

Co-regularized multiview nonnegative matrix factorization with correlation constraint for representation learning

Weihua Ou¹ · Fei Long² · Yi Tan³ · Shujian Yu⁴ · Pengpeng Wang³

Received: 29 September 2016 / Revised: 3 May 2017 / Accepted: 6 June 2017 /
Published online: 15 June 2017
© The Author(s) 2017. This article is an open access publication

Abstract With the increasing availability of multiview nonnegative data in real applications, multiview representation learning based on nonnegative matrix factorization (NMF) has attracted more and more attentions. However, existing NMF-based methods are sensitive to noises and are difficult to generate discriminative features with noisy views. To address these problems, we propose a co-regularized multiview nonnegative matrix factorization method with correlation constraint for nonnegative representation learning, which jointly exploits consistent and complementary information across different views. Different from previous works, we aim at integrating information from multiple views efficiently and making it more robust to the presence of noisy views. More specifically, we exploit the complementary information of multiple views through the co-regularization to accommodate the presence of the noisy views. Meanwhile, correlation constraint is imposed on the low-dimensional space to learn a common latent representation shared by different views. For the induced objective function, we derive an alternative algorithm to solve the optimization problem. The experimental results on four real datasets demonstrate the effectiveness and robustness of the proposed algorithm.

Keywords Nonnegative matrix factorization · Co-regularization · Correlation constraint · Consistent · Complementary

✉ Weihua Ou
ouweihuahust@gmail.com

¹ School of Big Data and Computer Science, Guizhou Normal University, Guiyang 550025, China

² School of Electrical and Information Engineering, Guizhou Institute of Technology, Guiyang 550003 China

³ College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China

⁴ Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA

1 Introduction

In real application, each object can be described by multiple different views or different features [30]. For example, as shown in Fig. 1, the objects can be represented by texture, color, shape, text and speech. These multiview representations provide complementary information to each other [47]. Integrating information from multiple views and uncovering the common latent structure shared by multiple views are the main concerning for the multiple views representation learning [55]. Generally, the traditional method is to concatenate all the features into a single vector, and then applies existing algorithms to this single vector. Obviously, this method ignores the differences of statistical properties between different views and also lacks physical meaning [39]. Actually, multi-view data contains consistent and complementary information simultaneously across different views [28, 31]. Leveraging the complementary information amongst views has better generalization ability than single view [7, 23, 25, 30, 46].

In the past year, multi-view learning algorithms have been proposed and applied successfully to image processing and computer vision [20, 42, 45, 48]. Those methods can be mainly categorized into three classes, i.e., co-training, multi-kernel learning and shared subspace learning. Co-training [2, 26] pursues to maximize the mutual agreement on two different views alternately. Based on the assumption that different kernels correspond to different views, multiple kernel learning [27, 51] combines different kernels to improve the performance. Different from co-training and multiple kernel method, the aim of subspace learning [6, 14, 19, 22, 32, 48] is to obtain a latent subspace based on the assumption that different views are generated from this latent common subspace. The classical subspace method includes canonical correlation analysis [17], which obtains the latent subspace via maximizing the correlation between different views. Though these approaches are successful in multi-view learning, they do not perform well on purely non-negative features such as pixel values or color histogram [21].

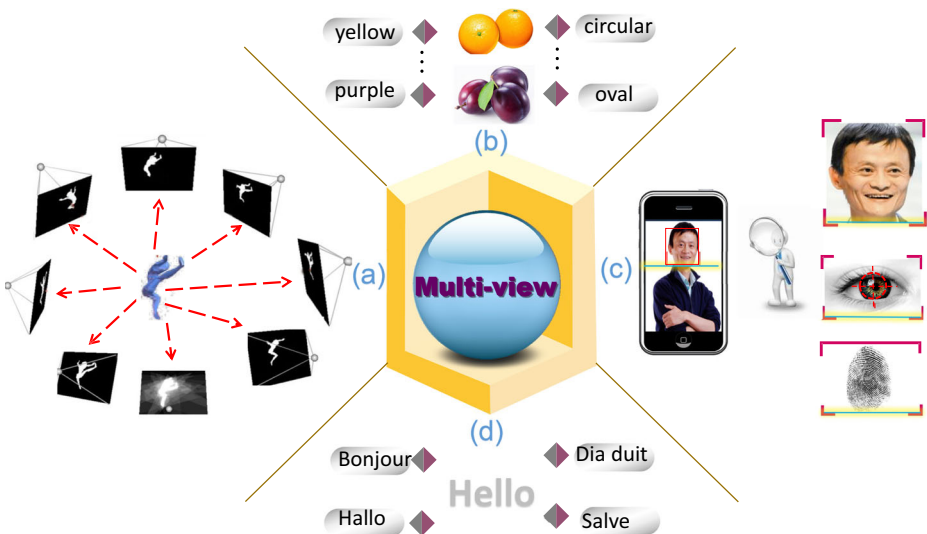


Fig. 1 Examples of multiview data. **a** an object can be described from different views; **b** the fruit can be depicted from color, shape and so on; **c** the identity of person can be represented by faces, iris and fingerprint; **d** a word can be expressed in different languages

As an effective technique for data analysis [10, 43], nonnegative matrix factorization (NMF) has been widely used in non-negative features extraction [13, 24, 41]. In recent years, many NMF-related feature extraction algorithms have been proposed. For example, LNMF [29], GNMF [4], DNMF [53], RDPNA[52] and so on. These algorithms can generate superior clustering results, but only deal with single-view data. In the past five years, several extensions of NMF to multi-view data have been proposed [5, 12, 34–36, 38, 44, 56]. For example, Liu et al. [12] presented a multi-view clustering approach via joint NMF, which aimed at finding a consensus matrix by minimizing the disagreement between the coefficient matrix and the consensus matrix. But, it performs well only for the homogeneous views. Akata et al. [1] presented an approach to learn common representation from image features and the associated tag via common coefficient matrix constraint. Xiangnan et al. [18] extended NMF for multi-view clustering by jointly factorizing the multiple matrices through the co-regularization, which has shown better performance for views with varying levels of quality. Sunil et al. [15] proposed a partial shared nonnegative subspace learning method for two views, which shows the effectiveness in social media retrieval. Furthermore, Liu et al. [31] generalized this idea into multiview nonnegative data, which can deal with more than two views. Recently, Shao et al. [37] proposed a online multi-view clustering method with incomplete view via imposing lasso regularization on the representation of each view. More references can be referred to [16, 33, 40, 50]. These methods are useful for the nonnegative multiview data analysis, however, they are not suitable for the noisy views and incomplete views, which are often encountered in real applications. For example, in the clustering of bi-lingual documents, two different languages can be regarded as two different views, however, many documents usually have only one language part.

In this paper, we propose a co-regularized nonnegative matrix factorization method with correlation constraint for robust multi-view feature learning, which provides an explicit latent representation via capturing complementary and consistent information across different views. As shown in Fig. 2, different views are represented in heterogeneous feature

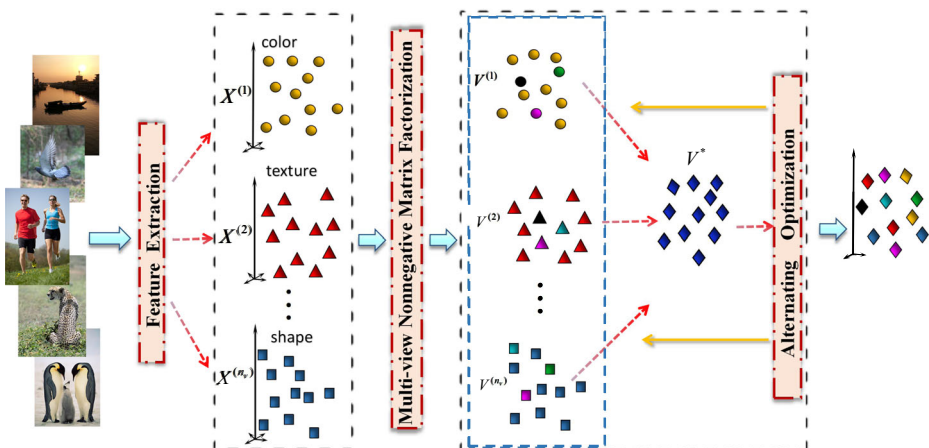


Fig. 2 Illustration of the proposed approach. Firstly, different features are extracted from the objects, such as color, texture, shape and so on. Then, the low-dimensional representations are obtained by multi-view nonnegative matrix factorization, in which the co-regularization on each pair of coefficient matrices and the correlation constraints are imposed during the factorization processing

spaces. The proposed method aims to learn robust features from all views simultaneously via exploiting the complementary and consistent information between different views. More specifically, we learn the latent representation shared by different views via maximizing the correlation between the coefficient matrix and consensus matrix. Meanwhile, we impose similarity constraints on the latent representation by co-regularizing each pair of views during the factorization process. The main contributions are summarized as follows:

- co-regularization: we exploit co-regularization for each pair of views, which is effective to accommodate the imbalance of the quality of multiple views;
- correlation constraint: we impose correlation constraint on the low-dimensional space to obtain the compact latent representation shared by different views;
- robustness to noisy views: the experimental results show the proposed method are more robust than existing methods, especially for the noisy views.

The remainder of this paper is organized as follows. In Section 2, we briefly review some related works. In Section 3, we present the proposed multi-view NMF via the co-regularization with correlation constraint. In Section 4, we give the details of optimization algorithm. Then, we report the experimental results in Section 5 and summarize this paper in Section 6.

2 Related works

In this section, we briefly review nonnegative matrix factorization (NMF) and multiview NMF.

2.1 NMF

Given a nonnegative matrix X , NMF decomposes X into the production of non-negative matrices U and V [10], i.e., $X \approx UV^T$. The objective function of NMF can be formulated as follows [5]:

$$\begin{aligned} \min_{U, V} \|X - UV^T\|_F^2 \\ \text{s.t. } U \geq 0, V \geq 0. \end{aligned} \quad (1)$$

NMF has shown good performance in pattern recognition and computer vision [41, 49].

2.2 Multi-view NMF

Given a multiview nonnegative dataset consisting of N samples with n_v different views as $\{X^{(1)}, X^{(2)}, \dots, X^{(n_v)}\}$. For each view $X^{(v)}$, multiview NMF [12] factorizes $X^{(v)} \approx U^{(v)}(V^{(v)})^T$, and learns a latent representation V^* across all the views via the following objective function:

$$\begin{aligned} \min \sum_{v=1}^{n_v} \left\{ \|X^{(v)} - U^{(v)}(V^{(v)})^T\|_F^2 + \lambda_v \|V^{(v)} - V^*\|_F^2 \right\}, \\ \text{s.t. } U^{(v)}, V^{(v)}, V^* \geq 0, \end{aligned} \quad (2)$$

where λ_v is the regularization parameter, which balances the importance of different views and the reconstruction error.

3 Co-regularized multiview NMF with correlation constraint

In this section, we present the co-regularized multiview NMF with correlation constraint for nonnegative representation learning. Given a multiview data set consisting of N samples with n_v views as $\Gamma = \{X^{(v)} \in \mathbb{R}_+^{m_v \times N}\}_{v=1}^{n_v}$, $X^{(v)} = \{\vec{x}_1^{(v)}, \dots, \vec{x}_N^{(v)}\}$, $X_i = \{\vec{x}_i^{(1)}, \dots, \vec{x}_i^{(n_v)}\}$, where $X^{(v)}$ denotes the N samples of the v th view with dimensionality m_v , X_i is the i th sample from different views. We want to learn the common latent representation V^* cross different views under the framework of NMF.

3.1 Co-regularization

The intuitive method for multiview representation learning is to learn a common representation by regularizing the representation matrices of different views. This idea works well for homogeneous views or all the views with similar quality. However, in real applications, the quality between views might vary drastically. Thus, the existing methods would be failed.

In this paper, we impose similarity constraints on each pair of views, which encourages the coefficients matrices learned from any pair of views to be complement with each other during the factorization processing. Given the contaminated data from view $X^{(v)}$ and the associated representation $V^{(v)}$, the corresponding clear data from view t is $X^{(t)}$ and the associated representation is $V^{(t)}$, the coefficients matrices $V^{(v)}$ and $V^{(t)}$ would be complement with each other by minimizing $\|V^{(v)} - V^{(t)}\|_F^2$. Thus, the problem of quality imbalance between different views can be addressed efficiently. Considering all the pair of views, the co-regularization term can be defined as follows:

$$\sum_{t=1}^{n_v} \lambda_{vt} \|V^{(v)} - V^{(t)}\|_F^2 \tag{3}$$

where λ_{vt} is the parameter to balance the importance of the similarity constraint between $V^{(v)}$ and $V^{(t)}$.

3.2 Correlation constraint

As we known, different views are complementary to each other, which capture the same latent structure of the same entity [12, 21]. To utilizing this information, in this paper, we propose correlation constraint on the low-dimensional representation to learn a compact and shared latent representation. Given the coefficient vector $V_{i,\cdot}^{(v)}$ and consensus vector $V_{i,\cdot}^*$ of the i th sample, we encourage the correlation between $V_{i,\cdot}^{(v)}$ and $V_{i,\cdot}^*$ to be as large as possible. This can be formulated as follows:

$$\max \left\{ V_{i,\cdot}^{(v)} (V_{i,\cdot}^*)^T \right\}.$$

Considering all the N sample, we have $\sum_{i=1}^N V_{i,\cdot}^{(v)} (V_{i,\cdot}^*)^T = Tr(V^{(v)}(V^*)^T)$, where Tr denotes the trace of a matrix. Thus, correlation constraints can be formulated as follows:

$$\min \left\{ Tr \left[V^* (V^*)^T - V^{(v)} (V^*)^T \right] \right\}. \tag{4}$$

Here, we impose constraints $V_{i,\cdot}^* (V_{i,\cdot}^*)^T$ on $V_{i,\cdot}^*$ in order to learn meaningful representation.

3.3 Objective function

Incorporating the co-regularization (3) and correlation constraint (4) into the NMF framework, we obtain the objective function for the proposed method:

$$\min \sum_{v=1}^{n_v} \left\{ \left\| X^{(v)} - U^{(v)}(V^{(v)})^T \right\|_F^2 + \sum_{t=1}^{n_v} \lambda_{vt} \left\| V^{(v)} - V^{(t)} \right\|_F^2 + \sigma_v Tr \left[V^*(V^*)^T - V^{(v)}(V^*)^T \right] \right\},$$

$$s.t. U^{(v)}, V^{(v)}, V^* \geq 0, \left\| U_{\cdot,k}^{(v)} \right\|_1 = 1, \forall 1 \leq k \leq K \tag{5}$$

where λ_{vt} and σ_v are the regularization parameters and K is the dimensionality of low dimensional subspace. $\|U_{\cdot,k}^{(v)}\|_1 = 1$ is the normalization with respect to the basis vector according to the relationship between NMF and probabilistic latent semantic analysis [11].

4 Optimization algorithm

To simplify the computation, we formulate the constraint on the basis matrix $U^{(v)}$ into following diagonal matrix:

$$Q^{(v)} = Diag \left(\sum_{i=1}^{m_v} U_{i,1}^{(v)}, \sum_{i=1}^{m_v} U_{i,2}^{(v)}, \dots, \sum_{i=1}^{m_v} U_{i,K}^{(v)} \right)$$

Thus, problem (5) can be reformulated as below:

$$\min \sum_{v=1}^{n_v} \left\{ \left\| X^{(v)} - U^{(v)}(V^{(v)})^T \right\|_F^2 + \sum_{t=1}^{n_v} \lambda_{vt} \left\| V^{(v)}Q^{(v)} - V^{(t)} \right\|_F^2 + \sigma_v Tr \left[V^*(V^*)^T - V^{(v)}Q^{(v)}(V^*)^T \right] \right\}$$

$$s.t. U^{(v)}, V^{(v)}, V^* \geq 0 \tag{6}$$

4.1 Optimize $U^{(v)}$ and $V^{(v)}$ for given V^*

We utilize the alternative update scheme, i.e., solving one variable with the others fixed. When V^* is fixed, for each given v , the computation of $U^{(v)}$ and $V^{(v)}$ is independent of view. Therefore, we use $X, U, V, \lambda_t, \sigma$ and Q to denote $X^{(v)}, U^{(v)}, V^{(v)}, \lambda_{vt}, \sigma_v$ and $Q^{(v)}$ for the brevity.

4.1.1 Optimize U for given V and V^*

Given V and V^* , the problem (6) can be solved by optimizing each row of U separately as follows:

$$L(U) = \left\| X - UV^T \right\|_F^2 + \sum_{t=1}^{n_v} \lambda_t \left\| VQ - V^{(t)} \right\|_F^2 + \sigma Tr \left[V^*(V^*)^T - VQ(V^*)^T \right] + Tr(\Theta^T U), \tag{7}$$

where $\Theta = [\theta_{i,k}] \in \mathbb{R}^{m \times K}$ is the Lagrange multipliers for the non-negative constraint $U \geq 0$. The partial derivatives of $L(U)$ with respect to $U_{i,k}$ is presented below:

$$\frac{\partial L(U)}{\partial U_{i,k}} = -2(XV)_{i,k} + 2(UV^T V)_{i,k} + S_{i,k} + \sigma T_{i,k} + \theta_{i,k}, \tag{8}$$

where $S_{i,k}$ is the derivative of $\sum_{t=1}^{n_v} \lambda_t \|VQ - V^{(t)}\|_F^2$, and $T_{i,k}$ is the derivative of $Tr[V^*(V^*)^T - VQ(V^*)^T]$ with respect to $U_{i,k}$. Their calculus formulations are shown below, respectively:

$$\begin{aligned} S_{i,k} &= \frac{\partial \sum_{t=1}^{n_v} \lambda_t \|VQ - V^{(t)}\|_F^2}{\partial U_{i,k}} \\ &= 2 \sum_{t=1}^{n_v} \lambda_t \left\{ \sum_{l=1}^m U_{l,k} \left(\sum_{j=1}^N V_{j,i} V_{j,k} \right) - \sum_{j=1}^N V_{j,i} V_{j,k}^{(t)} \right\} \\ T_{i,k} &= \frac{\partial Tr[V^*(V^*)^T - VQ(V^*)^T]}{\partial U_{i,k}} = - \sum_{j=1}^N V_{j,i} V_{j,k}^* \end{aligned} \tag{9}$$

Setting (8) to zero and utilizing the KKT conditions $\theta_{i,k} U_{i,k} = 0$, we can get following equation for $U_{i,k}$:

$$\left(-2(XV)_{i,k} + 2(UV^T V)_{i,k} + S_{i,k} + \sigma T_{i,k} + \theta_{i,k} \right) U_{i,k} = 0$$

This equation leads to the update rule below for $U_{i,k}$:

$$U_{i,k} \leftarrow U_{i,k} \frac{2(XV)_{i,k} + 2 \sum_{t=1}^{n_v} \lambda_t \sum_{j=1}^N V_{j,i} V_{j,k}^{(t)} + \sigma \sum_{j=1}^N V_{j,i} V_{j,k}^*}{2(UV^T V)_{i,k} + 2 \sum_{t=1}^{n_v} \lambda_t \sum_{l=1}^m U_{l,k} \sum_{j=1}^N V_{j,i} V_{j,k}} \tag{10}$$

4.1.2 Optimize V for given U and V*

To optimize V, we first normalize the columns of U using Q as following:

$$U \leftarrow UQ^{-1}, V \leftarrow VQ$$

Then, the problem (6) is equivalent to minimize following objective function:

$$\begin{aligned} L(V) &= \|X - UV^T\|_F^2 + \sum_{t=1}^{n_v} \lambda_t \|V - V^{(t)}\|_F^2 \\ &\quad + \sigma Tr[V^*(V^*)^T - V(V^*)^T] + Tr(\Psi^T V), \end{aligned} \tag{11}$$

where $\Psi = [\psi_{j,k}] \in \mathbb{R}^{N \times K}$ is the Lagrange multipliers for the non-negative constraints $V \geq 0$. The partial derivatives of $L(V)$ with respect to $V_{j,k}$ is below:

$$\begin{aligned} \frac{\partial L(V)}{\partial V_{j,k}} &= -2(X^T U)_{j,k} + 2(VU^T U)_{j,k} + 2 \sum_{t=1}^{n_v} \lambda_t V_{j,k} \\ &\quad - 2 \sum_{t=1}^{n_v} \lambda_t V_{j,k}^{(t)} - \sigma V_{j,k}^* + \psi_{j,k} \end{aligned} \tag{12}$$

Setting (12) to zero and utilizing the KKT conditions $\psi_{j,k}V_{j,k} = 0$, we can get following equation for $V_{j,k}$:

$$(-2(X^T U)_{j,k} + 2(VU^T U)_{j,k} + 2 \sum_{t=1}^{n_v} \lambda_t V_{j,k} - 2 \sum_{t=1}^{n_v} \lambda_t V_{j,k}^{(t)} - \sigma V_{j,k}^* + \psi_{j,k})V_{j,k} = 0 \tag{13}$$

Thus, the update rules for $V_{j,k}$ can be shown below:

$$V_{j,k} \leftarrow V_{j,k} \frac{2(X^T U)_{j,k} + 2 \sum_{t=1}^{n_v} \lambda_t V_{j,k}^{(t)} + \sigma V_{j,k}^*}{2(VU^T U)_{j,k} + 2 \sum_{t=1}^{n_v} \lambda_t V_{j,k}} \tag{14}$$

4.2 Optimize V^* for given $U^{(v)}$ and $V^{(v)}$

Taking the derivative of the objective function (6) with respect to V^* , we obtain

$$\frac{\partial R}{\partial V^*} = \sum_{v=1}^{n_v} 2\sigma_v V^* - \sum_{v=1}^{n_v} \sigma_v V^{(v)} Q^{(v)}, \tag{15}$$

where $R = \sum_{v=1}^{n_v} \sigma_v Tr [V^*(V^*)^T - V^{(v)} Q^{(v)}(V^*)^T]$. Setting (15) to 0, we get the closed solution for V^* :

$$V^* = \frac{\sum_{v=1}^{n_v} \sigma_v V^{(v)} Q^{(v)}}{\sum_{v=1}^{n_v} 2\sigma_v} \tag{16}$$

$U^{(v)}$, $V^{(v)}$ and V^* are updated alternatively via (10), (14) and (16). It can be seen that $U^{(v)}$, $V^{(v)}$ and V^* are non-negative after each update. Moreover, it is provable that the objective function is non-increasing under the above iterative updating rules, and the convergence is guaranteed. The proof can be demonstrated by constructing the auxiliary function similar to [8]. This procedure repeats until convergence. The complete algorithm is summarized in Algorithm 1.

Algorithm 1 Robust multiview nonnegative representation learning

Input:

Multiview data $\Gamma = \{X^{(v)} \in \mathbb{R}^{m_v \times N}\}_{v=1}^{n_v}$, regularization parameters $\lambda_{v,t}$, σ_v , and dimensionality of subspace K .

Output:

Basis matrix and the coefficients matrix $\{U^{(i)}, V^{(i)}\}$, $i = 1, 2, \dots, n_v$, and the consensus matrix V^* .

- 1: Normalize each view $X^{(v)}$ such that $\|X^{(v)}\|_1 = 1$.
 - 2: Initialize $U^{(v)}$, $V^{(v)}$ and V^* .
 - 3: **repeat**
 - 4: **for** $v = 1$ to n_v **do**
 - 5: **repeat**
 - 6: Fixing V^* and $V^{(v)}$, update $U^{(v)}$ by (10);
 - 7: Fixing V^* and $U^{(v)}$, update $V^{(v)}$ by (14);
 - 8: **until** The terminated condition is satisfied.
 - 9: **end for**
 - 10: Fixing $V^{(v)}$ and $U^{(v)}$, update V^* by (16);
 - 11: **until** The objective function converges.
-

4.3 Complexity analysis

We adopt the standard NMF as baseline to analyse the time complexity of the proposed method. It can be seen that the proposed method is essentially an extension of NMF for multiple view data. The complexity of basic NMF's update rules in each iteration is $O(mKN)$, where big O is the notation for complexity. For each update of U in our proposed method, its cost is $O(n_v mKN)$. For each update of V , the additional cost in the proposed method is the second term in the numerator and denominator, whose time complexity is $O(n_v KN)$. Therefore the time complexity of the proposed method for each view is $O(n_v mKN)$. Then, the total complexity of the proposed method in each iteration is $O(n_v mKN)$, where n_v is the number of the views.

5 Experimental results

In this section, we conduct experiments on four datasets to evaluate the performance of the proposed method compared to the following algorithms.

- **Single view.** This method runs each view separately using the NMF. Both the best and the worst single view results are reported, which are denoted by BSV and WSV respectively.
- **Feature concatenation (FC).** This method runs NMF directly on the concatenated features from all views.
- **Multi-view NMF (Multi-NMF) [12].** This method requires all the representation of different views to share a common latent one, i.e., $\sum_{v=1}^{n_v} \lambda_v \|V^{(v)} - V^*\|_F^2$. As the authors provided a NMF-based initialization, we use the same initialization method and set the regularization parameters as 0.01.
- **Multi-view RNMF (Multi-RNMF) [35].** This method learns the common latent representation under the nonnegative patch alignment framework and considers the local geometric structure for each view.
- **Co-regularization NMF (CoNMF) [18].** This method learns the common latent space via pair-wise co-regularization.
- **Our method.** This method learns latent representation by simultaneously exploiting the complementary and consistent information from all views through the co-regularization and correlation constraint.

5.1 Data sets and evaluation

ORL dataset The ORL dataset consists of 40 subjects and 10 different images for each subject with totally 400 images. The images are grayscale and have been normalized to 32×32 pixels, some of which are shown in Fig. 3a. We adopt two different views. The first view is the raw pixel values, i.e., $X^1 \in \mathbb{R}_+^{1024 \times 400}$, and the second view is the $LBP_{(8,1)}$ feature, i.e., $X^2 \in \mathbb{R}_+^{59 \times 400}$.

CMU-PIE dataset There are 41,368 images under 68 persons with 13 different poses, 43 different illumination, and 4 different expressions in the CMU-PIE dataset. In our experiment, we chose 42 images at pose 27 for each person at different illumination conditions with resolution 32×32 and add white random block occlusion with size 10×10 . There are 2856 images in all and some examples are shown in Fig. 3b. We consider two different

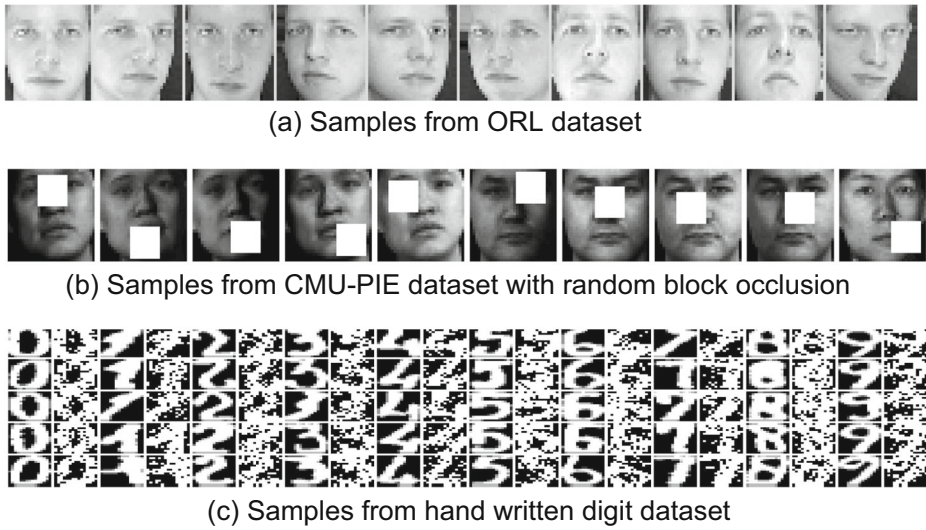


Fig. 3 **a** The samples from the ORL dataset. **b** The samples from CMU-PIE dataset with random block occlusion. **c** The samples from handwritten digit dataset. It consists of original images and noisy images with salt & pepper noise, in which the noise level is 25%

views: the first view is the raw pixel values $X^1 \in \mathbb{R}_+^{1024 \times 2856}$, and the second view is the local binary pattern $X^2 \in \mathbb{R}_+^{256 \times 2856}$.

UCI handwritten digit dataset This handwritten digits (0–9) data is from the UCI repository, which consists of 2000 samples, with the first view being the 76 fourier coefficients of the character shapes, the second view being the 240 pixel averages in 2×3 windows, the third view being the 216 profile correlations, and the fourth view being the 47 Zernike moments. In order to test the robustness of the proposed method, salt&pepper noises are added with noise level varied as {5%, 10%, 15%, 20%, 25%}. Some examples are shown in Fig. 3c.

OuluVS dataset Lipreading is a technology to interpret the utterance solely using the visual information of lip movements. The OuluVS dataset records the lipreading video of 20 subjects, with a total of 817 videos. Each subject was asked to sit in front of a camera and speaks 10 different sentences as shown in Table 1. The subjects are from four countries, with different speech habit and accent. Usually, multivariate time series are used to model the facial landmarks around mouth outer. Then, the extracted time series are formulated as texture images with a modified recurrence plot. The recognition is based on the texture images. In this dataset, we extract two different features as different views from the texture images. The first view is the uniform local binary pattern operator $LB P_{p,r}^{u2}$ with $p = 8$ and $r = \{1, 2, 3\}$ to generate a 177-dimensional feature vector. For the second view is the grey

Table 1 The ten different sentences in OuluVS dataset

“Excuse me”	“See you”	“Good bye”	“I am sorry”	“Hello”
“Thank you”	“How are you ”	“Have a good time”	“Nice to meet you”	“You are welcome”

Table 2 ACC with different K values on the ORL dataset

Method	ACC v.s. different K							Avg.
	10	15	20	25	30	35	40	
BSV	0.225±0.01	0.327±0.02	0.417±0.02	0.508±0.03	0.547±0.03	0.612±0.03	0.647±0.02	0.469±0.03
WSV	0.213±0.02	0.293±0.02	0.377±0.03	0.440±0.04	0.472±0.03	0.515±0.03	0.557±0.03	0.409±0.02
FC	0.212±0.02	0.280±0.03	0.345±0.03	0.373±0.03	0.407±0.04	0.422±0.05	0.447±0.04	0.355±0.03
Multi-NMF	0.217±0.01	0.292±0.03	0.347±0.03	0.424±0.02	0.442±0.03	0.455±0.03	0.493 ±0.03	0.381±0.02
CoNMF	0.220±0.01	0.320±0.02	0.422±0.03	0.494±0.03	0.542±0.05	0.603±0.05	0.645±0.04	0.463±0.02
Multi-RNMF	0.205±0.01	0.292±0.02	0.367±0.03	0.432±0.03	0.482±0.02	0.475±0.02	0.555±0.03	0.401±0.02
our method	0.222±0.01	0.330±0.02	0.430±0.02	0.507±0.02	0.473±0.04	0.607±0.03	0.660±0.03	0.499±0.02

Table 3 NMI with different K values on the ORL dataset

Method	NMI v.s. different K							Avg.
	10	15	20	25	30	35	40	
BSV	0.551±0.02	0.630±0.03	0.693±0.04	0.728±0.03	0.753±0.02	0.788±0.02	0.809±0.02	0.707±0.02
WSV	0.515±0.03	0.568±0.03	0.614±0.02	0.683±0.03	0.693±0.02	0.716±0.03	0.732±0.03	0.645±0.03
FC	0.502±0.02	0.563±0.03	0.591±0.03	0.616±0.02	0.630±0.04	0.644±0.03	0.656±0.03	0.600±0.03
Multi-NMF	0.525±0.02	0.594±0.02	0.604±0.02	0.642±0.03	0.656±0.03	0.678±0.02	0.694±0.02	0.627±0.02
CoNMF	0.534±0.03	0.640±0.02	0.702±0.02	0.732±0.02	0.748±0.02	0.783±0.02	0.804±0.02	0.706±0.02
Multi-RNMF	0.525±0.02	0.584±0.02	0.606±0.02	0.646±0.02	0.685±0.03	0.686±0.02	0.733±0.03	0.638±0.03
our method	0.548±0.03	0.651±0.02	0.710±0.02	0.749±0.02	0.760±0.03	0.791±0.02	0.824±0.01	0.720±0.02

Table 4 ACC with different K values on the CMU-PIE dataset with occlusion

Method	ACC v.s. different K									
	10	20	30	40	50	60	68	Avg.		
BSV	0.063±0.02	0.133±0.02	0.187±0.02	0.264±0.02	0.327±0.02	0.396±0.03	0.433±0.02	0.257±0.02		
WSV	0.045±0.02	0.083±0.02	0.121±0.02	0.162±0.01	0.208±0.02	0.258±0.03	0.288±0.03	0.166±0.02		
FC	0.053±0.02	0.084±0.02	0.120±0.02	0.165±0.02	0.185±0.03	0.255±0.03	0.270±0.03	0.152±0.02		
Multi-NMF	0.054±0.02	0.117±0.02	0.177±0.03	0.261±0.03	0.302±0.02	0.384±0.04	0.421±0.04	0.245±0.02		
CoNMF	0.059±0.02	0.126±0.02	0.172±0.03	0.271±0.03	0.311±0.02	0.396±0.04	0.447±0.02	0.255±0.02		
Multi-RNMF	0.061±0.02	0.122±0.02	0.183±0.02	0.282±0.02	0.334±0.02	0.405±0.02	0.450±0.03	0.262±0.02		
our method	0.067±0.02	0.140±0.02	0.203±0.02	0.297±0.02	0.356±0.02	0.445±0.02	0.488±0.02	0.285±0.02		

Table 5 NMI with different K values on the CMU-PIE dataset with occlusion

Method	NMI v.s. different K								Avg.
	10	20	30	40	50	60	68		
BSV	0.234±0.02	0.389±0.02	0.483±0.02	0.525±0.02	0.610±0.02	0.643±0.02	0.693±0.02	0.693±0.02	0.511±0.02
WSV	0.095±0.02	0.203±0.03	0.291±0.02	0.348±0.02	0.403±0.02	0.465±0.02	0.493±0.02	0.493±0.02	0.328±0.02
FC	0.089±0.02	0.223±0.02	0.309±0.02	0.260±0.02	0.401±0.03	0.465±0.03	0.489±0.02	0.489±0.02	0.321±0.02
Multi-NMF	0.188±0.02	0.363±0.02	0.443±0.03	0.493±0.03	0.561±0.03	0.612±0.03	0.636±0.03	0.636±0.03	0.472±0.02
CoNMF	0.242±0.02	0.365±0.02	0.467±0.03	0.542±0.02	0.594±0.03	0.664±0.03	0.685±0.02	0.685±0.02	0.509±0.02
Multi-RNMF	0.250±0.02	0.380±0.02	0.483±0.02	0.538±0.02	0.583±0.03	0.645±0.03	0.699±0.02	0.699±0.02	0.510±0.02
our method	0.258±0.02	0.398±0.02	0.495±0.02	0.568±0.02	0.607±0.02	0.672±0.01	0.695±0.02	0.695±0.02	0.526±0.02

Table 6 Clustering results on the handwritten digit dataset ($K = 10$)

Method	BSV	WSV	FC	Multi-NMF	CoNMF	Multi-RNMF	Our method
ACC	0.819±0.02	0.711±0.07	0.841±0.02	0.868±0.01	0.875±0.01	0.843±0.02	0.880±0.01
NMI	0.713±0.02	0.634±0.04	0.751±0.03	0.770±0.03	0.801±0.01	0.760±0.02	0.804±0.01

level co-occurrence matrix (GLCM). We use four direction ($0^\circ, 45^\circ, 90^\circ$ and 135°) and five distances ($d = \{1, 2, 3, 4, 5\}$) to calculate 20 GLCMs. Thus, a 400-dimensional feature vector is obtained.

Evaluation metrics For quantitative evaluation, the accuracy (ACC) and the normalized mutual information metric (NMI) are used to measure the clustering performance [3, 54]. The detailed definitions are shown below.

- **Clustering accuracy (ACC).** ACC compares the generated clusters with the ground truth. In details, given samples \mathbf{x}_i , let l_i and g_i be the clusters label and ground truth label. The definition of ACC is defined as below:

$$ACC = \frac{1}{n} \sum_{i=1}^n \delta(g_i, map(l_i)),$$

where n is the total number of samples, and $\delta(x, y)$ is the delta function that equals one if $x = y$, else $\delta(x, y) = 0$. And $map(\cdot)$ is the permutation mapping function, which maps each cluster label to the real label. Here, we used the Kuhn-Mukres algorithm [9]. It is easy to see the range of ACC is $[0, 1]$. The more large of value ACC is, the better of cluster results is.

- **Normalized mutual information (NMI).** Let \mathcal{C} denote the set of clusters obtained from the ground truth, and \mathcal{C}' be the cluster results, the mutual information is defined as below:

$$MI = \sum_{\mathbf{c}_i \in \mathcal{C}, \mathbf{c}'_j \in \mathcal{C}'} p(\mathbf{c}_i, \mathbf{c}'_j) \log \frac{p(\mathbf{c}_i, \mathbf{c}'_j)}{p(\mathbf{c}_i)p(\mathbf{c}'_j)},$$

where $p(\mathbf{c}_i)$ and $p(\mathbf{c}'_j)$ are the probabilities that a sample arbitrarily selected from the data set belongs to the clusters \mathbf{c}_i and \mathbf{c}'_j , respectively. $p(\mathbf{c}_i, \mathbf{c}'_j)$ is the joint probability density function of \mathcal{C} and \mathcal{C}' . In our experiments, the NMI is defined as below

$$NMI = \frac{MI(\mathcal{C}, \mathcal{C}')}{\max(H(\mathcal{C}), H(\mathcal{C}'))},$$

where $H(\mathcal{C})$ and $H(\mathcal{C}')$ denote the correntropy of \mathcal{C} and \mathcal{C}' , respectively. It is easy to see that NMI ranges from 0 to 1. The more large value of NMI is, the better result of clutering is.

Table 7 Clustering results on the OuluVS dataset ($K = 10$)

Method	BSV	WSV	FC	Multi-NMF	CoNMF	Multi-RNMF	Our method
ACC	0.500±0.03	0.370±0.07	0.510±0.06	0.510±0.06	0.550±0.05	0.540±0.05	0.580±0.05
NMI	0.490±0.06	0.390±0.09	0.524±0.07	0.513±0.04	0.537±0.05	0.525±0.04	0.569±0.04

5.2 Experimental results with two views

In this section, we conduct clustering on three real-word datasets with two views, respectively. For handwritten digit dataset, we select the first and second view. In order to evaluate the performance effectively, we runs each experiment 30 times, then the average clustering results and standard variation are reported.

The clustering results with different number of clusters K on ORL dataset are shown in Tables 2 and 3. From that, it can be seen the clustering performances of all algorithms get better with the increase of K and the proposed method performs better than other multi-view algorithms in most cases. It notes that only the best results are recorded in BSV, which is not stable for clustering.

The clustering results with different number of clusters K on CMU-PIE dataset are shown in Tables 4 and 5. It is obvious that the multi-view algorithms outperform the single feature method, even for the best results. Among all the multiview methods, Multi-RNMF and our method are better than that of all the other methods. Our method outperforms Multiview-RNMF method slightly.

For the handwritten digit dataset and OuluVS dataset, we fix the cluster number K at 10. Tables 6 and 7 show the comparison results of the average clustering performance on those two datasets, respectively. It is clear that the clustering performance of Multi-NMF and Multi-RNMF is better than that of single view NMF. Meanwhile, the performance of FC has

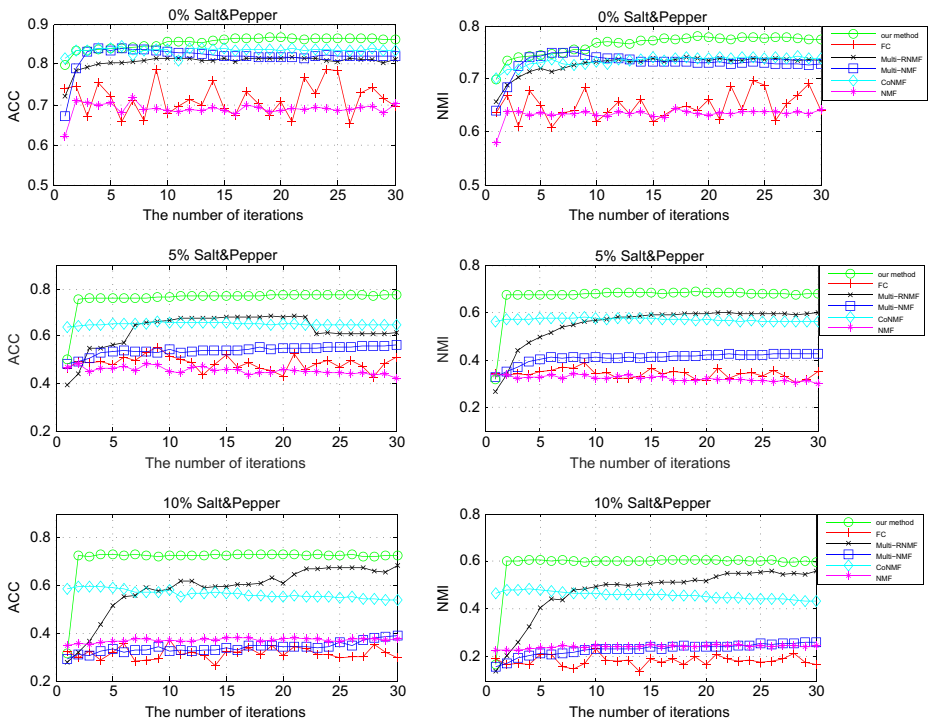


Fig. 4 Clustering results of different methods on the handwritten digit dataset with four views in the presence of salt&pepper noises varying from 5 to 10%

obviously different on the different datasets because it only integrates multiple features by concatenating of feature, which ignores the differences of statistical properties between different views. In addition, our method obtain impressive clustering performance in all cases due to the utilization of both consistent and complementary information across different views.

5.3 Experimental results with four views

In this section, we conduct clustering on the handwritten digit dataset with four views. To test the robustness, the salt&pepper noises are added with noise level varying from 5 to 25%. The clustering results are shown in Figs. 4 and 5. It is obvious that the performance of multi-view algorithms is better than that of single view NMF and FC. Both Multi-RNMF and Multi-RNMF achieve satisfactory clustering results, while Multi-RNMF performs better than that of Multi-NMF with increase of noise level. Meanwhile, the proposed method obtains the best clustering results compared to other algorithms. Specifically, the performance of all algorithms drops down sharply with the increase of the noise level, but the proposed method decreases slightly. This is mainly because we utilize the co-regularization and correlation constraints to exploit the complementarity and consistent information across different views.

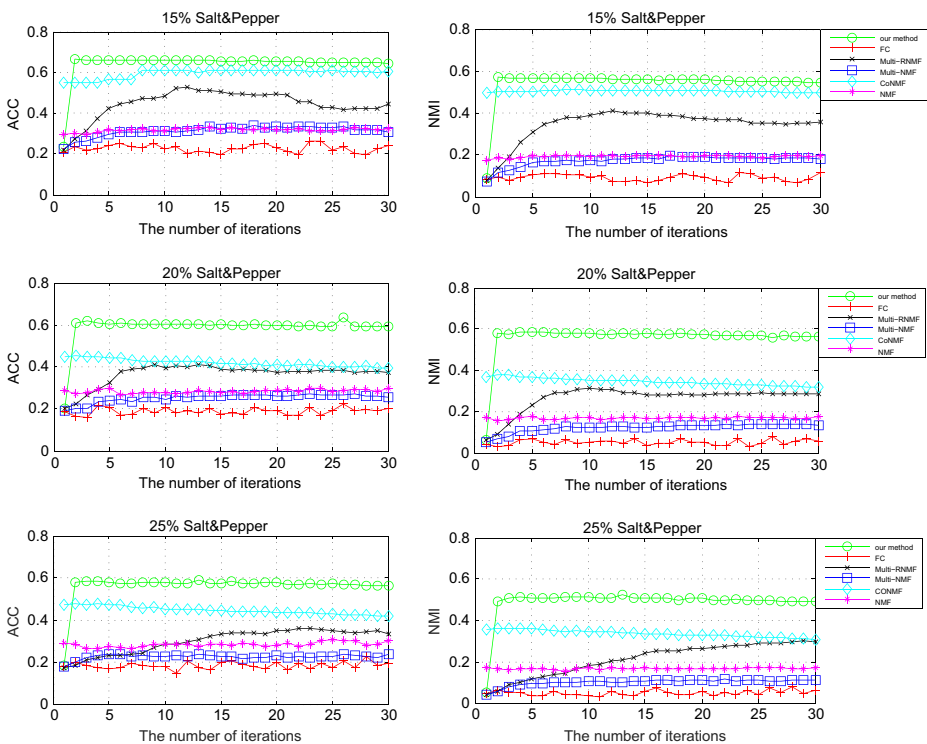


Fig. 5 Clustering results of different methods on the handwritten digit dataset with four views in the presence of salt&pepper noises varying from 15 to 25%

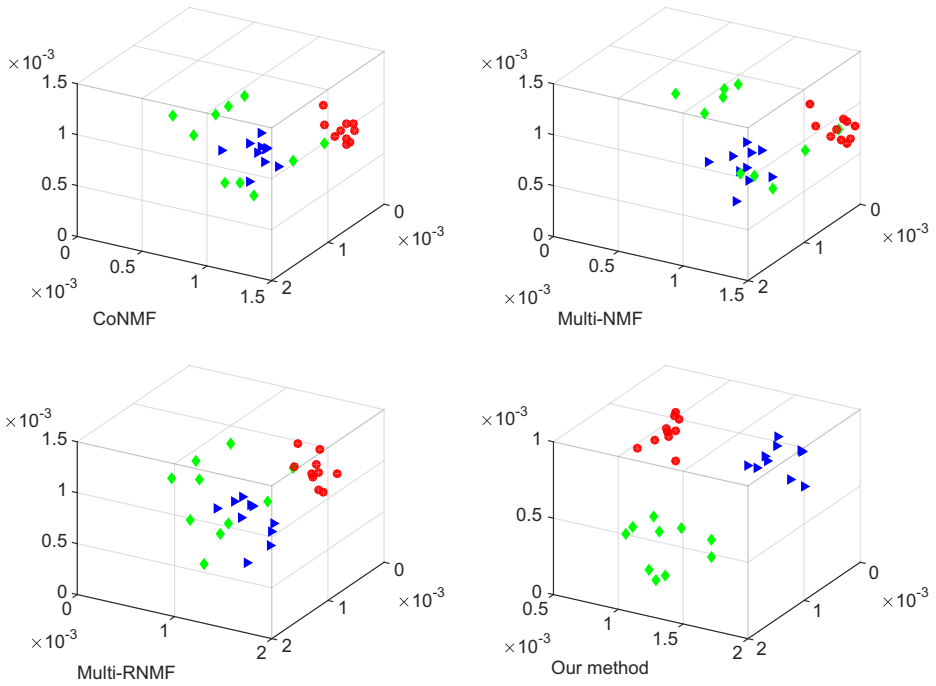


Fig. 6 The visualization of low-dimensional representation of three different classes, which are randomly selected from the ORL dataset. Different colors denote different classes, and the ten points in each class corresponds to the ten images in each subject

5.4 Visualization of clustering results

To visualize the clustering results, we randomly select three subjects from the ORL dataset with two views. There are 10 samples for each subject and 30 samples in all. The hidden

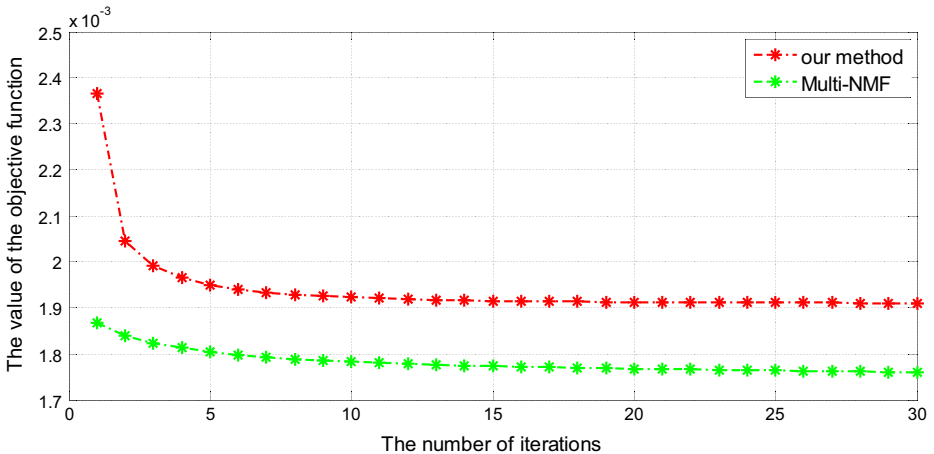


Fig. 7 The convergence curves of Multi-NMF and our method on handwritten digit dataset

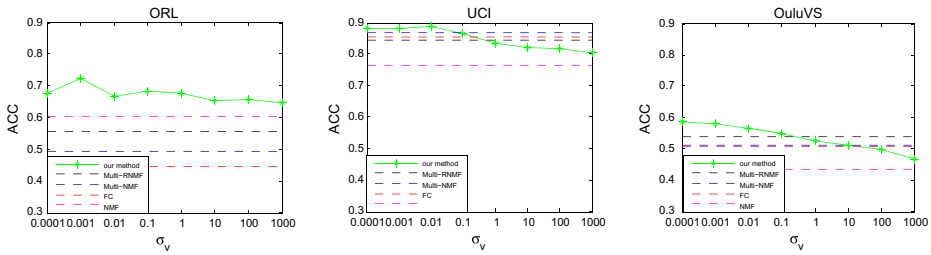


Fig. 8 The performance of different methods on three dataset with σ_v varying while fixing $\lambda_{vt} = 0.001$

factor k is set to 3. The learned common representation in the new space are shown in Fig. 6. It is obvious to see that the proposed method obtains more discriminative features.

5.5 Analysis of convergence

In this section, we demonstrate the convergence of our method by conducting experiment on handwritten digit dataset with the same initialization. As shown in Fig. 7, the objective function value is non-increasing under the proposed iterative update rules. Meanwhile, it is easy to see that the proposed method converges faster than that of Multi-NMF. Compared to Multi-NMF, the objective function value of the proposed method decreases fast within 5 iterations. This is because the co-regularization term and the correlation constraints limited the solution space. These also can be verified through the clustering results on the handwritten digit dataset in Figs. 4 and 5.

5.6 Parameters selection

Two kinds of parameters σ_v and λ_{vt} are needed to set. The parameter σ_v balances the correlation constraint and the latent representation, while λ_{vt} determines the importance of each pair of view in co-regularization. In this section, we conduct experiments on three dataset with two views to study the influence of them. Figures 8 and 9 show the performance of the proposed method with one parameter varying while the others fixed. It is clear that our method is relatively stable across a wide range of values, especially on the ORL dataset. For the OuluVS dataset, they vary dramatically compared to other datasets. According to the results the parameters are set to 0.001 and 0.05 respectively in our experiments.

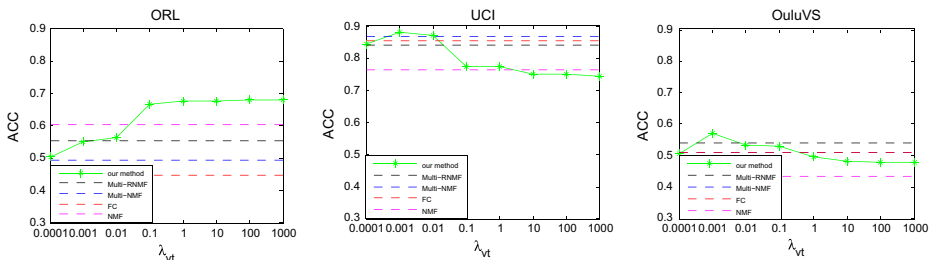


Fig. 9 The performance of different methods on three dataset with λ_{vt} varying while fixing $\sigma_v = 0.05$

6 Conclusion

In this paper, we proposed a co-regularized multiview nonnegative matrix factorization method with correlation constraint for nonnegative representation learning. We exploited the complementary information through the co-regularization to deal with the imbalance views. Thus, the latent representations were complemented to each other when one of views was contaminated. Meanwhile, we imposed correlation constraint on the common latent subspace to obtain the latent representation shared by different views. The experimental results show that the representation learned by proposed method is more compact and discriminative, especially for noisy view. In the future work, we will study the supervised multiview nonnegative representation learning for classification.

Acknowledgments This work is partly supported by the National Nature Science Foundation of China (Grant No. 61402122, 61263005, 61461008), the Natrual Science Fundataion of Guizhou Province No.[2017]1130, the Program for New Century Excellent Talents in Chinese University under Grant No.NCET-12-0657, the Innovation Fund for Graduate Student of Guizhou University (No.2015080), the 2014 Ph.D. Recruitment Program of Guizhou Normal University, the Outstanding Innovation Talents of Science and Technology Award Scheme of Education Department in Guizhou Province (Qianjiao KY word [2015]487), the China Scholarship Council (No.201508525007), Foundation of Guizhou Educational Department (KY[2016]027).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Akata C, Thurau Z, Bauckhage C (2011) Non-negative matrix factorization in multimodality data for segmentation and label prediction. In: Computer vision winter workshop
2. Blum A, Mitchell T (1998) Combining labeled and unlabeled data with co-training. In: Proceedings of the eleventh annual conference on computational learning theory. ACM, pp 92–100
3. Cai D, He X, Han J (2005) Document clustering using locality preserving indexing. *IEEE Trans Knowl Data Eng* 17(12):1624–1637
4. Cai D, He X, Han J, Huang T (2011) Graph regularized nonnegative matrix factorization for data representation. *IEEE Trans Pattern Anal Mach Intell* 33(8):1548–1560
5. Chang WY, Wei CP, Wang YCF (2014) Multi-view nonnegative matrix factorization for clothing image characterization. In: 2014 22nd international conference on pattern recognition (ICPR), pp 1272–1277
6. Chaudhuri K, Kakade SM, Livescu K, Sridharan K (2009) Multi-view clustering via canonical correlation analysis. In: Proceedings of the 26th annual international conference on machine learning. ACM, pp 129–136
7. Chen N, Zhu J, Sun F, Xing EP (2012) Large-margin predictive latent subspace learning for multiview data analysis. *IEEE Trans Pattern Anal Mach Intell* 34(12):2365–78
8. Chen X, Yu J, Sun W (2013) Area-correlated spectral unmixing based on bayesian nonnegative matrix factorization. *Open J Appl Sci* 3(1):365–372
9. Cornuéjols G, Hartvigsen D (1986) An extension of matching theory. *J Comb Theory Ser B* 40(3):285–296
10. Ding C, Kong D (2012) Nonnegative matrix factorization using a robust error function. In: IEEE International conference on acoustics, speech, and signal processing ICASSP, pp 2033–2036
11. Ding C, Li T, Peng W (2008) On the equivalence between non-negative matrix factorization and probabilistic latent semantic indexing. *Comput Stat Data Anal* 52(8):3913–3927
12. Gao J, Han J, Liu J, Wang C (2013) Multi-view clustering via joint nonnegative matrix factorization. In: Proceedings of the 13th SIAM international conference on data mining, pp 252–260

13. Guan N, Tao D, Luo Z, Shawe-Taylor J (2012) Mahnmf: Manhattan non-negative matrix factorization. *J Mach Learn Res* 1(5):11–43
14. Gui J, Tao D, Sun Z, Luo Y, You X, Tang YY (2014) Group sparse multiview patch alignment framework with view consistency for image classification. *IEEE Trans Image Process* 23(7):3126–3137
15. Gupta SK, Phung D, Adams B, Tran T, Venkatesh S (2010) Nonnegative shared subspace learning and its application to social media retrieval. In: Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 1169–1178
16. Gupta SK, Phung D, Adams B, Venkatesh S (2013) Regularized nonnegative shared subspace learning. *Data Min Knowl Disc* 26(1):57–97
17. Hardoon DR, Szedmak S, Shawe-Taylor J (2004) Canonical correlation analysis: an overview with application to learning methods. *Neural Comput* 16(12):2639–2664
18. He X, Kan MY, Xie P, Chen X (2014) Comment-based multi-view clustering of web 2.0 items. In: International conference on world wide web, pp 771–782
19. He Z, Yi S, Cheung Y-M, You X, Tang YY (2016) Robust object tracking via key patch sparse representation. *IEEE Transactions on Cybernetics* 99:1–11
20. Jing X-Y, Hu R, Zhu Y-P, Wu S, Liang C, Yang J-Y (2014) Intra-view and inter-view supervised correlation analysis for multi-view feature learning. In: AAAI, pp 1882–1889
21. Kalayeh MM, Idrees H, Shah M (2014) Nmf-knn: image annotation using weighted multi-view non-negative matrix factorization. In: Proceedings of the 2014 IEEE conference on computer vision and pattern recognition, IEEE computer society, pp 184–191
22. Kan M, Shan S, Zhang H, Lao S, Chen X (2012) Multi-view discriminant analysis. *Computer Vision–ECCV* 2012:808–821
23. Kan M, Shan S, Zhang H, Lao S, Chen X (2016) Multi-view discriminant analysis. *IEEE Trans Pattern Anal Mach Intell* 38(1):188–194
24. Kim H, Park H (2007) Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics* 23(12):1495–1502
25. Kumar A (2011) A co-training approach for multi-view spectral clustering. In: Proceedings of the 28th international conference on machine learning, ICML, pp 393–400
26. Kumar A, Rai P, Daume H (2011) Co-regularized multi-view spectral clustering. In: Advances in neural information processing systems 24, Curran Associates, Inc, pp 1413–1421
27. Lanckriet GRG, Christianini N, Bartlett PL, Ghaoui LE, Jordan MI (2002) Learning the kernel matrix with semi-definite programming. *J Mach Learn Res* 5(1):323–330
28. Li J, Wu Y, Zhao J, Lu K (2016) Low-rank discriminant embedding for multiview learning. *IEEE Transactions on Cybernetics* 1–14
29. Li SZ, Hou X, Zhang H, Cheng Q (2001) Learning spatially localized, parts-based representation. In: CVPR, IEEE, pp 207–212
30. Li X, Liu Q, He Z, Wang H, Zhang C, Chen W-S (2016) A multi-view model for visual tracking via correlation filters. *Knowl-Based Syst* 113:88–99
31. Liu J, Jiang Y, Li Z, Zhou Z-H, Lu H (2015) Partially shared latent factor learning with multiview data. *IEEE Trans Neural Netw Learning Syst* 26(6):1233–1246
32. Liu J, Jiang Y, Li Z, Zhou ZH, Lu H (2014) Partially shared latent factor learning with multiview data. *IEEE Trans Neural Netw Learning Syst* 26:1233–1246
33. Liu H, Ji R, Wu Y, Hua G Supervised matrix factorization for cross-modality hashing. arXiv:1603.05572
34. Ma Y, Hu X, He T, Jiang X (2016) Multi-view clustering microbiome data by joint symmetric nonnegative matrix factorization with Laplacian regularization. In: IEEE international conference on bioinformatics and biomedicine (BIBM), 2016, IEEE, pp 625–630
35. Ou W, Yu S, Li G, Lu J, Zhang K, Xie G (2016) Multi-view non-negative matrix factorization by patch alignment framework with view consistency. *Neurocomputing* 204:116–124
36. Rad R, Jamzad M (2017) Image annotation using multi-view non-negative matrix factorization with different number of basis vectors. *J Vis Commun Image Represent* 46:1–12
37. Shao W, He L, Lu C-T, Yu PS Online multi-view clustering with incomplete views. arXiv:1611.00481
38. Shao W, He L, Philip SY (2015) Multiple incomplete views clustering via weighted nonnegative matrix factorization with l21 regularization. In: Joint European conference on machine learning and knowledge discovery in databases. Springer, pp 318–334
39. Tang J, Hu X, Gao H, Liu H (2013) Unsupervised feature selection for multi-view data in social media. *Comput Sci Eng* 270–278

40. Wang J, Wang X, Tian F, Liu CH, Yu H, Liu Y (2016) Adaptive multi-view semi-supervised nonnegative matrix factorization. In: International conference on neural information processing. Springer, pp 435–444
41. Wang JJ-Y, Wang X, Gao X (2013) Non-negative matrix factorization by maximizing correntropy for cancer clustering. *BMC Bioinf* 14(1):1–11
42. Wang X, Wei B, Dacheng T (2013) Grassmannian regularized structured multi-view embedding for image classification. *IEEE Trans Image Process* 22(7):2646–2660
43. Wang Y, Zhang Y (2013) Non-negative matrix factorization: a comprehensive review. *IEEE Trans Knowl Data Eng* 25(6):1336–1353
44. Wang Z, Kong X, Fu H, Li M, Zhang Y (2015) Feature extraction via multi-view non-negative matrix factorization with local graph regularization. In: IEEE international conference on image processing (ICIP), 2015, IEEE, pp 3500–3504
45. Weifeng L, Dacheng T (2013) Multiview hessian regularization for image annotation. *IEEE Trans Image Process* 22(7):2676–2687
46. Wu F, Jing X, You X, Yue D, Hu R, Yang J (2016) Multi-view low-rank dictionary learning for image classification. *Pattern Recogn* 50:143–154
47. Xu C, Tao D, Xu C A survey on multi-view learning. [arXiv:1304.5634](https://arxiv.org/abs/1304.5634)
48. Xu C, Tao D, Xu C (2015) Multi-view intact space learning. *IEEE Trans Pattern Anal Mach Intell* 37(12):2531–2544
49. Xu W, Liu X, Gong Y (2003) Document clustering based on non-negative matrix factorization. In: Proceedings of the 26th annual international ACM SIGIR conference on research and development in informaion retrieval. ACM, pp 267–273
50. Yin Q, Wu S, He R, Wang L (2015) Multi-view clustering via pairwise sparse subspace representation. *Neurocomputing* 156:12–21
51. You X, Guo W, Yu S, Li K, Principe JC, Tao D (2016) Kernel learning for dynamic texture synthesis. *IEEE Trans Image Process* 25(10):4782–4795
52. You X, Ou W, Chen CLP, Li Q, Zhu Z, Tang Y (2015) Robust nonnegative patch alignment for dimensionality reduction. *IEEE Trans Neural Netw Learning Syst* 26(11):2760–2774
53. Zafeiriou S, Tefas A, Buciu I, Pitas I (2006) Exploiting discriminant information in nonnegative matrix factorization with application to frontal face verification. *IEEE Trans Neural Netw* 17(3):683–695
54. Zheng M, Bu J, Chen C, Wang C, Zhang L, Qiu G, Cai D (2011) Graph regularized sparse coding for image representation. *IEEE Trans Image Process* 20(5):1327–1336
55. Zhu Z, Du L, Zhang L, Zhao Y (2014) Shared subspace learning for latent representation of multi-view data. *Inform Hiding Multimedia Signal Proc* 5(3):546–554
56. Zong L, Zhang X, Zhao L, Yu H, Zhao Q (2017) Multi-view clustering via multi-manifold regularized non-negative matrix factorization. *Neural Netw* 88:74–89



Weihua Ou received the M.S. degree in Mathematics from the Southeast University, Nanjing, China in 2006 and the Ph.D. degree in Information and Communication Engineering from Huazhong University of Science and Technology (HUST), China in 2014, respectively. Currently, he is an Associate Professor at the School of Big data and Computer Science in Guizhou Normal University, Guiyang, China. His current research interests include sparse representation, multi-view learning, and cross-modal retrieval.



Fei Long is currently a professor in the School of Electrical and Information Engineering, Guizhou Institute of Technology. He received the Ph.D. degree from Southeast University in 2006. His research interest covers image processing and data mining.



Yi Tan was born in July 1990 in Sichuan, China. She is currently pursuing her M.S. Degree in the School of Information Science and Engineering, Guizhou University. Her research interests include machine learning.



Shujian Yu received the B.S. (Hons.) degree from the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China, in 2013. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA. His research interests include machine learning for signal processing, image analysis and understanding.



Pengpeng Wang was born in July 1992 in Yantai, China. She is currently pursuing her M.S. Degree in the School of Information Science and Engineering, Guizhou University. Her research interests include machine learning.