AUT Journal of Mathematics and Computing

AUT J. Math. Comput., 5(4) (2024) 289-304
https://doi.org/10.22060/AJMC.2024.23353.1252

Original Article

Feature representation via graph-regularized entropy-weighted nonnegative matrix factorization

Hazhir Sohrabi[a], Shahrokh Esmaeili*[a], Parham Moradi[b,c]

[a] *Department of Applied Mathematics, University of Kurdistan, Sanandaj, Iran*
[b] *Department of Computer Engineering, University of Kurdistan, Sanandaj, Iran*
[c] *School of Engineering, RMIT University, Melbourne, Australia*

ABSTRACT: Feature extraction plays a crucial role in dimensionality reduction in machine learning applications. Nonnegative Matrix Factorization (NMF) has emerged as a powerful technique for dimensionality reduction; however, its equal treatment of all features may limit accuracy. To address this challenge, this paper introduces Graph-Regularized Entropy-Weighted Nonnegative Matrix Factorization (GEWNMF) for enhanced feature representation. The proposed method improves feature extraction through two key innovations: optimizable feature weights and graph regularization. GEWNMF uses optimizable weights to prioritize the extraction of crucial features that best describe the underlying data structure. These weights, determined using entropy measures, ensure a diverse selection of features, thereby enhancing the fidelity of the data representation. This adaptive weighting not only improves interpretability but also strengthens the model against noisy or outlier-prone datasets. Furthermore, GEWNMF integrates robust graph regularization techniques to preserve local data relationships. By constructing an adjacency graph that captures these relationships, the method enhances its ability to discern meaningful patterns amid noise and variability. This regularization not only stabilizes the method but also ensures that nearby data points appropriately influence feature extraction. Thus, GEWNMF produces representations that capture both global trends and local nuances, making it applicable across various domains. Extensive experiments on four widely used datasets validate the efficacy of GEWNMF compared to existing methods, demonstrating its superior performance in capturing meaningful data patterns and enhancing interpretability.

Review History:
Received:17 July 2024
Revised:03 Septemberr 2024
Accepted:09 Septemberr 2024
Available Online:01 October 2024

Keywords:
Feature extraction
Subspace learning
Weighted NMF
Entropy regularizer

MSC (2020):
15A23; 68T05; 68T30

1. Introduction

The advancement of technology and the proliferation of social networks have led to the generation of high-dimensional data. Often, this data contains extraneous information or noise, which can hinder data mining and analysis techniques. Dimensionality reduction involves decreasing the number of variables under consideration, either by identifying a set of principal variables or by transforming the original variables into a new set. Among the

*Corresponding author.
E-mail addresses: h.sohrabi@uok.ac.ir, sh.esmaeili@uok.ac.ir, p.moradi@uok.ac.ir, parham.moradi@rmit.edu.au

methods for dimensionality reduction, matrix factorization has garnered considerable attention due to its strong mathematical foundations and interpretability.

NMF is a type of matrix factorization that restricts its components to non-negative elements. It approximates a non-negative data matrix by decomposing it into the product of two matrices. For instance, when applied to images of people's faces, NMF can identify key facial features such as eyes, lips, nose, and eyebrows [5]. Consequently, this method has found extensive use in various domains, including image analysis, pattern recognition, medical applications, and text classification.

NMF and its variants perform well in numerous fields and offer several advantages: 1) the implementation process of NMF is simple; 2) the factorization results of NMF are interpretable; and 3) it reduces storage space requirements. However, they also have some limitations. First, NMF-based methods often overlook the local geometric structure of the data, leading to feature representations that lack sufficient discriminative information. Second, many rely on Euclidean distance in the objective function, which is not robust to noise and outliers, resulting in poor generalization and less effective feature representation.

To address the aforementioned deficiencies, numerous enhancements to NMF have been proposed. Ding et al. [8] first presented the Semi-NMF method, in which the data and basis matrices could have negative entries. They then introduced the convex method [13, 26], in which the basis matrix is a convex combination of the data matrix. Recognizing that orthogonality constraints lead to sparsity, Ding et al. added orthogonality constraints to the base and representation matrices and proposed an orthogonal NMF (ONMF) method [9]. To enhance NMF's robustness to noise, Gao et al. [11] suggested using the capped norm for the objective function and applying a threshold to remove outliers. Similarly, Guan et al. [12] proposed the truncated Cauchy NMF method, which manages outliers by truncating large errors, thereby learning subspaces on datasets affected by significant noise or corruption.

Ma et al. [23] introduced layer-specific NMF (LSNMF) for multi-layer networks, implementing an orthogonality constraint on certain components to ensure feature vertex specificity. Dong et al. [10] proposed a method that sparsifies with the $L_{2,\log}$-(pseudo) norm to efficiently generate sparse solutions, obtain a better part-based representation, and enhance its robustness and applicability in complex noisy scenarios. Wei et al. [31] developed Entropy-Weighted Nonnegative Matrix Factorization (EWNMF) to reduce the impact of noise and outliers, using a weighting mechanism via the objective function of each sample.

Cai et al. [4] created a graph regularization constraint to learn the local structures of data and applied manifold learning, leading to the development of Graph Regularization NMF (GNMF). GNMF includes various graph construction methods, such as dot-product weighting, heat kernel weighting, and binary weighting. Huang et al. [16] reported $L_{2,1}$-NMF with Adaptive Neighbors ($L_{2,1}$-NMFAN) by integrating $L_{2,1}$-NMF and an adaptive local structure learning strategy [24]. This method can learn the similarity between data while decomposing the matrix, thereby mining a high-quality spatial network.

Deng et al. [7] used the L_1-norm and developed a graph-regularized sparse NMF (GSNMF) by integrating this norm with graph regularization. Additionally, Deng et al. [6] created tri-regularized NMTF (TRNMTF) by combining the Frobenius norm, graph regularization, and L_1-norm into a single framework. Hamza and Brady [14] proposed the $L_1 + L_2$ function to construct the cost function of NMF, but the solving algorithm is very time-consuming.

Huang et al. [15] introduced robust graph-regularized NMF (RGNMF) to mitigate the impact of noise and outliers. This method manages corrupted data by incorporating an error matrix and combining robust graph regularization with the L_1-norm. Robust structure NMF (RSNMF) was proposed by Huang et al. [17] to include both global and local structure learning, applying the $L_{2,1}$-norm to the basis matrix.

The aforementioned methods typically outperform standard NMF, but their effectiveness is sensitive to the regularization parameters and the quality of the adjacency graph. To address this issue, Ahmed et al. [1] proposed neighborhood structure-assisted NMF (NSNMF), which uses the minimum spanning tree (MST) to construct a sparse similarity matrix.

Recently, to obtain more discriminative feature representations, semi-supervised methods have been explored. For example, Wu et al. [32] developed positive and negative label-driven NMF (PNLD-NMF) by incorporating both positive and negative labels. Jia et al. [18] constructed a similarity matrix and a dissimilarity matrix based on labeled and unlabeled samples and proposed semi-supervised NMF (SNMF), which forms a pair of complementary regularization terms integrated into NMF. Additionally, Wu et al. [33] proposed dual embedding regularized NMF (DENMF), which jointly learns low-dimensional representations and an assignment matrix.

Most studies assign equal importance to all features of the data points. However, some researchers have recognized the varying significance of different features. Blondel et al. [2] introduced predetermined weights for each attribute of every data point, demonstrating that these weights can significantly enhance flexibility by emphasizing certain features in image approximation problems. Building on this, Kim and Choi [19] proposed a new Weighted Nonnegative Matrix Factorization (WNMF) method to handle incomplete data matrices with missing entries, incorporating binary weights into the NMF multiplication update. Wang et al. [29] developed a WNMF method that

assigns the weight of a data point as the product of column and row weights. These methods are referred to as "hard WNMF"; however, a major limitation is the need for predetermined weights.

Chen et al. [5] proposed a novel NMF that incorporates a transformation matrix and a graph regularization term to project samples into a subspace and learn the data manifold. They employed the constrained Laplacian rank algorithm to address these issues. The cost function includes two weighted matrices, and a complex algorithm is used to determine the weights, resulting in high algorithmic complexity.

This article introduces the Graph Regularized Entropy Weighted NMF (GEWNMF) method. This approach not only preserves the global structure through NMF but also utilizes optimizable weights to determine the importance of the extracted features, leading to the extraction of more significant features from the data space. Entropy is employed to enhance the likelihood of various features being represented in the new data space. To maintain the local structure of the data, a robust graph is used, improving the method's resilience to noise and extreme outliers. The main contributions of this paper are highlighted as follows:

- The proposed GEWNMF integrates dimensionality reduction, Laplacian regularization, and local structure preservation into a unified model that simultaneously learns both feature representation and graph structure. This approach addresses the limitations associated with pre-defined graph structures in feature extraction.
- A novel weighted nonnegative matrix factorization (WNMF) method, incorporating an entropy term, is used to enhance the representation of diverse features within the new data space.
- An adjacency graph with an optimal number of neighbors, designed to be robust against noise, is employed to effectively preserve the local structure.
- An iterative updating algorithm has been developed to optimize GEWNMF. The convergence of this algorithm is confirmed through both theoretical analysis and numerical experiments. The effectiveness of the proposed GEWNMF has been validated on four well-known datasets.

The remainder of this paper is organized as follows. Section 2 provides an overview of related work. Section 3 introduces the proposed GEWNMF framework, discusses the optimization algorithm, and details the algorithm updates. Section 4 assesses the performance of GEWNMF through extensive experimental results. Finally, Section 5 presents conclusions and suggests directions for future research.

2. Related works

In this section, research related to our proposed algorithm is introduced. Before discussing the formulation of various methods, the symbols used in this article are presented.

2.1. Notations

For an arbitrary matrix $A \in \mathbb{R}^{m\times n}$, A_{ij} denotes the i,j entry of A, while $A_{i:}$ and $A_{:j}$ represent the ith row and jth column of A, respectively. Let the sum of the entries of A be denoted by $\|A\|_\diamond = \sum_{i,j} A_{ij}$. The Hadamard product $A \odot B$ yields a matrix with entries defined as $(A \odot B)_{ij} = A_{ij}B_{ij}$. The trace of a square matrix A is represented by $\mathrm{Tr}(A)$ and I_k is the $k \times k$ matrix known as the identity. The Euclidean norm of $\boldsymbol{x} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$ is defined by $\|\boldsymbol{x}\| = (\sum_{i=1}^n x_i^2)^{1/2}$, and this notation is also applicable to row vectors. The vector $\mathbf{1}$ is a column vector with all entries equal to 1. The Frobenius norm of A is defined by $\|A\|_{\mathrm{F}} = (\sum_{i=1}^n \|A_{i:}\|^2)^{1/2}$, with $\|A\|_{\mathrm{F}}^2 = \mathrm{Tr}(AA^\top)$. We employ the notation $\mathbb{R}_+^{m\times n}$ to denote the space of $m \times n$ nonnegative matrices. Given a data set $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m\}$, the notation $X = [\boldsymbol{x}_1; \boldsymbol{x}_2; \ldots; \boldsymbol{x}_m] = [\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_n] \in \mathbb{R}_+^{m\times n}$ represents the data matrix, n signifies the number of samples and m denotes the number of features.

2.2. Nonnegative matrix factorization

Nonnegative Matrix Factorization (NMF) is a method that captures low-dimensional representations and reveals hidden structures and patterns in data. As a representative dimensionality reduction approach, NMF effectively addresses the curse of dimensionality problem [30]. First, we review the standard NMF methodology. The NMF expression is given by:

$$X \approx UV,$$

where $X \in \mathbb{R}_+^{m\times n}$ is a data matrix, with columns representing samples and rows representing features. The goal of NMF is to find two matrices, $U \in \mathbb{R}_+^{m\times k}$ and $V \in \mathbb{R}_+^{k\times n}$, called the basis matrix and the representation matrix, respectively, such that they can approximate X well, where $k \ll \min\{m, n\}$.

Frobenius norm and Kullback–Leibler divergence are often utilized to solve low-rank matrices [20]. The optimization problem based on Frobenius norm is:

$$\min F_1(U,V) \equiv \|X - UV\|_F^2 \quad \text{s.t.} \quad U \geq 0,\ V \geq 0. \tag{1}$$

The multiplicative iterative algorithm is widely applied to solve (1), and the variables U and V can be updated using the following rules:

$$U \leftarrow U \odot \frac{XV^\top}{UVV^\top}, \qquad V \leftarrow V \odot \frac{U^\top X}{U^\top UV}, \tag{2}$$

where the division is performed in an element-wise manner as well [29].

2.3. Weighted nonnegative matrix factorization

Nonnegative Matrix Factorization places equal importance on all features of data points. Therefore, some researchers have addressed the varying importance of these features. This led to the introduction of Weighted Nonnegative Matrix Factorization (WNMF), first proposed in [25] for the weighted Euclidean distance. Several algorithms, including Newton-related methods, have been used to solve this problem, but they tend to have high complexity. Simpler algorithms were introduced with the help of Lee and Seung's algorithm [20]. The WNMF problem, aiming to minimize the objective function:

$$F_2(U,V) = \|X - UV\|_W^2 = \sum_{i,j} [W \odot (X - UV) \odot (X - UV)]_{ij},$$

where $W \in \mathbb{R}_+^{m\times n}$ is a given weight matrix. The update rules for U and V are as follows [31]:

$$U \leftarrow U \odot \frac{(W \odot X)V^\top}{(W \odot [UV])V^\top}, \qquad V \leftarrow V \odot \frac{U^\top (W \odot X)}{U^\top (W \odot [UV])}.$$

The use of WNMF emphasizes learning specific parts of the data space.

2.4. Graph-regularized nonnegative matrix factorization

NMF aims to find two nonnegative matrices whose product provides a good approximation to the original matrix, and the NMF algorithm is proposed to learn the parts of objects, such as human faces and text documents [20]. However, it fails to consider the geometrical structure of the data space, which is essential for data clustering and classification problems. To overcome this limitation of NMF, an algorithm called Graph Regularized Nonnegative Matrix Factorization (GNMF) was proposed [4]. GNMF encodes the geometrical information of the data space by constructing a nearest neighbor graph. The goal is to ensure that if two samples are connected in the neighbor graph, they are close to each other in the new representation. The objective function of GNMF is as follows:

$$\|X - UV\|_F^2 + \alpha \mathrm{Tr}(VLV^\top).$$

Here, the matrix L is called the graph Laplacian and defined as $L = D - G$, where D is a diagonal matrix with entries that are the column or row sums of G, specifically $D_{ii} = \sum_j G_{ij}$. The matrix G is a weight matrix and can be defined in different ways. One common method, which has been used in most articles, is as follows:

$$G_{ij} = \begin{cases} 1 & \boldsymbol{x}_i \in \mathcal{N}_p(\boldsymbol{x}_j) \text{ or } \boldsymbol{x}_j \in \mathcal{N}_p(\boldsymbol{x}_i) \\ 0 & \text{otherwise} \end{cases}$$

where $\mathcal{N}_p(\boldsymbol{x}_i)$ denotes the set of p-nearest neighbors of $\boldsymbol{x}_i$. This method of constructing a neighborhood graph is called binary weighting. In addition, there are other methods for making graphs, such as heat kernel weighting and dot product weighting [21].

2.5. Methods for constructing neighborhood graphs

In this section, alternative methods for constructing the neighborhood graph are presented. These methods not only improve the learning of local structures but are also resistant to noise. Various exceptional adaptive strategies for learning local structures have been proposed to enhance clustering performance in resulting methods [24].

Lotfi et al. [22] utilize a fuzzy neighborhood measure to calculate local density values. Their approach involves clustering data through three primary steps: identifying cluster centers, constructing backbones, and assigning labels. Zhang et al. [35] anticipated that enhancing the detail and comprehensiveness of the graph could improve

the performance of NMF. To address this, they introduced a graph regularizer based on a linear projection of the rating matrix, resulting in a method they named Linear Projection and Graph Regularized Nonnegative Matrix Factorization (LPGNMF).

By defining the adaptive deep graph convolution method, Wu et al. [34] tried to maintain the local structure of space with the help of neural networks. They introduce an adaptive deep graph convolution method that incorporates the adaptive aggregation of local high-order neighborhoods, as opposed to the traditional approach of stacking single-order convolutional layers within the message-passing framework.

In recent years, methods such as LLSRFS [28] and MVSC [36], each with distinct characteristics, have attempted to learn local graph structures. Most of these approaches typically rely on information from a single data point. This makes it straightforward to form an inaccurate local structure graph in a data setting that includes outliers or noise. In fact, each individual data point should not possess a consistent number of neighbors [37]. The closer a data point is to a cluster, the higher the number of neighbors it has.

An advanced and intuitive method for constructing a neighborhood graph was proposed in [27], in which each point can have a unique number of neighbors according to its specific geometrical position. It preserves local structure and noise resistance with much lower computational cost than neural networks and with optimal accuracy.

The important issue is that, by setting different thresholds for the neighborhood, adaptive local structure learning strategies with varying robustness can be obtained.

3. Proposed framework

This section introduces the proposed method and its associated problem formulation, providing a detailed exposition of the methodological approach and the problem statement to facilitate comprehensive understanding.

3.1. Entropy-weighted nonnegative matrix factorization

The main goal of this research is to reduce the dimensions of the data while maintaining its global structure, which can be achieved using NMF. To differentiate between the values of different data features, we employ the WNMF algorithm, which gives rise to the following optimization problem:

$$\begin{aligned} &\min \ \|X - UV\|_W^2 \\ &\text{s.t.} \ \ U \geq 0, V \geq 0, W \geq 0, \ \sum_{i=1}^{m} W_{ij} = 1. \end{aligned} \tag{3}$$

For fixed matrices U and V, determining W_{ij} is straightforward: Define $E = X - UV$, then $W_{ij} = 1$ if $|E_{ij}| = \min\{|E_{1j}|, \ldots, |E_{mj}|\}$, otherwise $W_{ij} = 0$. This approach highlights that each column of W has only one element as 1 and the rest as 0, which simplifies the problem but may not fully address the real-world problem. To mitigate this issue, we introduce an entropy regularizer in the cost function, aiming to constrain weights to the range $[0, 1]$ rather than strictly 0 or 1. This approach utilizes information entropy to quantify the uncertainty associated with weights. The updated optimization problem is formulated as follows:

$$\begin{aligned} &\min \ \ \|X - UV\|_W^2 + \alpha \|W \odot \mathcal{LW}\|_\diamond \\ &\text{s.t.} \ \ U \geq 0, V \geq 0, W \geq 0, \ \sum_{i=1}^{m} W_{ij} = 1, \end{aligned} \tag{4}$$

where $\mathcal{LW} \in \mathbb{R}^{m \times n}$ is a matrix defined such that $[\mathcal{LW}]_{ij} = \ln(W_{ij})$, and $\alpha \geq 0$ is a specified hyperparameter. The first term in (4), similar to (3), represents the sum of weighted errors, while the second term corresponds to the negative entropy of the weights. Entropy, in theory, measures the level of disorder or randomness in the system. Minimizing the second term in the objective function suggests maximizing the dimensionality of feature extraction. The original objective function in (3) limits each data point to a single attribute for feature representation, whereas the entropy regularizer encourages the inclusion of more attributes to enhance feature representation.

All available methods must be employed to craft an appropriate low-dimensional data representation. Utilizing information about the local data structure, facilitated by the GNMF algorithm, is one such method. Integrating the Laplacian term into GNMF Algorithm enables leveraging inherent geometric relationships between data points, thereby enhancing the quality of representation. The resulting optimization problem is formulated as follows:

$$\begin{aligned} &\min \ \ \|X - UV\|_W^2 + \alpha \|W \odot \mathcal{LW}\|_\diamond + \beta \mathrm{Tr}(VLV^\top) \\ &\text{s.t.} \ \ U \geq 0, V \geq 0, W \geq 0, \ \sum_{i=1}^{m} W_{ij} = 1, \end{aligned} \tag{5}$$

where, as mentioned in Section 2.4, L is the graph Laplacian matrix.

3.2. Local structure graph

One innovation of this research is the utilization of the optimal neighborhood graph. As discussed in Section 2.5, when employing the GNMF algorithm, instead of the conventional neighborhood graph, we utilize a graph with optimal neighbors, resulting in improved noise robustness and structure preservation. One approach to constructing such a neighborhood graph involves using a global neighborhood threshold determined by the geometry of the data space, which is employed for neighbor detection.

In accordance with the findings presented in [27], we construct the optimal local structure graph $S \in \mathbb{R}_+^{n\times n}$. In particular, we initially establish $d_{ij} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2$, and arrange the set $\{d_{i1}, \ldots, d_{in}\}$ in ascending order. Secondly, unlike the conventional local structure learning approach that assigns an individual threshold for each node, we introduce a global threshold $\varepsilon_p = \frac{1}{n}\sum_{i=1}^{n} d_{i,p+2}$, where p is a parameter that controls the sparsity of S. Subsequently, we specify that only the pairs of nodes $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ for which the resulting value d_{ij} is less than the threshold ε_p are eligible to become neighbors, i.e., $S_{ij} = 0$ if $d_{ij} \geq \varepsilon_p$, and $S_{ij} > 0$ otherwise. Here, the matrix H is defined as $H_{ij} = d_{ij} - \varepsilon_p$. Then, the following strategy for learning an optimal local structure graph is introduced:

$$\min_{S\geq 0} \|H \odot S\|_\diamond + \eta\|S\|_F^2, \tag{6}$$

where the second term acts as a regularization, with $\eta \geq 0$ serving as a flexible parameter to control the density of the optimal graph S. Some implicit conditions, such as $\sum_{j=1}^{n} S_{ij} = 1$ and $S_{ij} \geq 0$, are implied by this problem. Based on these constraints, we partition problem (6) into n subproblems, each corresponding to an individual sample. Subsequently, to address each of these subproblems, we construct the Lagrangian function as follows:

$$\mathcal{L}_i(S_{i:}, \tau, \boldsymbol{\beta}_i) = \frac{1}{2}\left\| S_{i:} + \frac{1}{2\eta_i} H_{i:} \right\|^2 - \tau(S_{i:}\mathbf{1} - 1) - S_{i:}\boldsymbol{\beta}_i,$$

where $\tau \geq 0$ and $\boldsymbol{\beta}_i \geq 0$ are the Lagrangian multipliers. Applying the Karush-Kuhn-Tucker (KKT) conditions, the solution to the corresponding subproblem can be determined as $S_{ij} = \max(-H_{ij}/(2\eta_i) + \tau, 0)$. As indicated in [24], if the optimal $S_{i:}$ consist of only p nonzero elements, the multiplier τ and the parameter η_i can be derived. For computational convenience, the overall parameter η can be set as the mean of $\eta_1, \eta_2, \ldots, \eta_n$, denoted as:

$$\eta = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{p}{2}d_{i,p+2} - \frac{1}{2}\sum_{j=2}^{p+1} d_{ij}\right) = \frac{p}{2}\varepsilon_p - \frac{1}{2n}\sum_{i=1}^{n}\sum_{j=2}^{p+1} d_{ij}.$$

Certainly, this adjustment to the optimal graph S can also be achieved by assigning probabilistic interpretations, such that $\frac{1}{n}\sum_{i=1}^{n}\sum_{j=2}^{p+1} S_{ij} \approx 1$. Ultimately, the solution to optimization problem (6) is attained through

$$S_{ij} = \max\left(\frac{\varepsilon_p - d_{ij}}{2\eta}, 0\right). \tag{7}$$

The neighborhood matrix S, formed in this manner, lacks symmetry. Therefore, we employ the matrix G, obtained by $G = (S + S^\top)/2$, and then construct the Laplacian matrix as $L_S = D - G$.

As stated in [27], the optimal graph strategy (6) exhibits notable robustness against outliers. Testing this concept involves evaluating the effectiveness of the suggested robust optimal local structure learning approach (6) using a synthetic dataset with three data clusters distributed on separate arcs, including an outlier. The k-NN strategy is designated as the control group. Figure 1 presents specific results, indicating that only equation (6) accurately constructs the spatial network of non-outliers. Conversely, k-NN fails to correctly identify the spatial structure under the influence of an outlier.

3.3. Objective function

In this manner, in addition to the objective function of the optimization problem (5), the objective function derived for the introduced method, utilizing the proposed graph, is expressed as follows:

$$\mathscr{F}(U, V, W, S) = \frac{1}{2}\|X - UV\|_W^2 + \alpha\|W \odot \mathcal{W}\|_\diamond + \frac{\beta}{2}\mathrm{Tr}(VL_SV^\top) + \gamma\left(\|H \odot S\|_\diamond + \eta\|S\|_F^2\right). \tag{8}$$

In this context, positive balancing parameters, denoted as α, β, and γ, play a pivotal role as weighting factors within the objective function. These parameters intricately influence the trade-offs between various components of the optimization problem, allowing for fine-tuning and customization of the model's behavior. The derived objective function, along with the associated constraints, constitutes the following optimization problem:

$$\min\ \mathscr{F}(U, V, W, S) \qquad \text{s.t. } U \geq 0, V \geq 0, W \geq 0, \sum_{i=1}^{m} W_{ij} = 1. \tag{9}$$

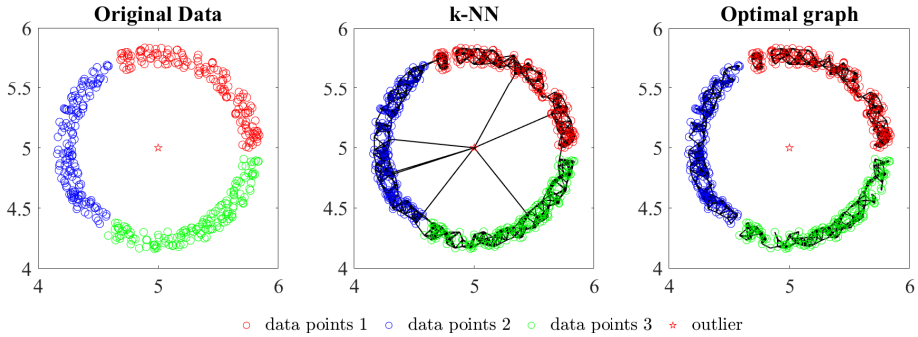

Figure 1: Comparison of robustness in optimal local structure learning strategies

Upon solving the optimization problem, we obtain the matrices U and V. Subsequently, using the relation $X_{\text{new}} = UV$, we derive a new representation of the data matrix, which is more compact and contains more useful information.

3.4. Optimization algorithm

In optimization problems like (9), where the objective function involves multiple variables, simultaneous optimization of all variables is necessary but can be computationally impractical. In such cases, an effective method is to optimize one variable at a time while keeping the others fixed, repeating this process until convergence is reached and no further improvement in the objective function is possible.

The variable S, which appears in the third and fourth terms of the objective function, is the first variable we will discuss. Initially, we determine the parameter p to construct the S matrix using (7). Consequently, the variable S will be treated as constant in the subsequent discussion, and we will temporarily omit the fourth term in the objective function. With the matrix S fixed, attention now shifts to updating other model variables within the objective function (8). As is customary, we commence the procedure by constructing the Lagrangian corresponding to the optimization problem (9), introducing the Lagrange multipliers Φ, Ψ, Ω, and Λ as follows:

$$\mathscr{L}(U,V,W,\Phi,\Psi,\Omega,\Lambda) = \mathscr{F}(U,V,W,S) + \mathrm{Tr}(\Phi U^\top) + \mathrm{Tr}(\Psi V^\top) + \mathrm{Tr}(\Omega W^\top) + \mathrm{Tr}(\Lambda \odot Z^\top).$$

Here, Λ and Z are same order diagonal matrices with diagonal entries $\Lambda_{jj} = \lambda_j$ and $Z_{jj} = \sum_{i=1}^{m} W_{ij} - 1$. Critical points of the Lagrangian $\mathscr{L}$ correspond to critical points of the optimization problem (9). Given that the objective and constraint functions have continuous first partial derivatives in (9), we apply the gradient form of the KKT Theorem as outlined below:

$$\begin{aligned} &\nabla\mathscr{F}(U,V,W,S) + \nabla\mathrm{Tr}(\Phi U^\top) + \nabla\mathrm{Tr}(\Psi V^\top) + \nabla\mathrm{Tr}(\Omega W^\top) + \nabla\mathrm{Tr}(\Lambda \odot Z^\top) = 0,\\ &\Phi \odot U = 0, \quad \Psi \odot V = 0, \quad \Omega \odot W = 0, \quad \Lambda \odot Z = 0, \quad \Phi,\Psi,\Omega,\Lambda \geq 0. \end{aligned} \tag{10}$$

By the KKT conditions (10), a minimizer (U,V,W) must satisfy

$$\begin{aligned} &W \odot (X - UV)(0 - V^\top) = \Phi,\\ &(0 - U^\top)W \odot (X - UV) + \beta V L_S = \Psi,\\ &\nabla_W\mathscr{F}(U,V,W,S) + \nabla_W\mathrm{Tr}(\Omega W^\top) + \nabla_W\mathrm{Tr}(\Lambda \odot Z^\top) = 0,\\ &\Phi \odot U = 0, \quad \Psi \odot V = 0, \quad \Omega \odot W = 0, \quad \Lambda \odot Z = 0, \quad \Phi,\Psi,\Omega,\Lambda \geq 0. \end{aligned} \tag{11}$$

By using the first and second equations in (11) and noting that $L_S = D - G$, the updating rules for U and V are derived as follows:

$$U \leftarrow U \odot \frac{(W \odot X)V^\top}{(W \odot [UV])V^\top)}, \tag{12}$$

$$V \leftarrow V \odot \frac{U^\top(W \odot X) + \beta VG}{U^\top(W \odot UV) + \beta VD}. \tag{13}$$

Given the matrices U and V, W in the third equation in (11) is a minimizer when

$$\sum_{i=1}^{m} W_{ij} - 1 = 0, \quad [(X - UV) \odot (X - UV)]_{ij} + \alpha \ln W_{ij} + \alpha = \lambda_j.$$

By solving these two equations for W and eliminating the multiplier λ_j, the result is as follows:

$$W_{ij} = \frac{\exp\left(-\frac{1}{\alpha}(X_{ij} - [UV]_{ij})^2\right)}{\sum_{\ell=1}^{m} \exp\left(-\frac{1}{\alpha}(X_{\ell j} - [UV]_{\ell j})^2\right)}. \tag{14}$$

The detailed iterative updating procedures are provided in Algorithm 1, which outlines the step-by-step process for updating the variables iteratively until convergence is achieved.

Algorithm 1 The training process GEWNMF.

Input: Data matrix $X \in \mathbb{R}_+^{m\times n}$; the number of reduced dimensions k; the parameters α, β and γ; the domain parameter p and the maximum number of iterations `maxIter`.

Output: the weight matrix W , the base matrix U, and the representation matrix V.

Creating neighborhood matrix $S \in \mathbb{R}^{n\times n}$ by (7) and affinity matrix G by $G = (S + S^\top)/2$.
Initialize $U \in \mathbb{R}_+^{m\times k}$ and $V \in \mathbb{R}_+^{k\times n}$.
while iteration $\leq$ `maxIter` **do**
    Update W by (14);
    Update U by (12);
    Update V by (13);
end while
return W, U, and V.

3.5. Convergence analysis and computational cost

The optimization of the objective function (8) involves three variables: U, V, and W. Therefore, we need to prove that the objective function (8) is convergent under the updating rules (12), (13), and (14), respectively. The discussion of the non-increasing behavior of the objective function for methods like GEWNMF is often based on the well-known work of Lee and Seung [20]. Specifically, the updating rules result in the following inequalities at each iteration:

$$\mathscr{F}(U^{t+1}, V^{t+1}, W^{t+1}, S) \leq \mathscr{F}(U^{t+1}, V^{t}, W^{t+1}, S) \leq \mathscr{F}(U^{t}, V^{t}, W^{t+1}, S) \leq \mathscr{F}(U^{t}, V^{t}, W^{t}, S).$$

However, since the updating rules in the GEWNMF algorithm are similar to those in the articles [31, 4], the theoretical review of the algorithm's convergence using the method of auxiliary functions is referenced in these two articles. The practical review of the convergence will be conducted at the end of the next section, i.e., section 4.5.

The time complexity of Algorithm 1 at each iteration includes the following components: the complexity of updating the variable W, governed by (14), is $\mathcal{O}(kmn)$; the complexity of updating the variable U, governed by (12), is $\mathcal{O}(knm)$; and the complexity of updating the variable V, as determined by (13), is $\mathcal{O}(kn^2 + kmn)$. Because the reduced rank k is significantly smaller than m and n, the overall time complexity of the proposed GEWNMF algorithm is $\mathcal{O}(n^2 + mn)$.

4. Experiments

In this section, we conduct experiments to validate the effectiveness of GEWNMF on public datasets. We focus on the following analyses: (1) comparisons with existing NMF methods; (2) ablation study; (3) convergence analysis; (4) parameter sensitivity analysis; and (5) clustering performance versus k (the number of dimensions of the representation space) on datasets. We use the k-means clustering algorithm to evaluate unsupervised feature extraction methods, setting the parameter k to correspond to the number of reduced ranks. To minimize the effect of initialization for factor matrices, we repeat each experiment 20 times and report the average results.

Experiments were conducted on four datasets. These include face image datasets (ORL and Yale), biological datasets (LUNG and lung-discrete), and spoken letter recognition data (Isolet). The details of these datasets are summarized in Table 1.

4.1. Comparison methods

We compare GEWNMF with eight established NMF methods: NMF [20], GNMF [4], ONMF [9], Semi-NMF [8], Convex-NMF [8], RNMF [11], Cauchy-NMF [12], and EWNMF [31]. To ensure fair performance comparisons, we tested a wide range of hyperparameter values from 2^{-8} to 2^8 for each method and reported the best results.

Table 1: Detailed information about the datasets

Dataset	instances	features	classes	Type of Data
Isolet	1560	617	26	Letter image
ORL	400	1024	40	Face image
Yale	165	1024	15	Face image
Lung-discrete	73	325	7	Biological microarray

Table 2: ACC of unsupervised feature extraction techniques

Dataset	Isolet	ORL	Yale	Lung-discrete
NMF	61.10	57.31	40.12	78.84
GNMF	60.78	57.10	39.94	77.80
ONMF	60.76	53.73	39.64	77.12
Semi - NMF	60.42	57.62	40.00	79.93
Convex - NMF	40.84	35.78	36.91	78.08
RNMF	46.64	44.15	34.06	73.42
Cauchy - NMF	59.32	64.73	42.18	67.67
EWNMF	59.97	57.25	46.10	77.67
GEWNMF	61.12	58.83	47.27	80.12

Table 3: NMI of unsupervised feature extraction techniques

Dataset	Isolet	ORL	Yale	Lung-discrete
NMF	75.42	75.55	43.07	71.06
GNMF	74.89	75.49	43.29	70.73
ONMF	74.98	72.76	42.85	69.60
Semi - NMF	75.60	75.57	43.15	73.30
Convex - NMF	56.00	55.56	39.81	70.85
RNMF	59.48	63.94	36.56	63.17
Cauchy - NMF	73.53	62.29	46.91	56.81
EWNMF	73.73	75.45	53.10	70.36
GEWNMF	76.07	75.92	53.66	75.21

Various unsupervised feature extraction methodologies were evaluated using clustering accuracy (ACC) and normalized mutual information (NMI) metrics [3, 21], with results summarized in Tables 2 and 3.

Given a set of true class labels y and the resulting cluster labels y', clustering accuracy is defined as follows:

$$\mathrm{ACC} = \frac{1}{N}\sum_{i=1}^{N}\delta(y_i, \mathrm{map}(y_i')),$$

where

$$\delta(a,b) = \begin{cases} 1 & a = b \\ 0 & \text{otherwise.} \end{cases}$$

Here, map$(\cdot)$ represents a permutation mapping function that aligns the obtained cluster labels with the true labels. A higher ACC value indicates better clustering performance.

NMI measures the agreement between two data distributions and is defined as:

$$\mathrm{NMI}(y, y') = \frac{\mathrm{MI}(y, y')}{\max(H(y), H(y'))},$$

where $H(y)$ denotes the entropy of y and $\mathrm{MI}(y, y')$ represents the mutual information between the two random variables y and y', defined as:

$$\mathrm{MI}(y, y') = \sum_{y_i \in y, y'_i \in y'} p(y_i, y'_i) \ln \left(\frac{p(y_i, y'_i)}{p(y_i)p(y'_i)} \right).$$

Here, $p(y_i)$ and $p(y'_j)$ are the probabilities of a data point belonging to clusters y_i and y'_j, respectively, while $p(y_i, y'_j)$ is the joint probability that a data point is assigned to both clusters y_i and y'_j simultaneously. The NMI score varies between 0 and 1, with higher values indicating better clustering performance. In this research, we report the NMI values as percentages.

4.2. Result and analysis

Illustrated in Figs. 2 and 3, the curves representing clustering accuracy and normalized mutual information are plotted across varying reduced ranks, encompassing different methodologies. An in-depth analysis of these curves reveals the performance of various feature extraction approaches. It becomes evident that, in the majority of instances, the GEWNMF method consistently outperforms its counterparts, as evidenced by its ability to achieve higher ACC and NMI scores across a diverse range of feature extraction scenarios.

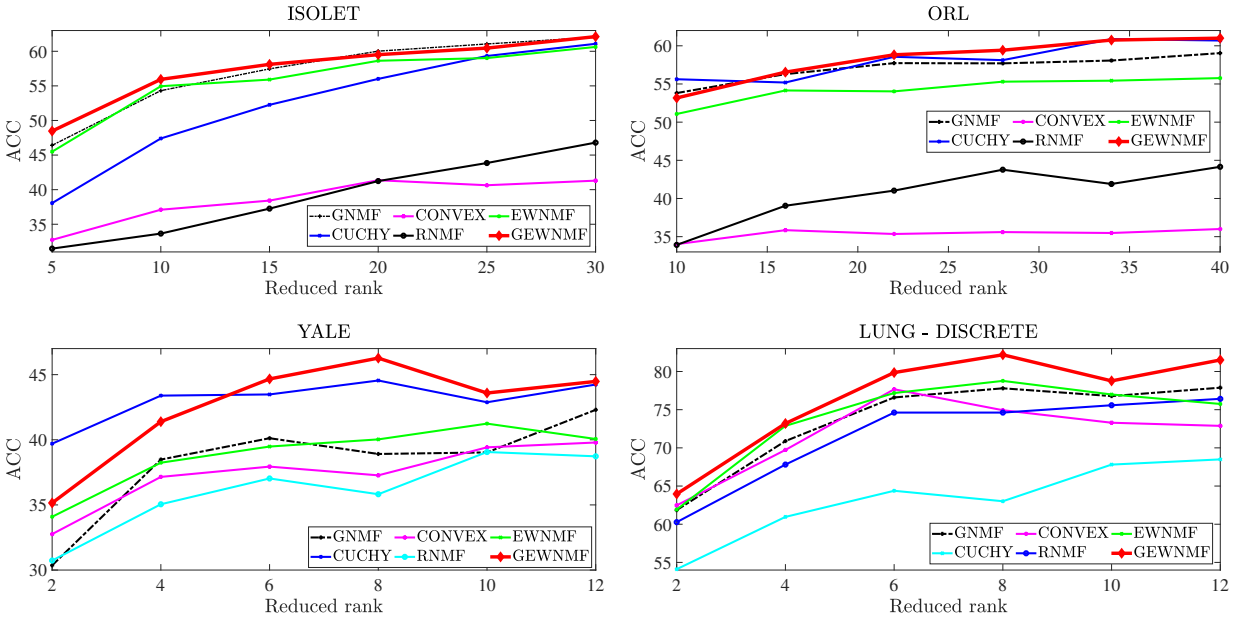

Figure 2: Results of ACC of various feature extraction methods on the four datasets

4.3. Parameter settings

Our proposed method includes three regularization parameters α, β, and γ, and its performance is affected by the values of these parameters. Therefore, analyzing the impact of these parameters on GEWNMF and identifying suitable values for them is crucial. As we demonstrated in Equations 1 to 3, the value of parameter γ is influenced by parameter β, so we perform the parameter analysis on α and β). We evaluate the performance of GEWNMF using different values for α and β.

The values of the parameters α and β are searched from $\{2^i : i = -8, -7, \ldots, 7, 8\}$, and the datasets in Table 1 are used to investigate the effect of variations in the main tuning parameters. Once a new representation is specified, the k-means algorithm is used to cluster all the samples according to the new representation.

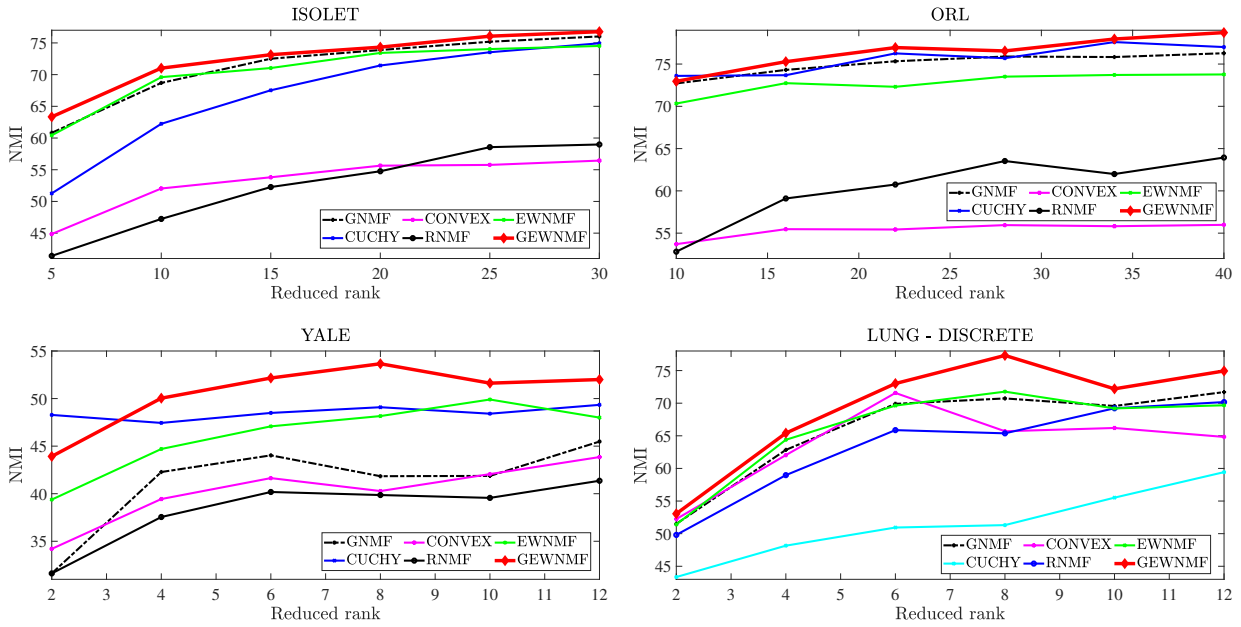

Figure 3: Results of NMI of various feature extraction methods on the four datasets

On the other hand, initialization significantly impacts the performance of k-means clustering. To account for this, we repeat the clustering process 20 times with different random initial values and present the average results along with their standard deviations. We evaluate the performance of GEWNMF using the ACC and NMI metrics. Figure 4 illustrates the ACC values for various settings of the α and β parameters, while Figure 5 shows the NMI values. The results reveal that GEWNMF exhibits only slight sensitivity to these primary parameters.

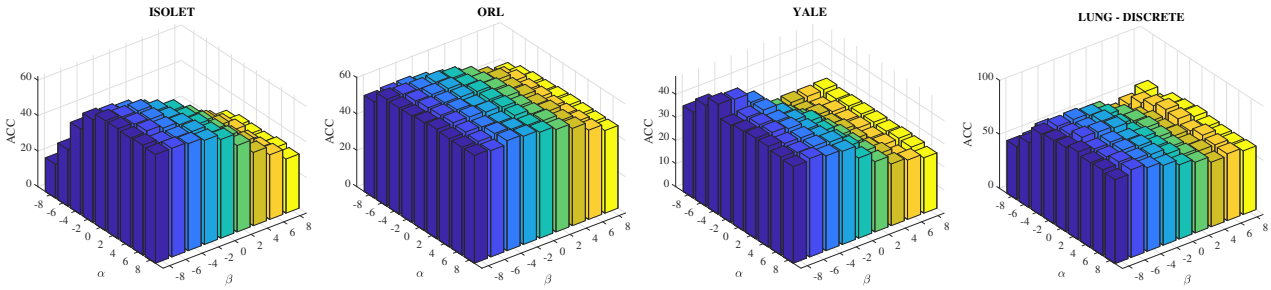

Figure 4: The ACC of GEWNMF with different values of parameters α and β on different datasets

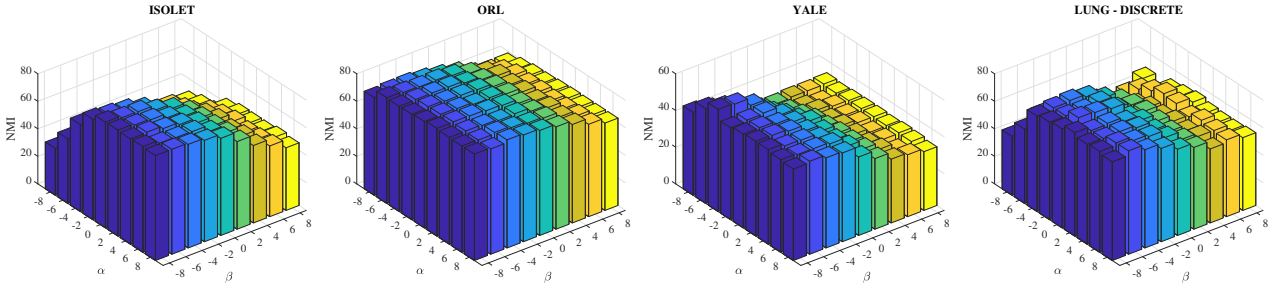

Figure 5: The NMI of GEWNMF with different values of parameters α and β on different datasets

4.4. Statistical test

In this section, the Friedman test, a nonparametric statistical test, is employed to assess the efficacy of different comparison methods, including the proposed GEWNMF method. To achieve this, the performance of GEWNMF across all four datasets is analyzed using the outcomes from Tables 2 and 3. In this context, datasets serve as

the subjects under study, comparison methods represent the various treatments applied, and evaluation metrics offer the measurements. The primary objective of this analysis is to determine the average ranking of the methods across all datasets based on the specified criteria. Specifically, the performance of the methods is evaluated by testing the null hypothesis that assumes no significant difference between them, using Friedman's test to confirm this hypothesis and determine the relative effectiveness of the methods.

The results of the Friedman test, as shown in Fig. 6, indicate the superiority of the GEWNMF method. It's important to note that a lower rank indicates better performance. Upon scrutinizing Fig. 6, it becomes evident that the GEWNMF method secures the topmost position in both ACC and NMI metrics. This suggests that GEWNMF demonstrates the utmost effectiveness when juxtaposed with alternative methodologies.

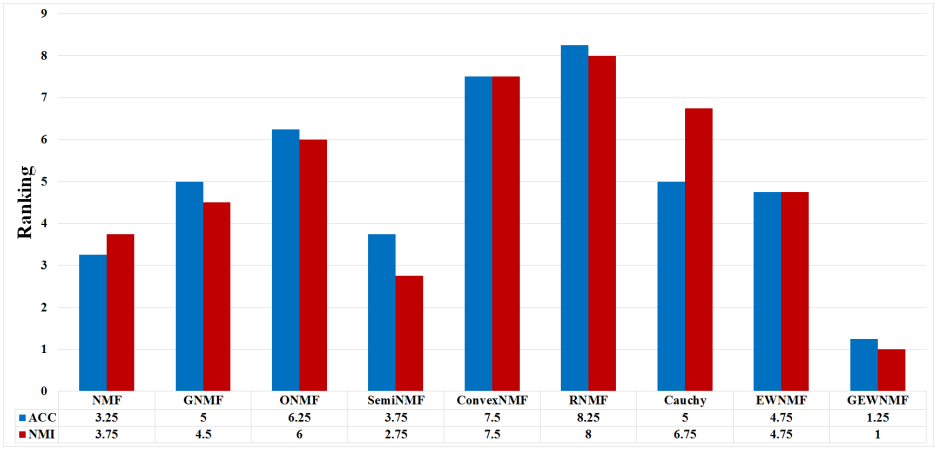

Figure 6: Average rankings obtained from the Friedman test

Moreover, to provide a more comprehensive insight into the evident enhancement in the clustering outcomes achieved by GEWNMF, as depicted in Tables 2 and 3, a rigorous statistical analysis of the GEWNMF results is conducted. Specifically, the paired t-test, a statistical hypothesis test, is employed. To conduct the t-test with a significance level of $\alpha = 0.05$, each algorithm was executed 20 times. We utilized the results from Tables 2 and 3, each representing the average of 20 repetitions. The statistical experiment yields two key indicators, denoted as h and p. The significance level, denoted by p, determines the threshold for statistical significance.

If the test result yields $h = 1$ and the p-value is minimal, it suggests that the null hypothesis cannot be rejected at the 5% level, indicating a significant difference between the two algorithms. Conversely, if the null hypothesis can be rejected at the 5% level, the test result will be $h = 0$. Tables 4 and 5 present the paired t-test outcomes of GEWNMF against other methods in all datasets. Table 4 reveals that the paired t-test results indicate significant discrepancies in the ACC values between GEWNMF and the other methods. Across most datasets, the paired t-tests yield $h = 1$ and very small p-values. Conversely, on a few datasets, $h = 0$, suggesting that the ACC values of GEWNMF do not exhibit a noticeable improvement compared to other algorithms. Overall, GEWNMF demonstrates a substantial enhancement in ACC across the majority of cases.

The findings from Table 5 reveal that, with few exceptions, the paired t-test results for NMI indicate $h = 1$ and small p-values across the majority of cases. This suggests notable discrepancies between the NMI values obtained by GEWNMF and those of the comparison algorithms, indicating a significant enhancement in NMI achieved by GEWNMF. The clustering results presented in Tables 4 and 5 consistently demonstrate a considerable improvement with GEWNMF compared to other algorithms, thereby confirming the superiority of GEWNMF.

4.5. Convergence behavior of the GEWNMF in practice

The experiments in this section are dedicated to examining the convergence behavior of the GEWNMF technique. As anticipated in Section 3.4, the GEWNMF cost function demonstrates a consistent decrease over multiple iterations until convergence is reached. This observation is further illustrated in Fig. 7, where we observe a similar trend across the datasets listed in Table 1, with the cost function steadily decreasing over iterations until it converges. These results demonstrate the effectiveness of the GEWNMF optimization algorithm.

4.6. Time Complexity

To further assess the effectiveness of the chosen methods, we compared their running times for all nine methods, as detailed in Table 6. GEWNMF surpasses both the RNMF and Cauchy-NMF methods across all datasets, exceeds

Table 4: The paired t-test outcome of ACC of GEWNMF and the comparison algorithm across all datasets

Algorithms	Lung-discrete		ORL		Isolet		Yale	
	p	h	p	h	p	h	p	h
NMF	$1.11\ e-15$	1	$2.63\ e-10$	1	$3.03\ e-05$	1	$1.03\ e-13$	1
GNMF	$1.14\ e-17$	1	$2.96\ e-10$	1	$6.39\ e-04$	1	$9.11\ e-16$	1
ONMF	$1.25\ e-19$	1	$1.99\ e-15$	1	$1.78\ e-01$	0	$6.98\ e-13$	1
Semi-NMF	$2.74\ e-21$	1	$4.43\ e-22$	1	$1.37\ e-21$	1	$1.73\ e-14$	1
Convex-NMF	$1.16\ e-22$	1	$3.09\ e-26$	1	$3.58\ e-25$	1	$4.98\ e-17$	1
RNMF	$1.69\ e-11$	1	$7.08\ e-07$	1	$3.84\ e-03$	1	$2.39\ e-19$	1
Cauchy-NMF	$8.02\ e-35$	1	$2.25\ e-01$	0	$1.35\ e-01$	0	$8.94\ e-11$	1
EWNMF	$5.91\ e-06$	1	$6.76\ e-10$	1	$3.56\ e-05$	1	$6.89\ e-02$	0

Table 5: The paired t-test outcome of NMI of GEWNMF and the comparison algorithm across all datasets

Algorithms	Lung-discrete		ORL		Isolet		Yale	
	p	h	p	h	p	h	p	h
NMF	$1.59\ e-14$	1	$2.86\ e-02$	1	$1.43\ e-02$	1	$1.26\ e-18$	1
GNMF	$1.08\ e-14$	1	$1.83\ e-08$	1	$1.96\ e-03$	1	$7.11\ e-18$	1
ONMF	$2.72\ e-15$	1	$1.28\ e-14$	1	$2.12\ e-03$	1	$4.25\ e-16$	1
Semi-NMF	$6.51\ e-08$	1	$9.85\ e-03$	1	$1.54\ e-01$	0	$7.15\ e-15$	1
Convex-NMF	$1.32\ e-14$	1	$1.35\ e-27$	1	$2.75\ e-25$	1	$7.63\ e-21$	1
RNMF	$7.10\ e-21$	1	$3.48\ e-22$	1	$7.61\ e-23$	1	$9.99\ e-22$	1
Cauchy-NMF	$2.54\ e-24$	1	$2.84\ e-23$	1	$5.06\ e-08$	1	$3.34\ e-15$	1
EWNMF	$4.66\ e-12$	1	$1.30\ e-04$	1	$1.43\ e-08$	1	$7.68\ e-02$	0

Table 6: Comparison of the running time of nine algorithms on four datasets for 200 iterations

Dataset	Yale	ORL	Isolet	Lung-discrete
NMF	0.22	0.7	1.53	0.08
GNMF	0.24	0.75	1.74	0.09
ONMF	1.74	3.48	6.73	1.41
Semi-NMF	0.42	1.23	2.66	0.13
Convex-NMF	0.48	1.66	21.32	0.09
RNMF	4.48	14.71	77.22	0.37
Cauchy-NMF	17.91	25.12	108.13	12..64
EWNMF	2.53	3.25	9.61	0.26
GEWNMF	2.58	3.33	9.7	0.27

the performance of the Convex-NMF method on the Isolet dataset, and outperforms ONMF on the ORL and Lung-discrete datasets. Overall, the time complexity of the proposed method is moderate.

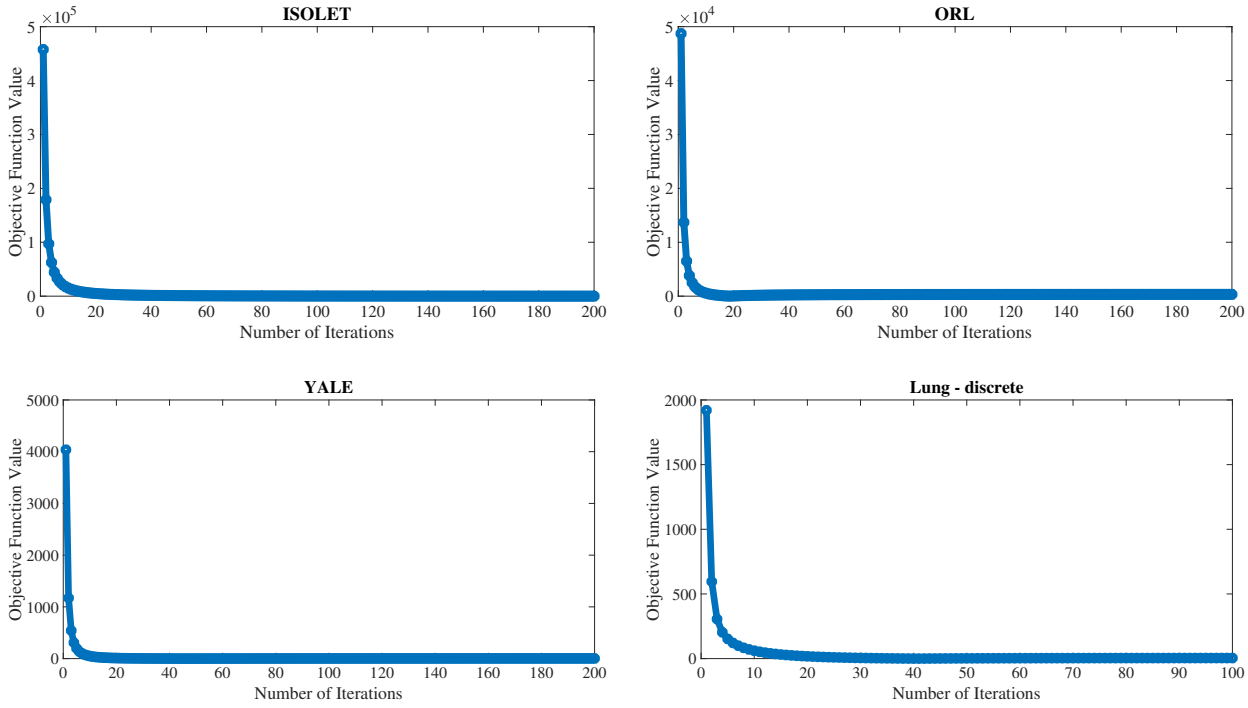

Figure 7: Convergence diagrams of the GEWNMF on different datasets

5. Conclusion

This article introduces a novel approach for dimensionality reduction called GEWNMF. The method employs a weighted non-negative matrix factorization framework that assigns adjustable weights to each feature based on its significance and reliability. This adaptive weighting scheme ensures that the extracted features accurately reflect the underlying data structure, thereby enhancing the interpretability and utility of the feature representation. In addition to feature weighting, the method incorporates a graph-based regularization technique by constructing an adjacency graph with an optimal number of neighbors. This step preserves the local relationships within the data, facilitating the extraction of meaningful patterns and improving the method's robustness to noise and outliers. Furthermore, the integration of an entropy term into the objective function enriches the feature representation by promoting the inclusion of diverse features that contribute uniquely to the data representation. This enhancement not only improves the model's ability to capture complex data distributions but also boosts its generalization capability across different datasets. The efficacy of the proposed approach is validated through extensive experiments conducted on multiple datasets, demonstrating its superiority over existing methods in terms of feature extraction and representation quality.

Future directions include exploring the interplay between feature weights and adjacency graph formation. Investigating how different weight assignments influence graph structure and subsequent feature extraction outcomes could provide deeper insights into further optimizing the method. Additionally, the potential integration of Distance Metric Learning (DML) techniques offers a promising avenue for enhancing feature discrimination and classification accuracy in practical applications. Furthermore, as computational resources advance, we plan to evaluate the method on more complex datasets such as "Orlraws10P," "CLL_SUB111," "Prostate_GE," and "warpPIE10P," to assess its scalability and performance in more challenging scenarios.

References

[1] I. Ahmed, T. Galoppo, X. Hu, and Y. Ding, *Graph regularized autoencoder and its application in unsupervised anomaly detection*, IEEE Trans. Pattern Anal. Mach. Intell., 44 (2022), pp. 4110–4124.

[2] V. D. Blondel, N. D. Ho, and P. V. Dooren, *Weighted nonnegative matrix factorization and face feature extraction*, Image Vis. Comput., 26 (2008), pp. 1–17.

[3] D. Cai, X. He, and J. Han, *Document clustering using locality preserving indexing*, IEEE Trans. Knowl. Data Eng., 17 (2005), pp. 1624–1637.

[4] D. Cai, X. He, J. Han, and T. S. Huang, *Graph regularized nonnegative matrix factorization for data representation*, IEEE Trans. Pattern Anal. Mach. Intell., 33 (2010), pp. 1548–1560.

[5] M. Chen, M. Gong, and X. Li, *Feature weighted non-negative matrix factorization*, IEEE Trans. Cybern., 53 (2021), pp. 1093–1105.

[6] P. Deng, T. Li, H. Wang, S. J. Horng, Z. Yu, and X. Wang, *Tri-regularized nonnegative matrix tri-factorization for co-clustering*, Knowl.-Based Syst., 226 (2021), p. 107101.

[7] P. Deng, T. Li, H. Wang, D. Wang, S. Horng, and R. Liu, *Graph regularized sparse non-negative matrix factorization for clustering*, IEEE Trans. Comput. Soc. Syst., (2022), pp. 1–12.

[8] C. Ding, T. Li, and M. I. Jordan, *Convex and semi-nonnegative matrix factorizations*, IEEE Trans. Pattern Anal. Mach. Intell., 32 (2010), pp. 45–55.

[9] C. Ding, T. Li, W. Peng, and H. Park, *Orthogonal nonnegative matrix t-factorizations for clustering*, in Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2006, pp. 126–135.

[10] Y. Dong, H. Che, M. F. Leung, C. Liu, and Z. Yan, *Centric graph regularized log-norm sparse non-negative matrix factorization for multi-view clustering*, Signal Process., 217 (2024), pp. 109–341.

[11] H. Gao, F. Nie, W. Cai, and H. Huang, *Robust capped norm nonnegative matrix factorization: Capped norm nmf*, in Proceedings of the 24th ACM International Conference on Information and Knowledge Management, 2015, pp. 871–880.

[12] N. Guan, T. Liu, Y. Zhang, D. Tao, and L. S. Davis, *Truncated cauchy non-negative matrix factorization*, IEEE Trans. Pattern Anal. Mach. Intell., 41 (2017), pp. 246–259.

[13] R. Hamon, V. Emiya, and C. Fevotte, *Convex nonnegative matrix factorization with missing data*, in IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), 2016, pp. 1–6.

[14] A. B. Hamza and D. J. Brady, *Reconstruction of reflectance spectra using robust nonnegative matrix factorization*, IEEE Trans. Signal Process., 54 (2006), pp. 3637–3642.

[15] S. Huang, H. Wang, T. Li, T. Li, and Z. Xu, *Robust graph regularized nonnegative matrix factorization for clustering*, Data Min. Knowl. Discov., 32 (2018), pp. 483–503.

[16] S. Huang, Z. Xu, Z. Kang, and Y. Ren, *Regularized nonnegative matrix factorization with adaptive local structure learning*, Neurocomputing, 382 (2020), pp. 196–209.

[17] Q. i. Huang, X. Yin, S. Chen, Y. Wang, and B. Chen, *Robust nonnegative matrix factorization with structure regularization*, Neurocomputing, 412 (2020), pp. 72–90.

[18] Y. Jia, S. Kwong, J. Hou, and W. Wu, *Semi-supervised non-negative matrix factorization with dissimilarity and similarity regularization*, IEEE Trans. Neural Netw. Learn. Syst., 31 (2019), pp. 2510–2521.

[19] Y. D. Kim and S. Choi, *Weighted nonnegative matrix factorization*, in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2009, pp. 19–24.

[20] D. D. Lee and H. S. Seung, *Learning the parts of objects by non-negative matrix factorization*, Nature, 401 (1999), pp. 788–791.

[21] H. Liu, Z. Wu, X. Li, D. Cai, and T. S. Huang, *Constrained nonnegative matrix factorization for image representation*, IEEE Trans. Pattern Anal. Mach. Intell., 34 (2012), pp. 1299–1311.

[22] A. Lotfi, P. Moradi, and H. Beigy, *Density peaks clustering based on density backbone and fuzzy neighborhood*, Pattern Recognit., 107 (2020), p. 107449.

[23] X. Ma, W. Zhao, and W. Wu, *Layer-specific modules detection in cancer multi-layer networks*, IEEE/ACM Trans. Comput. Biol. Bioinform., 20 (2022), pp. 1170–1179.

[24] F. Nie, X. Wang, and H. Huang, *Clustering and projected clustering with adaptive neighbors*, in Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 2014, Association for Computing Machinery, pp. 977–986.

[25] P. Paatero, *Least squares formulation of robust, non-negative factor analysis*, Chemom. Intell. Lab. Syst., 37 (1997), pp. 23–35.

[26] A. Suleman, *A convex semi-nonnegative matrix factorization approach to fuzzy k-means clustering*, Fuzzy Sets Syst., 270 (2015), pp. 90–110.

[27] J. Tang and H. Feng, *Robust local-coordinate non-negative matrix factorization with adaptive graph for robust clustering*, Inf. Sci., 610 (2022), pp. 1058–1077.

[28] C. Wang, J. Wang, Z. Gu, J. M. Wei, and J. Liu, *Unsupervised feature selection by learning exponential weights*, Pattern Recognit., 148 (2024), pp. 110–183.

[29] D. Wang, T. Li, and C. Ding, *Weighted feature subset non-negative matrix factorization and its applications to document understanding*, in IEEE International Conference on Data Mining, 2010, pp. 541–550.

[30] Y. X. Wang and Y. J. Zhang, *Nonnegative matrix factorization: a comprehensive review*, IEEE Trans. Knowl. Data Eng., 25 (2013), pp. 1336–1353.

[31] J. Wei, C. Tong, B. Wu, H. Qiang, Q. Shouliang, Y. Yudong, and T. Yueyang, *An entropy weighted nonnegative matrix factorization algorithm for feature representation*, IEEE Trans. Neural Netw. Learn. Syst., 34 (2022), pp. 1–11.

[32] W. Wu, Y. Jia, S. Wang, R. Wang, H. Fan, and S. Kwong, *Positive and negative label-driven nonnegative matrix factorization*, IEEE Trans. Circuits Syst. Video Technol., 31 (2021), pp. 2698–2710.

[33] W. Wu, S. Kwong, J. Hou, Y. Jia, and H. H. S. Ip, *Simultaneous dimensionality reduction and classification via dual embedding regularized nonnegative matrix factorization*, IEEE Trans. Image Process., 28 (2019), pp. 3836–3847.

[34] Y. Wu, Y. Wang, L. Hu, and J. Hu, *Dcgnn: Adaptive deep graph convolution for heterophily graphs*, Inf. Sci., 666 (2024), pp. 120–427.

[35] X. Zhang, D. Chen, H. Yu, G. Wang, H. Tang, and K. Wu, *Improving nonnegative matrix factorization with advanced graph regularization*, Inf. Sci., 597 (2022), pp. 125–143.

[36] J. Zhao and G. F. Lu, *Clean affinity matrix learning with rank equality constraint for multiview subspace clustering*, Pattern Recognit., 134 (2023), pp. 109–118.

[37] Z. Zhou, G. Si, H. Sun, K. Qu, and W. Hou, *A robust clustering algorithm based on the identification of core points and k-nn kernel density estimation*, Expert Syst. Appl., 195 (2022), pp. 116–573.

Please cite this article using:

Hazhir Sohrabi, Shahrokh Esmaeili, Parham Moradi, Feature representation via graph-regularized entropy-weighted nonnegative matrix factorization, AUT J. Math. Comput., 5(4) (2024) 289-304
https://doi.org/10.22060/AJMC.2024.23353.1252