Signal Processing 217 (2024) 109341

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Centric graph regularized log-norm sparse non-negative matrix factorization for multi-view clustering

Yuzhu Dong [a,b], Hangjun Che [b,c,d,*], Man-Fai Leung [e], Cheng Liu [f], Zheng Yan [g]

[a] College of Westa, Southwest University, Chongqing, China
[b] College of Electronic and Information Engineering, Southwest University, Chongqing, China
[c] Chongqing Key Laboratory of Nonlinear Circuits and Intelligent Information Processing, Chongqing, China
[d] Key Laboratory of Cyber–Physical Fusion Intelligent Computing (South-Central Minzu University), State Ethnic Affairs Commission, China
[e] School of Computing and Information Science, Faculty of Science and Engineering, Anglia Ruskin University, Cambridge, United Kingdom
[f] Department of Computer Science, Shantou University, Shantou, China
[g] University of Technology Sydney, Sydney, Australia

ARTICLE INFO

Keywords:
Multi-view learning
Non-negative matrix factorization
Pairwise co-regularization
Centric graph regularization

ABSTRACT

Multi-view non-negative matrix factorization (NMF) provides a reliable method to analyze multiple views of data for low-dimensional representation. A variety of multi-view learning methods have been developed in recent years, demonstrating successful applications in clustering. However, existing methods in multi-view learning often tend to overlook the non-linear relationships among data and the significance of the similarity of internal views, both of which are essential in multi-view tasks. Meanwhile, the mapping between the obtained representation and the original data typically contains complex hidden information that deserves to be thoroughly explored. In this paper, a novel multi-view NMF is proposed that explores the local geometric structure among multi-dimensional data and learns the hidden representation of different attributes through centric graph regularization and pairwise co-regularization of the coefficient matrix. In addition, the proposed model is further sparsified with $l_{2,log}$-(pseudo) norm to efficiently generate sparse solutions. As a result, the model obtains a better part-based representation, enhancing its robustness and applicability in complex noisy scenarios. An effective iterative update algorithm is designed to solve the proposed model, and the convergence of the algorithm is proven to be theoretically guaranteed. The effectiveness of the proposed method is verified by comparing it with nine state-of-the-art methods in clustering tasks of eight public datasets.

1. Introduction

Massive data has been gathered from various origins in recent years, encompassing heterogeneous properties from diverse feature perspectives, which is referred to as multi-view data [1]. Typically, compatible and complementary information can be found in such multiple representations [2]. As a result, the multi-view learning paradigm has been developed and investigated. The clustering approach, which effectively harnesses multiple views to jointly contribute to the results, enables accurate analysis of heterogeneous features and efficient utilization of abundant features from different perspectives [3]. However, multi-view data is often characterized by high dimensionality, making it unsuitable for direct use in pattern recognition tasks [4,5]. Consequently, obtaining a more suitable representation using multi-view data for downstream tasks like clustering remains a challenging endeavor [6].

By formulating a well-designed learning mechanism, the execution of multi-view clustering can effectively unveil the latent structures embedded within the multi-view data, thereby enhancing the clustering performance [7–9]. Many types of multi-view learning models have emerged over recent years [10]. The crucial aspect of employing multiple views to address the clustering problem involves rational fusion, aiming to generate outcomes that are both accurate and robust [11]. Current approaches of multi-view learning are generally categorized into multi-kernel learning [12–14], graph clustering [15,16] and subspace clustering [17–19]. Multi-kernel learning integrates multiple basis kernels through linear or non-linear combinations to derive the optimal clustering kernel. The graph-based clustering approach conceptualizes the multi-view task as a process of graph partitioning, where the process for multi-view learning corresponds to the fusion of multiple graphs. Subspace clustering endeavors to discover appropriate low-dimensional representations and structures for each view through matrix factorization. The representations are subsequently fused into a unified representation that encompasses complementary information,

* Corresponding author at: College of Electronic and Information Engineering, Southwest University, Chongqing, China.
E-mail address: hjche123@swu.edu.cn (H. Che).

https://doi.org/10.1016/j.sigpro.2023.109341
Received 11 August 2023; Received in revised form 19 November 2023; Accepted 22 November 2023
Available online 23 November 2023

with the goal of unveiling a shared underlying subspace among multiple views. Compared with other approaches, the method of matrix factorization is constructed in a way which has low computational complexity [20]. Among them, NMF is interpretable and is able to discover part-based representation from raw data [21]. It has been successfully applied in many areas, such as face recognition, social network analysis, text clustering, image retrieval and visual tracking [22].

NMF is extensively employed for multi-view clustering owing to its capability to handle high-dimensional data effectively. A multi-view NMF for clustering tasks is presented to achieve a common consensus for each view of the clustering solution while learning the basis and coefficient matrices [23]. Although NMF performs well in handling high-dimensional problems [24], it appears to be incapable of capturing the internal structure among multi-dimensional data [25]. Therefore, many researches have been made to explore the internal relationships within views by applying manifold regularization. A multi-manifold NMF is developed to consider the similarity of the coefficient matrix of each view to the consensus matrix by utilizing centroid-based co-regularization [26]. A diverse NMF is proposed with a novel regularization term, which encourages a sufficient diversity of representations from multiple views to capture comprehensive information [27]. A novel manifold regularization utilizes orthogonality to adequately capture the diversity within the views [28]. A new model is introduced that integrates high-level manifold consensus constraints to obtain the underlying clustering structure for each view [29]. A method based on the consistency of multi-view data is proposed, which employs a multi-manifold regularized NMF algorithm to obtain uniform manifold and global clustering [30]. Additionally, a new NMF method for clustering multi-view data with manifold regularization is proposed, which is able to evaluate the divergence terms between coefficient matrices and the consensus matrix [31].

In multi-view NMF, the objective is to approximate the input data matrix in each view using two non-negative factor matrices, allowing each observation to be explained as a linear combination of non-negative basis vectors. However, this approach ignores the non-linear relationships between multi-dimensional data, which are considerably more important in practical applications. To address this, the graph-based learning approaches are often introduced to deal with multi-dimensional data and explain non-linear relationships [32,33]. A multi-view clustering method, which utilizes graph regularized NMF, is proposed to enhance the extraction of useful information through graph embedding and remove the redundant information through orthogonality constraints for each view [34]. Additionally, a novel multi-view clustering method is developed, incorporating deep graph regularization into the NMF framework to extract a more abstract representation through the construction of a multi-layer NMF model [35]. A semi-supervised model is developed for adaptive learning of similarity graphs, which utilizes must-link and cannot-link scenarios to discover high-quality similarity graphs for achieving the final clustering result [36]. And a new multi-view graph learning framework is introduced, which explicitly addresses multi-view consistency and inconsistency, effectively tackling quality and noise issues in multi-view data [37]. However, it has been demonstrated that graph Laplace regularization only models the characteristics of each node in the graph without considering the relationships between the nodes [38]. Thus propagation regularization is proposed as a novel graph constraint, which is a variant of the regularization based on the graph Laplacian and provides new supervisory signals for the nodes in the graph [38].

Noise is inevitably generated during the sampling of data. To reduce the noise impact, some robust multi-view NMF methods have been proposed [27,39]. In [40], a robust algorithm is proposed by employing joint non-negative matrix factorization, which explores both discriminative and non-discriminative information present in common and specific components across different views. In [41], a novel robust multi-view NMF is proposed, which focuses on leveraging high-order similarity information to comprehensively uncover neighborhood structural details. Previous study has shown that introducing $l_{2,1}$ norm with manifold regularization in clustering tasks can reduce the effect of outliers [22]. Therefore, the $l_{2,1}$ norm has been broadly applied in instances to deal with the effects of noise [42,43]. In the context of multi-view tasks, robust manifold NMF models are proposed, which incorporate the $l_{2,1}$ norm as a quality measure for the factorization [22,44]. The $l_{2,1}$ norm with column-wise sparsity adds the l_2 norms of all columns with equal weights but often fails to achieve sufficient sparseness for columns [45]. To overcome this limitation, the $l_{2,log}$-(pseudo) norm is developed to enhance sparsity for noise reduction [45]. However, the column sparsity problem is not yet well addressed for multi-view clustering tasks.

While many multi-view clustering methods based on NMF have made significant progress, there still remain several issues: (1) Redundant information commonly exists in the data representations obtained from multiple views. (2) The similarity between views is not fully taken into account. (3) The graph Laplacian regularization only makes representations of neighboring nodes closer but does not emphasize mutual influences and information propagation between nodes. (4) The $l_{2,1}$ norm has limitations when addressing noise problems. To address the mentioned issues, a novel method for multi-view clustering is proposed in this paper, which is named centric graph regularized log-norm sparse NMF for multi-view clustering. Fig. 1 presents the overall framework of the proposed model and the contributions are summarized in the following four points.

(1) A novel centric graph regularization is designed to enable each node to capture information from more distant nodes, constructing a graph with improved spatial structure.

(2) A pairwise co-regularization is employed to measure the similarity between views, allowing more information between views and making multi-view data space compact.

(3) A log-based sparse constrained multi-view NMF model is proposed which utilizes $l_{2,log}$-(pseudo) norm to restrict column sparsity. The model ensures that sparse solutions are generated at each view to obtain a better part-based representation and reduce the mutually redundant information in multiple views.

(4) An iterative optimization algorithm is devised for the proposed optimization model. The objective function exhibits a non-increasing monotonicity after each iteration of the optimization algorithm. Extensive experiments are conducted on eight multi-view datasets. The experimental results demonstrate that the proposed method outperforms state-of-the-art methods in terms of several metrics.

The remainder of this paper is organized as follows. In Section 2, the NMF, propagation regularization, and sparsity-induction norm are briefly outlined. Section 3 details the proposed multi-view NMF model. Section 4 discusses the performance comparison and convergent results. In the last section, the conclusion is drawn.

2. Preliminaries

Before detailing the proposed method, it is important to review the closely related work about NMF, propagation regularization and sparsity-induction norm.

2.1. Non-negative matrix factorization

The principle of NMF [46] is as follows. In the given data matrix $\boldsymbol{X} \in R^{p\times q}$, where p and q represent the data dimension and sample size, respectively. The objective of the NMF model is to discover the non-negative matrices $\boldsymbol{U} = [u_{ij}] \in R^{p\times r}$ and $\boldsymbol{V} = [v_{ij}] \in R^{q\times r}$. Here, u_{ij} and v_{ij} represent the ijth element in the matrix $\boldsymbol{U}$ and $\boldsymbol{V}$, respectively, and r represents the desired reduced dimension. Using the basis matrix $\boldsymbol{U}$ and the coefficient matrix $\boldsymbol{V}$, the relationship between the matrices is approximated as $\boldsymbol{X} \approx \boldsymbol{U}\boldsymbol{V}^T$. The measure of similarity is accomplished by computing the distance. The commonly employed distance metric is

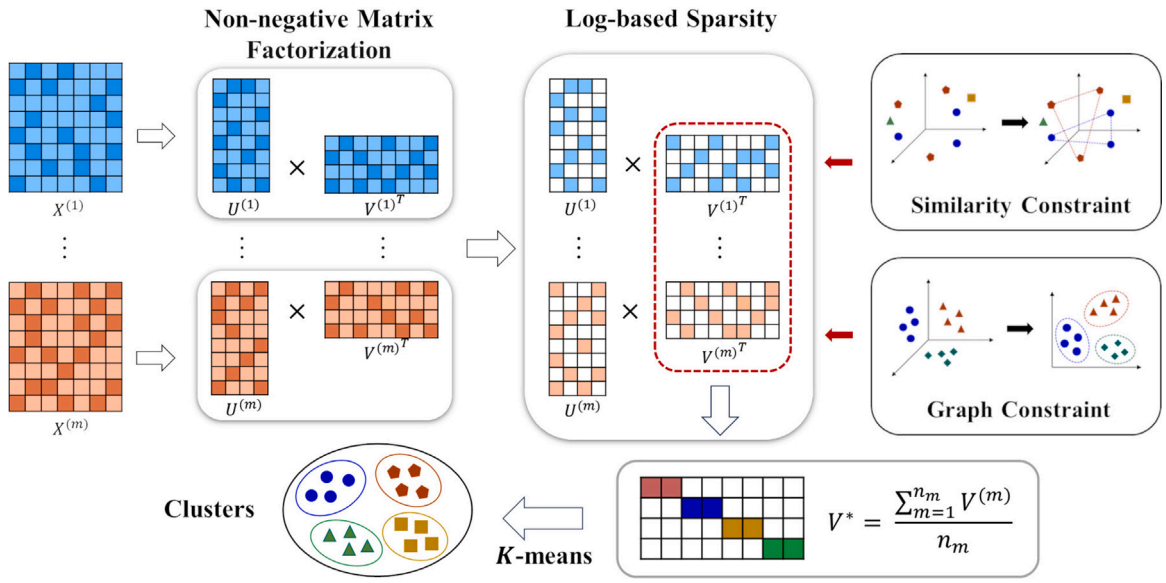


Fig. 1. The illustration of the proposed method. Firstly NMF is performed on the input matrix. Secondly, a novel graph constraint is applied to the coefficient matrix by centric graph regularization. Thirdly, a similarity constraint is implemented on the coefficient matrix through pairwise co-regularization. Then the sparseness of factorized matrices is enforced by $l_{2,log}$-(pseudo) norm. Finally, clustering results are obtained by performing K-means on the consensus coefficient matrix $\boldsymbol{V}^*$.

the Frobenius norm, which represents the squared Euclidean distance between two matrices [47]. Thus, the NMF model is expressed as an optimization problem:

$$\min_{\boldsymbol{U},\boldsymbol{V}} \|\boldsymbol{X} - \boldsymbol{U}\boldsymbol{V}^T\|_F^2, \quad \text{s.t.} \quad \boldsymbol{U},\boldsymbol{V} \geq 0 \tag{1}$$

where $\|\cdot\|_F$ represents the Frobenius norm. The constraints $\boldsymbol{U} \geq 0$ and $\boldsymbol{V} \geq 0$ indicate that all elements of matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ are non-negative. The objective function of problem (1) is known to be convex only in either $\boldsymbol{U}$ or $\boldsymbol{V}$, which makes it impossible to find the global minimum. Therefore, the following multiplicative update rules are proposed to achieve the optimal solutions:

$$u_{ij} \leftarrow u_{ij}\frac{(\boldsymbol{X}\boldsymbol{V})_{ij}}{(\boldsymbol{U}\boldsymbol{V}^T\boldsymbol{V})_{ij}}, \quad v_{ij} \leftarrow v_{ij}\frac{(\boldsymbol{X}^T\boldsymbol{U})_{ij}}{(\boldsymbol{V}\boldsymbol{U}^T\boldsymbol{U})_{ij}}. \tag{2}$$

2.2. Propagation regularization

The graph Laplace regularization is widely used for its ability to effectively discover the non-linear structural information of the data [48]. Moreover, to provide new supervisory signals to the nodes in the graph and supply additional information to improve model reliability, a propagation regularization is proposed based on the graph Laplacian regularization [38]. The propagation regularization is expressed as

$$L_{p-reg} = \frac{1}{q}\phi(\boldsymbol{V}, \hat{\boldsymbol{A}}\boldsymbol{V}) \tag{3}$$

where $\boldsymbol{A} = [a_{ij}] \in R^{q\times q}$ is the similarity matrix that describes the degree of similarity between data points. And $\hat{\boldsymbol{A}} = \boldsymbol{D}^{-1}\boldsymbol{A}$ is a normalized similarity matrix, where $\boldsymbol{D} = [d_{ij}] \in R^{q\times q}$ is a diagonal degree matrix with $d_{ij} = \sum_{j=1}^{q} a_{ij}$. $\hat{\boldsymbol{A}}\boldsymbol{V}$ is the further propagated output and $\phi(\boldsymbol{V}, \hat{\boldsymbol{A}}\boldsymbol{V})$ is a function to measure the difference between $\boldsymbol{V}$ and $\hat{\boldsymbol{A}}\boldsymbol{V}$ directly. While using squared error as ϕ and denote it as ϕ_{SE}, there is

$$\phi_{SE}(\boldsymbol{V}, \hat{\boldsymbol{A}}\boldsymbol{V}) = \frac{1}{2}\sum_{i=1}^{q}\|(\hat{\boldsymbol{A}}\boldsymbol{V})_i^T - (\boldsymbol{V})_i^T\|_2^2 = \frac{1}{2}\|\hat{\boldsymbol{A}}\boldsymbol{V} - \boldsymbol{V}\|_F^2 \tag{4}$$

where $(.)_i^T$ denotes the vector of the ith row in a matrix. $\phi_{SE}(\boldsymbol{V}, \hat{\boldsymbol{A}}\boldsymbol{V})$ is node-centric, which involves aggregating the information from a node's neighbors to serve as supervision targets [38]. This allows each node to acquire additional categorical information from its neighbors, aiding in better determining their positions and roles within the data, particularly in capturing complex non-linear structures.

2.3. Sparsity-induction norm

In NMF, sparse solutions always lead to better parts-based representations, and further improve the robustness [49]. Therefore, $l_{2,1}$ norm with column-wise sparsity is used instead of the l_2 norm, which is the sum of the l_2 norms of all column vectors in the matrix. For a matrix $\boldsymbol{G} = [g_{ij}] \in R^{p\times q}$, it is defined as

$$\|\boldsymbol{G}\|_{2,1} = \sum_{j=1}^{q}\|\boldsymbol{g}_j\|_2. \tag{5}$$

However, the $l_{2,1}$ norm and l_1 norm exhibit similar limitations in achieving adequate column sparsity [45]. Specifically, as the size of the input matrix increases, there is a tendency for the approximation error to grow, potentially leading to inaccurate approximations and non-optimal solutions. Hence, $\|\boldsymbol{G}\|_{log} = \sum_{i=1}^{p}\sum_{j=1}^{q} log(1 + |g_{ij}|)$ is proposed to enhance the smoothness and reduce the solving complexity. And it is further extended to the $l_{2,1}$ norm by designing the following novel $l_{2,log}$-(pseudo) norm:

$$\|\boldsymbol{G}\|_{2,log} = \sum_{j=1}^{q} log(1 + \|\boldsymbol{g}_j\|_2). \tag{6}$$

In terms of denoising, $l_{2,log}$-(pseudo) norm can lead to more sparseness than $l_{2,1}$ norm [45]. Due to the log-based value being closer to 0 than the l_2-based value, it provides a more accurate approximation of the actual sparsity. In order to visually compare the robustness of l_1 norm, l_2 norm, $l_{2,1}$ norm and $l_{2,log}$-(pseudo) norm to noise, using the norms to normalize the matrix with added noise respectively. The norms based loss functions are employed to measure the effect of a certain noise intensity, the variation of norm value with noise intensity is shown in Fig. 2. It can be observed that $l_{2,log}$-(pseudo) norm is significantly more robust to noise than the other norms.

3. Main results

In this section, centric graph regularized log-norm sparse NMF for multi-view clustering is formulated as an optimization problem in detail. Then an iterative algorithm is designed, and the convergence behavior as well as the computational complexity are analyzed.

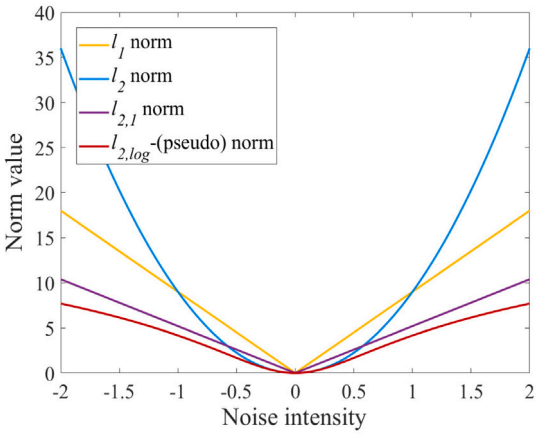


Fig. 2. Robust performance of l_1 norm, l_2 norm, $l_{2,1}$ norm and $l_{2,log}$-(pseudo) norm against noise.

3.1. Centric graph regularized NMF for multi-view clustering

Let $\boldsymbol{X}^{(m)} \in R^{p^{(m)}\times q}$ represents the mth view of the matrix, where $m = 1, 2, \ldots, n_m$. The data matrix for each view is factorized into $\boldsymbol{U}^{(m)} \in R^{p^{(m)}\times r}$ and $\boldsymbol{V}^{(m)} \in R^{q\times r}$, which are non-negative low-rank matrices. Thus, the problem of multi-view clustering based on NMF can be described as follows:

$$\min_{\boldsymbol{U}^{(m)},\boldsymbol{V}^{(m)}} \sum_{m=1}^{n_m} \|\boldsymbol{X}^{(m)} - \boldsymbol{U}^{(m)}\boldsymbol{V}^{(m)^T}\|_F^2, \quad \text{s.t.} \quad \boldsymbol{U}^{(m)}, \boldsymbol{V}^{(m)} \geq 0, m = 1, \ldots, n_m. \tag{7}$$

To capture the structures among distinct views, pairwise co-regularization is introduced to evaluate the similarity of the co-efficient matrices of the paired views [11,50]. The method aims to minimize $\|\boldsymbol{V}^{(m)} - \boldsymbol{V}^{(n)}\|_F^2$ for $m, n = 1, \ldots, n_m$ and $n \neq m$, to pursue the maximum similarity between $\boldsymbol{V}^{(m)}$ and $\boldsymbol{V}^{(n)}$. Higher coefficient matrix similarity serves a more important effect in clustering and the following cost function is adopted for measuring the similarity between views:

$$\min_{\boldsymbol{V}^{(m)}} \sum_{m,n=1,n\neq m}^{n_m} \|\boldsymbol{V}^{(m)} - \boldsymbol{V}^{(n)}\|_F^2, \quad \text{s.t.} \quad \boldsymbol{V}^{(m)} \geq 0, m = 1, \ldots, n_m. \tag{8}$$

The similarity of paired views captures the clustering structure and facilitates the part-based representation. However, it falls to fully consider non-linear associations among data. To solve the problem, a novel graph regularization is designed based on Eq. (4), which is named the centric graph regularization. The expression is as follows:

$$\frac{1}{2n_m} \sum_{m=1}^{n_m} \|\hat{\boldsymbol{A}}^{(m)}\boldsymbol{V}^{(m)} - \boldsymbol{V}^{(m)}\|_F^2, \quad \text{s.t.} \quad \boldsymbol{V}^{(m)} \geq 0, m = 1, \ldots, n_m. \tag{9}$$

In the above optimization problem, $\phi_{SE}(\boldsymbol{V}^{(m)}, \hat{\boldsymbol{A}}^{(m)}\boldsymbol{V}^{(m)})$ is used to discover the manifold structure of mth view. And it is normalized with the number of views, which aims to maintain multi-view spatial consistency. By incorporating centric graph regularization and pairwise co-regularization into the multi-view NMF framework, reformulate the problem (7) as follows:

$$\begin{aligned}\min_{\boldsymbol{U}^{(m)},\boldsymbol{V}^{(m)}} & \sum_{m=1}^{n_m} \|\boldsymbol{X}^{(m)} - \boldsymbol{U}^{(m)}\boldsymbol{V}^{(m)^T}\|_F^2 + \beta \sum_{m=1}^{n_m}\sum_{n=1,n\neq m}^{n_m} \|\boldsymbol{V}^{(m)} - \boldsymbol{V}^{(n)}\|_F^2 \\ & + \frac{\gamma}{2n_m} \sum_{m=1}^{n_m} \|\hat{\boldsymbol{A}}^{(m)}\boldsymbol{V}^{(m)} - \boldsymbol{V}^{(m)}\|_F^2, \\ \text{s.t.} \quad & \boldsymbol{U}^{(m)}, \boldsymbol{V}^{(m)} \geq 0, m = 1, \ldots, n_m\end{aligned} \tag{10}$$

where $\beta \geq 0$ and $\gamma \geq 0$ are the balancing parameters, which give the model greater degrees of freedom. The first term in the optimization problem is multi-view NMF. The second term corresponds to the pairwise co-regularization, which constrains the similarity between distinct views. The last term is the centric graph regularization, which enhances the non-linear properties of the data within the transformed low-dimensional space while incorporating additional remote node information to facilitate the model in capturing the spatial structure more efficiently.

3.2. Problem formulation

Due to the noises in data, sparse representation is needed for enhancing the robustness of the formulation (10), the optimization problem is expanded as:

$$\begin{aligned}\min_{\boldsymbol{U}^{(m)},\boldsymbol{V}^{(m)}} & \sum_{m=1}^{n_m} \|\boldsymbol{X}^{(m)} - \boldsymbol{U}^{(m)}\boldsymbol{V}^{(m)^T}\|_{2,1} + \beta \sum_{m=1}^{n_m}\sum_{n=1,n\neq m}^{n_m} \|\boldsymbol{V}^{(m)} - \boldsymbol{V}^{(n)}\|_F^2 \\ & + \frac{\gamma}{2n_m} \sum_{m=1}^{n_m} \|\hat{\boldsymbol{A}}^{(m)}\boldsymbol{V}^{(m)} - \boldsymbol{V}^{(m)}\|_F^2, \\ \text{s.t.} \quad & \boldsymbol{U}^{(m)}, \boldsymbol{V}^{(m)} \geq 0, m = 1, \ldots, n_m\end{aligned} \tag{11}$$

where $\|\cdot\|_{2,1}$ is the $l_{2,1}$ norm in Eq. (5). The $l_{2,1}$ norm remains fixed, ensuring the preservation of spatial information in the examples. However, optimizing the $l_{2,1}$ norm under the non-negativity constraint presents significant challenges. To facilitate the optimization process, it commonly further factorizes the data matrix considering noise [45]. In the mth view, it is factorized into $\boldsymbol{X}^{(m)} = \boldsymbol{U}^{(m)}\boldsymbol{V}^{(m)^T} + \boldsymbol{S}^{(m)}$, where $\boldsymbol{S}^{(m)}$ is a matrix with column-wise sparseness to interpret noise. Based on the above assumption, a closed-form solution is employed, and the optimization problem is relaxed as follows:

$$\begin{aligned}\min_{\boldsymbol{U}^{(m)},\boldsymbol{V}^{(m)},\boldsymbol{S}^{(m)}} & \sum_{m=1}^{n_m} \|(\boldsymbol{X}^{(m)} - \boldsymbol{S}^{(m)}) - \boldsymbol{U}^{(m)}\boldsymbol{V}^{(m)^T}\|_F^2 + \alpha \sum_{m=1}^{n_m} \|\boldsymbol{S}^{(m)}\|_{2,1} \\ & + \beta \sum_{m=1}^{n_m}\sum_{n=1,n\neq m}^{n_m} \|\boldsymbol{V}^{(m)} - \boldsymbol{V}^{(n)}\|_F^2 \\ & + \frac{\gamma}{2n_m} \sum_{m=1}^{n_m} \|\hat{\boldsymbol{A}}^{(m)}\boldsymbol{V}^{(m)} - \boldsymbol{V}^{(m)}\|_F^2, \quad \text{s.t.} \quad \boldsymbol{U}^{(m)}, \boldsymbol{V}^{(m)} \geq 0, m = 1, \ldots, n_m\end{aligned} \tag{12}$$

where $\alpha \geq 0$ is a balancing parameter. It can be observed that the relaxed optimization problem is easier to solve. However, it should be noted that the $l_{2,1}$ norm is computed in a closely related way to the l_1 norm, which sums the l_2 norm of all columns. Therefore it can be judged that the column-wise sparse property of $l_{2,1}$ norm has the same problem as the l_1 norm, which means that the regularization may be less efficient in approximating the real sparsity. Therefore, $l_{2,log}$-(pseudo) norm is introduced, which enhances the column sparsity property and better takes into account the noise impact [45]. The log-norm regularized sparse multi-view NMF model is obtained by utilizing the $l_{2,log}$-(pseudo) norm in Eq. (6) to measure the matrix $\boldsymbol{S}^{(m)}$, and is formulated as follows:

$$\begin{aligned}\min_{\boldsymbol{U}^{(m)},\boldsymbol{V}^{(m)},\boldsymbol{S}^{(m)}} & \sum_{m=1}^{n_m} \|(\boldsymbol{X}^{(m)} - \boldsymbol{S}^{(m)}) - \boldsymbol{U}^{(m)}\boldsymbol{V}^{(m)^T}\|_F^2 + \alpha \sum_{m=1}^{n_m} \|\boldsymbol{S}^{(m)}\|_{2,log} \\ & + \beta \sum_{m=1}^{n_m}\sum_{n=1,n\neq m}^{n_m} \|\boldsymbol{V}^{(m)} - \boldsymbol{V}^{(n)}\|_F^2 \\ & + \frac{\gamma}{2n_m} \sum_{m=1}^{n_m} \|\hat{\boldsymbol{A}}^{(m)}\boldsymbol{V}^{(m)} - \boldsymbol{V}^{(m)}\|_F^2, \quad \text{s.t.} \quad \boldsymbol{U}^{(m)}, \boldsymbol{V}^{(m)} \geq 0, m = 1, \ldots, n_m.\end{aligned} \tag{13}$$

Since the log-based approximation is closer to 0 than l_2 norm, the elements of the $\boldsymbol{S}^{(m)}$ matrix contain only essentially small values. It indicates that the matrix $\boldsymbol{S}^{(m)}$ represents noise and is indeed sparse.

3.3. The proposed optimization algorithm

In this subsection, the optimization problem of centric graph regularized log-norm sparse NMF for multi-view clustering is addressed.

3.3.1. $S^{(m)}$ Minimization

In the mth view, $m = 1, \dots, n_m$, for the minimization problem of the matrix $S^{(m)}$, the sub-problem is

$$\min_{S^{(m)}} \|(X^{(m)} - S^{(m)}) - U^{(m)}V^{(m)^T}\|_F^2 + \alpha\|S^{(m)}\|_{2,log}. \tag{14}$$

According to $l_{2,log}$ shrinkage operator theorem [45], in the mth view, given the matrix $H^{(m)} = X^{(m)} - U^{(m)}V^{(m)^T}$ and the non-negative parameter $\sigma = \frac{\alpha}{2}$, the problem is converted as follows:

$$\min_{S^{(m)}} \frac{1}{2}\|H^{(m)} - S^{(m)}\|_F^2 + \sigma\|S^{(m)}\|_{2,log}. \tag{15}$$

The closed-form solution of $s_i^{(m)}$ is

$$s_i^{(m)} = \begin{cases} \frac{\psi}{\|h_i^{(m)}\|_2} h_i^{(m)}, & \text{if } f_i^{(m)}(\psi) \le \frac{1}{2}\|h_i^{(m)}\|_2^2, (1 + \|h_i^{(m)}\|_2)^2 > 4\sigma, \psi > 0 \\ 0, & \text{otherwise,} \end{cases} \tag{16}$$

where $f_i^{(m)}(t) = \frac{1}{2}(t - \|h_i^{(m)}\|_2)^2 + \sigma log(1 + t)$ and $\psi = \frac{\|h_i^{(m)}\|_2 - 1}{2} + \sqrt{\frac{(1+\|h_i^{(m)}\|_2)^2}{4} - \sigma}$.

3.3.2. Lagrange function construction

In the optimization problem (13), fixing $S^{(m)}$, consider $U^{(m)}$ and $V^{(m)}$, the associated sub-problem of each view is formulated as follows:

$$\begin{aligned} &\min_{U^{(m)}, V^{(m)}} \|(X^{(m)} - S^{(m)}) - U^{(m)}V^{(m)^T}\|_F^2 \\ &+ \beta \sum_{n=1, n\neq m}^{n_m} \|V^{(m)} - V^{(n)}\|_F^2 + \frac{\gamma}{2n_m}\|\hat{A}^{(m)}V^{(m)} - V^{(m)}\|_F^2, \\ &\text{s.t.} \quad U^{(m)}, V^{(m)} \ge 0, m = 1, \dots, n_m. \end{aligned} \tag{17}$$

Previous study has shown that all columns of $X^{(m)} - S^{(m)}$ are the non-negative values, so the matrix $X^{(m)} - S^{(m)}$ is non-negative [45]. Therefore, the variables can be updated iteratively using the multiplication strategy. Let $\delta^{(m)} = [\delta_{ij}^{(m)}]$ and $\eta^{(m)} = [\eta_{ij}^{(m)}]$ be the Lagrange multipliers of constraint $U^{(m)} \ge 0$ and $V^{(m)} \ge 0$, respectively. The following Lagrange function L is obtained according to problem (17):

$$\begin{aligned} L = &tr((X^{(m)} - S^{(m)})(X^{(m)} - S^{(m)})^T - 2U^{(m)}V^{(m)^T}(X^{(m)} - S^{(m)})^T \\ &+ U^{(m)}V^{(m)^T}V^{(m)}U^{(m)^T}) \\ &+ \beta \sum_{n=1, n\neq m}^{n_m} tr(V^{(n)}V^{(n)^T} - 2V^{(n)}V^{(m)^T} + V^{(m)}V^{(m)^T}) \\ &+ \frac{\gamma}{2n_m} tr(\hat{A}^{(m)}V^{(m)}V^{(m)^T}\hat{A}^{(m)^T} \\ &- 2V^{(m)}V^{(m)^T}\hat{A}^{(m)^T} + V^{(m)}V^{(m)^T}) + tr(\delta^{(m)}U^{(m)}) + tr(\eta^{(m)}V^{(m)}). \end{aligned} \tag{18}$$

The iterative updating of $U^{(m)}$ and $V^{(m)}$ is discussed in the subsequent parts.

3.3.3. Fix $V^{(m)}$ and update $U^{(m)}$

Let the partial derivative of L with regard to $U^{(m)}$ be 0, gives

$$\delta^{(m)} = 2(X^{(m)} - S^{(m)})V^{(m)} - 2U^{(m)}V^{(m)^T}V^{(m)}. \tag{19}$$

Following the Karush-Kuhn–Tucker (KKT) condition [51] $\delta_{ij}^{(m)}u_{ij}^{(m)} = 0$, according to Eq. (19), we have:

$$(2(X^{(m)} - S^{(m)})V^{(m)} - 2U^{(m)}V^{(m)^T}V^{(m)})u_{ij}^{(m)} = 0. \tag{20}$$

Then we can obtain the updating rule of $u_{ij}^{(m)}$:

$$u_{ij}^{(m)} \leftarrow u_{ij}^{(m)} \frac{((X^{(m)} - S^{(m)})V^{(m)})_{ij}}{(U^{(m)}V^{(m)^T}V^{(m)})_{ij}}. \tag{21}$$

3.3.4. Fix $U^{(m)}$ and update $V^{(m)}$

Let the partial derivative of L with regard to $V^{(m)}$ be 0, it gives that

$$\begin{aligned} \eta^{(m)} = &2(X^{(m)} - S^{(m)})^T U^{(m)} - 2V^{(m)}U^{(m)^T}U^{(m)} + 2\beta \sum_{n=1, n\neq m}^{n_m} V^{(n)} \\ &- 2\beta(n_m - 1)V^{(m)} - \frac{\gamma}{n_m}(\hat{A}^{(m)^T}\hat{A}^{(m)}V^{(m)} - 2\hat{A}^{(m)^T}V^{(m)} + V^{(m)}). \end{aligned} \tag{22}$$

Equivalently make $\eta_{ij}^{(m)}v_{ij}^{(m)} = 0$, the following equation for $v_{ij}^{(m)}$ is obtained:

$$\begin{aligned} &(2(X^{(m)} - S^{(m)})^T U^{(m)} - 2V^{(m)}U^{(m)^T}U^{(m)} + 2\beta \sum_{n=1, n\neq m}^{n_m} V^{(n)} \\ &- 2\beta(n_m - 1)V^{(m)} - \frac{\gamma}{n_m}(\hat{A}^{(m)^T}\hat{A}^{(m)}V^{(m)} - 2\hat{A}^{(m)^T}V^{(m)} + V^{(m)}))v_{ij}^{(m)} = 0. \end{aligned} \tag{23}$$

The following updating rule is derived:

$$v_{ij}^{(m)} \leftarrow v_{ij}^{(m)} \frac{((X^{(m)} - S^{(m)})^T U^{(m)} + \beta\sum_{n=1, n\neq m}^{n_m} V^{(n)} + \frac{\gamma}{n_m}\hat{A}^{(m)^T}V^{(m)})_{ij}}{(V^{(m)}U^{(m)^T}U^{(m)} + \beta(n_m - 1)V^{(m)} + \frac{\gamma}{2n_m}(\hat{A}^{(m)^T}\hat{A}^{(m)}V^{(m)} + V^{(m)}))_{ij}}. \tag{24}$$

Algorithm 1 summarizes the optimization process with the convergence condition defined as $iter \le k_{max}$, where $iter$ represents the number of iterations, and k_{max} is the maximum allowed number of iterations.

Algorithm 1 The description of the proposed multi-view clustering algorithm.

Input:

Multi-view datasets $X^{(1)}, X^{(2)}, ..., X^{(n_m)}$.
Balancing parameters α, β, γ.
Maximum iteration k_{max}.

```
for m = 1 to n_m do
    Normalize X^(m).
    Initialize U^(m), V^(m), S^(m).
end for
while iter ≤ k_max do
    for m = 1 to n_m do
        Fix V^(m), update U^(m) by Eq. (21).
        Fix U^(m), update V^(m) by Eq. (24).
    end for
end while
Calculate the consensus coefficient matrix by V* = (Σ_{m=1}^{n_m} V^(m)) / n_m.
```

Output:

Consensus coefficient matrix V^*. Execute K-means on V^* to accomplish clustering.

3.4. Convergence analysis

In (17), the second and third terms relate only to $V^{(m)}$. Moreover, it is known that $X^{(m)} - S^{(m)}$ is non-negative if the initial values of $U^{(m)}$ and $V^{(m)}$ are non-negative. Therefore, the update formula for $U^{(m)}$ is treated in the method the same as in [46] by replacing $X^{(m)}$ with $X^{(m)} - S^{(m)}$. Now it needs to display that the objective function of (17) is non-increasing under the updating rule (24). Following the proof in [46], an auxiliary function is firstly defined as follows:

Definition 1. If the functions $G(v, v^t)$ and $F(v)$ satisfy the following conditions:

$$G(v, v^t) \geq F(v), \quad G(v, v) = F(v), \tag{25}$$

$G(v, v^t)$ is an auxiliary function of $F(v)$.

Lemma 1. *If $G(v, v^t)$ is an auxiliary function of $F(v)$, optimize the variable v according to the following rule:*

$$v^{t+1} = arg \min_{v} G(v, v^t), \tag{26}$$

$F(v)$ is non-increasing after each update.

Proof. The proof can be easily verified via following inequalities:

$$F(v^{t+1}) \leq G(v^{t+1}, v^t) \leq G(v^t, v^t) = F(v^t). \tag{27}$$

The next step is to construct a suitable auxiliary function such that the updating rule in (24) is equivalent to the update step in (26). The objective function of problem (17) is expressed in the following component form:

$$\begin{aligned} O_m &= \|(\boldsymbol{X}^{(m)} - \boldsymbol{S}^{(m)}) - \boldsymbol{U}^{(m)}\boldsymbol{V}^{(m)^T}\|_F^2 + \beta \sum_{n=1,n\neq m}^{n_m} \|\boldsymbol{V}^{(m)} - \boldsymbol{V}^{(n)}\|_F^2 \\ &+ \frac{\gamma}{2n_m}\|\hat{\boldsymbol{A}}^{(m)}\boldsymbol{V}^{(m)} - \boldsymbol{V}^{(m)}\|_F^2 \\ &= \sum_{i=1}^{p^{(m)}}\sum_{j=1}^{q}(x_{ij}^{(m)} - s_{ij}^{(m)} - \sum_{k=1}^{r} u_{ik}^{(m)} v_{kj}^{(m)})^2 + \beta \sum_{n=1,n\neq m}^{n_m}\sum_{j=1}^{q}\sum_{k=1}^{r}(v_{jk}^{(m)} - v_{jk}^{(n)})^2 \\ &+ \frac{\gamma}{2n_m}\sum_{j=1}^{q}\sum_{k=1}^{r}(\hat{a}_{jj}^{(m)} v_{jk}^{(m)} - v_{jk}^{(m)})^2. \end{aligned} \tag{28}$$

Given an element $v_{ab}^{(m)}$ in $\boldsymbol{V}^{(m)}$, $F_{ab}^{(m)}$ represents the part of Eq. (28) that is only relevant to $v_{ab}^{(m)}$. The first and second order derivatives of $F_{ab}^{(m)}$ with respect to $v_{ab}^{(m)}$ can be obtained directly:

$$\begin{aligned} F_{ab}{}' &= (\frac{\partial O_m}{\partial \boldsymbol{V}^{(m)}})_{ab} = (-2(\boldsymbol{X}^{(m)} - \boldsymbol{S}^{(m)})^T\boldsymbol{U}^{(m)} + 2\boldsymbol{V}^{(m)}\boldsymbol{U}^{(m)^T}\boldsymbol{U}^{(m)})_{ab} \\ &+ 2\beta((n_m - 1)\boldsymbol{V}^{(m)} - \sum_{n=1,n\neq m}^{n_m}\boldsymbol{V}^{(n)})_{ab} + \frac{\gamma}{n_m}(\hat{\boldsymbol{A}}^{(m)^T}\hat{\boldsymbol{A}}^{(m)}\boldsymbol{V}^{(m)} \\ &- 2\hat{\boldsymbol{A}}^{(m)^T}\boldsymbol{V}^{(m)} + \boldsymbol{V}^{(m)})_{ab}, \end{aligned} \tag{29}$$

$$\begin{aligned} F_{ab}'' &= (\frac{\partial F_{ab}^{(m)'}}{\partial \boldsymbol{V}^{(m)}})_{ab} = (2\boldsymbol{U}^{(m)^T}\boldsymbol{U}^{(m)})_{bb} + (2\beta(n_m - 1) + \frac{\gamma}{n_m})I_{bb} \\ &+ \frac{\gamma}{n_m}(\hat{\boldsymbol{A}}^{(m)^T}\hat{\boldsymbol{A}}^{(m)} - 2\hat{\boldsymbol{A}}^{(m)^T})_{aa}. \end{aligned} \tag{30}$$

As the updates are element-wise, it is adequate to establish that $F_{ab}^{(m)}$ remains non-increasing under the updating rule presented in (24).

Lemma 2. *The function*

$$\begin{aligned} G(v^{(m)}, v_{ab}^{(m)^t}) &= F_{ab}(v_{ab}^{(m)^t}) + F_{ab}'(v_{ab}^{(m)^t})(v^{(m)} - v_{ab}^{(m)^t}) \\ &+ \frac{(\boldsymbol{V}^{(m)}\boldsymbol{U}^{(m)^T}\boldsymbol{U}^{(m)})_{ab} + (\beta(n_m - 1) + \frac{\gamma}{2n_m})\boldsymbol{V}_{ab}^{(m)} + \frac{\gamma}{2n_m}(\hat{\boldsymbol{A}}^{(m)^T}\hat{\boldsymbol{A}}^{(m)}\boldsymbol{V}^{(m)})_{ab}}{v_{ab}^{(m)^t}} \\ &\times (v^{(m)} - v_{ab}^{(m)^t})^2 \end{aligned} \tag{31}$$

is an auxiliary function of F_{ab}.

Proof. It is obvious that $G(v^{(m)}, v^{(m)}) = F_{ab}(v^{(m)})$, therefore it is only necessary to prove that $G(v^{(m)}, v^{(m)^t}) \geq F_{ab}(v^{(m)})$. And the Taylor series expansion of $F_{ab}(v^{(m)})$ is

$$F_{ab}(v^{(m)}) = F_{ab}(v_{ab}^{(m)^t}) + F_{ab}'(v^{(m)} - v_{ab}^{(m)^t}) + F_{ab}''(v^{(m)} - v_{ab}^{(m)^t})^2. \tag{32}$$

Take Eqs. (29) and (30) into (32), the proving of $G(v^{(m)}, v^{(m)^t}) \geq F_{ab}(v^{(m)})$ can be reformulated as the demonstration of the following inequality

$$\begin{aligned} &\frac{(\boldsymbol{V}^{(m)}\boldsymbol{U}^{(m)^T}\boldsymbol{U}^{(m)})_{ab} + (\beta(n_m - 1) + \frac{\gamma}{2n_m})\boldsymbol{V}_{ab}^{(m)} + \frac{\gamma}{2n_m}(\hat{\boldsymbol{A}}^{(m)^T}\hat{\boldsymbol{A}}^{(m)}\boldsymbol{V}^{(m)})_{ab}}{v_{ab}^{(m)^t}} \\ &\geq (\boldsymbol{U}^{(m)^T}\boldsymbol{U}^{(m)})_{bb} + (\beta(n_m - 1) + \frac{\gamma}{2n_m})I_{bb} + \frac{\gamma}{2n_m}(\hat{\boldsymbol{A}}^{(m)^T}\hat{\boldsymbol{A}}^{(m)} - 2\hat{\boldsymbol{A}}^{(m)^T})_{aa}. \end{aligned} \tag{33}$$

Since we have

$$(\boldsymbol{V}^{(m)}\boldsymbol{U}^{(m)^T}\boldsymbol{U}^{(m)})_{ab} = \sum_{k=1}^{r} v_{ak}^{(m)^t}(\boldsymbol{U}^{(m)^T}\boldsymbol{U}^{(m)})_{kb} \geq v_{ab}^{(m)^t}(\boldsymbol{U}^{(m)^T}\boldsymbol{U}^{(m)})_{bb}, \tag{34}$$

$$\begin{aligned} (\beta(n_m - 1) + \frac{\gamma}{2n_m})\boldsymbol{V}_{ab}^{(m)} &= (\beta(n_m - 1) + \frac{\gamma}{2n_m})\sum_{j=1}^{q} v_{aj}^{(m)^t} I_{jb} \\ &\geq v_{ab}^{(m)^t}(\beta(n_m - 1) + \frac{\gamma}{2n_m})I_{bb} \end{aligned} \tag{35}$$

and

$$\begin{aligned} &\frac{\gamma}{2n_m}(\hat{\boldsymbol{A}}^{(m)^T}\hat{\boldsymbol{A}}^{(m)}\boldsymbol{V}^{(m)})_{ab} = \frac{\gamma}{2n_m}\sum_{j=1}^{q}(\hat{\boldsymbol{A}}^{(m)^T}\hat{\boldsymbol{A}}^{(m)})_{aj} v_{jb}^{(m)^t} \\ &\geq \frac{\gamma}{2n_m}(\hat{\boldsymbol{A}}^{(m)^T}\hat{\boldsymbol{A}}^{(m)})_{aa} v_{ab}^{(m)^t} \geq \frac{\gamma}{2n_m}(\hat{\boldsymbol{A}}^{(m)^T}\hat{\boldsymbol{A}}^{(m)} - 2\hat{\boldsymbol{A}}^{(m)^T})_{aa} v_{ab}^{(m)^t}. \end{aligned} \tag{36}$$

Therefore, Eq. (33) holds, which leads to $G(v^{(m)}, v^{(m)^t}) \geq F_{ab}(v^{(m)})$.

Theorem 1. *The objective function of (17) is non-increasing under the updating rule (24).*

Proof. Replacing $G(v^{(m)}, v_{ab}^{(m)^t})$ in Eq. (26) by Eq. (31), resulting in the iterative updating rule (see Eq. (37) in Box I).

This is essentially the updating rule of Eq. (24). As Eq. (31) is an auxiliary function of $F_{ab}(v^{(m)})$, according to Lemma 1, it can be concluded that $F_{ab}(v^{(m)})$ is non-increasing under Eq. (24). Combining the previous discussions about updating $\boldsymbol{U}^{(m)}$, it is evident that the proposed algorithm ensures the non-increasing change of objective function in problem (17).

3.5. Computational complexity analysis

This subsection discusses the computational complexity of the proposed model. The main computational cost of the model lies in the iterative updates of $\boldsymbol{U}^{(m)}$ and $\boldsymbol{V}^{(m)}$, where $m = 1, 2, \ldots, n_m$. The essential step involves solving the optimization problem in (21), which is performed for each column of $\boldsymbol{U}^{(m)}$. The computational cost of the updating rule (21) is $O(p^{(m)}q)$, resulting in a total cost of $O(rp^{(m)}q)$ for $\boldsymbol{U}^{(m)}$. Similarly, the computational cost of $\boldsymbol{V}^{(m)}$ is $O(rp^{(m)}q)$.

In conclusion, the computational cost of optimizing the mth view objective function is $O(rp^{(m)}q)$ in combination with $\boldsymbol{U}^{(m)}$ and $\boldsymbol{V}^{(m)}$. Therefore, the total computational complexity of the proposed model in all views is $O(rn_m pq)$, where $p = max\{p^{(1)}, p^{(2)}, \ldots, p^{(n_m)}\}$.

4. Experimental results

Extensive experiments are provided in this section to substantiate the effectiveness of the proposed method. The datasets and codes are available at https://github.com/dyz200219/CRLSNMF.

4.1. Description of datasets

The experiments utilize eight real-world datasets. These benchmark datasets serve as the basis for evaluating the method's performance, and the summary of eight datasets is presented in Table 1.

$$
\begin{aligned}
v_{ab}^{(m)t+1} &= v_{ab}^{(m)t} - v_{ab}^{(m)t} \frac{F'_{ab}(v_{ab}^{(m)t})}{2(\boldsymbol{V}^{(m)}\boldsymbol{U}^{(m)^T}\boldsymbol{U}^{(m)})_{ab} + (2\beta(n_m-1)+\frac{\gamma}{n_m})\boldsymbol{V}_{ab}^{(m)} + \frac{\gamma}{n_m}(\hat{\boldsymbol{A}}^{(m)^T}\hat{\boldsymbol{A}}^{(m)}\boldsymbol{V}^{(m)})_{ab}} \\
&= v_{ab}^{(m)t} \frac{2((\boldsymbol{X}^{(m)}-\boldsymbol{S}^{(m)})^T\boldsymbol{U}^{(m)})_{ab} + 2\beta\sum_{n=1,n\neq m}^{n_m}\boldsymbol{V}_{ab}^{(n)} + \frac{2\gamma}{n_m}(\hat{\boldsymbol{A}}^{(m)^T}\boldsymbol{V}^{(m)})_{ab}}{2(\boldsymbol{V}^{(m)}\boldsymbol{U}^{(m)^T}\boldsymbol{U}^{(m)})_{ab} + (2\beta(n_m-1)+\frac{\gamma}{n_m})\boldsymbol{V}_{ab}^{(m)} + \frac{\gamma}{n_m}(\hat{\boldsymbol{A}}^{(m)^T}\hat{\boldsymbol{A}}^{(m)}\boldsymbol{V}^{(m)})_{ab}}.
\end{aligned}
\tag{37}
$$

Box I.

Table 1
The summary eight datasets.

Dataset	Instances(N)	Views(P)	Cluster(C)
Yale	165	3	15
3Sources	169	3	6
BBCSport	544	2	5
Wisconsin	265	2	5
Handwritten	2000	5	10
20NewsGroups	500	3	5
Caltech20	2386	5	20
Scene	2688	3	8

- **Yale**[1]**:** The dataset consists of 165 black and white facial images representing 15 distinct individuals, resulting in 15 different clusters. It contains facial features captured from three distinct perspectives.
- **3Sources**[2]**:** The data comes from three different news sources, reporting a total of 169 stories. The stories are categorized into six subject tags.
- **BBCSport**[3]**:** The dataset consists of 544 documents fetched from websites, and each document is divided into two parts. The subject of each news item is assigned to one of five topic tags, resulting in five different clusters.
- **Wisconsin**[4]**:** The dataset consists of 265 web documents categorized into five groups. Each document is expressed by two feature bags.
- **Handwritten**[5]**:** The dataset contains 2000 instances of decimal numbers from 0 to 9, resulting in 10 different clusters. The data encompasses feature information from five different aspects.
- **20NewsGroups**[6]**:** This is a multi-view dataset of newsgroups, consisting of three views. Each view contains 500 instances, which are divided into five clusters.
- **Caltech20**[7]**:** The dataset contains 2386 images of objects belonging to 20 classes. All images are described using five types of features, which are Gabor, CENTRIST, HOG, GIST and LBP.
- **Scene**[8]: The dataset comprises 2688 images, divided into eight groups. For each image, three different feature vectors are used, including GIST, color moment, and HOG.

4.2. Compared methods

The proposed method is compared with the following state-of-the-art clustering methods to demonstrate its effectiveness.

- **NMF** [46]: The standard NMF achieves dimensionality reduction of the original matrix by seeking two non-negative matrices, resulting in a reduced-dimensional matrix of data features. Perform NMF on each view of the dataset and record the best results as the final experimental outcomes.
- **CoRegSPC**[9] [11]: This method makes different views appear to have a common consensus by normalizing their view-specific feature vectors. The graph Laplacian operators of all views are also combined so that each Laplacian produces a close to consistent underlying structure. We set the range for the weight parameter from 0.01 to 0.05, with an interval of 0.01.
- **GNMF**[10] [52]: GNMF is an improved model based on NMF that constructs an affine graph to encode geometric information in pursuit of matrix factorization with graph structure constraints. Execute GNMF on each view of the dataset, where the balancing parameter is selected from $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$, and record the best results as the final results.
- **MVKKM** [53]: The method executes the multi-view clustering task through unsupervised multi-kernel learning. It assigns different weights to each view's kernel matrix according to the view quality. According to the original paper, the value of parameter p is chosen from the set $\{1, 1.3, 1.5, 2, 4, 6\}$.
- **MultiNMF**[11] [23]: This is a manifold-based multi-view NMF method that explores the correlation among internal views and incorporates local graph regularization. Therefore, MultiNMF integrates the local geometric information for each view. In each view, the regularization parameter is set to 0.01, which is the parameter value recommended in the original paper.
- **RMKMC**[12] [54]: RMKMC extends the traditional K-means clustering to a robust multi-view K-means clustering. It integrates heterogeneous representations of large-scale data. The parameter γ can be configured, and according to the original paper, $\log_{10}\gamma$ varies within the range of 0.1 to 2 with an incremental step of 0.2.
- **DiNMF** [27]: The purpose of DiNMF is to enhance the diversity of data. It explores the diversity from different views and reduces redundancy between multi-view representations. Additionally, it makes the learning process in a linear execution time. For parameters α and β, values are chosen from $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$ for each parameter.
- **AMvDMD**[13] [20]: This method unveils the hierarchical semantics in data through a layered approach and captures hidden representations of distinct attributes. Consequently, it generates a new deep multi-view clustering model. We conduct experiments using the layer sizes as set in the original paper and record the best experimental results.

[1] https://cvc.yale.edu/projects/yalefaces/yalefaces.html.
[2] http://mlg.ucd.ie/datasets/3sources.html.
[3] https://mlg.ucd.ie/datasets/segment.html.
[4] https://lig-membres.imag.fr/grimal/data.html.
[5] http://archive.ics.uci.edu/ml/datasets/Multiple+Features.
[6] https://lig-membres.imag.fr/grimal/data.html.
[7] https://github.com/sudalvxin/2019-PR-Sparse-Multi-view-clustering/tree/master/Data.
[8] https://github.com/sudalvxin/SMSC/tree/master/data.
[9] https://sites.google.com/site/feipingnie/publications.
[10] http://www.cad.zju.edu.cn/home/dengcai/Data/GNMF.html.
[11] http://jialu.info/.
[12] https://sites.google.com/site/feipingnie/publications.
[13] https://github.com/huangsd/DeepMVC.

Table 2
Mean values and standard deviations of ACC, NMI, and Purity for clustering results with the proposed method and nine baseline methods on Yale (%).

Method	ACC	NMI	Purity
NMF	50.18(3.97)	53.39(3.15)	51.58(3.49)
CoRegSPC	60.24(5.61)	63.75(4.31)	61.58(5.41)
GNMF	46.85(2.37)	52.33(2.07)	48.18(2.56)
MVKKM	58.18(0.00)	61.99(0.00)	60.00(0.00)
MultiNMF	59.64(1.65)	61.97(1.40)	59.64(1.65)
RMKMC	44.61(4.01)	50.67(3.64)	45.70(3.88)
DiNMF	59.21(3.23)	61.72(2.87)	59.39(3.26)
AMvDMD	43.76(5.68)	48.93(5.73)	45.39(4.59)
DMFPA	56.42(6.33)	61.04(4.37)	57.52(5.65)
Ours	**64.36(3.54)**	**65.88(3.01)**	**64.85(3.37)**

Table 3
Mean values and standard deviations of ACC, NMI, and Purity for clustering results with the proposed method and nine baseline methods on 3Sources (%).

Method	ACC	NMI	Purity
NMF	59.05(3.84)	54.25(2.34)	74.02(0.76)
CoRegSPC	55.86(3.65)	50.74(1.88)	69.76(1.56)
GNMF	58.17(3.83)	51.73(2.74)	72.25(1.54)
MVKKM	35.50(0.00)	6.58(0.00)	39.05(0.00)
MultiNMF	50.36(3.53)	45.99(2.34)	61.18(2.54)
RMKMC	46.75(6.82)	32.59(6.00)	59.11(3.38)
DiNMF	54.97(6.78)	49.40(4.81)	68.34(4.13)
AMvDMD	56.75(9.67)	35.96(15.52)	60.53(9.70)
DMFPA	56.33(2.10)	55.81(3.11)	75.27(2.57)
Ours	**76.92(5.37)**	**66.62(4.91)**	**81.83(3.14)**

Table 4
Mean values and standard deviations of ACC, NMI, and Purity for clustering results with the proposed method and nine baseline methods on BBCSport (%).

Method	ACC	NMI	Purity
NMF	66.08(6.48)	48.16(5.71)	70.57(2.90)
CoRegSPC	48.92(1.97)	28.33(1.37)	52.39(0.61)
GNMF	72.21(2.65)	54.71(2.26)	75.62(2.52)
MVKKM	35.66(0.00)	1.23(0.00)	36.21(0.00)
MultiNMF	62.94(7.48)	45.43(3.91)	67.32(4.14)
RMKMC	65.13(6.03)	53.79(6.25)	70.96(4.88)
DiNMF	70.22(5.95)	53.74(4.82)	72.96(4.81)
AMvDMD	56.36(9.39)	41.95(9.11)	63.46(6.49)
DMFPA	56.89(1.63)	39.71(0.58)	63.36(1.36)
Ours	**78.03(3.69)**	**63.28(2.34)**	**79.08(1.68)**

- **DMFPA**[14] [25]: The method employs deep matrix factorization to gain the partition representation for each view and combines it with the optimal partition representation to achieve partition alignment. The parameters are set according to the original paper.

To perform a comprehensive assessment, three distinct evaluation metrics are employed: accuracy (ACC), normalized mutual information (NMI) and Purity. The detailed definitions for these metrics are given in [45]. Higher values for all three metrics indicate better clustering performance. By using different measurements, various aspects of the clustering performance are evaluated, allowing for a comprehensive assessment. Each comparison experiment is performed 10 times, and the mean values and standard deviations are recorded.

4.3. Comparison of clustering performance

Tables 2 to 9 present the clustering performance measured in terms of ACC, NMI, and Purity. The bold values indicate the best performance among the 10 advanced approaches. Through comparison, it can be observed that the proposed method is competitive among a series of advanced methods and outperforms other comparative methods in most cases. Based on the information presented in these tables, the following conclusions can be drawn.

[14] https://github.com/zhangchen234/MVC-DMF-PA.

Table 5
Mean values and standard deviations of ACC, NMI, and Purity for clustering results with the proposed method and nine baseline methods on Wisconsin (%).

Method	ACC	NMI	Purity
NMF	56.83(3.68)	35.83(4.47)	72.26(3.70)
CoRegSPC	55.36(2.55)	33.28(3.74)	71.70(2.55)
GNMF	57.32(1.20)	**42.64(1.64)**	74.08(1.31)
MVKKM	52.45(0.00)	13.45(0.00)	54.72(0.00)
MultiNMF	46.91(1.87)	11.48(3.55)	53.13(1.12)
RMKMC	45.21(4.69)	16.93(2.53)	60.38(2.97)
DiNMF	47.74(5.85)	18.67(6.93)	58.68(4.26)
AMvDMD	45.55(4.85)	16.08(6.41)	56.11(6.51)
DMFPA	44.79(0.84)	1.09(0.59)	46.19(0.48)
Ours	**60.60(1.84)**	42.45(3.42)	**75.92(2.64)**

Table 6
Mean values and standard deviations of ACC, NMI, and Purity for clustering results with the proposed method and nine baseline methods on Handwritten (%).

Method	ACC	NMI	Purity
NMF	68.57(5.74)	62.53(4.77)	70.66(4.59)
CoRegSPC	74.00(5.18)	70.36(2.57)	75.18(4.36)
GNMF	85.63(8.29)	82.69(5.10)	87.18(6.53)
MVKKM	73.15(0.00)	68.31(0.00)	73.15(0.00)
MultiNMF	79.02(4.37)	70.95(2.51)	79.38(3.73)
RMKMC	74.65(4.16)	72.98(2.93)	77.72(3.41)
DiNMF	71.35(4.84)	65.19(3.30)	72.07(4.46)
AMvDMD	79.57(7.46)	75.30(3.73)	80.70(5.39)
DMFPA	69.04(3.85)	64.06(1.43)	70.21(2.12)
Ours	**93.22(0.57)**	**86.96(0.91)**	**93.22(0.57)**

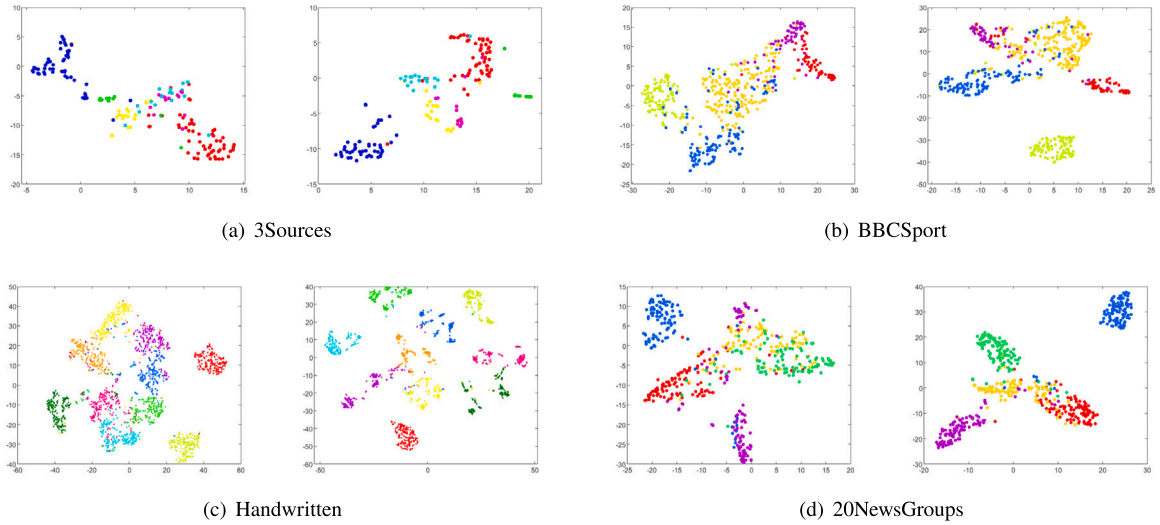
Table 7
Mean values and standard deviations of ACC, NMI, and Purity for clustering results with the proposed method and nine baseline methods on 20NewsGroups (%).

Method	ACC	NMI	Purity
NMF	60.48(6.47)	42.54(2.71)	61.68(5.62)
CoRegSPC	28.22(3.33)	4.85(2.47)	29.16(3.76)
GNMF	58.50(4.19)	38.37(2.42)	58.66(3.88)
MVKKM	21.00(0.00)	1.03(0.00)	21.00(0.00)
MultiNMF	23.40(1.48)	3.37(1.38)	23.92(1.73)
RMKMC	41.04(6.93)	15.32(6.30)	42.10(6.81)
DiNMF	52.66(10.08)	37.94(9.08)	55.54(8.75)
AMvDMD	39.60(7.38)	28.75(11.42)	40.58(8.02)
DMFPA	40.38(3.28)	17.32(3.69)	41.16(3.37)
Ours	**81.88(5.68)**	**65.67(4.33)**	**81.92(5.56)**

Table 8
Mean values and standard deviations of ACC, NMI, and Purity for clustering results with the proposed method and nine baseline methods on Caltech20 (%).

Method	ACC	NMI	Purity
NMF	36.70(1.65)	46.37(0.57)	69.58(0.80)
CoRegSPC	35.57(2.03)	41.09(0.66)	65.27(0.72)
GNMF	36.65(1.87)	39.42(0.91)	62.20(1.98)
MVKKM	40.53(0.00)	53.07(0.00)	**74.43(0.00)**
MultiNMF	37.26(1.93)	46.45(1.79)	69.91(1.30)
RMKMC	41.58(4.12)	45.18(2.61)	66.27(1.96)
DiNMF	33.09(1.93)	42.38(1.73)	67.92(1.35)
AMvDMD	41.04(2.92)	50.49(1.56)	71.75(2.10)
DMFPA	35.93(2.91)	15.66(4.70)	45.69(3.40)
Ours	**46.24(3.33)**	**53.78(1.35)**	73.90(1.55)

- As shown in Table 3, on the 3Sources, the improvement of the method over the second best method is about 17.87%, 12.37%, and 7.81% for ACC, NMI, and Purity, respectively. In Table 7, on the 20NewsGroups, the improvement is about 21.40%, 23.13%, and 20.24%. Despite a 0.19% decrease in NMI on the Wisconsin dataset and a 0.53% decrease in Purity on the Caltech20 dataset compared to the second best methods, the differences are small,



(a) 3Sources (b) BBCSport

(c) Handwritten (d) 20NewsGroups

Fig. 3. The proposed method and DiNMF use t-SNE [55] on 3Sources, BBCSport, Handwritten and 20NewsGroups. In each subfigure, the left part shows the effect of DiNMF visualized using t-SNE, and the right part shows the visualization of the proposed method. Different clusters for each dataset are indicated by different colors.

Table 9
Mean values and standard deviations of ACC, NMI, and Purity for clustering results with the proposed method and nine baseline methods on Scene (%).

Method	ACC	NMI	Purity
NMF	61.54(0.95)	44.90(0.88)	61.54(0.95)
CoRegSPC	64.04(1.08)	49.62(0.60)	64.14(0.99)
GNMF	41.03(2.22)	32.76(0.55)	44.90(2.03)
MVKKM	57.92(0.00)	45.26(0.00)	57.92(0.00)
MultiNMF	59.23(6.80)	46.54(3.25)	60.52(5.08)
RMKMC	48.31(5.22)	37.33(3.94)	50.41(4.57)
DiNMF	51.63(2.47)	39.27(1.29)	53.87(2.02)
AMvDMD	57.50(6.33)	45.65(3.40)	58.65(5.06)
DMFPA	52.93(1.45)	40.10(0.86)	56.48(1.41)
Ours	**67.23(2.24)**	**50.34(1.89)**	**67.23(2.24)**

and the other two metrics remain the best results as well. Overall, the proposed model generally outperforms the comparison methods.

- By comparing with the CoRegSPC and GNMF methods, which as well use graph regularization, evidently the proposed method demonstrates the best overall performance across all metrics. This indicates that the innovative utilization of the graph regularization variant, centric graph regularization, actually provides more valid information.
- On the majority of datasets, the proposed method outperforms other comparative approaches, providing evidence for the superiority of log-based regularization. In contrast to the Frobenius norm employed by most methods, after the novel and meaningful log-based sparse representation of the original objective function, a more part-based representation is obtained, which is of practical interest for NMF.

In addition, a two-sample t-test is used to determine if the proposed method significantly outperforms other clustering methods [56]. When significance level is set at 5%, the null hypothesis states that 'there is no difference between the proposed method and other methods', while the alternative hypothesis posits that 'the proposed method is superior to other methods'. The statistical test results for ACC, NMI, and Purity are displayed in Tables 10 to 12, respectively.

From Table 10, it can be observed that there is no significant difference between the proposed method and CoRegSPC on the Yale dataset. However, apart from this case, the p-values between the proposed method and each compared method are all less than 0.05, indicating that the proposed method achieves higher accuracy on these eight datasets. In Table 11, the NMI of the proposed method is lower than CoRegSPC on the Wisconsin dataset. Nevertheless, in most cases, the proposed method obtains higher NMI compared to the state-of-the-art methods. In Table 12, except for cases where there is no significant difference with specific comparison methods on the Yale, Wisconsin, and Caltech20 datasets, the proposed method achieves superior Purity in most cases.

4.4. Experiments of robustness

In this subsection, the proposed method is further evaluated on datasets with Gaussian and Poisson noise. Specifically, all methods are tested under different noise levels on the BBCSport dataset. Regarding the synthetic noise dataset for BBCSport, here is the provided information.

- **BBCSportG1**: This is a synthetic dataset where Gaussian noise with a mean of 0 and a variance of 0.005 is incorporated into each view of the original BBCSport dataset.
- **BBCSportG2**: This dataset introduces Gaussian noise to the original BBCSport. In this dataset, the Gaussian noise has an intensity characterized by a mean of 0 and a variance of 0.01.
- **BBCSportG3**: This dataset similarly incorporates Gaussian noise into the original BBCSport, with the Gaussian noise intensity characterized by a mean of 0 and a variance of 0.015.
- **BBCSportP**: This is a synthetic dataset where Poisson noise is added to each view of the original BBCSport dataset using functions from the MATLAB toolbox.

Tables 13 to 16 respectively provide the experimental results for four synthetic noise datasets. For four different types of noise, the proposed method exhibits better performance compared to the other methods. It can be observed that the proposed method is less affected by both Gaussian and Poisson noise. With increasing Gaussian noise, the performance of all methods generally decreases, confirming the adverse effects of noise. However, the proposed method is relatively less affected by the increase in noise. Therefore, the experiments on the four noise-inclusive datasets validate the robustness of the proposed method.

4.5. Visualization of the evolution of $\boldsymbol{V}^{(m)}$

To demonstrate the effectiveness of the coefficient matrix $\boldsymbol{V}^{(m)}$, where $m = 1, 2, \ldots, n_m$, the coefficient matrix $\boldsymbol{V}^{(m)}$ is visualized to

Table 10
P-values of comparisons between the proposed method and other clustering methods in terms of ACC.

Comparisons	Datasets							
	Yale	3Sources	BBCSport	Wisconsin	Handwritten	20NewsGroups	Caltech20	Scene
NMF	1.15e−07	9.31e−08	1.61e−04	1.22e−02	2.30e−07	3.13e−07	1.98e−07	3.24e−05
CoRegSPC	6.52e−02	6.05e−09	1.80e−14	7.55e−13	7.93e−07	1.16e−15	7.97e−08	7.37e−04
GNMF	8.08e−10	4.45e−08	7.39e−04	1.69e−04	1.78e−02	4.34e−09	2.75e−07	8.48e−16
MVKKM	3.72e−04	1.57e−09	4.48e−11	2.05e−07	1.91e−15	8.34e−11	4.21e−04	3.56e−07
MultiNMF	2.18e−03	1.27e−10	2.00e−05	2.64e−12	2.36e−06	3.37e−17	7.65e−07	4.70e−03
RMKMC	7.87e−10	2.05e−09	1.81e−05	1.49e−08	1.42e−07	2.49e−11	1.23e−02	1.76e−07
DiNMF	3.22e−03	2.34e−07	3.02e−03	4.12e−05	1.37e−07	2.51e−07	2.70e−09	1.62e−11
AMvDMD	1.34e−08	1.82e−05	2.18e−05	1.21e−06	2.57e−04	2.68e−11	1.61e−03	7.41e−04
DMFPA	2.78e−03	1.22e−07	2.37e−12	5.06e−12	5.96e−09	9.49e−14	7.68e−07	1.64e−12

Table 11
P-values of comparisons between the proposed method and other clustering methods in terms of NMI.

Comparisons	Datasets							
	Yale	3Sources	BBCSport	Wisconsin	Handwritten	20NewsGroups	Caltech20	Scene
NMF	3.95e−08	7.41e−06	5.38e−06	1.55e−03	3.03e−08	2.76e−11	1.67e−09	3.02e−02
CoRegSPC	2.16e−01	7.82e−07	2.20e−16	1.99e−05	5.88e−10	9.17e−19	6.44e−16	2.74e−01
GNMF	7.36e−10	1.26e−07	1.40e−07	8.77e−01	2.70e−02	1.05e−12	3.02e−16	3.10e−11
MVKKM	2.76e−03	2.57e−11	2.50e−14	6.68e−10	2.44e−13	4.27e−12	1.33e−01	1.37e−05
MultiNMF	2.69e−03	5.07e−10	3.08e−10	1.07e−13	6.13e−10	1.14e−19	5.47e−09	4.96e−03
RMKMC	6.88e−09	4.7e−11	8.23e−04	2.34e−13	2.34e−08	4.78e−14	2.95e−08	2.24e−08
DiNMF	5.42e−03	2.83e−07	2.43e−05	2.28e−07	1.22e−09	7.03e−08	2.92e−12	9.32e−12
AMvDMD	1.51e−07	1.03e−04	2.75e−05	2.03e−08	2.18e−06	7.99e−07	8.64e−05	1.87e−03
DMFPA	9.90e−03	1.44e−05	2.44e−11	1.05e−11	1.53e−19	5.54e−16	1.29e−10	6.81e−12

Table 12
P-values of comparisons between the proposed method and other clustering methods in terms of Purity.

Comparisons	Datasets							
	Yale	3Sources	BBCSport	Wisconsin	Handwritten	20NewsGroups	Caltech20	Scene
NMF	7.92e−08	1.72e−05	1.08e−06	2.02e−02	6.33e−08	2.06e−07	2.29e−06	2.41e−07
CoRegSPC	1.22e−01	5.87e−08	2.54e−20	1.85e−03	2.84e−07	2.18e−15	9.09e−10	8.75e−04
GNMF	2.75e−10	8.89e−07	2.01e−03	6.24e−02	1.69e−02	2.51e−09	1.78e−11	6.47e−15
MVKKM	1.38e−03	9.87e−12	3.53e−14	1.08e−09	1.91e−15	6.84e−11	3.07e−01	3.56e−07
MultiNMF	7.14e−04	3.74e−12	1.38e−07	2.89e−05	6.85e−07	3.37e−17	7.03e−06	2.30e−03
RMKMC	6.69e−10	6.85e−12	4.06e−04	3.02e−10	1.04e−07	2.78e−11	1.55e−08	1.01e−07
DiNMF	1.71e−03	1.66e−07	2.87e−03	2.36e−09	8.59e−08	2.25e−07	3.18e−08	4.08e−11
AMvDMD	2.69e−09	3.35e−06	2.15e−05	1.32e−06	4.03e−05	8.38e−11	1.80e−02	3.33e−04
DMFPA	2.42e−03	7.24e−05	8.63e−15	1.89e−11	8.47e−12	1.12e−13	7.12e−12	1.68e−10

Table 13
Mean values and standard deviations of ACC, NMI, and Purity for clustering results with the proposed method and nine baseline methods on BBCSportG1 (%).

Method	ACC	NMI	Purity
NMF	69.30(6.59)	48.73(6.49)	71.86(5.02)
CoRegSPC	54.61(7.52)	38.83(8.91)	63.86(8.56)
GNMF	56.45(8.82)	45.85(8.00)	65.72(7.60)
MVKKM	36.03(0.00)	1.11(0.00)	36.58(0.00)
MultiNMF	64.96(2.60)	50.37(1.98)	69.56(2.50)
RMKMC	63.71(9.27)	50.11(7.69)	69.89(6.37)
DiNMF	68.24(6.95)	48.08(4.32)	70.31(4.10)
AMvDMD	54.30(9.93)	30.74(12.55)	57.61(8.73)
DMFPA	56.64(3.17)	29.63(1.50)	60.70(0.83)
Ours	**77.28(2.69)**	**60.28(2.95)**	**78.31(2.14)**

Table 14
Mean values and standard deviations of ACC, NMI, and Purity for clustering results with the proposed method and nine baseline methods on BBCSportG2 (%).

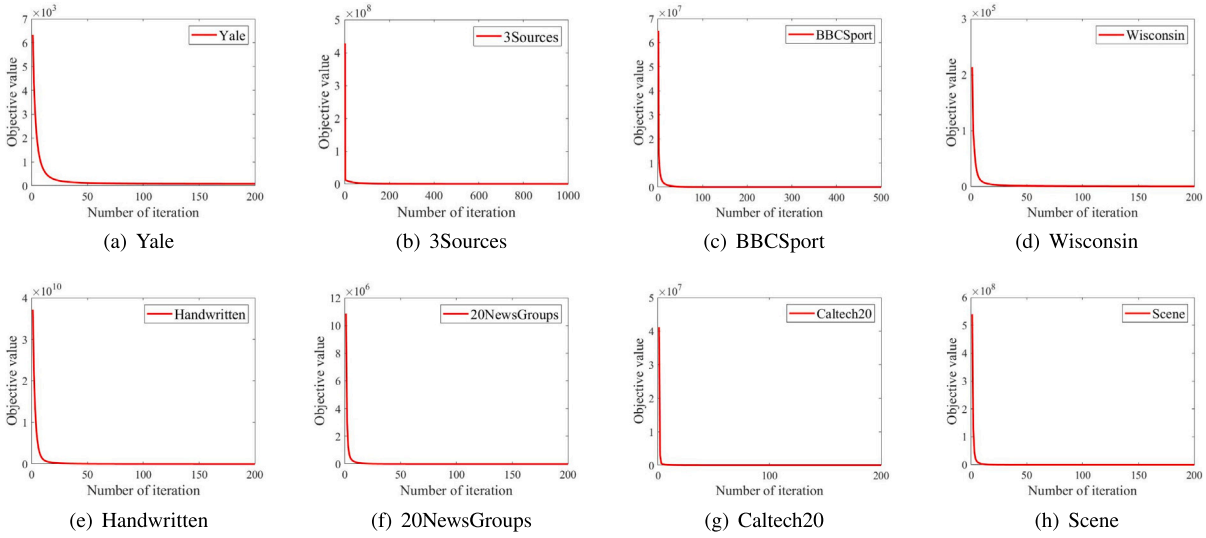
Method	ACC	NMI	Purity
NMF	66.78(5.81)	45.24(5.65)	70.53(4.55)
CoRegSPC	56.71(6.46)	42.66(5.22)	67.19(5.99)
GNMF	56.38(8.20)	42.48(4.93)	64.96(6.43)
MVKKM	36.03(0.00)	1.11(0.00)	36.58(0.00)
MultiNMF	64.19(1.26)	50.43(0.97)	68.84(1.27)
RMKMC	62.13(9.52)	43.61(11.24)	64.43(8.08)
DiNMF	66.47(8.46)	48.46(5.38)	69.23(8.30)
AMvDMD	54.61(9.03)	33.95(8.81)	59.26(7.32)
DMFPA	57.35(3.06)	36.32(1.30)	63.97(2.37)
Ours	**77.13(5.00)**	**62.84(2.65)**	**79.17(2.47)**

evaluate its learning effect. Therefore, the t-SNE algorithm is performed on $\boldsymbol{V}^{(m)}$ [55]. Based on data labels, clusters with different colors are generated.

Fig. 3 shows the comparative effect of $\boldsymbol{V}^{(m)}$ in the proposed method and DiNMF. 3Sources, BBCSport, Handwritten and 20NewsGroups are used in this experiment. The visualizations demonstrate that the coefficient matrix $\boldsymbol{V}^{(m)}$ of the proposed method acquires significant and clear clustering structures. Compared to DiNMF, the coefficient matrix of the proposed method is more centralized for categories and has less overlapping. Such results clearly show the usefulness of the learned $\boldsymbol{V}^{(m)}$ in clustering.

4.6. Convergence experiments

In Fig. 4, the evolution of the objective values during the iterations are plotted. It is evident that for the majority of datasets, the objective function values decrease dramatically within 100 iterations, then remain slight decreases after 200 iterations. However, for data with complex features, such as 3Sources, the model costs more iterations



(a) Yale (b) 3Sources (c) BBCSport (d) Wisconsin

(e) Handwritten (f) 20NewsGroups (g) Caltech20 (h) Scene

Fig. 4. Convergent results on eight datasets.

Table 15
Mean values and standard deviations of ACC, NMI, and Purity for clustering results with the proposed method and nine baseline methods on BBCSportG3 (%).

Method	ACC	NMI	Purity
NMF	68.44(6.23)	48.20(4.84)	71.31(5.26)
CoRegSPC	55.22(6.64)	40.20(7.76)	65.72(7.42)
GNMF	54.52(5.65)	44.92(3.51)	65.04(4.14)
MVKKM	36.03(0.00)	1.11(0.00)	36.58(0.00)
MultiNMF	63.81(5.89)	49.42(6.76)	68.93(4.65)
RMKMC	59.61(9.80)	39.32(8.60)	63.62(7.62)
DiNMF	64.08(7.49)	46.72(6.79)	67.21(7.15)
AMvDMD	51.51(7.77)	28.36(7.79)	55.02(7.61)
DMFPA	57.28(1.78)	39.53(0.54)	66.01(0.52)
Ours	**74.94(5.71)**	**59.20(4.37)**	**77.48(2.29)**

Table 16
Mean values and standard deviations of ACC, NMI, and Purity for clustering results with the proposed method and nine baseline methods on BBCSportP (%).

Method	ACC	NMI	Purity
NMF	67.32(6.09)	46.12(6.60)	69.69(4.98)
CoRegSPC	58.73(1.39)	44.60(1.13)	69.15(0.70)
GNMF	62.13(4.29)	42.58(3.92)	66.89(3.85)
MVKKM	38.24(0.00)	4.03(0.00)	38.42(0.00)
MultiNMF	65.75(0.17)	51.19(0.52)	69.85(0.17)
RMKMC	64.89(7.89)	49.63(9.08)	69.98(7.97)
DiNMF	66.87(6.57)	47.95(4.83)	69.01(5.29)
AMvDMD	53.92(5.82)	35.67(8.55)	59.85(5.38)
DMFPA	62.17(2.44)	40.67(0.93)	66.73(1.21)
Ours	**75.09(3.25)**	**57.69(3.38)**	**75.96(2.82)**

to find the solution. Therefore, it shows that the proposed method converges effectively on most datasets.

4.7. Parameter sensitivity

In practical applications, determining the optimal parameters for unsupervised learning methods can be challenging. Therefore, the insensitivity of the performance of unsupervised methods to parameter settings is crucial. The sensitivity to balancing parameters of the proposed method is demonstrated in this experiment. While preserving generality, the results for four datasets, including Yale, 3Sources, BBCSport, and Handwritten, are displayed. Similar patterns can be observed in other datasets as well.

The proposed model consists of three essential parameters, α, β and γ. The sensitivity analysis of a parameter is performed by first fixing the other two parameters and varying the value of the parameter within the range. Due to the transformation relationship of the balancing parameter between the formulation (14) and (15), the values of $\frac{\alpha}{2}$, β and γ are modified, and the range is $\{1, 10, 10^2, 10^3, 10^4, 10^5, 10^6, 10^7, 10^8\}$. Figs. 5–7 display the experimental results.

The outcomes indicate that the proposed method exhibits improvement when α is relatively larger. This indicates that a larger value of α means that the model can eliminate noises better. For β, the model achieves a higher performance over a wider range of parameter choices. The parameter γ obtains better performance at larger values. This indicates that graph structure has an important role in the clustering task. Similar patterns can be observed on other datasets, which suggests that larger values may be set for the balancing parameters in practical applications.

5. Conclusion

In this paper, a centric graph regularized log-norm sparse NMF is proposed for multi-view clustering. To reliably learn the underlying geometric structure which is embedded in multi-view data, the graph structure of the multi-view coefficient matrix is extracted by using centric graph regularization, which obtains more information about the spatial structure than using traditional graph regularization. Additionally, the pairwise co-regularization is used to reduce the redundant information within views. Finally, the log-based sparsification is introduced to improve the robustness of the model. An efficient update algorithm is derived for the formulated optimization problem, while its convergence and complexity are theoretically analyzed. Comparative experiments on eight real-world datasets with nine state-of-the-art methods demonstrate the effectiveness of the model.

CRediT authorship contribution statement

Yuzhu Dong: Software, Writing – original draft. **Hangjun Che:** Supervision, Methodology. **Man-Fai Leung:** Writing – review & editing, Investigation. **Cheng Liu:** Formal analysis, Investigation. **Zheng Yan:** Writing – review & editing.

Data availability

Data will be made available on request.

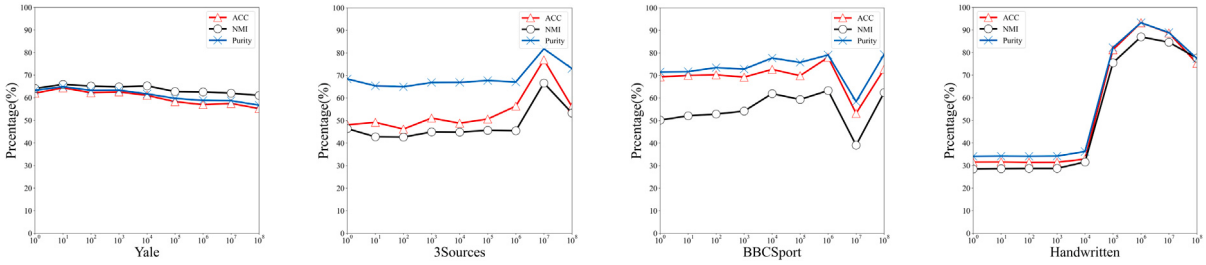


Fig. 5. The sensitivity analysis of $\frac{\alpha}{2}$ on four datasets.

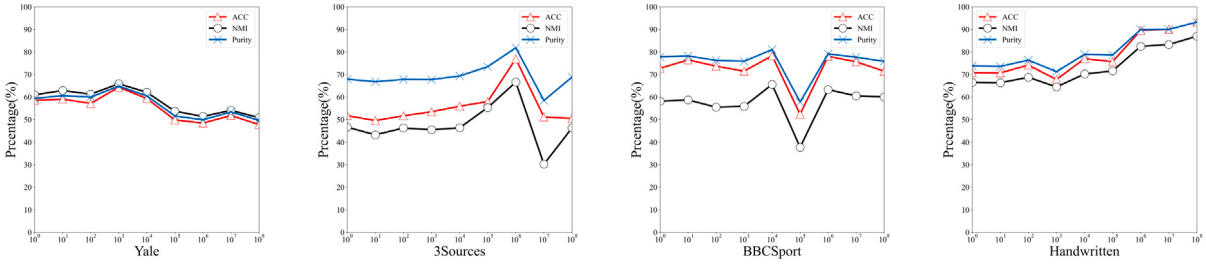


Fig. 6. The sensitivity analysis of β on four datasets.

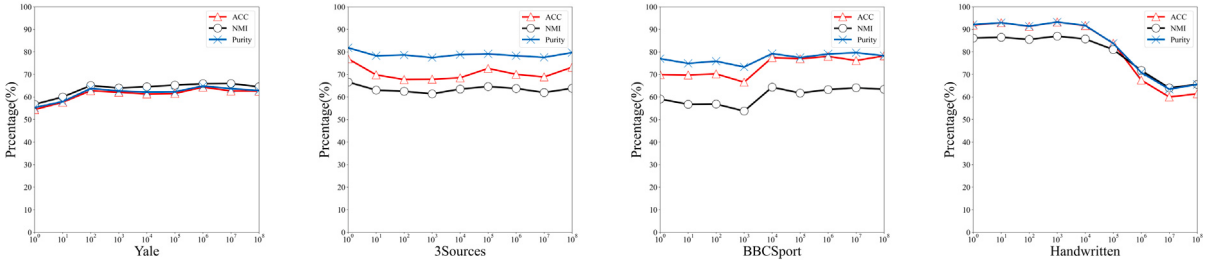


Fig. 7. The sensitivity analysis of γ on four datasets.

Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant No. 62003281), Natural Science Foundation of Chongqing, China (Grant No. cstc2021jcyj-msxmX1169), the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJQN202200207), and Open Fund of Key Laboratory of Cyber–Physical Fusion Intelligent Computing (South-Central Minzu University), State Ethnic Affairs Commission (Grant No. CPFIC202303).

References

[1] B. Pan, C. Li, H. Che, Nonconvex low-rank tensor approximation with graph and consistent regularizations for multi-view subspace learning, Neural Netw. 161 (2023) 638–658.

[2] Y. Kaloga, P. Borgnat, S.P. Chepuri, P. Abry, A. Habrard, Variational graph autoencoders for multiview canonical correlation analysis, Signal Process. 188 (2021) 108182.

[3] Z. Wang, Z. Shen, H. Zou, P. Zhong, Y. Chen, Retargeted multi-view classification via structured sparse learning, Signal Process. 197 (2022) 108538.

[4] T. Dokeroglu, A. Deniz, H.E. Kiziloz, A comprehensive survey on recent metaheuristics for feature selection, Neurocomputing 494 (2022) 269–296.

[5] X. Pu, H. Che, B. Pan, M.-F. Leung, S. Wen, Robust weighted low-rank tensor approximation for multi-view clustering with mixed noise, IEEE Trans. Comput. Soc. Syst. (2023) http://dx.doi.org/10.1109/TCSS.2023.3331366, early access.

[6] S. Ouadfel, M. Abd Elaziz, A multi-objective gradient optimizer approach-based weighted multi-view clustering, Eng. Appl. Artif. Intell. 106 (2021) 104480.

[7] S. Huang, Z. Kang, Z. Xu, Self-weighted multi-view clustering with soft capped norm, Knowl.-Based Syst. 158 (2018) 1–8.

[8] Z. Kang, Z. Guo, S. Huang, S. Wang, W. Chen, Y. Su, Z. Xu, Multiple partitions aligned clustering, in: Proceedings of the 28th International Joint Conference on Artificial Intelligence, IJCAI '19, AAAI Press, 2019, pp. 2701–2707.

[9] H. Che, B. Pan, M.-F. Leung, Y. Cao, Z. Yan, Tensor factorization with sparse and graph regularization for fake news detection on social networks, IEEE Trans. Comput. Soc. Syst. (2023) http://dx.doi.org/10.1109/TCSS.2023.3296479, early access.

[10] Y. Li, M. Yang, Z. Zhang, A survey of multi-view representation learning, IEEE Trans. Knowl. Data Eng. 31 (10) (2019) 1863–1883.

[11] A. Kumar, P. Rai, H. Daumé, Co-regularized multi-view spectral clustering, in: Proceedings of the 24th International Conference on Neural Information Processing Systems, NIPS '11, Curran Associates Inc., Red Hook, NY, USA, 2011, pp. 1413–1421.

[12] Y. Chen, X. Xiao, Y. Zhou, Jointly learning kernel representation tensor and affinity matrix for multi-view clustering, IEEE Trans. Multimed. 22 (8) (2019) 1985–1997.

[13] B. Huang, T. Xu, S. Jiang, Y. Chen, Y. Bai, Robust visual tracking via constrained multi-kernel correlation filters, IEEE Trans. Multimed. 22 (11) (2020) 2820–2832.

[14] H. Wang, Y. Wang, Z. Zhang, X. Fu, L. Zhuo, M. Xu, M. Wang, Kernelized multiview subspace analysis by self-weighted learning, IEEE Trans. Multimed. 23 (2020) 3828–3840.

[15] S. Wang, X. Liu, E. Zhu, C. Tang, J. Liu, J. Hu, J. Xia, J. Yin, Multi-view clustering via late fusion alignment maximization, in: Proceedings of the 28th International Joint Conference on Artificial Intelligence, IJCAI '19, AAAI Press, 2019, pp. 3778–3784.

[16] H. Zhang, D. Wu, F. Nie, R. Wang, X. Li, Multilevel projections with adaptive neighbor graph for unsupervised multi-view feature selection, Inf. Fusion 70 (2021) 129–140.

[17] M.-S. Chen, L. Huang, C.-D. Wang, D. Huang, J.-H. Lai, Relaxed multi-view clustering in latent embedding space, Inf. Fusion 68 (2021) 8–21.

[18] S. Zhou, E. Zhu, X. Liu, T. Zheng, Q. Liu, J. Xia, J. Yin, Subspace segmentation-based robust multiple kernel clustering, Inf. Fusion 53 (2020) 145–154.

[19] Y. Su, Z. Hong, X. Wu, C. Lu, Invertible linear transforms based adaptive multi-view subspace clustering, Signal Process. 209 (2023) 109014.

[20] S. Huang, Z. Kang, Z. Xu, Auto-weighted multi-view clustering via deep matrix decomposition, Pattern Recognit. 97 (2020) 107015.

[21] X. Yang, H. Che, M.-F. Leung, C. Liu, Adaptive graph nonnegative matrix factorization with the self-paced regularization, Appl. Intell. 53 (12) (2022) 15818–15835.

[22] B. Wu, E. Wang, Z. Zhu, W. Chen, P. Xiao, Manifold NMF with $l_{2,1}$-norm for clustering, Neurocomputing 273 (2018) 78–88.

[23] J. Liu, C. Wang, J. Gao, J. Han, Multi-view clustering via joint nonnegative matrix factorization, in: Proceedings of the 2013 SIAM International Conference on Data Mining, SIAM, 2013, pp. 252–260.

[24] H. Che, J. Wang, A nonnegative matrix factorization algorithm based on a discrete-time projection neural network, Neural Netw. 103 (C) (2018) 63–71.

[25] C. Zhang, S. Wang, J. Liu, S. Zhou, P. Zhang, X. Liu, E. Zhu, C. Zhang, Multi-view clustering via deep matrix factorization and partition alignment, in: Proceedings of the 29th ACM International Conference on Multimedia, MM '21, Association for Computing Machinery, New York, NY, USA, 2021, pp. 4156–4164.

[26] L. Zong, X. Zhang, L. Zhao, H. Yu, Q. Zhao, Multi-view clustering via multi-manifold regularized non-negative matrix factorization, Neural Netw. 88 (2017) 74–89.

[27] J. Wang, F. Tian, H. Yu, C.H. Liu, K. Zhan, X. Wang, Diverse non-negative matrix factorization for multiview data representation, IEEE Trans. Cybern. 48 (9) (2018) 2620–2632.

[28] N. Liang, Z. Yang, Z. Li, W. Sun, S. Xie, Multi-view clustering by non-negative matrix factorization with co-orthogonal constraints, Knowl.-Based Syst. 194 (2020) 105582.

[29] X. Zhu, J. Guo, W. Nejdl, X. Liao, S. Dietze, Multi-view image clustering based on sparse coding and manifold consensus, Neurocomputing 403 (2020) 53–62.

[30] J. Sun, Q. yu Zhang, Completion of multiview missing data based on multi-manifold regularised non-negative matrix factorisation, Artif. Intell. Rev. 53 (2020) 5411–5428.

[31] G.A. Khan, J. Hu, T. Li, B. Diallo, H. Wang, Multi-view data clustering via non-negative matrix factorization with manifold regularization, Int. J. Mach. Learn. Cybern. 13 (2022) 677–689.

[32] S. Wang, L. Chen, Y. Sun, F. Peng, J. Lu, Multiview nonnegative matrix factorization with dual HSIC constraints for clustering, Int. J. Mach. Learn. Cybern. 14 (6) (2023) 2007–2022.

[33] B. Pan, C. Li, H. Che, M.-F. Leung, K. Yu, Low-rank tensor regularized graph fuzzy learning for multi-view data processing, IEEE Trans. Consum. Electron. (2023) http://dx.doi.org/10.1109/TCE.2023.3301067, early access.

[34] X. Zhang, H. Gao, G. Li, J. Zhao, J. Huo, J. Yin, Y. Liu, L. Zheng, Multi-view clustering based on graph-regularized nonnegative matrix factorization for object recognition, Inform. Sci. 432 (2018) 463–478.

[35] J. Li, G. Zhou, Y. Qiu, Y. Wang, Y. Zhang, S. Xie, Deep graph regularized non-negative matrix factorization for multi-view clustering, Neurocomputing 390 (2020) 108–116.

[36] Y. Jia, H. Liu, J. Hou, S. Kwong, Semi-supervised adaptive symmetric non-negative matrix factorization, IEEE Trans. Cybern. 51 (5) (2021) 2550–2562.

[37] Y. Liang, D. Huang, C.-D. Wang, P.S. Yu, Multi-view graph learning by joint modeling of consistency and inconsistency, IEEE Trans. Neural Netw. Learn. Syst. (2022) http://dx.doi.org/10.1109/TNNLS.2022.3192445, early access.

[38] H. Yang, K. Ma, J. Cheng, Rethinking graph regularization for graph neural networks, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35, 2021, pp. 4573–4581.

[39] J. Ma, Y. Zhang, L. Zhang, Discriminative subspace matrix factorization for multiview data clustering, Pattern Recognit. 111 (2021) 107676.

[40] Z. Zhang, Z. Qin, P. Li, Q. Yang, J. Shao, Multi-view discriminative learning via joint non-negative matrix factorization, in: J. Pei, Y. Manolopoulos, S. Sadiq, J. Li (Eds.), Database Systems for Advanced Applications, Springer International Publishing, Cham, 2018, pp. 542–557.

[41] X. Liu, P. Song, C. Sheng, W. Zhang, Robust multi-view non-negative matrix factorization for clustering, Digit. Signal Process. 123 (2022) 103447.

[42] K. Liu, L. Brand, H. Wang, F. Nie, Learning robust distance metric with side information via ratio minimization of orthogonally constrained $l_{2,1}$-norm distances, in: Proceedings of the 28th International Joint Conference on Artificial Intelligence, IJCAI '19, AAAI Press, 2019, pp. 3008–3014.

[43] R. Li, X. Wang, W. Quan, Y. Song, L. Lei, Robust and structural sparsity auto-encoder with $l_{2,1}$-norm minimization, Neurocomputing 425 (2021) 71–81.

[44] C. Li, H. Che, M.-F. Leung, C. Liu, Z. Yan, Robust multi-view non-negative matrix factorization with adaptive graph and diversity constraints, Inform. Sci. 634 (2023) 587–607.

[45] C. Peng, Y. Zhang, Y. Chen, Z. Kang, C. Chen, Q. Cheng, Log-based sparse nonnegative matrix factorization for data representation, Knowl.-Based Syst. 251 (2022) 109127.

[46] D.D. Lee, H.S. Seung, Algorithms for non-negative matrix factorization, in: Proceedings of the 13th International Conference on Neural Information Processing Systems, NIPS '00, MIT Press, Cambridge, MA, USA, 2000, pp. 535–541.

[47] H. Mohanty, S. Champati, B.P. Barik, A. Panda, Cluster quality analysis based on SVD, PCA-based k-means and NMF techniques: an online survey data, Int. J. Reason. Intell. Syst. 15 (1) (2023) 86–96.

[48] K. Chen, H. Che, X. Li, M.-F. Leung, Graph non-negative matrix factorization with alternative smoothed L_0 regularizations, Neural Comput. Appl. 35 (14) (2022) 9995–10009.

[49] H. Che, J. Wang, A. Cichocki, Bicriteria sparse nonnegative matrix factorization via two-timescale duplex neurodynamic optimization, IEEE Trans. Neural Netw. Learn. Syst. 34 (8) (2023) 4881–4891.

[50] X. Wang, T. Zhang, X. Gao, Multiview clustering based on non-negative matrix factorization and pairwise measurements, IEEE Trans. Cybern. 49 (9) (2019) 3333–3346.

[51] W.-S. Chen, Q. Zeng, B. Pan, A survey of deep nonnegative matrix factorization, Neurocomputing 491 (2022) 305–320.

[52] D. Cai, X. He, J. Han, T.S. Huang, Graph regularized nonnegative matrix factorization for data representation, IEEE Trans. Pattern Anal. Mach. Intell. 33 (8) (2011) 1548–1560.

[53] G. Tzortzis, A. Likas, Kernel-based weighted multi-view clustering, in: 2012 IEEE 12th International Conference on Data Mining, 2012, pp. 675–684.

[54] X. Cai, F. Nie, H. Huang, Multi-view K-means clustering on big data, in: Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, IJCAI '13, AAAI Press, 2013, pp. 2598–2604.

[55] G. Zhong, C.-M. Pun, Improved normalized cut for multi-view clustering, IEEE Trans. Pattern Anal. Mach. Intell. 44 (12) (2022) 10244–10251.

[56] Y. Huang, Z. Shen, F. Cai, T. Li, F. Lv, Adaptive graph-based generalized regression model for unsupervised feature selection, Knowl.-Based Syst. 227 (2021) 107156.