ISSN: 2277-9655
Impact Factor: 1.852

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Multiview Clustering Based on Tensor Solutions

Dr.Bharathi.A[*1], Anitha.S[2]

[*1] Professor, Department of Information Technology, Bannari Amman Institute of Technology, Erode, India

[2] PG Scholar, Dept of Information Technology, Bannari Amman Institute of Technology, Erode, India

bharathia@bitsathy.ac.in

Abstract

Integrating multiview cluster is an crucial issue in heterogeneous environment. Spectral clustering is used for integrating cluster in heterogeneous environment. In this paper, we used tensor decomposition for identifying hidden pattern in the context of spectral clustering. This gives the good result when compared to other methods. Here synthetic datasets are used for evaluating the results. In most of the tensor solution two dimensional or three dimensional data are used. The higher order datasets are used in multiview clustering

Keywords: Component- Multiview clustering, Tensor decomposition, Higher Order Data, Spectral Clustering.

Introduction

Datasets consists of multiple similarities. For a group of people, we might know their height, weight, age, education, location, designation and their family related. For a set of authors have their own papers, citations and methods. For a set of drives, they have own files, documents, locations and their contents. Our approach is to cluster people, researchers, methods, or files, to treat all the similarities concurrently.

In a set of multiple networks, they share same set of nodes but possess different types of connection between nodes. Multiple relationship can be formed through individual activity is called as multiview learning [2]. The recent development in clustering is the spectral clustering. Spectral clustering is based on the Ncut algorithm [1]. This can work well in the single view data as it is based on matrix decompositions. Many clustering algorithms have been proposed in comparison with the single view data. Therefore, these algorithms have some limitation.

Tensors are the higher order generalization of matrices .They can be applied to several domains such as web searching, image processing, data mining, and image recognition. Here tensor based methods are used to model multiview data. This is also used to detect the hidden pattern in multiview data subspace by tensor analysis. It works based on the tensor decomposition [1] which captures the multilinear structures in higher order data, where data has more than two modes. From the above example, in tensor, similarity of researchers is one slice, and then similarity citations are one slice. Likewise all slices will be combined to form tensor.

Tensor decomposition is used to cluster all the similarity matrices into set of compilation feature vector. Many clustering algorithms like k Means, SVD, HOSVD [15] are used for many tensor methods. Spectral clustering [1] is used for clustering the similarity matrices based on tensor methods.

Materials and Methods

Multiview Clustering

A multiview clustering method that extends k-means and hierarchical clustering to deal with data as two conditionally independent views [13]. Canonical correlation analysis in multiview clustering assumes that the views are uncorrelated in the given cluster label. These algorithms can concentrate only on two view data. Long et al formulated a multiview spectral clustering method while investigating multiple spectral dimension reduction.

Zhou and Burges developed a multiview clustering strategy through generalizing the Ncut from a single view to multiple views and subsequently they build a multiview transductive inference. In tensor-based strategy, the multilinear relationship among multiview data is taken into account. The strategy focuses on the clustering of multitype interrelated data objects, rather than clustering of the same objects using multiple representations as in our research.

Spectral Clustering

Spectral clustering was derived based on relaxation of the Ncut formulation for clustering. Spectral clustering involves a matrix trace optimization

ISSN: 2277-9655
Impact Factor: 1.852

problem [14]. In this paper, we proposed that the spectral clustering formalism can be extended to deal with multiview problems based on tensor computations.

Given a set of N data points $\{x_i\}$ where $x_i \varepsilon IR^d$ is the i th data point, a similarity s_{ij} can be defined for each pair of data points x_i and x_j based on some similarity measure. An intuitive way for representing the data set by using a graph G= (V, E) in which the vertices V represents the data points and the edges characterize the similarity between data points which are quantified by s_{ij} the similarity measure of the graph is symmetric and undirected. The matrix of the graph G is the matrix S with entry in row i and column j equal to s_{ij}.The degree of the vertex can be written as

$$d_i = \sum_{j=1}^{N} s_{ij} \quad (1)$$

Where v_i is connected to the sum of all weight of the edges. The degree of the matrix D is a diagonal matrix containing the vertex degrees from $d_1 \ldots . d_N$ As the diagonal, It follows from the spectral formalism of embedding the Laplacian matrix can be defined as L=D-S and Ncut is defined by corresponding to the normalized Laplacian matrix

$$L_{Ncut} = D^{-1/2} L D^{-1/2} = \text{I-}S_N \quad (2)$$

Where s_N and L_{Ncut} are the normalized similarity eigenvector and their eigenvalues.

Multiview Spectral Clustering

In integration of multiview data in spectral clustering, there are two different strategies

Multiview Clustering by Trace Maximization (MC-TR-I)

Different views can be added to the objective function. The function can be as follows,

$$\max_{U} \sum_{K=1}^{K} trace\left(U^T S_N^{(K)} U\right).$$
$$= \left(U^T \left(\sum_{k=1}^{K} S_N^{(k)}\right) U\right), \quad (3)$$
$$U^T U = I,$$

Where $S_N^{(K)}$ a normalized matrix for kth value and U is the common factor shared by the views. This corresponds to the MKF with linear kernel. In alternative, weighted combination of objective functions, where the weights are learned from the data[6],[7]

$$= \max_{U,W} \sum_{K=1}^{K} W_K trace\left(U^T S_N^{(K)} U\right)$$
$$= \max_{U,W} \text{trace}\left(U^T \left(\sum_{K=1}^{K} W_K S_N^{(K)}\right) U\right), \quad (4)$$
$$U^T U = I,\ W \geq 0 \text{ and } \|W\|_F = 1.$$

Tensor

Scalars, vectors and matrices are facilitated by higher order tensors. Scalars denote lower case letters (A, B…). Vectors are written in italic capitals (, *A, B….)* matrices correspond to boldface capitals (A, B**…)** and tensors are written as casteller lettering (**A, B**…). This notation is consistently used for ordering lower parts of a given quantity such as a vector a, matrix **A**, and tensor *A*, correspondingly. The Kronecker product is denoted by ⊗. Vector is first order, matrix is a second order and tensors are the third order or the higher order tensor.

Tensor construction

Multiview data can be constructed from by several methods. Here tensor is constructed by stacking object by feature matrices derived from multiple views in a tensor. This is applicable only to the homogeneous data sources in which dimensions of the various features are same. In many multiview applications it deals with the heterogeneous data sources with the various dimensions with feature spaces.

In this paper, construction of the independent data dimensions is made by integrating heterogeneous data sources. We can work with similarity tensor $A \varepsilon IR^{N \times N \times N}$in which the similarity matrices $S_N^{(1)}, S_N^{(2)}, S_N^{(3)} \ldots . S_N^{(K)}$ are associated with different views. In similarity tensor, each similarity matrix view is computed in different space so normalization is required, so our method can be considered as normalization steps.

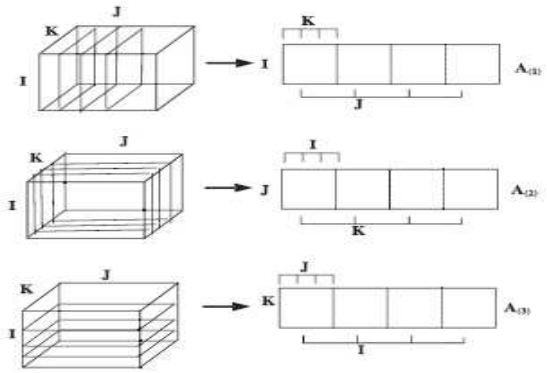

Fig 1. Unfolding of tensor into matrices

From the fig 1. The tensor can be unfolded into matrices. Then the similarity matrix are calculated. The

ISSN: 2277-9655
Impact Factor: 1.852

tensor can be constructed based on the below algorithm. Here it is constructed using following algorithms.

MC-TR-I

The pseudo code for the column space of the optimal matrix U in dominant eigen space of $\sum_{K=1}^{K} S_N^{(K)}$ is as follows:

Algorithm: MC-TR-I-EVD
Comment: M is the number of cluster
Step 1: Similarity matrix $\sum_{K=1}^{K} S_N^{(K)}$ is formed
Step 2: Using Eigen value decomposition U is formed
Step 3: The row of U is normalized to unit length.
Step 4: The cluster is calculated with k means on U
Return (clustering label)
Algorithm: MC-TR-IEVDIT
Step 1: Initialize MC-TR-I-EVD
While (! convergence>
Do
{
Obtain p (U)
Calculate weighting vector W
Relaxed assignment matrix U is obtained
}
Step 3: The rows of the U are normalized.
Step 4: The cluster is calculated with k means on U
Return (clustering label).

MC-FR-OI

Algorithm: MC-FR-OI-MLSVD
Comment: M is the number of clusters
Step1: Build a Similarity tensor *A*
Step2: Obtain the unfolding matrix $\boldsymbol{A}_{(1)}$
Step 3: Compute U from the subspace spanned
By M dominant left singular vectors of $\boldsymbol{A}_{(1)}$
Step 4: Normalize the rows of U to unit length
Step 5: Calculate the cluster with k-means on U
Return (clustering label)

The optimal solution can also be obtained through HOOI algorithm. Conditional updates for HOOI make iterates for U and V which are different in which *A* is symmentric in first two modes, then the U and V iterates will match again. Using the approximate of U for updating in both first and second mode may lead to deviation. The resulting algorithm is called MC-FR-OI-HOOI

Algorithm: MC-FR-OI-HOOI
Step 1: Build a similarity tensor $\boldsymbol{A}$
Step 2: Obtain the unfolding matrices $\boldsymbol{A}_{(1)}$, $\boldsymbol{A}_{(2)}$ and $\boldsymbol{A}_{(3)}$
Step 3: Obtain an initial U and V by MLSVD
While(!convergence)
Do
{
Iteration step 4.1.U in dominant subspace of $\boldsymbol{A}_{(1)}$ ($V_1 \otimes I$)
Iteration step 4.2 V in dominant subspace of $\boldsymbol{A}_{(1)}$($U_1 \otimes I$)
}
Comment: i is the counter of iteration.
Step 5: Normalize the rows of U to unit length
Step 6: Calculate the cluster with k-means on U
Return (clustering label)

Both MC-FR-OI-MLSVD and MC-FR-OI-HOOI involve a joint matrix compression.

MC-FR-MI

The objective function is not affected by the sign of W and all the weights can be nonnegative. The matrix U and the vector W yields the optimal subspace and the weights in different views.

Alogrithm: MC-FR-MI-HOOI
Step 1: Build a similarity tensor *A*
Step 2: Obtain the unfolding matrices $\boldsymbol{A}_{(1)}, \boldsymbol{A}_{(2)}, \boldsymbol{A}_{(3)}$
Step 3: Obtain an initial $\mathbf{U}_0$, $\mathbf{V}_0$ and $\mathbf{W}_0$ by MLSVD
While (! convergence)
Do
{
Iteration step 4.1. $\boldsymbol{A}_{(1)}(V_1 \otimes W_i)$
Iteration step 4.2. $\boldsymbol{A}_{(2)}(W_1 \otimes U_{i+1})$
Iteration step 4.3 $\boldsymbol{A}_{(3)}(U_1 \otimes V_{i+1})$
}
Comment i is the counter of iteration.
Step 5: Normalize the rows of U to unit length.
Step 6: Calculate the cluster with k-means on U
Return (clustering label)

Algorithm: MC-FR-MI-HOOI
Step 1: Build a similarity tensor *A*
Step 2: Obtain the unfolding matrices $\boldsymbol{A}_{(1)}, \boldsymbol{A}_{(2)}, \boldsymbol{A}_{(3)}$
Step 3: Obtain an initial $\mathbf{U}_0$ by MLSVD
While (! convergence)
Do
{
Iteration step.4.1.Calculate W_{i+1} as the dominant left singular vector of $A_{(3)}$ ($U_i \otimes U_i$)
Iteration step.4.2 Compute new integration matrix Š
Iteration step.4.3 Obtain U_{i+1} by eigenvalue Decomposition of Š
}
Comment i is the counter of iteration
Step 5: Normalize the rows of U to unit length
Step 6: calculate the cluster with k-means on U
Return (clustering label)

In MC-OI framework, MC-FR-OI-MLSVD and MC-FR-OI-HOOI are discussed. In MC-MI framework, only MC-FR-MI-HOOI discussed. The reason is that the

ISSN: 2277-9655
Impact Factor: 1.852

test indicated here is mere truncated MLSVD, which in third mode only one vector is retained in which the results are not retained.

Higher Order Data

In existing system, 3D projection of a data points are used. The projection of a data points are represented as X-Y projection, Y-Z projection and X-Z projection. The combinations of these three projections are the single view dataset [1].

Direct combination of these three projections will not give the proper clustering. By using tensor , all the projections can be simultaneously combined and clustering can be done. The performance of the clustering is high when compared to others.

In proposed system, multiview clustering can be done by using ten dimensional projections of a data points. Here the microarray dataset of gene in the body are taken as the input.

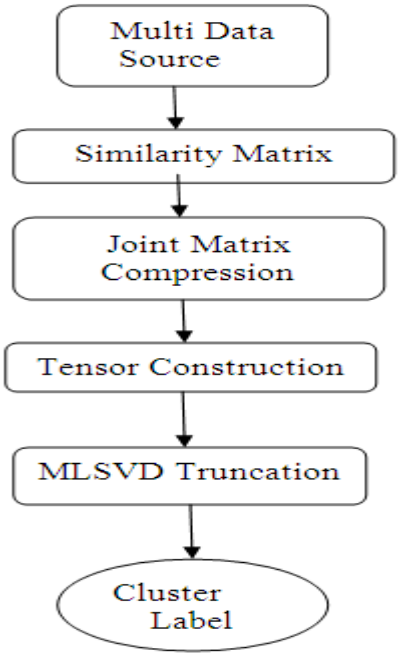

Fig 2. Flow chart for Multiview clustering

In multiview clustering, it is proposed based on the SVD and PCA analysis. This framework can be extended using spectral clustering. By using tensor decomposition framework, both homogeneous and heterogeneous information can be integrated simultaneously to form clustering. First dataset can be given as a input. Then ten dimensional similarity matrix is formed.

The similarity matrix can be given as a input to the Eigen value decomposition. It accepts square matrix alone as a input. The given square matrix can be ten dimensional data. Then the eigen value decomposition value can be determined. Then values are normalized. Based on that value, weight matrix is calculated. Then the normalization of the weight matrix is calculated.

The normalization is done to get the values in ones and zeroes. The tensor construction is made for achieving multiview clustering. The tensor framework is made for integrating the multiview clustering. It consists of two types of algorithm. It can be implemented in tensor framework and then the values can be normalized. The final eigen values are calculated and then the values can be normalized.

The variables of the clusters can be given as a input and then the nearer clusters can be calculated based on the similarities. The clustering can be done and the final cluster labels are formed. This is an efficient method for handling the high order data. The data can be in any form they can be either image or the large amount of input.

Experimental Evaluation

We evaluated and compared with micro array dataset networks.. This can be implemented in micro array dataset to increase the experimental power. The interaction probability is calculated for each group member. Based on the dimensions, interaction probability differs. When we add one member of the group to the other group with low probability. Normalized Mutual Information (NMI) is used to measure the performance. NMI value is 1 when two clusters are exactly same. Normally NMI values exist between 0 and 1.

Here tensor decomposition with higher order data is used to cover the hidden patterns and it also performs well, when compared with other two methods. For example if we include some noises in the second dimension, the performance is reduced from 0.5 to 0.1. For this dimension, the pattern cannot be analyzed. By using tensor based method, it achieves good performance. This method is more robust in noisy dimensions.

Table 1

	Methods	NMI	AMI
Single View	DNA mRNA RNA CDS	0.7606 0.8928 0.7197 0.6317	0.7994 0.9191 0.8195 0.5598

ISSN: 2277-9655
Impact Factor: 1.852

Multi View	MC-FR-OI-MLSVD MC-FR-OI-HOOI MC-FR-MI-HOOI MC-TR-I-EVDit	0.9320 0.9240 0.9430 0.9632	0.9507 0.9508 0.9668 0.9616

Performance on Micro array dataset with Multiview cluster

Here the genes in single dimensional community detection is low when compared with the multidimensional community. In multidimensional clustering, the performance of gene is higher when compared with the other maximization techniques. Single dimensions have the higher variance when compared with all multidimensional community.

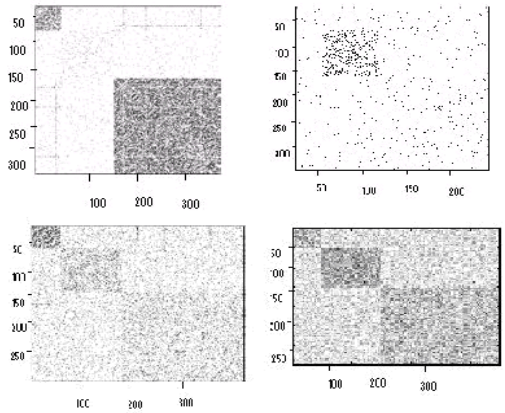

Fig 3. Example of Multi dimensional Network.

In fig 3. First dimension consists of two clusters and second dimension consists of one cluster. The first two dimensions are based on single view dimensional with low dimensionality. Therefore, the hidden patterns cannot be viewed properly. They are indicated with the low NMI. When comparing with the single view clustering, hidden patterns can be shown clearly. The NMI value exists between 0 and 1.

We first evaluate and compare the different clustering strategies and applied to the multiview clustering. Here clustering can be formed by the extension of the spectral clustering. Here dataset consist of various types of gene which consist of more cluster members.

We can generate different view of interaction that is in each view; network shares the same vertices but has a different interaction pattern. The group members within the group can interact with the others random manner. The probability of interaction differs with respect to distinct views. Two vertices are connected randomly in low probability by adding some noise. The different views demonstrate the different interaction pattern in them.

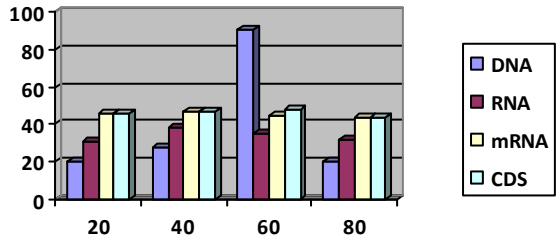

Fig 4. Performance of micro array dataset in Multiview cluster

In fig 4.The clustering evaluation for micro array dataset in multiview clustering is analyzed. In multiview clustering hidden patterns can be clearly viewed when compared to the single view clustering. Therefore, most of the multiview clustering results are better than the single view clustering. Multiview clustering helps to reduce the noise and shows the shared cluster.

Result and Discussion

In multidimensional networks, hidden pattern can be viewed clearly when compared to the single dimensional networks. Here we have proposed the extension of spectral clustering by implementing the spectral clustering algorithm. By using this algorithm, multidimensional networks can be included in heterogeneous environment. Tensor based solutions are proposed for including multiple networks in the form of tensor.

Multiview clustering performance can be analyzed using many tensor based strategies. Multiple similarities can be founded through many methods. In cluster ensemble method, single view partition clusters can be integrated. So, this is not efficient.

In LMF method, clustering performance get lower at initialization and the partition cluster at the end are unstable. The optimization results also consumes much time.

In tensor based multiview clustering, spectral clustering can be extended by spectral clustering using Higher Order Data where multiple networks can be linked together. This method is more efficient when compared with other methods. NMI is used to measure the performance of the clustering.

ISSN: 2277-9655
Impact Factor: 1.852

References

[1] Xinhai Liu, Shuiwang Ji, Wolfgang Glanzel, and Bart De moor, Fellow, "Multiview clustering via Tensor Methods" IEEE Trans. Knowledge and Data Engineering, vol.25, no. 5, May 2013.

[2] Lei Tang, Xufei Wang, Huan Liu, "Uncovering Groups via Heterogeneous Interaction Analysis",

[3] H.G. Ayad and M.S. Kamel, "Cumulative Voting Consensus Method for Partitions with Variable Number of Clusters," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 30, no. 1, pp. 160-173, Jan. 2008.

[4] Teresa M. Selee,, Tamara G. Kolda, W. Philip Kegelmeyer, And Joshuda. Griffin," Extracting Clusters from Large Datasets with Multiple Similarity Measures Using IMSCAND", summer proceedings, 2007

[5] S. Bickel and T. Scheffer," Multi-View Clustering", Proc. IEEE Fourth Int'l Conf. Data Mining (ICDM '04), pp. 19-26, 2004.

[6] J.D. Carroll and J.J. Chang," Analysis of Individual Differences in Multidimensional Scaling via an n-Way Generalization of Echart-Young, Decomposition", Psychometricka, vol. 35, pp. 283-319, 1970.

[7] K. Chaudhuri, S.M. Kakade, K. Livescu, and K. Sridharan," Multi-View Clustering Via Canonical Correlation Analysis", Proc. 26th Ann. Int'l Conf. Machine Learning (ICML '09), pp. 129-136, 2009.

[8] A. Cichocki, R. Zdunek, A.-H. Phan, and S. Amari," Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation". John Wiley, 2009.

[9] L. De Lathauwer, B.D. Moor, and J. Vandewalle, "A Multilinear Singular Value Decomposition, SIAM J. Matrix Analysis and Applications", vol. 21, no. 4, pp. 1253-1278, 2000.

[10] L. De Lathauwer, B.D. Moor, and J. Vandewalle, "On the Best Rank-1 and Rank Approximation of Higher-Order Tensors, SIAM J. Matrix Analysis and Applications", vol. 21, no. 4, pp. 1324-1342, 2000.

[11] D.M. Dunlavy, T.G. Kolda, and E. Acar, Poblano v1.0:" A MATLAB Toolbox for Gradient-Based Optimization", Technical Report SAND2010-1422, Sandia Nat'l Laboratories, Mar. 2010.

[12] D.M. Dunlavy, T.G. Kolda, and W.P. Kegelmeyer," Multilinear Algebra for Analyzing Data with Multiple Linkages," Technical Report SAND2006-2079, Sandia Nat'l Laboratories, 2006.

[13] H. Huang, C. Ding, D. Luo , and T. Li, "Simultaneous Tensor Subspace Selection and Clustering: The Equivalence of High Order SVD and k-Means Clustering," Proc. 14th ACM SIGKDDInt'l Conf. Knowledge Discovery and Data Mining, pp. 327-335, 2008.

[14] B. Long, Z.M. Zhang, X. Wu´ , and P.S. Yu, "Spectral Clustering for Multi-Type Relational Data," Proc. 23rd Int'l Conf. Machine Learning, pp. 585-592, 2006.

[15] U. Luxburg, "A Tutorial on Spectral Clustering," Statistics and Computing, vol. 17, no. 4, pp. 395-416, 2007