



## Link Analysis Extension for Mining Relational Databases

G.P.Narasaiah<sup>1</sup>, S. Jessica Saritha<sup>2</sup>

<sup>1</sup>M.Tech. student, Dept of CSE, JNTUACE, Pulivendula, A.P. India, narasaiah.mtech@gmail.com

<sup>2</sup>Asst. Professor, Dept of CSE, JNTUACE, Pulivendula, A.P. India.

### ABSTRACT

This paper on link analysis extension employed by data processing systems to mining associations between individual information records or information sets concerned within the same event. It demonstrates associate in Nursing implementation of the formula with custom modifications created to expand practicality and improve time and house quality. The algorithms area unit integrated in an exceedingly easy system which may be accustomed generate frequent itemsets and extract association rules on-line in real time. principally the formula refers to computer-based techniques utilized in characteristic, extracting, gathering, storing, examine (something) methodically and in detail, and providing access to information to assist enterprise users create higher business selections analyzing business information. The formula could also be divided into reportage systems and data processing applications. data processing is data Discovery in information and also the science of extracting helpful data from immense information repositories. data processing applications usually use subtle mathematical and applied math techniques to perform information analysis, seek for specific patterns or relationships, if they exist, and create future predictions.

**Key words :** Data mining, multi-relational data mining, k-means algorithm

### 1. INTRODUCTION

Data mining, the extraction of hidden predictive information from wide range or scope databases, is a powerful new technology with great potential. Data mining is a knowledge discovery process in large and complex data sets, refers to mining knowledge from large amounts of data. Moreover, data mining can be used to predict an outcome for a given entity (Hernández et al., 2006). Thus clustering algorithms in data mining are equivalent to the task of identifying groups of records that are similar between themselves but different from the rest. (Varan, 2006). Data mining is a multidisciplinary field with many techniques. With these techniques you can create a mining model that described the data that you will use. (see Figure 1).

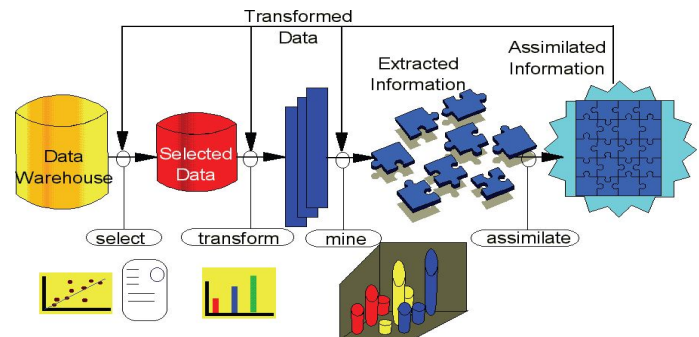


Figure 1 Data Mining Process

**A. Data Set.** It is a collection of data, routinely presented in tabular form. Each column represents a especially great or intense versatile. Each row corresponds to a given member of the data set in question. It lists values for each of the versatile, such as height Some elements in Data Mining Process are weight of an object or values of random numbers. Pronominal value is known as a datum. The data set may consist data for one or more members, match to the number of rows.

**B. Pre-processing.** Data mining requires substantial pre-processing of data. This was especially the case of the behavioral data. To make the data comparable, all data needs to be normalized.

**C. General Results.** Data mining gives general results by using data mining techniques and algorithms. Now a days number of corporate companies and different sector using Data mining algorithms for mining the large amount of data

**D. Decision Trees.** Decision trees are powerful and popular tools for classification and to foretell the future. The attractiveness of decision trees is due to the fact that, these decision trees represent rules. Decision tree induction is a typical inductive approach to learn knowledge on classification. The key requirements to do extracting with decision trees are: Attribute-value description, Predefined classes, Discrete classes and Sufficient data.

**E. Association Rules.** Association rules describe events that tend to happen together. They are formal statements in the organize of  $X \Rightarrow Y$ , where if X happens, Y is likely to happen (Márquez al., 2008).

The goal of this dissertation is to create efficient data mining algorithms that can find interesting (descriptive) patterns when the database in the knowledge discovery process is a relational database.

## 2. LINK ANALYSIS

A link-analysis-based technique permitting to find relationships existing between components of a electronic database or, additional usually, a graph. additional specifically, this work is predicated on a stochastic process through the information shaping a Marko chain having as many countries as components within the information. Suppose, as an example, we tend to have an interest in analyzing the relationships between components contained in 2 completely different tables of a electronic database. to the present finish, a ballroom dance procedure is developed. First, a far smaller, reduced, Markov chain, solely containing the weather of interest—typically the weather contained within the 2 tables—and conserving the most characteristics of the initial chain, is extracted by random complementation. associate economical algorithmic program for extracting the reduced Markov chain from the massive, sparse, Markov chain representing the information is projected. Then, the reduced chain is analyzed by, as an example, sticking the states within the mathematical space spanned by the proper eigenvectors of the transition matrix known as the fundamental diffusion map during this paper), or by computing a kernel principal part analysis on a diffusion map kernel computed from the reduced graph and visualizing the results. Indeed, a legitimate graph kernel supported the diffusion map distance, extending the fundamental diffusion map to graphs, is introduced.

## 3. RELATIONAL DATABASES

A database management system (DBMS) is a software system that enables the develop, maintenance and use of large amounts of data [Abiteboul *et al.*, 1995]. Relational Database Management Systems (RDBMS) are based on the relational model [Codd, 1970], which has been the dominant paradigm for industrial database applications during the final decades, and it is at the core of all major commercial database systems, making RDBMSs one of the most commonly available kinds of repositories. A relational database is a collection of tables called relations, each of which is assigned a unique name. Each tuple consists of a set of attributes and usually stores a bulky set of tuples. Each tuple in a relational table represents an object identified by a unique key and described by a set of attribute values. Often one uses a semantic model to represent relational databases, allowing one to describe and design the database without having to pay attention to the physical database. Such a model is often referred to as a database scheme. One of the most common models is the Entity-Relationship (ER) model, which we will use in this

dissertation. An Entity- Relationship model represents the database as entity sets and their relationships. An entity set is a collection of entities. Each entity set has a set of proper- ties, called attributes. A subset of these attributes is called the key, and is unique for each entity in the entity set.

### 3.1 Relational data

A relational database typically consists of several tables (relations) and not just one table. The example database in Table 1 has two relations: Customer and Married To. Note that relations can be defined extensionally (by tables, as in our example) or intentionally through database views (as explicit logical rules). The latter typically represent relationships that can be inferred from other relationships. For example, having extensional representations of the relations mother and father, we can intentionally define the relations grandparent, grandmother, sibling, and ancestor, among others. In tentional definitions of relations typically represent general knowledge about the domain of discourse. For example, if we have extensional relations listing the atoms that make a compound molecule and the bonds between them, functional groups of atoms can be defined intentionally. Such general knowledge is called domain knowledge or background knowledge.

## 4. DATA MINING TECHINQUES

Data Mining Technique can be broadly divided into two categories .These are as below

- Unique Table Approach Mining
- Structured Data Mining

### 4.1. Unique Table Approach Mining.

This unique table approach mining is popularly known as propositional data mining. The important task of the data mining is known as Propositional Data Mining. The aim assumption is that each individual is represented by a fixed set of characteristics which is known as attributes. Again the individual can be thought of as a collection of attribute –value pairs, which are represented as a vector format [7] In this approach the central database of individuals becomes a table: which consists of rows(or tuples) correspond to individuals and columns (attribute).If we want to establish the relation among the different tables (which may in the form of Primary and Foreign key relationships) then we have encounter the two problems .There are two ways to handle the problem

**First Problem:** Compute the join but this leads to data redundancy, Missing values. A single instance in the original database is mapped onto multiple instances in the new table, which is the problematic.

**Second Problem:** The second way is aggregate the information in different tuples representing the same individual into one tuples after computing the join. Which may be solved the problem but causes the loss of information.

## 4.2 Structured Data Mining

The term structured data mining means to handle the complex data. That is this kind of data that cannot store in the sensible in one table or doesn't have a single table representation in which table rows or columns are not related to each other. Now a day's everyone using multirelational databases i.e. databases of multiple tables. Some problems involving molecules, proteins, phylogenetic trees, social networks and web server access logs cannot be tackled if rows in a single table are not explicitly linked to each other. To deal with structured data all kinds of problems have to be tackled that are almost trivially solved in attribute-value mining algorithms. In present day's scenario, the amount of data is available hugely so human must take help from automatic computerized methods for mining the required information from relational databases. These types of data must be represented in the form of graph data structure which will represent the nodes and their attributes, relationships with the other entities.

## 5 STRUCTURED DATA MINING TYPES

There are four categories are available these are as below:

### 5.1. Graph Mining

Graph mining is the techniques which will extract the required information from data represented in the form of graph structured form. A graph can be defined as the equation  $G=\{V,E\}$ , Where  $V =\{v_1,v_2,v_3,\dots\dots\dots v_n\}$  is an ordered set of vertices in the graph and  $E=\{e_1,e_2,e_3,\dots\dots\dots e_n\}$  is the set of pair of edges. The term graph mining which can refer to discover the graph patterns. Chakrabarti and Faloutsos has defined the two scenarios: First: It is the typical of web domains and the second one is: database of chemical compounds. The main objective of the Graph Mining is the concept of frequent graph pattern. Graphs becomes progressively important in modeling elaborate structures ,such as circuits , chemical compounds , images, biological networks ,social networks, protein structures, XML documents and workflows. Many graphs search algorithms have been evolve in video list, computer vision, chemical informatics and text mining. With the increasing involve on the analysis of large amounts of organized information from relational databases, graph mining has become strategic and an active theme in data mining. Graph mining is used to mine frequent graph patterns and perform characterization, discrimination, classification [18] etc.

### 5.2 Inductive Logic Programming (ILP)

This ILP paradigm says that how the logic program will convert the patterns. The logic program induced from a database of logical facts which is called as ILP. ILP follows the top-down approach. The advantage of ILP is too expressive

and powerful rules are understandable the disadvantage of ILP is Inefficient for the database with the complex schema as well as not properly handled the continuous attributes.

The database consists of a collection of facts in first-order logic) [19], [20]. Each fact represents a separate, and individuals can be reconstructed by piecing together these facts. First-order logic (often Prolog) can be used to select subgroups. [12]

### 5.3 Semi-Structured Data Mining

Now a days the data mining is concerned with the discovery of patterns as well as the relations .Almost all the data mining algorithms handle the data with the fixed form but the schema is defined with advance When the data is on the web then its structure is regular form .We normally called as semi-structured data mining .To handle such type of data we use the XML language .XML not only handles the tabular data but also the arbitrary trees. Semi-structured data is naturally modeled in terms of graphs contain labels that give semantics to the underlying structure. The database consists of XML documents, which describe objects in a mixture of structural and free-text information. [13]

### 5.4 Multi-Relational Data Mining (MRDM)

The database consists of a collection of relational database. Records in each table represent separate , and individuals can be reconstructed by joining over the foreign key relations between the tables. [8][9][10][11]. the database consists of a collection of relational database. Records in each table Represent separate, and individuals can be reconstructed by joining over the foreign key relations between the tables. Subgroups can be defined by means of SQL. [14]

#### 5.4.1. Selection Graph:

When multi relational patterns are expressed in terms of graphical language then we called as the Selection Graph. Selection graph model is used in SQL (Database Language) to directly deal with tables. There are so many algorithms was developed which aims to convert the selection graph model into SQL query form. [15] . Anna Atramentov et.al also focused the graphical representation of graph. [16]

#### 5.4.2. TupleID Propagation:

It is a method of transferring information among different relations by virtually joins them. This method exhibit to search in the relational database and which is observed that less costly than physical joins in both time and space. It is the technique for performing virtual joins among relations which less expansive then physical joins. When we want to search a good predicate then we will use propagate Tuple Ids between any two relations which provides the less computations and storage cost compared with creating join requirements

[11].During this study we found there are 2 cases that such propagates could be counterproductive.

#### Drawbacks:

Propagates via large fan-outs

- Propagate via long weak links
- During this study we found the 2 most challenges these are as below:

**Challenge-1:**Finding the Useful Links :

#### Solution for Problem-I:

It is necessary that to find out the techniques that can estimate the usefulness of links across tables and then use the most useful links to achieve the better data mining task.

#### Challenge-II:

➤ Transferring information efficiently :

#### Solution for Problem-II:

We must create the strategies with as low inter database communication cost as possible

#### 5.4.3 Multi view Learning

The multi-view learning problem with n views can be seen as n inter-dependent relations and are thus applicable to multi-relational learning. This is the basic scenario in multi relational learning problem The Multi View Classification (MVC) approach employs the multi-view learning framework to operate directly on multi-relational databases with conventional data extracting methods. The approach works as follows

- Information Propagation Stage:
- Aggregation Stage:
- Multiple Views Construction Stage:
- View Validation Stage:
- View Combination Stage:

### 6. ANALYSIS OF MRDM APPROACHES

The below available comparative analysis shows (Figure-I) some strong points of Multi View Learning (MVL) compare to other approaches. We have compared with the three most important factors.

**Factor –I:** The relational database is able to keep its compact representation and normalized structure.

**Factor-II:** The second is that it uses a framework that is able to directly incorporate any traditional single-table data mining algorithm.

**Factor-III:** The multi-view learning framework is

extremely efficient for Mining relational databases in term of running time

### 7. K-MEANS ALGORITHM

This section reports the primary k-means algorithm. The idea is to categorize a given set of data into k number of disjoint clusters, where the value of k is nonmoving in progress. The algorithm correspond of two separate phases: first phase is to define k centroids, one for each cluster. Then next phase is to take each point be to the given data set and interact it to the nearest centroid. Euclidean distance is considered to ascertain the distance between data points and the centroids. When all the points are included in identical clusters, the first phase is completed and an early classify is done. At this point we need to recalculate the new centroids, as the inclusion of new points may leads to change in the cluster centroids. Once we determined k new centroids and new binding is to be created between the identical data points and the nearest new centroid, generating a loop. As a result of that loop, the k centroids may change their place in a stage by stage manner and situation will be reached where the centroids do not move more. This signifies the convergence criterion for clustering. The k-means algorithm is a popular method for automatically grouping vector based data. It is used in various applications such as density estimation and vector quantization, workload behavior characterization and automatic topic identification and image compression,, among many others. It is been identified as one of the top- 10 algorithms in data mining [6]. Pseudo code for k-means clustering algorithm is listed as

**Algorithm 1** [3].

Basic K-Means Algorithm:[4]

step1: Decide K points as initial centroids  
 step2: reproduce  
 step3: Depute every point to its closest centroid to form K clusters  
 step4: Recompute the centroid of every cluster  
 step5: until halt condition is met

To find out a distance from a centroid c to the given point x an Euclidean metric is used and although other metrics are also possible:

$$d(x; c) = \sqrt{[(x_i - c_i)^2]}$$

For algorithm halt few termination conditions that may be used

- Maximum number of reproduce is arrive at.
  - Centroids converged to minima,
- i.e. there is no further change in centroids positions.

#### 7.1 RESTRICTS OF K-MEANS ALGORITHM

K-Means Algorithm [16] suffers from two major restricts

1. K-Means algorithm is computationally very expensive as it involves several distance determinations of each data point from all the centroids in each reproduce.
2. The final cluster results heavily depend on the selection of initial centroids which causes it to converge at local best.

The first restrict has been solved using the enhanced K-means algorithm in our paper at an international conference on emerging trends in management and engineering .[4],[5] The second restrict states that the results of k-means clustering vary along the initial centres chosen . So, the solution of this problem is to determine a way to select the initial centres. This paper proposes the algorithm to chose initial centres. To apply k-means algorithm to certain clustering problem, first of all we must find out the initial centroids and the number of initial centroids is equal to the number of desired clusters (k) in which the given input data is to be partitioned. The number k is determined by the user i.e. this is an input parameter.

The primary k-means algorithms as initial centroids are taken just randomly out of the input data set. But the random choose of initial centroids leads the computation of the algorithm into local optima. Say, k is determined to be 3.If from a given data set we select first 3 points as initial cent roids and compute the k-means algorithm. Next time suppose we select the last 3 points as the initial centroids and further third time let we select any 3 arbitrary data points as initial centroids and compute the k-means algorithm. Every time the end clustering results will come out to be different.we have to analyze which one is the most assign result. Thus with the random selection of initial centroids there is no guarantee that the k-means algorithm will converge into best results. The limitation which needs to be dealt with in order to make the k-means algorithm more efficient.This can be shown with an example dataset of eight employees giving different clustering results with different initial centres.

### 8. DISCUSSION OF THE RESULTS

Let us now come back to our research questions as a first

Employees	Attribute1:X (experience in number of years)	Attribute2 :Y (salary in Lacs/annum)
Employee1	0.5	2
Employee2	1	3
Employee3	2	3.5
Employee4	3	4
Employee5	3.5	5
Employee6	4	6
Employee7	4.5	7
Employee8	5	7.5

observation, we can say that mining more than two relational database procedure so difficult and complexity to write sql query. To overcome this problem we can introduce link analysis extension for mining relational databases. Second, mining more than four relational databases using sql server as

Employees	Attribute1:X (experience in number of years)	Attribute2 :Y (salary in Lacs/annum)	Group
Employee1	0.5	2	1
Employee2	1	3	2
Employee3	2	3.5	2
Employee4	3	4	2
Employee5	3.5	5	3
Employee6	4	6	3
Employee7	4.5	7	3
Employee8	5	7.5	3

backend tool and Microsoft visual studio as front end tool. Microsoft visual studio is Computer-based techniques used in recognize, digging-out, and examining ' challenge' business data, such as sales and revenue by products or departments or

Employees	Attribute1:X (experience in number of years)	Attribute2 :Y (salary in Lacs/annum)	Group
Employee1	0.5	2	1
Employee2	1	3	1
Employee3	2	3.5	2
Employee4	3	4	2
Employee5	3.5	5	2
Employee6	4	6	3
Employee7	4.5	7	3
Employee8	5	7.5	3

associated incomes and cost. Objectives of a Business Intelligence exercise include and Understanding of a firm's internal and external strengths and Understanding of the relationship and weaknesses between different data for better decision making. Detection of opportunities for creation. Cost reduction and optimal deployment of resources and See also competitive intelligence [1].

**Table 1** Employee dataset

Now, after applying K-means clustering algorithm on this data set for k=3 and initial centres as c1 (0.5, 2) ,c2 (1,3) , c3(2, 3.5) the result is

**Table 2** clustering results with first set of initial centres

Now, when we take different initial centres with same k=3 and initial centres as c1(0.5,2),c2(3,4) and c3(5,7.5) the results are

**Table 3** second set of initial centres for clustering results

Thus various initial centres give different results and solve this problem in the algorithm to find initial centers is proposed.

### EXPANDING K-MEANS WITH IMPROVED INITIAL CENTER USING MID-POINT METHOD

In this algorithm a systematic method to find out the initial centroids is explained. This method is quite efficient to produce good clusters using k-mean algorithm and compared to taking the initial centroids randomly. The algorithm is as under:

**Algorithm1:**

Input: D:= Set of n data points , K:= call for number of clusters

Output: k number of initial centroids

Steps:

1. In this algorithm the given data set D, if the data points contain both the positive and negative attribute values then goto step 2, else goto step 4.
2. Find out the minimum attribute value in the given dataset D.
3. For each data point attribute, less with the minimum attribute value.
4. For each data point calculate the distance from origin.
5. Sort the distances get in step 4. Sort the data points in accordance with the distances.
6. Partition the sorted data points into k equal sets.
7. In each set , take the middle point as the initial centroid.

In the above algorithm, if the input data set contains the contradict value attributes and then all the attributes are transformed to positive space by contradict each data point attribute with the minimum attribute value in the data set. This transformation is required because in the algorithm the distance from origin to each data point is calculated. So if there are both positive and contradict values, then for different data points same Euclidean distance will be obtained which will result in incorrect selection of initial centroids. Then after transforming all attribute values to positive, next step is to calculate the distance of each point from the origin. The original data points are sorted into k equal sets and each set, the mid-point is calculated. All the mid-points are taken as the initial centroids.

**Sql server:**

Structure query language server is data base which design relational databases. They are creating multi relational data bases Store the records or tuples .we can create multiple relational databases such as personal relational database, school relational database, college relational database, job relational database, marriage relational database. Here create links between multiple relational databases. once link is established between multiple relational databases then mining relational databases is very easy and efficiency . Mining relational databases based on link analysis extension.

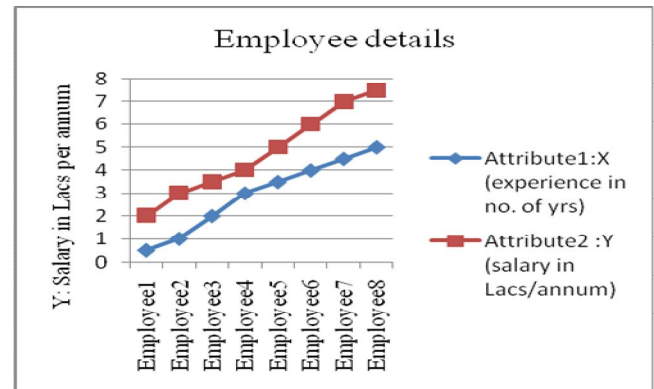


Figure 2 shows employees details with varying experience in no.of yrs and salary in lacs per annum

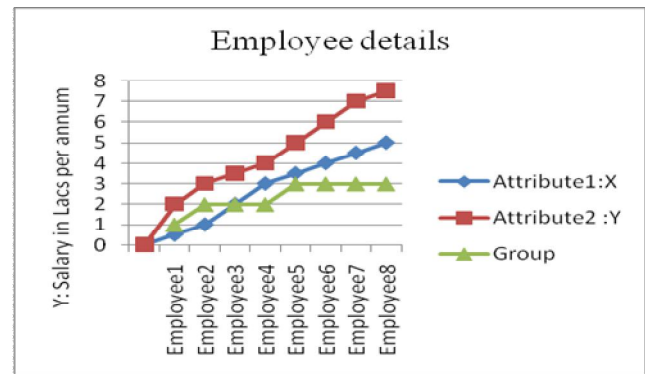


Figure 3 shows employees details with varying experience in no.of yrs and salary in lacs per annum and group

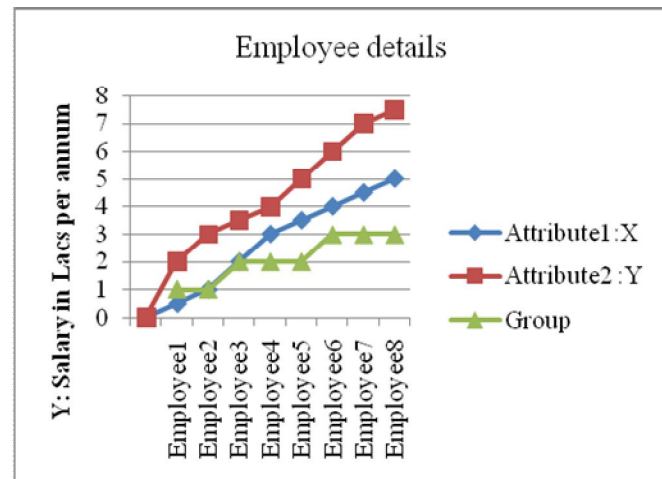


Figure 4 shows employees details with varying experience in no.of yrs and salary in lacs per annum and group

**CONCLUSION**

This work introduced a link-analysis-based technique allowing to analyze relationships existing in relational databases. The database is viewed as a graph and where the nodes match to the elements contained in the table and the think match to the relations between the tables.



## FURTHER WORK

Further work will be devoted to the application of this methodology to fuzzy SQL queries or fuzzy information restoration. The objective is to retrieve not only the elements strictly complying with the constraints of the SQL query, but also the elements almost comply with these constraints and are therefore close to the target elements. We will also evaluate the proposed methodology on real relational databases. My idea because of using business intelligence tool. We can increase efficiency and reducing time complexity for mining relational databases. Business Intelligence is Computer-based techniques used in make out, digging-out, and analyze challenging business data, such as sales and revenue by products, associated incomes, costs and departments,. Understanding of a firm's internal and external strengths and Understanding of the relationship between various data for better decision making and weaknesses. Detection of opportunities for creation, cost reduction and optimal deployment of resources. See also competitive intelligence [1]. Business data can be any type of data related to a business or organization but is usually sales data, customer records, transaction history, employee records or any kind of stored data which can be analyzed . [2].

## REFERENCES

- [1] Amir Ben-Dor, Ron Shamir and Zohar Yakini, "Clustering Gene Expression Patterns," Journal of Computational Biology, 6(3/4): 281-297, 1999
- [2] Chaturvedi J. C. A, Green P, "K-modes clustering," J. Classification, (18):35-55, 2001.
- [3] Tapas Kanungo, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu. An efficient kmeans
- [4] Kirti Aggarwal, Neha Aggarwal, Sunita Bhardwaj, Nikita Taneja: An Effective Enhanced k-mean Clustering Algorithm for Data Mining, international cofrence on emerging trends in engineering and management,"in press". clustering algorithm: Analysis and implementation. IEEE Trans. Pattern Anal. Mach. Intell., 24(7):881-892, 2002
- [5] Kirti Agarwal, Neha Aggarwal, Kanika Gupta: Comparative Analysis of k-means and Enhanced K-means clustering algorithm for data mining, intrnational journal of compuer applications,"unpublished"
- [6] Xindong Wu, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus F. M. Ng, Bing Liu, Philip S. Yu, Zhi-Hua Zhou, Michael Steinbach, David J. Hand, and Dan Steinberg. Top 10 algorithms in data mining. Knowl. Inf. Syst., 14(1):1-37, 2008.
- [7] Kirti Agarwal, Neha Aggarwal, Kanika Gupta: Comparative Analysis of k-means and Enhanced K-means clustering algorithm for data mining, intrnational journal of computer applications, "unpublished".

- [8] Neelamadhab, Rasmita Data warehousing and OAPL,MRDM technology In the decision support system in the21st century" , VSRD Technical Journal-VSRD-JCSIT, VOL.2(3),2012,2010-222
- [9] Dr.Pragnyaban Mishra, Neelamadhab Padhy, Rasmita Panigrahi "CIIT International Journal of Data Mining and Knowledge Engineering", Vol.4 No.5, May-2012.
- [10] A. Clare, H. E. Williams, and N. Lester. Scalable multi-relational association mining. In *ICDM*, 2004
- [11] S. Dzeroski and N. L. editors. *Relational Data Mining*. Springer, 2001
- [12] de Raedt, L. (Ed.) *Advances in Inductive Logic Programming*, IOS Press, 1996
- [13] Wang, K., Liu, H. *Discovering structural association of semi structured data*, IEEE Trans. Knowledge and Data Engineering (TKDE2000), 12(3):353-371, 2000
- [14] Arno Jan Knobbe, Multi Relational Data Mining Thesis - [www.kiminkii.com/thesis.pdf](http://www.kiminkii.com/thesis.pdf)
- [15] N.Lavrac and S.Dzeroski. *Inductive Logic Programming Techniques and Applications*. Ellis Horwood, 1994.
- [16] S.Muggleton .Inverse entailment and prolog. New Generation computing, Special Issue on Inductive Logic Programming, 3:245-286,1995
- [17] Neha Aggarwal and Kirti Aggarwal A Mid – Point based k-mean Clustering Algorithm for Data mining Vol. 4 No. 06 June 2012

## AUTHORS PROFILE.

<sup>1</sup>G.P.NARASIAH, received the MCA degree in 2011 from JNTU Anantapur. He is pursuing his M.Tech, final year (computer Science and Engineering) degree from JNTUA College of Engineering Pulivendula, Kadapa, and Andhra Pradesh, India. His research interests include Data Mining and clustering.



<sup>2</sup>Mrs. S. Jessica Saritha, received her B. Tech (ECE) from JNTU Anantapur in 2000. She received her M.Tech from JNTU College of Engineering, Kakinada in 2005 and pursuing Ph. D in JNTU Hyderabad. She has total of 8 years of experience in teaching and currently

Working as Assistant Professor at JNTUA College of Engineering, Pulivendula, Y.S.R. Dist., A.P, and India. Her areas of interest are data mining, and clustering.