This is the Pre-Published Version.
This is a post-peer-review, pre-copyedit version of an article published in International Journal of Machine Learning and Cybernetics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s13042-015-0396-8.

Nonnegative Matrix Factorization with Manifold Regularization and Maximum Discriminant Information

Wenjun Hu [a,b,c,*], Kup-Sze Choi [c], Jianwen Tao [a], Yunliang Jiang [d,e], Shitong Wang [b]

[a] School of Information and Engineering, Huzhou Teachers College, Huzhou, Zhejiang, China
[b] School of Digital Media, Jiangnan University, Wuxi, Jiangsu, China
[c] Centre for Integrative Digital Health, Hong Kong Polytechnic University, Hong Kong, China
[d] School of Information Science and Engineering, Ningbo
[e] Institute of Technology, Zhejiang University, Ningbo, Zhejiang, China

Abstract: Nonnegative Matrix Factorization (NMF) has been successfully used in different application areas including computer vision, pattern recognition and text mining. Unlike the ordinary NMF, we propose a novel NMF, denoted as MMNMF, which considers both geometrical information and discriminative information hidden in the data. The geometrical information is discovered by minimizing the distance among the representations of the data, while the discriminative information is uncovered by maximizing the distance among base vectors. Clustering experiments are performed on the real-world data sets of faces, images, and documents to demonstrate the effectiveness of the proposed algorithm.

Index Terms—Nonnegative matrix factorization, manifold regularization, maximum information, clustering.

1 INTRODUCTION

Nonnegative Matrix Factorization (NMF) [1], [2] is a popular matrix factorization technique, which decomposes a data matrix into the product of two matrices whose entries is constrained to be nonnegative. With this nonnegative constraint, NMF can be interpreted as a parts-based representation of the data that only allows additive combination but not subtractive, which makes it distinct from other matrix factorization methods, such as Singular Value Decomposition (SVD), Principal Component Analysis (PCA) and Independent Component Analysis (ICA) [3], [4]. A number of studies have shown that the success of NMF in various application fields, including computer vision, pattern recognition [7], text mining [5], [6]. The performance of NMF is especially remarkable in applications concerning face recognition, document representation and brain electromagnetic tomography [8], [9], [10], [11].

Research effort has been devoted to further improve NMF. Ding et al. proposed semi-NMF and convex-NMF to extend the applicability of NMF [11]. The semi-NMF also strengthens the connections between NMF and K-means clustering. More recently, Cai et al. proposed a graph regularized NMF (GNMF) by combining the manifold structure

with NMF [10]. In GNMF, among the two matrices obtained by NMF decomposition, one is regarded as a basis while the other as the new representation of the data matrix. To preserve the low-dimension manifold structure, the new representation is required to keep the local structure invariant. Previous studies have shown that that the resulting basis vectors from NMF can be simply considered as the clustering centroids [11]. It is natural to keep the centroids far away in order to yield more discriminant information.

Motivated by manifold learning, the NMF and the semi-NMF techniques [1], [11], we propose a novel NMF algorithm called the NMF with Manifold regularization and Maximum discriminant information (MMNMF) which takes both the geometrical structure and discriminant information of the data into account, thereby exhibiting the best discriminative power among other algorithms. The rest of the paper is organized as follows. Section 2 provides a brief description of the work related to the proposed algorithm. Section 3 introduces the proposed MMNMF algorithm and discusses the solving scheme. Section 4 presents the clustering experiments used to evaluate the proposed algorithm and the results. Finally, a conclusion is given in Section 5.

2 RELATED WORK

This section gives a brief review of NMF [1], [2], K-Means Clustering Method [12] and GNMF [10] which are related to the proposed MMNMF algorithm.

2.1 NMF

NMF is a matrix factorization technique in numerical linear algebra It decomposes a data matrix into a product of two matrices whose elements are nonnegative. Let $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N]$ be the data matrix with the column vector $\mathbf{x}_i \in \Re^D$. Then, NMF can be formally described by

$$\mathbf{X} \approx \mathbf{U}\mathbf{V}^T, \tag{1}$$

where $\mathbf{U} = [u_{ik}] \in \Re^{D \times K}$ and $\mathbf{V} = [v_{jk}] \in \Re^{N \times K}$ are two matrices with nonnegative entries, and the columns of $\mathbf{U}$ are called the basic vectors. To measure the quality of NMF in (1), Paatero et al. designed two mechanisms based on Euclidean and Divergence distance measurement respectively [1]. In this paper, we focus on the former and the objective function can be expressed as

$$O_1(\mathbf{U}, \mathbf{V}) = \left\| \mathbf{X} - \mathbf{U}\mathbf{V}^T \right\|^2, \tag{2}$$

where $\|\bullet\|$ denotes the Frobenius norm of a matrix. To minimize the objective function in (2), Lee and Seung [13] proposed a multiplicative update algorithm which is formulated as follows

$$u_{ik} \leftarrow u_{ik} \frac{(\mathbf{X}\mathbf{V})_{ik}}{(\mathbf{U}\mathbf{V}^T\mathbf{V})_{ik}} \tag{3}$$

and

$$v_{jk} \leftarrow v_{jk} \frac{(\mathbf{X}^T\mathbf{U})_{jk}}{(\mathbf{V}\mathbf{U}^T\mathbf{U})_{jk}}. \tag{4}$$

2.2 K-means Clustering

K-means clustering [12] aims to partition the data set into K clusters. The objective function is given by

$$O_2(\mathbf{C},\mathbf{V}) = \sum_{i=1}^{N}\sum_{k=1}^{K} v_{ik} \left\|\mathbf{x}_i - \mathbf{c}_k\right\|^2,$$

where $\mathbf{c}_k$ is the class centroid of the *k*th cluster, $v_{ik} \in \{0,1\}$ denotes the cluster indicator (i.e., $v_{ik} = 1$ if the data $\mathbf{x}_i$ belongs to the *k*th cluster; $v_{ik} = 0$ otherwise), $\mathbf{C} = [\mathbf{c}_1, \cdots, \mathbf{c}_K] \in \Re^{D\times K}$ and $\mathbf{V} = [u_{ik}] \in \Re^{N\times K}$. Following some simple algebraic steps, the objective function of K-means clustering can be rewritten as

$$O_2(\mathbf{C},\mathbf{V}) = \left\|\mathbf{X} - \mathbf{C}\mathbf{V}^T\right\|^2. \tag{5}$$

From the description of K-means clustering above, the following observations can be made.

1) K-means clustering has the same objective function as NMF. However, it is noteworthy that $\mathbf{C}$ is not constrained to be nonnegative in K-means clustering.
2) From the perspective of K-means clustering, it implies that the column vectors of $\mathbf{U}$ in NMF are the class centroids.

2.3 Graph Regularized NMF

The GNMF technique is developed by integrating the geometrically-based regularized term with NMF [10]. Its applications in image and document data sets have yielded good performance. In fact, the approximation of NMF in (1) can be considered on per-column basis as follows,

$$\mathbf{x}_j \approx \sum_{k=1}^{K} \mathbf{u}_k v_{jk},$$

where $\mathbf{u}_k$ is the *k*th column vector of $\mathbf{U}$. Clearly, the linear combination of the basis vectors and the entries of $\mathbf{V}$ can be used to approximate each data $\mathbf{x}_j$. This implies that $v_{jk}, \cdots, v_{jK}$ are the coordinates with respect to the $\mathbf{U}$, i.e., the vector $\mathbf{z}_j = [v_{jk}, \cdots, v_{jK}]^T$ can represent the original data $\mathbf{x}_j$ under the basis– $\mathbf{U}$. Thus, by introducing a regularized term in performing the learning of NMF, for inheriting and preserving the underlying manifold structure of the data space where $\mathbf{X}$ is sampled, the objective of GNMF [10] can be expressed as

$$O_3(\mathbf{U},\mathbf{V}) = \left\|\mathbf{X} - \mathbf{U}\mathbf{V}^T\right\|^2 + \lambda \frac{1}{2} \sum_{i,j=1}^{N} \left\|\mathbf{z}_i - \mathbf{z}_j\right\|^2 \mathbf{W}_{ij}, \tag{6}$$

where $\mathbf{W}_{ij}$ is the *ij*th entry of the weight matrix $\mathbf{W}$ on the adjacent graph [10], [17], and λ is a tradeoff parameter. There are many possible definitions for $\mathbf{W}$. Let $N(\mathbf{x}_i)$ denote a set of p nearest neighbors of $\mathbf{x}_i$. We give two common definitions of $\mathbf{W}$. One is called the 0-1 weights, which is given by

$$\mathbf{W}_{ij} = \begin{cases} 1, & \text{if } \mathbf{x}_j \in N(\mathbf{x}_i) \text{ or } \mathbf{x}_i \in N(\mathbf{x}_j) \\ 0, & \text{otherwise} \end{cases}.$$

The other is called the heat kernel weights, which is expressed as

$$\mathbf{W}_{ij} = \begin{cases} \exp(-\dfrac{\left\|\mathbf{x}_i - \mathbf{x}_j\right\|}{2\sigma^2}), & \text{if } \mathbf{x}_j \in N(\mathbf{x}_i) \text{ or } \mathbf{x}_i \in N(\mathbf{x}_j) \\ 0, & \text{otherwise} \end{cases},$$

where σ is the heat kernel parameter with a suitable constant. In (6), the first term on the right-hand side is to decrease the approximated tolerance between $\mathbf{X}$ and $\mathbf{U}\mathbf{V}^T$. The second term requires that the coordinates with respect to the basis $\mathbf{U}$ try to preserve the manifold structure of the data space. That is to say, if $\mathbf{x}_i$ and $\mathbf{x}_j$ are close to each other, so are $\mathbf{z}_i$ and $\mathbf{z}_j$. Further details of GNMF can be found in [10].

3 MANIFOLD REGULARIZED AND DISCRIMINANT INFORMATION MAXIMIZED NMF

While GNMP's performance is remarkably improved by incorporating the geometrically-based regularized term to NMF, it does not take into account the class centroids which are represented to some extent by the column vectors of $\mathbf{U}$. In this section, the MMNMF algorithm is proposed by applying the information between the basis vectors to maximize the discriminant information in the data.

3.1 NMF with Manifold Regularization and Maximum Discriminant Information

Denote the regularized term of GNMF with $R_1(\mathbf{V})$. Then, we have

$$\begin{aligned} R_1(\mathbf{V}) &= \frac{1}{2}\sum_{i,j=1}^{N}\left\|\mathbf{z}_i - \mathbf{z}_j\right\|^2 \mathbf{W}_{ij} \\ &= \sum_{i=1}^{N}\mathbf{z}_i^T\mathbf{z}_i\mathbf{D}_{ii} - \sum_{i,j=1}^{N}\mathbf{z}_i^T\mathbf{z}_j\mathbf{W}_{ij} \quad , \\ &= \mathrm{Tr}(\mathbf{V}^T\mathbf{D}\mathbf{V}) - \mathrm{Tr}(\mathbf{V}^T\mathbf{W}\mathbf{V}) \\ &= \mathrm{Tr}(\mathbf{V}^T\mathbf{L}\mathbf{V}) \end{aligned} \tag{7}$$

where $\mathrm{Tr}(\bullet)$ denotes the trace of a matrix, $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the Laplacian matrix [10], [14], [15], [16], [17], and $\mathbf{D}$ is a diagonal matrix whose entries along the diagonal are the column sum of $\mathbf{W}$, i.e. $\mathbf{D}_{ii} = \sum_j \mathbf{W}_{ij}$. To obtain the maximum geometrical information in the data, it is necessary to minimize the regularized term $R_1(\mathbf{V})$, as in GNMF.

However, the discriminative information in the data is yet ignored to some extent. To capitalize this information, we exploit the data belonging to different classes far away, which can be evaluated by maximizing the following objective function

$$\begin{aligned} R_2(\mathbf{U}) &= \sum_{i=1}^{K} \left\| \mathbf{u}_i - \frac{1}{K} \sum_{j=1}^{K} \mathbf{u}_j \right\|^2 \\ &= \mathrm{Tr}\left(\mathbf{U} \left(\mathbf{I} - \frac{1}{K}\mathbf{E} \right)^T \left(\mathbf{I} - \frac{1}{K}\mathbf{E} \right) \mathbf{U}^T \right), \\ &= \mathrm{Tr}\left(\mathbf{U}\mathbf{M}\mathbf{U}^T \right) \end{aligned} \tag{8}$$

where $\mathbf{E}$ is a $K \times K$ matrix of all 1s, and $\mathbf{M} = \left(\mathbf{I} - \frac{1}{K}\mathbf{E} \right)^T \left(\mathbf{I} - \frac{1}{K}\mathbf{E} \right) = \mathbf{I} - \frac{1}{K}\mathbf{E}$. In this paper, $R_2(\mathbf{U})$ in (8) is called the discriminative information term. By integrating both the regularized term $R_1(\mathbf{V})$ and the discriminative information term $R_2(\mathbf{U})$ with the original NMF, we obtain the proposed MMNMF. That is, the objective function in MMNMF, which is to be minimized, is given by

$$O_4(\mathbf{U},\mathbf{V}) = \left\| \mathbf{X} - \mathbf{U}\mathbf{V}^T \right\|^2 + \lambda_1 R_1(\mathbf{V}) - \lambda_2 R_2(\mathbf{U}), \tag{9}$$

where λ_1 and λ_2 control the tradeoff between three terms on the right-hand side, the first term is the approximated tolerance of $\mathbf{X}$ with $\mathbf{U}\mathbf{V}^T$, the second and the third terms are to exploit the geometrical information and the discriminative information respectively. An iterative algorithm is presented in the next section to solve (9).

3.2 Solution to MMNMF

To solve the objective function of MMNMF, we substitute the equations (7) and (8) into (9) to obtain

$$\begin{aligned} O_4(\mathbf{U},\mathbf{V}) &= \mathrm{Tr}\left((\mathbf{X} - \mathbf{U}\mathbf{V}^T)(\mathbf{X} - \mathbf{U}\mathbf{V}^T)^T \right) + \lambda_1 R_1(\mathbf{V}) - \lambda_2 R_2(\mathbf{U}) \\ &= \mathrm{Tr}(\mathbf{X}\mathbf{X}^T) - 2\mathrm{Tr}(\mathbf{X}\mathbf{V}\mathbf{U}^T) + \mathrm{Tr}(\mathbf{U}\mathbf{V}^T\mathbf{V}\mathbf{U}^T) \\ &\quad + \lambda_1 \mathrm{Tr}(\mathbf{V}^T\mathbf{L}\mathbf{V}) - \lambda_2 \mathrm{Tr}(\mathbf{U}\mathbf{M}\mathbf{U}^T) \end{aligned} . \tag{10}$$

Since all entries of $\mathbf{U}$ and $\mathbf{V}$ are nonnegative, we define the Lagrangian multipliers $\mathbf{U}$ and $\mathbf{V}$ with $\mathbf{\Theta} = [\theta_{ik}]$ and $\mathbf{\Phi} = [\phi_{jk}]$ respectively. Then, the Lagrangian function is given by

$$\begin{aligned} L(\mathbf{U},\mathbf{V}) &= O_4(\mathbf{U},\mathbf{V}) + \mathrm{Tr}(\mathbf{\Theta}\mathbf{U}^T) + \mathrm{Tr}(\mathbf{\Phi}\mathbf{V}^T) \\ &= \mathrm{Tr}(\mathbf{X}\mathbf{X}^T) - 2\mathrm{Tr}(\mathbf{X}\mathbf{V}\mathbf{U}^T) + \mathrm{Tr}(\mathbf{U}\mathbf{V}^T\mathbf{V}\mathbf{U}^T) + \lambda_1 \mathrm{Tr}(\mathbf{V}^T\mathbf{L}\mathbf{V}). \\ &\quad - \lambda_2 \mathrm{Tr}(\mathbf{U}\mathbf{M}\mathbf{U}^T) + \mathrm{Tr}(\mathbf{\Theta}\mathbf{U}^T) + \mathrm{Tr}(\mathbf{\Phi}\mathbf{V}^T) \end{aligned} \tag{11}$$

Setting the partial derivatives of $L(\mathbf{U},\mathbf{V})$ with respect to the primal variables $\mathbf{U}$ and $\mathbf{V}$ to zero will lead to the following formulae

$$\frac{\partial L(\mathbf{U},\mathbf{V})}{\partial \mathbf{U}} = -2\mathbf{XV} + 2\mathbf{UV}^T\mathbf{V} - 2\lambda_2\mathbf{UM} + \mathbf{\Theta} = 0, \tag{12}$$

$$\frac{\partial L(\mathbf{U},\mathbf{V})}{\partial \mathbf{V}} = -2\mathbf{X}^T\mathbf{U} + 2\mathbf{VU}^T\mathbf{U} + 2\lambda_1\mathbf{LV} + \mathbf{\Phi} = 0. \tag{13}$$

According to the Karush–Kuhn–Tucker (KKT) conditions, i.e. $\theta_{ik}u_{ik} = 0$ and $\phi_{jk}v_{jk} = 0$, the formulae in (12) and (13) become

$$-(\mathbf{XV})_{ik}u_{ik} + (\mathbf{UV}^T\mathbf{V})_{ik}u_{ik} - \lambda_2(\mathbf{UM})_{ik}u_{ik} = 0, \tag{14}$$

$$-(\mathbf{X}^T\mathbf{U})_{jk}v_{jk} + (\mathbf{VU}^T\mathbf{U})_{jk}v_{jk} + \lambda_1(\mathbf{LV})_{jk}v_{jk} = 0. \tag{15}$$

Finally, the equations in (14) and (15) yield the following iterative rules

$$u_{ik} \leftarrow u_{ik}\frac{(\mathbf{XV} + \lambda_2\mathbf{U})_{ik}}{(\mathbf{UV}^T\mathbf{V} + \frac{\lambda_2}{K}\mathbf{UE})_{ik}}, \tag{16}$$

$$v_{jk} \leftarrow v_{jk}\frac{(\mathbf{X}^T\mathbf{U} + \lambda_1\mathbf{WV})_{jk}}{(\mathbf{VU}^T\mathbf{U} + \lambda_1\mathbf{DV})_{jk}}. \tag{17}$$

The iterative rules can be implement by using the multiplicative algorithm in [13], [18].

4 EXPERIMENTS

Several experiments are carried out to investigate the effectiveness of the proposed MMNMF algorithm in clustering. We employ the K-means clustering method to evaluate the clustering performance of five algorithms, including K-means clustering in original space [12], normalized cut [19], NMF-based clustering [1], [2], GNMF-based clustering [10] and MMNMF-based clustering.

4.1 Data Preparation

NMF is a powerful technique especially for image clustering and document clustering. Two image data sets and one document data set are therefore used for the clustering experiments. The first data set is obtained from the PIE face database of the Carnegie Mellon University (CMU) (downloadable from http://www.zjucadcg.cn/dengcai). The face images are created under different poses, illuminations and expressions. The database contains 41,368 images of 68 subjects. The image size is 32×32 pixels, with 256 grey levels. From the database, 1428 images under different illumination conditions are selected for the clustering experiment. In the following sections, the classification experiment is first discussed, followed by the clustering experiment.

The second data set, the COIL20 image library[1] from the Columbia University, is used in the classification experiment. It contains 1440 images generated from 20 objects. Each image is represented by a 1024-dimensional vector, and the size is 32 × 32 pixels with 256 grey levels per pixel.

The third data set is the NIST Topic Detection and Tracking (TDT2) Corpus, which contains the corpus of text and transcribed speech and is originated from 6 sources, including 2 newswires, 2 radio programs and 2 television programs. It consists of 11,201 documents. In this experiment, we select part categories for comparison, which contain the 9,394 documents and can be downloaded from http://www.zjucadcg.cn/dengcai. Further details about the data set can be obtained from http://www.itl.nist.gov/iad/mig/tests/tdt/.

4.2 Clustering Experiment

For each data set, the experimental results are conducted repeatedly with different number of clusters K. For the PIE data set, $K = 10, 20, \cdots, 60, 68$; for the COIL20 data set, $K = 2, 4, \cdots, 20$; and for TDT2 data set, $K = 5,10, \cdots, 30$. For a given value of K, the experiment process is described as follows:

1) Randomly select K classes from the data set;
2) Running the corresponding algorithm (except for K-means clustering in original space);
3) Execute K-means clustering method for 20 times with different initialization settings by random and record the best results;
4) Repeat steps 1), 2) and 3) for 20 times (except when K is the maximum value, i.e. $K = 68$ for PIE data set);
5) Conduct the mean and standard error of performance for the given value of K;
6) Change the number of clusters K and repeat steps 1) to 5) until all the values of K have been selected.

Table 1 Clustering Results on PIE

K	Accuracy (%)					Normalized Mutual Information (%)				
	KM	NCut	NMF	GNMF	MMNMF	KM	NCut	NMF	GNMF	MMNMF
5	38.38±7.31	**97.74±6.07**	54.43±7.11	86.31±10.11	90.21±13.07	28.46±11.82	**97.03±3.66**	48.66±7.25	85.07±8.03	88.97±9.04
10	34.05±3.95	**91.29±7.85**	49.17±4.91	85.95±8.11	85.11±8.28	40.04±4.23	**93.74±3.97**	58.37±4.47	89.72±4.49	89.17±3.61
20	28.01±2.40	80.28±3.89	45.16±3.75	81.24±4.77	**81.44±5.10**	44.27±2.70	88.19±2.14	63.81±2.60	89.14±2.14	**89.35±2.15**
30	26.08±1.99	74.29±3.53	42.41±2.33	78.97±4.64	**79.23±3.89**	47.65±2.72	85.11±1.93	66.03±2.11	88.97±1.81	**89.52±1.79**
40	25.79±1.52	70.75±2.38	41.23±2.12	76.69±3.93	**78.29±2.83**	50.48±1.32	84.09±1.42	67.55±1.17	88.61±1.50	**89.65±1.16**
50	25.14±1.28	69.08±2.15	39.95±2.08	74.97±2.52	**76.22±2.93**	51.95±1.21	84.11±1.45	68.22±1.56	88.46±0.97	**88.84±1.03**
60	24.56±0.99	67.12±2.37	39.83±1.75	74.57±3.74	**75.87±2.02**	52.62±0.87	82.46±1.29	69.12±1.09	88.23±1.12	**88.63±0.85**
68	24.30	67.05	40.62	76.12	**79.62**	54.01	82.46	68.77	88.95	**89.42**

[1] http://www1.cs.columbia.edu/CAVE/software/softlib/coil-20.php.

Av.	28.29	77.20	44.10	79.35	**80.75**	46.19	87.15	63.82	88.40	**89.19**

Table 2 Clustering Results on COIL20

K	Accuracy (%)					Normalized Mutual Information (%)				
	KM	NCut	NMF	GNMF	MMNMF	KM	NCut	NMF	GNMF	MMNMF
2	88.47±16.87	**97.19±10.51**	87.95±16.71	95.73±11.98	93.51±14.71	70.52±36.69	**93.28±22.65**	68.29±35.66	89.55±26.61	84.54±32.84
4	82.36±17.35	90.07±14.12	79.01±16.90	93.14±10.67	**93.80±9.83**	75.60±20.25	89.85±12.99	72.10±19.52	91.25±11.16	**91.53±12.32**
6	78.61±10.41	84.53±12.22	76.33±10.47	90.79±8.84	**94.63±5.03**	75.04±11.02	88.77±7.75	72.06±10.90	89.34±9.25	**92.66±5.79**
8	71.57±9.63	83.71±8.76	70.51±7.76	87.60±9.06	**88.95±9.27**	72.67±8.53	88.85±6.11	71.82±8.15	89.67±7.05	**90.25±7.14**
10	72.08±7.14	75.74±5.60	69.84±7.12	85.87±5.95	**86.87±6.33**	74.43±6.40	86.08±3.29	72.38±6.75	89.94±4.20	**90.04±4.09**
12	66.10±6.38	72.42±6.64	66.60±5.53	79.76±5.53	**80.16±5.77**	72.07±4.82	83.83±3.40	71.71±4.07	87.42±3.47	**87.56±3.44**
14	65.98±4.76	74.30±7.24	67.42±6.42	**83.17±5.06**	82.97±5.11	73.46±3.65	85.77±4.16	73.92±4.80	**89.66±3.37**	89.56±3.71
16	65.67±4.69	71.32±5.82	64.31±5.37	78.62±4.55	**79.24±5.56**	73.68±3.97	84.34±3.19	72.97±3.82	87.29±2.79	**88.04±2.95**
18	63.65±4.04	70.55±4.90	65.00±4.44	**79.60±4.90**	79.46±4.02	73.84±2.60	84.45±2.43	73.69±2.43	**88.78±2.47**	88.54±1.94
20	67.78	65.83	70.28	79.79	**82.36**	77.45	82.28	77.51	**90.32**	90.05
Av.	72.23	78.57	71.73	85.41	**86.19**	73.88	86.75	72.64	**89.32**	89.28

Table 3 Clustering Results on TDT2

K	Accuracy (%)					Normalized Mutual Information (%)				
	KM	NCut	NMF	GNMF	MMNMF	KM	NCut	NMF	GNMF	MMNMF
5	84.36±14.68	**98.77±1.92**	83.79±15.64	96.80±6.40	98.23±1.87	76.52±20.50	**94.51±5.11**	75.60±23.34	89.83±13.15	91.86±6.84
10	74.84±13.78	86.56±6.58	73.26±12.47	92.74±5.66	**93.16±5.89**	76.69±12.91	83.50±8.23	73.17±13.00	86.63±9.14	**87.38±7.45**
15	64.98±5.82	82.92±6.87	58.90±6.28	90.26±4.46	**90.43±3.85**	71.87±4.72	81.68±5.39	65.21±4.65	85.52±4.91	**85.72±4.17**
20	66.13±8.89	77.76±9.35	58.57±6.90	87.70±3.34	**87.90±3.94**	74.07±7.10	80.99±6.74	68.63±6.61	85.42±4.25	**85.67±4.00**
25	63.27±4.00	72.32±4.90	53.43±3.09	**87.13±3.80**	86.91±3.60	72.49±1.65	78.10±2.77	66.68±2.49	**84.73±2.80**	84.46±2.79
30	60.52	78.66	52.18	81.59	**86.24**	71.80	81.43	66.88	82.28	**83.35**
Av.	69.02	82.83	63.35	89.37	**90.48**	73.91	83.37	69.36	85.73	**86.41**

The clustering performance on the data sets of PIE, COIL20 and TDT2 are reported in Tables 1, 2 and 3 respectively. In these tables, KM and NCut denote respectively the algorithm of K-means clustering in original space and the normalized cut algorithm. The findings of the experiments are highlighted as follows:

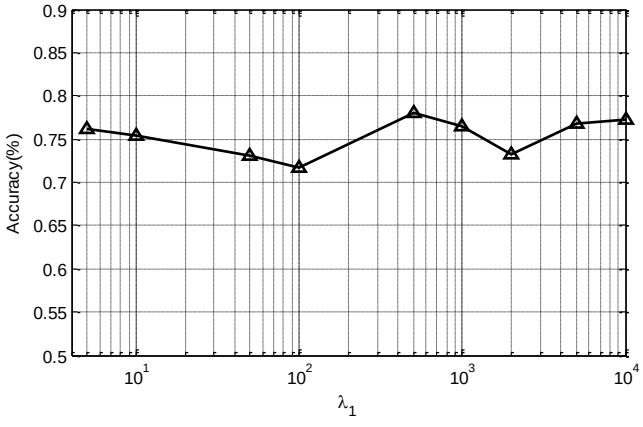
1) Among the three NMF-based methods, i.e. NMF, GNMF and MMNMF, GNMF and MMNMF outperform NMF, which suggests the importance of geometrical structure in discovering the hidden information. Besides, MMNMF achieves better performance than GNMF, as evident from almost all clustering results and the average results. This demonstrates that it is also important to exploit the discriminative information of the base vector in the NMF

techniques.

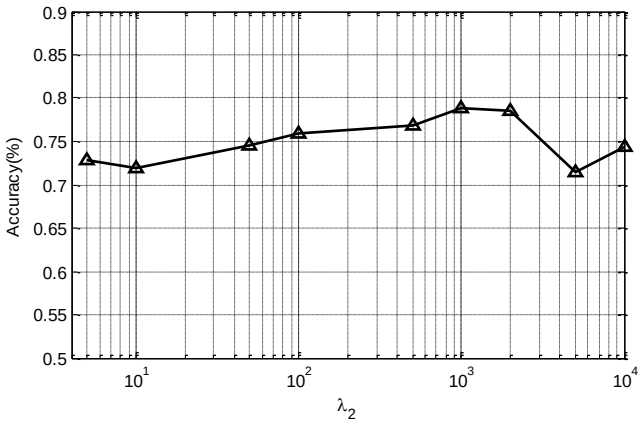
2) The NCut algorithm, which uses the geometrical structure to reveal the hidden information, achieves better results than the original NMF and KM. The result once again demonstrates the importance of the geometrical structure in the clustering process.

4.3 Parameters Selection

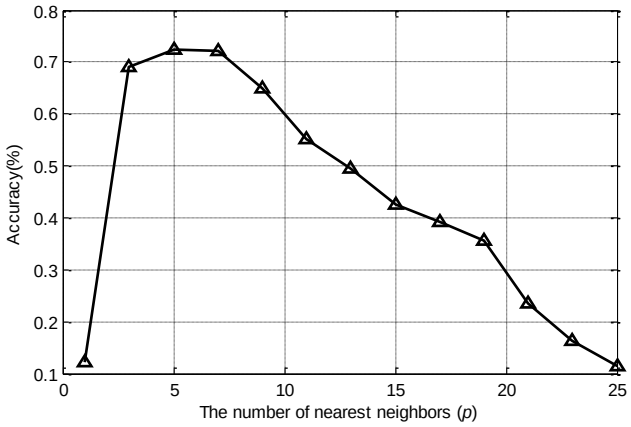
In the proposed MMNMF algorithm, the construction of the weight matrix $\mathbf{W}$ of the adjacent graph requires the setting of three parameters: the control parameters λ_1 and λ_2, and the number of nearest neighbors p. In our experiments, we adopt the 0-1 weights and set $p = 5$, $\lambda_1 = 100$ and $\lambda_2 = 500$ (the values of p and λ_1 are the same as that used in GNMF and MMNMF). To study the effect of each individual parameter on the clustering accuracy, the MMNMF algorithm is performed on the entire PIE data set. Here, we keep tow of the three parameters fixed and vary the third parameters within a certain range. Fig 1 shows how the performance of MMNMF varies with the parameters λ_1, λ_2 and p. In Fig 1(a), we set $p = 5$ and $\lambda_1 = 100$, while λ_2 varies within the grid {5e+0, 1e+1, 5e+1, 1e+2, 5e+2, 1e+3, 2e+3, 5e+3, 1e+4}. In Fig 1(b), $p = 5$, $\lambda_2 = 500$, and λ_1 varies with the same grid as in Fig1(a). In Fig 1(c), $\lambda_1 = 100$, $\lambda_2 = 500$ and p varies within the grid {1, 3, …, 25}. It can been seen that MMNMF is able to achieve good performance over a wide range of λ_1 and λ_2, but the number of nearest neighbors has to be in the range between 3 and 10 in order to achieve an accuracy of 60% or above.



(a)



(b)



(c)

Fig. 1 The impact of three parameters λ_1, λ_2 and p on the clustering performance.

5 CONCLUSION AND FUTURE WORK

This paper proposes a novel NMF algorithm called Manifold Regularization and Maximum Discriminant Information NMF (MMNMF), which is motivated by the notion that the geometrical structure and discriminant information are important for data clustering. The MMNMF algorithm exploits both geometrical and discriminative information in order to obtain better clustering performance than the ordinary NMF and GNMF. Experiments performed on the PIE, COIL20 and TDT2 data sets have demonstrated the advantage of MMNMF in data clustering.

Like GNMF, the proposed MMNMF also takes the geometric structure into account. They require the construction of an adjacent graph to discover the intrinsic structure information. However, theoretical selection of the suitable size of the *k*-nearest neighbors to match the local structure remains an issue for both algorithms, which affects their clustering performance. As discussed in Section 4.3, the number of nearest neighbors is required to be within 3 to 10, otherwise the performance decreases considerably. The theoretical selection of the two parameters λ_1 and λ_2 in MMNMF, which control the tradeoff between the geometric structure information and the discriminant information, is also an issue, although we have demonstrated experimentally that the performance of MMNMF is not sensitive to the two parameters. Further research effort will therefore be dedicated to the establishment of theoretical base for the selection of the MMNMF parameters.

ACKNOWLEDGEMENTS

This work was supported in part by General Research Fund of the Hong Kong Research Grants Council under Grant PolyU 5134/12E; the National Natural Science Foundation of China under Grants 61170122, 61272210 and 61202311; the Natural Science Foundation of Zhejiang Province under Grants LY13F020011, LY14F010010 and LY14F020009; the Humanities and Social Science Foundation of Ministry of Education of China under Grant 13YJAZH084; the Natural

Science Foundation of Jiangsu Province under Grants BK2011003 and BK2011417; the Natural Science Foundation of Huzhou City under Grant 2013YZ05; Huzhou University science research project under Grants KX24063 and KX24058.

REFERENCES

[1] P. Paatero, U. Tapper, "Positive Matrix Factorization—A Nonnegative Factor Model with Optimal Utilization of Error Estimates of Data Values," *Environmetrics*, vol. 5, no. 2, pp. 111-126, 1994.

[2] D. D. Lee, H. S. Seung, "Learning the Parts of Objects by Nonnegative Matrix Factorization," *Nature*, vol. 401, pp. 788-791, Oct. 1999.

[3] A. Hyvärinen, E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411-430, Dec. 2000.

[4] A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis. John Wiley and Sons, 2001.

[5] W. Xu, X. Liu, and Y. Gong, "Document Clustering Based on Non-Negative Matrix Factorization," *Proc. ACM Conf. Research and Development in Information Retrieval (SIGIR)*, pp. 267-273, 2003.

[6] V. P. Pauca, F. Shahnaz, M. Berry, R. Plemmons, "Text Mining Using Non-Negative Matrix Factorization," *Proc. SIAM Int'l Conf. Data Mining*, pp. 452-456, 2004.

[7] S. Li, X. Hou, H. Zhang, Q. Cheng, "Learning Spatially Localized, Parts-Based Representation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 207-212, 2001.

[8] F. Shahnaza, M.W. Berrya, V. Paucab, and R.J. Plemmonsb, "Document Clustering Using Nonnegative Matrix Factorization," *Information Processing and Management*, vol. 42, no. 2, pp. 373-386, Mar. 2006.

[9] A. Pascual-Montano, J. M. Carazo, K. Kochi, D. Lehmann, R. D. Pascual-Marqui, "Nonsmooth nonnegative matrix factorization (nsNMF)," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 403-415, Mar. 2006.

[10] D. Cai, X,He, J. Han, T. Huang, "Graph regularized non-negative matrix factorization for data representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1548-1560, Aug. 2011.

[11] C. Ding, T. Li, M. I. Jordan, "Convex and Semi-Nonnegative Matrix Factorizations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 45-55, Jan. 2010.

[12] J. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations," *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, pp. 281-297, 1967.

[13] D.D. Lee, H.S. Seung, "Algorithms for Non-Negative Matrix Factorization," *Advances in Neural Information Processing Systems 13*, MIT Press, 2001.

[14] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," *Advances in Neural Information Processing Systems 14*, pp. 585-591, MIT Press, 2001.

[15] Fan R. K. Chung, Spectral Graph Theory, *Regional Conference Series in Mathematics*, no. 92, 1997.

[16] X. He, D. Cai, S. Yan, H. Zhang, "Neighborhood preserving embedding," *Proceeding in International Conference on Computer Vision (ICCV)*, 2005.

[17] X. He, P. Niyogi, "Locality preserving projections," *Proceeding in Conference Advances in Neural Information Processing Systems (NIPS)*, 2003.

[18] C. J. Lin, "On the Convergence of Multiplicative Update Algorithms for Non-Negative Matrix Factorization," *IEEE Transactions on Neural Networks*, vol. 18, no. 6, pp. 1589-1596, Nov. 2007.

[19] J. Shi, J. Malik, "Normalized Cuts and Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888-905, Aug. 2000.