

Received 12 February 2023, accepted 30 March 2023, date of publication 5 April 2023, date of current version 18 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3264837

## METHODS

# Diversity Multi-View Clustering With Subspace and NMF-Based Manifold Learning

JIAMAN DING, XIAOJIANG FANG, LIANYIN JIA, YING JIANG, AND RUNXIN LI

Faculty of Information Engineering and Automation, Kunming University, Kunming, Yunnan 650500, China  
Key Laboratory of Computer Technologies Application of Yunnan Province, Kunming University, Kunming, Yunnan 650500, China

Corresponding author: Lianyin Jia (lianyinjia@kust.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 62262034 and Grant 62262035.

**ABSTRACT** Since the complementarity information among multiple views has been exploited to improve the clustering effect significantly, multi-view clustering has become a hot topic, and many multi-view clustering methods have emerged. Most of them only consider local features in each view, ignoring the differences in the manifold structure of the same class samples among different views. In addition, they need to balance the importance of respective views effectively, thus ignoring the diversity among views in the clustering process. To address these problems, we propose a new diversity multi-view clustering method with subspace and NMF-based manifold learning. Firstly, non-negative matrix factorization and manifold learning are used to obtain features and local geometric structures of samples. After that, the latent space representation facilitates the transfer of manifold structural features between views and improves the class consistency of the same sample in different views. Moreover, the Hilbert-Schmidt independence criterion is introduced to learn diversity for mutual learning and information fusion among views. Finally, experiments on seven datasets demonstrate the superiority of the proposed method compared to ten state-of-the-art methods.

**INDEX TERMS** Multi-view latent clustering (MVLC), Hilbert Schmidt independence criterion (HSIC), non-negative matrix factorization (NMF), manifold learning (ML).


## I. INTRODUCTION

With the rapid development of big data, real-world data generated from different sources or observed from different views is becoming richer in terms of semantics and structure. For instance, image data can be represented using a heterogeneous set of features such as local binary patterns, colour layout descriptors, etc. These datasets consisting of multiple views can provide much richer information. Multi-view clustering (MVC) is learning from various views simultaneously to produce consistency and general grouping information for datasets.

Multi-view clustering has achieved success in many fields, such as natural language processing [1], computer vision [2], big data [3], biomedical information analysis [4], image classification [5], [6], [7], face clustering [8], [9], community detection [10], [11], [12], and so on. So far, many multi-view

clustering methods based on spectrum [13], subspace [14], [15], and non-negative matrix factorization [16] have been proposed. Among them, the development of methods based on non-negative matrix factorization is remarkable. Multi-view clustering methods using NMF have received much attention because they can learn effective low-rank representations by mapping the original high-rank to low-rank space.

NMF generally produces inconsistent clustering results, partly due to the inability of NMF-based methods to preserve the geometric structure of the original data when mapping it into low-rank space [7]. In other words, adjacent samples in the original space may not remain adjacent in the new space. Cai et al. [17] introduced the concept of manifold learning in single-view clustering, aiming to preserve local or global geometric structure of the original data. Zhang et al. [18], Qian et al. [19] introduced the idea of manifold learning to multi-view clustering, combining manifold learning with a non-negative matrix factorization framework to explore and maintain the geometric structure of

The associate editor coordinating the review of this manuscript and approving it for publication was Zeev Zalevsky .

low-rank data. The multi-view clustering method based on NMF and manifold learning show effectiveness against other state-of-the-art methods. The main idea of manifold learning is to maintain similar manifold structures of the same class samples in low-rank space. However, another difficulty in multi-view clustering is learning the exact consensus matrix from multiple views and then using the valid consensus matrix for clustering. In learning the consensus matrix, NMF-based manifold learning is more likely to use the non-negative factorized data coefficient matrix as the consensus matrix, which ignores different manifold structures that exist in different views, even for the same class samples. Therefore, in recent years, many works have attempted to obtain more reliable coefficient matrices in various ways. For example, Guo et al. [20] proposed a matrix factorization LRR model based on product Grassmann manifolds, using the kernel norm of the matrix to obtain a low-rank representation of the original data and a more reliable coefficient matrix. Zhang et al. [21] attempted to fuse the clustering indicator matrices of different views into the consensus matrix, thus avoiding noise and redundancy in the original data. Wang et al. [22] restrict the number of connected components to be consistent with the sample categories and thus learn a uniform similarity matrix from each view. However, these methods may ignore the connection among the original data, as the manifold structure of the same class samples tends to differ among views. This difference makes the learned consensus matrix affect the next clustering operation.

There is complementarity among multi-view datas. Data in a single view can get more information from other views. Zhang et al. [23] proposed to use the complementary reconstructed data of multi-view to obtain latent representations. They assumed that multi-view data could be composed of exact latent representations according to different mapping relationships. The consensus matrix formation is natural and easy for subsequent clustering operations. In addition, Zhang et al. [24] extended it to nonlinear mappings, which experimentally proved simple but effective. However, this method ignores the manifold structure between sample points after mapping. Cao et al. [25] adopted the Hilbert-Schmidt independence criterion(HSIC) to facilitate rich representations of other view data, specifying that the learned subspace representations are novel to each other. Tan et al. [26] borrowed ideas from this and took complete account of low-rank factorisation and noise information to achieve data fusion.

Unlike reconstructing the latent representation in this paper, our method is not to reconstruct the entire latent space using prior knowledge. Firstly, our methods obtain the view features and geometric structures from the samples in each view using non-negative matrix factorization and then facilitate the interconnection of geometric structure information among different views through the latent space. At the same time, each sample can be considered a linear combination of latent representations and individual view features. We use latent representations to explore the manifold relationships

between the same class data points in different views and combine manifold learning with exploring the manifold structure among data points in each view. In addition, the Hilbert Schmidt Independence Criterion is also used for feature learning among views. We assume that each view's manifold structure of the same class samples is often formed with the exact latent feature representations. Based on this assumption, we propose a new method called Diversity Multi-view Clustering with Subspace and NMF-based Manifold Learning (DMVCSN). Using the idea of non-negative matrix factorization, our method learns the latent features of the manifold structure of the same class samples from the complementary information of each view. Meanwhile, manifold learning can maintain the manifold structure between samples in different views. Besides, HSIC can promote the interconnection of information among views and accelerate mutual learning in model optimization. We use the augmented Lagrangian multiplier to optimize the algorithm with an alternating direction minimization strategy efficiently. Finally, we conduct extensive experiments to compare our method with the state-of-the-art method to demonstrate its effectiveness.

The main contributions of the proposed DMVCSN algorithm are as follows:

- 1) Based on non-negative matrix factorization with subspace learning, we propose a method called Diversity Multi-view Clustering with Subspace and NMF-based-Manifold Learning, which can learn a more reliable coefficient matrix using the data manifold features of multiple views.
- 2) Unlike latent representation learning, DMVCSN employs data manifold feature representation to maintain the internal structure between data points in different views.
- 3) We introduce the Hilbert-Schmidt independence criterion to learn diversity for mutual learning, fuse the information among the views to construct a unified low-rank multi-view clustering method, and design an alternating optimization algorithm using an augmented Lagrangian multiplier. Experimental results show that the proposed method is competitive with some state-of-the-art methods.

The rest of this paper is organized as follows. We present the related work and technical background, including multi-view NMF, Manifold Learning, HSIC covariance constraints, and latent subspace clustering in Section II. The proposed method and its optimization procedure and convergence analysis are then described in Section III. The experimental results and analysis are shown in detail in Section IV. Finally, some conclusions are drawn in Section V.

## II. RELATED WORKS

This section briefly describes the background of the techniques used in our proposed method, including Non-negative matrix factorization, Manifold learning, HSIC covariance constraints, and latent subspace clustering.

### A. NON-NEGATIVE MATRIX FACTORIZATION (NMF)

NMF [27] is a widely popular method in multi-view learning [16], [19], which decomposes a given matrix  $\mathbf{X}$  into a non-negative basis matrix  $\mathbf{U}$ , and a non-negative coefficient matrix  $\mathbf{V}$ , and obtains a subspace representation in the form of a product. which is

$$\mathbf{X} \approx \mathbf{UV}^T$$

where  $\mathbf{X} = [x_1, x_2, \dots, x_n] \in \mathbb{R}^{m \times n}$ ,  $m$  is the number of sample features,  $n$  is the number of samples, and each column of  $\mathbf{X}$  represents one sample. The elements in both the base matrix  $\mathbf{U} = [u_{ij}] \in \mathbb{R}^{m \times l}$  and the coefficient matrix  $\mathbf{V} = [v_{ij}] \in \mathbb{R}^{n \times l}$  satisfy  $u_{ij} \geq 0, v_{ij} \geq 0$ .

To measure Euclidean distance of the following objective functions.

$$O_1 = \|\mathbf{X} - \mathbf{UV}^T\|_F^2$$

s.t.  $\mathbf{U} \geq \mathbf{0}, \mathbf{V} \geq \mathbf{0}$

NMF is used as matrix factorization and the coefficient matrix can be used for later clustering operations.

### B. MANIFOLD LEARNING (ML)

The NMF method decomposes higher order data matrices into smaller matrices while generating many approximate solutions, often producing unstable results. It does not consider the internal structure of the data matrix. Cai et al. [17] combined manifold learning with the NMF framework, which helps to produce data matrices that are conducive to clustering while maintaining the original data geometry. After, many works [22], [28] attempted to use k-nearest neighbors to capture local similarity relationships between data points. Based on KNN, there are three general representative methods for constructing similarity graph methods as follows.

- 1) *Binary weighting*: When constructing a similarity graph.  $w_{ij}$  is the weight between two nodes, which represents the similarity between them.
- 2) *Heat kernel weighting*:

$$w_{ij} = e^{-\frac{\|x_i - x_j\|^2}{\sigma}}$$

- 3) *Dotproduct weighting*:

$$w_{ij} = \mathbf{x}_j^T \cdot \mathbf{x}_i$$

In the low-dimensional representation  $\mathbf{v}_j = [v_{j1}, v_{j2}, \dots, v_{jl}]^T$  of the original sample point  $x_i$ . The Euclidean distance

$$d(\mathbf{v}_i, \mathbf{v}_j) = \|\mathbf{v}_i - \mathbf{v}_j\|^2$$

is used to measure the difference between the two data points relative to the original sample points. Using the above

weights, the smoothness of the low-dimensional representation can be measured using the following equation.

$$\begin{aligned} O_2 &= \sum_{i,j=1}^n \|\mathbf{v}_i - \mathbf{v}_j\|^2 w_{ij} \\ &= \sum_{i=1}^n \mathbf{v}_i^T \mathbf{v}_i D_{ii} - \sum_{i,j=1}^n \mathbf{v}_i^T \mathbf{v}_j w_{ij} \\ &= \text{Tr}(\mathbf{V}^T \mathbf{D} \mathbf{V}) - \text{Tr}(\mathbf{V}^T \mathbf{W} \mathbf{V}) \\ &= \text{Tr}(\mathbf{V}^T \mathbf{L} \mathbf{V}) \end{aligned}$$

### C. THE HILBERT SCHMIDT INDEPENDENCE CRITERION

Combining complementary information between views is a critical way to improve clustering. The Hilbert-Schmidt independence criterion (HSIC) allows data from other views to be used as new features to enrich the data representation. According to references [25], [29], the cross-covariance  $C_{xy}$  can be defined as:

$$C_{xy} = E_{xy}[(\phi(x) - \mu_x) \otimes (\varphi(y) - \mu_y)]$$

where  $\mu_x, \mu_y$  is the expectation of  $x, y$ , i.e.,  $\mu_x = E(\phi(x))$ ,  $\mu_y = E(\varphi(y))$ , and  $\otimes$  represents the matrix product.

*Definition 1:* Given two separable reproducing kernel Hilbert space  $\Gamma, \Upsilon$  and a joint probability distribution  $\rho_{xy}$ , we define the HSIC as the Hilbert Schmidt norm of the associated cross-covariance operator  $C_{xy}$

$$\mathbf{HSIC}(\rho_{xy}, \Gamma, \Upsilon) = \|C_{xy}\|_{HS}^2$$

where  $\|\mathbf{A}\|_{HS}$  denotes the Hilbert-Schmidt norm of a matrix as:

$$\|\mathbf{A}\|_{HS} = \sqrt{\sum_{i,j} a_{ij}^2}$$

However, the joint distribution  $\rho_{xy}$  is often unknown or hard to estimate. Thus, the empirical version of HSIC is induced as follow:

*Definition 2:* Consider a series of  $n$  independent observations drawn from  $\rho_{xy}$ ,  $\mathcal{Z} := \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \in \Gamma \times \Upsilon$ , written as  $\mathbf{HSIC}(\mathcal{Z}, \Gamma, \Upsilon)$ , is given by:

$$\mathbf{HSIC}(\mathcal{Z}, \Gamma, \Upsilon) = (n-1)^{-2} \text{Tr}(\mathbf{K}_1 \mathbf{H} \mathbf{K}_2 \mathbf{H})$$

where  $\mathbf{K}_1, \mathbf{K}_2$  are two Gram matrices, and  $\mathbf{H}$  is the central matrix. For more details about HSIC, please refer to the literature [25], [29]. In order for the data in different views to provide more complementary interest, HSIC is used to penalize the dependencies between the newly represented data.

### D. LATENT MULTI-VIEW SUBSPACE CLUSTERING

The complementary information in multiple views can reconstruct the latent space, and further clustering operations can be done in the latent space. This method assumes that the data in different views come from the exact latent representation,

as shown in Fig. 1. The objective function of subspace clustering based on latent representation  $\mathbf{H}$  can be defined as:

$$\min_{\mathbf{P}, \mathbf{H}} L_h(\mathbf{X}^{(v)}, \mathbf{P}^{(v)} \mathbf{H})$$

where the individual views  $\mathbf{X}^v$  are obtained from the shared latent representation  $\mathbf{H} = \{h\}_{i=1}^n$  using the respective projection models  $\{\mathbf{P}^{(1)}, \dots, \mathbf{P}^{(n_v)}\}$ .  $L_h(\cdot, \cdot)$  is the loss function associated with the latent representation. In general, by virtue of the complementarity between the views, the latent representation contains more comprehensive information than the single-view representation.

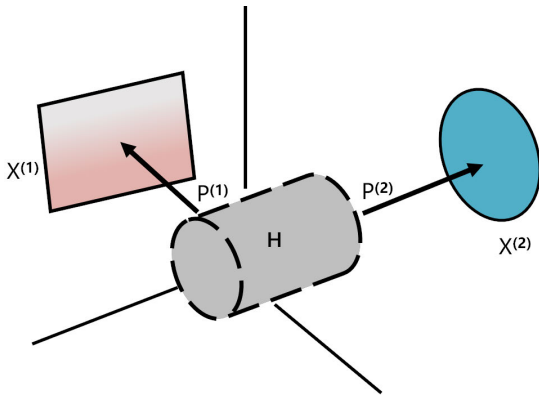


FIGURE 1. Schematic demonstration of multi-view latent representation.

### III. THE PROPOSED METHOD

In order to simultaneously considers both the geometric structures of the data manifold and fusion information among views, we propose a multi-view diversity clustering model based on non-negative matrix factorization(DMVCSN). In this section, we present the objective function, its optimization and solution scheme, and the convergence proof.

#### A. OBJECTIVE FUNCTION OF DMVCSN

The data are defined as  $\{\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(v)}, \dots, \mathbf{X}^{(n_v)}\}$ , where  $n_v$  denotes the number of total views and  $\mathbf{X}^{(v)} \in \mathbb{R}^{d_v \times n}$  denotes the data in the  $v$ -th view.  $n$  is the number of sample points, and  $d_v$  is the number of features of the  $v$ -th view.  $\mathbf{U}^{(v)} \in \mathbb{R}^{d_v \times l}$  and  $\mathbf{P}^{(v)} \in \mathbb{R}^{n \times m}$  are the feature matrix and projection matrix of the  $v$ -th view respectively. And  $\mathbf{V}^* \in \mathbb{R}^{n \times l}$  is the set of latent sample features, and different views share the same matrix variable  $\mathbf{V}^*$ . we transformed the problem into a non-negative matrix triple factorization [30], [31], and its objective function is

$$\min_{\mathbf{U} \geq 0, \mathbf{P} \geq 0, \mathbf{V}^* \geq 0} \sum_{v=1}^{n_v} \|\mathbf{X}^{(v)} - \mathbf{U}^{(v)}(\mathbf{P}^{(v)} \mathbf{V}^*)^T\|_F^2 \quad (1)$$

The above equation decomposes the original data matrix  $\mathbf{X}$  into three parts, where  $\mathbf{U}$  preserves the features of individual views, whereas matrix  $\mathbf{P}\mathbf{V}^*$  preserves the structure among the data points. Borrowing from latent space representation, the

matrix  $\mathbf{V}^*$  transfers the manifold structure features among the views.

Non-negative matrix triple factorization is an approximate process with a certain degree of error that ignores the connections between data points. We combine manifold learning to obtain an accurate data manifold structure and a latent representation of the data manifolds among views. Meanwhile, to avoid the projection matrix  $\mathbf{P}^{(v)}$  falling into the trivial solution, the regular term is added and the modified objective function is as follows.

$$\begin{aligned} \min_{\mathbf{U} \geq 0, \mathbf{P} \geq 0, \mathbf{V}^* \geq 0} & \sum_{v=1}^{n_v} \|\mathbf{X}^{(v)} - \mathbf{U}^{(v)}(\mathbf{P}^{(v)} \mathbf{V}^*)^T\|_F^2 \\ & + \alpha_v \sum_{v=1}^{n_v} \text{Tr}((\mathbf{P}^{(v)} \mathbf{V}^*)^T \mathbf{L}_{p^{(v)}}^{(v)} (\mathbf{P}^{(v)} \mathbf{V}^*)) \\ & + \beta_v \sum_{v=1}^{n_v} \|\mathbf{P}^{(v)}\|_F^2 \end{aligned} \quad (2)$$

where  $\mathbf{L}_{p^{(v)}}^{(v)} = \mathbf{D}^{(v)} - \mathbf{W}_{p^{(v)}}^{(v)}$ ,  $\mathbf{D}_{p^{(v)}}^{(v)}$  is the diagonal matrix of the  $v$ -th view, and the elements on the diagonal are  $D_{ii}^{(v)} = \sum_{j=1}^n w_{ij}^{(v)}$ . In the second term, each view collectively maintains a latent data manifold representation using its own data manifold. This method preserves the individual view coefficient matrix manifold structure and allows the latent data manifold space to be used to transfer manifold features among different views.

Compared to manifold learning works [18], [19], constraints of the above function may be more relaxed. Therefore, we added the Hilbert-Schmidt independence criterion to our function and got the final objective function of DMVCSN as follows.

$$\begin{aligned} \min_{\mathbf{U} \geq 0, \mathbf{P} \geq 0, \mathbf{V}^* \geq 0} & \sum_{v=1}^{n_v} \|\mathbf{X}^{(v)} - \mathbf{U}^{(v)}(\mathbf{P}^{(v)} \mathbf{V}^*)^T\|_F^2 \\ & + \alpha_v \sum_{v=1}^{n_v} \text{Tr}((\mathbf{P}^{(v)} \mathbf{V}^*)^T \mathbf{L}_{p^{(v)}}^{(v)} (\mathbf{P}^{(v)} \mathbf{V}^*)) \\ & + \beta_v \sum_{v=1}^{n_v} \|\mathbf{P}^{(v)}\|_F^2 \\ & + \gamma_v \sum_{w=1, v \neq w}^{n_v} \text{HSIC}(\mathbf{P}^{(v)} \mathbf{V}^*, \mathbf{P}^{(w)} \mathbf{V}^*) \end{aligned} \quad (3)$$

Refer to (3), the first term is the standard NMF factorization. The matrix is factorized into three non-negative parts; The second term is a smoothing term to maintain a common manifold feature space while capturing the manifold structure of each view itself, ensuring that adjacent samples in the original space remain adjacent in the new low-dimensional representation; The third term is the standard term, it is so as not to fall into a trivial solution. The fourth term is to use the idea of HSIC to fuse the complementarity among views to reduce redundancy and improve the accuracy of clustering results.

### B. MODEL OPTIMIZATION

The alternating direction multiplier can change the joint optimization of main variables into the separate alternating iterations, so our objective function is rewritten as (ignoring the index of the view)

$$\begin{aligned} O(\mathbf{U}, \mathbf{P}, \mathbf{V}^*) &= \text{Tr}(\mathbf{X} - \mathbf{UV}^{*T} \mathbf{P}^T)^T (\mathbf{X} - \mathbf{UV}^{*T} \mathbf{P}^T) \\ &+ \alpha \text{Tr}((\mathbf{PV}^*)^T \mathbf{L}_{pv^*} (\mathbf{PV}^*)) \\ &+ \beta \text{Tr}(\mathbf{P}^T \mathbf{P}) \\ &+ \gamma \text{HSIC}(\mathbf{P}^{(v)} \mathbf{V}^*, \mathbf{P}^{(w)} \mathbf{V}^*) \end{aligned} \quad (4)$$

In this paper, we use  $\mathbf{K}^v = (\mathbf{P}^v \mathbf{V}^*)(\mathbf{P}^v \mathbf{V}^*)^T$  as the inner product kernel. For the convenience of representation, the HSIC part is represented by:

$$\begin{aligned} &\sum_{w=1, w \neq v}^{n_v} \text{HSIC}(\mathbf{V}^{(v)}, \mathbf{V}^{(w)}) \\ &= \sum_{w=1, w \neq v}^{n_v} \text{Tr}(\mathbf{HK}^{(v)} \mathbf{HK}^{(w)}) \\ &= \sum_{w=1, w \neq v}^{n_v} \text{Tr}((\mathbf{P}^v \mathbf{V}^*)^T \mathbf{HK}^{(w)} \mathbf{H}(\mathbf{P}^v \mathbf{V}^*)) \\ &= \text{Tr}((\mathbf{P}^v \mathbf{V}^*)^T \mathbf{K}(\mathbf{P}^v \mathbf{V}^*)) \end{aligned} \quad (5)$$

where  $\mathbf{K} = \sum_{w=1, w \neq v}^{n_v} \mathbf{HK}^{(w)} \mathbf{H}$ . Equation (4) can be rewritten as

$$\begin{aligned} O(\mathbf{U}, \mathbf{P}, \mathbf{V}^*) &= \text{Tr}(\mathbf{X}^T \mathbf{X}) - 2 \text{Tr}(\mathbf{X}^T \mathbf{UV}^{*T} \mathbf{P}^T) \\ &+ \text{Tr}((\mathbf{UV}^{*T} \mathbf{P}^T)^T (\mathbf{UV}^{*T} \mathbf{P}^T)) \\ &+ \alpha \text{Tr}((\mathbf{PV}^*)^T \mathbf{L}_{pv^*} (\mathbf{PV}^*)) \\ &+ \beta \text{Tr}(\mathbf{P}^T \mathbf{P}) \\ &+ \gamma \text{Tr}((\mathbf{P}^v \mathbf{V}^*)^T \mathbf{K}(\mathbf{P}^v \mathbf{V}^*)) \end{aligned} \quad (6)$$

Notice that when we update the variables of the one view, the parameters of the other views are fixed, and the last term  $\mathbf{K}$  in (6) is fixed.

To solve (6) under the given constraints, we introduced the Lagrange multipliers to our objective function, including  $\Phi$ ,  $\Psi$  and  $\Omega$ . So (6) is rewritten as:

$$\begin{aligned} \mathcal{L} &= \text{Tr}(\mathbf{X}^T \mathbf{X}) - 2 \text{Tr}(\mathbf{X}^T \mathbf{UV}^{*T} \mathbf{P}^T) \\ &+ \text{Tr}((\mathbf{UV}^{*T} \mathbf{P}^T)^T (\mathbf{UV}^{*T} \mathbf{P}^T)) \\ &+ \alpha \text{Tr}((\mathbf{PV}^*)^T \mathbf{L}_{pv^*} (\mathbf{PV}^*)) \\ &+ \beta \text{Tr}(\mathbf{P}^T \mathbf{P}) + \gamma \text{Tr}((\mathbf{P}^v \mathbf{V}^*)^T \mathbf{K}(\mathbf{P}^v \mathbf{V}^*)) \\ &- \text{Tr}(\Phi \mathbf{U}) - \text{Tr}(\Psi \mathbf{P}) - \text{Tr}(\Omega \mathbf{V}^*) \end{aligned} \quad (7)$$

#### 1) UPDATING U AND P

By taking derivatives of (7) on U and P, we have

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{U}} &= \mathbf{UV}^{*T} \mathbf{P}^T \mathbf{PV}^* - \mathbf{XPV}^* - \Phi \\ \frac{\partial \mathcal{L}}{\partial \mathbf{P}} &= \mathbf{PV}^* \mathbf{U}^T \mathbf{UV}^{*T} + \beta \mathbf{P} + \gamma \mathbf{KPV}^* \mathbf{V}^{*T} \\ &- \mathbf{X}^T \mathbf{UV}^{*T} - \alpha \mathbf{L}_{pv^*} \mathbf{PV}^* \mathbf{V}^{*T} - \Psi \end{aligned}$$

Combined with the Karush-Kuhn-Tucker conditions  $\Phi_{ij} U_{ij} = 0$  and  $\Psi_{ij} P_{ij} = 0$ , setting  $\frac{\partial \mathcal{L}}{\partial \mathbf{U}} = 0$  and  $\frac{\partial \mathcal{L}}{\partial \mathbf{P}} = 0$  lead to

$$\begin{aligned} (\mathbf{UV}^{*T} \mathbf{P}^T \mathbf{PV}^* - \mathbf{XPV}^*)_{ij} U_{ij} &= 0 \\ (\mathbf{PV}^* \mathbf{U}^T \mathbf{UV}^{*T} + \beta \mathbf{P} + \gamma \mathbf{KPV}^* \mathbf{V}^{*T} \\ - \mathbf{X}^T \mathbf{UV}^{*T} - \alpha \mathbf{L}_{pv^*} \mathbf{PV}^* \mathbf{V}^{*T})_{ij} P_{ij} &= 0 \end{aligned}$$

Combining  $L_{pv^*} = D_{pv^*} - W_{pv^*}$ , the updating scheme is get as following

$$u_{ij} \leftarrow u_{ij} \frac{(\mathbf{XPV}^*)_{ij}}{(\mathbf{UV}^{*T} \mathbf{P}^T \mathbf{PV}^*)_{ij}} \quad (8)$$

$$p_{ij} \leftarrow p_{ij} \frac{(\mathbf{X}^T \mathbf{UV}^{*T} + \alpha \mathbf{W}_{pv^*} \mathbf{PV}^* \mathbf{V}^{*T})_{ij}}{(\mathbf{PV}^* \mathbf{U}^T \mathbf{UV}^{*T} + \alpha \mathbf{D}_{pv^*} \mathbf{PV}^* \mathbf{V}^{*T} + \beta \mathbf{P} + \gamma \mathbf{KPV}^* \mathbf{V}^{*T})_{ij}} \quad (9)$$

#### 2) UPDATING V\*

By taking derivatives of (7) on  $\mathbf{V}^*$ , we have

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{V}^*} &= \mathbf{P}^T \mathbf{PV}^* \mathbf{U}^T \mathbf{U} + \alpha \mathbf{P}^T \mathbf{L}_{pv^*} \mathbf{PV}^* + \gamma \mathbf{P}^T \mathbf{KPV}^* \\ &- \mathbf{P}^T \mathbf{X}^T \mathbf{U} - \Omega \end{aligned}$$

Combining  $L_{pv^*} = D_{pv^*} - W_{pv^*}$ , the updating scheme is get as following

$$v_{ij}^* \leftarrow v_{ij}^* \frac{\sum_v (\mathbf{P}^T \mathbf{X}^T \mathbf{U} + \alpha \mathbf{P}^T \mathbf{W}_{pv^*} \mathbf{PV}^*)_{ij}}{\sum_v (\mathbf{P}^T \mathbf{PV}^* \mathbf{U}^T \mathbf{U} + \alpha \mathbf{P}^T \mathbf{D}_{pv^*} \mathbf{PV}^* + \gamma \mathbf{P}^T \mathbf{KPV}^*)_{ij}} \quad (10)$$

The final clustering process is done by applying K-means algorithm on  $\mathbf{PV}^*$ .

### C. CONVERGENCE ANALYSIS

In this subsection, the convergence of the Algorithm 1 is discussed. A similar proof has been given in the related literature [32] for the updating scheme (8)(9). In the following, we provide related proof for (10).

*Definition 1:*  $Z(x, x')$  is an auxiliary function of  $F(x)$  when the following conditions are satisfied.

$$Z(x, x') \geq F(x) \quad (11)$$

*Lemma 1:*  $F(x)$  is non-increasing under the updating scheme when  $Z(x, x')$  is an auxiliary function of  $F(x)$ .

$$x^{(t+1)} = \arg \min_x Z(x, x^{(t)}) \quad (12)$$

Proof. From **Lemma 1** and **Definition 3**, we can get the following inequalities

$$F(x^{(t+1)}) \leq Z(x^{(t+1)}, x^{(t)}) \leq Z(x^{(t)}, x^{(t)}) = F(x^{(t)}) \quad (13)$$

the equality  $F(x^{(t+1)}) = F(x^{(t)})$  holds only when  $x^{(t)}$  is local minimum of  $Z(x, x^{(t)})$ .

**Algorithm 1** The Algorithm for DMVCSN**Require:**Multi-view dataset  $\{\mathbf{X}^{(v)}\}_{v=1}^{n_v}$ , parameters  $\alpha_v, \beta_v, \gamma_v, m, l$ **Ensure:** $[P^{(1)}V^*, P^{(2)}V^*, \dots, P^{(n_v)}V^*]$ **Initialize:** Random initialize  $U^{(v)}, P^{(v)}, V^*$ 

```

1: while not converge do
2:   for  $v$  to  $n_v$  do
3:      $k = 0$ 
4:     for  $w$  to  $n_v$  do
5:       Update  $k_v$  by (5)
6:       if  $v \neq w$  then
7:          $k = k + k_v$ 
8:       end if
9:     end for
10:    Update  $\mathbf{U}, \mathbf{P}$  by Eq(8), Eq(9)
11:  end for
12:  Update  $\mathbf{V}^*$  by Eq(10)
13: end while

```

Through continuous iterative updating (13), we can finally converge to obtain the estimation sequence as following inequalities:

$$F(x_{min}) \leq \dots \leq F(x^{(t)}) \leq \dots \leq F(x^{(0)}) \quad (14)$$

In order to prove the consistency between the updating rule (10) and (12), let  $F_{v_{ij}^*}$  are related to any element  $v_{ij}^*$  in  $V^*$ , we can get the following equations:

$$\begin{aligned}
F_{v_{ij}^*}(v_{ij}^*) &= \left\| X - U(PV^*)^T \right\|_F^2 \\
&\quad + \alpha \text{Tr}((PV^*)^T L_{pv^*}(PV^*)) \\
&\quad + \gamma \text{Tr}((PV^*)^T K(PV^*)) \\
F'_{v_{ij}^*}(v_{ij}^*) &= \left[ \frac{\partial F}{\partial v^*} \right]_{ij} = [-2P^T X^T U + 2P^T PV^* U^T U]_{ij} \\
&\quad + [2\alpha P^T L_{pv^*} PV^* + 2\gamma P^T K PV^*]_{ij} \\
F''_{v_{ij}^*}(v_{ij}^*) &= 2[P^T P U^T U + \alpha P^T L_{pv^*} P + \gamma P^T K P]_{ij}
\end{aligned}$$

An auxiliary function of  $v_{ij}^*$  is defined as following:

$$\begin{aligned}
Z(v^*, v_{ij}^{*(t)}) &= F_{v_{ij}^*}(v_{ij}^{*(t)}) + F'_{v_{ij}^*}(v_{ij}^{*(t)})(v^* - v_{ij}^{*(t)}) \\
&\quad + \frac{[P^T PV^* U^T U]_{ij}}{v_{ij}^{*(t)}} (v^* - v_{ij}^{*(t)})^2 \\
&\quad + \frac{[\alpha P^T L_{pv^*} PV^*]_{ij}}{v_{ij}^{*(t)}} (v^* - v_{ij}^{*(t)})^2 \\
&\quad + \frac{[\gamma P^T K PV^*]_{ij}}{v_{ij}^{*(t)}} (v^* - v_{ij}^{*(t)})^2 \quad (15)
\end{aligned}$$

Proof. The expansion of  $F_{v_{ij}^*}(v_{ij}^*)$  in Taylor series is defined:

$$\begin{aligned}
F_{v_{ij}^*}(v_{ij}^*) &= F_{v_{ij}^*}(v_{ij}^{*(t)}) + F'_{v_{ij}^*}(v_{ij}^{*(t)})(v_{ij}^* - v_{ij}^{*(t)}) \\
&\quad + (P^T P U^T U)(v_{ij}^* - v_{ij}^{*(t)})^2 \\
&\quad + (\alpha P^T L_{pv^*} P)(v_{ij}^* - v_{ij}^{*(t)})^2 \\
&\quad + (\gamma P^T K P)(v_{ij}^* - v_{ij}^{*(t)})^2 \quad (16)
\end{aligned}$$

From **Definition 3**, we can easily see that  $Z(v^*, v^*) = F_{v_{ij}^*}(v^*)$ . Comparing the formulas (15) and (16), to prove  $Z(v^*, v_{ij}^{*(t)}) \geq F_{v_{ij}^*}(v_{ij}^*)$ , we only need to prove

$$\begin{aligned}
\frac{[P^T PV^* U^T U]_{ij}}{v_{ij}^{*(t)}} &\geq [P^T P U^T U]_{ij} \\
\frac{[\alpha P^T L_{pv^*} PV^*]_{ij}}{v_{ij}^{*(t)}} &\geq [\alpha P^T L_{pv^*} P]_{ij} \\
\frac{[\gamma P^T K PV^*]_{ij}}{v_{ij}^{*(t)}} &\geq [\gamma P^T K P]_{ij}
\end{aligned}$$

Based on the above inequality, we can derive the following new inequality:

$$\begin{aligned}
[P^T PV^* U^T U]_{ij} &= \sum_l (P^T PV^*) (U^T U) \\
&\geq (P^T PV^*) (U^T U) \\
&\geq \sum_l (P^T P) v^{*(t)} (U^T U) \\
&\geq v^{*(t)} (P^T P) (U^T U) \\
\alpha [P^T L_{pv^*} PV^*]_{ij} &= \alpha \sum_l P^T L_{pv^*} P v^{(t)} \\
&\geq \alpha P^T L_{pv^*} P v^{(t)} \\
\gamma [P^T K PV^*]_{ij} &= \gamma \sum_l P^T K P v^{(t)} \\
&\geq \gamma P^T K P v^{(t)}
\end{aligned}$$

Therefore  $Z(v^*, v_{ij}^{*(t)}) \geq F_{v_{ij}^*}(v_{ij}^*)$ . so (15) is an auxiliary function of the variable  $v_{ij}^*$ .

**IV. EXPERIMENT AND RESULT ANALYSIS**

We employ seven multi-view clustering datasets widely used in recent work [33], [34], [35], [36] to evaluate our proposed method fully. They cover various domains and types, such as text, news, and facial images.

The table 1 summarizes general information and statistics about these seven datasets.

**TABLE 1.** Summary of the seven datasets used in our experiments.

Dataset	3Sources	BBCSport	Caltech101_20	Handwritten	MSRCv1	Yale	Extended-YaleB
Samples	169	544	2386	2000	210	165	640
Views	3	2	6	6	5	3	3
Clusters	6	5	20	10	7	15	10

**A. EVALUATION INDICATORS**

To quantitatively assess the effect, four evaluation metrics were used to measure the performance in the experiments:

accuracy (ACC), normalized mutual information (NMI), adjusted Rand index (ARI), and purity (Purity), respectively.

## B. COMPARED METHODS

To quantitatively assess the effect, we used four evaluation metrics to measure the performance in the experiments: accuracy (ACC), normalized mutual information (NMI), adjusted Rand index (ARI), and purity (Purity), respectively.

**GNMF [17]:** The graph regularized non-negative matrix factorization (GNMF) is a single-view clustering algorithm with graph constraints based on NMF. In the GNMF method, the manifold structure of similar graphs is considered.

**NMFCC [37]:** The Non-negative Matrix Factorization with Orthogonal Constraints (NMF-CC) algorithm is an orthogonally constrained multi-view clustering algorithm based on NMF.

**NSGL [34]:** The non-negative structural graph learning (NSGL) algorithm directly learns structural graphs from raw features by imposing rank constraints, while exploiting the complementarity of multi-view features to perform adaptive functions to select visual semantic information and guide the graph learning process, and finally, the information is selected by forcing the feature selection matrix to be sparse in the rows that line the sparse regression.

**LMSC [23]:** The Multi-View Latent Representation Clustering (LMSC) algorithm exploits the complementarity between views to explore latent representations of data points, resulting in more accurate subspace representations.

**GPMVC [38]:** The Graph Regularized Partially Multi-View Clustering (GPMVC) algorithm exploits the intrinsic geometry of the data distribution in each view as an extension of the non-negative matrix factorization-based PVC method while supporting graph-specific Laplacian regularization.

**MSC\_IAS [36]:** The Integrity-Aware Similarity Based Multi-View Subspace Clustering (MSC\_IAS) algorithm constructs a data similarity matrix by adaptively assigning neighborhoods to each contact spatial data point based on local connectivity.

**SMVSC [39]:** The Scalable Multi-View Subspace Clustering with Unified Anchors (SMVSC) method is a scalable matrix three factorization multi-view subspace clustering algorithm model based on unified anchors.

**ECMSC [40]:** Based on Complementary Representation and Consistent Multi-View Subspace Clustering Model (ECMSC), it exploits the complementary information between different representations.

**DIMSC [25]:** Diversity Multi-View Subspace Clustering (DIMSC) exploring the complementarity of multi-view representations using the Hilbert Schmidt Independence Criteria.

**JSMC [41]:** The subspace representation on each view is decomposed into two matrices, the view commonality matrix and the view inconsistency matrix. The multi-view local structure is also used to promote the learning of the common representation. In addition, the robustness of clustering is improved by introducing lower-order representations via kernel parametric.

## C. PARAMETER SETTING

The objective function of DMVCSN consists of many parameters, including parameters for manifold learning  $\alpha_v$ , regularization parameters  $\beta_v$ , parameters  $\gamma_v$  that hold the importance of the HSIC module, and the parameters  $m, l$  that hold the size of the latent matrix. We also discussed the influence of these parameters in this section.

First of all, the parameter  $\alpha_v$  is the one used to measure the importance of manifold learning. This parameter plays a crucial role in the manifold structure learning of the original data in the experiment. It can be seen from in Fig. 2 that the fluctuation range of the three evaluation metrics ACC, NMI, and ARI on the text and image datasets 3Sources, BBCSport, and Extended Yale-B is very small. There are similar effects in other datasets, which means the proposed model is more stable for this parameter. In experiments, we usually set this parameter to 1 or 10.

The second regularization parameter  $\beta_v$  measures the influence of each view on the common manifold structure. As shown in Fig. 3, this parameter has different effects on different datasets, and it is difficult to obtain this parameter value empirically. In experiments, we usually set this parameter to 1 or 10.

The third important parameter is  $\gamma_v$ , which ensures the diversity of representations in different subspaces: the larger the parameter's value, the more critical the common features among views. As shown in Fig. 4, all datasets are set to 1.

The two most important parameters are  $m, l$ . These two parameters control the size of the consensus matrix. As shown in Fig. 5, the significant difference in parameter settings between the text dataset and the image dataset may be because the non-negative factorization method can obtain more features in the image dataset. Specifically, if the consensus matrix dimension is too small, each views cannot get enough feature information about the manifold structure from the latent consensus representation to adjust their own manifold structure; when the matrix dimension is too large, redundant features will reduce the clustering effect. Following work [32], we usually set  $m = l$ . For different dataset types, its value is generally different. Text datasets can be obtained between [10, 50]; for image datasets, we can accept them between [100, 200].

## D. CLUSTERING RESULTS

The clustering performance and results on seven datasets are presented in tables 2 to 8. The proposed method outperforms the other algorithms in most of the datasets on four evaluation metrics. Detailed discussion is as follows:

- 1) Our method produced significantly improved results on several datasets, including 3Sources, BBCSport, Yale, and Extended-YaleB. Take the evaluation metrics ACC and NMI metrics as examples. Specifically, the ACC metrics of our method are higher than the second-best one by 5.33%, 5.34%, 4.87%, and 9.41% for the data sets 3Sources, BBCSport, Yale, and Extended-YaleB,

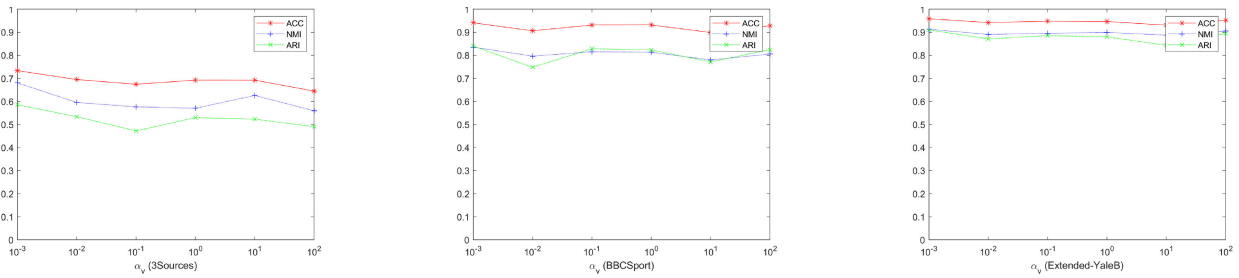


FIGURE 2. ACC, NMI and ARI are presented respectively according to the parameter  $\alpha_V$ .

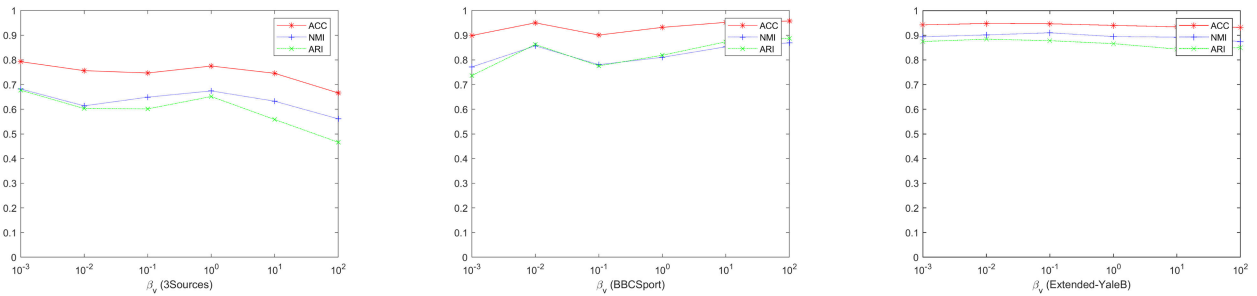


FIGURE 3. ACC, NMI and ARI are presented respectively according to the parameter  $\beta_V$ .

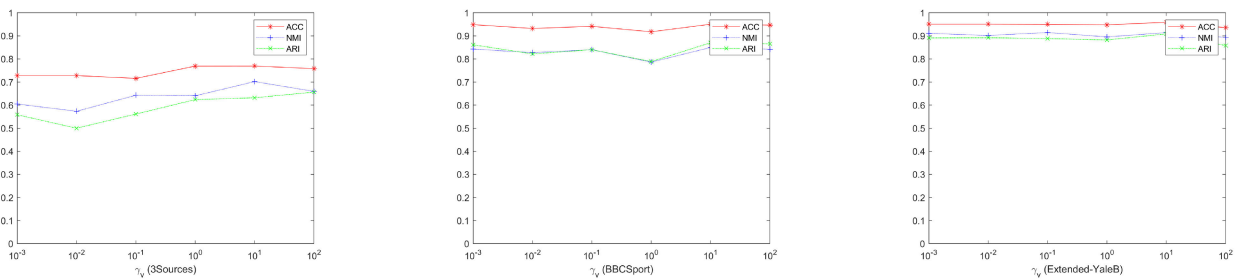


FIGURE 4. ACC, NMI and ARI are presented respectively according to the parameter  $\gamma_V$ .

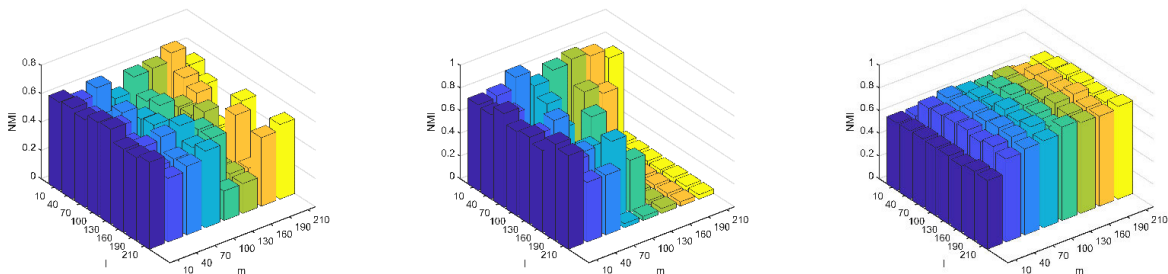


FIGURE 5. Visualization for the hyperparameter  $m$  and  $l$  selection procedure on datasets 3Sources, BBCSport and Yale.

respectively, and 0.82%, 7.47%, 8.1% and 2.75% for the NMI metrics, respectively. tables 2, 3, 7 and 8 shows a significant improvement in our method.

- 2) For the image datasets Caltech101\_20, Handwritten, MSRCv1, Yale, and Extended-YaleB, tables 4, and 6, 8 show the clustering effect on the image class datasets. According to tables 7, and 8, we can see that our

algorithm has excellent experimental results on the Yale and Extended-YaleB datasets, especially on the dataset Extended-YaleB, all evaluation metrics are improved significantly, and the lowest improvement of evaluation metrics is about 2.74%. Compared with the graph non-negative matrix factorization method (GNMF), the improvement in each evaluation metric



**TABLE 2. Results on the 3Sources dataset (best results in bold).**

Method	ACC	NMI	ARI	Purity
GNMF [17]	0.5446	0.5144	0.4156	0.5920
NMFCC [38]	0.6182	0.6382	0.4925	0.6992
NSGL [34]	0.4358	0.2798	0.0780	0.6965
LMSC [23]	0.6175	0.4957	0.4568	0.7008
GPMVC [39]	0.5373	0.4687	0.2919	0.6441
MSC_IAS [36]	0.5893	0.4765	0.3364	0.6498
SMVSC [39]	0.5678	0.5569	0.3713	0.6390
ECMSC [40]	0.3181	0.1032	0.0108	0.6138
DIMSC [25]	0.4084	0.3815	0.2410	0.4983
JSMC [41]	0.5644	0.6346	0.4582	0.6780
Ours	<b>0.7278</b>	<b>0.6410</b>	<b>0.5297</b>	<b>0.7570</b>

is 16.48%, 5.6%, 16.17%, and 3.81%, respectively. According to the table 4 on the dataset Caltech101\_20, the progress of the practical effect may not be significant and suboptimal on indicators ACC and ARI. tables 5, and 6 show that the experimental results are much worse than the best algorithm on the datasets Handwritten and MSRCv1. This result may be due to the significant differences between the data features of different views in these datasets. The non-negative matrix factorization is ineffective in separating the view features and data structure manifolds, resulting in significant differences between data manifold structures between the same class samples of different views. We can improve the experimental results by using a more accurate feature extraction method, such as the sparse matrix factorization method used in the works [26], [42].

- 3) For the text datasets 3Sources and BBCSport, tables 2, and 3 show the clustering results. Compared with other algorithms, our proposed algorithm significantly improves all metrics. In particular, on the dataset BBCSport, our algorithm improves about 5.33%, 5.34%, 11.02%, and 7.77% on the four evaluation metrics of ACC, NMI, ARI, and Purity, respectively, compared to the suboptimal algorithm JSMC. This is due to the strong correlation between the contexts of text datasets, the non-negative matrix factorization can better separate the features from the data manifolds, and the views adjust their manifold structure using the accurate data manifolds between them. Besides, compared to algorithm SMVSC, although our method uses the technique of matrix triple factorization, the HSIC module promotes the interconnection between different view data and enhances the mutual learning of information. In contrast to the algorithm DIMSC, although both adopt the HSIC module, our algorithm improves the accuracy of clustering by combining the idea of non-negative matrix factorization and manifold learning to explore the manifold structure between view data points deeply.

To better illustrate the characteristics of data manifold feature transfer in the algorithm, we show the adjacency matrix of the  $PV^*$  of sample points and the adjacency matrix of

**TABLE 3. Results on the BBCSport dataset (best results in bold).**

Method	ACC	NMI	ARI	Purity
GNMF [17]	0.6856	0.6628	0.5207	0.8600
NMFCC [38]	0.8058	0.7340	0.7167	0.8839
NSGL [34]	0.4240	0.1656	0.0490	0.8800
LMSC [23]	0.7476	0.4957	0.4568	0.7740
GPMVC [39]	0.6604	0.5495	0.3895	0.6886
MSC_IAS [36]	0.7207	0.6017	0.5548	0.8291
SMVSC [39]	0.8425	0.7936	0.7647	0.8809
ECMSC [40]	0.3629	0.0552	0.0065	0.9093
DIMSC [25]	0.8201	0.6705	0.6989	0.8401
JSMC [41]	0.9026	0.8149	0.7682	0.8780
Ours	<b>0.9559</b>	<b>0.8683</b>	<b>0.8784</b>	<b>0.9557</b>

**TABLE 4. Results on the Caltech101\_20 dataset (best results in bold).**

Method	ACC	NMI	ARI	Purity
GNMF [17]	0.4305	0.5753	0.3170	0.5029
NMFCC [38]	0.4650	0.6093	0.3485	0.4929
NSGL [34]	0.5174	0.3836	0.3775	0.6960
LMSC [23]	0.4775	0.6112	0.3555	0.5587
GPMVC [39]	0.4683	<b>0.6440</b>	0.3637	<b>0.7824</b>
MSC_IAS [36]	0.5265	0.5381	0.3814	0.6126
SMVSC [39]	<b>0.5855</b>	0.5664	<b>0.5182</b>	0.7033
ECMSC [40]	0.5063	0.6198	0.3940	0.6016
DIMSC [25]	0.2856	0.3047	0.1368	0.3127
JSMC [41]	0.4867	0.6354	0.4135	0.5826
Ours	0.5360	0.6087	0.4135	0.5826

**TABLE 5. Results on the Handwritten dataset (best results in bold).**

Method	ACC	NMI	ARI	Purity
GNMF [17]	<b>0.9401</b>	0.9041	<b>0.8872</b>	0.9532
NMFCC [38]	0.5755	0.5779	0.4288	0.6735
NSGL [34]	0.4721	0.6141	0.4414	0.8793
LMSC [23]	0.8341	0.8052	0.7443	0.9055
GPMVC [39]	0.7941	0.7796	0.7068	0.8208
MSC_IAS [36]	0.8820	<b>0.9190</b>	0.8591	<b>0.9780</b>
SMVSC [39]	0.7094	0.7416	0.6230	0.8908
ECMSC [40]	0.8138	0.8281	0.7520	0.8782
DIMSC [25]	0.6831	0.6965	0.4865	0.7291
JSMC [41]	0.8404	0.8419	0.7809	0.8846
Ours	0.8363	0.8316	0.7620	0.9283

**TABLE 6. Results on the MSRCv1 dataset (best results in bold).**

Method	ACC	NMI	ARI	Purity
GNMF [17]	0.5532	0.6225	0.4007	0.7934
NMFCC [38]	0.5584	0.5154	0.3553	0.6602
NSGL [34]	0.5429	0.4906	0.3615	0.7135
LMSC [23]	0.6973	0.6493	0.5310	0.7192
GPMVC [39]	<b>0.8612</b>	<b>0.7697</b>	<b>0.8612</b>	<b>0.8612</b>
MSC_IAS [36]	0.8104	0.7577	0.6708	0.8333
SMVSC [39]	0.7942	0.7080	0.6424	0.8142
ECMSC [40]	0.8201	0.7534	0.6912	0.8201
DIMSC [25]	0.6587	0.5667	0.4744	0.6949
JSMC [41]	0.8425	0.7367	0.6785	0.8425
Ours	0.7571	0.6224	0.5313	0.7571

the projection matrix  $\mathbf{P}$  for the datasets BBCSport, Yale, and Extended-YaleB. As shown in Fig. 6, we can see that the adjacency matrix of  $PV^*$  of sample points and the adjacency matrix of the projection matrix  $\mathbf{P}$  are very similar for both

TABLE 7. Results on the Yale dataset (best results in bold).

Method	ACC	NMI	ARI	Purity
GNMF [17]	0.5445	0.6271	0.3645	0.6855
NMFCC [38]	0.5146	0.6011	0.3285	0.6460
NSGL [34]	0.3898	0.4531	0.1778	0.4759
LMSC [23]	0.6770	0.7073	0.4830	0.7461
GPMVC [39]	0.6097	0.6719	0.4397	0.6175
MSC_IAS [36]	0.7397	0.7512	0.5754	0.7830
SMVSC [39]	0.4900	0.5499	0.3009	0.6132
ECMSC [40]	0.6875	0.6956	0.4350	0.7299
DIMSC [25]	0.8402	0.6706	0.6991	0.8402
JSMC [41]	0.7624	0.7003	0.6570	0.7468
Ours	<b>0.8889</b>	<b>0.8322</b>	<b>0.7391</b>	<b>0.8889</b>

TABLE 8. Results on the Extended-YaleB dataset (best results in bold).

Method	ACC	NMI	ARI	Purity
GNMF [17]	0.7906	0.8521	0.7407	0.9163
NMFCC [38]	0.3018	0.2751	0.1398	0.3013
NSGL [34]	0.1774	0.0699	0.0236	0.3478
LMSC [23]	0.5411	0.5439	0.2944	0.5929
GPMVC [39]	0.2883	0.2714	0.1227	0.2997
MSC_IAS [36]	0.7881	0.7778	0.6642	0.8112
SMVSC [39]	0.2879	0.2322	0.0672	0.3323
ECMSC [40]	0.7880	0.7580	0.5057	0.7907
DIMSC [25]	0.4095	0.3819	0.2408	0.4975
JSMC [41]	0.8613	0.8806	0.8750	0.8413
Ours	<b>0.9554</b>	<b>0.9081</b>	<b>0.9024</b>	<b>0.9544</b>

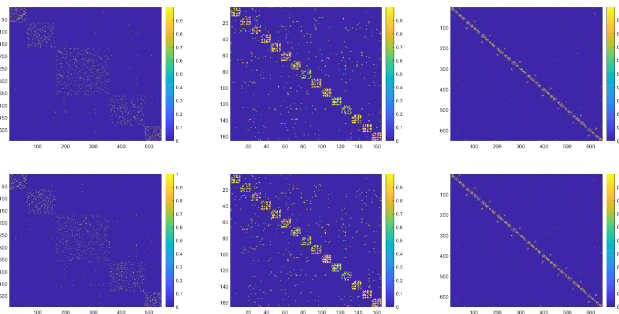


FIGURE 6. From left to right is adjacency matrix of datasets BBCSport, Yale and Extended-YaleB; From top to button is adjacency matrix of PV\* and P.

text-based and picture-based datasets, which is because the combination of similar latent features often forms the manifold features of the same class samples even in different views. That is also consistent with our idea of data manifold projection.

### E. SELECTION OF SIMILARITY MATRIX

There are three ways to define similarity matrices in our experiments: dot product weights, binary weights, and heat kernel weights. After several experiments, the best performance is often obtained by choosing the heat kernel weights. Besides, as shown in Fig. 7, different numbers of nearest neighbors tend to produce large differences in the clustering performance. Based on the results shown in Fig. 7, we usually choose  $k = 5$ . The clustering performance is always best when the nearest neighbors are between 3 and 12.

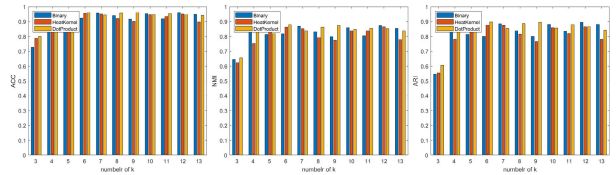


FIGURE 7. The performance of DMVCSN in the BBCSports datasets. ACC, NMI and ARI were used for comparison.  $k$  was chosen as 5.

### V. SUMMARY

To exploit the complementarity information among multiple views, Besides non-negative matrix factorization to get local geometric structure, we utilize manifold learning and latent representation to facilitate learning common geometric structures in different views. Moreover, the Hilbert-Schmidt independence criterion is introduced to learn diversity for mutual learning and information fusion among views. In addition, we conduct experiments on seven multi-view datasets, including two text datasets and five image datasets, comparing nine multi-view clustering algorithms and one single-view clustering algorithm. Compared to these state-of-the-art methods, our method significantly improves various datasets, which shows that our algorithm is quite competitive. In future work, we will extend the DMVCSN method to explore the sparse representation of different views.

### REFERENCES

- [1] Y.-M. Kim, M.-R. Amini, C. Goutte, and P. Gallinari, "Multi-view clustering of multilingual documents," in *Proc. 33rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Geneva, Switzerland, Jul. 2010, p. 821.
- [2] Y. Shi, G. Li, Q. Cao, K. Wang, and L. Lin, "Face hallucination by attentive sequence optimization with reinforcement learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 11, pp. 2809–2824, Nov. 2020.
- [3] X. Cai, F. Nie, and H. Huang, "Multi-view  $K$ -means clustering on big data," in *Proc. 23rd Int. Joint Conf. Artif. Intell.* Beijing, China: AAAI Press, 2013, pp. 2598–2604.
- [4] M. Zhang, Y. Yang, F. Shen, H. Zhang, and Y. Wang, "Multi-view feature selection and classification for Alzheimer's disease diagnosis," *Multimedia Tools Appl.*, vol. 76, no. 8, pp. 10761–10775, 2017.
- [5] Q. Gao, W. Xia, Z. Wan, D. Xie, and P. Zhang, "Tensor-SVD based graph learning for multi-view subspace clustering," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 4, pp. 3930–3937.
- [6] Y. Wang and L. Wu, "Beyond low-rank representations: Orthogonal clustering basis reconstruction with optimized graph structure for multi-view spectral clustering," *Neural Netw.*, vol. 103, pp. 1–8, Jul. 2018.
- [7] Q. Zheng, J. Zhu, Z. Li, S. Pang, J. Wang, and Y. Li, "Feature concatenation multi-view subspace clustering," *Neurocomputing*, vol. 379, pp. 89–102, Feb. 2020.
- [8] F. Yan, X.-D. Wang, Z.-Q. Zeng, and C.-Q. Hong, "Adaptive multi-view subspace clustering for high-dimensional data," *Pattern Recognit. Lett.*, vol. 130, pp. 299–305, Feb. 2020.
- [9] C. Zhou, C. Zhang, H. Fu, R. Wang, and X. Cao, "Multi-cue augmented face clustering," in *Proc. 23rd ACM Int. Conf. Multimedia*, Oct. 2015, pp. 1095–1098.
- [10] J. Li, H. Yong, F. Wu, and M. Li, "Online multi-view subspace learning with mixed noise," in *Proc. 28th ACM Int. Conf. Multimedia*, Seattle, WA, USA, Oct. 2020, pp. 3838–3846.
- [11] J. Li, J. Zhao, F. Zhao, H. Liu, J. Li, S. Shen, J. Feng, and T. Sim, "Robust face recognition with deep multi-view representation learning," in *Proc. 24th ACM Int. Conf. Multimedia*, Amsterdam The Netherlands, Oct. 2016, pp. 1068–1072.
- [12] B. Wu, E. Zhong, A. Horner, and Q. Yang, "Music emotion recognition by multi-label multi-layer multi-instance multi-view learning," in *Proc. 22nd ACM Int. Conf. Multimedia*, Orlando, FL, USA, Nov. 2014, pp. 117–126.
- [13] U. von Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.

[14] J. Feng, Z. Lin, H. Xu, and S. Yan, "Robust subspace segmentation with block-diagonal prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, Jun. 2014, pp. 3818–3825.

[15] M. Yin, Y. Guo, J. Gao, Z. He, and S. Xie, "Kernel sparse subspace clustering on symmetric positive definite manifolds," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 5157–5164.

[16] 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.

[17] 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.

[18] X. Zhang, L. Zhao, L. Zong, X. Liu, and H. Yu, "Multi-view clustering via multi-manifold regularized nonnegative matrix factorization," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2014, pp. 1103–1108.

[19] B. Qian, X. Shen, Y. Gu, Z. Tang, and Y. Ding, "Double constrained NMF for partial multi-view clustering," in *Proc. Int. Conf. Digit. Image Comput., Techn. Appl.*, Nov. 2016, pp. 1–7.

[20] J. Guo, Y. Sun, J. Gao, Y. Hu, and B. Yin, "Low rank representation on product Grassmann manifolds for multi-view subspace clustering," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 907–914.

[21] P. Zhang, X. Liu, J. Xiong, S. Zhou, W. Zhao, E. Zhu, and Z. Cai, "Consensus one-step multi-view subspace clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 10, pp. 4676–4689, Oct. 2022.

[22] Q. Wang, X. Jiang, M. Chen, and X. Li, "Autoweighted multiview feature selection with graph optimization," *IEEE Trans. Cybern.*, vol. 52, no. 12, pp. 1–12, Dec. 2022.

[23] 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)*, Honolulu, HI, USA, Jul. 2017, pp. 4333–4341.

[24] C. Zhang, H. Fu, Q. Hu, X. Cao, Y. Xie, D. Tao, and D. Xu, "Generalized latent multi-view subspace clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 1, pp. 86–99, Jan. 2020.

[25] X. Cao, C. Zhang, H. Fu, S. Liu, and H. Zhang, "Diversity-induced multi-view subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 586–594.

[26] J. Tan, Z. Yang, J. Ren, B. Wang, Y. Cheng, and W.-K. Ling, "A novel robust low-rank multi-view diversity optimization model with adaptive-weighting based manifold learning," *Pattern Recognit.*, vol. 122, Feb. 2022, Art. no. 108298.

[27] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.

[28] S. El Hajjar, F. Dornaika, F. Abdallah, and N. Barrena, "Consensus graph and spectral representation for one-step multi-view kernel based clustering," *Knowl.-Based Syst.*, vol. 241, Apr. 2022, Art. no. 108250.

[29] A. Gretton, O. Bousquet, A. Smola, and B. Schölkopf, "Measuring statistical dependence with Hilbert–Schmidt norms," in *Algorithmic Learning Theory (Lecture Notes in Computer Science)*, vol. 3734, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, S. Jain, H. U. Simon, and E. Tomita, Eds. Berlin, Germany: Springer, 2005, pp. 63–77.

[30] F. Shang, L. C. Jiao, and F. Wang, "Graph dual regularization non-negative matrix factorization for co-clustering," *Pattern Recognit.*, vol. 45, no. 6, pp. 2237–2250, 2012.

[31] Z. Lu, G. Liu, and S. Wang, "Sparse neighbor constrained co-clustering via category consistency learning," *Knowl.-Based Syst.*, vols. 201–202, Aug. 2020, Art. no. 105987.

[32] J. Sun, Z. Wang, F. Sun, and H. Li, "Sparse dual graph-regularized NMF for image co-clustering," *Neurocomputing*, vol. 316, pp. 156–165, Nov. 2018.

[33] L. Fei-Fei, R. Fergus, and P. Perona, "One-shot learning of object categories," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 594–611, Apr. 2006.

[34] X. Bai, L. Zhu, C. Liang, J. Li, X. Nie, and X. Chang, "Multi-view feature selection via nonnegative structured graph learning," *Neurocomputing*, vol. 387, pp. 110–122, Apr. 2020.

[35] M. Brbić and I. Kopriva, "Multi-view low-rank sparse subspace clustering," *Pattern Recognit.*, vol. 73, pp. 247–258, Jan. 2018.

[36] X. Wang, Z. Lei, X. Guo, C. Zhang, H. Shi, and S. Z. Li, "Multi-view subspace clustering with intactness-aware similarity," *Pattern Recognit.*, vol. 88, pp. 50–63, Apr. 2019.

[37] N. Liang, Z. Yang, Z. Li, W. Sun, and S. Xie, "Multi-view clustering by non-negative matrix factorization with co-orthogonal constraints," *Knowl.-Based Syst.*, vol. 194, Apr. 2020, Art. no. 105582.

[38] N. Rai, S. Negi, S. Chaudhury, and O. Deshmukh, "Partial multi-view clustering using graph regularized NMF," in *Proc. 23rd ICPR*, 2016, pp. 2192–2197.

[39] M. Sun, P. Zhang, S. Wang, S. Zhou, W. Tu, X. Liu, E. Zhu, and C. Wang, "Scalable multi-view subspace clustering with unified anchors," in *Proc. 29th ACM Int. Conf. Multimedia*, 2021, pp. 3528–3536.

[40] X. Wang, X. Guo, Z. Lei, C. Zhang, and S. Z. Li, "Exclusivity-consistency regularized multi-view subspace clustering," in *Proc. CVPR*, Honolulu, HI, USA, 2017, pp. 1–9.

[41] X. Cai, D. Huang, G.-Y. Zhang, and C.-D. Wang, "Seeking commonness and inconsistencies: A jointly smoothed approach to multi-view subspace clustering," *Inf. Fusion*, vol. 91, pp. 364–375, Mar. 2023.

[42] X. Yu, H. Liu, Y. Wu, and C. Zhang, "Fine-grained similarity fusion for multi-view spectral clustering," *Inf. Sci.*, vol. 568, pp. 350–368, Aug. 2021.



**JIAMAN DING** was born in 1974. He received the B.S. degree from the Faculty of Information Engineering and Automation, Kunming University of Science and Technology, in 1998, and the M.S. degree, in 2005. He is currently a Professor. His current research interests include data mining, cloud computing, and machine learning.



**XIAOJIANG FANG** is currently pursuing the M.S. degree. His current research interests include data mining and machine learning.



**LIANYIN JIA** received the Ph.D. degree from the South China University of Technology, in 2012. His current research interests include databases, data mining, information retrieval, and parallel computing. He is currently a member of the Youth Working Committee of the Artificial Intelligence Society and a member of the Computer Society. He has chaired and participated in several National Natural Science Foundation of China.



**YING JIANG** is currently a Supervisor of doctoral students with the Kunming University of Science and Technology, the Deputy Director of the Steering Committee of Software Engineering in Yunnan Higher Education, a member of the Steering Committee of Computer Science in Yunnan Higher Education, a member of the Software Engineering Committee of the Chinese Computer Society, and a member of the Education Committee of the Chinese Computer Society. She has presided over four National Natural Science Funds. In recent years, she has published more than 60 papers in critical academic journals and conferences at home and abroad, among which SCI, EI, and ISTP have indexed more than 30. Her current research interests include cloud computing, big data analysis, and intelligent software engineering.



**RUNXIN LI** is a Senior Member of CCF, a Member of the CCF Special Committee on Service Computing, and used to be the Vice President of the School of Aeronautics of the Kunming University of Technology. He has chaired the National Natural Science Foundation of China project. His main research areas are service computing, intelligent decision-making, big data, and software engineering.

...