ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

An Enhanced Spectral Clustering for Overlapping Data in Multiple Task Clustering

R. Renukadevi[1], Dr. S. Meenakshi[2]

Research Scholar, Department of Computer Science, Gobi Arts & Science College, Gobichettipalayam, Tamilnadu, India[1]

Associate Professor in Computer Science, Gobi Arts & Science College, Gobichettipalayam, Tamilnadu, India[2]

ABSTRACT: Clustering is one of the most widely used approaches for exploratory data analysis in data mining. In large scale data sources, multitask clustering is an important research work to handle overlapping data, negative and non-negative values among clustering of multiple tasks which is used to improve the learning relationship among related tasks and sharing of information across the tasks. Recent past spectral clustering has become the well accepted clustering algorithm to perform multitask clustering which relies on the eigenstructure of a similarity matrix and outperforms partitioning of data with more complicated structures than traditional clustering algorithms such as the K-means and Fuzzy C-means. This research work proposes an enhanced multitask spectral clustering (EMTSC) technique to perform multitask clustering without overlapping in data points among clustering of multiple tasks in large data sets for improving clustering performance.

KEYWORDS: Clustering, multitask clustering, data overlapping, spectral clustering algorithm, clustering performance.

I. INTRODUCTION

Clustering is one of the most widely used approaches for exploratory data analysis in data mining. Clustering aims at grouping the similar patterns and discovering the meaningful patterns of the data. Many clustering algorithms have developed which include K-means clustering, C-means clustering and Fuzzy C-means clustering and these traditional algorithms are depending upon the knowledge or acquisition of similarity information to relate data items to each other [2]. The traditional clustering algorithms are used to make cluster only by the single-task approaches [10].

Due to the rapid growth of data, the various challenges have posed on clustering are, partitioning the high dimensional data into different clusters and clustering of multiple tasks. In order to achieve the better clustering performance, recent past spectral clustering has become the well accepted clustering algorithm which relies on the eigenstructure of a similarity matrix and outperforms partitioning of high dimensional data with complicated structures than traditional clustering algorithms such as the K-means and Fuzzy C-means [13]. Spectral clustering is a promising approach to perform multiple task clustering [5]. Spectral clustering is simple to implement and outperforms efficiently for large data sets [14]. Spectral clustering is a powerful technique that has found increasing support and application in many areas. The concept, algorithms and application areas of spectral clustering are reviewed [9].

In large scale data sources, multitask clustering is an important research work to handle overlapping data, negative and non-negative values among clustering of multiple tasks which is used to improve the learning relationship among related tasks and sharing of information across the tasks. In multitask clustering, the multiple related prediction tasks are considered simultaneously and the same tasks belong to different clusters are allowed to overlap. Thus the development of an effective multitask spectral clustering algorithm is an essential research work to handle data overlapping in multiple task clusters. This research work proposes an enhanced multitask spectral clustering (EMTSC) technique to perform multitask clustering without overlapping in data points among clustering of multiple tasks in large data sets for improving clustering performance. The proposed EMTSC technique uses the Euclidean method to calculate the distance between two data points to find the similarity of data points are lie in a high dimensional space for eliminating the overlapping data in clustering of multiple tasks. The proposed system improves the clustering of

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

multiple tasks in large scale data sources with several advantages such as less complexity of partitioning of data, high accuracy and less time to perform clustering in multiple tasks.
The remaining structure of this paper is organized as follows. Section II describes the literature survey. Section III describes the proposed methodology. Section IV presents the experimental results and Section V concludes the research paper.

II. LITERATURE SURVEY

The development of an effective multitask spectral clustering algorithm is an important research work to handle various issues such as correlations, overlapping data, negative and non-negative values among clustering of multiple tasks. The various proposed multitask spectral clustering techniques in the literature are given below.
Xie, Lu and He [16] have proposed a multi-task co-clustering using non-negative matrix factorization (NMF) method for co-clustering both instance space and feature space simultaneously. In this work, the objective function which consists of two parts such as task-specific co-clustering for co-cluster the data individually and cross-task feature space regularization for learn and refine the relations of feature space among different tasks. In this work pairwise knowledge is transfer between task-specific feature spaces that is any two feature spaces is connected during the optimization procedure. The problem of learning multitasks with the assumption that tasks can be classified into various groups which are not known in prior is considered. The NMF method avoids the problem of negative information transfers among dissimilar and outlier tasks. The proposed work also captures the group specific similarities.
Tao Li and Chris Ding [7] have proposed a unique framework with the ability to cluster multiple related tasks based on the inter-task and intra-task relationships through geometric affine transformation of distance. The focus of this work is to perform multitasks via a three-factor non-negative matrix factorization. An intra-task orthogonal constraint is induced to a symmetric non-negative factorization form to acquire vectors that are close orthogonal within each task, which induces prior learning that a task must have independent clusters. The geometric transformation improves inter-task clustering without affecting the individual tasks acquired from other tasks. In this work the results have proved with several related tasks and different data distributions. This research work utilizes both task-specific and cross-task knowledge that significantly enhance the clustering performance.
Bonchi et al., [1] have developed a local search algorithm for solving the problem of overlapping clustering in which each data point is assigned a set of labels representing membership to different overlapping clusters. In this work, the mapping is finding by distances between data points agree with distances taken over their label sets and the distance between label sets are measured by using two methods such as set-intersection indicator function and Jaccard coefficient. In this work to solve the distance problems the greedy method is used to constrain the maximum number of cluster labels allowed, either globally or per data points. The authors apply the simple local-search algorithm for optimizes the labels of one object, given the labels of all other objects.
Yang et al., [17] have proposed an multitask spectral clustering algorithm (MTSC) for solving two types of correlations such as intertask correlation, which refers the relations among different clustering tasks and intratask learning correlation, which refers the process of learning cluster labels and learning mapping function to reinforce each other. This work proposes a novel $l_{2,p}$-norm regularizer to control the coherence of all tasks based on an assumption that related tasks should share a common low-dimensional representation. For each individual task an explicit mapping function is simultaneously learnt for predicting cluster labels by mapping features to the cluster label matrix. The MTSC algorithm performs clustering on multiple related tasks by exploring their intertask correlation and also learns a mapping function for predicting the cluster labels for the out-of-sample data. The advantage of MTSC algorithm is to improve the clustering performance on various real-world datasets and the limitations of this algorithm are overlapping that are occurred among clustering of multiple tasks.
Shang et al., [11] have proposed a global discriminative based non-negative spectral clustering algorithm for both normalized criterion and ratio cut criterion for detecting the geometrical information of the datasets and improving the cluster quality. This method is used to preserve the both global geometrical structure and global discriminative structure of datasets. This work has good ability to handle the out-of-sample data and the problem of proposed system is cannot achieve the good results on high dimensional datasets.
Lu, Fu and Shu [8] have proposed non-negative sparse spectral clustering algorithm to solve the problem of dimension reduction and part-based data representation. In this work the spectral clustering method using R-cut and N-cut as

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

objective functions for partitioning the graph. In this work the non-negative sparse spectral clustering is performed by imposing a sparse constraint on the cluster indicator matrix, and derives a multiplicative update rule to improve the clustering performance.

Yang et al., [18] have proposed the discriminative non-negative spectral clustering to solve the problems of grouping of out-of-sample data outside the training data, and to find the cluster indicator matrix under the discretization. The proposed method is used to explicitly impose an additional nonnegative constraint on the cluster indicator matrix to seek for a more interpretable solution and also learning an extra prediction function to grouping the out-of-sample data outside the training data to a large scale setting.

Nie et al., [2] have proposed the spectral embedded clustering (SEC) algorithm to solve the problems of minimize the normalized cut criterion in spectral clustering, control the mismatch between the cluster assignment matrix and low dimensional embedded representation of the data. In this work the cluster assignment matrix of high dimensional data can be represented by a low dimensional linear mapping of the high dimensional data and also discover the connection between SEC and other clustering methods. The proposed clustering algorithm gives outperform results for all high dimensional datasets when compared with other traditional clustering techniques.

Xiang and Gong [15] have proposed a novel spectral clustering algorithm for determining the number of clusters, the relevance learning method which measures the relevance of an eigenvector according to separate the dataset into different clusters and performing clustering. In this work the characteristic of eigenspace is analyzed and demonstrated that not every eigenvector of a data affinity matrix is informative and relevant for clustering. In this work the eigenvector selection is critical. The uninformative/irrelevant eigenvectors could lead to poor clustering results and the corresponding eigenvalues cannot be used for relevant eigenvector selection given a realistic data set. The algorithm correctly estimate the number of clusters by using BIC (Bayesian Information Criterion) with GMM (Gaussian Mixture Model) and also reveal the natural grouping of the input data/patterns even given sparse and noisy data.

Abhishek kumar et al., [6] have proposed a novel co-regularized spectral clustering algorithm for exploiting information from multiple views among clustering. In this co-regularized spectral clustering a joint optimization function is used for spectral embeddings of all the views. In multiple views the corresponding nodes in each graph should have the same cluster membership. In this work the novel spectral clustering objective functions that implicitly combine graphs from multiple views to achieve a better clustering. In this work some views have missing data for estimating the missing entry by simply average the similarities from views in which the data point is available.

Yeqing Li et al., [19] have proposed the novel large-scale multi-view spectral clustering (MVSC) approach based on the bipartite graph to solve the problem of clustering the large-scale data and out-of-sample data. This work first generates consensus uniform salient points among all the views to represent the manifold structures of all the features. For each view, one sub-bipartite-graph is constructed between the raw data points and the generated salient points. These generated points are used to capturing the manifold of the original views. In this work the local manifold fusion is used to generate a fused bipartite graph to integrate information of all the sub-graph for exploring the structure of the bipartite graph. For the clustering results, cluster labels are obtained for both the training data and salient points. The advantages of this proposed system is to handle the out-of-sample problem in low computational cost and high data samples.

In clustering of multiple tasks, the multiple related prediction tasks are considered simultaneously and the tasks belonging to different groups are allowed to overlap with each other. The data overlapping in multitask clusters degrades the clustering performance. This limitation requires to improve the multitask spectral clustering algorithm further.

III. PROPOSED METHODOLOGY

- **System design**

The proposed EMTSC algorithm has been designed to handle overlapping data occurred among clustering of multiple tasks to improve the clustering performance. In the proposed system, the input is taken from the dataset and partition the training data in each task into cluster. The proposed system first construct the Laplacian matrix based on the training data samples for partitioning the training data of each task into clusters. Then the diagonal matrix is computed based on the regression coefficient to simultaneously learn the cluster label matrix and the mapping function for each

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

individual task. While performing clustering of multiple tasks overlapping is occurred due to same tasks are belonging to different clusters.
The proposed EMTSC technique uses the Euclidean method to calculate the distance between two data points to find the similarity whose data points are lie in a high dimensional space for eliminating the occurrence of overlapping among clustering of multiple tasks. Finally, the clustering performance of proposed system has been analyzed using the metrics such as accuracy and running time of multitask clustering.

- **Enhanced multitask spectral clustering algorithm**

In clustering of multiple tasks overlapping can occur in multiple tasks when the same tasks are belonging to the different clusters. The proposed algorithm improves the clustering performance by using Euclidean method to eliminate these problems in clustering of multiple tasks. In this work the Euclidean method is used to calculate the distance between the data point for find the minimum distance with high similarity of the data points. The similar data points have similar labels and it is assigned to the same cluster. The dissimilar data points have dissimilar labels assigned to the different clusters. The Euclidean distance is also used to find the distance of the dissimilarity cluster for correspondingly placing clusters one by one.
The proposed EMTSC algorithm performs various steps includes computation of Laplacian matrix based on training samples, computation of diagonal matrix, compute Euclidean distance method and the output is the cluster indicator matrix. The proposed EMTSC algorithm for handling overlapping data among multiple task clustering is shown in Figure 3.1.

1. The input M dataset $\{X^{(t)}\}|^{M}_{t=1}$ and partition the training data in each task into cluster c_t $(1 \leq t \leq M)$
   //where c_t=cluster, t=task, M=unsupervised clustering tasks
   // initializing process
   for t = 1 to M do
2. Constructing the graph Laplacian matrix $L^{(t)}$ based on training data samples $X^{(t)}$
   end for
3. Initializing the regression coefficient for all tasks $\{W^{(t)}\}|^{M}_{t=1}$
   // where $W = [W^{(1)}, W^{(2)}, \ldots, W^{(M)}]$
4. Computing the diagonal matrix U based on regression coefficient W
   for t = 1 to M do
5. Compute $A^{(t)} = \alpha(\alpha X^{(t)}(X^{(t)})^{T} + \beta U)^{-1}$ and $H^{(t)} = (L^{(t)} - \alpha(X^{(t)})^{T}A^{(t)}X^{(t)} + \alpha I_t)$
   // where α = new balance parameter, β = balance parameter
6. Update cluster indicator matrix $F^{(t)}$ by performing eigen-decomposition over $H^{(t)}$
7. Update regression coefficient $W^{(t)} = A^{(t)}X^{(t)}F^{(t)}$
   end for
   until convergence
8. Compute Euclidean method $d(x_i, y_i) = \text{sqrt}(\sum^{n}_{i=1}(x_i - y_i)^2)$ // calculate distance between two points, where x_i and y_i are two data points
   for i = 2 to N
9. If (d (x, C) > θ) and (m < q) then create new clusters
   else unnecessary clusters are formed
   end if
   end for
   // where x = vector point, C = cluster, q = maximum number of cluster allowed
   θ = threshold value of dissimilarity of the cluster, N = number of data points
10. return cluster indicator matrix $\{F^{(t)}\}|^{M}_{t=1}$

After applying these steps get a number of clusters without the the problem of overlapping in the multitask cluster.

➢ **Compute Laplacian matrix**

In this work, after partitioning the training data of each task the graph Laplacian matrix is constructed based on the training data samples according to the data local structure using different strategies. For each task construct an undirected graph which can be represented by the weighted adjacency matrix $S = (s_{ij})_{i,j=1,2,\ldots,n}$. where $s_{ij} > 0$ indicates

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

that data points x_i and x_j are connected and $s_{ij} = 0$ means they are not connected. The graph Laplacian matrix is constructed based on data points is defined as, $L = D - S$.
Then the normalized graph Laplacian is defined as, $L_n = D^{-1/2} LD^{-1/2} = I_n - D^{-1/2} SD^{-1/2}$.
Where, L denotes graph Laplacian matrix, D denotes diagonal matrix with its diagonal $d_{ij} = \sum_j D_{ij}$ and S denotes weighted adjacency matrix.

- **Initialize the regression coefficients**

After constructing the graph Laplacian matrix, an explicit mapping function $f_t: R^{dx1} \quad \{0, 1\}^{ct\ x1}$ is learn for each task. By using the proposed spectral clustering approach the cluster indicator matrix is obtained for the *t*th task and initialize the regression coefficient to co-learn the cluster indicator matrix and mapping functions to make them reinforce each other is defined as,

$$S.\ t\ (F^{(t)})^T F^{(t)} = I_t$$ Where I_t is the c_t x c_t identity matrix, $F^{(t)}$ is the cluster indicator matrix

- **Computation of diagonal matrix**

The diagonal matrix is computed based on the regression coefficient of each individual task. For each task the regression coefficient and cluster label matrix is updated for solving the optimization problem occur in proposed system by performing eigen-decomposition over $H^{(t)}$.
Where $H^{(t)} = (L^{(t)} - \alpha(X^{(t)})^T A^{(t)} X^{(t)} + \alpha I_t)$, $A^{(t)} = \alpha(\alpha X^{(t)} (X^{(t)})^T + \beta U)^{-1}$, α is a new balance parameter, β is a balance parameter, U is a d x d diagonal matrix with its jth diagonal element, $L^{(t)}$ is the Laplacian matrix for the *t*th task and I_t is a c_t x c_t identity matrix.

- **Computation of Euclidean method**

The proposed system has been used the Euclidean distance method for calculating the distance of the two data points. If the similarity value is equal to 1, the data points are similar and the similarity value is equal to 0, the data points are dissimilar. By using cluster label matrix, the unique labels are assigned to the data points and similar data points have similar label number assigned to same cluster. Due to the calculation of similarity of the data points overlapping are eliminated among clustering of multiple tasks. In proposed system also calculate the dissimilarity of the cluster d(x, C) based on the distance between vector point and a cluster by using Euclidean distance method, where x is the vector point and C is the cluster. The vectors in different tasks belong to Euclidean center which implies no sharing of information between tasks apart from the size of the different vectors. If the Euclidean center is unknown, the low-dimensional space is also unknown. In proposed system every newly formed vector point is assigned to an existing cluster and the new cluster is created depend on the distance to the already defined clusters. The Euclidean distance method is defined as the following:

$$d(x_i, y_i) = \text{sqrt}\ (\sum^n_{i=1} (x_i - y_i)^2)$$ Where x_i and y_i are two data points

- **Cluster indicator matrix**

The cluster indicator matrix is used to returns the output of number of clusters are formed without the problem of overlapping in the multitask clusters. The cluster indicator vector F indicates which data points belong to the jth cluster C_j.

IV. EXPERIMENTAL RESULTS

In this work four real-world datasets have been used for performance evaluation. The first dataset is WebKB4, which contains four frequently-used classes such as student, project, faculty, and courses. The second one is the 3-D Motion Data [12] and the other two domain datasets are Comp.vs.Sci and Rec.vs.Talk [3], [4] generated from 20 Newsgroup.
For WebKB4, Comp.vs.Sci and Rec.vs.Talk sequentially divide each task into training part and test part, while for HumanEva 3-D Motion, 1000 data points has been selected from each task for training and the remaining data used for test. The training clustering refers to the process of grouping training data into several clusters, while testing clustering is the process of measuring the distance of the data points to find similarity of the data points in the cluster by using Euclidean method to eliminate overlapping in clustering of multiple tasks. In training clustering perform comparison algorithms to cluster training data and compare their clustering performance.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

- **Performance analysis of proposed EMTSC algorithm**

Experiments have been performed to evaluate the performance of the proposed EMTSC technique based on the four datasets. The performance of EMTSC algorithm has been compared with existing MTSC algorithm based on the metrics such as accuracy for clustering performance of multiple tasks and running time for clustering of multiple tasks

➢ **Accuracy for multitask clustering performance**

The experiments have been performed with the parameters of comparing the existing and proposed algorithms are consistently tuned from the range of 10^{-2}, 10^{-1}, 10^{0}, 10^{1} and 10^{2}. The clustering accuracy results are shown in Figure 4.1.

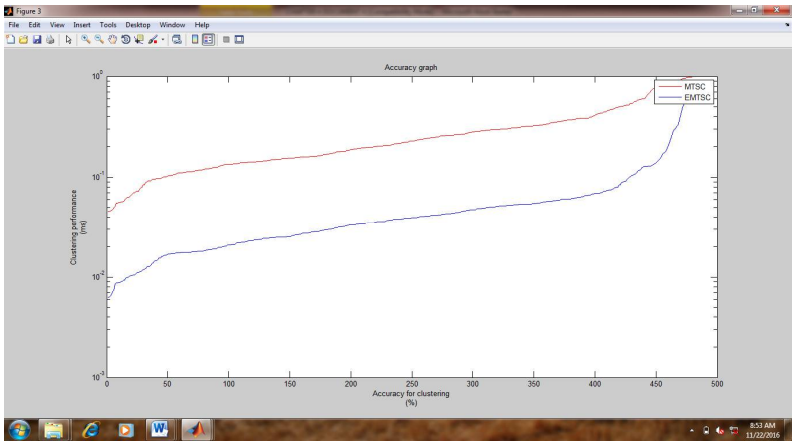

Figure 4.1 Comparison of accuracy for multitask clustering performance

In the X-axis represents the accuracy for number of clusters formed and Y-axis represents the range of parameters to perform clustering. The performance graph shows that the proposed EMTSC method improves clustering performance with better accuracy than the existing MTSC method based on the four datasets with two tasks.

- **Running time for multitask clustering**

The experiments have been performed with four datasets to evaluate the performance of the running time to perform multitask clustering. The running time for multitask clustering performance is shown in Figure 4.2.

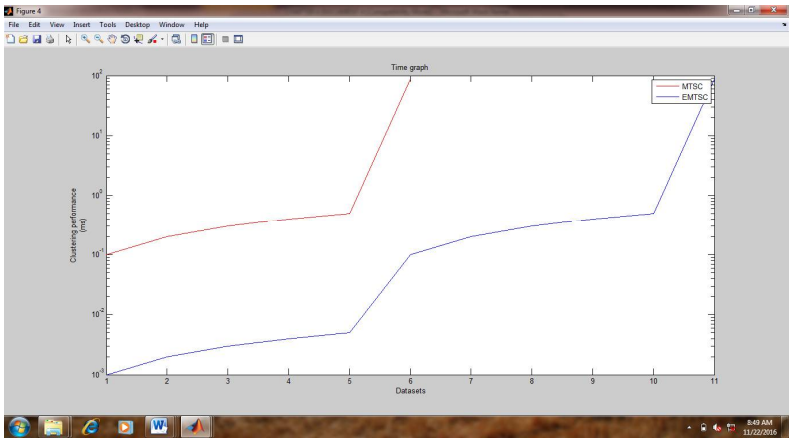

Figure 4.2 Comparison of running time for multitask clustering

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

In the graph X-axis represents the datasets and Y-axis represents the clustering performance of the data points. The graph shows that the proposed EMTSC method provides better clustering performance with less running time than the existing MTSC method based on the four datasets with two tasks.

V. CONCLUSION

Spectral clustering has become the most important technique for multiple clustering tasks and it shows more capability in partitioning of data with more complicated structures compared to traditional clustering approaches. By using spectral clustering, exploitation of data local structures can improve clustering performance. However the task-specific property of the multiple clustering requires an effective spectral clustering scheme for simultaneously performing multiple clustering tasks on different data. In multitask clustering, the multiple related prediction tasks are considered simultaneously and the same tasks belong to different clusters are allowed to overlap which degrades the clustering performance. In this research work an enhanced multitask spectral clustering technique (EMTSC) has been proposed to perform multitask clustering without overlapping in data points of multiple tasks for improving clustering performance. The experimental results show that the proposed system improves the performance of multiple tasks clustering in large scale data sources with less complexity of partitioning of data, high accuracy and less running time for multitask clustering.

REFERENCES

[1] F. Bonchi, A. Gionis and A. Ukkonen, "Overlapping correlation clustering", In 11th International Conference on Data Mining (ICDM), IEEE, pp. 51-60, Dec 2011.
[2] Feiping Nie, Dong Xu, I. W. Tsang and C. Zhang, "Spectral Embedded Clustering", In Proceedings of the 21st International Conference on Artificial Intelligence, pp. 1181-1186, Jul 2009.
[3] Q. Gu and J. Zhou, "Learning the shared subspace for multi-task clustering and transductive transfer classification", In Proceedings of the 9th IEEE International Conference of Data Mining (ICDM), Miami, FL, USA, pp. 159–168, 2009.
[4] Q. Gu, Z. Li, and J. Han, "Learning a kernel for multi-task clustering", In Proceedings of the 25th AAAI Conference of Artificial Intelligence, San Francisco, CA, USA, Aug 2011.
[5] H. Jia, S. Ding, X. Xu and R. Nie, "The latest research progress on spectral clustering", Neural Computing and Applications, Vol. 22, No.7-8, pp. 1477-86, Jun 2013.
[6] A. Kumar, H. Daumé, "A co-training approach for multi-view spectral clustering", In Proceedings of the 28th International Conference on Machine Learning (ICML-11), pp. 393-400, 2011.
[7] T. Li, C. Ding and M. I. Jordan, "Solving consensus and semi-supervised clustering problems using nonnegative matrix factorization", In 7th IEEE International Conference on Data Mining (ICDM), pp. 577-582, Oct 2007.
[8] H. Lu, Z. Fu and X. Shu, "Non-negative and sparse spectral clustering", Pattern Recognition, Vol. 47, No. 1, pp. 418-426, Jan 2014.
[9] S. Meenakshi, R. Renukadevi, A Review on Spectral Clustering and its Applications, International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE), Vol. 4, No. 8, pp. 14840-14845, Aug 2016.
[10] A.Y. Ng, M.I. Jordan and Y. Weiss, "On Spectral Clustering: Analysis and an algorithm", In Advances in Neural Information Processing Systems, Vol.13, No. 7, pp. 849-856, 2001.
[11] R. Shang, Z. Zhang, L. Jiao, W. Wang and S. Yang, "Global discriminative-based nonnegative spectral clustering", Pattern Recognition, Vol. 55, pp. 172-82, Jul 2016.
[12] L. Sigal, A. O. Balan and M. J. Black, "HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion", International Journal of Computer Vision, Vol. 87, No. 1, pp. 4–27, 2010.
[13] Ulrike von Luxburg, "A tutorial on spectral clustering, Statistics and computing", Vol. 17, No. 4, pp. 395-416, 2007.
[14] Verma, D., and Meila, M., "A comparison of spectral clustering algorithms", University of Washington Tech Rep UWCSE030501, No. 1, pp. 1-18, 2003.
[15] T. Xiang, S. Gong, "Spectral clustering with eigenvector selection, Pattern Recognition", Vol. 41, No. 3, pp. 1012-1029, Mar 2008.
[16] S. Xie, H. Lu and Y. He, "Multi-task co-clustering via nonnegative matrix factorization", In Pattern Recognition (ICPR), 21st International Conference, IEEE, pp. 2954-2958, Nov 2012.
[17] Y. Yang, Z. Ma, Y. Yang, F. Nie, H. T. Shen, "Multitask spectral clustering by exploring inter-task correlation", IEEE transactions on cybernetics, Vol. 45, No. 5, pp. 1083-1094, May 2015.
[18] Y. Yang, H. T. Shen, F. Nie, R. Ji and X. Zhou, "Nonnegative spectral clustering with discriminative regularization", In 25th AAAI Conference on Artificial Intelligence, Aug 2011.
[19] Yeqing Li, Feiping Nie, Heng Huang, Jun Zhou Huang, "Large-Scale Multi-View Spectral Clustering via Bipartite Graph", In Proceedings of the 29th AAAI Conference on Artificial Intelligence, pp. 2750-2756, 2010.