Orthogonal Dual Graph Regularized
Nonnegative Matrix Factorization

Jinrong He[1,2,3,4], Yanxin Shi[2], Zongwen Bai[4(✉)], and Zeyu Bao[5]

[1] Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs,
Northwest A&F University, Yangling 712100, Shaanxi, China
[2] College of Mathematics and Computer Science, Yan'an University, Yan'an 716000, China
[3] Shaanxi Key Laboratory of Agricultural Information Perception and Intelligent Service,
Northwest A&F University, Yangling 712100, Shaanxi, China
[4] Shaanxi Key Laboratory of Intelligent Processing for Big Energy Data, Yanan University,
Yan'an 716000, Shaanxi, China
ydbzw@yau.edu.cn
[5] University of Science and Technology of China, Hefei 230026, Anhui, China

Abstract. Nonnegative matrix factorization (NMF) is a classical low-rank approximation method of data matrix, which decomposes a high-dimensional data matrix into two nonnegative low-rank matrices, namely basis matrix and coefficient matrix. In order to capture the local geometric structure of original dataset, manifold learning methods are incorporated into NMF framework. Motivated by recent progress in dual graph regularization, by considering the geometric structures of both the data manifold and the feature manifold, Orthogonal Dual-graph NMF (ODNMF) algorithms were proposed, which imposed orthogonality constraints on basis matrix or coefficient matrix. Since the projection directions were mutually orthogonal, the representation power of data samples was enhanced, thus ODNMF methods are more robust for data clustering. The extensive experimental results on UCI, text and face image data sets have demonstrated the superiority of the proposed methods.

Keywords: Data representation · Nonnegative matrix factorization · Dual-graph regularization · Orthogonal projection

1 Introduction

Non-negative Matrix Factorization (NMF) is one of the most powerful tools in numerical computing, machine learning and data mining. By decomposing original high dimensional non-negative data matrix X into two low dimensional non-negative factors U and V, namely basis matrix and coefficient matrix, such that $X \approx UV^T$. Moreover, the additive reconstruction with nonnegative constraints can lead to a parts-based representation for images [1], texts [2], and microarray data [3] and so on.

To decompose the given data into a component part that encodes low-rank sparse principal features and a noise-fitting error part, Zhang et al. proposed a transductive low-rank and sparse principal feature coding (LSPFC) model to recover the low-rank and

sparse subspaces jointly for robust data representation [4]. By imposing orthogonality constraints on NMF, Ding et al. proposed orthogonal non-negative matrix factorization (ONMF) method [5], which has been shown to work remarkably well for clustering tasks. Pompili et al. show that ONMF is mathematically equivalent to a weighted variant of spherical k-means [6]. Motivated by recent progress in manifold learning and orthogonal projection, in this paper we propose a novel algorithm, which called Orthogonal Graph Regularized Nonnegative Matrix Factorization (OGNMF), which combining orthogonal constraints and manifold regularization in a parts-based low-dimensional space.

From geometric perspective, the data samples are usually sampled from a low-dimensional manifold embedded in a high-dimensional ambient space. Based on such assumption, Cai et al. proposed graph regularized NMF (GNMF [7]) method, which incorporates the manifold learning into NMF for finding a compact low-dimensional representation to discover the latent semantics and intrinsic geometric structure in the dataset simultaneously. Since the sparse hypergraph inherits the merits of both the hypergraph model and sparse representation, S. Huang et al. proposed Sparse Hypergraph regularized NMF (SHNMF [8]) to exploit the high-order discriminant manifold information for data representation. In order to depict complex relations of samples, J. Wang et al. encoded different sample relations into multiple graphs and proposed Multiple Graph regularized NMF (MGNMF [9]). However, the process of multiple graphs construction is very time consuming. Motivated by recent study on sample diversity learning [10], C. Wang et al. proposed Graph regularized NMF with Sample Diversity (GNMFSD [11]), which incorporated label information into the graph to encode the intrinsic geometrical structures of the data space, then the discriminant power of the basis vectors is enhanced. By introducing an additional regression level, Mei et al. proposed a general method for including side information on the columns and rows in NMF, which can be used for time series recovery and prediction [12].

Recent works have shown that not only the original data samples are found to lie on a nonlinear low-dimensional manifold, which is called data manifold, but also the feature vectors lie on a manifold, which is called feature manifold [13]. Considering the duality between data samples and feature vectors, several dual-graph based data representation algorithms have been proposed and shown to be superior to traditional data manifold regularization based methods. Shang et al. [14] proposed graph Dual regularization Non-negative Matrix Factorization (DNMF). In order to improve the overall performance of recommender systems, by encoding the side information from both users and items as two graph regularization terms, Yao et al. [15] proposed a dual-regularized model for one-class collaborate filtering. In order to estimate fraction of different land cover types from remote sensing imagery, Tong et al. [16] applied DNMF in hyperspectral unmixing and showed that DNMF has better performance than GNMF. Yang et al. [17] applied DNMF to discover real-world events from Flickr data. By imposing the L2,1-norm constraint on the self-representation coefficients matrix in data space, Shang et al. proposed self-representation based dual-graph regularized feature selection clustering (DFSC, [18]) and non-negative spectral learning and sparse regression-based dual-graph regularized feature selection (NSSRD, [19]). For incorporating the graph topology, Yankelevsky et al. [20] proposed a dual regularized dictionary learning method, which imposed a quadratic smoothness constraint on the dictionary bases and a manifold smoothness regularization

on the sparse representations. Luo et al. proposed Dual-regularized Multi-view Non-negative Matrix Factorization (DMvNMF [21]) to deal with multi-view data, which can extract compatible and complementary information contained in multiple modality datasets, while preserve the geometric structures in both the data space and the feature space. In order to reduce the redundancy between bases and representations in GNMF model, He et al. proposed Orthogonal GNMF (OGNMF [22]), which incorporates three kinds of orthogonal constraints into GNMF model.

In summary, our main contributions in this paper are listed below.

(1) Orthogonal dual-graph regularized NMF models are proposed, in which three types of orthogonal constraints are added, including U orthogonal, V orthogonal and bi-orthogonal. In this way, the potential geometrical structural information can be preserved during the process of data representation, which can effectively enhance the discriminative ability of clustering.
(2) The multiplicative iterative updating rules derived from original NMF are used to optimize the proposed DNMF models.
(3) Comprehensive experiments on ten real world datasets are conducted to show the effectiveness of the proposed three algorithms, and demonstrate its advantage over other state-of-the-art methods.

The remainder of this paper is organized as follows. Section 2 briefly reviews the related works, such as NMF, ONMF and DNMF. In Sect. 3, we proposed three types of ODNMF models and their optimization algorithms. Extensive experiments are presented in Sect. 4. Finally, Sect. 5 concludes the paper.

2 Related Works

2.1 NMF

NMF aims to express each sample x_i as a linear combination of the basis vectors which are columns of U, namely, where v_j is the jth column $x_i = Uv_j$ of V. Such a factorization is generally obtained by minimizing the following cost function:

$$\min_{U,V} \left\| X - UV^T \right\|_F^2$$
$$s.t. \ u_{ij} \geq 0, v_{ij} \geq 0 \tag{1}$$

where $X = [x_1, x_2, \cdots, x_n] \in R^{d \times n}$ denotes the original data matrix, which contains n data samples with the dimensionality d, and $U \in R^{d \times r}$ denotes the basis matrix, $V \in R^{r \times n}$ denotes the coefficient matrix, where r is the reduced target dimensionality of projected low dimensional data. The NMF model (1) can be optimized by using following multiplicative updating rules [1]:

$$u_{ij} = u_{ij} \frac{(XV)_{ij}}{\left(UV^TV\right)_{ij}} \tag{2}$$

$$v_{ij} = v_{ij} \frac{(X^T U)_{ij}}{(VU^T U)_{ij}} \tag{3}$$

Since the objective function in (1) is not convex in both U and V, the above iterative updating rules can only find the local minimum of the objective function. The convergence of the optimization algorithm is proved in [1].

2.2 ONMF

ONMF is a variant of NMF, which approximate the data matrix with the product of two low-rank nonnegative matrices, one of which has orthonormal columns. For example, when the coefficient matrix V has orthonormal columns, the corresponding optimization model is

$$\begin{aligned} &\min_{U,V} \ \|X - UV^T\|_F^2 \\ &s.t. \quad \begin{aligned} &u_{ij} \geq 0, \ v_{ij} \geq 0 \\ &V^T V = I \end{aligned} \end{aligned} \tag{4}$$

Since the column vectors of coefficient matrix V is orthonormal, its power lies in the potential to generate sparser part-based decompositions of data with smaller basis vectors, that are easier to interpret.

The updating rules to optimize the above objective function (4) are given as follows

$$u_{ij} = u_{ij} \frac{(XV)_{ij}}{(UV^T V)_{ij}} \tag{5}$$

$$v_{ij} = v_{ij} \frac{(X^T U)_{ij}}{(VU^T XV)_{ij}} \tag{6}$$

2.3 GNMF

GNMF adds manifold learning into the classical NMF for finding a compact representation, in which the local geometric structure of data manifold is simulated by a neighborhood graph.

Let $G(X, W)$ denote a graph with vertex set $X = \{x_1, x_2, \cdots, x_n\}$ and weight matrix W whose entry W_{ij} denotes the similarity between sample x_i and x_j. There several ways to compute the similarity weight matrix W, and we use the Gauss kernel function as follows:

$$W_{ij} = \begin{cases} e^{-\frac{\|x_i - x_j\|_2^2}{t}} & x_i \in N_k(x_j) \vee x_j \in N_k(x_i) \\ 0 & otherwise \end{cases} \tag{7}$$

where $N_k(x_i)$ is the sample points set that contains the k nearest neighbors of x_i, and t is heat kernel parameter. Since G is an undirected graph, the weight matrix W is symmetric and the diagonal entries in W are all 0.

GNMF aims to learning a collection of nonnegative basis which can not only minimize the reconstruction error but also preserve the similarities between pairwise samples that encoded in neighborhood graph. Therefore, the optimal nonnegative basis U can be obtained via minimizing the following objective:

$$\min_{U,V} \left\| X - UV^T \right\|_F^2 + \lambda tr\left(V^T LV\right)$$
$$s.t. \ \ u_{ij} \geq 0, v_{ij} \geq 0 \tag{8}$$

where $L = D$–W is the graph Laplacian matrix and D is a diagonal matrix whose entry is $D_{ii} = \sum_{j=1}^{n} W_{ij}$, W is the weight matrix to measure the similarity between the nearby data samples. The λ is a regularization parameter which balance the reconstruction error and graph embedding term. Since the graph embedding regularizer could smooth the variation between two samples with large similarity in the latent low-dimensional space, the performance of clustering based on GNMF can be effectively improved.

Similar to Eq. (2) and (3), the iterative updating rules to solve (8) are presented as follows:

$$u_{ij} = u_{ij} \frac{(XV)_{ij}}{\left(UV^T V\right)_{ij}} \tag{9}$$

$$v_{ij} = v_{ij} \frac{\left(X^T U + \lambda WV\right)_{ij}}{\left(VU^T U + \lambda DV\right)_{ij}} \tag{10}$$

2.4 DNMF

DNMF is an extension of GNMF, which imposed graph regularization on basis matrix and coefficient matrix to discover the geometrical structure of data. Similar to GNMF, DNMF constructed k nearest neighbor graph on data samples and feature vectors. Suppose the columns of data matrix X are $\{X_{:1}, X_{:2}, ..., X_{:N}\}$ and the weight between two neighborhood samples can be defined as:

$$[W^V]_{ij} = \begin{cases} 1 \text{ if } X_{:,j} \in N(X_{:,i}) \\ 0 \text{ otherwise} \end{cases} i, j = 1, ..., N \tag{11}$$

where $N(X_{:,i})$ is the k-th neighborhood sample of $X_{:,i}$. The graph Laplacian matrix is $L_V = D^V - W^V$, here D^V is diagonal matrix, whose diagonal elements are rows sum of W, i.e., $[D^V]_{ii} = \sum_j [W^V]_{ij}$.

Similarly, the rows of data matrix X are feature vectors, i.e., $\{X_{1,:}^T, X_{2,:}^T, ..., X_{N,:}^T\}$, and the weight between feature vectors can be defined as:

$$[W^U]_{ij} = \begin{cases} 1, \text{ if } X_{j,:}^T \in N(X_{i,:}^T) \\ 0 \ \ otherwise \end{cases} i, j = 1, ..., M \tag{12}$$

where $N(X_{i,:}^T)$ is the k-th feature vector of $X_{i,:}^T$., and the corresponding graph Laplacian matrix is $L_U = D^U - W^U$, here D^U is a diagonal matrix, and its element is row sum of W, i.e. $[D^V]_{ii} = \sum_j [W^V]_{ij}$.

Similar to GNMF, the DNMF model can be formulated as

$$J_{DNMF} = ||X - UV^T||^2 + \lambda \text{Tr}(V^T L_V V) + \mu \text{Tr}(U^T L_U U)$$
$$\text{s.t. } U \geq 0, V \geq 0 \tag{13}$$

where λ, μ are regularization parameters. When μ tends to 0, DNMF is equivalent to GNMF. When both the λ and μ tend to 0, DNMF is equivalent to NMF. The minimization process of objective function in Eq. (9) can be achieved through the following update rules:

$$U_{ij} = U_{ij} \frac{(XV)_{ij} + \mu (W^U U)_{ij}}{(UV^T V)_{ij} + \mu (D^U U)_{ij}} \tag{14}$$

$$V_{ij} = V_{ij} \frac{(X^T U)_{ij} + \lambda (WV)_{ij}}{(VU^T U)_{ij} + \lambda (DV)_{ij}} \tag{15}$$

3 Proposed Methods

Orthogonal projections have been proved to be more discriminative in most cases, since it can also release the correlation of the projection directions or low-dimensional representations. With orthogonal constraints on basis matrix and coefficient matrix, the performance of DNMF model on data clustering is improved. ODNMF inherits the advantages of ONMF and DNMF, it can capture the local similarity structures of data space and feature which are helpful for extracting the discriminative features, while preserving the global geometrical structure of data space which aims to contain the intrinsic information as much as possible.

3.1 ODNMF-U

Adding the orthogonal constraint on basis matrix U, we have the following orthogonal dual-graph regularized NMF with U orthogonality (ODNMF-U):

$$J_{ODNMF} = ||X - UV^T||^2 + \lambda \text{Tr}(V^T L_V V) + \mu \text{Tr}(U^T L_U U)$$
$$\text{s.t. } U \geq 0, V \geq 0, \ U^T U = I \tag{16}$$

The objective function in model (16) can be rewritten as:

$$\begin{aligned} J_{ODNMF} &= tr((X - UV^T)(X - UV^T)^T) \\ &+ \lambda tr(V^T L_V V) + \mu tr(U^T L_U U) \\ &= tr(XX^T) - 2tr(XVU^T) + tr(UV^T VU^T) \end{aligned}$$

$$+ \lambda tr(V^T L_V V) + \mu tr(U^T L_U U) \tag{17}$$

The partial derivatives of objective function (17) with respect to U and V are:

$$\nabla_U J = [\nabla_U J]^+ - [\nabla_U J]^-$$
$$= UV^T V - XV + \mu L_U U$$
$$= \nabla_U \varepsilon + \mu L_U U \tag{18}$$

$$\nabla_V J = [\nabla_V J]^+ - [\nabla_V J]^-$$
$$= VU^T U - X^T U + \lambda L_V V$$
$$= \nabla_V \varepsilon + \lambda L_V V \tag{19}$$

where $L_U = D^U - W^U$ and $L_V = D^V - W^V$, and their definitions are same as Eq. (11) and (12). Let $\nabla_U \varepsilon = UV^T V - XV$ and $\nabla_V \varepsilon = VU^T U - X^T U$, when imposing orthogonal constraint on U, then

$$\nabla_U \varepsilon = UV^T X^T U - XV \tag{20}$$

Substituting Eq. (20) into Eq. (18), we have the following updating rule on U,

$$U_{ij} = U_{ij} \frac{(XV)_{ij} + \mu \left(W^U U \right)_{ij}}{\left(UV^T X^T U \right)_{ij} + \mu \left(D^U U \right)_{ij}} \tag{21}$$

Similarly, the updating rule on V is,

$$V_{ij} = V_{ij} \frac{\left(X^T U \right)_{ij} + \lambda \left(W^V V \right)_{ij}}{\left(VU^T U \right)_{ij} + \lambda \left(D^V V \right)_{ij}} \tag{22}$$

3.2 ODNMF-V

By incorporating orthogonal constraint of the coefficient matrix V in DNMF model, we have the following ODNMF-V model,

$$J_{ODNMF} = ||X - UV^T||^2 + \lambda tr(V^T L_V V) + \mu tr(U^T L_U U)$$
$$\text{s.t. } U \geq 0, V \geq 0, \ V^T V = I \tag{23}$$

The partial derivatives of objective function (23) with respect to U and V are same as in Eq. (18) and (19). Similarly, the orthogonal constraint on V can lead to the following formula,

$$\nabla_V \varepsilon = [\nabla_V \varepsilon]^+ - [\nabla_V \varepsilon]^-$$
$$= VU^T XV - X^T U \tag{24}$$

Substituting Eq. (24) into Eq. (19), we have,

$$
\begin{aligned}
\nabla_v J &= [\nabla_v J]^+ - [\nabla_v J]^- \\
&= \nabla_v \varepsilon + \mu L_v V \\
&= V U^T X V - X^T U + \mu L_v V
\end{aligned}
\tag{25}
$$

Then the updating rules on V and U can be formulated as follows,

$$
V_{ij} = V_{ij} \frac{\left(X^T U\right)_{ij} + \lambda\left(W^V V\right)_{ij}}{\left(V U^T X V\right)_{ij} + \lambda\left(D^V V\right)_{ij}}
\tag{26}
$$

$$
U_{ij} = U_{ij} \frac{(XV)_{ij} + \mu\left(W^U U\right)_{ij}}{\left(U V^T V\right)_{ij} + \mu\left(D^U U\right)_{ij}}
\tag{27}
$$

3.3 ODNMF

ODNMF imposes the orthogonal constraints on both the basis matrix U and coefficient matrix V, which involves the following minimization problem,

$$
J_{ODNMF} = ||X - UV^T||^2 + \lambda tr(V^T L_V V) + \mu tr(U^T L_U U)
$$
$$
\text{s.t. } U \geq 0, V \geq 0, \ V^T V = I, \ U^T U = I
\tag{28}
$$

The orthogonal constraints on the basis matrix U and coefficient matrix V provide a strong capability of simultaneously clustering rows and columns, which we call bi-orthogonal constraints. Combining update rules (21) in OGNMF-U and (26) in OGNMF-V, the update rules of OGNMF are given as follows

$$
V_{ij} = V_{ij} \frac{\left(X^T U\right)_{ij} + \lambda\left(W^V V\right)_{ij}}{\left(V U^T X V\right)_{ij} + \lambda\left(D^V V\right)_{ij}}
\tag{29}
$$

$$
U_{ij} = U_{ij} \frac{(XV)_{ij} + \mu\left(W^U U\right)_{ij}}{\left(U V^T X^T U\right)_{ij} + \mu\left(D^U U\right)_{ij}}
\tag{30}
$$

4 Experimental Results

In this section, we evaluate the effectiveness of our proposed ODNMF framework compared with the related matrix factorization methods including NMF, ONMF, GNMF.

4.1 Experimental Setting

To investigate the image clustering performances, two popular evaluation metric are used in the experiments, i.e., the clustering accuracy (AC) and the normalized mutual information (NMI) [10], that are defined as follows

$$AC = \frac{\sum\limits_{i=1}^{n} \delta(l_i, \tau_i)}{n}$$

where δ is a cluster label indicator function, which equals 1 if the two entries are the same and equals 0 otherwise. l_i and τ_i denote the true label and predicted label.

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

where C is the set of clusters from the ground truth, and C' is predicted clusters by clustering method. $H(\cdot)$ is the entropy of a set, and the mutual information $MI(C, C')$ between two sets of clusters C and C' is defined as

$$MI(C, C') = \sum_{c_i \in C, c'_j \in C'} p\left(c_i, c'_j\right) \log \frac{p\left(c_i, c'_j\right)}{p(c_i)p\left(c'_j\right)}$$

where $p(c_i)$ and $p\left(c'_j\right)$ are the probabilities of a sample belonging to the clusters c_i and c'_j, respectively. $p\left(c_i, c'_j\right)$ denotes the joint probability that the selected sample belongs to the clusters c_i as well as c'_j at the same time. The larger AC and higher NMI indicate a better clustering performance [6].

For the ODNMF methods, the neighborhood graph on data samples are set the same as those for GNMF, where the number of nearest neighbors of each sample is fixed as 5 and the regularization parameter in model (9), (12), (19) and (24) are empirically set as 100. After the low-dimensional data representations were produced by matrix factorization methods, the k-means clustering method was performed in the low-dimensional data space. All the experiments in this paper run in MATLAB R2015b on Win7 with 8G RAM and 3.4 GHz CPU.

4.2 Datasets Description

In our experiments, we use three types of datasets, including UCI datasets[1], text datasets, and face image datasets. For UCI datasets, MADELON is an artificial dataset, which was part of the NIPS 2003 feature selection challenge and containing data points grouped in 32 clusters placed on the vertices of a five dimensional hypercube and randomly labeled +1 or −1. CNAE9 is a highly sparse text dataset, which contains 1080 documents with 9 groups. Semeion is a handwritten digital dataset, which contains 1593 binary images

---

[1] http://archive.ics.uci.edu/ml/ .

created by 80 persons, and each image is with the size of 16×16. Hillvalley is a graphical dataset, and each sample contains 100 points of a 2D plane, and there are 606 samples in two groups, hill or valley. For text dataset, Tr31 dataset is a subset of TREC datasets[2] and Re1 is a subset of Reuters21578 dataset[3]. For face image dataset, UMIST dataset[4] consists of 575 images with 20 different persons, the number of which varies from 19 to 36. XM2VST dataset[5] contains 2950 samples with 295 persons, and YaleB dataset[6] used in our experiments is a subset of original YaleB dataset, containing 2414 images of 38 persons under 64 lighting conditions. ORL dataset[7] contains 400 images of 10 individuals, each of which contains 40 images that captured at different conditions, including shooting time, illuminations, expressions and some other facial details, such as with/without glasses. Some statistics of the four face image datasets are summarized in Table 1.

Table 1. The description of datasets used in the experiments

Datasets		#Samples	#Dimensionality	#Classes
UCI datasets	MADELON	4400	500	2
	CNAE9	1080	856	9
	Semeion	1593	256	10
	Hillvalley	606	100	2
Text dataset	Tr31	927	10128	7
	Re1	1657	3758	25
Face image dataset	UMIST	575	2576	20
	XM2VST	2950	2576	295
	YaleB	2414	1024	38
	ORL	400	1024	40

4.3 Results Discussion

The optimal clustering results on different datasets are reported in this section. The clustering accuracies are summarized in Table 2 and the NMI metrics are summarized in Table 3. The value in the parentheses is the optimal target dimensionality under which the method achieved the maximum evaluation metric. As can be shown, the proposed orthogonal DNMFs have better clustering performances than others in most cases. It

---

[2] https://trec.nist.gov/data.html.

[3] http://www.daviddlewis.com/resources/testcollections/reuters21578/.

[4] http://www.sheffield.ac.uk/eee/research/iel/research/face.

[5] http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/.

[6] http://vision.ucsd.edu/~D.Leekc/ExtYaleDatabase/ExtYaleB.html.

[7] https://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.

is worth noting that the original DNMF and orthogonal DNMF perform perfect on the Hillvalley dataset. This is because that the feature or attribute neighborhood information plays more role on identifying a Hill and a Valley on a 2D graph. However, on CNAE9, Tr31 and Re1 datasets, the traditional NMF method performs better than other methods, which indicates that dual-graph regularized NMFs are not suitable for text clustering tasks.

Table 2. The optimal clustering results (AC%)

Datasets	NMF	ONMF	DNMF	ODNMF-U	ODNMF-V	ODNMF
MADELON	54.20(78)	53.25(20)	52.40(3)	51.20(24)	58.45(2)	**59.00(2)**
CNAE9	**70.74(93)**	54.44(43)	63.89(9)	56.39(92)	64.26(37)	65.00(23)
Semeion	60.39(64)	52.48(14)	69.49(59)	69.93(16)	71.50(19)	**72.44(12)**
Hillvalley	68.65(2)	592.4(23)	**100.0(10)**	**100.00(10)**	**100.00(10)**	**100.0(10)**
Tr31	**58.36(64)**	50.05(6)	54.05(12)	54.48(12)	53.83(8)	52.64(8)
Re1	**43.39(94)**	42.31(74)	39.35(38)	38.32(76)	39.77(50)	40.43(16)
XM2VTS	17.29(31)	12.71(4)	15.80(44)	15.63(87)	**17.39(96)**	16.92(71)
UMIST	48.35(77)	33.91(8)	62.96(19)	63.30(23)	**63.65(69)**	63.13(34)
YaleB	16.20(91)	6.96(1)	32.97(18)	**33.35(16)**	13.09(21)	10.65(5)
ORL	56.50(76)	36.75(27)	55.50(65)	58.50(41)	**60.25(73)**	58.75(70)

Table 3. The optimal clustering results (NMI%)

Datasets	NMF	ONMF	DNMF	ODNMF-U	ODNMF-V	ODNMF
MADELON	0.51(78)	0.31(20)	0.17(36)	0.04(24)	2.10(2)	**2.36(2)**
CNAE9	**68.88(93)**	48.12(43)	63.26(64)	53.88(71)	64.84(6)	64.67(6)
Semeion	53.01(86)	47.32(14)	64.96(59)	63.83(16)	**66.70(20)**	65.73(12)
Hillvalley	10.29(2)	2.52(23)	**100.0(10)**	**100.0(10)**	**100.0(10)**	**100.0(10)**
Tr31	41.07(4)	36.79(6)	44.01(38)	42.93(38)	**44.58(8)**	43.98(14)
Re1	44.63(70)	**48.09(28)**	46.40(52)	45.68(52)	46.74(40)	47.21(16)
XM2VTS	61.17(10)	58.75(5)	60.20(66)	60.16(3)	**61.45(98)**	61.17(96)
umist	66.86(79)	48.13(8)	79.73(36)	79.69(52)	**80.18(35)**	79.66(83)
YaleB	27.63(86)	8.64(1)	**45.36(18)**	44.89(18)	20.88(5)	17.86(5)
ORL	74.18(59)	59.31(27)	73.03(65)	74.66(80)	**76.05(73)**	74.05(82)

5 Conclusions

Existing data representation methods are all carried out in data space. However, the information of feature space cannot be fully exploited. To compensate for this drawback, orthogonal dual-graph regularized NMF (ODNMF) is proposed by incorporating DNMF and ONMF. Orthogonal constraints on the basis matrix U and coefficient matrix V are incorporated as the additional condition, which can not only make full use of geometrical structural information underlying the data manifold, but also enhance the generalization performance and robustness of the proposed methods. Since ONMF are more robust to noise, and DNMF has good performance on data clustering, therefore, ODNMF can combine the advantages of ONMF and DNMF together, which has been confirmed on data clustering experiments. The theoretical analysis of the models will be investigated deeply in the future work.

References

1. Gao, H., Nie, F., Cai, W., Huang, H.: Robust capped norm nonnegative matrix factorization. In: Proceedings of the 24th ACM International Conference on Information and Knowledge Management, Melbourne, Australia, October 19–23 (2015)
2. Shi, T., Kang, K., Choo, J., et al.: Short-text topic modeling via non-negative matrix factorization enriched with local word-context correlations. In: Proceedings of the 2018 World Wide Web Conference on World Wide Web. International World Wide Web Conferences Steering Committee, pp. 1105–1114 (2018)
3. Kim, H., Park, H.: Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. Bioinformatics **23**(12), 1495–1502 (2007)
4. Zhang, Z., Li, F., Zhao, M., et al.: Joint low-rank and sparse principal feature coding for enhanced robust representation and visual classification. IEEE Trans. Image Process. **25**(6), 2429–2443 (2016)
5. Ding, C., Li, T., Peng, W., Park, H.: 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. ACM (2006)
6. Li, B., Zhou, G., Cichocki, A.: Two efficient algorithms for approximately orthogonal nonnegative matrix factorization. IEEE Signal Process. Lett. **22**(7), 843–846 (2015)
7. Cai, D., He, X., Han, J., et al.: Graph regularized nonnegative matrix factorization for data representation. IEEE Trans. Pattern Anal. Mach. Intell. **33**(8), 1548–1560 (2011)
8. Huang, S., Wang, H., Ge, Y., et al.: Improved hypergraph regularized Nonnegative Matrix Factorization with sparse representation. Pattern Recogn. Lett. **102**(15), 8–14 (2018)

9. Wang, J.Y., Bensmail, H., Gao, X.: Multiple graph regularized nonnegative matrix factorization. Pattern Recogn. **46**(10), 2840–2847 (2013)
10. Xu, Y., Li, Z., Zhang, B., Yang, J., You, J.: Sample diversity, representation effectiveness and robust dictionary learning for face recognition. Inf. Sci. **375**(1), 171–182 (2017)
11. Wang, C., Song, X., Zhang, J.: Graph regularized nonnegative matrix factorization with sample diversity for image representation. Eng. Appl. Artif. Intell. **68**(2), 32–39 (2018)
12. Mei, J., De Castro, Y., Goude, Y., et al.: Nonnegative matrix factorization with side information for time series recovery and prediction. IEEE Trans. Knowl. Data Eng. **31**(3), 493–506 (2018)
13. He, J., Bi, Y., Liu, B., et al.: Graph-dual Laplacian principal component analysis. J. Ambient Intell. Humaniz. Comput. 1–14 (2018)
14. Shang, F., Jiao, L.C., Wang, F.: Graph dual regularization non-negative matrix factorization for co-clustering. Pattern Recogn. **45**(6), 2237–2250 (2012)
15. Ye, J., Jin, Z.: Dual-graph regularized concept factorization for clustering. Neurocomputing **138**, 120–130 (2014)
16. Yin, M., Gao, J., Lin, Z., et al.: Dual graph regularized latent low-rank representation for subspace clustering. IEEE Trans. Image Process. **24**(12), 4918–4933 (2015)
17. Yao, Y., Tong, H., Yan, G., et al.: Dual-regularized one-class collaborative filtering. In: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, pp. 759–768. ACM (2014)
18. Shang, R., Zhang, Z., Jiao, L., et al.: Self-representation based dual-graph regularized feature selection clustering. Neurocomputing **171**, 1242–1253 (2016)
19. Shang, R., Wang, W., Stolkin, R., et al.: Non-negative spectral learning and sparse regression-based dual-graph regularized feature selection. IEEE Trans. Cybern. **48**(2), 793–806 (2017)
20. Yankelevsky, Y., Elad, M.: Dual graph regularized dictionary learning. IEEE Trans. Signal Inf. Process. Over Netw. **2**(4), 611–624 (2016)
21. Luo, P., Peng, J., Guan, Z., Fan, J.: Dual regularized multi-view non-negative matrix factorization for clustering. Neurocomputing **294**(6), 1–11 (2018)
22. He, J., He, D., Liu, B., Wang, W.: Orthogonal graph regularized nonnegative matrix factorization for image clustering. In: Jin, H., Lin, X., Cheng, X., Shi, X., Xiao, N., Huang, Y. (eds.) BigData 2019. CCIS, vol. 1120, pp. 325–337. Springer, Singapore (2019). https://doi.org/10.1007/978-981-15-1899-7_23