ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

Spectral Cluster Analysis for Efficient Mining of Overlapped Multi Dimensional Data

[I]Dr R.S. Vetrivel, [II]B. Rizwan, [III]Dr S.G.Shrinivas
[I,III]Professor, Computer Science, Subramanya College of Arts & Science, Palani
[II]Research Scholar, Computer Science, Subramanya College of Arts & Science, Palani

Abstract

In World Wide Web (WWW), data mining software is a one of the most important tool for analyzing data. Data mining is developed to discover and examines the reliable data from large data sets in data base. Data mining is the process of extracting the possible information which determines the significant pattern and rules.
The existing system was developed to resolve the problem of clustering with high-dimensional data. The Global Hubness-Proportional K-Means (GHPKM) method mainly focuses on selecting accurate cluster prototypes for identifying approximate hyper spherical clusters. The major hubs are utilized as cluster prototypes and guides for centroid-based cluster configurations while the searching of data in web. The GHPKM method was established to be more robust than k-Means++ baseline on both synthetic and real-world data.
However, Hubness of the clustering objects is greater than predefined threshold which exhibit too low hubness to clustering elements.
To overcome these problems, we propose a technique as spectral clustering analysis for mining of overlapped multi dimensional data. With the help of spectral clustering analysis, the similar data are identified through the similarity matrix. From the similar data, high density overlapped data are discovered by parametric density measure. The K-means algorithm is performed to organize the nearest data which helps to mining the overlapped multidimensional data from data base.

Keywords

Data mining, Multi Dimensional Data, Spectral Cluster Analysis, Overlapped data point, K-means algorithm

I. Introduction

A. Data Mining
The main objective of the data mining is the detecting correlations or patterns between numbers of fields in huge relational databases. In data mining, unusual data records are discovered by identification of patterns and trends in data collected using classification and clustering rules.
Data mining process is employed for identifying the data from various perspectives and summarizing it into valuable data. It permits to users for evaluating data from many different dimensions or angles and categorize it and to identify the summarized data.
The useful information which gets from data mining is used to increase revenue, cuts costs, or both. Data mining is mostly exploits by more companies for builds a strong consumer focus - retail, financial, communication, and marketing organizations. It aids to companies to decide relationships among interior factors such as price, product positioning, or staff skills, and exterior factors such as economic indicators, competition, and customer demographics.

B. Multi Dimensional Data
The multi dimensional data analysis is one of the types of data analysis technique that optimized for data warehouse. Multi dimensional data analysis is developed to groups the data into categories as data dimensions and measurements.
Handling of high-dimensional data with high efficient is the one of the great challenging task in traditional data-mining techniques. Clustering of multi dimesional data becomes difficult because enhacement of data sparsity rising the difficulty in distinguishing distances between data points.

C. Clustering
Cluster analysis technique is developed for grouping a set of objects which objects in the same group are more similar to each other groups. For the summarization and enhanced understanding purposes, clustering analysis are developed in the mining techniques. Cluster analysis is employed to divides the data into valuable groups which is known as cluster.
The main objective of clustering in data mining, used for statistical data analysis in which includes the machine learning, pattern recognition, image analysis and information retrieval bioinformatics. Cluster analysis is an unsupervised process of grouping elements together. The groupings elements are assumed to the same cluster are more similar to each other than to the remaining data points in other groups.
According to the information found in the data describing of data (objects), cluster analysis groups the objects. The group of data (objects) is similar which is related to one another and different from the objects in other groups. Based on the degree of association between data and groups, clustering mechanism is introduced to generate groups automatically.
The clustering and classification mechanisms examines document collection as group which aids to analyzing other documents containing some of those same data. Clustering is compact with identifying a *structure* in a collection of unlabeled data. The clustering is the process of organizing objects into groups whose members are similar in some way. The main intension of clustering in the field of mining is helps to improve the efficiency and effectiveness of the results.

II. Literature Survey

1. Handling Multi-Dimensional Complex Queries in Key-Value Data Stores
In this paper [1], the author introduces a technique for handling multi-dimensional complex queries. To key-value data stores, the SPIKE, a sliced Pyramid-based index system was developed

in this paper. The highly accessible and scalable distributed key value data stores are established by SPIKE which bridges the gap between the data scale and querying functionality.

Complex queries are handled effectively by SPIKE because that is a well-organized two level multi dimensional index system. The author introduces a SP index for enhancing the original pyramid index through splitting the multi dimensional space much finer. It helps to allow the data stores to better support complex queries on large-scale datasets or datasets with less distinct values. The SP-Index is developed as a two-level index mechanism which contains sliced pyramid space partition index and a distributed B-Tree index. Based on SP-Index, SPIKE was implemented on Cassandra, which helps to perform multi-dimensional complex query processing effectively.

2. Improving classification rate constrained to imbalanced data between overlapped and non-overlapped regions by hybrid algorithms

The author proposes a technique as Soft Hybrid algorithm in this paper [2] for enhancing the performance of classification for highly imbalanced huge data sets with overlapped regions. It helps to resolve the problem in the new data imbalanced situation. The computational complexity in large data sets is reduced by this method.

The Soft hybrid algorithm was developed to discover the over lapped and non overlapped regions to resolve the new data imbalanced situation. The technique was divided into two main phases as

- Boundary region determination
- Response classification algorithm

The first phase data clustered as non overlapping data, overlapping data and borderline data. Based on statistical values from classification phase, the modified Kernel learning method, modified DBSCAN and RBF network were applied to organize the data into correct groups.

3. On Evolutionary Spectral Clustering

In evolutionary spectral clustering, the author develops a technique with two frameworks in this paper [3] which incorporate with temporal smoothness. The solutions to the evolutionary spectral clustering were developed to attain more constant and reliable clustering results. The results provided by this method which are short-term noises with less sensitive while at the same time are adaptive to long term cluster drifts.

The Evolutionary clustering is the fundamental technique for the important applications such as clustering dynamic Web and blog contents and clustering data streams. The excellent clustering results were produced by evolutionary clustering which fit the current data well while keeping out the deviation from history data..

4. Robust path – based spectral clustering

In this paper [4], based on Mestimation from robust statistics, the author introduces a robust path-based spectral clustering method. The robust path-based similarity measure is determined for performing efficient spectral clustering under both unsupervised and semi-supervised settings. This method was extensively more robust than spectral clustering and path-based clustering. Based on both synthetic and real-world data, robust path – based spectral clustering carry out the experiments. For the image segmentation experiments, color images are taken from the Berkeley segmentation data set and benchmark.

5. A Nonparametric Bayesian Model for Multiple Clustering with Overlapping Feature Views

In this paper [5] the author designs probabilistic nonparametric Bayesian model of clustering algorithms to offer a single clustering solution. Probabilistic nonparametric Bayesian model was introduce to identify the multiple clustering solutions from data. It also aids to find the feature subsets that are related for the clusters in each analysis. For different analysis purpose, the features are shared. It permits overlapping to the sets of relevant features.

The model feature relevance in each view was analyzed by using an Indian Buffet Process and the cluster membership was examined by using Chinese Restaurant Process. According to multiple partitioning problems, they offer an inference approach for learning latent parameters. The model not only established to learn the features and clusters in each view but also designed for learning the number of clusters, number of views and number of features in each view automatically.

6. Overlapping correlation clustering

The author presents a new approach to resolve the problem of overlapping clustering by local search algorithm in this paper [6]. The purpose of this paper to originate the overlapping clustering as an optimization problem in which every data point is mapped to a small set of labels. They represent the membership to different clusters.

While finding of mapping which distances between data points agree virtually possible with distances taken over their label sets. This approach considers two measures, for identifying the distances between label sets. i) set-intersection indicator function and ii) Jaccard coefficient.

7. Model based Overlapping Clustering

In this paper [7], the author proposes a basic approach to afford an alternative interpretation of the model as a generalization of mixture models, which makes it easily interpretable. The original model maximized likelihood over constant variance Gaussians was established it to effort among regular exponential family distribution, and related Bregman divergences.

It helps to assume the model applicable for a wide range of clustering distance functions. The general model is suitable to some domains, which contains high-dimensional sparse domains as text and recommender systems. The author applied some algorithmic modifications for enhancing both the performance and applicability of the model.

8. Noise Robust Spectral Clustering

In this paper [8], the author proposes a technique as spectral clustering algorithm to introduce the robustness against noise. Basis of regularization, the author develops a warping model to mining the data into a new space. Every point extends easily its spatial information to other points while warping of data.

After the warping, experimental studies are demonstrates that the clusters become reasonably compact and well separated. They comprises the noise cluster which established by the noise points. The numbers of clusters are determined with the help of eigen value analysis. The spectral mapping was applied to the data by achieving the low-dimensional data representation. Finally, the clustering was performed with the help of K-means algorithm. The Noise Robust Spectral Clustering has the following advantages.

- Due to the noise points are grouped into one new cluster, this technique develops robustness against the noise.
- This technique was used to automatically determine the number of clusters and the parameters of the algorithm.

9. On spectral clustering: analysis and an algorithm

The author introduces a new approach as simple spectral clustering algorithm in this paper [9]. This approach was executed by using a few lines of Matlab. The matrix perturbation theory was established in this method for examining the algorithm. On the number of challenging problems, this approach provides excellent experimental results successfully.

10. A random walks view of spectral segmentation

The author proposes a technique as clustering and segmentation through the pair wise similarities. In this paper [10], the similarities are described as edge ows in a Markov random walk and analysis the eigenvalue and eigenvectors of the walk's transition matrix. This technique demonstrates spectral methods for clustering and segmentation which includes probabilistic foundation. From the framework, we desired as Normalized Cut method arises logically. When the Normalized Cut algorithm is exact, the method achieves a complete characterization of the cases.

11. Weighted Overlap Dominance – A procedure for interactive selection on multidimensional interval data

In this paper [11], the author introduces an outranking procedure as Weighted Overlap Dominance procedure (WOD) to execute the selection of alternatives by multiple attributes with interval valued data. Through the determination of weights of the different attributes, parameters which are associates with risk attitude, weighted dominance, decision maker was established to direct the search for preferred alternatives. The weighted dominance relations estimated by volume based measures which makes a relation on the pairwise comparisons between optimistic and pessimistic weighted values.

12. A Space-Filling Multidimensional Visualization (SFMDVis) for Exploratory Data Analysis

In this paper [12], the author proposes a technique as Space-Filling Multidimensional Visualization (SFMDVis) for the multidimensional data analysis. This proposed technique was developed with horizontal lines which help to signify the multidimensional data items. Based on the color mapping in order to indicate the data item with its values, every line was divided into segments logically. The SFMDVis technique aids to resolving the visual clutter and over plotting problems.

13. Statistically-driven generation of multidimensional analytical schemas from linked data

In this paper [13], the author introduces an automatic generation of multidimensional analytical stars (MDAS) technique to generate the multidimensional (MD) patterns from Linked Data (LD) sources. Through this proposed method, MD conceptual patterns are discovered and summarize LD. These patterns are similar to the MD star schema distinctive of relational data warehousing.

14. Overlapping Slicing: A Novel Approach for Data Anonymization

In this paper [15], the author introduces a technique as overlapped slicing. For high dimensional data, this technique was implemented to attain better data utility while protecting against privacy threats by using l-diversity requirement. For attribute correlation, the author uses an efficient algorithm called as chi_matrix.

15. Automatic Sleep Staging using Multi-dimensional Feature Extraction and Multi-kernel Fuzzy Support Vector Machine

In this paper [14], the author proposes a work for Automatic Sleep Staging. This method established with the clinical Polysomnographic (PSG) data, all-night Electroencephalogram (EEG), Electrooculogram (EOG) and Electromyogram (EMG) signals of subjects, and assumes the American Academy of Sleep Medicine (AASM) clinical staging manual as standards to recognize automatic sleep staging.

To build vectors, the author designs a method with eighteen various features of EEG, EOG and EMG in time domains and frequency domains. The linear combination of weights and parameters of multiple kernels of the fuzzy support vector machine (FSVM) were learned and the multi-kernel FSVM (MK-FSVM) was constructed through the adoption of sleep samples self-learning.

III. Methodology

In this proposed method, the overlapped multidimensional data are mined efficiently from web by using spectral cluster analysis technique. The main goal of cluster analysis is to discover fundamental structures of data and to organize them into significant subgroups for future study and analysis purpose.

In the existing system, Global Hubness-Proportional K-Means (GHPKM) method was established with hub-based algorithms for high-dimensional data. The GHPKM method was designed for hubs which used to approximate local data centers is not only a feasible option, but also frequently leads to improvement over the centroid-based approach.

The hubness is a good measure of point centrality i.e., the tendency of high-dimensional data to contain points (hubs) which frequently occur in k-nearest neighbor lists of other points are successfully exploited in clustering. The using of hubs as cluster prototypes and points guiding by centroid-based search is a promising new idea in clustering high-dimensional and noisy data. The global hubness estimates are generally to be preferred with respect to the local ones.

However, Global Hubness-Proportional K-Means (GHPKM) method has some disadvantages. The Hubness of the clustering objects greater than predefined threshold. GHPKM method was failed to handle clustering of overlapped multi dimensional data.

Due to the some disadvantages in the existing method, we introduce a Spectral Cluster Analysis technique in this proposed paper. The Spectral Cluster Analysis (SCA –OMDDM) technique is employed for mining the overlapped multi dimensional data. SCA –OMDDM technique is helps to handle the multi dimensional data more effectively than the existing method. The efficiency of overlapped data mining is improved by this proposed technique. By using of standard K-means cluster algorithm , the data (object) are classified which aids to identify the nearest data successfully.

From the figure 3.1, we analysis the SCA –OMDDM technique. The spectral cluster analysis is designed in this proposed work for divides the data into clusters. The spectral cluster analysis is introduced to discovered the similar data from the multi dimensional database.The similar data are identified with the help

of formation of similarity matrix by using graph.
Based on the parametric density measure,the overlapped data point is used to find outs the high density overlapped data.The high density data are classified by the K-means algorithm.The classification assists to K-means algorithm which extract the nearest multi dimensional data.The spectral cluster analysis is developed with K-means algorithm for mining the multidimensional data.

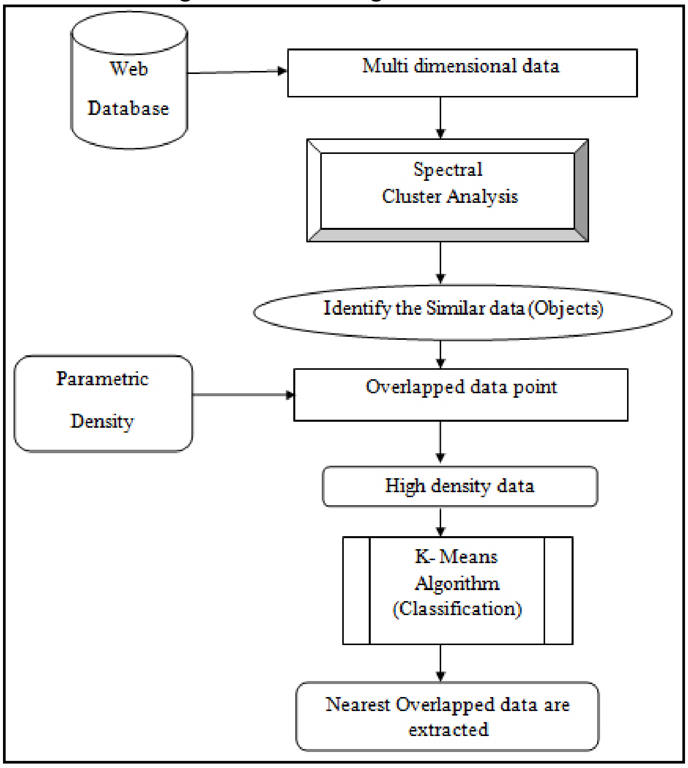

Fig. 3.1 : Architecure of spectral cluster analysis for overlapped multi dimensional data mining.

A. Modules

a) Spectral cluster analysis
b) Overlapped data point
c) Data classification using K-means algorithm

B. Modules Description

1. Spectral cluster analysis

Spectral cluster analysis is the simple method to implement and reasonably fast especially for sparse data sets up to several thousands. Spectral clustering estimates the data clustering as a graph partitioning without make any assumption on the form of the data clusters. In spectral clustering the data points are considered as nodes of the connected graph. Based on Nodes are grouped (clustered) by partitioning the graph.
Spectral cluster analysis is divided into three categories

- Pre-processing,
- Spectral representation,
- Clustering.

In the Pre-processing step, graph is created with partition. From the graph, the similarity matrix is determined which represents the dataset. Spectral representation is used to form the associated Laplacian matrix and also compute eigen values and eigenvectors of the Laplacian matrix. Clustering assign points to two or more classes, based on the new representation. The spectral cluster analysis is used to cluster the similar data from the database with the help of forming the similarity matrix. We assume the data points as $(x_1,x_2,\ldots x_n)$. The similarity S_{ij} is determined by comparing the x_i and x_j from the matrix.
$S_{ij}=1$ when two points are similar
$S_{ij}=0$ when two points are dissimilar
From the partition of the graph , we analysis the data as the edges between different groups have very low weights which means that points in different clusters are dissimilar from each other and the edges within a group have high weights which means that points within the same cluster are similar to each other.

2. Overlapped Data Point

The Parametric density is the estimation which is needed for finding the overlapped in data points. The density based process is an essential method for clustering data while grouping the irregular data, noise and outliers are present in the network. The detect mode of the overlapped points are established by inspection of multi dimensional density estimates.Based on the density values, the overlapped data are classified as high density data and low density data.
The High density data construct methods for visualization of multivariate functions to amplify usefulness of overlapped density data points. It's dealing with large data sets cluster hubness is extremely useful to partition the data set into subsets of similar observations and each subset is a cluster .
The structure of overlapped data set is reflected by its density function Density estimator is computed based on grid partitioning in multidimensional data space. It histogram methods a multidimensional data space is quantized into finite number of cells that form a grid structure.

3. Data classification using K-means algorithm

Spectral cluster analysis uses the K means algorithm for grouping the related data points. The K- means clustering is the most widely used method for cluster analysis in data mining. It aids to analysis of data which each data belongs to the cluster with the nearest mean.
The K-means clustering is developed with spectral clustering to divide the objects as data (n) into clusters (k). From the graph, we determines the details as if two points are close then Dij=1 and when two points are far apart then Dij=0. The K-means clustering graph identifies the every node as data which is only connected to its nearest neighbors.
After achieving clustering, the classification also performed to discover the overlapped data from network. Here the hierarchical type of clustering is used to find the data items effectively which are sequentially linked.
The hierarchical clustering is successfully performed in the field of large data sets. The Hierarchical based clustering is particularly useful for multi dimensional data in the web search. Finally the multidimensional data are discovered efficiently by using spectral clustering analysis with K-means algorithm.

Input: Database 'D', Number of regions 'R'
Output: Extract the overlapped multi dimensional data
Step 1: **Begin** Step 2: Multidimensional data taken from the database Step 3: Spectral cluster analysis is used to groups the similar data into cluster through the similarity matrix Step 4: Based on parametric density, the high density data are identified from the clustering of similar data Step 5: K-means algorithm is perform classification to find out the nearest overlapped data Step 6: the overlapped multi dimensional data are extracted Step 7: End

Fig. 3.2 : Algorithm for mining of overlapped multi dimensional data by using Spectral clustering analysis

IV. Experimental Evaluation

In this paper worked on performance evaluation in terms of Noise level, Cluster object size, Accuracy. The performance metric to evaluates and analyze the value in java environment simulations. This technique improves the effectiveness of mining of overlapped multi dimensional data. The performance measures of the proposed work are analyzed with following metrics:

- Noise level
- Cluster object size
- Accuracy

1. Noise level

Noise is a random fluctuation in an electrical signal, a characteristic of all electronic circuits. Noise generated by electronic devices varies greatly, as it is produced by several different effects. Noise level is measured in terms of percentage (%).

Table 4.1: Tabulation for Noise level

Number of multi dimensional data	Noise level (%)	
	Global Hubness-Proportional K-Means (GHPKM)	Spectral Cluster Analysis for Overlapped Multi Dimensional Data Mining (SCA -OMDDM)
4	48	35
8	54	41
12	62	49
16	78	56
20	86	63

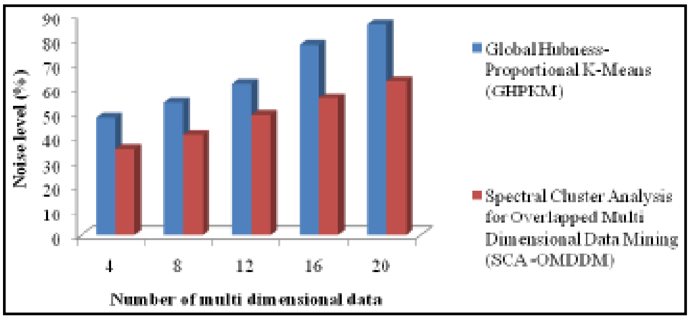

Fig. 4.1: Measure of Noise level

Figure 4.1 demonstrates the Noise level. The multi dimensional data taken for the experimental consideration is varied from 4 to 20. From the figure X axis represents the Number of multi dimensional data whereas Y axis denotes Noise level using SCA -OMDDM technique. From the figure it is clearly evident that the proposed SCA -OMDDM technique reduces the Noise level than the Existing ERST method. Hence, the Noise level reduced to 25% by the proposed SCA -OMDDM technique than the existing GHPKM method.

2. Cluster object size

The cluster object size is measured as the size of the clustering of similar data (multi dimensional data) from data by using the SCA -OMDDM technique. The cluster object level is the important value of clustering hubness based incrimination level. The cluster object size is measured in terms of Kilobytes (Kb).

Table 4.2: Tabulation for Cluster object size

Number of multi dimensional data	Cluster object size (Kb)	
	Global Hubness-Proportional K-Means (GHPKM)	Spectral Cluster Analysis for Overlapped Multi Dimensional Data Mining (SCA -OMDDM)
4	50	34
8	55	39
12	60	42
16	66	56
20	72	60

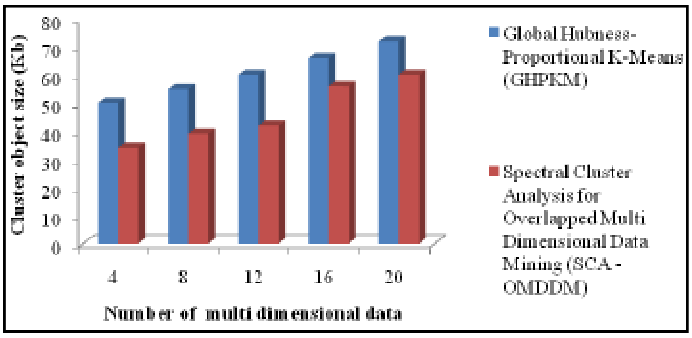

Fig. 4.2: Measure of Cluster object size

Figure 4.2 demonstrates the Cluster object size. The multi dimensional data taken for the experimental consideration is varied from 4 to 20. From the figure X axis represents the Number of multi dimensional data whereas Y axis denotes Cluster object size using SCA -OMDDM technique. From the figure it is clearly evident that the proposed SCA -OMDDM technique reduces the Cluster object size than the Existing ERST method. Hence, the Cluster object size reduced to 25% by the proposed SCA -OMDDM technique than the existing GHPKM method.

3. Accuracy

The accuracy of mining data is defined as the ratio of the number of multi dimensional data are taken from the database to the number of extracted overlapped multi dimensional data by using SCA -OMDDM technique.

The accuracy of mining the data is measured in terms of percentage (%).

Table 4.2: Tabulation for Cluster object size

Number of multidimensional data	Accuracy (%)	
	Global Hubness-Proportional K-Means (GHPKM)	Spectral Cluster Analysis for Overlapped Multi Dimensional Data Mining (SCA -OMDDM)
4	45	50
8	49	54
12	53	60
16	58	64
20	62	77

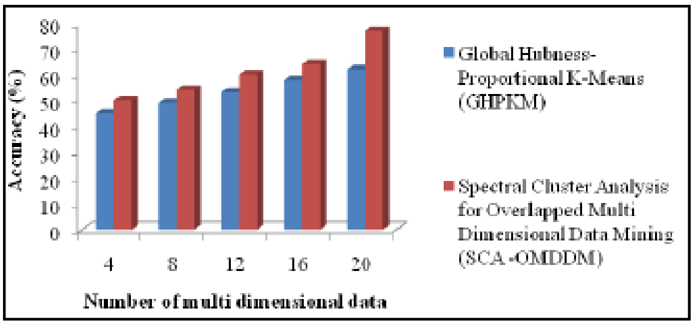

Fig. 4.3: Measure of Accuracy

Figure 4.3 demonstrates the Accuracy of mining. The multi dimensional data taken for the experimental consideration is varied from 4 to 20. From the figure X axis represents the Number of multi dimensional data whereas Y axis denotes Accuracy using SCA -OMDDM technique. From the figure it is clearly evident that the proposed SCA -OMDDM technique improves the Accuracy than the Existing ERST method. Hence, the Accuracy of mining is increased up to 14% by the proposed SCA -OMDDM technique than the existing GHPKM method.

V. Conclusion & Future Scope

1. Conclusion

In this paper, we have described in detail architecture for efficient mining of overlapped multi dimensional data. In this architecture, Spectral cluster analysis is organizing the overlapped multi dimensional data are collected from any number of sources, formatted according to the data points obscured using the K-means algorithm methods. Finally, we applied our analysis results to the design of a Spectral cluster analysis for efficient mining and apply the best design parameter settings in java environment. We implemented the proposed scheme, and conducted comprehensive performance analysis and evaluation, which showed its efficiency and advantages over existing schemes.

The future work enhances the important of hubness by using formatted according to the data points obscured using the kernel methods.

References

[1] Yu Tang, Hailong Sun, Qi Wang, Xudong Liu "Handling Multi-Dimensional Complex Queries in Key-Value Data Stores" ELSEVIER, Information Systems, Volume 66, June 2017, Pages 82–96

[2] Piyanoot Vorraboot, Suwanna Rasmequan, Krisana Chinnasarn, Chidchanok Lursinsap, "Improving classification rate constrained to imbalanced data between overlapped and non-overlapped regions by hybrid algorithms", ELSEVIER, Neurocomputing, Volume 152, 2015, Pages 429-443

[3] Yun chi, Xiaodan song, Dengyong zhou, Koji hino, Belle L.Tseng "On Evolutionary Spectral Clustering", ACM Transactions on Knowledge Discovery from Data (TKDD), Volume 3,Issue 4,November 2009, Article No.17

[4] Hong Chang, Dit-Yan Yeungb, " Robust path-based spectral clustering", ELSEVIER, Pattern Recognition ,Volume 41, Issue 1, January 2008, Pages 191–203

[5] Donglin Niu, Jennifer G. Dy, Zoubin Ghahramani, "A Nonparametric Bayesian Model for Multiple Clustering with Overlapping Feature Views", JMLR W&CP 22 , 2012, Pages 814-822

[6] Francesco Bonchi, Aristides Gionis, Antti Ukkonen "Overlapping correlation clustering",IEEE 11th International Conference on Data Mining, 2011, Pages: 51 - 60

[7] Arindam Banerjee Chase Krumpelman, Joydeep Ghosh, Sugato Basu ,Raymond J. Mooney, "Model based Overlapping Clustering", Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, pp. 532--537, August 2005

[8] Zhenguo Li, Jianzhuang Liu, Shifeng Chen, and Xiaoou Tang, "Noise Robust Spectral Clustering", Oct. 2007

[9] Andrew Y. Ng, Michael I. Jordan, and Yair Weiss "On spectral clustering: analysis and an algorithm", Advances in Neural Information Processing Systems 14, pages 849–856, 2002.

[10] Marina Meila, Jianbo Shi, "A Random Walks View of Spectral Segmentation", 2001

[11] Jens Leth Hougaard, Kurt Nielsen, "Weighted Overlap Dominance – A procedure for interactive selection on multidimensional interval data", ELSEVIER, Applied Mathematical Modelling 35 (2011, Pages 3958–3969

[12] Tze-Haw Huang, Mao Lin Huang, Quang Vinh Nguyen, Laiping Zhao, "A Space-Filling Multidimensional Visualization (SFMDVis) for Exploratory Data Analysis" ELSEVIER,Information Sciences,Volume 390, June 2017, Pages 32–53

[13] Victoria Nebot, Rafael Berlanga , "Statistically-driven generation of multidimensional analytical schemas from linked data"ELSEVIER, Knowledge-Base d Systems 0 0 0 (2016), Pages 1–15

[14] Yanjun Zhang, Xiangmin Zhang, Wenhui Liu, Yuxi Luo, Enjia Yu, Keju Zou1 and Xiaoliang Liu, "Automatic Sleep Staging using Multi-dimensional Feature Extraction and Multi-kernel Fuzzy Support Vector Machine", Journal of Healthcare Engineering ,Vol. 5 · No. 4 · 2014 Page 505–520

[15] K.Sirisha, , Mr.J.Nagamuneiah, "Overlapping Slicing: A Novel Approach for Data Anonymization", International Journal of Engineering Research & Technology (IJERT),Vol. 2 Issue 9, September - 2013