ROC Space to Analyze Item Set Costs/Benefits in Association Rules", Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction, IGIGlobal, 2009 Pages: 133-148 pp.
4. McNicholas, P. D: "Association rule analysis of CAO data (with discussion). Journal of the Statistical and Social Inquiry Society of Ireland, 2007
5. Prati, R., & Flach, P. "ROCCER: an algorithm for rule learning based on ROC analysis". Proceeding of the 19th International Joint Conference on Artificial Intelligence, 2005
6. Fawcett, T: "PRIE: A system to generate rulelists to maximize ROC performance". Data Mining and Knowledge Discovery Journal, 2008, 17(2), 207-224.
7. Kawano, H., & Kawahara, M: "Extended Association Algorithm Based on ROC Analysis for Visual Information Navigator". Lecture Notes In Computer Science, 2002, 2281, 640-649.
8. Piatetsky-Shapiro, G., Piatetsky-Shapiro, G., Frawley, W. J., Brin, S., Motwani, R., Ullman, J. D., et al: "Discovery, Analysis, and Presentation of Strong Rules". Proceedings of the 11th international symposium on Applied Stochastic Models and Data Analysis ASMDA, 2005, 16, 191-200
9. Liu, B., Hsu, W., & Ma, Y: "Pruning and summarizing the discovered associations". Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining 1999, (pp. 125-134)
10. Hacene Cherfi, Amedeo Napoli, Yannick Toussaint: "A Conformity Measure Using Background Knowledge for Association Rules: Application to Text Mining", Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction, IGIGlobal, 2009 Pages: 100-115 pp.
11. Liu, B., Hsu, W., & Ma, Y.: "Pruning and summarizing the discovered associations". Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, 1999
12. Hetal Thakkar, Barzan Mozafari, Carlo Zaniolo: "Continuous Post-Mining of Association Rules in a Data Stream Management System", Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction, IGIGlobal, 2009 Pages: 116-132 pp.
13. Kuntz, P., Guillet, F., Lehn, R., & Briand, H: "A User-Driven Process for Mining Association Rules". In D. Zighed, H. Komorowski, & J. Zytkow (Eds.), 4th Eur. Conf. on Principles of Data Mining and Knowledge Discovery 2000
14. Huawen Liu, Jigui Sun, Huijie Zhang: "Post-Processing for Rule Reduction Using Closed Set", Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction, IGIGlobal,2009 Pages: 81-99 pp.
15. Jaroszewicz, S., & Simovici, D. A: "Interestingness of Frequent Itemsets using Bayesian networks as Background Knowledge". ACM SIGKDD Conference on Knowledge Discovery in Databases, 2004
16. Jaroszewicz, S., & Scheffer, T: "Fast Discovery of Unexpected Patterns in Data, Relative to a Bayesian Network". ACM SIGKDD Conference on Knowledge Discovery in Databases, 2005
17. Faure, C., Delprat, D., Boulicaut, J. F., & Mille, A: "Iterative Bayesian Network Implementation by using Annotated Association Rules." 15th Int'l Conf. on Knowledge Engineering and Knowledge Management – Managing Knowledge in a World of Networks, 2006
18. [18]Liu, B., & Hsu, W: "Post-Analysis of Learned Rules". AAAI/IAAI, 1996
19. Lent, B., Swami, A. N., & Widom, J: "Clustering Association Rules". Proceedings of the Thirteenth International Conference on Data Engineering, 1997, Birmingham U.K., IEEE Computer Society
20. Savasere, A., Omiecinski, E., & Navathe, S. B: "Mining for strong negative associations in a large database of customer transactions". Proceedings of the 14th International Conference on Data Engineering, Washington DC, USA, 1998
21. Basu, S., Mooney, R. J., Pasupuleti, K. V., & Ghosh J: "Evaluating the Novelty of Text-Mined Rules using Lexical Knowledge". 7th ACM SIGKDD International Conference on Knowledge Discovery in Databases, 2001
22. Claudia Marinica and Fabrice Guillet: "Knowledge-Based Interactive Postmining of Association Rules Using Ontologies", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 22, NO. 6, JUNE 2010

23. M.J. Zaki and C.J. Hsiao, "Charm: An Efficient Algorithm for Closed Itemset Mining," Proc. Second SIAM Int'l Conf. Data Mining, pp. 34-43, 2002.
24. J. Pei, J. Han, and R. Mao, "Closet: An Efficient Algorithm for Mining Frequent Closed Itemsets," Proc. ACM SIGMOD Workshop Research Issues in Data Mining and Knowledge Discovery, pp. 21-30, 2000.
25. M.J. Zaki and M. Ogihara, "Theoretical Foundations of Association Rules," Proc. Workshop Research Issues in Data Mining and Knowledge Discovery (DMKD '98), pp. 1-8, June 1998.
26. D. Burdick, M. Calimlim, J. Flannick, J. Gehrke, and T. Yiu, "Mafia: A Maximal Frequent Itemset Algorithm," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 11, pp. 1490-1504, Nov. 2005.
27. J. Li, "On Optimal Rule Discovery," IEEE Trans. Knowledge and Data Eng., vol. 18, no. 4, pp. 460-471, Apr. 2006.
28. M. Hahsler, C. Buchta, and K. Hornik, "Selective Association Rule Generation," Computational Statistic, vol. 23, no. 2, pp. 303-315, Kluwer Academic Publishers, 2008.
29. H. Toivonen, M. Klemettinen, P. Ronkainen, K. Hatonen, and H. Mannila, "Pruning and Grouping of Discovered Association Rules," Proc. ECML-95 Workshop Statistics, Machine Learning, and Knowledge Discovery in Databases, pp. 47-52, 1995.
30. J. Bayardo, J. Roberto, and R. Agrawal, "Mining the Most Interesting Rules," Proc. ACM SIGKDD, pp. 145-154, 1999.
31. M. Klemettinen, H. Mannila, P. Ronkainen, H. Toivonen, and A.I. Verkamo, "Finding Interesting Rules from Large Sets of Discovered Association Rules," Proc. Int'l Conf. Information and Knowledge Management (CIKM), pp. 401-407, 1994.
32. R. Srikant and R. Agrawal, "Mining Generalized Association Rules," Proc. 21st Int'l Conf. Very Large Databases, pp. 407-419,

    http://citeseer.ist.psu.edu/srikant95mining.html, 1995.
33. Barzan Mozafari, Hetal Thakkar, Carlo Zaniolo. Verifying and Mining Frequent Patterns from Large Windows over Data Streams. In Proceedings of the 24th International Conference on Data Engineering (ICDE 2008), Cancún, México, April 7-12, 2008
34. Parallel Edge Projection and Pruning (PEPP) Based Sequence Graph protrude approach for Closed Itemset Mining kalli Srinivasa Nageswara Prasad, Sri Venkateswara University, Tirupati, Andhra Pradesh , India. Prof. S. Ramakrishna, Department of Computer Science, Sri Venkateswara University, Tirupati, Andhra Pradesh , India ol. 9 No. 9 September 2011 International Journal of Computer Science and Information Security Publication September 2011, Volume 9 No. 9
35. Mining Closed Itemsets for Coherent Rules: An Inference Analysis Approach By Kalli Srinivasa Nageswara Prasad, Prof. S. Ramakrishna, Global Journal of Computer Science and Technology Volume 11 Issue 19 Version 1.0 November 2011 Type: Double Blind Peer Reviewed International Research Journal Publisher: Global Journals Inc. (USA) Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Global Journals Inc. (US) Guidelines Handbook 2011

www.GlobalJournals.org

FELLOW OF ASSOCIATION OF RESEARCH SOCIETY IN ENGINEERING (FARSE)

- 'FARSE' title will be awarded to the person after approval of Editor-in-Chief and Editorial Board. The title 'FARSE" can be added to name in the following manner. eg. Dr. John E. Hall, Ph.D., FARSE or William Walldroff Ph. D., M.S., FARSE
- Being FARSE is a respectful honor. It authenticates your research activities. After becoming FARSE, you can use 'FARSE' title as you use your degree in suffix of your name. This will definitely will enhance and add up your name. You can use it on your Career Counseling Materials/CV/Resume/Visiting Card/Name Plate etc.
- 60% Discount will be provided to FARSE members for publishing research papers in Global Journals Inc., if our Editorial Board and Peer Reviewers accept the paper. For the life time, if you are author/co-author of any paper bill sent to you will automatically be discounted one by 60%
- FARSE will be given a renowned, secure, free professional email address with 100 GB of space eg.johnhall@globaljournals.org. You will be facilitated with Webmail, Spam Assassin, Email Forwarders, Auto-Responders, Email Delivery Route tracing, etc.
- FARSE member is eligible to become paid peer reviewer at Global Journals Inc. to earn up to 15% of realized author charges taken from author of respective paper. After reviewing 5 or more papers you can request to transfer the amount to your bank account or to your PayPal account.
- Eg. If we had taken 420 USD from author, we can send 63 USD to your account.
- FARSE member can apply for free approval, grading and certification of some of their Educational and Institutional Degrees from Global Journals Inc. (US) and Open Association of Research, Society U.S.A.
- After you are FARSE. You can send us scanned copy of all of your documents. We will verify, grade and certify them within a month. It will be based on your academic records, quality of research papers published by you, and 50 more criteria. This is beneficial for your job interviews as recruiting organization need not just rely on you for authenticity and your unknown qualities, you would have authentic ranks of all of your documents. Our scale is unique worldwide.
- FARSE member can proceed to get benefits of free research podcasting in Global Research Radio with their research documents, slides and online movies.
- After your publication anywhere in the world, you can upload you research paper with your recorded voice or you can use our professional RJs to record your paper their voice. We can also stream your conference videos and display your slides online.
- FARSE will be eligible for free application of Standardization of their Researches by Open Scientific Standards. Standardization is next step and level after publishing in a journal. A team of research and professional will work with you to take your research to its next level, which is worldwide open standardization.

- FARSE is eligible to earn from their researches: While publishing his paper with Global Journals Inc. (US), FARSE can decide whether he/she would like to publish his/her research in closed manner. When readers will buy that individual research paper for reading, 80% of its earning by Global Journals Inc. (US) will be transferred to FARSE member's bank account after certain threshold balance. There is no time limit for collection. FARSE member can decide its price and we can help in decision.

MEMBER OF ASSOCIATION OF RESEARCH SOCIETY IN ENGINEERING (MARSE)

- 'MARSE' title will be awarded to the person after approval of Editor-in-Chief and Editorial Board. The title 'MARSE" can be added to name in the following manner. eg. Dr. John E. Hall, Ph.D., MARSE or William Walldroff Ph. D., M.S., MARSE
- Being MARSE is a respectful honor. It authenticates your research activities. After becoming MARSE, you can use 'MARSE' title as you use your degree in suffix of your name. This will definitely will enhance and add up your name. You can use it on your Career Counseling Materials/CV/Resume/Visiting Card/Name Plate etc.
- 40% Discount will be provided to MARSE members for publishing research papers in Global Journals Inc., if our Editorial Board and Peer Reviewers accept the paper. For the life time, if you are author/co-author of any paper bill sent to you will automatically be discounted one by 60%
- MARSE will be given a renowned, secure, free professional email address with 30 GB of space eg.johnhall@globaljournals.org. You will be facilitated with Webmail, SpamAssassin, Email Forwarders, Auto-Responders, Email Delivery Route tracing, etc.
- MARSE member is eligible to become paid peer reviewer at Global Journals Inc. to earn up to 10% of realized author charges taken from author of respective paper. After reviewing 5 or more papers you can request to transfer the amount to your bank account or to your PayPal account.
- MARSE member can apply for free approval, grading and certification of some of their Educational and Institutional Degrees from Global Journals Inc. (US) and Open Association of Research,Society U.S.A.
- MARSE is eligible to earn from their researches: While publishing his paper with Global Journals Inc. (US), MARSE can decide whether he/she would like to publish his/her research in closed manner. When readers will buy that individual research paper for reading, 40% of its earning by Global Journals Inc. (US) will be transferred to MARSE member's bank account after certain threshold balance. There is no time limit for collection. MARSE member can decide its price and we can help in decision.

ANNUAL MEMBER

- Annual Member will be authorized to receive e-Journal GJHSS for one year (subscription for one year).
- The member will be allotted free 1 GB Web-space along with subDomain to contribute and participate in our activities.
- A professional email address will be allotted free 500 MB email space.

PAPER PUBLICATION

- The members can publish paper once. The paper will be sent to two-peer reviewer. The paper will be published after the acceptance of peer reviewers and Editorial Board.

Process of submission of Research Paper

The Area or field of specialization may or may not be of any category as mentioned in 'Scope of Journal' menu of the GlobalJournals.org website. There are 37 Research Journal categorized with Six parental Journals GJCST, GJMR, GJRE, GJMBR, GJSFR, GJHSS. For Authors should prefer the mentioned categories. There are three widely used systems UDC, DDC and LCC. The details are available as 'Knowledge Abstract' at Home page. The major advantage of this coding is that, the research work will be exposed to and shared with all over the world as we are being abstracted and indexed worldwide.

The paper should be in proper format. The format can be downloaded from first page of 'Author Guideline' Menu. The Author is expected to follow the general rules as mentioned in this menu. The paper should be written in MS-Word Format (*.DOC,*.DOCX).

The Author can submit the paper either online or offline. The authors should prefer online submission.Online Submission: There are three ways to submit your paper:

(A) (I) First, register yourself using top right corner of Home page then Login. If you are already registered, then login using your username and password.

(II) Choose corresponding Journal.

(III) Click 'Submit Manuscript'. Fill required information and Upload the paper.

(B) If you are using Internet Explorer, then Direct Submission through Homepage is also available.

(C) If these two are not convenient, and then email the paper directly to dean@globaljournals.org.

Offline Submission: Author can send the typed form of paper by Post. However, online submission should be preferred.

MANUSCRIPT STYLE INSTRUCTION **<u>(Must be strictly followed)</u>**

Page Size: 8.27" X 11'"

- Left Margin: 0.65
- Right Margin: 0.65
- Top Margin: 0.75
- Bottom Margin: 0.75
- Font type of all text should be Swis 721 Lt BT.
- Paper Title should be of Font Size 24 with one Column section.
- Author Name in Font Size of 11 with one column as of Title.
- Abstract Font size of 9 Bold, “Abstract” word in Italic Bold.
- Main Text: Font size 10 with justified two columns section
- Two Column with Equal Column with of 3.38 and Gaping of .2
- First Character must be three lines Drop capped.
- Paragraph before Spacing of 1 pt and After of 0 pt.
- Line Spacing of 1 pt
- Large Images must be in One Column
- Numbering of First Main Headings (Heading 1) must be in Roman Letters, Capital Letter, and Font Size of 10.
- Numbering of Second Main Headings (Heading 2) must be in Alphabets, Italic, and Font Size of 10.

You can use your own standard format also.
Author Guidelines:

1. General,

2. Ethical Guidelines,

3. Submission of Manuscripts,

4. Manuscript’s Category,

5. Structure and Format of Manuscript,

6. After Acceptance.

1. GENERAL

Before submitting your research paper, one is advised to go through the details as mentioned in following heads. It will be beneficial, while peer reviewer justify your paper for publication.

Scope

The Global Journals Inc. (US) welcome the submission of original paper, review paper, survey article relevant to the all the streams of Philosophy and knowledge. The Global Journals Inc. (US) is parental platform for Global Journal of Computer Science and Technology, Researches in Engineering, Medical Research, Science Frontier Research, Human Social Science, Management, and Business organization. The choice of specific field can be done otherwise as following in Abstracting and Indexing Page on this Website. As the all Global

Journals Inc. (US) are being abstracted and indexed (in process) by most of the reputed organizations. Topics of only narrow interest will not be accepted unless they have wider potential or consequences.

2. ETHICAL GUIDELINES

Authors should follow the ethical guidelines as mentioned below for publication of research paper and research activities.

Papers are accepted on strict understanding that the material in whole or in part has not been, nor is being, considered for publication elsewhere. If the paper once accepted by Global Journals Inc. (US) and Editorial Board, will become the copyright of the Global Journals Inc. (US).

Authorship: The authors and coauthors should have active contribution to conception design, analysis and interpretation of findings. They should critically review the contents and drafting of the paper. All should approve the final version of the paper before submission

The Global Journals Inc. (US) follows the definition of authorship set up by the Global Academy of Research and Development. According to the Global Academy of R&D authorship, criteria must be based on:

1) Substantial contributions to conception and acquisition of data, analysis and interpretation of the findings.

2) Drafting the paper and revising it critically regarding important academic content.

3) Final approval of the version of the paper to be published.

All authors should have been credited according to their appropriate contribution in research activity and preparing paper. Contributors who do not match the criteria as authors may be mentioned under Acknowledgement.

Acknowledgements: Contributors to the research other than authors credited should be mentioned under acknowledgement. The specifications of the source of funding for the research if appropriate can be included. Suppliers of resources may be mentioned along with address.

Appeal of Decision: The Editorial Board's decision on publication of the paper is final and cannot be appealed elsewhere.

Permissions: It is the author's responsibility to have prior permission if all or parts of earlier published illustrations are used in this paper.

Please mention proper reference and appropriate acknowledgements wherever expected.

If all or parts of previously published illustrations are used, permission must be taken from the copyright holder concerned. It is the author's responsibility to take these in writing.

Approval for reproduction/modification of any information (including figures and tables) published elsewhere must be obtained by the authors/copyright holders before submission of the manuscript. Contributors (Authors) are responsible for any copyright fee involved.

3. SUBMISSION OF MANUSCRIPTS

Manuscripts should be uploaded via this online submission page. The online submission is most efficient method for submission of papers, as it enables rapid distribution of manuscripts and consequently speeds up the review procedure. It also enables authors to know the status of their own manuscripts by emailing us. Complete instructions for submitting a paper is available below.

Manuscript submission is a systematic procedure and little preparation is required beyond having all parts of your manuscript in a given format and a computer with an Internet connection and a Web browser. Full help and instructions are provided on-screen. As an author, you will be prompted for login and manuscript details as Field of Paper and then to upload your manuscript file(s) according to the instructions.

To avoid postal delays, all transaction is preferred by e-mail. A finished manuscript submission is confirmed by e-mail immediately and your paper enters the editorial process with no postal delays. When a conclusion is made about the publication of your paper by our Editorial Board, revisions can be submitted online with the same procedure, with an occasion to view and respond to all comments.

Complete support for both authors and co-author is provided.

4. MANUSCRIPT'S CATEGORY

Based on potential and nature, the manuscript can be categorized under the following heads:

Original research paper: Such papers are reports of high-level significant original research work.

Review papers: These are concise, significant but helpful and decisive topics for young researchers.

Research articles: These are handled with small investigation and applications

Research letters: The letters are small and concise comments on previously published matters.

5.STRUCTURE AND FORMAT OF MANUSCRIPT

The recommended size of original research paper is less than seven thousand words, review papers fewer than seven thousands words also.Preparation of research paper or how to write research paper, are major hurdle, while writing manuscript. The research articles and research letters should be fewer than three thousand words, the structure original research paper; sometime review paper should be as follows:

Papers: These are reports of significant research (typically less than 7000 words equivalent, including tables, figures, references), and comprise:

(a)Title should be relevant and commensurate with the theme of the paper.

(b) A brief Summary, "Abstract" (less than 150 words) containing the major results and conclusions.

(c) Up to ten keywords, that precisely identifies the paper's subject, purpose, and focus.

(d) An Introduction, giving necessary background excluding subheadings; objectives must be clearly declared.

(e) Resources and techniques with sufficient complete experimental details (wherever possible by reference) to permit repetition; sources of information must be given and numerical methods must be specified by reference, unless non-standard.

(f) Results should be presented concisely, by well-designed tables and/or figures; the same data may not be used in both; suitable statistical data should be given. All data must be obtained with attention to numerical detail in the planning stage. As reproduced design has been recognized to be important to experiments for a considerable time, the Editor has decided that any paper that appears not to have adequate numerical treatments of the data will be returned un-refereed;

(g) Discussion should cover the implications and consequences, not just recapitulating the results; conclusions should be summarizing.

(h) Brief Acknowledgements.

(i) References in the proper form.

Authors should very cautiously consider the preparation of papers to ensure that they communicate efficiently. Papers are much more likely to be accepted, if they are cautiously designed and laid out, contain few or no errors, are summarizing, and be conventional to the approach and instructions. They will in addition, be published with much less delays than those that require much technical and editorial correction.

The Editorial Board reserves the right to make literary corrections and to make suggestions to improve briefness.

It is vital, that authors take care in submitting a manuscript that is written in simple language and adheres to published guidelines.

Format

Language: The language of publication is UK English. Authors, for whom English is a second language, must have their manuscript efficiently edited by an English-speaking person before submission to make sure that, the English is of high excellence. It is preferable, that manuscripts should be professionally edited.

Standard Usage, Abbreviations, and Units: Spelling and hyphenation should be conventional to The Concise Oxford English Dictionary. Statistics and measurements should at all times be given in figures, e.g. 16 min, except for when the number begins a sentence. When the number does not refer to a unit of measurement it should be spelt in full unless, it is 160 or greater.

Abbreviations supposed to be used carefully. The abbreviated name or expression is supposed to be cited in full at first usage, followed by the conventional abbreviation in parentheses.

Metric SI units are supposed to generally be used excluding where they conflict with current practice or are confusing. For illustration, 1.4 l rather than 1.4 × 10-3 m3, or 4 mm somewhat than 4 × 10-3 m. Chemical formula and solutions must identify the form used, e.g. anhydrous or hydrated, and the concentration must be in clearly defined units. Common species names should be followed by underlines at the first mention. For following use the generic name should be constricted to a single letter, if it is clear.

Structure

All manuscripts submitted to Global Journals Inc. (US), ought to include:

Title: The title page must carry an instructive title that reflects the content, a running title (less than 45 characters together with spaces), names of the authors and co-authors, and the place(s) wherever the work was carried out. The full postal address in addition with the e-mail address of related author must be given. Up to eleven keywords or very brief phrases have to be given to help data retrieval, mining and indexing.

Abstract, used in Original Papers and Reviews:

Optimizing Abstract for Search Engines

Many researchers searching for information online will use search engines such as Google, Yahoo or similar. By optimizing your paper for search engines, you will amplify the chance of someone finding it. This in turn will make it more likely to be viewed and/or cited in a further work. Global Journals Inc. (US) have compiled these guidelines to facilitate you to maximize the web-friendliness of the most public part of your paper.

Key Words

A major linchpin in research work for the writing research paper is the keyword search, which one will employ to find both library and Internet resources.

One must be persistent and creative in using keywords. An effective keyword search requires a strategy and planning a list of possible keywords and phrases to try.

Search engines for most searches, use Boolean searching, which is somewhat different from Internet searches. The Boolean search uses "operators," words (and, or, not, and near) that enable you to expand or narrow your affords. Tips for research paper while preparing research paper are very helpful guideline of research paper.

Choice of key words is first tool of tips to write research paper. Research paper writing is an art.A few tips for deciding as strategically as possible about keyword search:

- One should start brainstorming lists of possible keywords before even begin searching. Think about the most important concepts related to research work. Ask, "What words would a source have to include to be truly valuable in research paper?" Then consider synonyms for the important words.
- It may take the discovery of only one relevant paper to let steer in the right keyword direction because in most databases, the keywords under which a research paper is abstracted are listed with the paper.
- One should avoid outdated words.

Keywords are the key that opens a door to research work sources. Keyword searching is an art in which researcher's skills are bound to improve with experience and time.

Numerical Methods: Numerical methods used should be clear and, where appropriate, supported by references.

Acknowledgements: Please make these as concise as possible.

References

References follow the Harvard scheme of referencing. References in the text should cite the authors' names followed by the time of their publication, unless there are three or more authors when simply the first author's name is quoted followed by et al. unpublished work has to only be cited where necessary, and only in the text. Copies of references in press in other journals have to be supplied with submitted typescripts. It is necessary that all citations and references be carefully checked before submission, as mistakes or omissions will cause delays.

References to information on the World Wide Web can be given, but only if the information is available without charge to readers on an official site. Wikipedia and Similar websites are not allowed where anyone can change the information. Authors will be asked to make available electronic copies of the cited information for inclusion on the Global Journals Inc. (US) homepage at the judgment of the Editorial Board.

The Editorial Board and Global Journals Inc. (US) recommend that, citation of online-published papers and other material should be done via a DOI (digital object identifier). If an author cites anything, which does not have a DOI, they run the risk of the cited material not being noticeable.

The Editorial Board and Global Journals Inc. (US) recommend the use of a tool such as Reference Manager for reference management and formatting.

Tables, Figures and Figure Legends

Tables: Tables should be few in number, cautiously designed, uncrowned, and include only essential data. Each must have an Arabic number, e.g. Table 4, a self-explanatory caption and be on a separate sheet. Vertical lines should not be used.

Figures: Figures are supposed to be submitted as separate files. Always take in a citation in the text for each figure using Arabic numbers, e.g. Fig. 4. Artwork must be submitted online in electronic form by e-mailing them.

Preparation of Electronic Figures for Publication

Even though low quality images are sufficient for review purposes, print publication requires high quality images to prevent the final product being blurred or fuzzy. Submit (or e-mail) EPS (line art) or TIFF (halftone/photographs) files only. MS PowerPoint and Word Graphics are unsuitable for printed pictures. Do not use pixel-oriented software. Scans (TIFF only) should have a resolution of at least 350 dpi (halftone) or 700 to 1100 dpi (line drawings) in relation to the imitation size. Please give the data for figures in black and white or submit a Color Work Agreement Form. EPS files must be saved with fonts embedded (and with a TIFF preview, if possible).

For scanned images, the scanning resolution (at final image size) ought to be as follows to ensure good reproduction: line art: >650 dpi; halftones (including gel photographs) : >350 dpi; figures containing both halftone and line images: >650 dpi.

Color Charges: It is the rule of the Global Journals Inc. (US) for authors to pay the full cost for the reproduction of their color artwork. Hence, please note that, if there is color artwork in your manuscript when it is accepted for publication, we would require you to complete and return a color work agreement form before your paper can be published.

Figure Legends: Self-explanatory legends of all figures should be incorporated separately under the heading 'Legends to Figures'. In the full-text online edition of the journal, figure legends may possibly be truncated in abbreviated links to the full screen version. Therefore, the first 100 characters of any legend should notify the reader, about the key aspects of the figure.

6. AFTER ACCEPTANCE

Upon approval of a paper for publication, the manuscript will be forwarded to the dean, who is responsible for the publication of the Global Journals Inc. (US).

6.1 Proof Corrections

The corresponding author will receive an e-mail alert containing a link to a website or will be attached. A working e-mail address must therefore be provided for the related author.

Acrobat Reader will be required in order to read this file. This software can be downloaded

(Free of charge) from the following website:

www.adobe.com/products/acrobat/readstep2.html. This will facilitate the file to be opened, read on screen, and printed out in order for any corrections to be added. Further instructions will be sent with the proof.

Proofs must be returned to the dean at dean@globaljournals.org within three days of receipt.

As changes to proofs are costly, we inquire that you only correct typesetting errors. All illustrations are retained by the publisher. Please note that the authors are responsible for all statements made in their work, including changes made by the copy editor.

6.2 Early View of Global Journals Inc. (US) (Publication Prior to Print)

The Global Journals Inc. (US) are enclosed by our publishing's Early View service. Early View articles are complete full-text articles sent in advance of their publication. Early View articles are absolute and final. They have been completely reviewed, revised and edited for publication, and the authors' final corrections have been incorporated. Because they are in final form, no changes can be made after sending them. The nature of Early View articles means that they do not yet have volume, issue or page numbers, so Early View articles cannot be cited in the conventional way.

6.3 Author Services

Online production tracking is available for your article through Author Services. Author Services enables authors to track their article - once it has been accepted - through the production process to publication online and in print. Authors can check the status of their articles online and choose to receive automated e-mails at key stages of production. The authors will receive an e-mail with a unique link that enables them to register and have their article automatically added to the system. Please ensure that a complete e-mail address is provided when submitting the manuscript.

6.4 Author Material Archive Policy

Please note that if not specifically requested, publisher will dispose off hardcopy & electronic information submitted, after the two months of publication. If you require the return of any information submitted, please inform the Editorial Board or dean as soon as possible.

6.5 Offprint and Extra Copies

A PDF offprint of the online-published article will be provided free of charge to the related author, and may be distributed according to the Publisher's terms and conditions. Additional paper offprint may be ordered by emailing us at: editor@globaljournals.org .

the search? Will I be able to find all information in this field area? If the answer of these types of questions will be "Yes" then you can choose that topic. In most of the cases, you may have to conduct the surveys and have to visit several places because this field is related to Computer Science and Information Technology. Also, you may have to do a lot of work to find all rise and falls regarding the various data of that subject. Sometimes, detailed information plays a vital role, instead of short information.

2. Evaluators are human: First thing to remember that evaluators are also human being. They are not only meant for rejecting a paper. They are here to evaluate your paper. So, present your Best.

3. Think Like Evaluators: If you are in a confusion or getting demotivated that your paper will be accepted by evaluators or not, then think and try to evaluate your paper like an Evaluator. Try to understand that what an evaluator wants in your research paper and automatically you will have your answer.

4. Make blueprints of paper: The outline is the plan or framework that will help you to arrange your thoughts. It will make your paper logical. But remember that all points of your outline must be related to the topic you have chosen.

5. Ask your Guides: If you are having any difficulty in your research, then do not hesitate to share your difficulty to your guide (if you have any). They will surely help you out and resolve your doubts. If you can't clarify what exactly you require for your work then ask the supervisor to help you with the alternative. He might also provide you the list of essential readings.

6. Use of computer is recommended: As you are doing research in the field of Computer Science, then this point is quite obvious.

7. Use right software: Always use good quality software packages. If you are not capable to judge good software then you can lose quality of your paper unknowingly. There are various software programs available to help you, which you can get through Internet.

8. Use the Internet for help: An excellent start for your paper can be by using the Google. It is an excellent search engine, where you can have your doubts resolved. You may also read some answers for the frequent question how to write my research paper or find model research paper. From the internet library you can download books. If you have all required books make important reading selecting and analyzing the specified information. Then put together research paper sketch out.

9. Use and get big pictures: Always use encyclopedias, Wikipedia to get pictures so that you can go into the depth.

10. Bookmarks are useful: When you read any book or magazine, you generally use bookmarks, right! It is a good habit, which helps to not to lose your continuity. You should always use bookmarks while searching on Internet also, which will make your search easier.

11. Revise what you wrote: When you write anything, always read it, summarize it and then finalize it.

12. Make all efforts: Make all efforts to mention what you are going to write in your paper. That means always have a good start. Try to mention everything in introduction, that what is the need of a particular research paper. Polish your work by good skill of writing and always give an evaluator, what he wants.

13. Have backups: When you are going to do any important thing like making research paper, you should always have backup copies of it either in your computer or in paper. This will help you to not to lose any of your important.

14. Produce good diagrams of your own: Always try to include good charts or diagrams in your paper to improve quality. Using several and unnecessary diagrams will degrade the quality of your paper by creating "hotchpotch." So always, try to make and include those diagrams, which are made by your own to improve readability and understandability of your paper.

15. Use of direct quotes: When you do research relevant to literature, history or current affairs then use of quotes become essential but if study is relevant to science then use of quotes is not preferable.

16. Use proper verb tense: Use proper verb tenses in your paper. Use past tense, to present those events that happened. Use present tense to indicate events that are going on. Use future tense to indicate future happening events. Use of improper and wrong tenses will confuse the evaluator. Avoid the sentences that are incomplete.

17. Never use online paper: If you are getting any paper on Internet, then never use it as your research paper because it might be possible that evaluator has already seen it or maybe it is outdated version.

18. **Pick a good study spot:** To do your research studies always try to pick a spot, which is quiet. Every spot is not for studies. Spot that suits you choose it and proceed further.

19. Know what you know: Always try to know, what you know by making objectives. Else, you will be confused and cannot achieve your target.

20. Use good quality grammar: Always use a good quality grammar and use words that will throw positive impact on evaluator. Use of good quality grammar does not mean to use tough words, that for each word the evaluator has to go through dictionary. Do not start sentence with a conjunction. Do not fragment sentences. Eliminate one-word sentences. Ignore passive voice. Do not ever use a big word when a diminutive one would suffice. Verbs have to be in agreement with their subjects. Prepositions are not expressions to finish sentences with. It is incorrect to ever divide an infinitive. Avoid clichés like the disease. Also, always shun irritating alliteration. Use language that is simple and straight forward. put together a neat summary.

21. Arrangement of information: Each section of the main body should start with an opening sentence and there should be a changeover at the end of the section. Give only valid and powerful arguments to your topic. You may also maintain your arguments with records.

22. Never start in last minute: Always start at right time and give enough time to research work. Leaving everything to the last minute will degrade your paper and spoil your work.

23. Multitasking in research is not good: Doing several things at the same time proves bad habit in case of research activity. Research is an area, where everything has a particular time slot. Divide your research work in parts and do particular part in particular time slot.

24. Never copy others' work: Never copy others' work and give it your name because if evaluator has seen it anywhere you will be in trouble.

25. Take proper rest and food: No matter how many hours you spend for your research activity, if you are not taking care of your health then all your efforts will be in vain. For a quality research, study is must, and this can be done by taking proper rest and food.

26. Go for seminars: Attend seminars if the topic is relevant to your research area. Utilize all your resources.

27. Refresh your mind after intervals: Try to give rest to your mind by listening to soft music or by sleeping in intervals. This will also improve your memory.

28. Make colleagues: Always try to make colleagues. No matter how sharper or intelligent you are, if you make colleagues you can have several ideas, which will be helpful for your research.

29. **Think technically:** Always think technically. If anything happens, then search its reasons, its benefits, and demerits.

30. Think and then print: When you will go to print your paper, notice that tables are not be split, headings are not detached from their descriptions, and page sequence is maintained.

31. Adding unnecessary information: Do not add unnecessary information, like, I have used MS Excel to draw graph. Do not add irrelevant and inappropriate material. These all will create superfluous. Foreign terminology and phrases are not apropos. One should NEVER take a broad view. Analogy in script is like feathers on a snake. Not at all use a large word when a very small one would be sufficient. Use words properly, regardless of how others use them. Remove quotations. Puns are for kids, not grunt readers. Amplification is a billion times of inferior quality than sarcasm.

32. Never oversimplify everything: To add material in your research paper, never go for oversimplification. This will definitely irritate the evaluator. Be more or less specific. Also too, by no means, ever use rhythmic redundancies. Contractions aren't essential and shouldn't be there used. Comparisons are as terrible as clichés. Give up ampersands and abbreviations, and so on. Remove commas, that are, not necessary. Parenthetical words however should be together with this in commas. Understatement is all the time the complete best way to put onward earth-shaking thoughts. Give a detailed literary review.

33. Report concluded results: Use concluded results. From raw data, filter the results and then conclude your studies based on measurements and observations taken. Significant figures and appropriate number of decimal places should be used. Parenthetical remarks are prohibitive. Proofread carefully at final stage. In the end give outline to your arguments. Spot out perspectives of further study of this subject. Justify your conclusion by at the bottom of them with sufficient justifications and examples.

34. After conclusion: Once you have concluded your research, the next most important step is to present your findings. Presentation is extremely important as it is the definite medium though which your research is going to be in print to the rest of the crowd. Care should be taken to categorize your thoughts well and present them in a logical and neat manner. A good quality research paper format is essential because it serves to highlight your research paper and bring to light all necessary aspects in your research.

Informal Guidelines of Research Paper Writing

Key points to remember:

- Submit all work in its final form.
- Write your paper in the form, which is presented in the guidelines using the template.
- Please note the criterion for grading the final paper by peer-reviewers.

Final Points:

A purpose of organizing a research paper is to let people to interpret your effort selectively. The journal requires the following sections, submitted in the order listed, each section to start on a new page.

The introduction will be compiled from reference matter and will reflect the design processes or outline of basis that direct you to make study. As you will carry out the process of study, the method and process section will be constructed as like that. The result segment will show related statistics in nearly sequential order and will direct the reviewers next to the similar intellectual paths throughout the data that you took to carry out your study. The discussion section will provide understanding of the data and projections as to the implication of the results. The use of good quality references all through the paper will give the effort trustworthiness by representing an alertness of prior workings.

Writing a research paper is not an easy job no matter how trouble-free the actual research or concept. Practice, excellent preparation, and controlled record keeping are the only means to make straightforward the progression.

General style:

Specific editorial column necessities for compliance of a manuscript will always take over from directions in these general guidelines.

To make a paper clear

· Adhere to recommended page limits

Mistakes to evade

- Insertion a title at the foot of a page with the subsequent text on the next page

- Separating a table/chart or figure - impound each figure/table to a single page
- Submitting a manuscript with pages out of sequence

In every sections of your document

· Use standard writing style including articles ("a", "the," etc.)

· Keep on paying attention on the research topic of the paper

· Use paragraphs to split each significant point (excluding for the abstract)

· Align the primary line of each section

· Present your points in sound order

· Use present tense to report well accepted

· Use past tense to describe specific results

· Shun familiar wording, don't address the reviewer directly, and don't use slang, slang language, or superlatives

· Shun use of extra pictures - include only those figures essential to presenting results

Title Page:

Choose a revealing title. It should be short. It should not have non-standard acronyms or abbreviations. It should not exceed two printed lines. It should include the name(s) and address (es) of all authors.

Abstract:

The summary should be two hundred words or less. It should briefly and clearly explain the key findings reported in the manuscript--must have precise statistics. It should not have abnormal acronyms or abbreviations. It should be logical in itself. Shun citing references at this point.

An abstract is a brief distinct paragraph summary of finished work or work in development. In a minute or less a reviewer can be taught the foundation behind the study, common approach to the problem, relevant results, and significant conclusions or new questions.

Write your summary when your paper is completed because how can you write the summary of anything which is not yet written? Wealth of terminology is very essential in abstract. Yet, use comprehensive sentences and do not let go readability for briefness. You can maintain it succinct by phrasing sentences so that they provide more than lone rationale. The author can at this moment go straight to shortening the outcome. Sum up the study, with the subsequent elements in any summary. Try to maintain the initial two items to no more than one ruling each.

- Reason of the study - theory, overall issue, purpose
- Fundamental goal
- To the point depiction of the research
- Consequences, including definite statistics - if the consequences are quantitative in nature, account quantitative data; results of any numerical analysis should be reported
- Significant conclusions or questions that track from the research(es)

Approach:

- Single section, and succinct
- As a outline of job done, it is always written in past tense
- A conceptual should situate on its own, and not submit to any other part of the paper such as a form or table
- Center on shortening results - bound background information to a verdict or two, if completely necessary
- What you account in an conceptual must be regular with what you reported in the manuscript
- Exact spelling, clearness of sentences and phrases, and appropriate reporting of quantities (proper units, important statistics) are just as significant in an abstract as they are anywhere else

Introduction:

The **Introduction** should "introduce" the manuscript. The reviewer should be presented with sufficient background information to be capable to comprehend and calculate the purpose of your study without having to submit to other works. The basis for the study should be offered. Give most important references but shun difficult to make a comprehensive appraisal of the topic. In the introduction, describe the problem visibly. If the problem is not acknowledged in a logical, reasonable way, the reviewer will have no attention in your result. Speak in common terms about techniques used to explain the problem, if needed, but do not present any particulars about the protocols here. Following approach can create a valuable beginning:

- Explain the value (significance) of the study
- Shield the model - why did you employ this particular system or method? What is its compensation? You strength remark on its appropriateness from a abstract point of vision as well as point out sensible reasons for using it.
- Present a justification. Status your particular theory (es) or aim(s), and describe the logic that led you to choose them.
- Very for a short time explain the tentative propose and how it skilled the declared objectives.

Approach:

- Use past tense except for when referring to recognized facts. After all, the manuscript will be submitted after the entire job is done.
- Sort out your thoughts; manufacture one key point with every section. If you make the four points listed above, you will need a least of four paragraphs.
- Present surroundings information only as desirable in order hold up a situation. The reviewer does not desire to read the whole thing you know about a topic.
- Shape the theory/purpose specifically - do not take a broad view.
- As always, give awareness to spelling, simplicity and correctness of sentences and phrases.

Procedures (Methods and Materials):

This part is supposed to be the easiest to carve if you have good skills. A sound written Procedures segment allows a capable scientist to replacement your results. Present precise information about your supplies. The suppliers and clarity of reagents can be helpful bits of information. Present methods in sequential order but linked methodologies can be grouped as a segment. Be concise when relating the protocols. Attempt for the least amount of information that would permit another capable scientist to spare your outcome but be cautious that vital information is integrated. The use of subheadings is suggested and ought to be synchronized with the results section. When a technique is used that has been well described in another object, mention the specific item describing a way but draw the basic principle while stating the situation. The purpose is to text all particular resources and broad procedures, so that another person may use some or all of the methods in one more study or referee the scientific value of your work. It is not to be a step by step report of the whole thing you did, nor is a methods section a set of orders.

Materials:

- Explain materials individually only if the study is so complex that it saves liberty this way.
- Embrace particular materials, and any tools or provisions that are not frequently found in laboratories.
- Do not take in frequently found.
- If use of a definite type of tools.
- Materials may be reported in a part section or else they may be recognized along with your measures.

Methods:

- Report the method (not particulars of each process that engaged the same methodology)
- Describe the method entirely
- To be succinct, present methods under headings dedicated to specific dealings or groups of measures
- Simplify - details how procedures were completed not how they were exclusively performed on a particular day.
- If well known procedures were used, account the procedure by name, possibly with reference, and that's all.

Approach:

- It is embarrassed or not possible to use vigorous voice when documenting methods with no using first person, which would focus the reviewer's interest on the researcher rather than the job. As a result when script up the methods most authors use third person passive voice.
- Use standard style in this and in every other part of the paper - avoid familiar lists, and use full sentences.

What to keep away from

- Resources and methods are not a set of information.
- Skip all descriptive information and surroundings - save it for the argument.
- Leave out information that is immaterial to a third party.

Results:

The principle of a results segment is to present and demonstrate your conclusion. Create this part a entirely objective details of the outcome, and save all understanding for the discussion.

The page length of this segment is set by the sum and types of data to be reported. Carry on to be to the point, by means of statistics and tables, if suitable, to present consequences most efficiently.You must obviously differentiate material that would usually be incorporated in a study editorial from any unprocessed data or additional appendix matter that would not be available. In fact, such matter should not be submitted at all except requested by the instructor.

Content

- Sum up your conclusion in text and demonstrate them, if suitable, with figures and tables.
- In manuscript, explain each of your consequences, point the reader to remarks that are most appropriate.
- Present a background, such as by describing the question that was addressed by creation an exacting study.
- Explain results of control experiments and comprise remarks that are not accessible in a prescribed figure or table, if appropriate.
- Examine your data, then prepare the analyzed (transformed) data in the form of a figure (graph), table, or in manuscript form.

What to stay away from

- Do not discuss or infer your outcome, report surroundings information, or try to explain anything.
- Not at all, take in raw data or intermediate calculations in a research manuscript.

- Do not present the similar data more than once.
- Manuscript should complement any figures or tables, not duplicate the identical information.
- Never confuse figures with tables - there is a difference.

Approach

- As forever, use past tense when you submit to your results, and put the whole thing in a reasonable order.
- Put figures and tables, appropriately numbered, in order at the end of the report
- If you desire, you may place your figures and tables properly within the text of your results part.

Figures and tables

- If you put figures and tables at the end of the details, make certain that they are visibly distinguished from any attach appendix materials, such as raw facts
- Despite of position, each figure must be numbered one after the other and complete with subtitle
- In spite of position, each table must be titled, numbered one after the other and complete with heading
- All figure and table must be adequately complete that it could situate on its own, divide from text

Discussion:

The Discussion is expected the trickiest segment to write and describe. A lot of papers submitted for journal are discarded based on problems with the Discussion. There is no head of state for how long a argument should be. Position your understanding of the outcome visibly to lead the reviewer through your conclusions, and then finish the paper with a summing up of the implication of the study. The purpose here is to offer an understanding of your results and hold up for all of your conclusions, using facts from your research and generally accepted information, if suitable. The implication of result should be visibly described. Infer your data in the conversation in suitable depth. This means that when you clarify an observable fact you must explain mechanisms that may account for the observation. If your results vary from your prospect, make clear why that may have happened. If your results agree, then explain the theory that the proof supported. It is never suitable to just state that the data approved with prospect, and let it drop at that.

- Make a decision if each premise is supported, discarded, or if you cannot make a conclusion with assurance. Do not just dismiss a study or part of a study as "uncertain."
- Research papers are not acknowledged if the work is imperfect. Draw what conclusions you can based upon the results that you have, and take care of the study as a finished work
- You may propose future guidelines, such as how the experiment might be personalized to accomplish a new idea.
- Give details all of your remarks as much as possible, focus on mechanisms.
- Make a decision if the tentative design sufficiently addressed the theory, and whether or not it was correctly restricted.
- Try to present substitute explanations if sensible alternatives be present.
- One research will not counter an overall question, so maintain the large picture in mind, where do you go next? The best studies unlock new avenues of study. What questions remain?
- Recommendations for detailed papers will offer supplementary suggestions.

Approach:

- When you refer to information, differentiate data generated by your own studies from available information
- Submit to work done by specific persons (including you) in past tense.
- Submit to generally acknowledged facts and main beliefs in present tense.

Administration Rules Listed Before Submitting Your Research Paper to Global Journals Inc. (US)

Please carefully note down following rules and regulation before submitting your Research Paper to Global Journals Inc. (US):

Segment Draft and Final Research Paper: You have to strictly follow the template of research paper. If it is not done your paper may get rejected.

- The **major constraint** is that you must independently make all content, tables, graphs, and facts that are offered in the paper. You must write each part of the paper wholly on your own. The Peer-reviewers need to identify your own perceptive of the concepts in your own terms. NEVER extract straight from any foundation, and never rephrase someone else's analysis.

- Do not give permission to anyone else to "PROOFREAD" your manuscript.

- Methods to avoid Plagiarism is applied by us on every paper, if found guilty, you will be blacklisted by all of our collaborated research groups, your institution will be informed for this and strict legal actions will be taken immediately.)
- To guard yourself and others from possible illegal use please do not permit anyone right to use to your paper and files.

Criterion for Grading a Research Paper (Compilation) by Global Journals Inc. (US)

Please note that following table is only a Grading of "Paper Compilation" and not on "Performed/Stated Research" whose grading solely depends on Individual Assigned Peer Reviewer and Editorial Board Member. These can be available only on request and after decision of Paper. This report will be the property of Global Journals Inc. (US).

	Grades		
	A-B	**C-D**	**E-F**
Abstract	Clear and concise with appropriate content, Correct format. 200 words or below	Unclear summary and no specific data, Incorrect form Above 200 words	No specific data with ambiguous information Above 250 words
Introduction	Containing all background details with clear goal and appropriate details, flow specification, no grammar and spelling mistake, well organized sentence and paragraph, reference cited	Unclear and confusing data, appropriate format, grammar and spelling errors with unorganized matter	Out of place depth and content, hazy format
Methods and Procedures	Clear and to the point with well arranged paragraph, precision and accuracy of facts and figures, well organized subheads	Difficult to comprehend with embarrassed text, too much explanation but completed	Incorrect and unorganized structure with hazy meaning
Result	Well organized, Clear and specific, Correct units with precision, correct data, well structuring of paragraph, no grammar and spelling mistake	Complete and embarrassed text, difficult to comprehend	Irregular format with wrong facts and figures
Discussion	Well organized, meaningful specification, sound conclusion, logical and concise explanation, highly structured paragraph reference cited	Wordy, unclear conclusion, spurious	Conclusion is not cited, unorganized, difficult to comprehend
References	Complete and correct format, well organized	Beside the point, Incomplete	Wrong format and structuring

Index

A

B

C

D

E

F

G

H

I

K

M

P

Q

R

S

T

U

V

W

save our planet

Global Journal of Researches in Engineering

Visit us on the Web at www.GlobalJournals.org | www.EngineeringResearch.org
or email us at helpdesk@globaljournals.org

ISSN 9755861