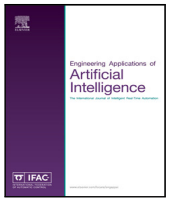
Engineering Applications of Artificial Intelligence 121 (2023) 106043

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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



Semi-supervised non-negative matrix tri-factorization with adaptive neighbors and block-diagonal learning

Songtao Li [a,d], Weigang Li [a,b,*], Hao Lu [c], Yang Li [a,b]

[a] School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, China
[b] Engineering Research Center for Metallurgical Automation and Measurement Technology Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081, China
[c] School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China
[d] School of Artificial Intelligence, Jianghan University, Wuhan 430056, China

ARTICLE INFO

Keywords:
Non-negative matrix tri-factorization
Semi-supervised learning
Adaptive neighbors learning
Block-diagonal structure
Clustering

ABSTRACT

Graph-regularized non-negative matrix factorization (GNMF) is proved to be effective for the clustering of nonlinear separable data. Existing GNMF variants commonly improve model performance by adding different additional constraints or refining the model factorization form, which can lead to problems such as increased algorithm complexity or insufficient performance release. In this paper, we propose semi-supervised non-negative matrix tri-factorization with adaptive neighbors and block-diagonal (ABNMTF). Different from existing methods, in ABNMTF the similarity graph matrix is learned from the original data by adaptive neighbors k-nearest model, and a block diagonal matrix is constructed based on a few labeled data to update the similarity matrix. Our approach reconstructs the block diagonal structure into the adaptive similarity matrix, which enables simultaneous learning of the similarity matrix and label binding during factorization, engendering a distinguishable subspace representation matrix and therefore improving the clustering performance without significantly increasing the complexity of the algorithm. We also represent an optimization method to solve the ABNMTF and provide analyses of convergence and computational complexity. Extensive experiments on 8 real image datasets show that the proposed algorithm reports superior performance against several state-of-the-art approaches. *Code has been made available at:* https://github.com/LstinWh/ABNMTF.

1. Introduction

With the explosion of information, data have grown in both dimensionality and quantity, leading to a decline in the use of raw data for decision-making in practice (Sharma and Seal, 2020; Zubaroğlu and Atalay, 2021). It therefore requires techniques to extract effective low-dimensional data from high-dimensional ones to improve the performance in various perception tasks (Li et al., 2021). There are many dimensionality reduction methods, such as local linear embedding (LLE) (Roweis and Saul, 2000), singular value decomposition (SVD) (Wall et al., 2003), and Principal Component Analysis (PCA) (Abdi and Williams, 2010). Among them, Non-negative Matrix Factorization(NMF) explores the underlying structure of the data to obtain a meaningful representation of high-dimensional data (Lee and Seung, 1999). Compared with other dimensionality reduction algorithms, NMF is imposed by additional non-negative constraints, resulting in better physical meanings and theoretical explanations of the learned low-dimensional representations. The motivation of NMF is consistent with the fact that many data in reality are non-negative. Therefore, NMF has received much attention in many application scenarios, such as hyperspectral unmixing (Zhou et al., 2020; Xu et al., 2020; Dong et al., 2020), document clustering (Lydia et al., 2020), community detection (He et al., 2021; Ye et al., 2018), source separation (Yamamoto et al., 2020), data representation (Min et al., 2022; Wei et al., 2022; Shu et al., 2022) and gene analysis (Yu et al., 2020; Jiao et al., 2020).

While NMF is well developed, there is still room for optimization in many directions. In the past two decades, many researchers have proposed a large number of NMF variants to deal with various challenges. Ding et al. (2006) proposed a non-negative matrix tri-factorization (NMTF) model by introducing new factors into NMF, which is more flexible than NMF and is used for co-clustering. Li et al. (2010) proposed orthogonal non-negative matrix factorization (ONMF). The model additionally adds orthogonality constraints, which improves subspace learning of the algorithm and is also conducive to the interpretation of clustering. Cai et al. (2010) presented a graph regularized nonnegative matrix factorization (GNMF) method. It considers the spatial geometric information of the original data by constructing the manifold structure of the data graph, and incorporates it into the model

* Corresponding author.
E-mail address: liweigang.luck@foxmail.com (W. Li).

https://doi.org/10.1016/j.engappai.2023.106043
Received 17 November 2022; Received in revised form 10 January 2023; Accepted 21 February 2023
Available online 27 February 2023

through regularization terms. Liu et al. (2011) proposed a constrained non-negative matrix factorization (CNMF), which considered part of the labeled data and constructed a global auxiliary matrix as hard constraints. These variants all improve NMF models in a certain way with few additional parameters, while the performance improvement is not significant due to single-factor modeling.

In recent years, more and more researchers focus on NMF models with hybrid constraints (Wu et al., 2018b, 2019, 2018a). Since the advantages of multiple variants of NMF models are combined, high performance NMF algorithms can be obtained. Meng et al. (2018) proposed dual graphs to regularize non negative matrix factorization with sparse and orthogonal constraints (SODNMF). This framework integrates the dual graph regularization, sparse constraints, and orthogonal constraints into a unified framework, which significantly improves subspace learning. Jia et al. (2019) gave a semi-supervised non-negative matrix factorization with dissimilarity and similarity regularization (GSSNMF) method. This method constructs similarity and dissimilarity graph structures through partial label information, which can accurately describe the spatial geometric structure of samples. Peng et al. (2020) presented correntropy based orthogonal nonnegative matrix tri-factorization (CNMTF). It also integrates dual graph regularization and orthogonal constraints into the NMTF algorithm, and uses the correlation entropy to measure the loss value, which robustizes the model. Wu et al. (2021) used the antagonism of labeled data in semi-supervised learning and proposed Positive and Negative Label-Driven Nonnegative Matrix Factorization (PNLD-NMF), which unifies label information constraints and dual graph regularization. Peng et al. (2021b) found that most graph regularization are based on global data to construct a similarity matrix, while ignoring the local characteristics of the data. Therefore, an effective model is proposed by constructing global and local similarity matrices and combining orthogonal constraints. It is called Nonnegative matrix factorization with local similarity learning (KLS-NMF). It is obvious that most mixed constrained NMF models combine the graph regularization method into their own modules, which not only improves the performance of the model, but also shows that the graph regularization method is of great significance to the NMF model. However, the incorporation of additional constraints and graph regularization will increase many additional parameters (in theory, a graph regularizer will bring more than two additional parameters), which is costly in computation and unfriendly to implementation.

To overcome the aforementioned issues, the goal of this work is to improve the performance of the model with as few additional parameters as possible. Fig. 1 shows the flow of ABNMTF in detail. Specifically, on one hand, the similarity graph structure is reconstructed for the original data such that it can accurately express the spatial connotation of the original data. On the other hand, to improve the utilization of label information, we construct a global label constraint matrix, and use the label information to construct a block diagonal matrix to reconstruct the above similarity matrix, which not only reuses label information, but also relaxes the hard constraints of the algorithm. With these ideas, we define the objective function of the algorithm and derive an effective and convergent multiplication iterative algorithm to update the model under the KKT condition. We also conducted a large number of controlled experiments on eight real image datasets to demonstrate the effectiveness of the algorithm. Our main contributions are four-fold:

- We unify the adaptive neighbors graph structure and diagonal block induction method into NMTF, and proposes the ABNMTF method. This method elegantly adjusts the relationship between partial label information and adaptive neighbors graph structure, which is representative and applicable.
- To the best of our knowledge, this work is the first to integrate semi-supervised block diagonal induction into the adaptive neighbors similarity graph matrix of NMTF. It allows the algorithm to exploit the label information in block diagonals and into the adaptive neighbors graph regularization, which can increase each other and strengthens subspace learning of NMTF.
- The mainstream manifold NMF algorithm is sensitive to the input similarity graph matrix, which means that the similarity graph matrix determines the performance of the algorithm. In the process of constructing the similarity graph matrix, ABNMTF allocates adaptive neighbors to each data point, which enables adaptive changes between the data and accurately describes the similarity relationship of the original data.
- In the ABNMTF algorithm, the labeled data are reused twice to construct a global label matrix and block diagonal induction. Compared with other semi-supervised NMF methods, ABNMTF fully uses label information.

The rest of this paper is organized as follows: In Section 2, we first define the notations and briefly review related variant NMF methods. In Section 3, the ABNMTF algorithm is described in detail, and the proof of convergence is given mathematically. In Section 4, a variety of controlled experiments are designed to verify the effectiveness of the ABNMTF algorithm. Finally, we draw a conclusion and provide suggestions for future work in Section 5.

2. Preliminaries

2.1. Notations

Here we introduce the symbols used in this article. We suppose that a capital X denote matrix, a lowercase x is a vector and scalar. For example, given an arbitrary $m \times n$ matrix A, we define a_j and a_{ij} as the jth column vector and the (i,j)th entry of matrix A, respectively. Furthermore, $Tr(\cdot)$ represents the trace of a matrix, $(\cdot)^T$ denotes the transpose operator, and $\odot$ indicates the elementwise multiplication operator. $\|\cdot\|_F$, $\|\cdot\|_1$ and $\|\cdot\|_2$ are respectively the Frobenius, L_1 and L_2 norms.

2.2. Related work

In this subsection, we review some work closely related to our research.

2.2.1. Non-negative matrix tri-factorization

Given a nonnegative data matrix $X = [x_1, x_2, \ldots, x_n], X \in \mathbf{R}_+^{m\times n}$ consisting of n samples, each column x_i of X is an m-dimensional sample. The goal of standard NMTF (Ding et al., 2006) is to decompose X into three non-negative matrices: a column-coefficient matrix $V \in \mathbf{R}^{n\times k}$, a block value matrix $S \in \mathbf{R}^{k\times k}$ and a row-coefficient matrix $U \in \mathbf{R}^{m\times k}$, where k is the rank that you want to factorize in NMTF. The objective function of the standard NMTF is as follows:

$$\min_{U\geq 0, V\geq 0, S\geq 0} \|X - USV^T\|_F^2 \tag{1}$$

2.2.2. Graph laplacian

The viewpoints studied in manifold learning show that the real data in the real world are non-linear and are all sampled from a low-dimensional manifold, and the manifold structure is embedded in a high-level space (Belkin and Niyogi, 2008). For example, detecting correlation in community detection can be seen as mining the structural relationships in the graph manifold hidden in the context of the text content (Fortunato, 2010). The graph Laplacian is a commonly used manifold structure to measure the similarity between data. Its definition is:

$$\begin{aligned}\frac{1}{2}\sum_{i,j=1}^{n}\|x_i - x_j\|_2^2 w_{ij} &= \sum_{i=1}^{n} x_i D_{ii} x_i^T - \sum_{i,j=1}^{n} x_j w_{ij} x_j^T \\ &= Tr\left(X^T DX\right) - Tr\left(X^T WX\right) = Tr\left(X^T LX\right)\end{aligned} \tag{2}$$

where W represents the similarity matrix, which measures the pairwise similarity of the original data points and D is a diagonal matrix with $D_{ii} = \sum_j W_{ij}$, x_i and x_j represent a pair of data points in the original data, the Laplacian matrix is $L = D - W$. It is worth noting that the

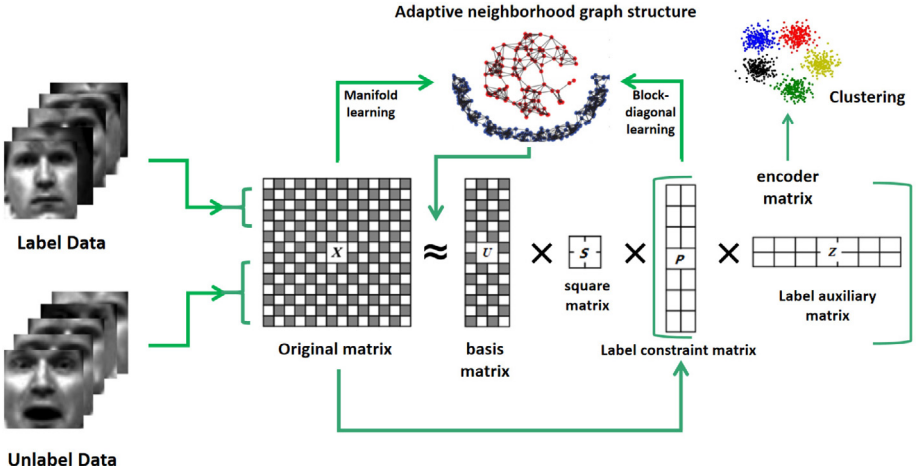


Fig. 1. Illustration of the proposed framework.

manifold enforces the smoothness of data in linear and nonlinear spaces by minimizing the above formula. This results in that if two data points are close in the inner geometry of the data distribution, then the two data points in the new representation space are also similar (Singer, 2006).

2.2.3. Graph regularized nonnegative matrix factorization

Cai et al. (2010) obtained the similarity matrix of the original data based on the thermal kernel weighting method, and its construction rules are as follows:

$$W_{i,j} = \begin{cases} \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right), & x_i \in N_p(x_j) \quad \text{or} \quad x_j \in N_p(x_i) \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where $W_{i,j}$ stands for the similarity between ith and jth samples, $x_i \in N_p(x_j)$ indicates that sample x_i belongs to the p-nearest samples of sample x_j and σ is Gaussian kernel parameter to control the values of similarity. Then, the $Tr(V^T LV)$ regularization term can be constructed to measure the smoothness of the low-dimensional representation, so the GNMF can be described as follows:

$$\min_{U\geq 0, V\geq 0} \|X - UV^T\|_F^2 + \lambda \mathrm{Tr}(V^T LV) \tag{4}$$

where λ is the regular weight parameter of the graph and $\lambda > 0$. GNMF successfully incorporates the geometric structure of the original data into the NMF method and has achieved a good non-linear dimensionality reduction effect. However, each construction and use of a similarity matrix requires the introduction of 2–3 additional adjustable parameters (for example: λ, σ and p).

2.2.4. Constrained nonnegative matrix factorization

NMF is an unsupervised linear dimensionality reduction method. It cannot use label information to improve its spatial discrimination. In order to overcome the above problems, CNMF (Liu et al., 2011) tries to use a small part of the label information to construct a label constraint matrix to implement the global constraints of NMF. The core of the CNMF model is to use the product of the above-mentioned constraint matrix P with label information and the label auxiliary matrix Z to replace the basic matrix V in the NMF, so CNMF can be described as follows:

$$\min_{U\geq 0, Z\geq 0} \|X - UZ^T P^T\|_F^2 \tag{5}$$

It can be clearly seen that the P matrix with the label information of the original data is included in the objective function. As the subsequent update iterations continue to constrain the entire model, CNMF as a semi-supervised algorithm will be better than NMF in performance. This method also has a significant advantage, that is, it does not increase the adjustable parameters of the algorithm while introducing additional constraints.

3. ABNMTF algorithm

In this section, we introduce the ABNMTF algorithm in detail. The algorithm uses the local graph adaptive structure to construct a similarity matrix, and then reconstructs the block diagonal matrix constructed by the labeled data into the similarity matrix, which greatly improves the graph regularization performance of the algorithm. Furthermore, the iterative update rule scheme is used to solve the objective function of ABNMTF. Finally, the convergence of the proposed algorithm is proved by the auxiliary function method and the computational complexity analysis of the algorithm is given.

3.1. Local structure graph regularization with adaptive neighbors

The traditional Laplacian graph does not consider the number of data clusters in the construction process, which leads to many connections in the graph structure, and it is impossible to accurately construct the internal geometric structure of the original data (Sh et al., 2020). In order to overcome the above problems, we do not use the traditional Laplacian graph structure, but learn a Euclidean distance-induced probability domain matrix based on the relationship between the original data and the number of clusters.

It can be known from Eq. (2) that the Laplacian diagram has only trivial solution; only the data points near x_i are its neighbors, and the other points are not its neighbors. Therefore, we can use the following formulation to describe its characteristics:

$$\min_{w_i^T 1=1, 0\leq w_i\leq 1} \sum_{j=1}^{n} \|x_i - x_j\|_2^2 w_{ij} \tag{6}$$

Nie et al. (2014) proposed a solution that does not involve any distance information to solve the trivial solution in the Laplace diagram as:

$$\min_{w_i^T 1=1, 0\leq w_i\leq 1} \sum_{j=1}^{n} w_{ij}^2 \tag{7}$$

The Eq. (7) describes the area where all the optimal solutions near the x_i data point have the same probability of $\frac{1}{n}$, which can be regarded as finding the prior information for the assignment of the neighborhood.

Combining Eq. (6) and Eq. (7) can obtain graph information with probability information:

$$\min_{w_i^T 1=1, 0\leq w_i\leq 1} \sum_{j=1}^{n} \left(\|x_i - x_j\|_2^2 w_{ij} + \gamma w_{ij}^2\right) \tag{8}$$

where γ is the regularization parameter used to control the size of the adaptive neighbors of local structure similarity graph matrix W_a. Let

$d_i^x \in \mathbf{R}^{n\times 1}$ be a vector and its jth element is $d_{ij}^x = \|x_i - x_j\|_2^2$, then the Eq. (8) can be rewritten as:

$$\min_{w_i^T 1=1, 0\le w_i\le 1} \|w_i + \frac{1}{2\gamma} d_i^x\|_2^2 \tag{9}$$

Newman et al. (1991) proposed a theorem: if the similarity matrix is non-negative, the multiplicity g of the eigenvalue 0 of the Laplacian matrix L is equal to the number of connected components in the graph with the similarity matrix. Therefore, a constraint can be added to the above similarity matrix generation rules, the new similarity matrix can reflect the ideal neighborhood allocation of clustering structure. This structure is called the adaptive neighborhood graph manifold structure:

$$\begin{aligned} &\min_{W_a} \sum_{i,j=1}^{n} \left(\|x_i - x_j\|_2^2 w_{ij} + \gamma w_{ij}^2\right) \\ &s.t.\ \forall i, w_i^T 1 = 0, 0 \le w_i \le 1, rank(L) = m - g \end{aligned} \tag{10}$$

where g denotes the cluster number, and m is the number of rows of input data X.

3.2. Efficient use of labeled data

3.2.1. Global structural constraints

In order to make full use of the label information, the low-dimensional representation V learned by NMTF is projected to the corresponding label space. The goal can be defined as:

$$V = PZ \tag{11}$$

where Z is an auxiliary matrix and P is the label constraint matrix, the matrix P is composed of two parts, one part is the label index matrix S, constructed according to the following rules: if the sample x_i belongs to the jth category, then $s_{ij} = 1$, otherwise $s_{ij} = 0$, and the other part is an $(n-l)\times(n-l)$-dimensional identity matrix, n means the number of samples of the input data, and l means the number of labeled samples. Note that P is naturally defined as:

$$P = \begin{pmatrix} C_{l\times c} & 0 \\ 0 & I_{n-l} \end{pmatrix} \tag{12}$$

where I_{n-l} is an $(n-l)\times(n-l)$ identity matrix and $C_{l\times c}$ denotes an $l\times c$ indicator matrix with the following rules:

$$c_{ij} = \begin{cases} 1, & x_i \in j\text{th} \quad \text{class} \\ 0, & x_i \notin j\text{th} \quad \text{class} \end{cases} \tag{13}$$

by incorporating the above label constraint matrix P into the NMTF algorithm, a basic semi-supervised NMTF algorithm can be transformed as:

$$\min_{U,S,Z\ge 0} \|X - USZ^T P^T\|_F^2 \tag{14}$$

where S is a block value matrix, and it is described in this paper as an identity matrix, which in NMTF (Ding et al., 2006) can provide more flexible decomposition choices. It can be observed from Eq. (14) that this method can embed the label information into the NMTF model in the form of global hard constraints, and can also avoid the introduction of new tunable parameters, which is a stable and efficient solution.

3.2.2. Block diagonal reconstruction of similarity matrix

Many studies have proved the value of block diagonal structure (Lu et al., 2018; Xie et al., 2017). In this subsection, we repeatedly use the known labels to construct a set of block diagonal matrices and reconstruct them into the similarity matrix. Specifically, we use the label information to construct a block diagonal matrix M, which is defined as:

$$M = \begin{pmatrix} J & 0 & \cdots & 0 \\ 0 & J & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J \end{pmatrix}_{l\times l} \tag{15}$$

where the label block matrix $J \in \mathbf{R}^{\frac{l}{k}\times\frac{l}{k}}$ and all elements of 1, l represents the number of labeled samples in the data sample and k represents the number of types of labeled samples in the data sample (the size of the decomposed rank). Suppose the input original data $X \in \mathbf{R}_+^{m\times n}$, then the constructed similarity matrix $W_a \in \mathbf{R}^{n\times n}$, which can be expressed as:

$$W_a = \begin{pmatrix} w_{11} & w_{21} & \cdots & w_{n1} \\ w_{12} & w_{22} & \cdots & w_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ w_{1n} & w_{2n} & \cdots & w_{nn} \end{pmatrix}_{n\times n} \tag{16}$$

It can be known that $l \le n$, so $\frac{l}{k} \le n$, so we reconstruct the W_{ab} matrix from the M matrix according to the following rules:

$$W_{ab} = \begin{pmatrix} J & 0 & \cdots & 0 & w_{1(l+1)} & \cdots & w_{1n} \\ 0 & J & \cdots & 0 & w_{2(l+1)} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J & w_{l(l+1)} & \cdots & w_{ln} \\ w_{(l+1)1} & w_{(l+1)2} & \cdots & w_{(l+1)l} & w_{(l+1)(l+1)} & \cdots & w_{(l+1)n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nl} & w_{n(l+1)} & \cdots & w_{nn} \end{pmatrix}_{n\times n} \tag{17}$$

Obviously, the reconstructed similarity matrix W_{ab} has a block diagonal form of label structure and an adaptive nearest neighbor graph that can more accurately describe the internal geometric structure of the original data. This matrix can better describe the association between labeled data and unlabeled data in semi-supervised learning, so as to make the low-dimensional representation more discriminative.

3.3. Objective function of ABNMTF

As mentioned above, the research of variant algorithms based on NMF is mostly developed through various mixed information (such as sparse constraints, supervisory information constraints, graph structure embedding (Meng et al., 2018; Shu et al., 2017; Peng et al., 2021b), etc.). Most of these methods are only linear combination constraints, and do not produce better coupling effects, and introduce more additional adjustable parameters, which is a great harm to the stability of the algorithm, and seriously reduces the applicability of the algorithm. Therefore, when a highly discriminative regularization term is constructed in the objective function, the entire algorithm framework has only one adjustable parameter, which can avoid multiple adjustment parameter problems and coupling problems caused by multiple constraints.

By combining the above factors, we define the semi-supervised non-negative matrix tri-factorization with adaptive neighbors and block-diagonal objective function as follows:

$$F_{ABNMTF} = \min_{U,S,Z,P\ge 0} \|X - USZ^T P^T\|_F^2 + \lambda \mathrm{Tr}\left(Z^T P^T L_{ab} PZ\right) \tag{18}$$

where λ is a hyperparameter that adjusts the contribution of the similarity matrix, L_{ab}=$D_{ab} - W_{ab}$ is Laplacian matrix, and D_{ab} is diagonal matrix with $d_{ii} = \sum_{j=1}^{n} w_{ij}$. It can be clearly observed that there is only one adjustable parameter in the objective function of our algorithm, which also means that our model can be directly extended to solve the problem of multi-constraint and graph clustering.

3.4. Optimization procedure of ABNMTF

It is obvious that the decomposed matrix U, S, P and Z in the objective function F_{ABNMTF} of ABNMTF are non-convex together. In order to solve the optimization problem, Eq. (18) can be converted into the following form:

$$\begin{aligned} &F_{ABNMTF} = Tr(XX^T) - 2Tr(XPZS^T U^T) \\ &+Tr(USZ^T P^T PZS^T U^T) + \lambda Tr((D_{ab} - W_{ab})(PZZ^T P^T)^T) \\ &= Tr(XX^T) - 2Tr(XPZS^T U^T) + Tr(USZ^T P^T PZS^T U^T) \\ &+\lambda Tr(Z^T P^T D_{ab} PZ) - \lambda Tr(Z^T P^T W_{ab} PZ) \\ &\text{s.t.} U \ge 0, S \ge 0, Z \ge 0, P \ge 0 \end{aligned} \tag{19}$$

Introduce three Lagrangian multipliers $\Psi = [\psi_{ij}], \Phi = [\phi_{ij}], \Gamma = [\tau_{ij}]$, $\Xi = [\xi_{ij}]$ to satisfy the following conditions $U_{ij} \geq 0, S_{ij} \geq 0$, $Z_{ij} \geq 0$ and $P_{ij} \geq 0$, we can convert the Lagrange function of the objective function as follows:

$$\begin{aligned} L_{ABNMTF} &= F_{ABNMTF} + Tr(\Psi U^T) + Tr(\Phi S^T) + Tr(\Gamma Z^T) \\ &= Tr(XX^T) - 2Tr(XPZS^TU^T) + Tr(USZ^TP^TPZS^TU^T) \\ &+\lambda Tr(Z^TP^TD_{ab}PZ) - \lambda Tr(Z^TP^TW_{ab}PZ) \\ &+Tr(\Psi U^T) + Tr(\Phi S^T) + Tr(\Gamma Z^T) + Tr(\Xi P^T) \end{aligned} \tag{20}$$

In order to obtain the effective iterative rules of the four decomposition matrices in the objective function, we should take the respective partial derivatives of L:

$$\frac{\partial L}{\partial U} = -2XPZS^T + 2USZ^TP^TPZS^T + \Psi \tag{21}$$

$$\frac{\partial L}{\partial S} = -2U^TXPZ + 2U^TUSZ^TP^TPZ + \Phi \tag{22}$$

$$\frac{\partial L}{\partial P} = -2X^TUSZ^T + 2PZS^TU^TUSZ^T + 2\lambda D_{ab}PZZ^T - 2\lambda W_{ab}PZZ^T + \Xi \tag{23}$$

$$\frac{\partial L}{\partial Z} = -2P^TX^TUS + 2P^TPZS^TU^TUS + 2\lambda P^TD_{ab}PZ - 2\lambda P^TW_{ab}PZ + \Gamma \tag{24}$$

Taking into account $\psi_{ij}u_{ij} = 0$, $\phi_{ij}s_{ij} = 0$, $\tau_{ij}z_{ij} = 0$ and $\xi_{ij}p_{ij} = 0$ in the KKT condition, we can eliminate the constraint conditions in the objective function and get the corresponding iteration rules as follows:

$$u_{ij} \leftarrow u_{ij}\frac{(XPZS^T)_{ij}}{(USZ^TP^TPZS^T)_{ij}} \tag{25}$$

$$s_{ij} \leftarrow s_{ij}\frac{(U^TXPZ)_{ij}}{(U^TUSZ^TP^TPZ)_{ij}} \tag{26}$$

$$p_{ij} \leftarrow p_{ij}\frac{(X^TUSZ^T + \lambda W_{ab}PZZ^T)_{ij}}{(PZS^TU^TUSZ^T + \lambda D_{ab}PZZ^T)_{ij}} \tag{27}$$

$$z_{ij} \leftarrow z_{ij}\frac{(P^TX^TUS + \lambda P^TW_{ab}PZ)_{ij}}{(P^TPZS^TU^TUS + \lambda P^TD_{ab}PZ)_{ij}} \tag{28}$$

According to the above iteration rules and the similarity matrix construction process, we summarize the process of ABNMTF algorithm, as shown in Algorithm 1.

Algorithm 1 ABNMTF algorithm

Input: The dataset X, labeled data l, graph regularization parameter λ, cluster number c, the number of neighbors to determine the initial graph K;
Output: The new data representation $V = PZ$;
1:Normalize the data matrix X and Initialize U and V;
2:According to the label information, the label constraint matrix P and the block diagonal matrix M are constructed;
3:Construct similarity matrix W_a based on adaptive neighbors local structure graph;
4:Reconstruct W_{ab} according to the information of M matrix by using (17);
5:**repeat**
6: Update U by using (25);
7: Update S by using (26);
8: Update P by using (27);
9: Update Z by using (28);
10:**until** Algorithm convergence
11:return $V = PZ$

3.5. Convergence analysis of ABNMTF

In this subsection, we rationalize the use of auxiliary functions to prove that the objective function (18) is monotonically decreasing under the update rules (25)–(28). Firstly, we analyze that the objective function is monotonically non-increasing in the case of update rule (28).

Definition 4.1. If the following conditions are satisfied:

$$G(z, z') \geq F(z), G(z, z) = F(z) \tag{29}$$

$G(z, z')$ is an auxiliary function for $F(z)$.

Lemma 4.1. *If $G(z, z')$ is an auxiliary function of $F(z)$, then $F(z)$ is nonincreasing under the following update steps from iteration t to $t+1$:*

$$z^{t+1} = \arg\min_z G(z, z^t) \tag{30}$$

Proof. From Lemma 4.1, it can be easily obtained that when $z = z^t$, $G(z, z')$ can obtain a local minimum. From Definition 4.1, the following inequality $G(z^{t+1}, z^t) \geq F(z^{t+1})$ can be derived:

$$F(z^{t+1}) \leq G(z^{t+1}, z^t) \leq G(z^t, z^t) = F(z^t) \tag{31}$$

The first-order derivative and second-order derivative on $F(z)$ are:

$$F'_{z_{ij}}(z_{ij}) = \begin{pmatrix} -2P^TX^TUS + 2P^TPZS^TU^TUS \\ +2\lambda P^TD_{ab}PZ - 2\lambda P^TW_{ab}PZ \end{pmatrix}_{ij} \tag{32}$$

$$\begin{aligned} F''_{z_{ij}}(z_{ij}) &= 2(P^TP)_{ii}(U^TUS)_{jj} \\ &+2\lambda(P^TD_{ab}P)_{ii} - 2\lambda(P^TW_{ab}P)_{ii} \end{aligned} \tag{33}$$

The Taylor expansion of $F(z)$ can be expressed as:

$$\begin{aligned} F_{z_{ij}}(z_{ij}) &= F_{z_{ij}}(z^t_{ij}) + F'_{z_{ij}}(z^t_{ij})(z - z^t_{ij}) + \frac{1}{2}F''_{z_{ij}}(z^t_{ij})(z - z^t_{ij})^2 \\ &= F_{z_{ij}}(z^t_{ij}) + F'_{z_{ij}}(z^t_{ij})(z - z^t_{ij}) \\ &+\frac{(P^TPZS^TU^TUS+\lambda P^TD_{ab}PZ)_{ij}}{z^t_{ij}}(z - z^t_{ij})^2 \end{aligned} \tag{34}$$

To ensure that $G(z, z^t_{ij}) \geq F_{z_{ij}}(z_{ij})$, we obtain by using (33) and (34).

$$\begin{aligned} &\frac{(P^TPZS^TU^TUS+\lambda P^TD_{ab}PZ)_{ij}}{z^t_{ij}} \geq (P^TP)_{ii}(U^TUS)_{jj} \\ &+\lambda(P^TD_{ab}P)_{ii} - \lambda(P^TW_{ab}P)_{ii} \end{aligned} \tag{35}$$

so that $G(z, z^t_{ij}) \geq F_{z_{ij}}(z_{ij})$. Based on Lemma 4.1, it can be proved as follows: Taking Eq. (34) as the objective function and combining with formula (30), it can be concluded that the update rule of objective function Z is:

$$\begin{aligned} z^{t+1}_{ij} &= z^t_{ij} - z^t_{ij}\frac{F'_{z_{ij}}(z^t_{ij})}{(2P^TPZS^TU^TUS+2\lambda P^TD_{ab}PZ)_{ij}} \\ &= u_{ij}\frac{(P^TX^TUS+\lambda P^TW_{ab}PZ)_{ij}}{(P^TPZS^TU^TUS+\lambda P^TD_{ab}PZ)_{ij}} \end{aligned} \tag{36}$$

Obviously, Eq. (36) is equivalent to Eq. (28), which also means that the objective function is monotonically decreasing and convergent under the update rule (20). The proof of Theorem 4.1 is completed.

With the same method, we can prove $G(u, u) = F(u)$, $G(s, s) = F(s)$ and $G(p, p) = F(p)$ decreases monotonically in the iterative updating rules (25), (26) and (27).

3.6. Computational complexity of ABNMTF

In this section, we analyze the computational complexity of the ABNMTF model, because it is theoretically important for the evaluation of algorithm performance. We use the commonly used representation method big O to express the complexity of the proposed method.

Table 1
Comparison of computational complexity of each algorithm.

	Model	Overall cost	Number of graph regularization
1	NMF (Lee and Seung, 1999)	$O(tmnc)$	0
2	GNMF (Cai et al., 2010)	$O\left(tmnc + mn^2\right)$	1
3	CDONMTF (Ge et al., 2019)	$O\left(tmnc + mn^2 + m^2n\right)$	2
4	PAMGNMF (Shu et al., 2017)	$O(tmnc + dmn^2)$	d
5	SODNMF (Meng et al., 2018)	$O\left(tmnc + mn^2 + m^2n\right)$	2
6	CSNMF (Peng et al., 2021a)	$O\left(tmnc + mn^2 + cn^2\right)$	2
7	GSSNMF (Jia et al., 2019)	$O\left(tmnc + 2mn^2\right)$	2
8	ABNMTF	$O(tmnc + mn^2)$	1

Table 2
Benchmark dataset.

Datasets	# Samples	# Dimensions	# Classes	Type
FEI	600	768	50	Face
JAFFE	213	676	10	Face
PIE	2856	1024	68	Face
UMIST	575	644	20	Face
COIL20	1440	1024	20	Object
COIL100	7200	1024	100	Object
Optdigits	5620	64	10	Handwritten
MNIST	6996	784	10	Handwritten

Construct adaptive neighbors similarity graph structure requires $O(n^2m)$, the cyclic multiplication update complexity of the model is $O(tmnc)$. When ABNMTF stops after t iterations and c is the number of clusters, the overall cost of ABNMTF is $O(tmnc + mn^2)$.

According to Lee and Seung (1999), Cai et al. (2010), Ge et al. (2019), Shu et al. (2017), Meng et al. (2018), Peng et al. (2021a), Jia et al. (2019), the overall computational costs of NMF, GNMF, CDONMTF, PAMGNMF, SODNMF, CSNMF, GSSNMF, ABNMTF and the number of graph regularization are summarized in Table 1, where d denotes the number of similarity matrices constructed in the PAMGNMF and $d \geq 2$. We can clearly see that the method proposed in this paper is more complex than the NMF model, equal to the GNMF model, but simpler than other mixed-constrained NMF models. It is worth noting that the complexity of the proposed model is better than that of PAMGNMF, which is famous for its few parameters and low complexity.

4. Experiments

In this section, we evaluate the effectiveness of the proposed method from multiple aspects, including the visualization of the similarity matrix W_{ab}, parameter analysis, evaluation of the proportion of label information, convergence experiment, cluster evaluation (numerical comparison and visual comparison). In order to reflect the fairness of the comparative experiment, the operating environment of ABNMTF algorithm is the same as that of other NMF algorithms.

4.1. Experimental settings

4.1.1. Datasets

In the clustering experiment, 8 real public datasets were used to verify the effectiveness of the ABNMTF method. Specifically, FEI, JAFFE, PIE, and UMIST are face image datasets, COIL20 and COIL100 are object image datasets, and MNIST and Optdigits are handwritten image datasets. Some examples of these datasets are visually shown in Fig. 2. In addition, each dataset is formatted as an $m \times n$ size matrix, where m is the dimension of each image and n is the number of samples in the dataset, and then the normalization process is performed uniformly. The detailed characteristics of these datasets can be found in Table 2.

4.1.2. Evaluation metrics

In the experimental comparison, we use clustering accuracy (ACC), normalized mutual information (NMI) and cluster purity (Cai et al., 2010; Peng et al., 2020, 2021a) as three evaluation metrics to measure the clustering results of all clustering methods.

4.1.3. Comparing methods

In order to reflect the effectiveness of ABNMTF subspace learning, we compared it with the benchmark algorithm and several state-of-the-art NMF variants, including NMF (Lee and Seung, 1999), GNMF (Cai et al., 2010), CNMF (Liu et al., 2011), SODNMF (Meng et al., 2018), PAMGNMF (Shu et al., 2017), CSNMF (Peng et al., 2021a), MSNMF (Zhang et al., 2021), GSSNMF (Jia et al., 2019), RSNMF (Li et al., 2017), TRNMTF (Deng et al., 2021), PNMTF (Chen et al., 2022). We briefly describe these methods as follows:

- **NMF** is the basic NMF model.
- **GNMF** constructs an affinity graph to encode geometric information, which successfully improves the non-linear space learning ability.
- **CNMF** uses the label information as an additional constraint to guide the factorization of high-dimensional data to lower dimensions.
- **SODNMF** is an excellent semi-supervised mixed-constrained NMF method, which combines constraints such as dual graph regularization, label constraint, bi-orthogonal and sparseness.
- **PAMGNMF** uses a linear combination of multiple graphs to replace the previous manifold method to approximate the manifold structure of the data, and it can automatically learn the optimal weight of each graph.
- **CSNMF** uses correlation entropy to measure the loss of non-negative matrix factorization, and retains dual graph regularization and biorthogonal constraints in the model.
- **MSNMF** combines the label information into multi graph regularization in the form of pairwise constraints, and the pairwise constraints spread to all samples.
- **GSSNMF** impose restrictions on both the similarity and dissimilarity of the low-dimensional representations of data samples with labels as well as a small number of unlabeled ones, and incorporates them into traditional NMF to guide factorization.
- **RSNMF** cleverly uses the label information to construct a structured matrix and embed it in the same constrained subspace.
- **TRNMTF** combines graph regularization, Frobenius norm and L_1-norm to optimize the objective function simultaneously. It can perform feature selection well, enhance the sparsity of the model, adjust the eigenvalues in the low-dimensional matrix.
- **PNMTF** is based on the NMTF model and is capable of updating the factor matrix in parallel while processing in a distributed memory system.

4.1.4. Other settings

In order to perform clustering comparison experiments fairly, all semi-supervised NMF models randomly extract 10% of images and their label information from each category in the original data. Moreover, the best parameters of all comparison algorithms are derived from the settings in the original paper. For all dataset experiments, we take the maximum number of categories of the dataset as the clustering number c, which means that if the input data of all comparative experiments are consistent, the maximum number of iterations t of each algorithm is 100, all methods run 20 times independently, calculate the average value of all evaluation measures as the final experimental results, and display the best results in bold.

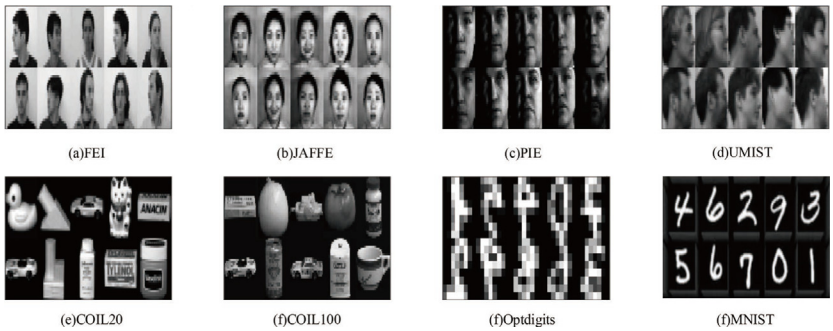


Fig. 2. Examples selected from the dataset.

Table 3
Clustering results on eight datasets.

Datasets	ACC(%)											
	NMF	GNMF	CNMF	SODNMF	PAMGNMF	CSNMF	MSNMF	GSSNMF	RSNMF	TRNMTF	PNMTF	ABNMTF
FEI	40.75	58.70	55.38	64.81	63.17	68.02	70.03	65.63	59.75	68.12	43.40	**74.08**
JAFFE	68.12	73.18	72.01	93.27	89.73	91.66	95.21	95.58	85.47	93.37	75.14	**98.50**
PIE	35.22	75.44	65.12	80.13	72.24	84.75	82.98	77.80	74.63	83.17	49.03	**87.64**
UMIST	50.97	63.28	61.33	81.74	53.32	70.43	69.11	74.52	66.87	79.54	55.00	**88.42**
COIL20	53.26	74.27	71.82	81.23	75.19	83.09	82.78	78.17	68.17	78.90	61.33	**96.24**
COIL100	38.67	53.17	51.34	73.23	55.38	69.36	68.43	59.03	49.33	74.19	52.24	**81.01**
Optdigits	55.02	71.04	61.71	86.71	81.29	88.21	**98.46**	74.01	74.26	90.11	68.71	97.26
MNIST	38.76	53.67	58.20	79.00	63.41	79.63	85.55	54.57	59.76	77.22	54.70	**87.93**
Datasets	NMI(%)											
	NMF	GNMF	CNMF	SODNMF	PAMGNMF	CSNMF	MSNMF	GSSNMF	RSNMF	TRNMTF	PNMTF	ABNMTF
FEI	66.12	77.84	59.12	78.52	75.01	76.13	78.89	76.98	70.20	74.14	69.90	**88.04**
JAFFE	71.84	80.99	78.90	91.89	90.88	93.75	96.30	93.38	83.72	93.77	75.53	**97.94**
PIE	63.67	88.17	62.71	81.12	84.02	91.81	91.47	83.80	87.63	84.41	69.81	**97.64**
UMIST	63.06	68.07	65.30	84.79	70.86	82.13	81.36	77.05	78.59	82.11	68.73	**93.31**
COIL20	66.27	86.98	80.11	84.02	85.37	90.75	90.32	89.96	77.31	81.70	73.40	**98.44**
COIL100	65.81	55.77	57.03	78.62	75.13	86.54	85.95	61.13	76.35	81.19	69.77	**95.07**
Optdigits	54.46	72.91	62.47	88.70	84.69	90.57	**96.08**	74.01	78.63	86.12	60.10	94.57
MNIST	30.73	60.17	62.10	77.84	62.26	80.82	**82.10**	63.82	68.32	76.35	60.72	79.56
Datasets	Purity(%)											
	NMF	GNMF	CNMF	SODNMF	PAMGNMF	CSNMF	MSNMF	GSSNMF	RSNMF	TRNMTF	PNMTF	ABNMTF
FEI	46.04	60.73	51.90	65.07	65.15	70.25	73.29	71.13	62.84	69.13	51.03	**77.36**
JAFFE	70.02	75.01	72.74	92.75	89.90	92.37	92.80	95.84	83.66	93.30	76.43	**98.50**
PIE	40.14	77.10	58.11	83.63	84.11	86.78	91.58	82.91	80.04	84.15	53.43	**90.70**
UMIST	53.00	62.94	60.80	82.17	71.59	82.72	80.05	75.07	76.78	81.12	58.11	**89.69**
COIL20	55.31	71.81	68.82	83.73	85.39	86.72	90.97	75.00	80.31	84.33	59.17	**96.93**
COIL100	41.04	52.95	54.22	80.12	71.20	70.65	79.93	55.53	64.25	79.37	51.78	**84.87**
Optdigits	56.78	70.05	60.81	84.09	92.03	87.10	96.43	74.52	75.65	89.13	62.72	**97.26**
MNIST	38.24	54.47	53.70	80.02	73.93	77.70	**89.37**	56.84	69.38	84.13	43.12	87.93

4.2. Clustering comparison

4.2.1. Numerical comparison

In order to highlight the effectiveness of ABNMTF, we use the evaluation indicators mentioned above to measure the quality of clustering. The experimental results of all comparison methods on 8 image datasets (FEI, JAFFE, PIE, UMIST, COIL20, COIL100, Optdigits, MNIST) as shown in Table 3, the best performance of each category is highlighted in bold.

It can be seen from the data shown in Table 3 that ABNMTF achieves the best clustering performance on multiple datasets, which shows that ABNMTF's subspace learning ability is better than other methods. In the comparison of the above data, the following phenomena can also be observed:

(1) In all face datasets and object datasets, ABNMTF model is significantly ahead of other models. Specifically, on the face dataset UMIST, compared to the second best algorithm, ABNMTF improves the clustering evaluation metrics by 8.2%, 10% and 8.4%. On the object dataset COIL20, compared to other algorithms, ABNMTF is improved by at least 15.8%, 8.5% and 6.6%.

(2) For handwritten datasets, the overall performance of ABNMTF is slightly weaker than that of MSNMF, which may be that the adaptive weight method of MSNMF has better matching on handwritten datasets. However, as an NMF model with a small number of adjustment parameters, ABNMTF is still better than other mixed constraint NMF models, such as SODNMF, RSNMF, etc.

(3) A very interesting phenomenon is that most NMF variant algorithms explore the fusion of Manifold Regularization. In fact, the NMF model with manifold learning has better general performance, which also proves the significance of our reconstruction research on the similarity matrix of manifold regularization term.

Based on the above clustering comparison experiment, we deeply explored the superiority of ABNMTF from two aspects:

(1) In order to study the effectiveness of similarity matrix construction, Fig. 3 shows the similarity matrix construction methods used by different NMF algorithms. To better illustrate the difference, we mark the highlighted area with a red rectangle at the same position in each figure. From the comparison of the above red matrices, we can see that the reconstruction method based on label information can help the similarity matrix obtain exact connection attributes, so as to obtain more correct connections. In addition, the similarity matrix

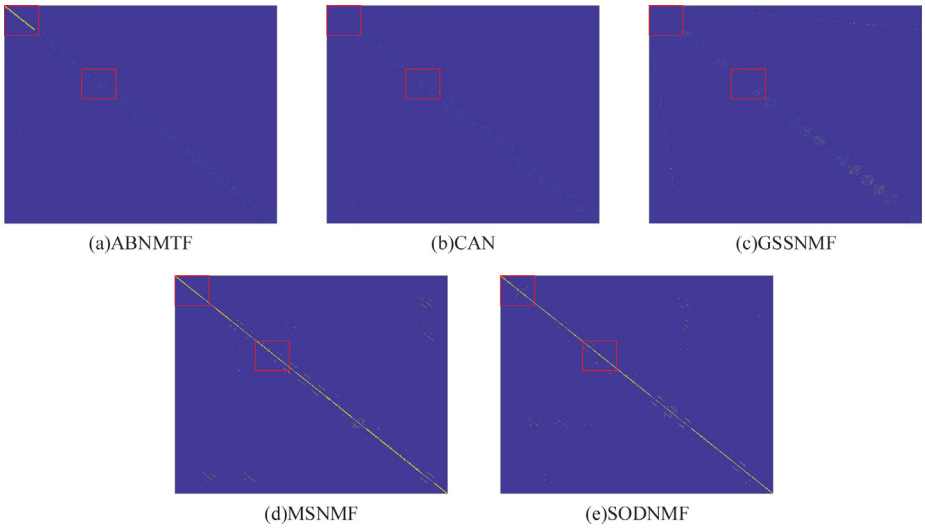


Fig. 3. Similarity matrix visualization comparison.

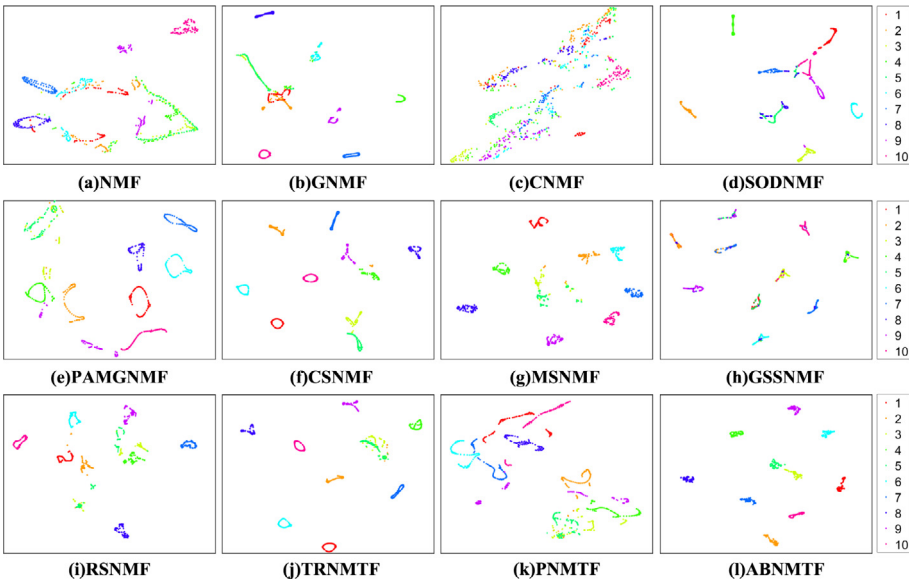


Fig. 4. 2-D representations of COIL20 dataset using t-SNE on the results of different methods.

learning method based on adaptive neighbors can obtain more standard diagonal connection methods. Compared with GSSNMF, the diagonal connection of the similarity matrix of our proposed method will be more balanced (without dense and sparse). Compared with MSNMF and SODNMF, our proposed method can produce more correct connections and reduce noise points.

(2) The effective use of label information to construct more constraints is particularly important for semi supervised methods. This paper uses label information very efficiently, constructs label constraint matrix and reconstructs accurate similarity matrix respectively, improves the discrimination of the method, and has strong popularization value.

4.2.2. Visualization comparison

In order to intuitively reflect the subspace learning ability of ABNMTF, we intuitively project several NMF models and the low-dimensional data learned by ABNMTF into a two-dimensional space with t-SNE. In order to show the comparison results of cluster visualization more clearly, we have selected three datasets COIL20, UMIST, and MNIST for comparison. Among them, 10 types of data are intercepted from the COIL20 and UMIST datasets as input data. The results of the visual comparison are shown in Figs. 4–6.

Figs. 4–6 shows that the low dimensional subspace expression learned by ABNMTF algorithm can better distinguish the gap between data samples, especially in the visual comparison of COIL20 dataset, ABNMTF can distinguish sample clusters, while other methods cannot perfectly separate samples from other clusters. The same conclusion can also be observed from the comparison of the other two groups. This method not only strengthens the reliability of the above clustering data comparison experiment, but also verifies that the contribution contained in ABNMTF can improve the learning ability of algorithm subspace.

4.3. Ablation experiment

In the ABNMTF model, only one regularizer is used, that is, the adaptive neighbors block diagonal graph regularizer, so the regularization parameters cannot be simply adjusted to reflect the effectiveness of the method proposed in this paper. We use two methods: model degradation and regularization parameter adjustment to discuss the ablation experiment in this section.

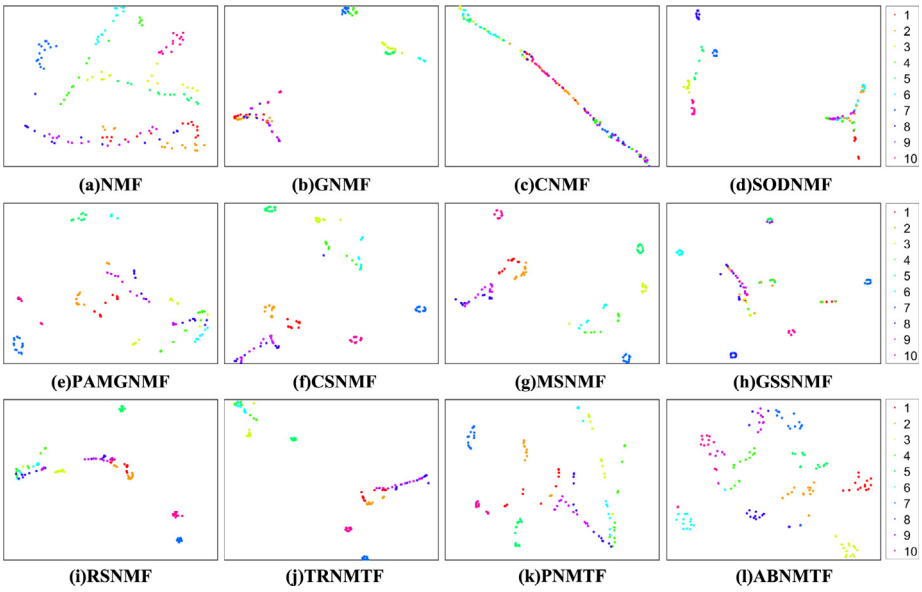


Fig. 5. 2-D representations of UMIST dataset using t-SNE on the results of different methods.

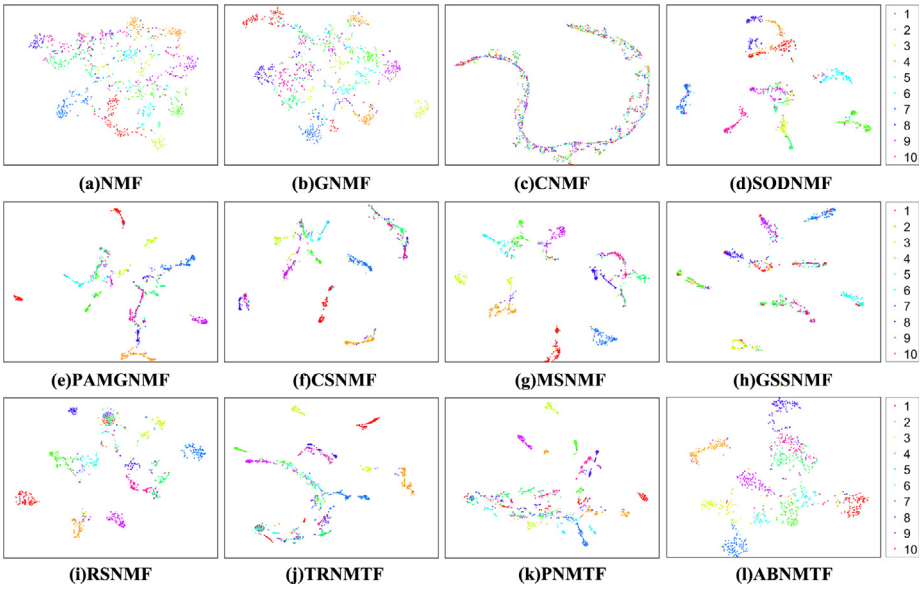


Fig. 6. 2-D representations of MNIST dataset using t-SNE on the results of different methods.

Table 4
Clustering results under different settings (ACC% and NMI%).

	Metric	FEI	JAFFE	PIE	UMIST	COIL20	COIL100	Optdigits	MNIST
NMTF	ACC	42.18	70.01	33.94	49.72	54.12	39.72	58.98	36.61
	NMI	64.10	72.33	60.17	63.36	65.91	66.77	59.00	32.45
Case "1"	ACC	55.97	72.12	48.32	59.08	67.13	55.61	65.12	54.00
	NMI	59.44	75.40	59.17	63.17	70.90	63.62	68.35	56.48
Case "2"	ACC	68.25	92.63	85.05	71.20	81.87	68.27	89.14	80.17
	NMI	78.21	94.09	91.27	83.17	90.02	84.12	92.75	78.24
Case "3"	ACC	71.45	95.42	84.24	83.28	91.20	76.88	93.34	85.12
	NMI	85.20	95.07	91.07	90.72	93.09	86.57	90.17	76.51
ABNMTF	ACC	**74.08**	**98.50**	**87.64**	**88.42**	**96.24**	**81.01**	**97.26**	**87.93**
	NMI	**88.04**	**97.94**	**97.64**	**93.31**	**98.44**	**95.07**	**94.57**	**79.56**

When the model loses all improvements, the model degenerates to NMTF; When λ=0, the model will lose the adaptive neighbors and block diagonal regularization, but retains the three-factor factorization and label constraint matrix, which is marked as "1"; When the model loses the block diagonal matrix reconstruction, it is marked as "2"; When the model does not construct the label constraint matrix, but retains the regularization of the adaptive neighbors block diagonal graph, it is marked as "3". Table 4 shows the ACC and NMI of the proposed model

under various settings. From Table 4, the following conclusions can be drawn:

(1) In all comparisons, ABNMTF produced the highest ACC and NMI, which shows the effectiveness of this paper on the similarity matrix reconstruction.

(2) Case "3" produced the second highest clustering index, which illustrates the importance of manifold regularization for the NMF model. Even if the manifold regularizer is kept alone, the model performance is still very strong.

(3) Case "1" only constructs the label constraint matrix, but it still performs better than NMTF, which shows that the hard constraint of labeled data has a very positive meaning for NMF.

(4) Comparing case "1", case "2" and case "3" comprehensively, it is obvious that the adaptive neighbor block diagonal regularizer improves the algorithm the most, followed by the label constraint method. However, all the methods are non-linearly added to ABNMTF, and the final result further confirms this conclusion.

4.4. Parameters sensitivity

How to select the optimal parameter combination of the algorithm has always been a difficult problem in machine learning. This paper explores the graph structure of NMF model and puts forward the corresponding improvement scheme. Therefore, only two adjustable parameters, λ and K, are retained in ABNMTF. Next, we discuss how to select the optimal parameters on eight experimental datasets, and analyze and obtain a simple method to deal with the parameter selection problem of most scenarios.

The adaptive neighbors block diagonal graph regularization parameter λ selected from {0,0.001,0.01,0.1,1,10,100} and the number of neighbors to determine the initial graph parameter K selected from {3,5,7,9,11,13,15}. Since the clustering accuracy can more intuitively reflect the clustering effect, in the parameter selection comparison, the comparison experiment of the parameters with ACC as the evaluation system is shown in Fig. 7.

Fig. 7 vividly shows the clustering performance of different parameter combinations on the experimental dataset. The higher the histogram and the whiter the color, the better the performance of the parameter combination, and the better performance area has adjacency. Specifically, the K value determines the number of neighbors of the initial graph and can adjust the final similarity graph structure, but reflected in the clustering performance, the adjustment ability of K value is weak, which also shows that the adaptive neighbors similarity graph structure proposed in this paper can better deal with the number of p-nn neighbors in the graph structure. As an adjustable parameter, λ determines the strength of graph regularization. From the above comparison results, it can be seen that a moderate graph regularization can greatly improve the performance of the algorithm and the value range is generally between 0.1 and 1.

As a semi-supervised algorithm, ABNMTF will naturally be affected by the ratio of training samples. Theoretically, the more training samples, the better the algorithm performance, but in fact, it is necessary to balance the balance between the label size and the algorithm performance. {10%,20%,30%,40%,50%,60%} label information for each class was successively used to generate pairwise constraints, and the results are recorded as the mean values over 10 repetitions for each case. Fig. 8 shows the curves in terms of clustering performance and the percentage of data labels for different algorithms. It should be noted that in this experiment, the settings of other parameters refer to the optimal parameters discussed above as input.

From Fig. 8, we can draw the following conclusions.

(1) The clustering quality of low dimensional data learned by the algorithm on all datasets is not positively correlated with the learning rate. 10% of the training samples can properly reflect the advantages of the algorithm, its performance is excellent, and the number of labels to be labeled is less.

(2) It can be observed that the clustering performance of different semi-supervised algorithms on the dataset is similar. For example, on the COIL100 dataset, the clustering accuracy of all semi-supervised algorithms decreases as the proportion of training sample ratio increases. Furthermore, the current mainstream semi supervised NMF algorithms are all based on label information to construct hard constraint matrix, which is not really learning the existing features of label information, but using known label information to construct hard constraint matrix to induce the NMF algorithm to reduce dimensions, which is disadvantageous to the high proportion of label NMF dimension reduction.

(3) The clustering accuracy of ABNMTF outperforms other advanced semi-supervised NMF algorithms in the vast majority of scenarios, which also illustrates the effectiveness of ABNMTF in label information utilization.

4.5. Convergence experiments

In order to prove Theorem 1 from an experimental point of view, we illustrate the convergence of the model through experiments. Fig. 9 describes the actual convergence curve of GNMF, SODNMF and ABNMTF on eight datasets. The X-axis represents the number of iterations, and the Y-axis represents the value of the objective function.

It can be easily seen from Fig. 9 that the model has achieved smooth and fast convergence on eight datasets, which also shows that the objective function can converge effectively under the multiplication iteration rules proposed in this paper, and explains the correctness of the proof of Theorem 1. Specifically, ABNMTF can quickly converge on the PIE, COIL20 and JAFFE datasets, and it stabilizes around 10 generations. It can also achieve stable convergence within 40 generations on other datasets. Compared with GNMF and SODNMF, ABNMTF has faster convergence speed and lower convergence limit, which also means that the convergence quality of the proposed method is better.

4.6. Running time

In this subsection, we designed a set of fair experiments to compare the time consuming of each algorithm on six datasets, and run them independently for 20 times. All codes are implemented on MATLAB 2018B and run on Intel Core I7-9700K and 16G memory. The final time-consuming average value was used as the comparison value. The detailed comparison data is shown in Table 5.

Table 5 shows that the running time of all methods is consistent with the computational complexity analysis. Theoretically, NMTF with the lowest computational complexity has the fastest computing speed. ABNMTF method has very excellent computational complexity, so its computing time is still very fast in all comparison algorithms. The PNMTF algorithm runs most efficiently among all comparisons, which benefits from the parallel update rules in the model, but the algorithm lacks the ability to uncover the implicit structural information in the original data leading to weak subspace learning. It is worth noting that the NMF method based on manifold regularization is still a hot research topic. It brings excellent subspace learning capabilities while not significantly increasing the computational complexity of the algorithm.

5. Conclusions

This paper proposes a novel high-performance semi-supervised algorithm called semi-supervised non-negative matrix tri-factorization with adaptive neighbors and block-diagonal (ABNMTF). This method cleverly uses the label information to construct a block diagonal structure to reconstruct the adaptive neighbor graph similarity matrix, and efficiently realizes the dissemination and utilization of label information. From the perspective of algorithm model, ABNMTF has achieved high performance and high operating efficiency of the algorithm through only one adjustable regularization parameter, which

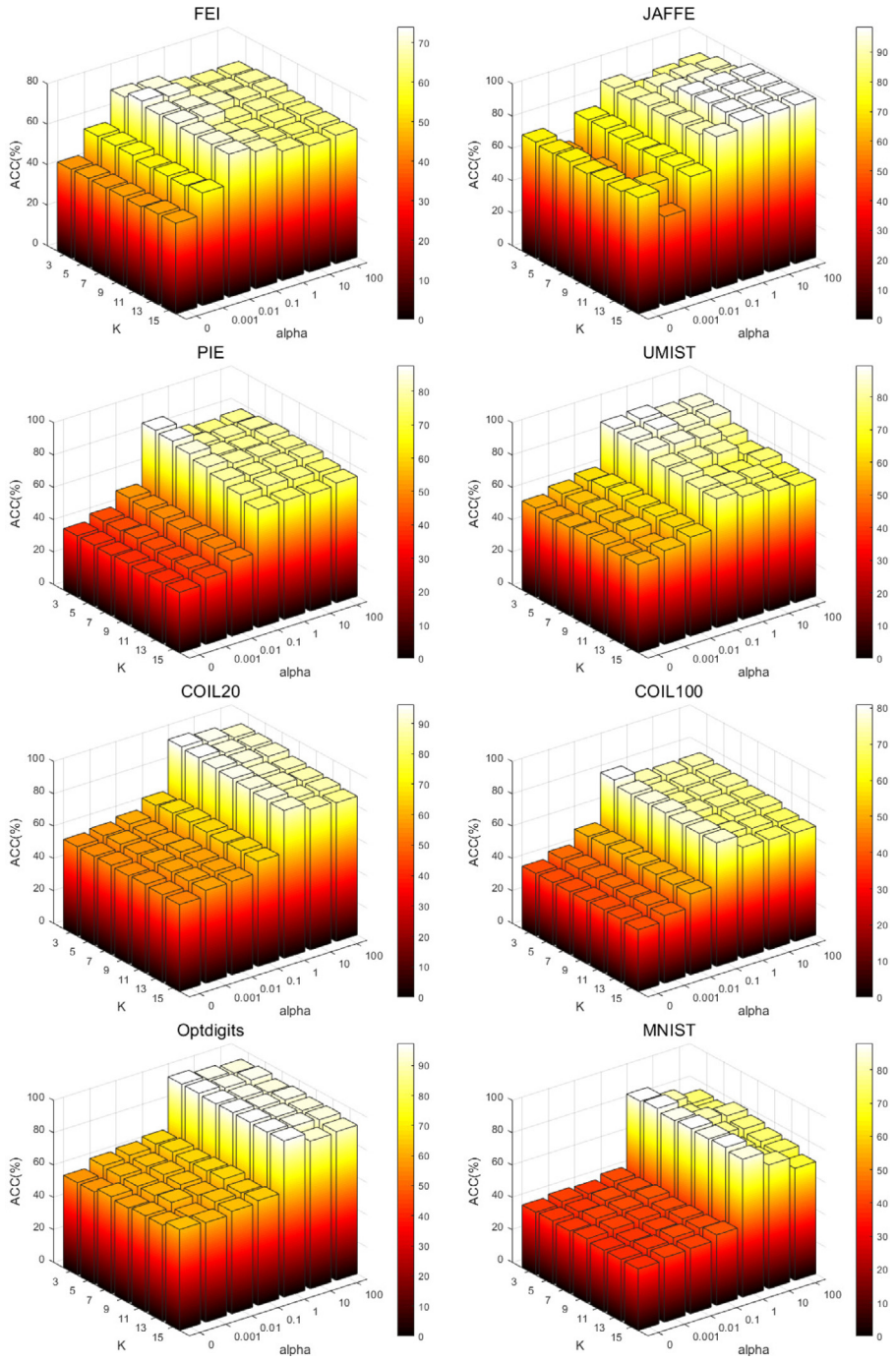


Fig. 7. Parameter comparison.

Table 5
Running time comparison of all NMF methods (seconds).

Datasets	NMF	GNMF	CNMF	SODNMF	PAMGNMF	CSNMF	MSNMF	GSSNMF	RSNMF	TRNMTF	PNMTF	ABNMTF
FEI	3.67	4.51	3.82	15.53	10.82	7.98	13.25	8.83	6.72	19.31	2.95	4.99
JAFFE	0.38	0.48	0.41	2.30	1.41	1.27	1.83	1.32	0.94	3.08	0.35	0.52
PIE	133.02	177.25	152.07	490.77	383.15	300.83	248.10	312.47	257.00	597.70	105.33	196.23
UMIST	1.33	1.68	1.40	5.51	3.92	2.96	4.77	2.98	2.50	7.01	1.20	1.79
COIL20	8.83	12.18	10.08	38.10	27.89	22.14	29.74	28.10	22.27	42.75	7.41	13.94
COIL100	2824.06	3452.22	3101.27	11872.04	8004.27	5867.11	10821.94	5607.11	5017.84	13572.19	2475.39	3878.69
Optigits	0.13	0.16	0.15	0.41	0.28	0.23	0.32	0.28	0.23	0.50	0.11	0.17
MNIST	36.04	43.87	39.41	169.79	97.10	61.99	119.12	63.31	58.61	184.43	31.77	47.17

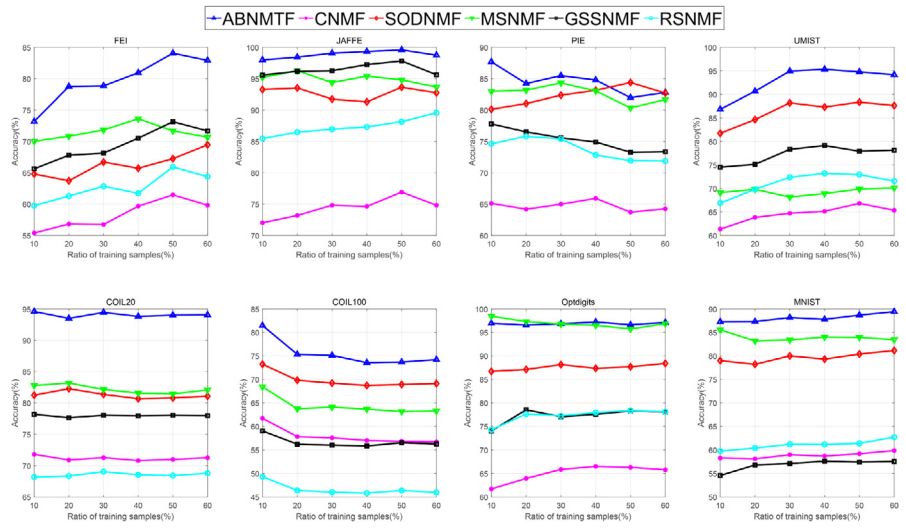


Fig. 8. Clustering accuracy of different training sample ratio.

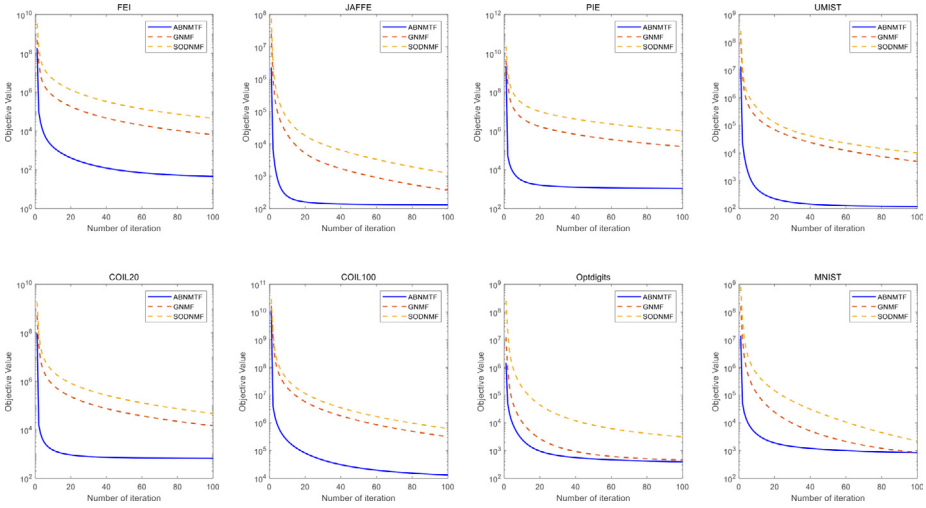


Fig. 9. Convergence curves of ABNMTF on eight datasets.

has excellent promotion significance. Different experiments on multiple datasets demonstrate the effectiveness of ABNMTF, moreover, the similarity matrix learned by ABNMTF is analyzed from a visual perspective to explore the significance of correct correlation between samples for subspace learning.

Although ABNMTF method performs well in clustering tasks, it can be further improved in two possible directions in the future. First of all, the semi-supervised NMF model is still a conservative learning method, which also limits the application of the NMF method. In the future, research work should be carried out in the direction of weakly supervised NMF or self-supervised NMF. Secondly, the NMF method is still a single-task method, and it is necessary to explore a more efficient multi-task NMF method.

CRediT authorship contribution statement

Songtao Li: Conceptualization, Methodology, Software, Writing – original draft. **Weigang Li:** Writing – review & editing. **Hao Lu:** Visualization. **Yang Li:** Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was funded by the National Natural Science Foundation of China,under Grants 51774219, Key R&D Projects in Hubei Province under Grants 2020BAB098 and Science and Technology project of Hubei Province under Grants 2020BAB098.

References

Abdi, H., Williams, L.J., 2010. Principal component analysis. Wiley Interdiscip. Rev. Comput. Stat. 2 (4), 433–459.

Belkin, M., Niyogi, P., 2008. Towards a theoretical foundation for Laplacian-based manifold methods. J. Comput. System Sci. 74 (8), 1289–1308.

Cai, D., He, X., Han, J., Huang, T.S., 2010. Graph regularized nonnegative matrix factorization for data representation. IEEE Trans. Pattern Anal. Mach. Intell. 33 (8), 1548–1560.

Chen, Y., Lei, Z., Rao, Y., Xie, H., Wang, F.L., Yin, J., Li, Q., 2022. Parallel non-negative matrix tri-factorization for text data co-clustering. IEEE Trans. Knowl. Data Eng..

Deng, P., Li, T., Wang, H., Horng, S.-J., Yu, Z., Wang, X., 2021. Tri-regularized nonnegative matrix tri-factorization for co-clustering. Knowl.-Based Syst. 226, 107101.

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

Dong, L., Yuan, Y., Luxs, X., 2020. Spectral–spatial joint sparse NMF for hyperspectral unmixing. IEEE Trans. Geosci. Remote Sens. 59 (3), 2391–2402.

Fortunato, S., 2010. Community detection in graphs. Phys. Rep. 486 (3–5), 75–174.

Ge, S., Li, H., Luo, L., 2019. Constrained dual graph regularized orthogonal nonnegative matrix tri-factorization for co-clustering. Math. Probl. Eng. 2019 (1), 1–17.

He, C., Fei, X., Cheng, Q., Li, H., Hu, Z., Tang, Y., 2021. A survey of community detection in complex networks using nonnegative matrix factorization. IEEE Trans. Comput. Soc. Syst. 1–18. http://dx.doi.org/10.1109/TCSS.2021.3114419.

Jia, Y., Kwong, S., Hou, J., Wu, W., 2019. Semi-supervised non-negative matrix factorization with dissimilarity and similarity regularization. IEEE Trans. Neural Netw. Learn. Syst. 31 (7), 2510–2521.

Jiao, C.-N., Gao, Y.-L., Yu, N., Liu, J.-X., Qi, L.-Y., 2020. Hyper-graph regularized constrained NMF for selecting differentially expressed genes and tumor classification. IEEE J. Biomed. Health Inf. 24 (10), 3002–3011.

Lee, D.D., Seung, H.S., 1999. Learning the parts of objects by non-negative matrix factorization. Nature 401 (6755), 788–791.

Li, S., Li, W., Hu, J., Li, Y., 2021. Semi-supervised bi-orthogonal constraints dual-graph regularized NMF for subspace clustering. Appl. Intell. 1–22.

Li, Z., Tang, J., He, X., 2017. Robust structured nonnegative matrix factorization for image representation. IEEE Trans. Neural Netw. Learn. Syst. 29 (5), 1947–1960.

Li, Z., Wu, X., Peng, H., 2010. Nonnegative matrix factorization on orthogonal subspace. Pattern Recognit. Lett. 31 (9), 905–911.

Liu, H., Wu, Z., Li, X., Cai, D., Huang, T.S., 2011. Constrained nonnegative matrix factorization for image representation. IEEE Trans. Pattern Anal. Mach. Intell. 34 (7), 1299–1311.

Lu, C., Feng, J., Lin, Z., Mei, T., Yan, S., 2018. Subspace clustering by block diagonal representation. IEEE Trans. Pattern Anal. Mach. Intell. 41 (2), 487–501.

Lydia, E.L., Kumar, P.K., Shankar, K., Lakshmanaprabu, S., Vidhyavathi, R., Maseleno, A., 2020. Charismatic document clustering through novel K-means non-negative matrix factorization (KNMF) algorithm using key phrase extraction. Int. J. Parallel Program. 48 (3), 496–514.

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

Min, W., Xu, T., Wan, X., Chang, T.-H., 2022. Structured sparse non-negative matrix factorization with l2, 0-norm. IEEE Transactions on Knowledge and Data Engineering 1–13. http://dx.doi.org/10.1109/TKDE.2022.3206881.

Newman, M.W., Libraty, N., On, O., On, K.A., On, K.A., 1991. The Laplacian spectrum of graphs. Graph Theory Comb. Appl. 18 (7), 871–898.

Nie, F., Wang, X., Huang, H., 2014. Clustering and projected clustering with adaptive neighbors. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 977–986.

Peng, S., Ser, W., Chen, B., Lin, Z., 2020. Robust orthogonal nonnegative matrix tri-factorization for data representation. Knowl.-Based Syst. 201, 106054.

Peng, S., Ser, W., Chen, B., Lin, Z., 2021a. Robust semi-supervised nonnegative matrix factorization for image clustering. Pattern Recognit. 111 (3), 107683.

Peng, C., Zhang, Z., Kang, Z., Chen, C., Cheng, Q., 2021b. Nonnegative matrix factorization with local similarity learning. Inform. Sci. (ISSN: 0020-0255) 562, 325–346.

Roweis, S.T., Saul, L.K., 2000. Nonlinear dimensionality reduction by locally linear embedding. Science 290 (5500), 2323–2326.

Sh, A., Zxa, B., Zhao, K.A., Yr, A., 2020. Regularized nonnegative matrix factorization with adaptive local structure learning. Neurocomputing 382, 196–209.

Sharma, K.K., Seal, A., 2020. Clustering analysis using an adaptive fused distance. Eng. Appl. Artif. Intell. 96, 103928.

Shu, Z., Weng, Z., Yu, Z., You, C., Liu, Z., Tang, S., Wu, X., 2022. Correntropy-based dual graph regularized nonnegative matrix factorization with Lp smoothness for data representation. Appl. Intell. 52 (7), 7653–7669.

Shu, Z., Wu, X., Fan, H., Huang, P., Wu, D., Hu, C., Ye, F., 2017. Parameter-less auto-weighted multiple graph regularized nonnegative matrix factorization for data representation. Knowl.-Based Syst. 131, 105–112.

Singer, A., 2006. From graph to manifold Laplacian: The convergence rate. Appl. Comput. Harmon. Anal. 21 (1), 128–134.

Wall, M.E., Rechtsteiner, A., Rocha, L.M., 2003. Singular value decomposition and principal component analysis. In: A Practical Approach to Microarray Data Analysis. Springer, pp. 91–109.

Wei, J., Tong, C., Wu, B., He, Q., Qi, S., Yao, Y., Teng, Y., 2022. An entropy weighted nonnegative matrix factorization algorithm for feature representation. IEEE Trans. Neural Netw. Learn. Syst. 1–11. http://dx.doi.org/10.1109/TNNLS.2022.3184286.

Wu, W., Jia, Y., Kwong, S., Hou, J., 2018a. Pairwise constraint propagation-induced symmetric nonnegative matrix factorization. IEEE Trans. Neural Netw. Learn. Syst. 29 (12), 6348–6361.

Wu, W., Jia, Y., Wang, S., Wang, R., Kwong, S., 2021. Positive and negative label-driven nonnegative matrix factorization. IEEE Trans. Circuits Syst. Video Technol. 31 (7), 2698–2710.

Wu, W., Kwong, S., Hou, J., Jia, Y., Ip, H.H.S., 2019. Simultaneous dimensionality reduction and classification via dual embedding regularized nonnegative matrix factorization. IEEE Trans. Image Process. 28 (8), 3836–3847.

Wu, W., Kwong, S., Zhou, Y., Jia, Y., Gao, W., 2018b. Nonnegative matrix factorization with mixed hypergraph regularization for community detection. Inform. Sci. 435, 263–281.

Xie, X., Guo, X., Liu, G., Wang, J., 2017. Implicit block diagonal low-rank representation. IEEE Trans. Image Process. 27 (1), 477–489.

Xu, X., Li, J., Li, S., Plaza, A., 2020. Curvelet transform domain-based sparse nonnegative matrix factorization for hyperspectral unmixing. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 4908–4924.

Yamamoto, T., Uenohara, S., Nishijima, K., Furuya, K., 2020. Sound source separation based on multichannel non-negative matrix factorization with weighted averaging. In: Conference on Complex, Intelligent, and Software Intensive Systems. Springer, pp. 177–187.

Ye, F., Chen, C., Zheng, Z., 2018. Deep autoencoder-like nonnegative matrix factorization for community detection. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management. pp. 1393–1402.

Yu, N., Wu, M.-J., Liu, J.-X., Zheng, C.-H., Xu, Y., 2020. Correntropy-based hypergraph regularized NMF for clustering and feature selection on multi-cancer integrated data. IEEE Trans. Cybern. 51 (8), 3952–3963.

Zhang, K., Zhao, X., Peng, S., 2021. Multiple graph regularized semi-supervised nonnegative matrix factorization with adaptive weights for clustering. Eng. Appl. Artif. Intell. 106, 104499.

Zhou, L., Zhang, X., Wang, J., Bai, X., Tong, L., Zhang, L., Zhou, J., Hancock, E., 2020. Subspace structure regularized nonnegative matrix factorization for hyperspectral unmixing. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 4257–4270.

Zubaroğlu, A., Atalay, V., 2021. Data stream clustering: a review. Artif. Intell. Rev. 54 (2), 1201–1236.