This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Co-Learning Non-Negative Correlated and Uncorrelated Features for Multi-View Data

Liang Zhao, Tao Yang, Jie Zhang, Zhikui Chen, Yi Yang, and Z. Jane Wang, *Fellow, IEEE*

***Abstract*— Multi-view data can represent objects from different perspectives and thus provide complementary information for data analysis. A topic of great importance in multi-view learning is to locate a low-dimensional latent subspace, where common semantic features are shared by multiple data sets. However, most existing methods ignore uncorrelated items (i.e., view-specific features) and may cause semantic bias during the process of common feature learning. In this article, we propose a non-negative correlated and uncorrelated feature co-learning (CoUFC) method to address this concern. More specifically, view-specific (uncorrelated) features are identified for each view when learning the common (correlated) feature across views in the latent semantic subspace. By eliminating the effects of uncorrelated information, useful inter-view feature correlations can be captured. We design a new objective function in CoUFC and derive an optimization approach to solve the objective with the analysis on its convergence. Experiments on real-world sensor, image, and text data sets demonstrate that the proposed method outperforms the state-of-the-art multiview learning methods.**

***Index Terms*— Co-learning, correlated features, multi-view data, uncorrelated features.**

I. Introduction

IN RECENT years, multi-view learning [1], [2] has been attracting increasing research attention in machine learning. In practical applications, multi-view data are common. Data about the same entity are often collected from different sources or channels with different descriptions, called modalities or views [3]. For example, in intelligent transportation systems, acoustic and seismic sensors are deployed to record different signals of one vehicle. In image recognition tasks, each image can be represented by diverse visual modalities, such as color histograms and texture structures. Also, in the dissemination of news, a specific piece of news is often written in different languages [4]–[6]. It is useful to take advantage of the information from different views about the same entity, as they represent the entity from different perspectives and thus can provide complementary information to enhance learning effectiveness. Multi-view learning is a useful technique to achieve this objective [7].

Multi-view learning has been an important machine learning research topic over the past decade [8]. Existing multi-view learning methods mostly aim to learn a unified function which combines all views and jointly optimizes them to improve the generalization performance [9], [10]. In particular, there has been some significant developments in the use of non-negative matrix factorization (NMF) for multi-view learning [11]–[14]. Nevertheless, these methods often ignore view-specific information, also called uncorrelated features, in each individual view during common subspace learning. In our previous work [15], this problem has been attempted by learning two projection matrices to transform the data matrix of each view to the common feature matrix and its individual information matrix in the latent subspace. Thus, the uncorrelated features are separated from the inter-view correlations, and a promising description of multi-view data can be obtained. However, there exist some major drawbacks in this approach: regarding the correlated feature matrix across views, it only depends on the corresponding projection matrix in each view. Besides, the method separates the uncorrelated information from the common features at the initial stage only, and the corresponding optimization of correlated feature matrix is independent of the uncorrelated feature matrix. This neglects the mutual interaction between them and cannot separate the remaining uncorrelated items in common features during the iterative optimization process.

To tackle these problems, we propose a new non-negative correlated and uncorrelated feature co-learning (CoUFC) method for multi-view data. It couples the correlated feature matrix and the uncorrelated ones together to reconstruct data matrices by corresponding basis matrices. Thus, they can be mutually updated in the iterative optimization process. In particular, the proposed CoUFC extends the multi-view NMF (multi-NMF) model and jointly factorizes data matrices of different views. Each data matrix is factorized into a consensus encoding matrix (correlated features), a view-specific encoding matrix (uncorrelated features) and their corresponding basis matrices. Through joint optimization and updating between inter-view correlated features and uncorrelated features in each individual view, the uncorrelated information for each view can be separated from the correlated features step by step during the iterative optimization process (not just at the initial

Manuscript received August 27, 2019; revised February 19, 2020; accepted March 28, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61906030 and in part by the Science and the Fundamental Research Funds for the Central Universities under Grant DUT18RC(3)025. *(Corresponding author: Yi Yang.)*

Liang Zhao, Tao Yang, Jie Zhang, and Zhikui Chen are with the Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, Dalian University of Technology, Dalian 116620, China, and also with the School of Software Technology, Dalian University of Technology, Dalian 116620, China.

Yi Yang is with the School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China (e-mail: yiyang@buaa.edu.cn).

Z. Jane Wang is with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC V6T 1Z4 Canada.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2020.2984810

stage), and thus more comprehensive multi-view correlations in the latent subspace can be obtained. This article contributes to the multi-view learning literature as follows.

1) In CoUFC, both correlated features and uncorrelated features are considered within the whole optimization process of learning the latent shared representation for multi-view data.
2) Local invariant graph regularization and the structured sparse regularizer are adopted to further improve the quality of feature correlations.
3) A new object function is designed and an efficient optimization approach is proposed to jointly transform and update correlated and uncorrelated features. Moreover, the convergence proof is given.

The proposed CoUFC method is evaluated on four real-world multi-view data sets, and the experimental results demonstrate its superiority when compared to state-of-the-art multi-view learning methods.

The remainder of this article is organized as follows. Section II gives a brief review of multi-NMF and some related works of this article. Section III describes the proposed CoUFC model and presents the optimization processes in detail. Section IV analyzes the convergence and the complexity of the proposed method. Experimental results and discussion are presented in Section V. Section VI gives the conclusion remarking.

II. Related Works

In this section, we briefly review the basic multi-NMF and some related works of this article.

A. Multi-NMF

Given the set of data instances with H views $\left\{X^{(v)}\right\}_{v=1}^{H}$, where $X^{(v)} \in R_{+}^{d_v \times n}$. The basic multi-NMF [16] tries to learn a latent common subspace by the shared encoding matrix V

$$\min \sum_{v=1}^{H} \|X^{(v)} - U^{(v)}V\|_F^2 \quad \text{s.t. } U^{(v)} \geq 0, \quad V \geq 0. \tag{1}$$

In this way, each instance across views is forced to have the same encoding in V and all the basis matrices for different views are coupled together by V. Thus, through joint optimization of different variables in (1) as that for NMF [17], the common features across views in the latent subspace can be obtained. The updating rules for $U^{(v)}$ in each view and V across views are as follows:

$$U_{ij}^{(v)} = \frac{(X^{(v)}V^T)_{ij}}{(U^{(v)}VV^T)_{ij}} U_{ij}^{(v)} \tag{2}$$

$$(V)_{ij} \leftarrow \frac{\left(\sum_{v=1}^{H} U^{(v)}X^{(v)}\right)_{ij}}{\left(\sum_{v=1}^{H} (U^{(v)})^T U^{(v)}V\right)_{ij}} (V)_{ij}. \tag{3}$$

Recently, many variants of basic multi-NMF have been proposed to address multi-view feature learning problems and promising results are achieved. In Section II-B, some related works will be described.

B. Multi-View Learning

A straightforward solution for multi-view learning is to concatenate all views into one single view, and exploit self-defined functions to solve it. However, this approach ignores the inherent structures and specific statistical properties of different views, and therefore makes sharing complementary information across multiple views disadvantageous. Besides, this naive combination usually incurs the curse of dimensionality, costing huge computational resources [16]. Another category of multi-view learning methods is called late fusion methods, which combine the learning results in each view to obtain the final decision [18]. Unfortunately, such methods may fail to sufficiently fuse the complementary information among views.

Multi-view latent subspace learning is an emerging area in multi-view learning. Its goal is to compute a low-dimensional common subspace (or feature representation) shared by all views [19], [20]. Many techniques can be generalized into multi-view subspace learning. They include canonical correlation analysis [10], NMF [11], low-rank constraint [21], and spectral embedding [22].

NMF is a promising technique for latent feature representation learning. It is based on the intuition of integrating the parts into the whole. In other words, the learned latent features in NMF can be interpreted by the corresponding basis components, and thus each view can be reconstructed in the latent subspace [17]. Recently, researchers proposed NMF-based multi-view learning methods, which achieved competitive results among existing methods [12]–[14]. They usually learn a variety of mapping matrices to transform all views into the common feature subspace and employ the regularization or consistency items to improve the comprehension of the latent features. For example, Liu *et al.* [23] proposed a typical NMF-based multi-view learning framework which formulates a joint matrix factorization process with the constraint that pushes the data in each view toward a common consensus instead of fixing it directly. After that, Ou *et al.* [24] developed the multi-view NMF by patch alignment framework with view consistency. It constructs a local patch to preserve the local geometric structure of each view and penalizes the disagreement across views via considering the view consistency. Zong *et al.* [14] designed a multi-manifold regularized NMF framework, in which the consensus manifold and consensus coefficient matrix are combined to preserve local geometric structures of the manifolds. Useful latent features and cluster patterns can then be obtained. In [25], a diverse NMF method is proposed to reduce the redundancy among multi-view representations with a novel defined diversity term, and thus the comprehensive and accurate multi-view features are learned. Besides, graph regularized NMF-based multi-view learning methods [26] have been designed, which can achieve promising performance.

Nevertheless, the above NMF-based multi-view approaches often ignore noises and uncorrelated information in the process of common feature learning. Some researchers [15], [27]–[31] proposed to tackle the noisy information in views by locating error matrices or uncorrelated features. For example,

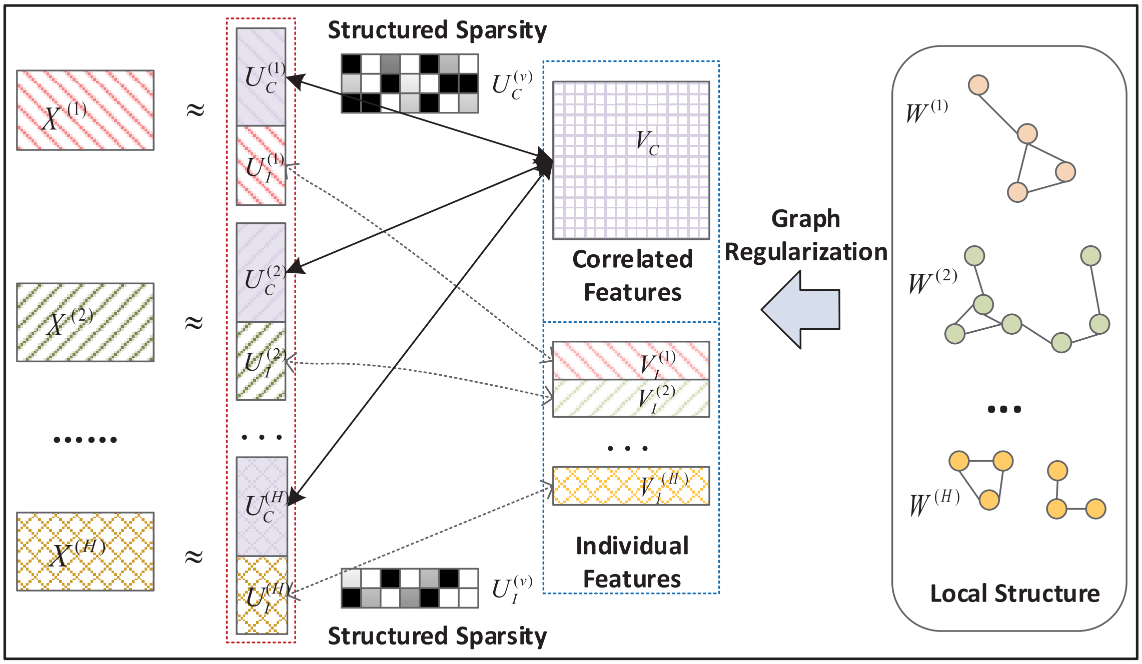


Fig. 1. Illustration of the work flow of the proposed CoUFC.

Ou *et al.* [29] developed a coregularized multi-NMF model with correlation constraint, which employs complementary information of multiple views to address the presence of the noisy views. In [30], a comprehensive latent factor learning (CLFL) method is proposed based on [23] to identify the specific latent information in each view. Moreover, Zhao *et al.* [15] attempted to identify the individual items in views by employing projection matrices to transform data matrices to correlated features and uncorrelated features in the latent subspace. Qiu *et al.* [31] extended [15] for multi-view correlated feature learning with dual graph-regularization. However, they only separated the uncorrelated information from the common features at the initial stage, and they cannot be jointly updated in the iterative optimization process. This article is an attempt to address this limitation. It employs the correlated features and uncorrelated features jointly to reconstruct data matrices by corresponding basis matrices, thus totally different models, optimization, and results are obtained. In particular, the uncorrelated information for each view is separated from the correlated features step by step during the iterative optimization process (not just at the initial stage), and thus more favorable results can be achieved. In Section III, we will describe our proposed model in detail.

III. Proposed CoUFC Method

In this section, we introduce our non-negative CoUFC method and give the optimization processes in detail. Fig. 1 illustrates the work flow of the proposed CoUFC. Regarding the data in each view $X^{(v)}$, it may contain some view-specific information, which cannot be shared across views. Thus, in the transformed subspace, the latent feature represented by consensus encoding matrix should be divided into the inter-view shared correlated feature V_C and the individual feature $V_I^{(v)}$ of each view. Correspondingly, using the basis matrices $U_C^{(v)}$ and $U_I^{(v)}$, the data matrix $X^{(v)}$ in each view is reconstructed. In this way, the view-specific information can be separated from the latent common features across views step by step by jointly optimizing the variables. Moreover, to preserve the local invariance structure of each view data $X^{(v)}$ and the shared feature V_C, the nearest neighbor graph is employed. Besides, the structured sparse regularizer is used for the basis matrices $U_C^{(v)}$ and $U_I^{(v)}$ to further improve the effectiveness of the proposed CoUFC.

A. CoUFC

NMF is able to combine the parts to form the whole, which is consistent with the cognitive process in human brains from psychological and physiological perspectives when dealing with multi-view data [17]. Inspired by this, we propose the CoUFC method by extending the traditional multi-NMF. Different from the traditional multi-NMF, the consensus encoding matrix in CoUFC is divided into two parts: 1) the inter-view shared encoding matrix and 2) the view-specific encoding matrix, as presented in Fig. 1. Therefore, two corresponding basis matrices are modeled for each view in CoUFC to reconstruct the original data matrix. Here, our new NMF-based multi-view learning objective is

$$\min \sum_{v=1}^{H} \left\| X^{(v)} - \left[U_C^{(v)}; U_I^{(v)} \right] \begin{bmatrix} V_C \\ V_I^{(v)} \end{bmatrix} \right\|_F^2$$
$$\text{s.t.:}\quad U_C^{(v)}, U_I^{(v)}, V_C, \quad V_I^{(v)} \geq 0 \tag{4}$$

where $X^{(v)} \in R^{d_v \times n}$ is the input data matrix of view v, with n instances and d_v-dimensions of attributes. $V_C \in R^{m_c \times n}$ and $V_I^{(v)} \in R^{m_v \times n}$ are the inter-view shared encoding matrix (correlated feature) and view-specific encoding matrix (uncorrelated feature), respectively. Correspondingly, $U_C^{(v)} \in R^{d_v \times m_c}$ and $U_I^{(v)} \in R^{d_v \times m_v}$ are the basis matrices. In this way, the input data of each view is modeled to have the same correlated feature V_C and the view-specific feature $V_I^{(v)}$ in the latent subspace. By jointly optimizing and updating these matrices coupled by V_C, a common feature representation across views can be obtained.

Similar to the traditional multi-NMF, this revised model still cannot preserve the local geometrical structure of the data space, which has been shown to be essential for preserving the data patterns [32]. We construct a local structure invariance graph to regularize the data matrix $X^{(v)}$ of each view and the shared feature matrix V_C. As presented in [32], the weight matrix $W^{(v)}$ for each view is firstly constructed based on a p-nearest neighbor graph. Then, the weight matrix and the similarity matrix, which measures the similarity between the low-dimensional representations of two data instances, are employed to model the local invariant graph regularization

$$\begin{aligned}\Re &= \frac{1}{2}\sum_{i,j=1}^{n} ||Z_i{}^T - Z_j{}^T||^2 W_{ij}^{(v)} \\ &= \sum_{i=1}^{n} Z_i Z_i{}^T D_{ii}^{(v)} - \sum_{i,j=1}^{n} Z_i Z_j{}^T W_{ij}^{(v)} \\ &= \mathrm{Tr}(Z^T D^{(v)} Z) - \mathrm{Tr}(Z^T W^{(v)} Z) \\ &= \mathrm{Tr}(Z^T L^{(v)} Z) \end{aligned} \tag{5}$$

where $\mathrm{Tr}(.)$ denotes the trace of a matrix. $Z = (V_C)^T$ and Z_i is the ith row of the matrix Z. $D^{(v)}$ is the diagonal matrix of $W^{(v)}$, the items of which are column or row sums of $W^{(v)}$. $L^{(v)} = D^{(v)} - W^{(v)}$ is the corresponding graph Laplacian matrix for view v. By minimizing the regularization item $\Re$, if the two instances $X_i^{(v)}$ and $X_j^{(v)}$ in view v are close (i.e., $W_{ij}^{(v)}$ is large), V_{Ci} and V_{Cj} are also close to each other in the latent subspace. Thus, the local geometric structure of the data can be preserved.

Regarding $U_C^{(v)}$ and $U_I^{(v)}$ for each view, we explore the structured sparse regularizer to make some basis items 0. This makes each view independent of the latent dimensions and can improve the flexibility and scalability of the proposed CoUFC. As presented in [3] and [15], $\ell_{2,1}$-norm is a useful way for measuring structured sparsity. Therefore, we define the sparsity constraint on $U_C^{(v)}$ and $U_I^{(v)}$ in each view as

$$\begin{aligned}\|U_C^{(v)}\|_{21} &= \sum_j \|U_{Cj}^{(v)}\|^2 \\ \|U_I^{(v)}\|_{21} &= \sum_j \|U_{Ij}^{(v)}\|^2. \end{aligned} \tag{6}$$

Finally, by substituting the local invariant graph regularization and the structured sparse regularizer into (4), the proposed CoUFC model can be expressed as

$$\begin{aligned}\min: \ &\sum_{v=1}^{H}\left(\left\| X^{(v)} - \left[U_C^{(v)}; U_I^{(v)}\right]\begin{bmatrix} V_C \\ V_I^{(v)} \end{bmatrix}\right\|_F^2 \right. \\ &\left. + \alpha \mathrm{Tr}\left(V_C L^{(v)} (V_C)^T\right) + \beta\left(\|U_C^{(v)}\|_{21} + \|U_I^{(v)}\|_{21}\right)\right) \\ \text{s.t.:}\ & U_C^{(v)}, U_I^{(v)}, V_C, V_I^{(v)} \geq 0. \end{aligned} \tag{7}$$

Herein, the regularization parameters $\alpha \geq 0$ and $\beta \geq 0$ control the smoothness of the common representation and the sparsity of the basis matrices, respectively. Through jointly optimizing and updating the inter-view correlated feature matrix and the individual-view uncorrelated feature matrices, comprehensive multi-view correlations in the latent subspace can be obtained.

B. Optimization

The optimization objective function in (7) is not convex with $U_C^{(v)}$, $U_I^{(v)}$, V_C and $V_I^{(v)}$ coupled together. Therefore, we can only find its local minima. To solve (7), we propose to alternatively optimize the block of variables ($U_C^{(v)}$, $U_I^{(v)}$, V_C or $V_I^{(v)}$) while keeping other blocks fixed.

Step 1: Fixing $U_C^{(v)}$, $U_I^{(v)}$,$V_I^{(v)}$ and updating the correlated feature matrix V_C. The minimization objective function over V_C is simplified as

$$\min_{V_C \geq 0} \sum_{v=1}^{H}\left(\left\| X^{(v)} - \left[U_C^{(v)}; U_I^{(v)}\right]\begin{bmatrix} V_C \\ V_I^{(v)} \end{bmatrix}\right\|_F^2 + \alpha \mathrm{Tr}\left(V_C L^{(v)} (V_C)^T\right)\right). \tag{8}$$

To optimize (8), the Lagrangian function is employed as

$$\begin{aligned} L = &\sum_{v=1}^{H}\Big(\mathrm{Tr}\Big(X^{(v)^T} X^{(v)}\Big) - 2\mathrm{Tr}\Big((V_C)^T U_C^{(v)^T} X^{(v)}\Big) \\ &- 2\mathrm{Tr}\Big(V_I^{(v)^T} U_I^{(v)^T} X^{(v)}\Big) + \mathrm{Tr}\Big((V_C)^T U_C^{(v)^T} U_C^{(v)} V_C\Big) \\ &+ 2\mathrm{Tr}\Big((V_C)^T U_C^{(v)^T} U_I^{(v)} V_I^{(v)}\Big) + \mathrm{Tr}\Big(V_I^{(v)T} U_I^{(v)^T} U_I^{(v)} V_I^{(v)}\Big) \\ &+ \alpha \mathrm{Tr}\big(V_C L^{(v)} (V_C)^T\big) + \mathrm{Tr}\big(\varphi^{(v)} V_C\big)\Big) \end{aligned} \tag{9}$$

where $\varphi^{(v)}$ is the Lagrangian multiplier for the constraint $V_C \geq 0$. By taking derivative of L with respect to V_C, we obtain

$$\begin{aligned}\frac{\partial L}{\partial V_C} = &\sum_{v=1}^{H}\Big(- 2U_C^{(v)^T} X^{(v)} + 2U_C^{(v)^T} U_C^{(v)} V_C \\ &+ 2U_C^{(v)^T} U_I^{(v)} V_I^{(v)} + 2\alpha V_C L^{(v)} + \varphi^{(v)}\Big). \end{aligned} \tag{10}$$

By applying the KKT condition $(\varphi^{(v)})_{ij}(V_C)_{ij} = 0$, the following updating rule for V_C is obtained:

$$(V_C)_{ij} \leftarrow \frac{\left(\sum_{v=1}^{H}\left(U_C^{(v)^T} X^{(v)} + \alpha V_C W^{(v)}\right)\right)_{ij} (V_C)_{ij}}{\left(\sum_{v=1}^{H}\left(U_C^{(v)^T} U_C^{(v)} V_C + U_C^{(v)^T} U_I^{(v)} V_I^{(v)} + \alpha V_C D^{(v)}\right)\right)_{ij}}. \tag{11}$$

Step 2: Fixing $U_C^{(v)}$, $U_I^{(v)}$,V_C and updating the uncorrelated feature matrix $V_I^{(v)}$ for each view. It can be observed from (7) that the optimization of $V_I^{(v)}$ is independent of the vth view when other variables are fixed. Thus, only the following objective function needs to be minimized for the vth view:

$$\min_{V_I^{(v)} \geq 0} \left\| X^{(v)} - \left[U_C^{(v)}; U_I^{(v)}\right]\begin{bmatrix} V_C \\ V_I^{(v)} \end{bmatrix}\right\|_F^2. \tag{12}$$

When employing the Lagrangian function and the KKT condition to optimize (12), we can update $V_I^{(v)}$ for each view as

$$\left(V_I^{(v)}\right)_{ij} \leftarrow \frac{\left(U_I^{(v)^T} X^{(v)}\right)_{ij} \left(V_I^{(v)}\right)_{ij}}{\left(U_I^{(v)^T} U_I^{(v)} V_I^{(v)} + U_I^{(v)^T} U_C^{(v)} V_C\right)_{ij}}. \tag{13}$$

Step 3: Fixing $U_I^{(v)}$, $V_I^{(v)}$,V_C and updating the correlated basis matrix $U_C^{(v)}$ for each view. Similar to (12), the objective function with respect to $U_C^{(v)}$ for each view is independent, which can be rewritten as

$$\min_{U_C^{(v)} \geq 0} \left\| X^{(v)} - \left[U_C^{(v)}; U_I^{(v)}\right] \begin{bmatrix} V_C \\ V_I^{(v)} \end{bmatrix} \right\|_F^2 + \beta \left\| U_C^{(v)} \right\|_{21}. \tag{14}$$

The Lagrangian function is constructed as was in Step 1. Then, its gradient with respect to $U_C^{(v)}$ is obtained by taking the first-order derivative

$$\begin{aligned} -2X^{(v)}(V_C)^T + 2U_C^{(v)} V_C (V_C)^T& \\ + \quad 2U_I^{(v)} V_I^{(v)} (V_C)^T + \beta D_C^{(v)} U_C^{(v)} + \phi^{(v)}& \end{aligned} \tag{15}$$

where $D_C^{(v)}$ is the diagonal matrix with the kth diagonal element defined by $(D_C^{(v)})_{kk} = 1/(||(U_C^{(v)})_k||^2)$, and $(U_C^{(v)})_k$ is the kth row of $U_C^{(v)}$. Using the KKT condition, we obtain the update rule for $U_C^{(v)}$ as

$$\left(U_C^{(v)}\right)_{ij} \leftarrow \frac{\left(X^{(v)}(V_C)^T\right)_{ij} \left(U_C^{(v)}\right)_{ij}}{\left(U_C^{(v)} V_C (V_C)^T + U_I^{(v)} V_I^{(v)} (V_C)^T + \frac{1}{2}\beta D_C^{(v)} U_C^{(v)}\right)_{ij}}. \tag{16}$$

Step 4: Fixing $U_C^{(v)}$, $V_I^{(v)}$, and V_C and updating the uncorrelated basis matrix $U_I^{(v)}$ for each view. We have

$$\min_{U_I^{(v)} \geq 0} \left\| X^{(v)} - \left[U_C^{(v)}; U_I^{(v)}\right] \begin{bmatrix} V_C \\ V_I^{(v)} \end{bmatrix} \right\|_F^2 + \beta \left\| U_I^{(v)} \right\|_{21}. \tag{17}$$

Similar to the optimized approach in (14), the uncorrelated basis matrix $U_I^{(v)}$ for each view can be updated by

$$\left(U_I^{(v)}\right)_{ij} \leftarrow \frac{\left(X^{(v)} V_I^{(v)^T}\right)_{ij} \left(U_I^{(v)}\right)_{ij}}{\left(U_I^{(v)} V_I^{(v)} V_I^{(v)^T} + U_C^{(v)} V_C V_I^{(v)^T} + \frac{1}{2}\beta D_I^{(v)} U_I^{(v)}\right)_{ij}} \tag{18}$$

where $D_I^{(v)}$ is the diagonal matrix and its elements are calculated following the same approach as computing $D_C^{(v)}$.

Following these four steps, $U_C^{(v)}$, $U_I^{(v)}$, V_C as well as $V_I^{(v)}$ are alternatively updated and uncorrelated features are thus dissociated from correlated features. When the objective function converges, useful multi-view correlations in the subspace are obtained. The proposed method is summarized in Algorithm 1.

It can be seen from Algorithm 1 that the proposed model can couple the correlated feature matrix and the uncorrelated feature matrix for each view together, which can be mutually updated in the process of optimization. Thus, the uncorrelated information for each view can be separated from the correlated features step by step (not just at the initial stage), and more favorable results can be obtained. These points of novelty bring significant advances compared to our previous work [15].

Algorithm 1 CoUFC

Input: 1) Data for H views $\{X^{(v)}\}_{v=1}^H$; 2) The expected dimensions of the common feature V_C and the view-specific feature $V_I^{(v)}$ for each view; and 3) The regularization parameters α and β.

Output: The common feature V_C, and the view-specific feature $V_I^{(v)}$ for each view

$t = 0$.
Initialize the common feature matrix V_C and the view-specific feature matrix $V_I^{(v)}$, the corresponding basis matrices $U_C^{(v)}$, $U_I^{(v)}$ for each view randomly.
Calculate the weight distance matrix $W^{(v)}$ and it diagonal matrix $D^{(v)}$ for each view.
repeat
  Update the common feature matrix V_C via Eq. (11), and the view-specific feature matrix $V_I^{(v)}$ via Eq. (13) for each view.
  Calculate the diagonal matrices $D_C^{(v)}$ and $D_I^{(v)}$.
  Update the basis matrices $U_C^{(v)}$, $U_I^{(v)}$ for each view by Eqs. (16) and (18), respectively.
  $t = t + 1$.
until Converges
return V_C, $V_I^{(v)}$ for each view.

IV. Analysis

A. Convergence

The objective function for CoUFC is not jointly convex with $U_C^{(v)}$, $U_I^{(v)}$, V_C and $V_I^{(v)}$ coupled together. Thus, we divide (7) into four subproblems [i.e., (8), (12), (14), and (17)]. If the convergence of the four subproblems can be proved, a local minima of the objective function can be achieved by iterative updating as presented in Algorithm 1 [33]. When update V_C with $U_C^{(v)}$, $U_I^{(v)}$, and $V_I^{(v)}$ fixed, the decrease of the objective function can be proved by the following theorem.

Theorem 1: Suppose $\{U_C^{(v)}\}_{v=1}^H$, $\{U_I^{(v)}\}_{v=1}^H$ and $\{V_I^{(v)}\}_{v=1}^H$ are fixed. When updating $V_C^{(t)}$ to $V_C^{(t+1)}$ by using (11), the objective function value of (7) monotonically decreases

$$\begin{aligned} \sum_{v=1}^H \Bigg(\Bigg(& \left\| X^{(v)} - U_C^{(v)} V_C^{(t+1)} - S_I^{(v)} \right\|_F^2 \\ & - \left\| X^{(v)} - U_C^{(v)} V_C^{(t)} - S_I^{(v)} \right\|_F^2 \Bigg) \\ & + \alpha \mathrm{Tr}\left(V_C^{(t+1)} L^{(v)} \left(V_C^{(t+1)}\right)^T - V_C^{(t)} L^{(v)} (V_C^{(t)})^T\right)\Bigg) \leq 0 \end{aligned} \tag{19}$$

in which $S_I^{(v)} = U_I^{(v)} \times V_I^{(v)}$. t stands for the tth update.

Proof: Let $P(V_C)$ be

$$\begin{aligned} \sum_{v=1}^H \mathrm{Tr}\Big(& X^{(v)^T} X^{(v)} - 2(V_C)^T U_C^{(v)^T} X^{(v)} + (V_C)^T U_C^{(v)^T} U_C^{(v)} V_C \\ & -2S_I^{(v)^T} X^{(v)} + 2(V_C)^T U_C^{(v)^T} S_I^{(v)} + S_I^{(v)^T} S_I^{(v)} \\ & + \alpha (V_C)^T V_C L^{(v)}\Big). \end{aligned} \tag{20}$$

Then, we can reformulate (19) to be

$$P\left(V_C^{(t+1)}\right) - P(V_C) \leq 0. \tag{21}$$

To prove the inequality in (21), an auxiliary function about V_C is introduced from [17] as

$$\begin{aligned} F(V_C, \widehat{V_C}) &= \sum_{v=1}^{H} \mathrm{Tr}\Big(X^{(v)^T} X^{(v)} - 2(V_C)^T U_C^{(v)^T} X^{(v)} \\ &\quad - 2S_I^{(v)^T} X^{(v)} + 2(V_C)^T U_C^{(v)^T} S_I^{(v)} + S_I^{(v)^T} S_I^{(v)}\Big) \\ &\quad + \sum_{v=1}^{H} \left(\sum_{i=1}^{m_c} \sum_{j=1}^{n} \left(\frac{\left(U_C^{(v)^T} U_C^{(v)} \widehat{V_C}\right)_{ij}}{\left(\widehat{V_C}\right)_{ij}} (V_C)_{ij}^2 \right.\right. \\ &\quad \left.\left. + \frac{\alpha\left(\widehat{V_C} L^{(v)}\right)_{ij}}{\left(\widehat{V_C}\right)_{ij}} (V_C)_{ij}^2 \right)\right). \end{aligned} \tag{22}$$

According to [34], the following matrix inequality can be achieved:

$$\mathrm{Tr}((V_C)^T A V_C B) \leq \sum_{i,j} \frac{\left(A\widehat{V_C}B\right)_{ij}}{\left(\widehat{V_C}\right)_{ij}} (V_C)_{ij}^2 \tag{23}$$

where A, B, and V_C are non-negative matrices, and $A = A^T$, $B = B^T$. If $A = U_C^{(v)^T} U_C^{(v)}$, $B = I_n$, in which I_n is an $n \times n$ identity matrix, then we can get

$$\mathrm{Tr}\Big((V_C)^T U_C^{(v)^T} U_C^{(v)} V_C\Big) \leq \sum_{i=1}^{m_c}\sum_{j=1}^{n} \frac{\left(U_C^{(v)^T} U_C^{(v)} \widehat{V_C}\right)_{ij}}{\left(\widehat{V_C}\right)_{ij}} (V_C)_{ij}^2. \tag{24}$$

If $A = I_{m_c}$, $B = L^{(v)}$, where I_{m_c} is an $m_c \times m_c$ identity matrix, the following matrix inequality is obtained:

$$\mathrm{Tr}\Big((V_C)^T V_C L^{(v)}\Big) \leq \sum_{i=1}^{m_c}\sum_{j=1}^{n} \frac{\left(\widehat{V_C} L^{(v)}\right)_{ij}}{\left(\widehat{V_C}\right)_{ij}} (V_C)_{ij}^2. \tag{25}$$

Therefore, we can achieve $P(V_C) \leq F(V_C, \widehat{V_C})$. When and only when $V_C = \widehat{V_C}$, $P(V_C) = F(V_C, \widehat{V_C})$. Moreover, the optimal V_C can be obtained by minimizing $F(V_C, \widehat{V_C})$. Let $f(V_C) = F(V_C, \widehat{V_C})$, we can see that it is a convex function and it has a global minima. By setting $(\partial f(V_C)/\partial (V_C)_{ij}) = 0$, we can get

$$(V_C)_{ij} = \frac{\left(\sum_{v=1}^{H}\left(U_C^{(v)^T} X^{(v)}\right)\right)_{ij} \left(\widehat{V_C}\right)_{ij}}{\left(\sum_{v=1}^{H}\left(U_C^{(v)^T} U_C^{(v)} \widehat{V_C} + U_C^{(v)^T} S_I^{(v)} + \alpha \widehat{V_C} L^{(v)}\right)\right)_{ij}}. \tag{26}$$

If setting $V_C^{(t+1)} = V_C$, and $V_C^{(t)} = \widehat{V_C}$, (26) is equal to the updating rule (11) for V_C. Also, $f(V_C^{(t+1)}) \leq f(V_C^{(t)})$, that is, $F(V_C^{(t+1)}, V_C^{(t)}) \leq F(V_C^{(t)}, V_C^{(t)})$, thus we can get

$$\begin{aligned} P\left(V_C^{(t+1)}\right) &= F\left(V_C^{(t+1)}, V_C^{(t+1)}\right) \leq F\left(V_C^{(t+1)}, V_C^{(t)}\right) \\ &\leq F\left(V_C^{(t)}, V_C^{(t)}\right) = P\left(V_C^{(t)}\right) \end{aligned} \tag{27}$$

which means that $P(V_C)$ is monotonically decreasing, namely the inequality (21) is hold. Hence, Theorem 1 is proved.

Similarly, we can prove the updating rules for $U_C^{(v)}$, $U_I^{(v)}$, and $V_I^{(v)}$ are monotonically decreasing. Therefore, we can show that the objective function value of (7) will converge to local minima under the update steps in (11), (13), (16) and (18), respectively. In the experiments, we will further verify the convergence of CoUFC when tested on the real data sets.

B. Complexity

In CoUFC, we need to allocate storage buffers for the data matrix $X^{(v)}$, the view-specific feature matrix $V_I^{(v)}$, the basis matrices $[U_C^{(v)}, U_I^{(v)}]$, the weight distance matrix $W^{(v)}$ for each view, the common feature matrix V_C, and the regularization parameters α and β. The corresponding space consumptions are $O(d_v n)$, $O(m_v n)$, $O(d_v m_c + d_v m_v)$, $O(n^2)$, $O(m_c n)$ and $O(1)$, respectively. In total, the space complexity for V views approximates to $O(H(d_v n + m_v n + d_v m_c + d_v m_v + n^2) + m_c n + 1) \approx O(n^2)$.

The time complexity consists of five parts, corresponding to the subproblems for updating V_C, $\{V_I^{(v)}\}_{v=1}^H$, $\{U_C^{(v)}\}_{v=1}^H$, $\{U_I^{(v)}\}_{v=1}^H$ and calculating the weight distance matrices $\{W^{(v)}\}_{v=1}^H$. In each iteration, the time costs are close to $O(Hm_c d_v n)$, $O(Hm_v d_v n)$, $O(Hm_c d_v n)$ and $O(Hm_v d_v n)$ for V_C, $\{V_I^{(v)}\}_{v=1}^H$, $\{U_C^{(v)}\}_{v=1}^H$ and $\{U_I^{(v)}\}_{v=1}^H$, respectively. Besides the multiplicative updates, constructing the p-nearest neighbor graph for $\{W^{(v)}\}_{v=1}^H$ costs $O(Hd_v n^2)$. Suppose that the updating process stops after t iterations, the overall time complexity for CoUFC is approximately $O(tH(m_c d_v n + m_v d_v n + m_c d_v n + m_v d_v n) + Hd_v n^2) \approx O(tHm_c d_v n + Hd_v n^2)$. Regarding the existing state-of-the-art methods, the overall time complexity of multi-NMF [23], multi-manifold regularization NMF (MMNMF) [14], unsupervised multi-view non-negative correlated feature learning (UMCFL) [15] and adaptive dual graph-regularized multi-view non-negative feature learning (ADMFL) [31] approximates to $O(tHm_c d_v n)$, $O(tHm_c d_v n + Hd_v n^2)$, $O(tHm_c d_v n + Hd_v n^2)$ and $O(tHm_c d_v n + Hd_v n^2)$, respectively. Herein, m_c is dimension of common features and d_v is the dimension of the vth view. When the number of data instances n is large enough, the running time of multi-NMF is linear, whereas other methods are quadratic. This is because MMNMF, UMCFL, ADMFL, and our CoUFC need to calculate the graph Laplacian matrix. However, the extra time cost will improve the effectiveness of the methods to some extent.

V. Experiments

In this section, we evaluate CoUFC by clustering problem as the compared methods [14], [15], [23], [30]–[32] on four real-world multi-view data sets: 1) SensIT vehicle [6]; 2) multiple features,[1]; 3) ALOI[2]; and 4) 3Sources.[3] The performance is evaluated using three clustering metrics: 1) normalized mutual information (NMI); 2) purity (PUR); and 3) accuracy (ACC) [35].

A. Data Sets

We give a brief description of the four multi-view data sets tested in the experiments in Table I.

[1] https://archive.ics.uci.edu/ml/datasets/Multiple+Features
[2] http://elki.dbs.ifi.lmu.de/wiki/datasets/MultiView
[3] http://mlg.ucd.ie/datasets/3sources.html

TABLE I
DESCRIPTION OF THE DATA SETS

Datasets	# of instances	# of views	# of classes
SensIT Vehicle	300	2	3
Multiple Features	2,000	6	10
ALOI	11,040	2	100
3Sources	169	3	6

SensIT Vehicle data set is collected from wireless sensor networks in intelligent transportation systems, in which the acoustic and seismic sensors are used to record the signals of three types of vehicles. As presented in [6], we randomly sampled 100 data instances for each class. Thus, there are 300 samples with two views in three classes for the experiments.

Multiple Feature data set consists of 2000 images in six views for ten handwritten digit classes (0–9). Each class contains 200 images. In the experiments, the 240-D pixel averages, 216-D profile correlations, 76-D Fourier coefficients, 64-D Karhunen-Love coefficients, and 47-D Zernike moment features are used for common feature learning.

ALOI data set is a collection of images with 110 250 instances for 1000 small objects. Similar to the settings in [14], we select the 11 040 instances with two views, RGB color histograms and HSV/HSB color histograms, for 100 classes to do the experiments.

3Sources data set is collected from three well-known online news sources, BBC, Reuters, and The Guardian. Each source is regarded as one view on certain news events and presented in text format. In total, the 169 stories that reported in all three sources are selected as in [14].

B. Evaluation Criteria

Three popular evaluation criteria [35], including NMI, PUR, and ACC, are employed to evaluate the effectiveness of compared methods.

NMI, which uses information theoretic measure, evaluates the effectiveness of clustering methods by calculating the mutual information between the real cluster labels and actual clustering results. It can be defined as

$$\mathrm{NMI} = \frac{\sum_{j=1}^{c}\sum_{i=1}^{k} n_{ij}\log\left(\frac{nn_{ij}}{n_j n_i}\right)}{\sqrt{\left(\sum_{j=1}^{c} n_j \log\left(\frac{n_j}{n}\right)\right)\left(\sum_{i=1}^{k} n_i \log\left(\frac{n_i}{n}\right)\right)}} \tag{28}$$

in which n_{ij} is the number of common instances in class (refer to real cluster labels) $j \in \{1, 2, \ldots, c\}$ and cluster (refer to actual clustering results) $i \in \{1, 2, \ldots, k\}$. n is the size of the data set, n_i and n_j are the numbers of objects in ith cluster and jth class, respectively.

PUR is formulated as

$$\mathrm{PUR}=\sum_{i=1}^{k}\frac{\max_j(n_{ij})}{n} \tag{29}$$

where n_{ij}, $j \in \{1, 2, \ldots, c\}$, $i \in \{1, 2, \ldots, k\}$, and n are the same to that in the definition of NMI.

ACC is a more direct measure to reflect the effectiveness of clustering results, which is defined as

$$\delta_j(m_i) = \max(n_{ij}) \tag{30}$$

$$\mathrm{ACC} = \sum_{t=1}^{c}\frac{\mathrm{map}(t, \delta_j(m_i))}{n}. \tag{31}$$

Herein, $\delta_j(m_i)$ indicates the maximum number of common instances in cluster i corresponding to different class $j \in \{1, 2, \ldots, c\}$, and $\mathrm{map}(t, \delta_j(m_i))$ is the tth largest value for all $\delta_j(m_i), i \in \{1, 2, \ldots, k\}$ with no duplication of corresponding classes label $j \in \{1, 2, \ldots, c\}$.

Higher value of NMI, PUR, and ACC implies that better clustering performance is obtained.

C. Experimental Settings

Since NMF-based multi-view learning has shown the competitiveness among existing methods, we compare CoUFC, also an NMF based method, with them, such as: 1) best single view NMF (BSNMF) [17]; 2) feature concatenation NMF (ConcatNMF) [17]; 3) multi-NMF [23]; 4) multi-view graph regularized NMF (multi-GNMF) [32]; 5) NMF-based CLFL [30]; 6) MMNMF [14]; 7) UMCFL [15]; and 8) ADMFL [31]. In BSNMF, the single view NMF [17] is employed to learn the latent feature for each view and the best result is selected. ConcatNMF constructs the input data matrix by concatenating the features of all views. Then the NMF is used to learn the latent feature. Multi-NMF extends NMF to multi-view data and regularizes the coefficient matrices of different views toward a common consensus. Multi-GNMF combines the graph regularized NMF [32] with multi-view data to preserve the local geometrical structure of data instances for common feature learning. CLFL uses multi-NMF to learn the common latent factor and integrates view-specific latent factor together for optimal representation of multi-view data. MMNMF incorporates consensus manifold and consensus coefficient matrix to preserve the local geometrical structure of the manifolds for multi-view learning. UMCFL projects view-specific features in each view and correlated features across views into different latent subspace with the constraint of local invariance graph regularization for multi-view common feature learning. ADMFL extends UMCFL with the dual graph-regularization of both data and feature manifolds to model the distribution of data points in the common subspace.

The parameters for these methods are all set to be the same as that in their original papers. For example, $\lambda_v = 0.01$ for multi-NMF [23], $\lambda = 100$ for multi-GNMF [32], $\beta = 0.7, \gamma = 0.005$ for CLFL [30] and $\lambda = \eta = 50, \gamma = 1.3, u = 1.2$ for ADMFL [31]. In [14], there are four kinds of MMNMF methods. We select MMNMF-R-C which has superior performance to the other three, for comparison. The parameters are $\beta = 2$, $\gamma = 1$, and $\eta = 1$.

To evaluate the methods, K-means is employed to cluster the learned features, and the NMI, PUR, and ACC values are used to measure the clustering performance. For NMF-based methods, we need to specify the dimension of the latent subspace. To make a fair comparison, we set it to be the

TABLE II
CLUSTERING PERFORMANCE COMPARISONS ON THE SENSIT VEHICLE DATA SET

Methods	NMI	PUR	ACC
BSNMF	0.008±0.005	0.380±0.012	0.379±0.013
ConcatNMF	0.018±0.009	0.413±0.007	0.407±0.007
MultiNMF	0.035±0.023	0.435±0.045	0.424±0.049
MultiGNMF	0.064±0.016	0.483±0.017	0.473±0.023
CLFL	0.032±0.022	0.422±0.035	0.417±0.038
MMNMF	0.086±0.023	0.550±0.029	0.548±0.027
UMCFL	0.075±0.027	0.467±0.033	0.461±0.038
ADMFL	0.098±0.002	0.523±0.003	0.525±0.002
CoUFC	**0.255**±0.020	**0.645**±0.023	**0.630**±0.024

TABLE III
CLUSTERING PERFORMANCE COMPARISONS ON THE 3SOURCES DATA SET

Methods	NMI	PUR	ACC
BSNMF	0.503±0.019	0.513±0.025	0.502±0.029
ConcatNMF	0.560±0.063	0.575±0.082	0.554±0.099
MultiNMF	0.481±0.030	0.503±0.045	0.478±0.043
MultiGNMF	0.609±0.047	0.646±0.075	0.634±0.077
CLFL	0.441±0.032	0.639±0.036	0.645±0.040
MMNMF	**0.657**±0.023	0.644 ±0.047	0.633±0.041
UMCFL	0.559±0.016	0.551±0.007	0.535±0.008
ADMFL	0.513±0.021	0.707±0.012	0.563±0.017
CoUFC	0.586±0.037	**0.718**±0.036	**0.692**±0.030

same as the number of the classes (which has been shown to be appropriate in [14], [17], and [32]) for all methods. Regarding the dimension of the uncorrelated feature for each view in CoUFC, UMCFL, and ADMFL, the values are selected through the experiments to achieve the best performance on all data sets. Moreover, the parameters α and β in CoUFC are empirically selected as in UMCFL to achieve the best clustering performance as well. Since all the compared methods depend on initialization, each experiment is repeated 20 times by random initialization and the average performance is reported.

D. Results

In the first group of experiments, we validate the effectiveness of the proposed method on SensIT Vehicle and 3Sources data sets, containing uncorrelated items. Specifically, in SensIT Vehicle, some noises caused by sensor faults are collected. In 3Sources, the same story is usually represented by different word vectors, which contain source-specific information. Both the noises and the source-specific information are the uncorrelated items. Therefore, we use these two data sets to asses if our proposed method can truly deal with correlated and uncorrelated features. The clustering performance of the methods on the tested data sets is reported in Tables II and III.

It can be observed that, compared with single-view methods (BSNMF and ConcatNMF), the multi-view methods (multi-NMF, multi-GNMF, CLFL, MMNMF, UMCFL, ADMFL, and CoUFC) can improve clustering results dramatically. Regarding the multi-view methods, CoUFC significantly outperforms other methods on these data sets. For example, on SensIT Vehicle, CoUFC can improve the clustering performances by 15.7%, 9.5%, and 8.2% in terms of NMI, PUR, and ACC

TABLE IV
CLUSTERING PERFORMANCE COMPARISONS ON THE MULTIPLE FEATURES DATA SET

Methods	NMI	PUR	ACC
BSNMF	0.663±0.037	0.724±0.046	0.704±0.056
ConcatNMF	0.692±0.021	0.744±0.039	0.735±0.050
MultiNMF	0.789±0.030	0.776±0.034	0.747±0.051
MultiGNMF	0.722±0.030	0.749±0.039	0.731±0.049
CLFL	0.837±0.023	0.828±0.028	0.804±0.026
MMNMF	0.783±0.042	0.816±0.034	0.810±0.047
UMCFL	0.848±0.012	0.842±0.006	0.822±0.011
ADMFL	**0.871**±0.003	0.847±0.016	**0.844**±0.021
CoUFC	0.855±0.009	**0.850**±0.009	**0.844**±0.020

compared to the second best method ADMFL on NMI or MMNMF on PUR and ACC (see Table II), respectively. This is because the view-specific information for each view is integrated to the latent subspace for multi-view learning in CoUFC. Therefore, by cross optimization, the uncorrelated items can be removed from the correlated feature step by step (not just at the initial stage), and promising results can be obtained. Even though UMCFL and ADMFL can also learn the uncorrelated feature and correlated feature together [15], [31], they separate the uncorrelated information from the common features at the initial stage only, and the corresponding optimization of common feature matrix is independent of the uncorrelated feature matrix. This neglects the mutual effect between them, and thus some uncorrelated items still exist in the latent common subspace. Moreover, we can see from Table II that CoUFC has facorable standard deviation compared to other methods, which indicates that its performance is robust. Even if at the worst case, it is superior to the second best method MMNMF on PUR and ACC or ADMFL on NMI.

From the experimental results on 3Sources data set, it can be seen that our proposed CoUFC also outperforms other compared methods, which once again proves that CoUFC is suitable for multi-view data with uncorrelated information. It is shown in Table III that UMCFL and ADMFL are unfavorable for processing text data. They have much lower clustering performance (NMI and ACC) than multi-GNMF and MMNMF. However, CoUFC overcomes this problem successfully. It achieves comparable performance in terms of NMI and almost 7% improvement in terms of PUR and ACC compared to multi-GNMF and MMNMF. Moreover, its performance is robust during the 20 times tests.

In the second group of experiments, we validate the performance of the methods on two widely used multi-view image data sets, multiple features and ALOI, which are relatively clean compared to SensIT Vehicle and 3Sources. The clustering performance of the methods on these two data sets is reported in Tables IV and V.

It can be seen when the data sets contain little noise, UMCFL or ADMFL is a promising choice for multi-view learning as well. As shown in Table IV, the performance of UMCFL and ADMFL are comparable to that of CoUFC and superior to other methods in terms of NMI, PUR and ACC on the multiple feature data set. Also, CoUFC, UMCFL, and ADMFL are more robust to other methods on the 20 times

TABLE V
CLUSTERING PERFORMANCE COMPARISONS ON THE ALOI DATA SET

Methods	NMI	PUR	ACC
BSNMF	0.494±0.009	0.382±0.009	0.309±0.010
ConcatNMF	0.659±0.008	0.504±0.010	0.447±0.011
MultiNMF	0.717±0.009	0.574±0.017	0.504±0.017
MultiGNMF	0.757±0.008	0.594±0.021	0.549±0.019
CLFL	0.764±0.007	0.603±0.011	0.587±0.013
MMNMF	0.789±0.008	0.586±0.016	0.556±0.020
UMCFL	0.754±0.004	0.598±0.003	0.560±0.007
ADMFL	0.772±0.004	0.626±0.009	0.594±0.014
CoUFC	**0.791**±0.003	**0.653**±0.012	**0.618**±0.017

tests. Overall, there is not much difference in the average performance of NMI, PUR or ACC for all the multi-view methods. This is because the data set Multiple Features is clean. However, our proposed method CoUFC is still the best compared to other methods.

From the results on data set ALOI shown in Table V, we can see that the multi-view methods, multi-NMF, multi-GNMF, CLFL, MMNMF, UMCFL, and ADMFL, achieve similar performance in terms of NMI, PUR and ACC. However, CoUFC is much better than them, which once again verifies that CoUFC can favorably separate the uncorrelated items from the common feature, though only a few noises are embedded in the data set.

In summary, the proposed CoUFC outperforms the state-of-the-art methods on all the tested multi-view data sets. Moreover, CoUFC can get significant performance improvement on multi-view data sets (SensIT Vehicle and 3Sources) containing uncorrelated items.

E. Parameter Selection

CoUFC has two essential parameters: the regularization parameters α and β. In this section, we investigate how they affect the performance of CoUFC in terms of NMI on all data sets. In the experiments, the two parameters are sampled and tuned jointly to obtain the best performance. Fig. 2 presents the results. It can be observed that, with increasing α and β, the NMI value is stable at first and then decreases dramatically on 3Sources. However, on other data sets, the NMI values achieved remain relatively stable with different α and β value pairs, except a rapid start on the SensIT Vehicle data set and the Multiple Features data set. We greedily assign the parameters to be [10, 1], [1, 0.01], [100, 1] and [100000, 10] for the corresponding data sets (see Fig. 2) in the experiments.

F. Convergence Analysis

In this article, each subproblem is iteratively updated to compute the local minima. We have proven that the subproblems are convergent. Here, we illustrate how they converge. Fig. 3 shows the convergence curves of the objective functions on four tested data sets, in which the x-axis means the number of iterations and the y-axis are means the objective function values. The objective function values converge very rapidly on all data sets, usually within ten iterations for SensIT Vehicle, Multiple Features, ALOI, and 50 iterations for 3Sources.

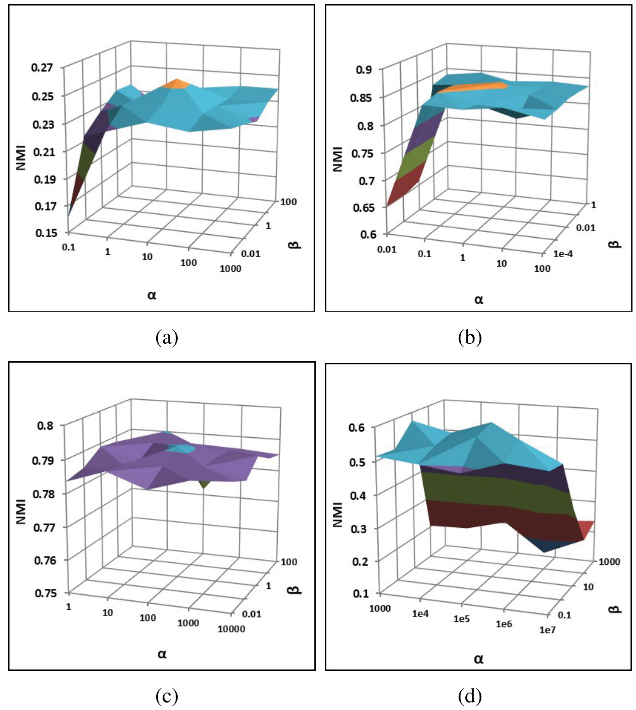


Fig. 2. Performances of CoUFC versus parameters α and β on (a) SensIT Vehicle, (b) Multiple Features, (c) ALOI, and (d) 3Sources data sets.

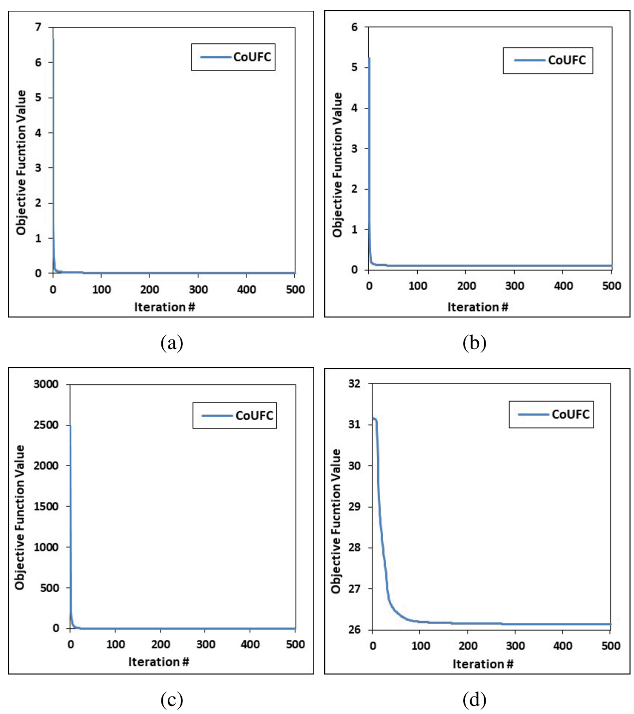


Fig. 3. Convergence curves of CoUFC on (a) SensIT Vehicle, (b) Multiple Features, (c) ALOI, and (d) 3Sources data sets.

VI. CONCLUSION

In this article, we propose a novel non-negative CoUFC method for multi-view data. Different from existing subspace learning methods, we split the latent features into view-specific items and inter-view correlations in the subspace, and propose a new multi-view learning model. To improve its quality and scalability, local invariant graph regularization and structured sparse regularizer are integrated into the model. We further develop an optimization algorithm to iteratively solve the proposed nonsmooth problem with proven convergence. Experiments on four real-word multi-view data sets demonstrate the superiority of the proposed method when compared with eight state-of-the-art approaches.

REFERENCES

[1] S. Sun, "A survey of multi-view machine learning," *Neural Comput. Appl.*, vol. 23, nos. 7–8, pp. 2031–2038, Dec. 2013.

[2] J. Zhao, X. Xie, X. Xu, and S. Sun, "Multi-view learning overview: Recent progress and new challenges," *Inf. Fusion*, vol. 38, pp. 43–54, Nov. 2017.

[3] W. Shao, L. He, C.-T. Lu, X. Wei, and P. S. Yu, "Online unsupervised multi-view feature selection," in *Proc. IEEE 16th Int. Conf. Data Mining (ICDM)*, Dec. 2016, pp. 1203–1208.

[4] Q. Zhang, L. T. Yang, X. Liu, Z. Chen, and P. Li, "A tucker deep computation model for mobile multimedia feature learning," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 13, no. 3s, pp. 1–18, Aug. 2017.

[5] L. Zhao, Z. Chen, Y. Yang, Z. Jane Wang, and V. C. M. Leung, "Incomplete multi-view clustering via deep semantic mapping," *Neurocomputing*, vol. 275, pp. 1053–1062, Jan. 2018.

[6] X. Cai, F. Nie, and H. Huang, "Multi-view k-means clustering on big data," in *Proc. 23rd Int. Joint Conf. Artif. Intell. IJCAI*, Jun. 2013, pp. 2598–2604.

[7] Z. Tao, H. Liu, S. Li, Z. Ding, and Y. Fu, "From ensemble clustering to multi-view clustering," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 2843–2849.

[8] Y. Li, M. Yang, and Z. Zhang, "A survey of multi-view representation learning," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 10, pp. 1863–1883, Oct. 2019.

[9] J. He, C. Du, C. Du, F. Zhuang, Q. He, and G. Long, "Nonlinear maximum margin multi-view learning with adaptive kernel," in *Proc. 26Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 1830–1836.

[10] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 188–194, Jan. 2016.

[11] H. Zhao, Z. Ding, and Y. Fu, "Multi-view clustering via deep matrix factorization," in *Proc. AAAI*, 2017, pp. 2921–2927.

[12] X. Zhang *et al.*, "Multi-view clustering based on graph-regularized nonnegative matrix factorization for object recognition," *Inf. Sci.*, vol. 432, pp. 463–478, Mar. 2018.

[13] C. Xu, Z. Guan, W. Zhao, Y. Niu, Q. Wang, and Z. Wang, "Deep multi-view concept learning," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 2898–2904.

[14] L. Zong, X. Zhang, L. Zhao, H. Yu, and Q. Zhao, "Multi-view clustering via multi-manifold regularized non-negative matrix factorization," *Neural Netw.*, vol. 88, pp. 74–89, Apr. 2017.

[15] L. Zhao, Z. Chen, and Z. J. Wang, "Unsupervised multiview nonnegative correlated feature learning for data clustering," *IEEE Signal Process. Lett.*, vol. 25, no. 1, pp. 60–64, Jan. 2018.

[16] Z. Guan, L. Zhang, J. Peng, and J. Fan, "Multi-view concept learning for data representation," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 11, pp. 3016–3028, Nov. 2015.

[17] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2001, pp. 556–562.

[18] E. Bruno and S. Marchand-Maillet, "Multiview clustering: A late fusion approach using latent models," in *Proc. 32nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. SIGIR*, 2009, pp. 736–737.

[19] C. Zhang, Q. Hu, H. Fu, P. Zhu, and X. Cao, "Latent multi-view subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4333–4341.

[20] S. Luo, C. Zhang, W. Zhang, and X. Cao, "Consistent and specific multi-view subspace clustering," in *Proc. 32nd AAAI Conf. Artif. Intell.*, Apr. 2018, pp. 3730–3737.

[21] Z. Ding and Y. Fu, "Robust multi-view subspace learning through dual low-rank decompositions," in *Proc. AAAI*, 2016, pp. 1181–1187.

[22] J. W. Son, J. Jeon, A. Lee, and S.-J. Kim, "Spectral clustering with brainstorming process for multi-view data," in *Proc. AAAI*, 2017, pp. 2548–2554.

[23] J. Liu, C. Wang, J. Gao, and J. Han, "Multi-view clustering via joint nonnegative matrix factorization," in *Proc. SIAM Int. Conf. Data Mining*, May 2013, pp. 252–260.

[24] W. Ou, S. Yu, G. Li, J. Lu, K. Zhang, and G. Xie, "Multi-view nonnegative matrix factorization by patch alignment framework with view consistency," *Neurocomputing*, vol. 204, pp. 116–124, Sep. 2016.

[25] J. Wang, F. Tian, H. Yu, C. H. Liu, K. Zhan, and X. Wang, "Diverse nonnegative matrix factorization for multiview data representation," *IEEE Trans. Cybern.*, vol. 48, no. 9, pp. 2620–2632, Sep. 2018.

[26] Y. Meng, R. Shang, L. Jiao, W. Zhang, and S. Yang, "Dual-graph regularized non-negative matrix factorization with sparse and orthogonal constraints," *Eng. Appl. Artif. Intell.*, vol. 69, pp. 24–35, Mar. 2018.

[27] C. Bhattacharya, N. Goyal, R. Kannan, and J. Pani, "Non-negative matrix factorization under heavy noise," in *Proc. Int. Conf. Mach. Learn.*, Jun. 2016, pp. 1426–1434.

[28] L. Yang, L. Jing, and M. K. Ng, "Robust and non-negative collective matrix factorization for text-to-image transfer learning," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4701–4714, Dec. 2015.

[29] W. Ou, F. Long, Y. Tan, S. Yu, and P. Wang, "Co-regularized multiview nonnegative matrix factorization with correlation constraint for representation learning," *Multimedia Tools Appl.*, vol. 77, no. 10, pp. 12955–12978, May 2018.

[30] H. Zheng, Z. Liang, F. Tian, and Z. Ming, "NMF-based comprehensive latent factor learning with multiview da," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 489–493.

[31] X. Qiu, Z. Chen, L. Zhao, and C. Hu, "Unsupervised multi-view nonnegative for law data feature learning with dual graph-regularization in smart Internet of things," *Future Gener. Comput. Syst.*, vol. 100, pp. 523–530, Nov. 2019.

[32] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.

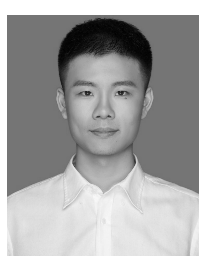
[33] L. Yang, L. Jing, B. Liu, and J. Yu, "Common latent space identification for heterogeneous co-transfer clustering," *Neurocomputing*, vol. 269, pp. 29–39, Dec. 2017.

[34] C. H. Q. Ding, T. Li, and M. I. Jordan, "Convex and semi-nonnegative matrix factorizations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 45–55, Jan. 2010.

[35] L. Zhao, Z. Chen, Y. Yang, L. Zou, and Z. J. Wang, "ICFS clustering with multiple representatives for large data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 3, pp. 728–738, Mar. 2019.

Liang Zhao received the M.S. and Ph.D. degrees in software engineering from the Dalian University of Technology, Dalian, China, in 2014 and 2018, respectively.

He is currently a Teacher with the School of Software Technology, Dalian University of Technology. His research interests include big data and AI.



Tao Yang is currently pursuing the B.S. degree in network engineering with the Dalian University of Technology, Dalian, China.

His research interests include incremental data mining and multiview data fusion.

Jie Zhang received the B.S. degree in software engineering from Northeast Normal University, Changchun, China, in 2019. She is currently pursuing the master's degree with the School of Software Technology, Dalian University of Technology, Dalian, China.

Her research interests include machine learning and data fusion.

Zhikui Chen received the B.S. degree from the Department of Mathematics and Computer Science, Chongqing Normal University, Chongqing, China, and the M.S. degree in mechanics and the Ph.D. degree in digital signal processing from Chongqing University, Chongqing, in 1990, 1993, and 1998, respectively.

He is currently a Full Professor with the Dalian University of Technology, Dalian, China, where he is leading the Institute of Ubiquitous Network and Computing. His research interests include big data processing, mobile cloud computing, and ubiquitous networks and its computing.

Yi Yang received the Ph.D. degree from the Nanjing University of Science and Technology, Nanjing, China, in 2008.

She was involved in post-doctoral research at the School of Reliability and Systems Engineering, Beihang University, Beijing, China, where she is currently a Professor with the School of Reliability and Systems Engineering. Her main research interests include reliability analysis and design, repairable systems, and control science and engineering.

Z. Jane Wang (Fellow, IEEE) received the B.Sc. degree from Tsinghua University, Beijing, China, in 1996, and the Ph.D. degree from the University of Connecticut, Storrs, CT, USA, in 2002.

Since August 2004, she has been with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC, Canada, where she is currently a Professor. Her research interests are in the broad areas of statistical signal processing theory and applications.