MULTI-VIEW CLUSTERING VIA SIMULTANEOUSLY LEARNING GRAPH REGULARIZED LOW-RANK TENSOR REPRESENTATION AND AFFINITY MATRIX

Yongyong Chen, Xiaolin Xiao, and Yicong Zhou

Department of Computer and Information Science, University of Macau, Macau, China
YongyongChen.cn@gmail.com, shellyxiaolin@gmail.com, yicongzhou@um.edu.mo

ABSTRACT

Low-rank tensor representation-based multi-view clustering has become an efficient method for data clustering due to the robustness to noise and the preservation of the high order correlation. However, existing algorithms may suffer from two common problems: (1) the local view-specific geometrical structures and the various importance of features in different views are neglected; (2) the low-rank representation tensor and the affinity matrix are learned separately. To address these issues, we propose a novel framework to learn the Graph regularized Low-rank Tensor representation and the Affinity matrix (GLTA) in a unified manner. Besides, the manifold regularization is exploited to preserve the view-specific geometrical structures, and the various importance of different features is automatically calculated when constructing the final affinity matrix. An efficient algorithm is designed to solve GLTA using the augmented Lagrangian multiplier. Extensive experiments on six real datasets demonstrate the superiority of GLTA over the state-of-the-arts.

Index Terms— Multi-view clustering, low-rank tensor representation, Tucker decomposition, adaptive weights, local manifold.

1. INTRODUCTION

Multi-view data is ubiquitous in data mining and machine learning areas. For example, images are often described by different kinds of features, such as color, texture, and edge; documents can be translated into different languages. Multi-view clustering [1, 2, 3, 4, 5, 6] takes advantages of these multi-view data to boost the clustering performance, and has shown the superiority over its single-view counterparts.

Considerable efforts have been made to develop many efficient algorithms for multi-view clustering. The core of multi-view clustering is how to integrate the information of multi-view features to form a reliable affinity matrix of all samples. For example, Xia *et al.* [7] proposed to recover a shared low-rank transition probability matrix by low-rank and sparse decomposition. One limitation of [7] is that the high order correlation of multiple features may be ignored by only learning the common information among all views [2]. To tackle this issue, the work in [8] proposed a novel essential tensor learning method by tensor robust principal component analysis framework [9, 10]. Under the assumption that the observed data points usually reside in some low-dimensional spaces [11, 12, 13, 14, 15], Zhang *et al.* [2] extended the low-rank representation [13] into the multi-view setting. Besides, the high order correlation is captured by the tensor nuclear norm defined as the sum of the nuclear norms of all unfolding matrices [16]. However, it may yield unsatisfactory performance in real applications, since the unfolding-based tensor nuclear norm is a loose approximation of Tucker rank [9, 17]. To mitigate this issue, the tensor-SVD-based tensor nuclear norm [9] is exploited in [4] to ensure the consensus among multiple views.

While the low-rank tensor representation-based multi-view clustering methods have achieved incredibly encouraging performance, they still suffer from the following limitations: (1) the ignorance of the local view-specific geometrical structures. The methods in [2, 4, 7, 8] extended the (tensor) robust principal component analysis or (tensor) low-rank representation [13, 9, 17] for multi-view clustering by only taking the global low-rank property of the representation matrices/tensors into consideration. Thus the locality and similarity information of samples may be ignored in the learning process [18, 3, 19]. (2) To obtain the clustering results, the methods in [2, 4] first pursue the subspace representation tensor using different tensor rank approximations and then construct the final affinity matrix. This way, the two critical factors in spectral clustering, *i.e.*, the representation tensor and the affinity matrix, are learned separately. Thus there is no guarantee of recovering an optimal clustering result. Besides, this scheme treats the representation matrices of different views equally, which may lead to unsatisfactory performance in real applications. It is mainly because different features characterize specific and partly independent information of the datasets and thus may have distinct contributions for the final clustering results [20].

Such two concerns are not well solved in existing low-

This work was supported in part by the Macau Science and Technology Development Fund under Grant FDCT/189/2017/A3, and by the Research Committee at University of Macau under Grants MYRG2016-00123-FST and MYRG2018-00136-FST. (Corresponding author: Yicong Zhou.)

978-1-5386-9552-4/19/$31.00 ©2019 IEEE
DOI 10.1109/ICME.2019.00234

rank tensor representation-based multi-view clustering methods. Motivated by this observation, we propose a novel multi-view clustering method by learning the Graph regularized Low-rank Tensor representation and Affinity matrix (GLTA) in a unified framework.

Contribution. The main contributions of this work can be summarized as follows. The proposed GLTA not only learns both the representation tensor and the affinity matrix, but also considers the local view-specific geometrical structures and the various contributions of different features. Specifically, GLTA exploits the Tucker decomposition to encode the low-rank property, adopts the manifold regularization to depict the local view-specific geometrical structures, and assigns adaptively various weights for different features when constructing the final affinity matrix. This way, the high order correlation among different views as well as the local view-specific geometrical structures can be explicitly captured. Extensive experimental evaluations on six real datasets demonstrate that GLTA outperforms the state-of-the-art approaches.

2. GLTA

In this section, we propose a novel multi-view clustering method, named GLTA, and then solve GLTA by the augmented Lagrangian multiplier (ALM).

2.1. The proposed GLTA

Suppose that $X^v \in \mathbb{R}^{d_v \times n}$ $(v = 1, 2, \cdots, V)$ denotes the v-th feature matrix, $Z^v \in \mathbb{R}^{n \times n}$ is the corresponding representation matrix, d_v is the dimension of a sample vector in the v-th feature matrix, n is the total number of data points, and V is the number of views. Then, the proposed GLTA is formulated as

$$\left\{\begin{array}{l} \min\limits_{\mathcal{Z},E,S,\omega} \underbrace{\|E\|_{2,1}}_{\text{noise}} + \sum\limits_{v=1}^{V}\Big(\underbrace{\alpha * tr\big(Z^v L^v Z^{v'}\big)}_{\text{local manifold}} + \underbrace{\beta\omega_v\|Z^v - S\|_F^2}_{\text{consensus representation}}\Big) \\ s.t.\ \ X^v = X^v Z^v + E^v,\ \ v = 1, 2, \cdots, V, \\ \qquad \underbrace{\mathcal{Z} = \mathcal{C} \times_1 U_1 \times_2 U_2 \times_3 U_3,\ U_i' * U_i = I, (i = 1, 2, 3),}_{\text{low-rank tensor representation}} \\ \qquad E = [E^1; E^2; \cdots; E^V],\ \ \omega \geq 0,\ \Sigma_v \omega_v = 1, \end{array}\right. \tag{1}$$

where $\|\cdot\|_{2,1}$ is the $l_{2,1}$-norm to remove the noise E. All representation matrices $\{Z^v\}$ are merged to construct a 3-order tensor $\mathcal{Z}$. Then, the low-rank property of $\mathcal{Z}$ is described by the Tucker decomposition, that is, $\mathcal{Z} = \mathcal{C} \times_1 U_1 \times_2 U_2 \times_3 U_3$. $\mathcal{C}$ is the core tensor and U_i $(i = 1, 2, 3)$ is the orthogonal factor matrix. L^v is the normalized graph Laplacian matrix. $tr(\cdot)$ denotes the trace of a matrix. S is the final affinity matrix. α and β are nonnegative parameters. $\omega = [\omega_1, \omega_2, \cdots, \omega_V]$ is the weight vector, whose entry ω_v is the relative weight of the v-th view. In Eq. (1), we can learn both the representation tensor $\mathcal{Z}$ encoded by the Tucker decomposition and the local manifolds, and the affinity matrix S by assigning different weights on different features.

2.2. Optimization of GLTA

This section aims to develop an efficient optimization algorithm to solve Eq. (1). Due to the inseparable property of variable $\mathcal{Z}$, we adopt the variable-splitting technique and introduce one auxiliary tensor variable $\mathcal{Y}$. Then, Eq. (1) can be reformulated as the following optimization problem:

$$\begin{array}{l} \min\limits_{\mathcal{Y},\mathcal{Z},E,S,\omega} \|E\|_{2,1} + \sum\limits_{v=1}^{V}\Big(\alpha * tr\big(Y^v L^v Y^{v'}\big) + \beta\omega_v\|Y^v - S\|_F^2\Big) \\ s.t.\ \ X^v = X^v Y^v + E^v,\ \ v = 1, 2, \cdots, V, \\ \qquad \mathcal{Z} = \mathcal{C} \times_1 U_1 \times_2 U_2 \times_3 U_3,\ U_i' * U_i = I, (i = 1, 2, 3), \\ \qquad E = [E^1; E^2; \cdots; E^V],\ \omega \geq 0,\ \Sigma_v \omega_v = 1,\ \mathcal{Z} = \mathcal{Y}. \end{array} \tag{2}$$

Using the ALM methodology, the corresponding augmented Lagrangian function of Eq. (2) is obtained by

$$\begin{aligned} \mathcal{L}_\rho = &\|E\|_{2,1} + \sum_{v=1}^{V}\Big(\alpha * tr\big(Y^v L^v Y^{v'}\big) + \beta\omega_v\|Y^v - S\|_F^2\Big) \\ &+ \frac{\rho}{2}\Big(\sum_{v=1}^{V}\|X^v - X^v Y^v - E^v + \frac{\Theta^v}{\rho}\|_F^2 + \|\mathcal{Z} - \mathcal{Y} + \frac{\Pi}{\rho}\|_F^2\Big), \end{aligned} \tag{3}$$

where Θ and Π of size $\mathbb{R}^{n \times n \times V}$ are Lagrange multipliers with respect to two equal constraints, respectively. $\rho > 0$ is the penalty parameter. Borrowing the idea of alternative update strategy [2, 1, 3, 21, 10], Eq. (3) can be divided into the following five subproblems[1].

Update $\mathcal{Y}$: Given other variables in their previous iteration, we can update $\mathcal{Y}$ by solving the following problem:

$$\begin{aligned} \mathcal{Y}^* = &\operatorname*{argmin}_{\mathcal{Y}} \sum_{v=1}^{V}\Big(\alpha * tr\big(Y^v L^v Y^{v'}\big) + \beta\omega_v\|Y^v - S\|_F^2\Big) \\ &+ \frac{\rho}{2}\Big(\sum_{v=1}^{V}\|X^v - X^v Y^v - E^v + \frac{\Theta^v}{\rho}\|_F^2 + \|\mathcal{Z} - \mathcal{Y} + \frac{\Pi}{\rho}\|_F^2\Big). \end{aligned} \tag{4}$$

Actually, Eq. (4) can be separated into V independent minimization problems and the v-th minimization problem is

$$\begin{aligned} Y^{v^*} = &\operatorname*{argmin}_{Y^v} \alpha * tr\big(Y^v L^v Y^{v'}\big) + \beta\omega_v\|Y^v - S\|_F^2 \\ &+ \frac{\rho}{2}\Big(\|X^v Y^v - A^v\|_F^2 + \|Y^v - B^v\|_F^2\Big), \end{aligned} \tag{5}$$

where $A^v = X^v - E^v + \frac{\Theta^v}{\rho}$ and $B^v = Z^v + \frac{\Pi^v}{\rho}$. By setting the derivative of Eq. (5) with respect to Y^v to zero, we can yield a Sylvester equation $M * Y^v + Y^v * N = C$, where $M = (2\beta\omega_v + \rho)I + \rho X^{v'} X^v$, $N = \alpha(L^v + L^{v'})$, and

[1]For simplicity, the iteration number k is omitted in the updates of all variables.

$C = 2\beta\omega_v S + \rho\Big(X^{v'}A^v + B^v\Big)$. Then, the optimal solution of Y^v can be obtained by

$$Y^{v^*} = lyap(M, N, -C), \tag{6}$$

where "$lyap(M, N, -C)$" is the Matlab function to efficiently solve the Sylvester equation.

Update $\mathcal{Z}$: Fixing other variables in their previous iteration, $\mathcal{Z}$ can be updated by

$$\mathcal{Z}^* = \underset{\mathcal{Z}=\mathcal{C}\times_1 U_1\times_2 U_2\times_3 U_3,\ U_i'*U_i=I}{\operatorname{argmin}} \|\mathcal{Z} - (\mathcal{Y} - \frac{\Pi}{\rho})\|_F^2, \tag{7}$$

which can be equivalently transformed into the following optimization problem:

$$\mathcal{Z}^* = \underset{U_i'*U_i=I}{\operatorname{argmin}} \|\mathcal{C} \times_1 U_1 \times_2 U_2 \times_3 U_3 - (\mathcal{Y} - \frac{\Pi}{\rho})\|_F^2. \tag{8}$$

The above problem can be easily solved by the HOOI algorithm [22] to obtain the core tensor $\mathcal{C}$ and the orthogonal factor matrices U_i. Then, the low-rank representation tensor $\mathcal{Z}^*$ is obtained by

$$\mathcal{Z}^* = \mathcal{C} \times_1 U_1 \times_2 U_2 \times_3 U_3. \tag{9}$$

Update E: Minimizing the augmented Lagrangian function Eq. (3) with respect to E, we have

$$\begin{aligned} E^* &= \underset{E}{\operatorname{argmin}} \|E\|_{2,1} + \frac{\rho}{2}\sum_{v=1}^{V}\|E^v - F^v\|_F^2 \\ &= \underset{E}{\operatorname{argmin}} \frac{1}{\rho}\|E\|_{2,1} + \frac{1}{2}\|E - F\|_F^2, \end{aligned} \tag{10}$$

where $F^v = X^v - X^vY^v + \frac{\Theta^v}{\rho}$ and F is constructed by vertically concatenating the matrices $\{F^v\}$ along the column. The j-th column of optical solution E^* can be obtained by

$$E^*(:,j) = \begin{cases} \frac{\|F(:,j)\|_2 - \frac{1}{\rho}}{\|F(:,j)\|_2}F(:,j), & if\ \frac{1}{\rho} < \|F(:,j)\|_2; \\ 0, & otherwise. \end{cases} \tag{11}$$

Update S: To obtain the optimal solution S^*, we can minimize the augmented Lagrangian function Eq. (3) with respect to S as

$$S^* = \underset{S}{\operatorname{argmin}} \sum_{v=1}^{V}\omega_v\|Y^v - S\|_F^2. \tag{12}$$

We also set the derivative of Eq. (12) with respect to S to zero. The closed-form solution S^* is

$$S^* = \frac{\sum_v \omega_v Y^v}{\sum_v \omega_v} = \sum_v \omega_v Y^v, \tag{13}$$

which is based on the constraint $\sum_v \omega_v = 1$.

Update ω: The optimization of ω is transformed into the following problem

$$\begin{aligned} \omega^* &= \underset{\omega}{\operatorname{argmin}} \sum_{v=1}^{V}\omega_v a^v + \gamma\|\omega\|_2^2, \\ s.t.\ &\omega \geq 0,\ \Sigma_v\omega_v = 1, \end{aligned} \tag{14}$$

where $\gamma\|\omega\|_2^2$ is used to avoid the futile solution [3] and $a^v = \|Y^v - S\|_F^2$. Then, Eq. (14) can be rewritten into the following quadratic programming formulation

$$\begin{aligned} \omega^* &= \underset{\omega}{\operatorname{argmin}} \|\omega + \frac{a}{2\gamma}\|_2^2, \\ s.t.\ &\omega \geq 0,\ \Sigma_v\omega_v = 1. \end{aligned} \tag{15}$$

Here, a Matlab function *i.e.*, *quadprog*, is exploited to solve the above problem.

Update Θ, Π, and ρ: The Lagrangian multipliers Θ, Π and the penalty parameter ρ can be updated by

$$\begin{aligned} \Theta^{v^*} &= \Theta^v + \rho(X^v - X^vY^v - E^v); \\ \Pi^* &= \Pi + \rho(\mathcal{Z} - \mathcal{Y}); \\ \rho^* &= \min\{\lambda * \rho, \rho_{max}\}. \end{aligned} \tag{16}$$

where $\lambda > 1$ is to facilitate the convergence speed [23, 24]. ρ_{max} is the max value of the penalty parameter ρ. The whole procedure of Eq. (1) is summarized in Algorithm 1.

Algorithm 1 GLTA for multi-view clustering

Input: multi-view features: $\{X^v\}$; parameters: $\alpha,\ \beta,\ \gamma = 10$; nearest neighbors number 5; cluster number K;
Initialize: $\mathcal{Y},\ \mathcal{Z},\ E,\ S,\ \Theta,\ \Pi$ initialized to **0**; weight $\omega_v = \frac{1}{V}$; $\rho = 10^{-3}$, $\lambda = 1.5$, $\epsilon = 10^{-7}$;
1: **while** not converged **do**
2: **for** $v = 1$ to V **do**
3: Update Y^{v^*} by Eq. (6);
4: **end for**
5: Update $\mathcal{Z}^*$, E^*, S^*, ω^*, Θ^{v^*}, Π^*, and ρ^* by Eqs. (9), (11), (13), (15), and (16), respectively;
6: Check the convergence condition
7: $\sum_v \|X^v - X^vY^{v^*} - E^{v^*}\|_F / \sum_v \|X^v\|_F \leq \epsilon$;
8: **end while**
Output: Affinity matrix S^*.

3. EXPERIMENTAL RESULTS

To verify the effectiveness of the proposed GLTA, in this section, we first conduct experiments on six real datasets over six state-of-the-art clustering methods. Then, to provide a comprehensive study of GLTA, we analyze the proposed model with respect to two important parameters and report the empirical convergence of GLTA.

3.1. Experimental settings

(1) Datasets: Following [3], we evaluate the performance of GLTA on four news store datasets: BBC4view, BBCSport[2], 20news[3], 3Sources[4]; one article dataset: Wikipedia[5], and one

[2]http://mlg.ucd.ie/datasets/segment.html
[3]http://lig-membres.imag.fr/grimal/data.html
[4]http://mlg.ucd.ie/datasets/3sources.html
[5]http://www.svcl.ucsd.edu/projects/crossmodal/

Table 1. Details summary of the six real multi-view datasets.

Category	Dataset	Instance	View	cluster
News store	BBC4view	685	4	5
	BBCSport	544	2	5
	20news	500	3	5
	3Sources	169	3	6
Article	Wikipedia	693	2	10
Generic Object	COIL_20	1440	3	20

generic object dataset: COIL_20[6]. The statistics of these datasets are summarized in Table 1.

(2) Compared methods and evaluation measures: We compared GLTA with six state-of-the-art methods, including $\mathbf{SSC_{best}}$ [14]: the l_1-norm regularized representation matrix construction with the most informative view; $\mathbf{LRR_{best}}$ [13]: the nuclear norm regularized representation matrix construction with the most informative view; **DiMSC** [1]: multi-view subspace clustering with the Hilbert-Schmidt Independence criterion; **LT-MSC** [2]: multi-view subspace clustering by low-rank tensor constraint; **MVCC** [3]: multi-view clustering via concept factorization with local manifold regularization; **t-SVD-MSC** [4]: multi-view clustering via tensor multi-rank minimization. The first two methods belong to single-view clustering baselines, while, the others belong to multi-view clustering ones.

For $\mathbf{SSC_{best}}$ and $\mathbf{LRR_{best}}$, each single feature is independently used and the best single view clustering result is reported. For **DiMSC**, **LT-MSC**, and **t-SVD-MSC**, they all need to firstly learn the representation matrix or tensor and then construct the affinity matrix. For all methods, the spectral clustering algorithm is performed to find the clustering result.

Following [2, 3, 7], we exploit six popular clustering measures, *i.e.*, accuracy (ACC), normalized mutual information (NMI), adjusted rank index (AR), F-score (F-s), Precision (Pre), and Recall (Re), to evaluate the clustering performance. Generally, the higher values these six measures have, the better the clustering quality is. Since the clustering step is based on K-means for all methods and different initializations may yield different results, we run 10 trials for each experiment and report their average performance to avoid the randomness perturbation.

3.2. Clustering performance comparison

All clustering results on six datasets are reported in Table 2. The best results are highlighted in boldface with respect to each dataset. We can observe that (1) multi-view clustering approaches DiMSC, t-SVD-MSC, and our GLTA achieve better performance than the single-view clustering approaches SSC_{best} and LRR_{best}. This is mainly because the high-order cross information among multiple views is well captured by DiMSC, t-SVD-MSC, and our GLTA; (2) LT-MSC achieves

[6]http://www.cs.columbia.edu/CAVE/software/softlib/

Table 2. Mean clustering results on all real datasets.

Method	BBC4view					
	ACC	NMI	AR	F-s	Pre	Re
$\text{SSC}_{\mathbf{best}}$	0.660	0.494	0.470	0.599	0.578	0.622
$\text{LRR}_{\mathbf{best}}$	0.802	0.560	0.621	0.712	0.697	0.727
DiMSC	0.892	0.728	0.752	0.810	0.811	0.810
LT-MSC	0.591	0.442	0.400	0.546	0.525	0.570
MVCC	0.745	0.587	0.550	0.656	0.654	0.658
t-SVD-MSC	0.858	0.685	0.725	0.789	0.800	0.778
GLTA	**0.910**	**0.771**	**0.810**	**0.854**	**0.864**	**0.845**
	BBCSport					
$\text{SSC}_{\mathbf{best}}$	0.627	0.534	0.364	0.565	0.427	0.834
$\text{LRR}_{\mathbf{best}}$	0.836	0.698	0.705	0.776	0.768	0.784
DiMSC	0.922	0.785	0.813	0.858	0.846	0.872
LT-MSC	0.460	0.222	0.167	0.428	0.328	0.629
MVCC	0.928	0.816	0.831	0.870	0.889	0.853
t-SVD-MSC	0.879	0.765	0.784	0.834	0.863	0.807
GLTA	**0.939**	**0.825**	**0.849**	**0.885**	**0.890**	**0.880**
	20news					
$\text{SSC}_{\mathbf{best}}$	0.545	0.461	0.289	0.481	0.357	0.749
$\text{LRR}_{\mathbf{best}}$	0.804	0.578	0.568	0.655	0.645	0.665
DiMSC	0.978	0.937	0.945	0.956	0.954	0.957
LT-MSC	0.990	0.965	0.975	0.979	0.979	0.980
MVCC	0.891	0.766	0.763	0.810	0.808	0.813
t-SVD-MSC	0.992	0.972	0.980	0.984	0.984	0.984
GLTA	**0.996**	**0.983**	**0.990**	**0.993**	**0.996**	**0.990**
	3Sources					
$\text{SSC}_{\mathbf{best}}$	0.762	0.694	0.658	0.743	0.769	0.719
$\text{LRR}_{\mathbf{best}}$	0.647	0.542	0.486	0.608	0.594	0.636
DiMSC	0.795	0.727	0.661	**0.748**	0.711	0.788
LT-MSC	0.781	0.698	0.651	0.734	0.716	0.754
MVCC	0.761	0.698	0.626	0.731	0.607	**0.916**
t-SVD-MSC	0.781	0.678	0.658	0.745	0.683	0.818
GLTA	**0.846**	**0.728**	**0.665**	0.736	**0.805**	0.678
	Wikipedia					
$\text{SSC}_{\mathbf{best}}$	0.561	0.527	0.418	0.481	0.491	0.471
$\text{LRR}_{\mathbf{best}}$	0.554	0.523	0.417	0.479	0.490	0.468
DiMSC	0.547	0.500	0.397	0.461	0.478	0.445
LT-MSC	0.532	0.496	0.407	0.471	0.480	0.461
MVCC	**0.600**	0.515	0.434	0.494	0.513	0.476
t-SVD-MSC	0.527	0.480	0.393	0.458	0.470	0.447
GLTA	0.595	**0.549**	**0.455**	**0.514**	**0.522**	**0.506**
	COIL_20					
$\text{SSC}_{\mathbf{best}}$	0.803	0.935	0.798	0.809	0.734	0.804
$\text{LRR}_{\mathbf{best}}$	0.761	0.829	0.720	0.734	0.717	0.751
DiMSC	0.778	0.846	0.732	0.745	0.739	0.751
LT-MSC	0.804	0.860	0.748	0.760	0.741	0.776
MVCC	0.732	0.845	0.675	0.692	0.647	0.744
t-SVD-MSC	0.830	0.884	0.786	0.800	0.785	0.808
GLTA	**0.878**	**0.945**	**0.869**	**0.875**	**0.856**	**0.895**

Table 3. Comparison among different view feature by LRR.

Index	20news/3Sources		
	View 1	View 2	View 3
ACC	0.804/0.580	0.684/0.647	0.479/0.618
NMI	0.578/0.516	0.454/0.542	0.276/0.511

unsatisfactory results on the first two datasets which may come from the fact that the unfolding-based tensor nuclear norm is a loose surrogate of Tucker rank; (3) the performance of GLTA is better than or comparable to that of all competing methods in most cases, especially on the 3Source and COIL_20 datasets. The main reason is that [1, 2, 4] both construct the low-rank representation matrix or tensor and the affinity matrix in two separated steps without consideration of the various contributions of different features. Besides, the promising performance of GLTA also benefits from the preservation of the local geometrical structures.

3.3. Model analysis

In this section, we aim to present a comprehensive study of the proposed GLTA. We first explain the necessity of the last term (the consensus representation) in Eq. (1) and then analyze the parameters and empirical convergence of the proposed GLTA.

(1) The necessity of the consensus representation term: The clustering results by LRR [13] on each view feature are reported in Table 3. From this table, we can see that the values of ACC and NMI on 20news vary from 47.9% to 80.4% and 27.6% to 57.8%, respectively. For 3Sources, differences among these three views with respect to ACC and NMI are 6.7% and 2.6%, respectively. Therefore, we can draw a conclusion that different features have various contributions for clustering results, which is the fundamental motivation of this paper. It is of vital importance to fully consider the various contributions of different features in the multi-view clustering procedure.

(2) Parameter selection: We set the number of nearest neighbors as 5 and $\gamma = 10$ for all experiments. Two free parameters α and β should be tuned. Specifically, they are empirically selected from the sets of $[0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.4, 0.5, 1, 2, 5, 10, 50, 100, 500]$ and $[0.01, 0.1, 0.5, 1, 3, 5, 7, 10, 50, 100]$, respectively. Due to page limitation, we only show the ACC and NMI results of our GLTA with different combinations of α and β on 3Sources dataset in Fig. 1. From this figure, we can see that with smaller α and β, GLTA is inadequate for clustering. The main reason is that when α and β are too small, the local manifolds and the consensus representation terms of Eq. (1) have mere contribution to the construction of the affinity matrix. When α and β are selected from the interval $[0.05,\ 0.4]$ and $[1,\ 10]$ respectively, our GLTA can yield promising results.

(3) The effect of $\mathcal{Z}$, S: Recall that the proposed GLTA

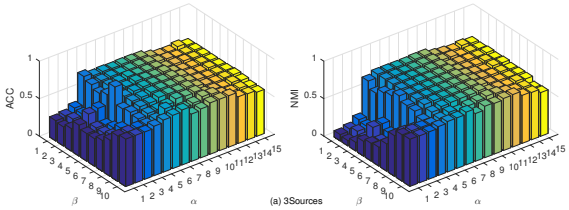

Fig. 1. ACC and NMI values of our GLTA with different combinations of α and β on 3Sources.

can learn the representation tensor $\mathcal{Z}$ and the affinity matrix S simultaneously. Following [1, 2, 4], once the representation tensor $\mathcal{Z}$ is learned, the affinity matrix, *i.e.*, $\Sigma_v\left(|\mathcal{Z}^v| + |\mathcal{Z}^{v'}|\right)$ can be constructed as the input of spectral clustering algorithm to segment all data points into K clusters. However, the above scheme fails to consider the various contributions of different features. To validate this, we also report the clustering results using the representation tensor $\mathcal{Z}$ in Table 4.

Note that we can approximately regard $\mathcal{Z}$ obtained by setting $\beta = 0$, which means that the representation tensor is encoded by the Tucker decomposition and local manifolds without the consideration of various contributions of different features. From Table 4, it can be seen that the clustering results using the affinity matrix S are dramatically higher than those using $\mathcal{Z}$. This directly verifies the importance of the consensus representation in Eq. (1).

Table 4. Comparison of the clustering performance of $\mathcal{Z}, S$.

	ACC / NMI		
	3Sources	BBC4view	BBCSport
$\mathcal{Z}$	0.654/0.592	0.875/0.704	0.892/0.782
S	**0.846/0.728**	**0.910/0.771**	**0.939/0.825**

(4) Convergence analysis: It is intractable to derive the theoretical convergence proof of the proposed GLTA. Instead, we just provide the empirical convergence analysis in Fig. 2, in which the vertical axis denotes the error defined as $\sum_v \|X^v - X^v Y^{v^*} - E^{v^*}\|_F / \sum_v \|X^v\|_F$. After 20 iterations, the error yields a stable value, which means that GLTA can converge within a few iterations.

4. CONCLUSIONS

In this paper, we developed a novel method, named graph regularized low-rank tensor representation and affinity matrix (GLTA) for multi-view clustering. GLTA can learn the low-rank representation tensor, which is encoded by the Tucker decomposition and the local manifolds, and the affinity matrix, which is constructed by assigning different weights on different features, simultaneously. Extensive experiments on

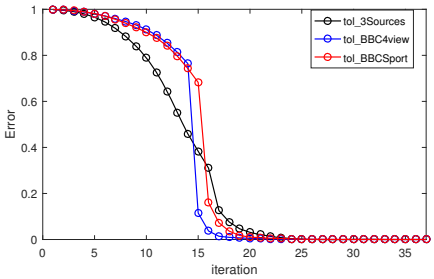

Fig. 2. Convergence versus iteration.

six real datasets demonstrate that our method outperforms the state-of-the-arts.

5. REFERENCES

[1] Xiaochun Cao, Changqing Zhang, Huazhu Fu, Si Liu, and Hua Zhang, "Diversity-induced multi-view subspace clustering," in *CVPR*, 2015, pp. 586–594.

[2] Changqing Zhang, Huazhu Fu, Si Liu, Guangcan Liu, and Xiaochun Cao, "Low-rank tensor constrained multiview subspace clustering," in *ICCV*, 2015, pp. 1582–1590.

[3] Hao Wang, Yan Yang, and Tianrui Li, "Multi-view clustering via concept factorization with local manifold regularization," in *ICDM*, 2016, pp. 1245–1250.

[4] Yuan Xie, Dacheng Tao, Wensheng Zhang, Yan Liu, Lei Zhang, and Yanyun Qu, "On unifying multi-view self-representations for clustering by tensor multi-rank minimization," *IJCV*, vol. 126, no. 11, pp. 1157–1179, 2018.

[5] Chang Xu, Dacheng Tao, and Chao Xu, "A survey on multi-view learning," *arXiv preprint arXiv:1304.5634*, 2013.

[6] Canyi Lu, Shuicheng Yan, and Zhouchen Lin, "Convex sparse spectral clustering: Single-view to multi-view," *IEEE TIP*, vol. 25, no. 6, pp. 2833–2843, 2016.

[7] Rongkai Xia, Yan Pan, Lei Du, and Jian Yin, "Robust multi-view spectral clustering via low-rank and sparse decomposition.," in *AAAI*, 2014, pp. 2149–2155.

[8] Jianlong Wu, Zhouchen Lin, and Hongbin Zha, "Essential tensor learning for multi-view spectral clustering," *arXiv preprint arXiv:1807.03602*, 2018.

[9] Zemin Zhang, Gregory Ely, Shuchin Aeron, Ning Hao, and Misha Kilmer, "Novel methods for multilinear data completion and de-noising based on tensor-svd," in *CVPR*, 2014, pp. 3842–3849.

[10] Yongyong Chen, Shuqin Wang, and Yicong Zhou, "Tensor nuclear norm-based low-rank approximation with total variation regularization," *IEEE JSTSP*, vol. 12, no. 6, pp. 1364–1377, 2018.

[11] Yuting Su, Xu Bai, Wu Li, Peiguang Jing, Jing Zhang, and Jing Liu, "Graph regularized low-rank tensor representation for feature selection," *J. Vis. Commun. Image R.*, vol. 56, pp. 234–244, 2018.

[12] Lance Parsons, Ehtesham Haque, and Huan Liu, "Subspace clustering for high dimensional data: a review," *ACM SIGKDD Explorations Newsletter*, vol. 6, no. 1, pp. 90–105, 2004.

[13] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE TPAMI*, vol. 35, no. 1, pp. 171–184, 2013.

[14] Ehsan Elhamifar and Rene Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE TPAMI*, vol. 35, no. 11, pp. 2765–2781, 2013.

[15] Yongyong Chen, Yongli Wang, Mingqiang Li, and Guoping He, "Augmented lagrangian alternating direction method for low-rank minimization via non-convex approximation," *Signal, Image and Video Processing*, pp. 1–8, 2017.

[16] Ji Liu, Przemyslaw Musialski, Peter Wonka, and Jieping Ye, "Tensor completion for estimating missing values in visual data," *IEEE TPAMI*, vol. 35, no. 1, pp. 208–220, 2013.

[17] Canyi Lu, Jiashi Feng, Yudong Chen, Wei Liu, Zhouchen Lin, and Shuicheng Yan, "Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization," in *CVPR*, 2016, pp. 5249–5257.

[18] Chong Peng, Zhao Kang, Yunhong Hu, Jie Cheng, and Qiang Cheng, "Robust graph regularized nonnegative matrix factorization for clustering," *ACM TKDD*, vol. 11, no. 3, pp. 33, 2017.

[19] Ming Yin, Junbin Gao, and Zhouchen Lin, "Laplacian regularized low-rank representation and its applications," *IEEE TPAMI*, vol. 38, no. 3, pp. 504–517, 2016.

[20] Feiping Nie, Jing Li, and Xuelong Li, "Self-weighted multiview clustering with multiple graphs," in *IJCAI*, 2017, pp. 2564–2570.

[21] Yongyong Chen, Yanwen Guo, Yongli Wang, Dong Wang, Chong Peng, and Guoping He, "Denoising of hyperspectral images using nonconvex low rank matrix approximation," *IEEE TGRS*, vol. 55, no. 9, pp. 5366–5380, 2017.

[22] Tamara G Kolda and Brett W Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[23] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends® Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[24] Shuqin Wang, Yongli Wang, Yongyong Chen, Peng Pan, Zhipeng Sun, and Guoping He, "Robust PCA using matrix factorization for background/foreground separation," *IEEE Access*, vol. 6, pp. 18945–18953, 2018.