Dual graph-regularized Constrained Nonnegative Matrix Factorization for Image Clustering

Jing Sun[1], Xibiao Cai[1], Fuming Sun[1*] and Richang Hong[2]
[1] School of Electronics and Information Engineering, Liaoning University of Technology
Jinzhou, 121001 - China
[e-mail: sunwenfriend@hotmail.com]
[2] School of Computer and Information, Hefei University of Technology
Hefei, 230009 - China
[e-mail: hongrc@hfut.edu.cn]
*Corresponding author: Fuming Sun

Received December 5, 2016; revised February 20, 2017; accepted March 12, 2017; published May 31, 2017

Abstract

Nonnegative matrix factorization (NMF) has received considerable attention due to its effectiveness of reducing high dimensional data and importance of producing a parts-based image representation. Most of existing NMF variants attempt to address the assertion that the observed data distribute on a nonlinear low-dimensional manifold. However, recent research results showed that not only the observed data but also the features lie on the low-dimensional manifolds. In addition, a few hard priori label information is available and thus helps to uncover the intrinsic geometrical and discriminative structures of the data space. Motivated by the two aspects above mentioned, we propose a novel algorithm to enhance the effectiveness of image representation, called Dual graph-regularized Constrained Nonnegative Matrix Factorization (DCNMF). The underlying philosophy of the proposed method is that it not only considers the geometric structures of the data manifold and the feature manifold simultaneously, but also mines valuable information from a few known labeled examples. These schemes will improve the performance of image representation and thus enhance the effectiveness of image classification. Extensive experiments on common benchmarks demonstrated that DCNMF has its superiority in image classification compared with state-of-the-art methods.

Keywords: Nonnegative matrix factorization, dual graph-regularized, manifold, feature manifold, label information

This research was supported by a research grant from the National Natural Science Foundation of China under Grants 61572244, the program for Liaoning Excelent Talents in University(LR2015030). We express our thanks to Dr. Richard Berke who checked our manuscript.

https://doi.org/10.3837/tiis.2017.05.017 **ISSN : 1976-7277**

1. Introduction

With increasing technological sophistication, especially intelligent mobile phones, image data continues to grow dramatically. As a result, it becomes a huge challenge of retrieving, classifying and extracting valuable information from the massive image data quickly and efficiently. To resolve this problem, clustering technology [1, 2], as one of important tools in machine learning [3] and data mining [4], has been widely adopted between academia and industry. However, clustering procedures refer to the process of dividing the collection of physical or abstract objects into a number of classes, which are composed of similar objects. From the traditional view, the cluster generated by clustering method is a collection of data objects that are similar to the other objects in the same cluster and have nothing in common with the objects in other clusters, which is what we often say 'Like attracts like' in our daily life. In the other hand, a data set can be viewed as a set of discrete sampling on a continuous manifold, from the geometrical view, while the purpose of clustering is to find the intrinsic geometry of the manifold.

Clustering technology has been widely studied for many years. Many scholars have proposed a number of clustering methods according to the different needs of the clustering results. The clustering algorithms can be roughly divided into five categories: partitioning methods, hierarchical methods, density-based methods, grid-based methods and model-based methods. K-means [5], spectral clustering [6] and non-negative matrix factorization (NMF) [7] are belonging to the commonly used clustering methods. Among them, NMF was proposed by Lee and Seung [8, 9] in 1999. The major purpose of the NMF algorithm is to find two low-rank nonnegative matrices, which can produce a better approximation to the original nonnegative matrix. Moreover, the introduction of nonnegative constraints makes NMF has a purely additive structure, which also leads to a parts-based representation of the decomposition results for human face images [10], text documents [11] and so on. As a result, NMF has been widely applied in data mining, computer vision, biomedical engineering and other fields.

Most of existing NMF variants attempt to address that the observed data usually distributed on a nonlinear low-dimensional manifolds in high-dimensional environmental space [12, 13]. In view of this, a large number of manifold learning methods have been proposed to detect the potential manifold structure [14], such as ISOMAP [15], Laplacian Eigenmaps (LE) [16], Locally Linear Embedding (LLE) [17]. However, recent research results showed that not only the observed data but also the features lie on the low-dimensional manifolds. To enhance the performance of classical NMF, a few researchers proposed many NMF variants. Cai et al. [18] proposed a graph regularized non-negative matrix factorization (GNMF), which considers the intrinsic geometric structure and simultaneously uncovers a compact representation of the hidden semantic information. What's more, it indicated that the structure information of the data can be greatly improved the learning performance. However, the GNMF algorithm only considers the spatial distribution of the data and does not take into account the structural information of the feature space. In essence, the GNMF algorithm belongs to the one-sided clustering method. Several co-clustering algorithms [19, 20] have proposed to solve the one-sided clustering problems [21, 22] and the superiority of co-clustering is proved by the experimental results. Gu et al. [23] presented a dual regularized co-clustering (DRCC) algorithm, which constructs two graphs of data manifold and feature manifold based on the semi-nonnegative matrix tri-factorization. Experimental results on benchmark datasets

showed that DRCC method outperforms many state-of-the-art clustering algorithms. This further demonstrates that the clustering effect of co-clustering is better than one-sided clustering. Shang et al. [24] introduced a graph dual regularization non-negative matrix factorization (DNMF) algorithm by combining the matrix factorization and the idea of dual regularization. At the same time, the graph dual regularization non-negative matrix tri-factorization (DNMTF) algorithm as an extension of DNMF is proposed. Experimental results on several datasets demonstrated the effectiveness of both DNMF and DNMTF.

In order to further boost the performance of NMF, we propose a novel algorithm in this paper, called Dual graph-regularized Constrained Nonnegative Matrix Factorization (DCNMF), which not only considers the geometric structures of data manifold and feature manifold, but also introduces the label information of known labeled samples into NMF. The proposed method reflects the geometric structure information of data and feature by constructing two nearest neighbor graphs. The iterative updating scheme for the objective function of DCNMF and its convergence proofs are also given.

2. Related Work

2.1 NMF

NMF algorithm is described as follows: given a data matrix $X = [x_1, x_2, \cdots, x_n] \in R^{m \times n}$, each column of X is an m-dimensional data point, where $x_i \in R^{m \times n}$ is a sample vector, and the elements of each sample vector are nonnegative. The goal of NMF is to seek two nonnegative matrices $U \in R^{m \times k}$ and $V \in R^{n \times k}$ whose product approximates X as closely as possible

$$X \approx UV^T \tag{1}$$

where $k \le mn/(m+n)$ and U denotes the basic matrix, V can be considered as the coefficient matrix. Especially, the elements of two matrices are all negative.

In order to find an approximate decomposition process, an objective function must be defined to guarantee the effect of the approximation. To measure the quality of the approximation defined in (1), a common metric is the Euclidean distance, given by

$$O_F = \left\| X - UV^T \right\|_F^2 = \sum_{i,j} \left(x_{ij} - \sum_{k=1}^{K} u_{ik} v_{jk} \right)^2 \qquad s.t. U \ge 0, V \ge 0 \tag{2}$$

where $\|\cdot\|_F$ is referred as the Frobenius norm.

Although the objective function of NMF is convex with respect to one variable matrix U or V, it is not convex in both matrices together. Therefore, the global minimum of the objective function is quite difficult to obtain. Multiplicative update method is a well-known NMF method, which can find a local minimum of the objective function in (2). The updating rules for the Euclidean distance objective function are as follows:

$$u_{ik} \leftarrow u_{ik} \frac{(XV)_{ik}}{(UV^TV)_{ik}}, \; v_{jk} \leftarrow v_{jk} \frac{(X^TV)_{jk}}{(V^TVU^T)_{jk}} \tag{3}$$

where $U=[u_{ik}]$, $V=[v_{jk}]$.

At the very beginning of the iterative update process, the two nonnegative matrices U_0 and V_0 are initialized at random. The iterative update procedure is executed repeatedly according to the updating rules until the given terminal condition is met. Ultimately, the final U and V can be obtained.

2.2 CNMF

NMF is an unsupervised method for matrix decomposition, not taking into account the known label information of the samples, while the label of a few examples in many real-word applications is relatively easy to get. Aiming at this case, Liu et al. [25] proposed the constrained nonnegative matrix factorization (CNMF) algorithm. This algorithm not only considers the label information as additional constraints, but also constrains it to the NMF framework. By CNMF, it can be guaranteed that the data points sharing the same label are projected into the same class in the low-dimensional space.

Assume that there are n nonnegative samples $\{x_i\}_{i=1}^{n}$ in the data set X, we know the label information of the first l samples which consists of c classes, while the rest n-l samples are not labeled. An indicator matrix C is first defined, where $c_{ij}=1$ if x_i is labeled with the jth class; $c_{ij}=0$ otherwise. With the indicator matrix C, we define a label constraint matrix A as below:

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

where I_{n-l} is an identity matrix.

The label constraint matrix A can be introduced into the objective function of NMF and $X\approx UV^T=U(AZ)^T$ can be gotten, where $V=AZ$ and $Z\in R^{(c+n-l)\times k}$ is the auxiliary matrix. It is easier to learn that if x_i and x_j share with the same label, then the weighted coefficient vector of them is identical.

The objective function of CNMF with the label constraints as (4).

$$O_F=\left\|X-UZ^TA^T\right\|_F^2 \qquad s.t. U\geq 0, Z\geq 0 \tag{4}$$

The updating rules for the objective function (4) are as follows:

$$u_{ik}\leftarrow u_{ik}\frac{(XAZ)_{ik}}{(UZ^TA^TAZ)_{ik}}\ ,\ z_{jk}\leftarrow z_{jk}\frac{(A^TX^TU)_{jk}}{(A^TAZU^TU)_{jk}} \tag{5}$$

Liu et al. provided the convergence proofs of objective function (4) under the updating rules as (5) in [25], so that the known label information is well introduced into NMF and improves the discriminative ability.

2.2 DNMF

NMF can not preserve the geometrical information of data space and feature space because it

does not take into account the geometric structure of data and features. Motivated by recent progress in dual regularization and matrix factorization, Shang et al. proposed a novel algorithm called graph dual regularization non-negative matrix factorization (DNMF), which simultaneously considers the geometric structure of the data manifold as well as feature manifold and respectively creates two graphs to reflect the structure of manifold distribution. Compared with one-side clustering algorithm, DNMF has better clustering performance.

The objective function of DNMF is formulated as follows:

$$O_F = \left\| X - UV^T \right\|_F^2 + \lambda Tr\left(V^T L_V V\right) + \mu Tr\left(U^T L_U U\right) \quad s.t. U \geq 0, V \geq 0 \tag{6}$$

The updating formulas for the objective function of DNMF as (7):

$$u_{ik} \leftarrow u_{ik} \frac{\left(XV + \mu W^U U\right)_{ik}}{\left(UV^T V + \mu D^U U\right)_{ik}} \quad , \quad v_{ik} \leftarrow v_{ik} \frac{\left(X^T U + \lambda W^V V\right)_{ik}}{\left(VU^T U + \lambda D^V V\right)_{ik}} \tag{7}$$

The objective function of the DNMF algorithm is convergent in iterative updating formulas (7), the detailed proof of convergence has been given in [24] and is not listed here.

3. DCNMF

There are some shortcomings in the improved NMF algorithms mentioned above. Although DNMF algorithm preserves the geometric structure of data and feature spaces by introducing both data manifold and feature manifold in the NMF framework, it obviously does not consider the known label information in the data set. Thus, the application of DNMF algorithm is limited. The CNMF algorithm integrates the known label information into the NMF framework to improve the discriminative ability of NMF. However, CNMF is unable to maintain the geometric structures of the space where the data samples and sample features are located.

To address these problems mentioned above, we propose a dual graph-regularized constrained nonnegative matrix factorization (DCNMF) algorithm extended from CNMF and DNMF to boost the performance of NMF. The proposed method inherits the advantages of two methods simultaneoulsy and has the folloing three aspects. Firstly, the prior constraint is applied in NMF, which can improve the discriminative ability of the coefficient matrix by making full use of the known label information. Secondly, the geometric structures of the data manifold as well as feature manifold are integrated into the NMF algorithm, and the structure of manifold distribution is reflected by creating two nearest neighbor graphs, which solves the clustering ineffective problem of the one-side clustering and effectively improves the feature representation of images and clustering performance. Finally, the above two parts are integrated into the same objective function and DCNMF algorithm proposed in this paper is further described by mathematical theory. **Table 1** gives an intuitive comparison of NMF, CNMF, DNMF and DCNMF.

Table 1. The comparison of algorithms

Advantages	Label Information	Data Graph	Feature Graph
NMF	×	×	×
CNMF	√	×	×
DNMF	×	√	√
DCNMF	√	√	√

3.1 Data Graph and Feature Graph

We create two nearest neighbor graphs to describe the data manifold and feature manifold, so that the data graph, feature graph and NMF can be well integrated into the same objective function. We first construct a *p*-nearest neighbor data graph whose vertices correspond to $\{x_1, x_2, \cdots, x_n\}$. In order to avoid the selection of heat-kernel parameters in Gaussian function, we choose 0-1 weighting method to construct the weight matrix of data graph, which is defined as follows:

$$W_{ij}^V = \begin{cases} 1, & if \quad x_j \in N_p(x_i) \\ 0, & otherwise \end{cases} \qquad i, j = 1,2,\cdots,n$$

where $N_p(x_i)$ represents the set of *p*-nearest neighbor of data point x_i. The graph Laplacian matrix of the data graph is defined as $L_v = D^V - W^V$, where D^V is a diagonal matrix, namely $D_{ii}^V = \sum_j W_{ij}^V$.

Let $V = \left[v_1^T, v_2^T, \cdots, v_n^T\right] \in R^{n \times k}$ denotes the data representation in the low-dimensional space, so the label smoothness of the data points as follows:

$$\begin{aligned} R_1 &= \frac{1}{2}\sum_{i,j=1}^{n} \left\| v_i - v_j \right\|^2 W^V \\ &= \left(\sum_{i=1}^{n} v_i v_i^T D_{ii}^V - \sum_{i,j=1}^{n} v_i v_j^T W_{ij}^V \right) \\ &= Tr\left(V^T D^V V\right) - Tr\left(V^T W^V V\right) \\ &= Tr\left(V^T L_V V\right) \end{aligned}$$

For convenience, a p-nearest neighbor feature graph is also constructed by using the 0-1 weighting method, whose vertices correspond to $\{x_1^T, x_2^T, \cdots, x_m^T\}$, and the weight matrix of feature graph is defined as follows:

$$W_{ij}^U = \begin{cases} 1, & if \quad x_j^T \in N_p\left(x_i^T\right) \\ 0, & otherwise \end{cases} \qquad i, j = 1,2,\cdots,m$$

where $N_p\left(x_i^T\right)$ represents the set of *p*-nearest neighbor of data point x_i^T. The graph Laplacian

matrix of the feature graph is defined as $L_u = D^U - W^U$, where D^U is a diagonal matrix, namely $D_{ii}^U = \sum_j W_{ij}^U$.

Let $U = \left[u_1^T, u_2^T, \cdots, u_m^T\right] \in R^{m \times k}$ represents the basis matrix to be solved, so the label smoothness of the features as follows:

$$\begin{aligned} R_2 &= \frac{1}{2}\sum_{i,j=1}^{m}\left\|u_i - u_j\right\|^2 W^U \\ &= \left(\sum_{i=1}^{m} u_i u_i^T D_{ii}^U - \sum_{i,j=1}^{m} u_i u_j^T W_{ij}^U\right) \\ &= Tr\left(U^T D^U U\right) - Tr\left(U^T W^U U\right) \\ &= Tr\left(U^T L_U U\right) \end{aligned}$$

3.2 Objective Function of DCNMF

We integrate the label smoothness of data and features into NMF objective function, and we get the DNMF algorithm. The objective function of DNMF algorithm can be defined as follows:

$$\begin{gathered} O_F = \left\|X - UZ^T A^T\right\|_F^2 + \lambda Tr\left(Z^T A^T L_Z A Z\right) + \mu Tr\left(U^T L_U U\right) \\ s.t. U \geq 0, Z \geq 0 \end{gathered} \tag{8}$$

where $L_Z = L_V = D^Z - W^Z$; A represents the label constraint matrix, while Z denotes an auxiliary matrix, and the coefficient matrix $V = AZ$. $\lambda \geq 0$ and $\mu \geq 0$ are the regularization parameters which balance the reconstruction error of DCNMF in the first term and graph regularizations in second and third terms. When letting $\mu = 0$, DCNMF degenerates to the GRCNMF method proposed by Shu et al. [26]; And when letting $\lambda = \mu = 0$, DCNMF degenerates to the CNMF algorithm proposed by Liu et al..

3.3 Optimization

It can be seen that the objective function in Eq. (8) is convex in *U* only or *Z* only and it is not convex in both variables together, so we can not find a closed-form solution. To solve this problem, we will present an iterative scheme to optimize the objective function. In other words, we will optimize the objective with respect to one variable when fixing the other one. According to the property of trace in a matrix, $Tr\left(A^T\right) = Tr\left(A\right)$ and $Tr\left(AB\right) = Tr\left(BA\right)$, the objective function in Eq. (8) can be rewritten as:

$$\begin{aligned} O_F &= Tr\left(\left(X - UZ^T A^T\right)\left(X - UZ^T A^T\right)^T\right) + \lambda Tr\left(Z^T A^T L_Z AZ\right) + \mu Tr\left(U^T L_U U\right) \\ &= Tr\left(XX^T\right) - 2Tr\left(XAZU^T\right) + Tr\left(UZ^T A^T AZU^T\right) \\ &+ \lambda Tr\left(Z^T A^T L_Z AZ\right) + \mu Tr\left(U^T L_U U\right) \\ &\qquad s.t. U \geq 0, Z \geq 0 \end{aligned}$$

Let δ and φ be the Lagrange multiplier for constraints u_{ij} and z_{ij}, respectively. Then the Lagrange function L is defined as:

$$L = O_F + Tr\left(\delta U^T\right) + Tr\left(\varphi Z^T\right) \tag{9}$$

(1). Updating U

The partial derivation of L with respect to U is

$$\frac{\partial L}{\partial U} = -2XAZ + 2UZ^T A^T AZ + 2\mu L_U U + \delta$$

Using the KKT condition $\delta_{ij} u_{ij} = 0$, we can get

$$\left(-XAZ + UZ^T A^T AZ + \mu L_U U\right)_{ij} u_{ij} = 0 \tag{10}$$

Since $L_u = D^U - W^U$ and the elements of D^U and W^U are nonnegative, then the above function can be rewritten as:

$$\left(-XAZ + UZ^T A^T AZ + \mu D^U U - \mu W^U U\right)_{ij} u_{ij} = 0 \tag{11}$$

According to E.q (11), we can get the following updating formula of variable u_{ij},

$$u_{ij} \leftarrow u_{ij} \frac{\left(XAZ + \mu W^U U\right)_{ij}}{\left(UZ^T A^T AZ + \mu D^U U\right)_{ij}} \tag{12}$$

(2). Updating Z

The partial derivation of L with respect to Z is

$$\frac{\partial L}{\partial Z} = -2A^T X^T U + 2A^T AZU^T U + 2\lambda A^T L_Z AZ + \varphi$$

Using the KKT condition $\varphi_{ij} z_{ij} = 0$, we obtain

$$\left(-A^T X^T U + A^T AZU^T U + \lambda A^T L_Z AZ\right)_{ij} z_{ij} = 0 \tag{13}$$

Since $L_z = D^Z - W^Z$ and the elements of D^Z and W^Z are nonnegative, then E.q (13) can be rewritten as:

$$\left(-A^T X^T U + A^T AZU^T U + \lambda A^T D^Z AZ - \lambda A^T W^Z AZ\right)_{ij} z_{ij} = 0 \tag{14}$$

According to E.q (14), we can get the following updating formula of variable z_{ij},

$$z_{ij} \leftarrow z_{ij} \frac{\left(A^T X^T U + \lambda A^T W^Z AZ\right)_{ij}}{\left(A^T AZU^T U + \lambda A^T D^Z AZ\right)_{ij}} \tag{15}$$

We summarize the optimization way of the proposed method in Algorithm 1.

Algorithm 1 Dual graph-regularized Constrained Nonnegative Matrix Factorization

Input: Matrix *X*, the number of classes *k*, regularization parameters λ, μ, maximum number of iterations *I*;

Output: Matrices *U* and *Z*;

Initialize *U* and *Z* using K-means;

while not convergent **and** *i* < *I* **do**

Update $u_{ij} \leftarrow u_{ij} \frac{\left(XAZ + \mu W^U U\right)_{ij}}{\left(UZ^T A^T AZ + \mu D^U U\right)_{ij}}$

Update $z_{ij} \leftarrow z_{ij} \frac{\left(A^T X^T U + \lambda A^T W^Z AZ\right)_{ij}}{\left(A^T AZU^T U + \lambda A^T D^Z AZ\right)_{ij}}$

end while

3.4 Convergence Analysis

Regarding these two updating rules as (12) and (15), we have the following theorem:

Theorem 1. For $U \geq 0$ and $Z \geq 0$, the objective function in Eq. (8) is convergent under the updating formulas in Eq. (12) and (15), namely non-increasing.

In order to prove Theorem 1, we first define the auxiliary function.

Definition 1. If $G(x, x')$ is an auxiliary function of $F(x)$, then the conditions $G(x, x') \geq F(x)$ and $G(x, x) = F(x)$ are satisfied.

Lemma 1. If $G(x, x^t)$ is an auxiliary function of $F(x)$, then $F(x)$ is non-increasing under the following updating rule,

$$x^{t+1} = \arg\min_x G(x, x^t) \tag{16}$$

Proof. Obviously, the function $G(x, x^t)$ can achieve the minimum value when $x = x^t$. From the Definition 1, it is known that $G(x^{t+1}, x^t) \geq F(x^{t+1})$ can be represented by the following inequality:

$$F\left(x^{t+1}\right) \overset{def}{\leq} G\left(x^{t+1}, x^{t}\right) \overset{\min}{\leq} G\left(x^{t}, x^{t}\right) \overset{def}{\leq} F\left(x^{t}\right)$$

From the above formula we can obtain that the equation $F\left(x^{t+1}\right) = F\left(x^{t}\right)$ holds only if x^t is a local minimum of $G\left(x, x^{t}\right)$. If the derivative of $F\left(x\right)$ exist and is continuous in a small neighborhood of x^t, then there is $\nabla F\left(x^{t}\right) = 0$.

From Eq. (16), we can get the sequence that converging to the local minimum $x_{\min}$ of the objective function:

$$F\left(x_{\min}\right) \leq \cdots \leq F\left(x^{t+1}\right) \leq F\left(x^{t}\right) \leq \cdots \leq F\left(x^{1}\right) \leq F\left(x^{0}\right)$$

Therefore, the iterative rules of the objective function in (8) can be satisfied with the Eq. (16) by defining the auxiliary function $G\left(x, x^{t}\right)$.

We first need to prove that the updating rule (15) is consistent with the formula (16). Therefore, we use $F_{Z_{ij}}$ to represent the part of the objective function that is related to any element z_{ij} in *Z*, we can get

$$\begin{aligned} F_{Z_{ij}}\left(z_{ij}\right) &= \left(-2XAZU^{T} + UZ^{T}A^{T}AZU^{T} + \lambda Z^{T}A^{T}L_{z}AZ\right)_{ij} \\ F_{Z_{ij}}{}'\left(z_{ij}\right) &= \left(-2A^{T}X^{T}U + 2A^{T}AZU^{T}U + 2\lambda\lambda A^{T}L_{z}AZ\right)_{ij} \\ F_{Z_{ij}}{}''\left(z_{ij}\right) &= \left(2A^{T}AU^{T}U + 2\lambda A^{T}L_{z}A\right)_{ij} \end{aligned}$$

where $F_{Z_{ij}}{}'\left(z_{ij}\right)$ and $F_{Z_{ij}}{}''\left(z_{ij}\right)$ denote the first-order partial derivative and the second-order partial derivative of the function $F_{Z_{ij}}$ to the variable z_{ij}, respectively.

Lemma 2. Define the auxiliary function of variable z_{ij} in the objective function O_F as follows:

$$G\left(z, z_{ij}^{t}\right) = F_{Z_{ij}}\left(z_{ij}^{t}\right) + F_{Z_{ij}}{}'\left(z_{ij}^{t}\right)\left(z - z_{ij}^{t}\right) + \frac{\left(A^{T}AZU^{T}U + \lambda A^{T}D^{Z}AZ\right)_{ij}}{z_{ij}^{t}}\left(z - z_{ij}^{t}\right)^{2} \tag{17}$$

Proof. Obviously, $G\left(z, z\right) = F_{Z_{ij}}\left(z\right)$. According to the Definition 1, we only need to prove $G\left(z, z_{ij}^{t}\right) \geq F_{Z_{ij}}\left(z_{ij}\right)$. The Taylor series expansion of $F_{Z_{ij}}\left(z_{ij}\right)$ can be defined as (18):

$$\begin{aligned} F_{Z_{ij}}\left(z_{ij}\right) &= F_{Z_{ij}}\left(z_{ij}^{t}\right) + F_{Z_{ij}}{}'\left(z_{ij}^{t}\right)\left(z - z_{ij}^{t}\right) + \frac{1}{2}F_{Z_{ij}}{}''\left(z_{ij}^{t}\right)\left(z - z_{ij}^{t}\right)^{2} \\ &= F_{Z_{ij}}\left(z_{ij}^{t}\right) + F_{Z_{ij}}{}'\left(z_{ij}^{t}\right)\left(z - z_{ij}^{t}\right) + \left(A^{T}AU^{T}U + \lambda A^{T}L_{z}A\right)\left(z - z_{ij}^{t}\right)^{2} \end{aligned} \tag{18}$$

After comparing Eq. (17) and (18), it is not difficult to find that $G\left(z,z_{ij}^{t}\right)\geq F_{Z_{ij}}\left(z_{ij}\right)$ is equivalent to inequality (19)

$$\frac{\left(A^{T}AZU^{T}U+\lambda A^{T}D^{Z}AZ\right)_{ij}}{z_{ij}^{t}}\geq\left(A^{T}AU^{T}U+\lambda A^{T}L_{z}A\right)_{ij} \tag{19}$$

To prove the above inequality, we have the following two inequalities as (20) and (21):

$$\begin{aligned}&\left(A^{T}AZU^{T}U\right)=\sum_{l}\left(A^{T}AZ\right)\left(U^{T}U\right)\geq\left(A^{T}AZ\right)\left(U^{T}U\right)\\&\geq\sum_{l}\left(A^{T}A\right)z^{t}\left(U^{T}U\right)\geq z^{t}\left(A^{T}A\right)\left(U^{T}U\right)\end{aligned} \tag{20}$$

$$\begin{aligned}&\left(\lambda A^{T}D^{Z}AZ\right)=\lambda\sum_{l}A^{T}D^{Z}Az^{t}\geq\lambda A^{T}D^{Z}Az^{t}\\&\geq\lambda A^{T}\left(D^{Z}-W^{Z}\right)Az^{t}=\lambda A^{T}L_{z}Az^{t}\end{aligned} \tag{21}$$

It can be seen from Eq. (20) and (21) that inequality (19) holds and $G\left(z,z_{ij}^{t}\right)\geq F_{Z_{ij}}\left(z_{ij}\right)$ is established. (Lemma 2 is certified.)

Next we will prove that the updating rule (12) is exactly the formula (16) with a proper auxiliary function. Therefore, we use $F_{u_{ij}}$ to represent the part of the objective function that is associated with any element u_{ij} in *U*, we obtain

$$\begin{aligned}F_{u_{ij}}\left(u_{ij}\right)&=\left(-2XAZU^{T}+UZ^{T}A^{T}AZU^{T}+\mu U^{T}L_{u}U\right)_{ij}\\F_{u_{ij}}{}'\left(u_{ij}\right)&=\left(-2XAZ+2UZ^{T}A^{T}AZ+2\mu L_{u}U\right)_{ij}\\F_{u_{ij}}{}''\left(u_{ij}\right)&=\left(2Z^{T}A^{T}AZ+2\mu L_{u}\right)_{ij}\end{aligned}$$

where $F_{u_{ij}}{}'\left(u_{ij}\right)$ and $F_{u_{ij}}{}''\left(u_{ij}\right)$ represent the first-order partial derivative and the second-order partial derivative of the function $F_{u_{ij}}$ to the variable u_{ij}, respectively.

Lemma 3. Define the auxiliary function of variable u_{ij} in the objective function O_F as follows:

$$G\left(u,u_{ij}^{t}\right)=F_{u_{ij}}\left(u_{ij}^{t}\right)+F_{u_{ij}}{}'\left(u_{ij}^{t}\right)\left(u-u_{ij}^{t}\right)+\frac{\left(UZ^{T}A^{T}AZ+\mu D^{U}U\right)_{ij}}{u_{ij}^{t}}\left(u-u_{ij}^{t}\right)^{2} \tag{22}$$

Proof. It is easy to see $G\left(u,u\right)=F_{u_{ij}}\left(u\right)$ from the Definition 1, we just need to prove $G\left(u,u_{ij}^{t}\right)\geq F_{u_{ij}}\left(u_{ij}\right)$. We can get the Taylor series expansion of $F_{u_{ij}}\left(u_{ij}\right)$ as Eq. (23):

$$\begin{aligned} F_{u_{ij}}\left(u_{ij}\right) &= F_{u_{ij}}\left(u_{ij}^{t}\right)+F_{u_{ij}}{}'\left(u_{ij}^{t}\right)\left(u-u_{ij}^{t}\right)+\frac{1}{2}F_{u_{ij}}{}''\left(u_{ij}^{t}\right)\left(u-u_{ij}^{t}\right)^{2} \\ &= F_{u_{ij}}\left(u_{ij}^{t}\right)+F_{u_{ij}}{}'\left(u_{ij}^{t}\right)\left(u-u_{ij}^{t}\right)+\left(Z^{T}A^{T}AZ+\mu L_{u}\right)\left(u-u_{ij}^{t}\right)^{2} \end{aligned} \tag{23}$$

By comparing equations (22) and (23), we can see that the inequality is equivalent to inequality (24)

$$\frac{\left(UZ^{T}A^{T}AZ+\mu D^{U}U\right)_{ij}}{u_{ij}^{t}}\geq\left(Z^{T}A^{T}AZ+\mu L_{u}\right) \tag{24}$$

According to the theory of linear algebra, we can get the following inequality

$$\begin{aligned} &\left(UZ^{T}A^{T}AZ\right)=\sum_{l}\left(UZ^{T}A^{T}\right)\left(AZ\right)\geq\left(UZ^{T}A^{T}\right)\left(AZ\right) \\ &\geq\sum_{l}u^{t}\left(Z^{T}A^{T}\right)\left(AZ\right)\geq u^{t}\left(Z^{T}A^{T}\right)\left(AZ\right) \end{aligned} \tag{25}$$

$$\mu D^{U}U=\mu\sum_{l}D^{U}u^{t}\geq\mu D^{U}u^{t}\geq\mu\left(D^{U}-W^{U}\right)u^{t}=\mu L_{u}u^{t} \tag{26}$$

It can be seen from Eq. (25) and (26) that inequality (24) holds and $G\left(u,u_{ij}^{t}\right)\geq F_{u_{ij}}\left(u_{ij}\right)$ is established. (Lemma 3 is certified.)

Finally, we will prove the convergence of Theorem 1.

Proof. Replacing $G\left(x,x^{t}\right)$ in Eq. (16) by auxiliary function equation (17), we can get

$$\begin{aligned} z_{ij}^{(t+1)} &= z_{ij}^{(t)}-z_{ij}^{(t)}\frac{F_{Z_{ij}}{}'\left(z_{ij}^{(t)}\right)}{2\left(A^{T}AZU^{T}U+\lambda A^{T}D^{Z}AZ\right)_{ij}} \\ &= z_{ij}^{(t)}\frac{\left(A^{T}X^{T}U+\lambda A^{T}W^{Z}AZ\right)_{ij}}{\left(A^{T}AZU^{T}U+\lambda A^{T}D^{Z}AZ\right)_{ij}} \end{aligned}$$

Similarly, we substitute the auxiliary function (22) for $G\left(x,x^{t}\right)$ in Eq. (16) to obtain the following update rule:

$$\begin{aligned} u_{ij}^{(t+1)} &= u_{ij}^{(t)}-u_{ij}^{(t)}\frac{F_{u_{ij}}{}'\left(u_{ij}^{(t)}\right)}{2\left(UZ^{T}A^{T}AZ+\mu D^{U}U\right)_{ij}} \\ &= u_{ij}^{(t)}\frac{\left(XAZ+\mu W^{U}U\right)_{ij}}{\left(UZ^{T}A^{T}AZ+\mu D^{U}U\right)_{ij}} \end{aligned}$$

Since equations (17) and (22) are both auxiliary functions for O_F, O_F is non-increasing

under the updating rules in Eq. (12) and (15). At the same time, the objective function (8) has a lower bound. In conclusion, the convergence of Theorem 1 is proved.

4. Experiments and Results Analysis

4.1 Data Sets

We will evaluate the performance of the DCNMF algorithm on the following two data sets which are widely used as benchmark data sets in clustering.

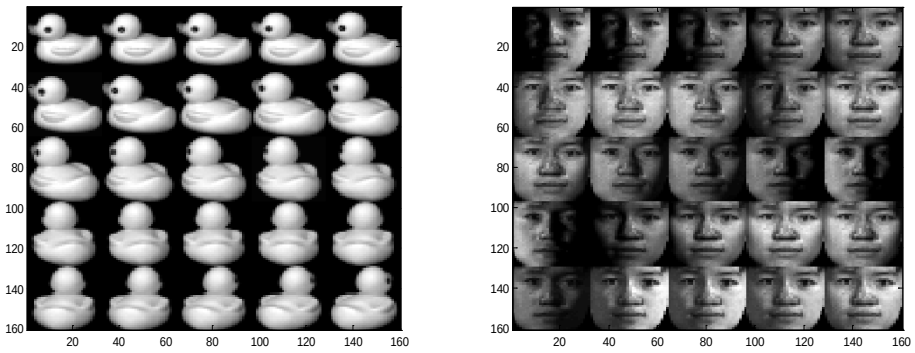
COIL20 This database is collected and produced by Columbia University, which contains 20 kinds of different objects in the image (toy duck, cups, etc.), each object is rotated in the level of 360 degrees, every 5 degrees shooting a picture, so each object possesses 72 pictures and there are 1440 pictures for 20 different objects.

PIE_pose27 This database is created by Carnegie Mellon University in the United States and PIE-pose27 consists of 2586 images with 68 volunteers, each of which is collected under the strict conditions of posture and illumination variations. There are 2856 pictures with 42 different lighting conditions in this dataset.

Table 2 gives the specifics of two databases in size, dimensionality and class. Besides, some example images from two datasets are displayed as **Fig. 1**.

Table 2. Statistics of the Dataset

Dataset	Size	Dimensionality	Classes
COIL20	1440	1024	20
PIE-pose27	2856	1024	68

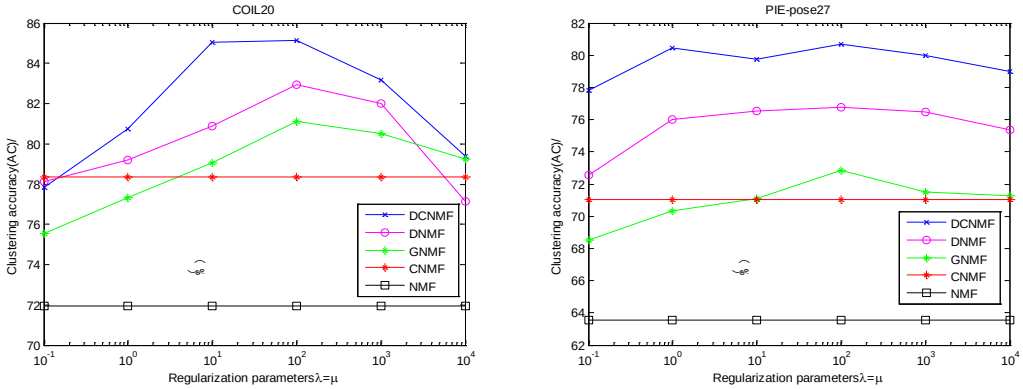


(a) COIL20 (b) PIE-pose27

Fig. 1. Instances from two datasets

4.2 Parameters Selection

There are mainly two parameters in our proposed DCNMF algorithm: the regularization parameters λ and μ. Since DCNMF is very robust to the value of λ and μ, we set $\lambda = \mu$ in our experiments.



(a) COIL20 (b) PIE-pose27

Fig. 2. Clustering accuracy with respect to the regularization parameters

From the results shown in **Fig. 2**, we can observe that the influence of clustering accuracy AC on DCNMF is not great when λ and μ are changed from 10 to 1000. Empirically, we set $\lambda = \mu = 100$.

In this experiment, we will select the first 20% of each class samples as the labeled samples, the remaining as the unlabeled samples, and randomly select k(1,2, ⋯,10) class samples from it for clustering experiments. We set the number of clusters equal to the true number of classes k for all clustering algorithms. We will repeat the experiment twenty times for each k and the average result is recoded as the final result. The maximum iteration number for our experiments is 300.

4.3 Clustering Results

In this section, we will evaluate the performance of the proposed algorithm on COIL20 and PIE-pose27 datasets. We compare our method with NMF, CNMF, GNMF and DNMF, and we use two popular evaluation metrics, the clustering accuracy (AC) and the normalized mutual information (NMI), to measure the performance of all the clustering algorithms. **Table 3** and **Table 4** show the best average clustering accuracy of all methods on two datasets, while **Table 5** and **Table 6** show the normalized mutual information.

Table 3. The clustering accuracy results on the COIL20 dataset

Clustering Accuracy（AC）/（%）					
k	**NMF**	**CNMF**	**GNMF**	**DNMF**	**DCNMF**
2	87.96	90.18	93.53	93.88	95.46
3	82.91	87.99	89.33	91.67	92.95
4	79.00	85.79	87.07	89.90	91.08
5	74.40	80.20	81.86	85.36	88.14
6	70.76	79.50	83.52	83.89	85.71
7	67.89	76.86	79.25	80.15	83.19
8	61.33	72.75	75.89	77.09	79.39

9	62.26	65.00	69.68	73.10	74.14
10	61.22	67.01	70.09	71.45	76.19
Avg.	71.95	78.37	81.12	82.94	85.14

Table 4. The clustering accuracy results on the PIE-pose27 dataset

Clustering Accuracy（AC）/（%）					
k	**NMF**	**CNMF**	**GNMF**	**DNMF**	**DCNMF**
2	75.31	78.00	80.38	88.54	92.84
3	67.57	74.88	78.85	85.42	89.72
4	65.88	75.59	76.68	80.95	85.43
5	64.29	72.24	73.19	77.95	83.89
6	63.25	70.85	72.44	75.49	78.58
7	61.32	70.00	70.03	73.38	77.06
8	59.51	69.76	70.92	72.45	75.21
9	56.27	67.29	68.75	69.94	73.38
10	58.65	60.68	64.46	66.72	70.14
Avg.	63.56	71.03	72.86	76.76	80.69

From **Table 3** and **Table 4**, we can see that the average clustering accuracy of DCNMF algorithm is better than that of other methods on two datasets. On the COIL20 dataset, the AC of DCNMF is 13.19% higher than NMF, 6.77% higher than CNMF, 4.02% higher than GNMF and 2.2% higher than DNMF. On the PIE-pose27 dataset, the AC of DCNMF is 17.13% higher than NMF, 9.66% higher than CNMF, 7.83% higher than GNMF and 3.93% higher than DNMF.

Table 5. The NMI results on the COIL20 dataset

Normalized Mutual Information（NMI）/（%）					
k	**NMF**	**CNMF**	**GNMF**	**DNMF**	**DCNMF**
2	85.89	87.34	89.42	91.65	92.84
3	77.14	80.77	83.09	89.06	89.61
4	80.02	81.01	85.39	88.39	91.81
5	73.71	79.89	80.92	81.51	86.41
6	70.32	77.90	79.58	83.42	83.20
7	63.02	76.34	78.06	79.23	81.68
8	66.97	72.98	73.33	74.46	77.78

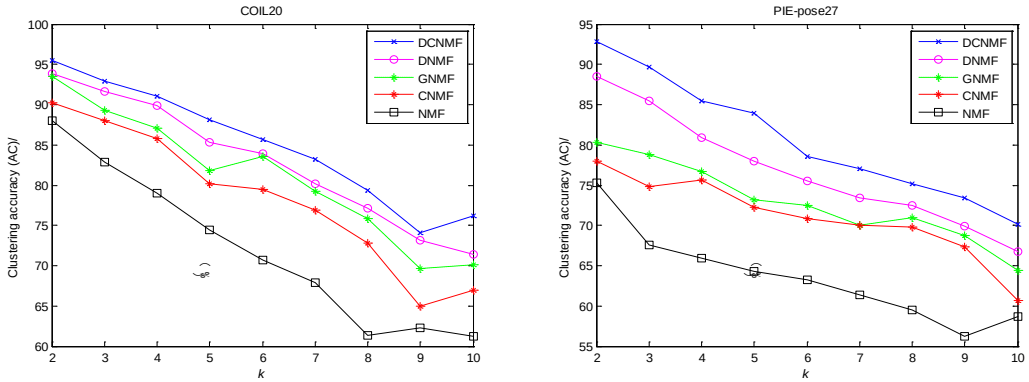
9	62.85	73.71	74.90	75.23	75.51
10	64.58	72.42	72.94	73.62	74.86
Avg.	71.61	78.04	79.74	81.84	83.74

Table 6. The NMI results on the PIE-pose27 dataset

Normalized Mutual Information（NMI）/（%）					
k	**NMF**	**CNMF**	**GNMF**	**DNMF**	**DCNMF**
2	70.13	75.12	80.85	85.89	90.45
3	63.30	71.45	74.51	77.14	83.19
4	66.17	65.37	71.64	80.02	88.90
5	65.89	68.15	74.05	79.33	81.10
6	60.54	68.85	69.38	73.71	76.32
7	61.29	69.30	70.35	68.99	74.11
8	58.00	70.61	69.21	64.58	67.34
9	60.02	65.91	67.25	63.02	69.00
10	58.46	66.72	65.48	62.85	65.17
Avg.	62.64	69.05	71.41	72.84	77.29

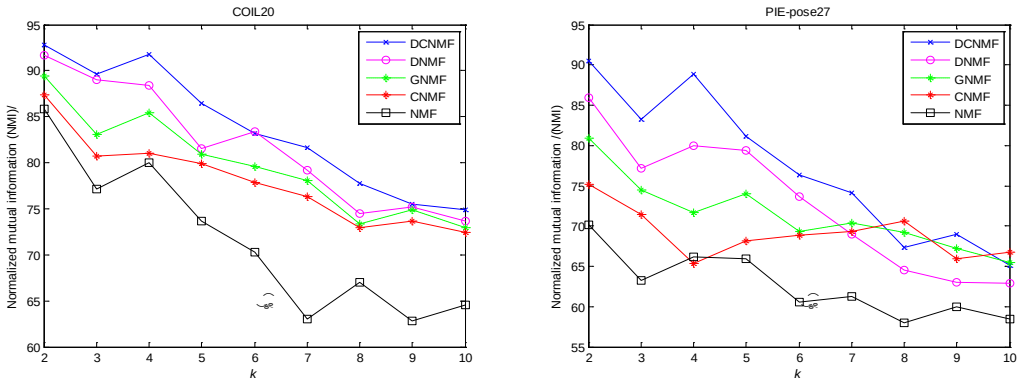
Similarly, it can be seen from **Table 5** and **Table 6** that the average normalized mutual information of DCNMF, DNMF, GNMF and CNMF are much higher than NMF on two datasets. On the COIL20 dataset, the NMI of DCNMF is 12.13% higher than NMF, 5.7% higher than CNMF, 4% higher than GNM and 1.9% higher than DNMF. On the PIE-pose27 dataset, the NMI of DCNMF is 14.65% higher than NMF, 8.24% higher than CNMF, 5.88% higher than GNMF, and 4.45% higher than DNMF.

In order to visually demonstrate the clustering effectiveness of DCNMF algorithm, **Fig. 3** shows the graphical clustering accuracy results with different *k* values for COIL20 and PIE_pose27 respectively, while **Fig. 4** shows the graphical normalized mutual information.



(a) COIL20 (b) PIE-pose27

Fig. 3. The clustering accuracy of two datasets



(a) COIL20 (b) PIE-pose27

Fig. 4. The normalized mutual information of two datasets

From **Fig. 3** and **Fig. 4** we can get: (1) the clustering accuracy and normalized mutual information of DCNMF, DNMF, GNMF and CNMF are much higher than that of NMF, although sometimes the results have some fluctuation; (2) NMF, CNMF and GNMF belong to the one-side clustering method, while DCNMF and DNMF belong to the co-clustering method, whose results of AC and NMI are relatively higher than other three methods, so we can conclude that the co-clustering method has better performance than one-side clustering method; (3) Compared with the other four algorithms, the DCNMF algorithm proposed in this paper achieves the best clustering results.

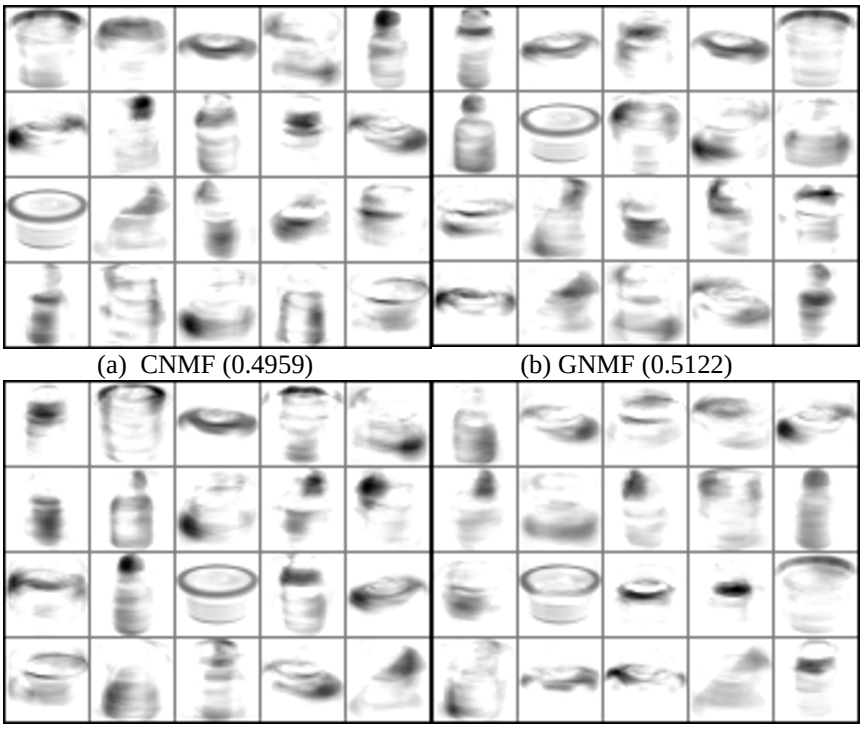
4.4 Sparseness Study

A sparseness measure based on the relationship between L_1 norm and L_2 norm [27] can be defined as (27):

$$sparseness(x) = \frac{1}{n-1}\left[n - \left(\|x\|_1 / \|x\|_2\right)^2\right] \tag{27}$$

where n is the dimensionality of the vector X, $sparseness(x) \in [0,1]$, $\|\cdot\|_1$ denotes L_1 norm and $\|\cdot\|_2$ denotes L_2 norm. This metric evaluates to unity if and only if X contains only a single non-zero component, and takes a value of zero if and only if all components are equal (up to signs), interpolating smoothly between the two extremes.

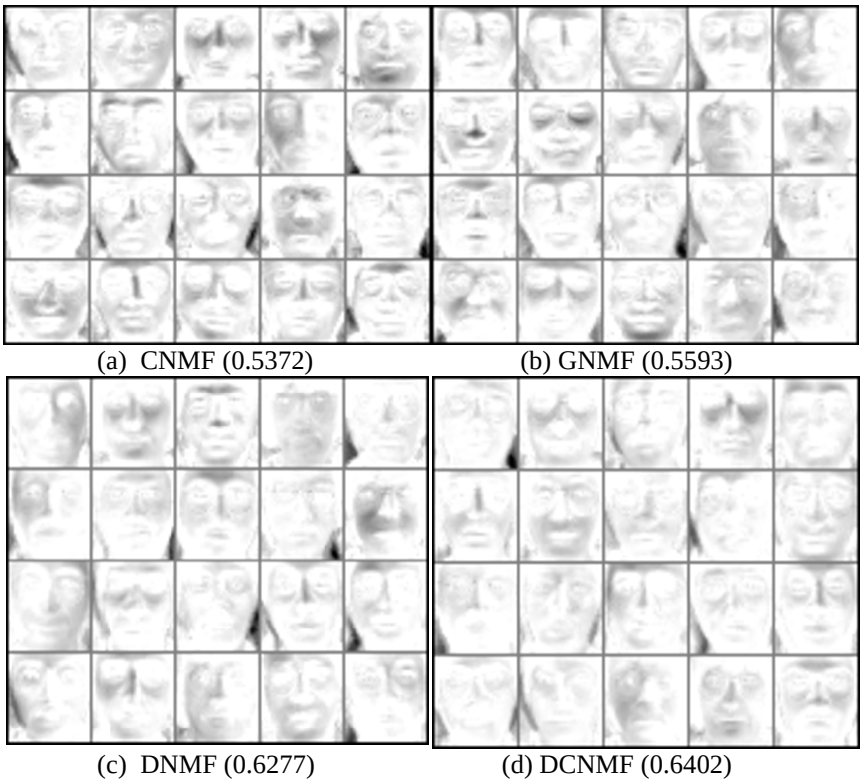
Next, we compute the sparseness of basis vectors learned by CNMF, GNMF, DNMF and DCNMF according to (27) on two databases, respectively. The partial basis images on COIL20 and PIE-pose27 are shown in **Fig. 5** and **Fig. 6**.



(a) CNMF (0.4959) (b) GNMF (0.5122)

(c) DNMF (0.5572) (d) DCNMF (0.5757)

Fig. 5. Basis images in COIL20



(a) CNMF (0.5372) (b) GNMF (0.5593)

(c) DNMF (0.6277) (d) DCNMF (0.6402)

Fig. 6. Basis images in PIE-pose27

From the above figures, it can be observed that the basis vectors generated by DCNMF approach are sparser than that generated by the other methods. These results indicate that DCNMF can learn a better parts-based representation than other algorithms.

5. Conclusion

In this paper, we proposed a dual graph-regularized constrained nonnegative matrix factorization algorithm (DCNMF), and gave the corresponding iterative updating rules and convergence proof. We used two popular evaluation metrics, the clustering accuracy (AC) and the normalized mutual information (NMI), to evaluate the clustering performance of DCNMF algorithm on COIL20 and PIE-pose27 datasets. From the experimental results, we can see that the DCNMF algorithm is obviously better in comparison with the others. The sparseness characteristics of DCNMF have illustrated on benchmark data sets. We also can draw a conclusion that the proposed algorithm has the highest sparseness degrees and could learn a better parts-based representation than other methods. As a result, the basis images have stronger discriminating power. However, the regularization parameters in DCNMF need to optimize by continuous searching. Therefore, how to choose the regularization parameters is one of the key points in our future work.

References

[1] J. Wright, A. Y. Yang and A. G. Thuy, "Robust face recognition via sparse representation," in *Proc. of 8th Int. IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 1-2, September 17-19, 2008. Article (CrossRef Link)

[2] Z. Li, J. Liu, Y. Yang and X. Zhou, "Clustering-guided sparse structural learning for unsupervised feature selection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 9, pp. 2138-2150, September, 2014. Article (CrossRef Link)

[3] A. K. Jain, M. N. Murty and P. J. Flynn, "Data clustering: a review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264-323, September, 1999. Article (CrossRef Link)

[4] T. M. Mitchell, J. G. Garbonell and R. S. Michalski, "Machine learning," McGraw-Hill, New York, 1986. Article (CrossRef Link)

[5] H. Zha, X. He, C. H. Q. Ding, H. D. Simon and M. Gu, "Spectral relaxation for k-means clustering," in *Proc. of 14th Int. Neural Information Processing Systems*: *Natural and Synthetic*, pp.1057-1064, December 3-8, 2001. Article (CrossRef Link)

[6] I. Dhillon, Y. Guan and B. Kulis, "Weighted graph cuts without eigenvectors: a multilevel approch," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1944-1957, September, 2007. Article (CrossRef Link)

[7] C. H. Q. Ding and X. He, "On the equivalence of non-negative matrix factorization and spectral clustering," in *Proc. of 5th Int. SIAM Conference on Data Mining*, pp.606-610, April 21-23, 2005. Article (CrossRef Link)

[8] D. D. Lee and H. S. Seung, "Algorithms for nonnegative matrix factorization," in *Proc. of 13th Int. Advances in Neural Information Processing Systems* (*NIPS*), pp.556-562, November 27-30, 2000. Article (CrossRef Link)

[9] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 6755, no. 401, pp. 788-791, October, 1999. Article (CrossRef Link)

[10] F. Sun, M. Xu, X. Hu and X. Jiang, "Graph regularized and sparse nonnegative matrix factorization with hard constraints for data representation," *Neurocomputing*, vol. 173, no. 2, pp. 233-244, January, 2016. Article (CrossRef Link)

[11] P. Paatero and U. Tapper, "Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, no. 2, pp. 111-126, June, 1994. Article (CrossRef Link)
[12] F. Shang, L. C. Jiao, J. Shi and J. Chai, "Robust positive semidefinite L-Isomap ensemble," *Pattern Recognition Letters*, vol. 32, no. 4, pp. 640-649, March, 2011. Article (CrossRef Link)
[13] X. B. Shu, J.H. Tang, G. J. Qi, Z. C. Li, Y. G. Jiang and S. C. Yan, "Image classification with tailored fine-grained dictionaries," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1-1, September, 2016. Article (CrossRef Link)
[14] J. B. Tenenbaum, V. de Silva and J.C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319-2323, December, 2000. Article (CrossRef Link)
[15] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323-2326, December, 2000. Article (CrossRef Link)
[16] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," *Advances in Neural Informatioon Processing Systems*, vol. 14, no. 9, pp. 585-591, April, 2002. Article (CrossRef Link)
[17] M. Belkin, P. Niyogi and V. Sindhwani, "Manifold regularization: a geometric framework for learning from labeled and unlabeled examples," *Journal of Machine Learning Research*, vol. 7, no. 1, pp. 2399-2434, December, 2006. Article (CrossRef Link)
[18] C. Deng, X. He, J. Han and T. S. Huang, "Graph regularized non-negative matrix factorization for data representation," *IEEE Trans. Pattern Analysis and Machine Intelligence* (*TPAMI*), vol. 33, no. 8, pp. 1548-1560, December, 2010. Article (CrossRef Link)
[19] V. Sindhwani, J. Hu and A. Mojsilovic, "Regularized co-clustering with dual supervision," in *Proc. of 21th Int. Advances in Neural Information Processing Systems* (*NIPS*), pp.1505-1512, December 8-11, 2008. Article (CrossRef Link)
[20] A. Narita, K. Hayashi, R. Tomioka and H. Kashima, "Tensor factorization using auxiliary information," *Data Mining and Knowledge Discovery*, vol. 25, no. 2, pp. 501-516, September, 2012. Article (CrossRef Link)
[21] F. Sun, J. Tang, H. Li, G. J. Qi and T. S. Huang, "Multi-label image categorization with sparse factor representation," *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp.1028-1037, March, 2014. Article (CrossRef Link)
[22] X. Jia, F. Sun, H. Li, Y. Cao and X. Zhang, "Image multi-label annotation based on supervised nonnegative matrix factorization with new matching measurement," *Neurocomputing*, September, 2016. Article (CrossRef Link)
[23] Q. Gu and J. Zhou, "Co-clustering on manifolds," in *Proc. of 15th Int. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (*KDD*), pp.359-368, June 28-July 01, 2009. Article (CrossRef Link)
[24] F. H. Shang, L. C. Jiao and F. Wang, "Graph dual regularization non-negative matrix factorization for co-clustering," *Pattern Recognition*, vol. 45, no. 6, pp. 2237-2250, June, 2012. Article (CrossRef Link)
[25] H. Liu, Z. Wu, D. Cai and T. S. Huang, "Constrained non-negative matrix factorization for image representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1299-1311, July, 2012. Article (CrossRef Link)
[26] Z. Q. Shu and C. X. Zhao, "Graph-regularized constrained non-negative matrix factorization algorithm and its application to image representation," *Pattern Recognition and Artificial Intelligence*, vol. 26, no. 3, pp. 300-306, March, 2013. Article (CrossRef Link)
[27] W. Michael, A. Shakhina and W. G. Stewart, "Computing sparse reduced-rank approximations to sparse matrices," *ACM Transactions on Mathematical Software* (*TOMS*), vol. 31, no.2, pp. 252-269, June, 2005. Article (CrossRef Link)

Jing Sun is currently a M.S. candidate at the School of Electronic and Information Engineering, Liaoning University of Technology, Jinzhou, China. She received the B. E. degree in communication engineering from Liaoning University of Technology, Jinzhou, China, in 2014. Her major research interests include pattern recognition and wireless cooperative communication.

Xibiao Cai is currently an associate Professor working with School of Electronic and Information Engineering in Liaoning University of Technology, Jinzhou, China. He received the Ph. D. degree from Beijing University of Posts and Telecommunications, Beijing, China, in 2012. He has authored over 20 journal and conference papers. His current research interests focus on wireless cooperative communication, cooperative multicast/broadcast, etc.

Fuming Sun is currently a Professor with the School of Electronic and Information Engineering, Liaoning University of Technology, Jinzhou, China. He received the Ph. D. degree from the University of Science and Technology of China, Hefei, China, in 2007. From 2012 to 2013, he was a Visiting Scholar with the Department of Automation, Tsinghua University. His current research interests include content-based image retrieval, image content analysis, and pattern recognition. He has authored over 30 journal and conference papers, including IEEE Transactions on Image Processing, Multimedia Systems, Neurocomputing, Multimedia Tools and Applications, etc.

Richang Hong received the PhD degree from the University of Science and Technology of China, Hefei, China, in 2008 and a research fellow in the School of Computing, National University of Singapore from September 2008 to December 2010. He is a professor in Hefei University of Technology, Hefei, China. He has coauthored more than 70 publications in the areas of his research interests, which include multimedia content analysis and social media. He received the Best Paper Award in the ACM Multimedia 2010, Best Paper Award in the ACM ICMR 2015, and the Honorable Mention of the IEEE Transactions on Multimedia Best Paper Award. He served as the associate editor of the Information Sciences and Signal Processing, Elsevier and the Technical Program chair of the MMM 2016. He is a member of the IEEE, ACM, and the executive committee member of the ACM SIGMM China Chapter.