Diffused Kernel DMMI Approach for Theoretic Clustering using Data Mining

Ankita G. Joshi, Prof. R. R Shelke

Computer Science and Engineering Department, Sant Gadge Baba Amravati University, India, ankitajoshi_10@yahoo.co.in, 7719832327

Abstract— Data mining is the process of finding anomalies, patterns and correlations within large data sets to predict outcomes. Data mining and knowledge discovery in databases (KDD) are treated as synonyms. Knowledge discovery in databases (KDD) is a research area that considers the analysis of large databases in order to identify valid, useful, meaningful, unknown, and unexpected relationships. The main objective of the data mining process is to extract information from a large data set and transform it into an understandable structure for further use. Clustering is a main task of exploratory data analysis and data mining applications. Theoretic Clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). Various algorithms based on simple data structures that can be captured until first and second order statistics are studied. Centralizing the whole data at one source node is not possible so distributed clustering algorithm are in great demand. Distributed clustering is to explore the hidden structure of the data collected/stored in geographically distributed nodes. We incorporate an information theoretic measure into the cost function of the distributed clustering since information theoretic measures take the whole distribution of cluster data into account for better clustering results. We interpret the motivation for choosing the MMI (Maximum Mutual Information) criterion to develop distributed clustering algorithms. The proposed Diffused Kernel DMMI algorithms can achieve excellent clustering results on both text and numeric data. Our proposed system generates optimized clusters in less duration and capable of removing empty clusters.

Keywords — Data Mining, Knowledge Discovery in Database, Clustering, Theoretic clustering, Information Theory, Divergence, Mutual Information.

INTRODUCTION

The volume of data produced is doubling every two years. Unstructured data alone makes up 90 percent of the digital universe. But more information does not necessarily mean more knowledge. Data mining allows us to sift through all the chaotic and repetitive noise, understand what is relevant and then make good use of that information to assess likely outcomes. Although a user often has a vague understanding of his data and their meaning and can usually formulate hypotheses and guess dependencies, user rarely knows: where to find the “interesting” or “relevant” pieces of information, whether these pieces of information support his hypotheses and models, whether (other) interesting phenomena are hidden in the data, which methods are best suited to find the needed pieces of information in a fast and reliable way, and how the data can be translated into human notions that are appropriate for the context in which they are needed. In reply to these challenges a new area of research has emerged, which has been named “knowledge discovery in databases” or “data mining”.

Data Mining, also popularly known as Knowledge Discovery in Databases or KDD for short, refers the nontrivial extraction of implicit, valid, novel, previously unknown and potentially useful and ultimately understandable information from data in databases. While data mining and knowledge discovery in databases (KDD) are frequently treated as synonyms, data mining is actually core part of the knowledge discovery process.

The unifying goal of the KDD process is to extract knowledge from data in the context of large databases. This valuable information can help the decision maker to make accurate future decisions. KDD applications deliver measurable benefits, including reduced cost of doing business, enhanced profitability, and improved quality of service. Therefore Knowledge Discovery in Databases has become one of the most active and exciting research areas in the database community [1].

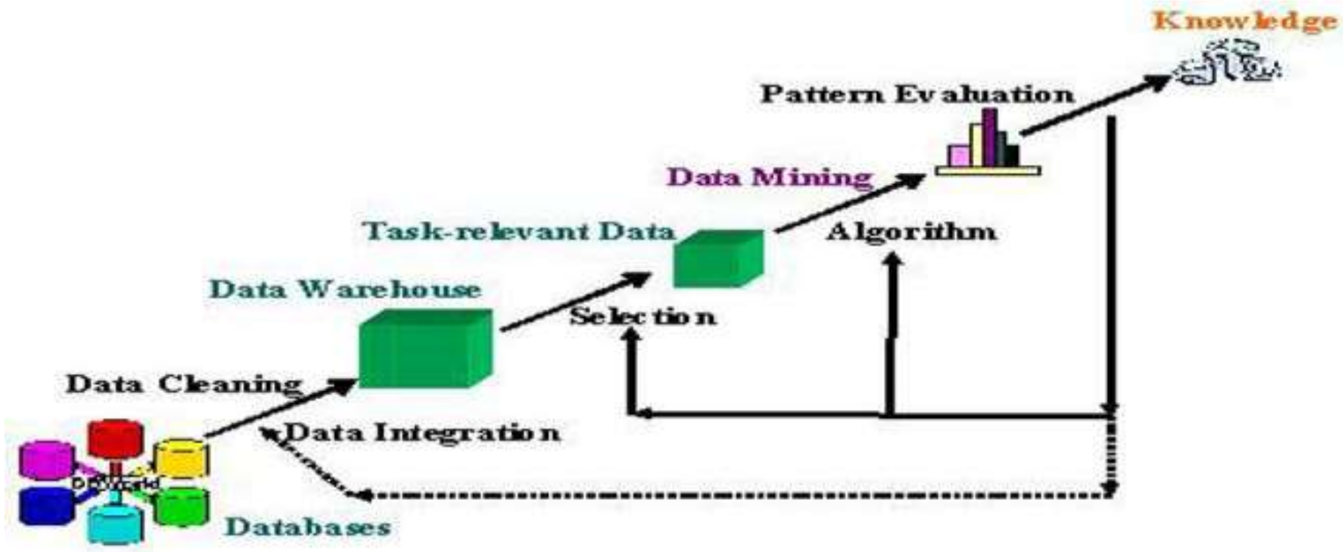

Figure 1: Data mining is 4^{th} core step in Knowledge Discovery in databases

The **Knowledge Discovery in Databases (KDD) process** comprises of a few steps leading from raw data collections to some form of new knowledge. The iterative process consists of the following steps:

- **Creating a target data set:**
  Selecting a data set, or focusing on a subset of variables, or data samples, on which discovery is to be performed.

- **Data cleaning & Pre-processing**:
  It also known as data cleansing, it is a phase in which noise data, missing data and irrelevant data are removed from the collection. This is a very important preprocessing step because your outcome would be dependent on the quality of selected data**.**

- **Data integration**:
  At this stage, multiple data sources, often heterogeneous, may be combined in a common source.

- **Data selection**:
  Data mining is done on your current or past records. Thus, you should select a data set or subset of data, in other words data samples, on which you need to perform data analysis and get useful knowledge. At this step, the data relevant to the analysis is decided on and retrieved from the data collection.

- **Data transformation**:
  It is also known as data consolidation, it is a phase in which the selected data is transformed into forms appropriate for the mining procedure. Using Transformation methods, the number of effective variables is reduced and only useful features are selected to depict data more efficiently based on the goal of the task. In short, data is transformed into appropriate form making it ready for data mining step.

- **Data mining process**:
  It is the crucial step in which clever techniques are applied to extract patterns potentially useful or Selecting method(s) to be used for searching for patterns in the data. Deciding which models and parameters may be appropriate. Matching a particular data mining method with the overall criteria of the KDD process.

- **Pattern evaluation**:
  This is a post processing step in KDD which interprets mined patterns and relationships. If the pattern evaluated is

not useful, then the process might again start from any of the previous steps, thus making KDD an iterative process. In this step, strictly interesting patterns representing knowledge are identified based on given measures.

➢ Knowledge representation

It is the final phase in which the discovered knowledge is visually represented to the user. This essential step uses visualization techniques to help users understand and interpret the data mining results. It is common to combine some of these steps together. For instance, data cleaning and data integration can be performed together as a pre-processing phase to generate a data warehouse. Data selection and data transformation can also be combined where the consolidation of the data is the result of the selection, or, as for the case of data warehouses, the selection is done on transformed data. The KDD is an iterative process. Once the discovered knowledge is presented to the user, the evaluation measures can be enhanced, the mining can be further refined, new data can be selected or further transformed, or new data sources can be integrated, in order to get different, more appropriate results.

Information theory provides a general framework to establish clustering criteria. With information theoretic measures (e.g. divergence and mutual information), data structure can be captured beyond the first and the second order statistics, by taking the whole probability distribution function (pdf) of cluster data into consideration. Divergence to measure the 'distance' between distributions of data belonging to different clusters. For a clustering result, large divergence means there are obvious differences or boundaries between data items belonging to different clusters. Hence, their goal is to maximize the divergence, by adjusting the assignment of cluster/class label on each data item. In this kind of method, calculating divergence relies on unknown *conditional pdfs of cluster data*, , which need to be estimated during clustering. Mutual information can be used to measure the information shared by data items and cluster labels. Large mutual information means that the structure information contained in data items is well preserved by the clustering result. Hence, MMI-based clustering algorithms seek the clustering result that maximizes the mutual information.

❖ Association Rule Mining

Association rules are one of the major techniques of data mining. Association rules are if/then statements that help uncover relationships between seemingly unrelated data in a relational database or other information repository. An example of an association rule would be "If a customer buys a dozen eggs, he is 80% likely to also purchase milk." Association rule mining finding frequent patterns, associations, correlations, or causal structures among sets of items or objects in transaction databases, relational databases, and other information repositories [5]. An association rule has two parts, an antecedent (if) and a consequent (then). An antecedent is an item found in the data. A consequent is an item that is found in combination with the antecedent.

Association rules are created by analyzing data for frequent if/then patterns and using the criteria *support* and *confidence* to identify the most important relationships. *Support* is an indication of how frequently the items appear in the database. *Confidence* indicates the number of times the if/then statements have been found to be true.

In data mining, association rules are useful for analyzing and predicting customer behavior. They play an important part in shopping basket data analysis, product clustering, and catalog design and store layout.

Examples of areas in which association rules have been used include:

- Credit card transactions: items purchased by credit card give insight into other products the customer is likely to purchase.
- Supermarket purchases: common combinations of products can be used to inform product placement on supermarket shelves.
- Telecommunication product purchases: commonly associated options (call waiting, caller display, etc) help determine how to structure product bundles which maximize revenue
- Banking services: the patterns of services used by retail customers are used to identify other services they may wish to purchase.
- Insurance claims: unusual combinations of insurance claims can be a sign of fraud.
- Medical patient histories: certain combinations of conditions can indicate increased risk of various complications.

❖ Clustering

Clustering is a division of data into groups of similar objects. Each group, called a cluster, consists of objects that are similar to one another and dissimilar to objects of other groups. It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics [4].

What distinguishes clustering from classification is that clustering does not rely on predefined classes. In clustering, there are no predefined classes. the records are grouped together on the basis of self similarity.

Clustering plays an important role in a broad range of applications, from information retrieval to CRM. Such applications usually deal with large datasets and many attributes[5]. Exploration of such data is a subject of data mining. A "clustering" is essentially a set of such clusters, usually containing all objects in the data set. Additionally, it may specify the relationship of the clusters to each other, for example a hierarchy of clusters embedded in each other[18].
`
Clustering can be roughly distinguished as:

- Hard clustering: each object belongs to a cluster or not
- Soft clustering: each object belongs to each cluster to a certain degree (e.g. a likelihood of belonging to the cluster)

The main requirements that a clustering algorithm should satisfy are:

- scalability;
- dealing with different types of attributes;
- discovering clusters with arbitrary shape;
- minimal requirements for domain knowledge to determine input parameters;
- ability to deal with noise and outliers;
- insensitivity to order of input records;
- high dimensionality;
- Interpretability and usability.

❖ Theoretic Clustering

Theoretic clustering is to partition a data set into "clusters" of data points that are "close" to each other but relatively "far from" other data points. Theoretic clustering is to explore the hidden structure of data and group data items into a few clusters in an unsupervised way (Unsupervised learning is to discover unknown structures that exist within a data set). Theoretic Clustering can be considered the most important *unsupervised learning* problem; so, as every other problem of this kind, it deals with finding a *structure* in a collection of unlabeled data.

A simple definition of theoretic clustering could be "the process of organizing objects into groups whose members are similar in some way". It is the organization of a collection of patterns (usually represented as a vector of measurements, or a point in a multidimensional space) into clusters based on similarity [2]. Intuitively, patterns within a valid cluster are more similar to each other than they are to a pattern belonging to a different cluster. Figure 2 shows a simple graphical example of clustering:

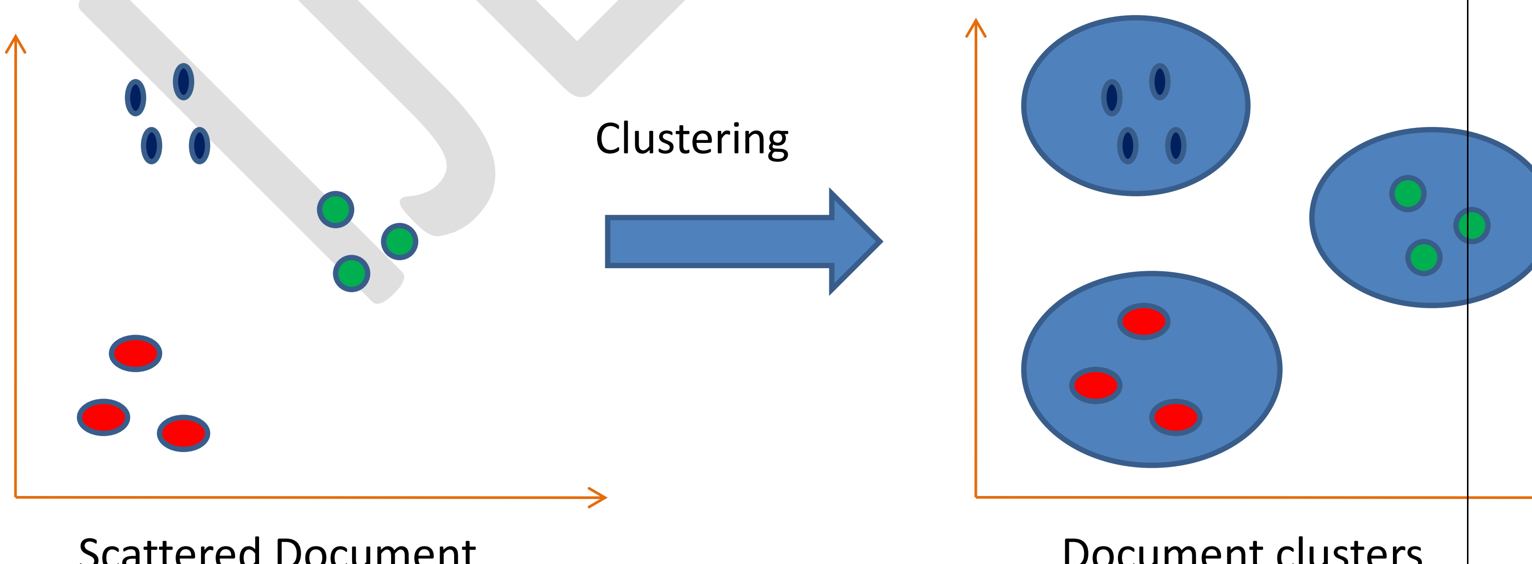

Figure 2: An example of clustering

In this case we easily identify the 3 clusters into which the data can be divided; the similarity criterion is *distance*: two or more objects belong to the same cluster if they are "close" according to a given distance (in this case geometrical distance). This is called *distance-based clustering [6], [11]*.

Theoretic Clustering is the process of making a group of abstract objects into classes of similar objects. It can be described as:

- A cluster of data objects can be treated as one group.
- While doing theoretic clustering, we first partition the set of data into groups based on data similarity and then assign the labels to the groups.
- The main advantage of theoretic clustering over classification is that, it is adaptable to changes and helps single out useful features that distinguish different groups.

❖ Theoretic Clustering Based Approach

Theoretic clustering approach applicable to data mining mostly belongs to unsupervised classification in general. Clustering is a division of data into groups of similar objects. Clustering is unsupervised learning of a hidden data concept. Data mining deals with large databases that impose on clustering additional severe computational requirements. These challenges led to the emergence of applicable data mining clustering methods.

❖ K-Means based Clustering Algorithm

One of the most common iterative algorithms is the K-means algorithm [2], [10], [15] broadly used for its simplicity of implementation and convergence speed.

K-mean algorithm creates clusters by determining a central mean for each cluster.

- Input: n objects (or points) and a number k
- Algorithm

1. Randomly place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.
2. Assign each object to the group that has the closest centroid.
3. When all objects have been assigned, recalculate the positions of the K centroids.
4. Repeat Steps 2 and 3 until the stopping criteria is met.

❖ Gaussian Mixture Model based Algorithm

A Gaussian mixture model is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. One can think of mixture models as generalizing k-means clustering to incorporate information about the variance structure of the data as well as the centers of the latent Gaussians. Gaussian is the probability given in a mixture of K Gaussians.

A Gaussian Mixture Model (GMM) is a parametric probability density function represented as a weighted sum of Gaussian component densities. GMMs are commonly used as a parametric model of the probability distribution of continuous measurements or features in a biometric system, such as vocal-tract related spectral features in a speaker recognition system. GMM parameters are estimated from training data using the iterative Expectation-Maximization (EM) algorithm.

GMMs are often used in biometric systems, most notably in speaker recognition systems, due to their capability of representing a large class of sample distributions. One of the powerful attributes of the GMM is its ability to form smooth approximations to arbitrarily shaped densities [14], [16].

❖ Divergence and Maximum Mutual Information based Clustering

For the divergence-based clustering, there are roughly two types of algorithms, which are the parametric type and the nonparametric type, respectively. The Bregman soft clustering algorithm is a representative and typical sample for the former [19], [24]. In [19], the authors model the data source with a mixture of exponential family distributions (one component for one cluster), and pose the clustering problem as a parameter estimation problem for the mixture model. They find the correspondence between exponential families and regular Bregman divergences, and thereby bring up a Bregman divergence viewpoint for learning the maximum likelihood parameters of the mixture model.

The algorithm provides a framework for clustering different datasets by using different Bregman divergences (or equivalently, parametric models of different exponential distributions). For a given application (dataset), to obtain good clustering performance, it is expected to artificially choose a specific Bregman divergence (or equivalently, parametric model of a specific exponential distribution) which matches the generative model of current data. However, the prior knowledge for the generative models of real datasets can be lacking, which makes it hard to choose an appropriate parametric model.

As for mutual information, in the context of clustering, it can be used to measure the information shared by data items and cluster labels. In more detail, it measures the uncertainty about cluster labels reduced by knowing the data items, or the uncertainty about data items reduced by knowing the corresponding cluster labels. Large mutual information means that the structure information contained in data items is well preserved by the clustering result. Hence, MMI-based clustering algorithms seek the clustering result that maximizes the mutual information [24].

❖ Linear and Kernel Distributed Maximum Mutual Information (DMMI) algorithm

Linear DMMI works accordingly to its name, is appreciated for linearly separable problems. Linearly separable problem can be understood with the help of a simple example. Suppose that there is a hyperplane (which splits your input space into two half-spaces) such that all points of the first class are in one half-space and those of the second class are in the other half-space[16], [17], [18]. In two dimensions, that means that there is a line which separates points of one class from points of the other class. In three dimensions, it means that there is a plane which separates points of one class from points of the other class. In higher dimensions, it's similar: there must exist a hyperplane which separates the two sets of points. For example, in this image, if blue circles represent points from one class and red circles represent points from the other class, then these points are linearly separable.

✓ Architecture of Proposed System

One of the challenges in data mining is how to extract important information from huge customer databases and cluster them effectively, in order to gain competitive advantage and user satisfaction.

Proposed system uses a diffused kernel DMMI based approach with association rule mining with improved similarity measures for clustering datasets from huge database; association rule mining is used to find association between various items present in the database.

List of optimized clusters of data help to improve business communication and can assist users to gain useful, valid and interesting information for their further requirements. It is helpful to examine the user clustering behavior and assists in increasing the need for getting potentially useful information.

Association rule mining identifies the remarkable association or relationship between a large set of data items. With huge quantity of data constantly being obtained and stored in databases, several industries are becoming concerned in mining association rules from their databases.

For example, the detection of interesting association relationships between large quantities of analysis, and various business decision making processes. A typical example of association rule mining is market basket analysis. This method examines customer buying patterns by identifying associations among various items that customers place in their shopping baskets. Following architecture shows steps for clustering datasets:

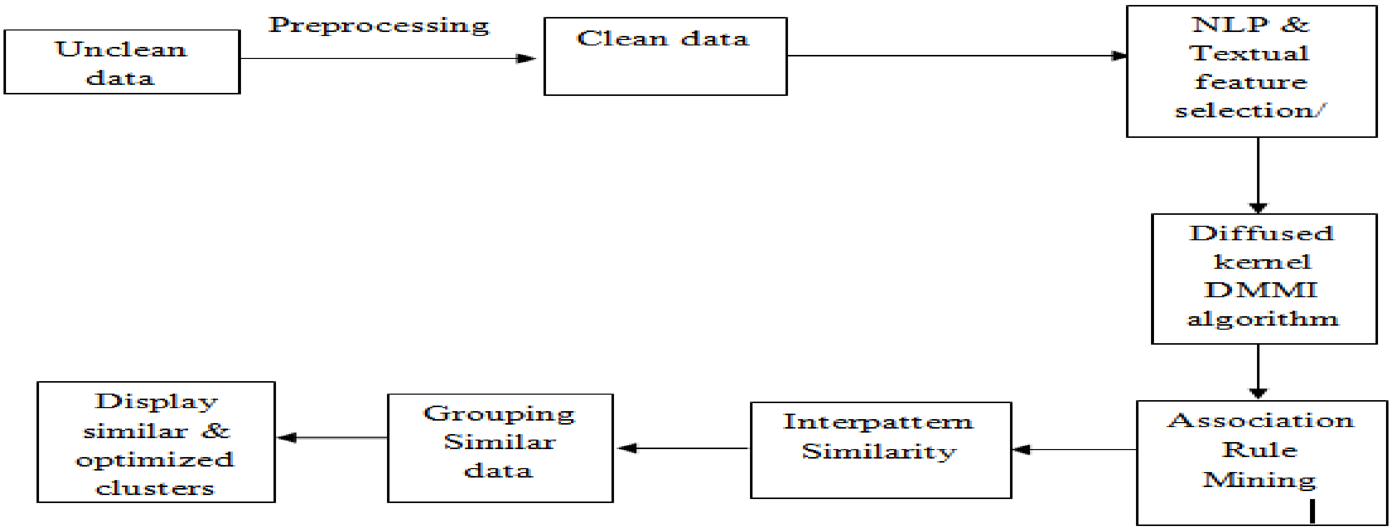

Figure 3: Architecture of Proposed System Diagram

This architecture shows steps to achieve approach to effectively obtain optimized clusters for given data. The main objective of the research is to develop a strategy which helps user using mining association between various datasets. The given data is clustered efficiently for users for their further use. The proposed scheme is used for theoretic clustering using data mining. To achieve the research objective successfully, a series of sequence progresses and analysis steps have been adopted. Figure 3 depicts the methodologies to extract similar datasets from given data.

Unclean Data

Unclean data refers to data that contains erroneous information. It is incomplete or erroneous data, especially in a computer system or database. In reference to databases, this is data that contain errors. Unclean data can contain such mistakes as spelling or punctuation errors, incorrect data associated with a field, incomplete or outdated data. The complete removal of dirty data from a source is impractical or virtually impossible.

Preprocessing

Data cleansing, data cleaning or data scrubbing is the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database. Used mainly in databases, the term refers to identifying incomplete, incorrect, inaccurate, irrelevant, etc. parts of the data and then replacing, modifying, or deleting this dirty data.

Clean Data

After cleansing, a data set will be consistent with other similar data sets in the system. The inconsistencies detected or removed may have been originally caused by user entry errors, by corruption in transmission or storage.

Natural language processing and Textual Feature Selection/ Extraction

Natural language processing (NLP) is a field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human (natural) languages. As such, NLP is related to the area of human–computer interaction. Many challenges in NLP involve natural language understanding, that is, enabling computers to derive meaning from human or natural language input. Identifying the most effective subset of the textual features to use in clustering and transformations of the input features to produce new salient features.

Diffused Kernel DMMI algorithm

Diffused kernel DMMI algorithm is proposed to handle complicated structures of cluster data by using modified kernel discriminative clustering function:

p(K| x; W) proportional to exp $(\sum_{h=1}^{D} ak,h\ G(xh,x) +\ bk)$

Where G (.,.) is positive kernel function which evaluates the inner product of two vectors in a high-dimensional space. Weight coefficient is $\{a_{k,h}\}$ and bias coefficient is $\{b_k\}$.

.X_h belongs to R^D is a D-dimensional base vector. The set of base vectors $\{X_h\}$ is constrained to be the same for all nodes. By this modification, the weight coefficients $\{a_{k,h,j}\}$of different nodes share the common acting-objects and thus could be directly fused among neighboring nodes. The authors suggested several feasible approaches to design the base vectors, including the grid-based design and the random design. In the former approach, grid points in the value range of data are chosen as the base vectors, while in the latter approach, base vectors are randomly sampled from the value range of data. Both of the two methods are suitable for the kernel DMMI. The appropriate value of depends on specific problems. Intuitively, it increases with the complexity of between-cluster boundaries and the dimension of data.

Proposed algorithm is capable to generate optimized clusters in less time, remove empty clusters formed by less matching between datasets, data structures are captured by measures beyond the first and the second order statistics(i.e first and second most minimum value in given data) which are used for feature extraction[25].

Association Rule Mining

Association is one of the best known data mining technique. In association, a pattern is discovered based on a relationship between items in the same transaction. That is the reason why association technique is also known as relation technique.

The association technique is used in market basket analysis to identify a set of products that customers frequently purchase together. Retailers are using association technique to research customer's buying habits. Based on historical sale data, retailers might find out that customers always buy crisps when they buy beers, and therefore they can put beers and crisps next to each other to save time for customer and increase sales.

Interpattern Similarity

Interpattern similarity measured by a distance function defined on pairs of patterns. In a wide sense, it measures the score to which a pair of objects is alike. Distance function is the Euclidean distance between clusters for given data which is automatically calculated when user want to find theoretic scores for given data. The Euclidean distance between point's p and q is the length of the line segment connecting them ($\overline{\mathbf{pq}}$).

In Cartesian coordinates, if **p** = (p_1, p_2,..., p_n) and **q** = (q_1, q_2,..., q_n) are two points in Euclidean *n*-space, then the distance (d) from **p** to **q**, or from **q** to **p** is given by the Pythagorean formula:

$$d(p,q) = d(q,p) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \ldots + (q_n - p_n)^2}$$

$$= \sqrt{\sum_{i=1}^{n}(qi - pi)^2}$$

The ranking model's purpose is to rank, i.e. produce a permutation of items in new, unseen lists in a way which is "similar" to rankings in the training data in some sense. Ranking is performed to find support value for given data. This support value is carry forwarded to scoring mechanism to find similarity between given data.

Scoring mechanism:

Once a model has been created by a data mining application, the model can then be used to make predictions for new data. The process of using the model is distinct from the process that creates the model. Typically, a model is used multiple times after it is created to score different databases. For example, consider a model that has been created to predict the probability that a customer will purchase something from a catalog if it is sent to them. The model would be built by using historical data from customers and prospects that were sent catalogs, as well as information about what they bought (if anything) from the catalogs. During the model-building process, the data mining application would use information about the existing customers to build and validate the model. In the end, the result is a model that would take details about the customer (or prospects) as inputs and generate a number between 0 and 1 as the output.

After a model has been created based on historical data, it can then be applied to new data in order to make predictions about unseen behavior. This is what data mining (and more generally, predictive modeling) is all about. The process of using a model to make predictions about behavior that has yet to happen is called "scoring." The output of the model, the prediction, is called a score. Scores can take just about any form, from numbers to strings to entire data structures, but the most common scores are numbers (for example, the probability of responding to a particular promotional offer). The scoring mechanism is used to get the max distance, support and confidence values.

Grouping Similar Data

The similar data gained after interpattern similarity process is grouped together in same cluster. Similarly grouped datasets or each group, called a cluster, consists of objects that are similar to one another and dissimilar to objects of other groups.

Display similar and Optimized Clusters

It displays list of clusters which are valid, useful and similar in some context. For user convenience, optimized clusters are generated which means more data is segregated in clusters.

Large amounts of data have been collected routinely in the course of day-to-day management in business. Such data is primarily used for accounting and for management of the customer's database. Typically, management data sets are very large and constantly growing and contain a large number of complex features. While these data sets reflect properties of the managed subjects and relations, and are thus potentially of some use to their owner, they often have relatively low information density. One requires robust, simple and computationally efficient tools to extract information from such data sets. The development and understanding of such tools is the core business of data mining. These tools are based on ideas from computer science, mathematics and statistics. Mining useful information and helpful knowledge from these large databases has thus evolved into an important research area [23], [25].

Consider a shopping website with a large collection of items. Typical business decisions that the management of the shopping website has to make include what to put on sale, how to design coupons, etc. Analysis of past transaction data is a commonly used approach in order to improve the quality of such decisions.

- ✓ **Proposed work include following modules:**

Module 1: Collection of Data Sets.

In this module, various datasets will be collected which will contain documents for theoretic clustering. This datasets will be used for the evaluation of the project.

Module 2: Implementation of Diffusion DMMI algorithm.

In this module, the diffusion kernel DMMI algorithm will be implemented, which will use the distributed maximum mutual information and used this for the cost function in distributed clustering. The accuracy of algorithm will be evaluated on all the datasets.

Module 3: Study of Various Data Mining Techniques for Clustering.

In this module, various data mining techniques will be studied and results will be checked from the base papers of the techniques.

Module 4: Development of the Desired/ Best Data Mining Techniques for Clustering.

In this module, the best technique for data mining will be implemented for the desired clustering algorithm and result will be checked for future evaluation.

Module 5: Integration of Diffusion Kernel DMMI and Data Mining.

In this module, the Data mining technique will be combined with diffusion kernel DMMI technique, in order to improve the clustering output.

Module 6: System Integration and Optimization.

In this module, the results will be evaluated, compared and optimized if required.

- ✓ **Proposed Algorithm for Theoretic Clustering**

Input: - Data from database

Output: - List of similar & optimized clusters

Step 1: Enter number of clusters

Step 2: Specify the data (Specify data which user want to cluster)

Step 3: Perform Natural language processing

Step 4: Find similarity between data
a. Perform ranking (Support value is calculated)
b. Find theoretic scores (for rules)
c. Perform theoretic clustering

Step 5: Group data with similar values of similarity together

Step 6: Find mean of the similar values to get final optimized clusters
(Association rules are applied in step 5 & step 6)

ACKNOWLEDGMENT

The making of the proposed system needed co-operation and guidance of a number of people. I therefore consider it my prime duty to thank all those who had helped me through their venture. It is my immense pleasure to express my gratitude to **Prof. R. R. Shelke** as guide who provided me constructive and positive feedback during the preparation of this seminar. I express my sincere thank to the head of department **Prof. A. B. Raut** and all other staff members of CSE department for their kind co-operation.

I would like to thank **Dr. A. B. Marathe**, Principal of our institution for providing necessary facility during the period of working on this report.

I am thankful to my friends and library staff members whose encouragement and suggestion helped me to complete my seminar.

I am also thankful to my parents whose best wishes are always with me.

CONCLUSION

The proposed system discussed an improved approach for theoretic clustering in order to generate optimized clusters from database in which data mining with association rule mining is used. Moreover user can extract valid, potentially useful and meaningful results using diffused kernel DMMI algorithm and association rule mining for mining association between different datasets in the database. We have considered the Maximum Mutual Information (MMI) criterion in the context of distributed data clustering, leading to more satisfactory clustering results for datasets with complicated data structures. The proposed improved Diffused Kernel DMMI algorithms generates optimized clusters for users in less duration and maintain good clustering performance in the cases beyond first & second order statistics, rates/ samples and removes empty clusters from which further reflects the semantic meaning, optimality, flexibility and applicability of the algorithms for practical cases.

REFERENCES:

[1] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, eds., *"Advances in Knowledge Discovery and Data Mining", in MIT Press*, Menlo Park, CA, USA 1996.

[2] N. Slonim, G. S. Atwal, G. Tkacik, and W. Bialek, "Information based clustering," *Proc. Nat. Acad. Sci. USA*, vol. 105, no. 51, pp. 18297–18302, 2005.

[3] A. K. Jain, "Data clustering: 50 years beyond K-means," *Patter Recogn. Lett.*, vol. 31, no. 8, pp. 651–666, 2010.

[4] A. K. Jain, A. Topchy, M. H. C. Law, and J. M. Buhmann, "Landscape of clustering algorithms," in *Proc. Int. Conf. Pattern Recogniy*, 2004, vol. 1, pp. 260–263.

[5] R. Xu and D. Wunsch-II, "Survey of clustering algorithms," *IEEE Trans. Neural Netw.*, vol. 16, no. 3, pp. 645–678, May 2005.

[6] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Comput. Surveys*, vol. 31, no. 3, pp. 265–323, 1999.

[7] J.Grabmeier and A. Rudolph, “Techniques of cluster algorithms in data mining,” *Data Mining and Knowl. Discov.*, vol. 6, no. 4, pp. 303–360, 2002.

[8] P. A. Forero, A. Cano, and G. B. Giannakis, “Distributed clustering using wireless sensor networks,” *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 4, pp. 707–724, Aug. 2011.

[9] I. Dhillon and D. Modha, “A data-clustering algorithm on distributed memory multiprocessors,” *Large-Scale Parallel Data Mining, Lecture Notes in Artif. Intell.*, pp. 245-260, 2000.

[10] P. A. Forero, A. Cano, and G. B. Giannakis, “Consensus-based k-means algorithm for distributed learning using wireless sensor networks,” in *Proc. Workshop Sens., Signal, Inf. Process.*, Sedona, AZ, USA, May 11–14, 2008.

[11] C. Bohm, C. Faloutsos, J. Y. Pan, and C. Plant, “Robust information theoretic clustering,” in *Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*,2005, vol. 17, no. 12, pp. 65–75.

[12] C. G. Lopes and A. H. Sayed, “Distributed processing over adaptive networks,” in *Proc. Adapt. Sens. Array Process. Workshop*, Lexington, MA, USA, Jun. 2006.

[13] M. Klusch, S. Lodi, and G. Moro, “Distributed clustering based on sampling local density estimates,” in *Proc. Int. Joint Conf. Artif. Intell.*, 2003, pp. 485–490.

[14] D. Gu, “Distributed EM algorithm for Gaussian mixtures in sensor networks,” *IEEE Trans. Neural Netw.*, vol. 19, no. 7, pp. 1154–1166, Jul. 2008.

[15] S. Datta, C. R. Giannella, and H. Kargupta, “Approximate distributed k-means clustering over a peer-to-peer network,” *IEEE Trans. Knowl.Data Eng., vol. 21, no. 10, pp. 1372–1388, Oct. 2009.*

[16] M. Klusch, S. Lodi, and G. Moro, “Distributed clustering based on sampling local density estimates,” in *Proc. Int. Joint Conf. Artif. Intell.*, 2003, pp. 485–490.

[17] H. Kargupta, W. Huang, S. Krishnamoorthy, and E. Johnson, “Distributed clustering using collective principal component analysis,” *Knowl. Inf. Syst. J.*, vol. 3, no. 4, pp. 422–448, 2001.

[18] S. Bandyopadhyay, C. Giannella, U. Maulik, H. Kargupta, K. Liu, and S. Datta, *Inf.* “Clustering distributed data streams in peer-to-peer environments,” *Sci.*, vol. 176, no. 14, pp. 1952–1985, 2006.

[19] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh, “Clustering with Bregman divergences,” *J. Mach. Learn. Res.*, vol. 6, pp. 1705–1749, 2005.

[20] J. C. Principe, *Information Theoretic Learning: Renyi’s Entropy and Kernel Perspectives*. New York, NY, USA: Springer-Verlag, 2010.

[21] E. Januzaj, H. P. Kriegel, and P. Martin, “Scalable density-based distributed *clustering*,” in *Proc. Knowl. Discov. Databases: PKDD 2004*, 2004, vol. 19, no. 7, pp. 231–244.

[22] T. M. Cover and J.A. Thomas, *Elements of Information Theory. New* York, NY, USA, *Wiley*, 1991.

[23] C. Li, P. Shen, Y. Liu, and Z. Zhang, “Diffusion information theoretic learning for distributed estimation over network,” *IEEE Trans. Signal Process., vol. 61, no. 16, pp. 4011–4024, Aug 15, 2013.*

[24] A. Krause, P. Perona, and R. G. Gomes, “Discriminative clustering by regularize information maximization,” in *Proc. Adv. Neural Inf. Process. Syst., 2010, pp. 775 – 783.*

[25] S. Merugu and J. Ghosh, “A privacy-sensitive approach to distributed clustering,” *Pattern Recognit. Lett., vol. 26, no. 4, pp. 399–410, 2005*