Multiview Clustering based on Robust and Regularized Matrix Approximation

Jiameng Pu*, Qian Zhang†, Lefei Zhang‡, Bo Du*, and Jane You‡
*State Key Laboratory of Software Engineering, School of Computer, Wuhan University, Wuhan 430079, China
{jpu, remoteking}@whu.edu.cn
† Beijing Samsung Telecom R&D Center, Beijing 100028, China, q0712.zhang@samsuang.com
‡Department of Computing, The Hong Kong Polytechnic University, Hong Kong, {cslfzhang, csyjia}@comp.polyu.edu.hk

***Abstract*—Pattern recognition tasks such as the data classification and clustering usually can be represented by the perspective of multiple views or feature spaces. Obviously, the accuracy of the classification and clustering should be greatly improved if we carefully consider the discriminabilities from multiple views and explore the complementary information among them. However, multiple features also bring new challenges to handle them. In the literature, many existed multiview feature learning methods dealt with different views equally, thus they couldn't optimally utilize the complementary property of them. On the other hand, the matrix factorization based clustering algorithms usually adopt the conventional ℓ_2-norm based squared residue minimization to measure the loss, which is easily influenced by the outliers and noises from the multiple sources of input. In this paper, we propose a novel multiview data clustering algorithm based on the matrix factorization to relieve the above issues. The basic idea of the proposed Robust and Regularized Matrix Approximation (RRMA) is that the observed data matrix could be low-rank approximated by a cluster centroid matrix and a cluster indicator matrix, respectively, and the major contributions of our work lie in the introduction of the robust $\ell_{2,1}$-norm and ensemble manifold regularization to regularize the matrix factorization and make the model more discriminative for multiview data clustering. We properly adjust the importance of different views by assigning a set of trainable weights on the views. Moreover, we propose an efficient solution featured with impactful updating rules to seek the local optimal parameters. Encouraging experimental results on numerous public multiview datasets demonstrate the superiority of our model compared to some state-of-the-art methods.**

I. Introduction

Unlabeled data is common and plentiful in modern life. It has been reported that about 72 hours of videos every minute and 6 billion new photos every month are uploaded on the YouTube and Facebook, respectively. Unsupervised learning methods which learn certain knowledge from these unlabeled data such as clustering, have attracted great interest of researchers in recent years. A great amount of clustering algorithms have been proposed and shown encouraging performances, e.g., the k-means and nonnegative matrix factorization (NMF) [1]. However, these methods are mainly developed for data clustering from a single view of feature representation.

With the rapid development of technology and the increasing amount of information, in many pattern recognition tasks, the users are able to collect data samples with multiple sources and from various perspectives [2][3][4]. For example, both the color and texture are useful feature representations for the task of image classification, while in the human identification, we may consider the rich features from the multiple sources such as face, iris, fingerprint, and palm print. In daily life people can get access to news covering the same social events from the BBC, the Reuters, the Guardian and many different media [5]. However, the conventional clustering algorithms assume that samples are extracted from a single feature space, thus they cannot deal with data from multiple views direcly. Although a straightforward solution to address this problem is to concatenate all the views and then apply a single-view clustering algorithm, however, due to the complementary and redundant properties of features from different views, the concatenation is not physical meaningful and the improvement is non-significant in most cases [6][7].

To overcome the drawbacks of single-view clustering algorithms, growing multiview algorithms had been proposed [8][9][10][11]. Since the multiview clustering are more challenging than single-view clustering with more complicated data to handle, the main challenge in multiview clustering is how to properly address the complementary as well as redundant proprieties to partition samples.

For instance, Wang proposed a constrained spectral clustering (CSP-P) [12], which took one view as the similarity matrix and incorporated the other view as the constraints, but it was only limited to two-view data. Kumar *et al.* [13][14] proposed two spectral clustering frameworks which utilized co-regularization and co-training respectively to make the eigenvectors agreed on all the views. However, these two spectral-based algorithms could only achieve an implicit clustering consistency by merely pushing consistence eigenvectors across different views [15]. Recently, some multiview clustering algorithms based on NMF, which was an effective technique firstly applied in the conventional data analysis and then showed superiority in the single-view clustering [16], had been proposed. Liu *et al.* [17] proposed a MultiNMF, which suggested a joint matrix factorization process but with the constraint that constrained clustering solution of each view towards a common consensus. Cai *et al.* [18] proposed a robust multiview k-means model, in which the relaxed k-means was actually equivalent to G-orthogonal nonnegative matrix factorization. However, most existing NMF-based clustering algorithms cannot fully address the following problems: 1) The model's performance is easily influenced by the outliers and noises in the input data [19]. 2) Since NMF learns a part-based representation and implements the factorization in the Euclidean feature space, The model fails to uncover the intrinsic geometrical structure of the multiview data [20][21]. 3) The importance of each view is not properly considered. Since different views provide complementary as well as redundant proprieties, it's not appropriate to encode the intrinsic structure of each view equally. In [17], the importance of different view was manually tuned, which indicated the performance of algorithm was limited by prior knowledge of views' qualities.

In order to fully address the above issues, we propose a robust and regularized matrix approximation (RRMA), which adopts $\ell_{2,1}$-norm to improve the model's robustness and encodes the geometrical structure of real-world data with manifold regularization. Specifically, we approximate the original feature matrix represented from multiple views into the product of cluster centroid matrix by the cluster indicator matrix which derives the final clustering label. We balance the importance of each view by maintaining a set of learnable weights for each view in the manifold regularization, hence features from different views may have the optimal complementarity to each other. Furthermore, we propose an elegant and efficient optimization procedure for solve the objective function.

The remainder of this paper is organized as follows: Section 2 provides the objective formulation of RRMA in detail. In Section 3, we propose an efficient optimization procedure for RRMA. Then the experimental results of RRMA compared with some state-of-the-art clustering approaches on typical datasets are reported in Section 4, followed by our conclusions in Section 5.

II. RRMA Model

In this section, we formulate the Robust and Regularized Matrix Approximation (RRMA) algorithm. Assume that we have got M views and let $X^{(m)}$ and $V^{(m)}$ represent the original feature matrix and the cluster centroid for the m-th view (m = [1, . . . , M]), respectively. In RRMA, we expect that each view will have an unique cluster indicator as represented by U. Let α donates a balanced parameter and $\beta^{(m)}$ controls the weight of the m-th view. The objective function of RRMA algorithm is shown in Eq. 1.

$$\begin{aligned} \arg\min_{U,V^{(m)},\beta} \sum_{m=1}^{M} \left\| X^{(m)} - V^{(m)}U^{\mathrm{T}} \right\|_{2,1} \\ + \alpha \sum_{m=1}^{M} (\beta_m)^r \mathrm{tr}(U^{\mathrm{T}} L^{(m)} U) \\ \text{s.t.} U^{\mathrm{T}}U = I, U \geq 0 \end{aligned} \tag{1}$$

In the following, we describe our model in detail. We first formulate the objective function of the basic NMF for multiview clustering, then add two additional regularization terms to the main objective function, i.e., the multiview matrix approximation via $\ell_{2,1}$-norm and the multiview manifold regularization to form the robust and regularized matrix approximation.

A. Multiview NMF

Given that we have a data matrix $X (X \in \mathrm{R}^{p\times n}) = [x_1, x_2, \ldots, x_n]$ consists of n data column vectors, NMF aims to decompose X into two nonnegative matrices, whose product can share the maximum degree of similarity with the original matrix X. The matrix V and U have lower rank compared with the matrix X. We obtain the optimal V, U by minimizing the following cost function:

$$J = \min \left\| X - VU^{\mathrm{T}} \right\|_F^2 \qquad \text{s.t.} U, V \geq 0 \tag{2}$$

Then we logically extend this standard NMF to a multiview version, in which the original matrices from each view share the same low-rank representation U:

$$\sum_{m=1}^{M} \left\| X^{(m)} - V^{(m)}U^{\mathrm{T}} \right\|_F^2 \tag{3}$$

B. Multiview matrix approximation via $\ell_{2,1}$-norm

In the formulations Eq. 2 and Eq. 3, the error for each data point is a squared residue error, so the algorithms can be easily affected by the noises and outliers with large errors. In this part, we employ the $\ell_{2,1}$-norm to overcome the unstable property of the standard NMF, and note that the $\ell_{2,1}$-norm of X is defined as [22][23]:

$$\|X\|_{2,1} = \sum\nolimits_{i=1}^{n} \sqrt{\sum\nolimits_{j=1}^{p} X_{ji}^2} = \sum\nolimits_{i=1}^{n} \|x_i\| \tag{4}$$

where n is the number of samples and p is the data dimension. Then we have:

$$\begin{aligned} \left\| X - VU^{\mathrm{T}} \right\|_{2,1} &= \sum_{i=1}^{n} \sqrt{\sum_{j=1}^{p} (X - VU^{\mathrm{T}})_{ji}^2} \\ &= \sum_{i=1}^{n} \left\| x_i - VU_i^{\mathrm{T}} \right\| \end{aligned} \tag{5}$$

Then the objective function becomes:

$$\begin{aligned} &\sum_{m=1}^{M} \left\| X^{(m)} - V^{(m)}U^{\mathrm{T}} \right\|_{2,1} \\ &= \sum_{m=1}^{M} \sum_{i=1}^{n} \sqrt{\sum_{j=1}^{p} (X^{(m)} - V^{(m)}U^{\mathrm{T}})_{ji}^2} \\ &= \sum_{m=1}^{M} \sum_{i=1}^{n} \left\| x_i^{(m)} - V^{(m)}U_i^{\mathrm{T}} \right\| \end{aligned} \tag{6}$$

By employing the $\ell_{2,1}$-norm, our method is no longer sensitive to outliers inside the data and in the meanwhile maintains the rotation invariance property within data vectors.

C. Multiview manifold regularized matrix approximation

In fact, different views featured with distinct physical and statistical characteristics are complementary for each other, thus the algorithm should incorporate the geometrical correlation among different views as much as possible. In RRMA, we employ the multiview manifold regularization to regularize our matrix approximation to incorporate the intrinsic and nonlinear structure of data across all different views.

Based on the previous formulations, we define the undirected graph of the m-th view as $G^{(m)} = \{X^{(m)}, W^{(m)}\}$, in which $X^{(m)}$ is the original feature matrix of the m-th view and the relation matrix $W^{(m)}$ is weighted by the heat kernels [24]. The formulation of $W^{(m)}$ is defined as following:

$$W_{ij}^{(m)} = \begin{cases} e^{(-\left\|x_i^{(m)} - x_j^{(m)}\right\|/t)}, \\ \qquad x_i^{(m)} \in N(x_j^{(m)}) \text{or} x_j^{(m)} \in N(x_i^{(m)}) \\ 0, \text{otherwise} \end{cases} \tag{7}$$

where t is the parameter in Gaussian function and $N(x_i^{(m)})$ is the k-nearest-neighbors set of data sample $x_i^{(m)}$. With the support of previous patch alignment framework [25], the manifold structure of data in the m-th feature space can be maximally preserved in the U through the Eq. 8. In this formulation, $L^{(m)} = D^{(m)} - W^{(m)}$ is a Laplacian matrix of m-th view, note that $D^{(m)}$ is a diagonal matrix and its entries are column sums of $W^{(m)}$ [26].

$$\arg\min_U \sum_{i \neq j} W_{ij}^{(m)} \left\|u_i - u_j\right\|_2^2 = \arg\min_U \text{tr}(U^{\mathrm{T}} L^{(m)} U) \tag{8}$$

Then we extend this manifold regularization to a multiview version and obtain the formulation of multiview manifold regularization as following:

$$\alpha \sum_{m=1}^{M} (\beta_m)^r \text{tr}(U^{\mathrm{T}} L^{(m)} U) \tag{9}$$

Thus, we can see that the factor matrix U, which is the cluster indicator in our matrix approximation framework, actually can also be understood as a special low dimensional feature representation of the original feature matrices. In Eq. 9, apart from trace operator tr, $\beta = [\beta_1, \beta_2, ..., \beta_m](\beta > 0$ and $\sum_{m=1}^{M} \beta_m = 1)$ is a set of nonnegative weight parameters that control the importance of each view, and scale parameter r controls the weights of multiple features. By weighting the features of different views, proper contribution of each view and their complementary as well as redundant properties are appropriately encoded so that we can get better performance [8][27][28]. In the next section, efficient updating rules are employed in our optimization procedure.

III. Optimization

The objective function in above is not convex in four variables but is convex if we update the $2M + 1$ variables alteratively. Thus, we use an efficient augmented lagrangian method (ALM) to optimize the objective function. By introducing two auxiliary variables $E^{(m)} = X^{(m)} - V^{(m)} U^{\mathrm{T}}$ and $Z_1 = U$, the objective function can be rewritten as the following equivalent formulation:

$$\begin{aligned} \arg \min_{U, V^{(m)}, E_m, Z_1} \sum_{m=1}^{M} \left\|E^{(m)}\right\|_{2,1} + \alpha \sum_{m=1}^{M} (\beta_m)^r \text{tr}(Z_1^{\mathrm{T}} L^{(m)} U) \\ \text{s.t.} E^{(m)} = X^{(m)} - V^{(m)} U^{\mathrm{T}}, Z_1 = U, U^{\mathrm{T}} U = I, Z_1 \geq 0 \end{aligned} \tag{10}$$

which can be solved by solving the following ALM problem:

$$\begin{aligned} \min_{U, V^{(m)}, E^{(m)}, Z_1, \lambda^{(m)}, \mu} \sum_{m=1}^{M} \left\|E^{(m)}\right\|_{2,1} \\ + \alpha \sum_{m=1}^{M} (\beta_m)^r \text{tr}(Z_1^{\mathrm{T}} L^{(m)} U) \\ + \sum_{m=1}^{M} < \lambda^{(m)}, X^{(m)} - V^{(m)} U^{\mathrm{T}} - E^{(m)} > + < \lambda_1, Z_1 - U > \\ + \frac{\mu}{2} (\|Z_1 - U\|_F^2 + \sum_{m=1}^{M} \left\|X^{(m)} - V^{(m)} U^{\mathrm{T}} - E^{(m)}\right\|_F^2) \\ \text{s.t.} U^{\mathrm{T}} U = I, Z_1 \geq 0 \end{aligned} \tag{11}$$

Update $E^{(m)}$: To update $E^{(m)}$, we fixed other variables except $E^{(m)}$ and remove the terms that are irrelevant to $E^{(m)}$. Then the objective function becomes:

$$\min_{E^{(m)}} \frac{1}{2} \left\|E^{(m)} - (X^{(m)} - V^{(m)} U^{\mathrm{T}} + \frac{1}{\mu} \lambda^{(m)})\right\|_F^2 + \frac{1}{\mu} \left\|E^{(m)}\right\|_{2,1} \tag{12}$$

Let $B = X^{(m)} - V^{(m)} U^{\mathrm{T}} + \frac{1}{\mu} \lambda^{(m)}$, then $E^{(m)}$ can be updated as:

$$e_{mi} = \begin{cases} (1 - \frac{1}{\mu \|b_i\|}) b_i, \text{if } \|b_i\| \geq \frac{1}{\mu} \\ 0, \text{otherwise} \end{cases} \tag{13}$$

Update $V^{(m)}$: To update $V^{(m)}$, we fix other variables except $V^{(m)}$, we obtain the following objective function:

$$\min_{V^{(m)}} \frac{\mu}{2} \left\|X^{(m)} - V^{(m)} U^{\mathrm{T}} - E^{(m)} + \frac{1}{\mu} \lambda^{(m)}\right\|_F^2 \tag{14}$$

Considering that $U^{\mathrm{T}} U = I$, we can rewrite the above objective function as:

$$\min_{V^{(m)}} \frac{1}{2} \left\|V^{(m)} - (X^{(m)} - E^{(m)} + \frac{1}{\mu} \lambda^{(m)}) U\right\|_F^2 \tag{15}$$

then $V^{(m)} = (X^{(m)} - E^{(m)} + \frac{1}{\mu} \lambda^{(m)}) U$.

Update Z_1:

$$\begin{aligned} \min_{Z_1 \geq 0} \frac{\mu}{2} \|Z_1 - U\|_F^2 + < \lambda_1, Z_1 - U > \\ + \alpha \sum_{m=1}^{M} (\beta_m)^r \text{tr}(Z_1^{\mathrm{T}} L^{(m)} U) \end{aligned} \tag{16}$$

we obtain:

$$\min_{Z_1 \geq 0} \|Z_1 - K\|_F^2 \tag{17}$$

where $K = (U - \frac{1}{\mu}\lambda_1 - \frac{\alpha}{\mu}\sum_{m=1}^{M}(\beta_m)^r L^{(m)} U)$. The above object function can be further decomposed to element-wise optimization problem as:

$$\min_{Z_{1ij}\geq 0} \|Z_{1ij} - K_{ij}\|_F^2 \tag{18}$$

Therefore, the optimal solution of above problems is:

$$Z_{1ij} = max(K_{ij}, 0) \tag{19}$$

Update U:

$$\begin{aligned} &\min_{U^{\mathrm{T}}U=I} < \lambda_1, Z_1 - U > \\ &+ \sum_{m=1}^{M} < \lambda^{(m)}, X^{(m)} - V^{(m)}U^{\mathrm{T}} - E^{(m)} > \\ &+ \frac{\mu}{2}(\| Z_1 - U \|_F^2 + \sum_{(m=1)}^{M} \left\| X^{(m)} - V^{(m)}U^{\mathrm{T}} - E^{(m)} \right\|_F^2) \\ &+ \alpha \sum_{m=1}^{M} (\beta_m)^r \mathrm{tr}(Z_1^{\mathrm{T}} L^{(m)} U) \end{aligned} \tag{20}$$

removing the irrelevant terms, we have:

$$\min_{U^{\mathrm{T}}U=I} \frac{\mu}{2} \| U \|_F^2 - \mu < H, U > \tag{21}$$

where

$$\begin{aligned} H = &\frac{1}{\mu}\lambda_1 + Z_1 - \frac{\alpha}{\mu}\sum_{m=1}^{M} L Z_1 \\ &+ \sum_{m=1}^{M} (X^{(m)} - E^{(m)} + \frac{1}{\mu} * \lambda^{(m)})^{\mathrm{T}} V^{(m)} \end{aligned} \tag{22}$$

Thus, the objective function is equivalent to:

$$\min_{U^{\mathrm{T}}U=I} \| U - H \|_F^2 \tag{23}$$

Denote:

$$L(U, \Lambda) = \| U - H \|_F^2 + \Lambda(UU^{\mathrm{T}} - I) \tag{24}$$

we then obtain:

$$U = N_u Q_u^{\mathrm{T}} \tag{25}$$

where N_u and Q_u are the left and right singular vectors of the economic singular value decomposition of H.

Update β_m: Denote $p^{(m)} = \mathrm{tr}(Z_1^{\mathrm{T}} L^{(m)} U)$

$$\beta_m = (rp^{(m)})^{\frac{1}{1-r}} / \sum_{m=1} M(rp^{(m)})^{\frac{1}{1-r}} \tag{26}$$

Update $\lambda^{(m)}$,λ_1, **and** μ:

$$\lambda^{(m)} = \lambda^{(m)} + \mu(X^{(m)} - V^{(m)}U^{\mathrm{T}} - E^{(m)}) \tag{27}$$

$$\lambda_1 = \lambda_1 + \mu(Z_1 - U) \tag{28}$$

$$\mu = \rho\mu \tag{29}$$

IV. Experiment

In this section, we first introduce the a few existed data clustering algorithms for comparison and the experimental settings of these methods, then we provide experimental results on the multiple feature dataset as well as some representative and challenging real-world multiview datasets, to evaluate the performance of our method.

A. Comparison algorithms

To test the effectiveness of RRMA algorithm, we show the performance of the following data clustering methods: (1) Stack: we directly stack all the feature vectors and perform k-means algorithm for clustering; (2) MSE [8]: we firstly perform MSE on the multiple features and then employ the k-means algorithm for clustering; (3) Canonical Correlation Analysis: we firstly perform CCA under the Generalized Multiview Analysis (GMA) framework [29] and then apply k-means algorithm; (4) ConcatRMNMF: we run RMNMF (single-view clustering algorithm) [22] on the concatenated feature representation; and we also compare RRMA with the state-of-the-art multiview clustering learning algorithms: (5) Co-regspectral [13], (6) Co-trainspectral [14], and (7) Constrained multiview spectral clustering (CSP-P, which only limited to two-view datasets [12]), and (8) MultiNMF [17].

B. Experimental Settings

In the experiment, some algorithms need to perform k-means after processing multiview feature. When performing k-means, we repeat clustering process for 20 times with random initializations and then report the average results with standard deviations. However, since our proposed RRMA directly outputs clustering result without any random procedure, which indicates our RRMA method could be strictly reproduced, we only need to set the appropriate parameter ranges and then find the optimal clustering result. In RRMA, we set the parameter range of α as $10^{[-4,-3,-2,-1,0,1,2,3,4]}$ and r as $10^{[-4,-3,-2,-1,0,1,2,3,4]}$. Evaluation accuracy (ACC) is utilized in the experiment to test the performance and the code can be found on the website[1].

C. Experimental results on Multiple Feature Dataset

In this subsection, to demonstrate how clustering performance can be improved by our RRMA, we evaluate the clustering algorithms on the Multiple Feature Dataset [2], which consists of features of 2000 handwritten digits (from 0 to 9) extracted from a collection of Dutch utility maps. In this dataset, there are two hundred examples per digit in biniary images (from 0 to 9) and six feature sets which respectively represent different views of digit. We utilize four different feature views: Mfeatkar, Mfeatpix, Mfeatmor, Mfeatzer. To make experiment more objective and convincing, we randomly pick two views out of aforementioned views to construct view X and view Y respectively, so that there are 6 pairs of two-view datasets in total. We summarize the experimental results

[1]http://www.cad.zju.edu.cn/home/dengcai/Data/Clustering.html
[2]http://archive.ics.uci.edu/ml/machine-learning-databases/mfeat/

TABLE II
STATISTICS OF THE FIVE REAL-WORLD DATASETS

Dataset	size	view	cluster
Wiki	2866	2	10
Cora	2708	2	7
Citeseer	3312	2	6
3sources	169	3	6
Synthetic	1000	3	2

of different methods on Multiple Feature Dataset in Table I by the measurement of ACC. We make the following observations: our RRMA outperforms other clustering methods under the most circumstances, which ultimately demonstrates the feasibility and superiority of our algorithm. In particular, it can be seen that the ACC of our proposed algorithm improves about 15% over that of MSE and 20% over that of Co-regspectral on the pix&zer pair. On the kar&zer pair, the ACC of our method is 96% and outperforms the ACC in the second place about 14%, which shows the great advantage of the proposed algorithm. Our method ranks second with a slight drop of 1% on the pix&kar pair when compared with MSE. It can be observed that the performance of Stack is mediocre on all the datasets, thus we can see only concatenating the feature from different views can largely limit the potential of multiview data. Although the performance of MSE is encouraging on datasets such as pix&kar pair, it's obviously not such stable compared with the performance of our method on all the datasets: MSE only ranks in the sixth place on the mor&pix pair and in the seventh place on the mor&zer pair. The performances of Co-regspectral and Co-trainspectral on each dataset are considerably stable, but the overall clustering results of RRMA are higher than those of Co-regspectral and Co-trainspectral. By simply concatenating all the feature, the ConcatRMNMF is not encouraging on most multiview datasets. For MultiNMF, by only manually weighting the importance of each view, the clustering results can be limited if there is no prior knowledge of views' qualities. In summary, our algorithm shows encouraging superiority with the robust $\ell_{2,1}$-norm and learnable weighted multiview manifold regularization.

D. Experimental results on representative real-world datasets

In this subsection, we provide experimental results on the clustering of Wiki text-image data, Cora, Citeseer, 3sources and synthetic data, to test the performance of our proposed RRMA algorithm. Detailed information of these representative and challenge datasets is listed in Table II.

(1) **Wiki text-image data**: consists of 2866 image-text pairs with 2 views, i.e., 10 dimensional latent Dirichlet allocation model based text features and 128 dimensional SIFT histogram image features.

(2) **Cora**[3]: The Cora dataset consists of 2708 machine learning papers. It can be taken as two views: one view indicates whether each word is presented in the paper, the other shows the citing relation among papers.

(3) **Citeseer**: This dataset consists of 3312 publications linked via 4732 citations. All these publications are annotated by six different labels: DB, IR, ML, Agents, AIand and HC. It can be taken as two views organized the way same as the Cora dataset.

(4) **3sources**[4]: This is a multiview text dataset, collected from three online news sources: BBC, Reuters, and The Guardian. In the dataset, 169 news articles are reported in all three sources, and we take these articles from BBC, Reuters, and the Guardian as view 1, view 2, and view 3, respectively.

(5) **Synthetic data**: This synthetic dataset is proposed in [13]. It has three views and two clusters. Each view is generated by a two component Gaussian mixture model. The features of each view are correlated.

Table III summarizes the clustering performances of each algorithm on these five datasets. We can see that the proposed RRMA algorithm outperforms other clustering methods on these challenging real-world datasets. In detail, we have the following observations: The advantage of our method is greatly highlighted on the 3sources dataset, which is a well-known dataset for clustering experiment. On 3sources dataset, the ACC of RRMA improves about 15% over that of ConcatRMNMF and 27% over that of MultiNMF, which greatly shows the superiority of RRMA. We can see the performance of Stack is weak on most datasets, which again indicates that it's not a wise solution to merely concatenate all the features. On Cora and Citeseer datasets, the performance of Co-regspectral and MSE is slightly lower than RRMA's respectively, but it can be observed that RRMA shows a relatively larger advantage on other datasets than the two algorithms. Since real-world datasets with more noise and outliers are always complicated, the stable and encouraging performance on all these challenge real-world datasets largely indicates the robustness and superiority of RRMA. However, the performances of MSE, Co-regspectral and CCA are less robust on the five datasets. In the same way, MultiNMF and ConcatRMNMF shows encouraging performance only on specific datasets. To summarize, our proposed algorithm has shown significant advantages compared with other clustering algorithms.

V. CONCLUSION

In this paper, we propose a novel RRMA algorithm based on the $\ell_{2,1}$-norm and manifold regularization for multiview clustering. To summarize, our algorithm not only alleviates the influence of noises and outliers but also encodes the intrinsic geometric information among features of different views. In addition, we properly consider the contribution of each different view with complementary and redundant features. To make our algorithm applicable, we further utilize an augmented lagrangian based method to optimize the proposed objective function. Extensive experiments and analyses on five representative real-world datasets demonstrate the effectiveness of our algorithm compare to the state-of-the-art methods.

[3]http://www.research.whizbang.com/data

[4]http://mlg.ucd.ie/datasets/3sources.html

TABLE I
CLUSTERING RESULTS OF DIFFERENT METHODS ON MULTIPLE FEATURE DATASET BY THE MEASUREMENT OF ACC.

Datasets	mor&pix	mor&zer	pix&zer	mor&kar	pix&kar	kar&zer
Stack	0.3904±0.0040	0.4187±0.0163	0.5092±0.0436	0.3915±0.0034	0.7020±0.0625	0.5307±0.0352
MSE	0.5515±0.0203	0.5170±0.0205	0.8165±0.0689	0.5525±0.0128	**0.8545±0.0942**	0.8280±0.0372
Co-regspectral	0.7741±0.0194	0.6499±0.0151	0.7154±0.0210	0.7812±0.0414	0.7307±0.0311	0.7342±0.0636
Co-trainspectral	0.8241±0.0102	0.6768±0.0203	0.7661±0.0122	0.8569±0.0293	0.7415±0.0098	0.7577±0.0746
CSP-P	0.4318±0.0241	0.4159±0.0277	0.6392±0.0436	0.2128±0.0211	0.3050±0.0198	0.3637±0.0449
CCA	0.6240±0.0363	0.6490±0.0386	0.2822±0.0073	0.6250±0.0488	0.7280±0.0479	0.5965±0.0189
MultiNMF	0.6860±0.0291	0.6270±0.0054	0.7240±0.0108	0.4512±0.0320	0.7585±0.0061	0.5415±0.0190
ConcatRMNMF	0.5725	0.5460	0.5590	0.5195	0.8020	0.5575
RRMA	**0.8360**	**0.6835**	**0.9695**	**0.8835**	0.8405	**0.9660**

TABLE III
CLUSTERING RESULTS OF DIFFERENT METHODS ON TYPICAL REAL-WORLD DATASETS BY THE MEASUREMENT OF ACC.

Datasets	Wiki	Cora	Citeseer	3sources	Synthetic
Stack	0.5433±0.0374	0.1963±0.0068	0.2517±0.0162	0.5148±0.0659	0.9900±0.0000
MSE	0.5456±0.0356	0.2471±0.0043	0.3623±0.0007	0.4533±0.0740	0.9941±0.0000
Co-regspectral	0.5497±0.0379	0.2612±0.0011	0.2507±0.0250	0.6302±0.0240	0.9949±0.0000
Co-trainspectral	0.5527±0.0259	0.2284±0.0017	0.2393±0.0113	0.6562±0.0190	0.9937±0.0000
CSP-P	0.1731±0.0050	0.2227±0.0022	0.2045±0.0108	——	——
CCA	0.5337±0.0056	0.2611±0.0002	0.3510±0.0011	0.3432±0.0000	0.9928±0.0001
MultiNMF	0.5621±0.0217	0.2522±0.0019	0.2646±0.0140	0.6789±0.0060	0.9921±0.0000
ConcatRMNMF	0.5557	0.2194	0.3397	0.7929	0.9300
RRMA	**0.5687**	**0.2629**	**0.3643**	**0.9408**	**0.9990**

ACKNOWLEDGMENT

This paper is supported in part by the National Natural Science Foundation of China under Grants 61401317, 61471274, 91338111 and U1536204, and by the Hong Kong Scholars Program under Grant XJ2015036.

REFERENCES

[1] C. Ding, X. He, H. D. Simon, and R. Jin, "On the equivalence of nonnegative matrix factorization and k-means-spectral clustering," in *Proc. SIAM ICDM*, 2005, pp. 606–610.
[2] C. Xu, D. Tao, and C. Xu, "Multi-view intact space learning," *IEEE TPAMI*, vol. 37, no. 12, pp. 2531–2544, 2015.
[3] Q. Zhang, L. Zhang, B. Du, W. Zheng, W. Bian, and D. Tao, "Mmfe: Multitask multiview feature embedding," in *Proc. ICDM*, 2015, pp. 1105–1110.
[4] L. Zhang, L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE TGRS*, vol. 50, no. 3, pp. 879–893, 2012.
[5] Y. Luo, D. Tao, K. Ramamohanarao, C. Xu, and Y. Wen, "Tensor canonical correlation analysis for multi-view dimension reduction," *IEEE TKDE*, vol. 27, no. 11, pp. 3111–3124, 2015.
[6] J. Xu, J. Han, and F. Nie, "Discriminatively embedded k-means for multi-view clustering," in *Proc. CVPR*, 2016, pp. 5356–5364.
[7] J. Yu, D. Tao, Y. Rui, and J. Cheng, "Pairwise constraints based multiview features fusion for scene classification," *PR*, vol. 46, no. 2, pp. 483–496, 2012.
[8] T. Xia, D. Tao, T. Mei, and Y. Zhang, "Multiview spectral embedding," *IEEE TSMCB*, vol. 60, no. 6, pp. 1438–1446, 2010.
[9] C. Xu, D. Tao, and C. Xu, "Large-margin multi-view information bottleneck," *IEEE TPAMI*, vol. 36, no. 8, pp. 1559–1572, 2014.
[10] W. Wang, R. Arora, K. Livescu, and J. Bilmes, "On deep multi-view representation learning," in *Proc. ICML*, 2015, pp. 1083–1092.
[11] A. M. Elkahky, Y. Song, and X. Song, "A multi-view deep learning approach for cross domain user modeling in recommendation systems," in *Proc. WWW*, 2015, pp. 278–288.
[12] X. Wang, B. Qian, and I. Davidson, "Improving document clustering using automated machine translation," in *Proc. ACM CIKM*, 2012, pp. 645–653.
[13] A. Kumar, P. Rai, and H. Daume, "Co-regularized multi-view spectral clustering," in *Proc. NIPS*, 2011, pp. 1413–1421.
[14] A. Kumar and H. Daume, "A co-training approach for multi-view spectral clustering," in *Proc. ICML*, 2011, pp. 393–400.
[15] C.-D. Wang, J. Lai, and P. Yu, "Multi-view clustering based on belief propagation," *IEEE TKDE*, vol. 28, no. 4, pp. 1007–1021, 2016.
[16] J.-P. Brunet, P. Tamayo, T. R. Golub, and J. P. Mesirov, "Metagenes and molecular pattern discovery using matrix factorization," *PNAS*, vol. 101, no. 12, pp. 4164–4169, 2004.
[17] J. Liu, C. Wang, J. Gao, and J. Han, "Multi-view clustering via joint nonnegative matrix factorization," in *Proc. SIAM ICDM*, vol. 13, 2013, pp. 252–260.
[18] X. Cai, F. Nie, and H. Huang, "Multi-view k-means clustering on big data," in *Proc. IJCAI*, 2013, pp. 2598–2604.
[19] X. Zhang, L. Zhao, L. Zong, X. Liu, and H. Yu, "Multi-view clustering via multi-manifold regularized nonnegative matrix factorization," in *Proc. ICDM*, 2014, pp. 1103–1108.
[20] Z. Zhang and K. Zhao, "Low-rank matrix approximation with manifold regularization," *IEEE TPAMI*, vol. 35, no. 7, pp. 1717–1729, 2013.
[21] L. Zhang, Q. Zhang, L. Zhang, D. Tao, X. Huang, and B. Du, "Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding," *PR*, vol. 48, no. 10, pp. 3102–3112, 2015.
[22] J. Huang, F. Nie, H. Huang, and C. Ding, "Robust manifold nonnegative matrix factorization," *ACM TKDD*, vol. 8, no. 3, pp. 1–21, 2014.
[23] S. Mei, G. Guan, Z. Wang, S. Wan, M. He, and D. D. Feng, "Video summarization via minimum sparse reconstruction," *PR*, vol. 48, no. 2, pp. 522–533, 2015.
[24] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.
[25] T. Zhang, D. Tao, X. Li, and J. Yang, "Patch alignment for dimensionality reduction," *IEEE TKDE*, vol. 21, no. 9, pp. 1299–1313, 2009.
[26] X. He and P. Niyogi, "Locality preserving projections," in *Proc. NIPS*, 2004, pp. 153–160.
[27] B. Geng, D. Tao, C. Xu, L. Yang, and X.-S. Hua, "Ensemble manifold regularization," *IEEE TPAMI*, vol. 34, no. 6, pp. 1227–1233, 2012.
[28] J. Gui, D. Tao, Z. Sun, Y. Luo, X. You, and Y. Y. Tang, "Group sparse multiview patch alignment framework with view consistency for image classification," *IEEE TIP*, vol. 23, no. 7, pp. 3126–3137, 2014.
[29] A. Sharma, A. Kumar, H. D. III, and D. W. Jacobs, "Generalized multiview analysis: A discriminative latent space," in *Proc. CVPR*, 2012, pp. 2160–2167.