Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)

Local Centroids Structured Non-Negative Matrix Factorization

Hongchang Gao,[1] **Feiping Nie,**[2,1*] **Heng Huan**[1*]
[1]Department of Computer Science and Engineering, University of Texas at Arlington, Texas, USA
[2]School of Computer Science, OPTIMAL, Northwestern Polytechnical University, Xian 710072, Shaanxi, P. R. China
{hongchanggao, feipingnie}@gmail.com, heng@uta.edu

Abstract

Non-negative Matrix Factorization (NMF) has attracted much attention and been widely used in real-world applications. As a clustering method, it fails to handle the case where data points lie in a complicated geometry structure. Existing methods adopt single global centroid for each cluster, failing to capture the manifold structure. In this paper, we propose a novel local centroids structured NMF to address this drawback. Instead of using single centroid for each cluster, we introduce multiple local centroids for individual cluster such that the manifold structure can be captured by the local centroids. Such a novel NMF method can improve the clustering performance effectively. Furthermore, a novel bipartite graph is incorporated to obtain the clustering indicator directly without any post process. Experiments on both toy datasets and real-world datasets have verified the effectiveness of the proposed method.

Introduction

Non-negative Matrix Factorization (NMF) has attracted extensive attention during the past two decades. It is to factorize a non-negative matrix into two non-negative factor matrices. Such a non-negativity property makes it easy to interpret for real-world data mining applications, compared with the general matrix factorization method, such as SVD and QR whose elements can be both positive and negative.

The seminal works (Lee and Seung 1999; 2001) give a parts-based-representation explanation about the factor matrices and propose how to compute them efficiently, which facilitates the development of NMF. Furthermore, the relationship between NMF and K-means, spectral clustering is disclosed in (Ding, He, and Simon 2005) where NMF is proved to be a clustering method. To improve its clustering performance, (Ding et al. 2006) proposed the orthogonal NMF where the indicator matrix has the orthogonal constraint. Such a constraint makes the indicator matrix unique and be easily interpreted in a clustering view (Ding et al. 2006). Then, NMF is widely used for clustering problems.

*Corresponding authors. This work was partially supported by the following grants: NSF-IIS 1302675, NSF-IIS 1344152, NSF-DBI 1356628, NSF-IIS 1619308, NSF-IIS 1633753, NIH R01 AG049371.

To address different problems, numerous variants have been proposed in the past ten years. For instance, semi-NMF and convex-NMF are proposed in (Ding, Li, and Jordan 2010). Semi-NMF relaxes the non-negative constraint of data matrix and basis matrix, while convex-NMF restricts the basis is a convex combination of the data points. Tri-Factorization is proposed in (Ding et al. 2006) to factorize a non-negative data matrix into three non-negative factor matrices. This model performs clustering on both row space and column space simultaneously, which makes it suitable for document analysis. To tackle noises and outliers existing in data matrix, $\ell_{2,1}$-norm NMF and capped norm NMF are proposed to solve different applications (Kong, Ding, and Huang 2011; Gao et al. 2015; Huang et al. 2014; Wang, Nie, and Huang 2015). With these models, NMF has been widely used for many data mining applications, such as face recognition (Gao et al. 2015), text and document mining (Ding, Li, and Peng 2006), human action analysis (Wang, Nie, and Huang 2016) and so on.

As a clustering method, the two factor matrices of NMF correspond to the clustering centroid and indicator respectively. The clustering result is indicated by the maximal element of each indicator vector, which means the corresponding centroid can be viewed as the representative of that cluster. Almost all the existing NMF methods use only one centroid to represent each cluster. Such a centroid can be viewed as a global representative and it depicts its cluster very coarsely due to losing a lot of local manifold structure. Data points in real-world applications usually lie in a complicated geometry structure, the global centroid is easy to fail in depicting its cluster. Therefore, how to capture the complicated local manifold structure is important to NMF.

In this paper, we propose a novel local centroids structured NMF to handle data points lying in a complicated geometry structure. The novelty lies in three aspects. First, instead of using single centroid for each cluster, we introduce multiple local centroids for individual cluster such that the manifold structure can be captured by the local centroids. Second, although there exists multiple local centroids in each cluster, we do not employ all of them to represent each data point. To preserve the local manifold information, we only adopt the nearby centroids to represent each data point. Third, since each data point is encoded by a few nearby centroids, the clustering indicator cannot be obtained

directly. Our method employs a bipartite graph and enforces it with exact c (c is the number of clusters) connected components to get the clustering indicator directly. Extensive experiments have been performed to verify the correctness and effectiveness of our novel method.

Related Work

In this section, we will review the related work and give the motivation of our method.

Non-negative Matrix Factorization

The classic NMF (Lee and Seung 1999; 2001) is to factorize X into two factor matrices as follows:

$$\min_{F\geq 0, G\geq 0} ||X - FG^T||_F^2 \,, \tag{1}$$

where $X \in \Re^{d\times n}$ is a non-negative data matrix, $F \in \Re^{d\times k}$ and $G \in \Re^{n\times k}$. As a clustering method (Ding, He, and Simon 2005), k is set as the number of clusters. Therefore, each column of F corresponds to a cluster centroid, and each row of G is the clustering indicator of each sample. The final clustering result is indicated by the index of maximal element in each row of G.

Graph Regularized NMF

Actually, data points in real applications usually lie in a low dimensional manifold rather than distribute uniformly in the high dimensional ambient space, which means that data points have complicated relationships, i.e. the similarity should be measured via geometry structure or manifold, not the standard Euclidean distance. If the method can capture the local manifold structure, the similarity between data points can be properly measured such that the clustering results are enhanced. Graph regularized NMF (GNMF) proposed in (Cai et al. 2011) tackles the local manifold information by incorporating a nearest neighbor graph into NMF as follows:

$$\min_{F\geq 0, G\geq 0} ||X - FG^T||_F^2 + \lambda Tr(G^T LG) \,, \tag{2}$$

where $L \in \Re^{n\times n}$ is Laplacian matrix defined as $L = D - W$. $W \in \Re^{n\times n}$ is the affinity matrix of data points, and $D \in \Re^{n\times n}$ is a diagonal matrix defined as $D_{ii} = \sum_{j=1}^{n} W_{ij}$. With the graph regularization, two close data points in the original space are enforced to be close in the low dimensional space which is spanned by columns of F. In other words, GNMF preserves the local manifold information during factorization so that the clustering performance will be improved.

However, data points in real applications usually lie in a complicated geometry structure, such as two crescent shapes in Figure 1, which is difficult to existing NMF methods. In Figure 2, we show the clustering result of NMF and GNMF running on this toy data set. Both NMF and GNMF fail to cluster these data points correctly. Although GNMF incorporates local manifold information and achieves better result, yet it still fails to find the correct clustering result.

Why do these methods fail? From Figure 2, we can find the centroids found by NMF and GNMF are not proper to represent its own group. Thus, some data points near to the opposite centroid are mis-clustered. In more details, there is only one centroid to represent each cluster globally. Such a global centroid can only represent a cluster very coarsely so that the local manifold structure cannot be captured. That is the reason of GNMF's failure. GNMF only incorporates local manifold information into the new representation G but not the centroid F. In this paper, we will propose a novel NMF method to address this difficult problem so that data points in Figure 1 can be grouped correctly.

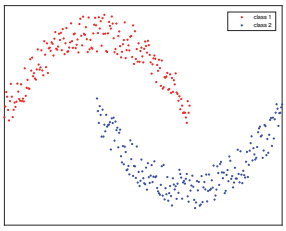
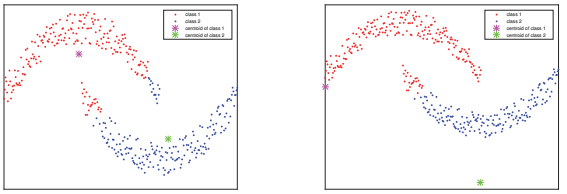


Figure 1: A toy dataset with complicated geometry structure.



(a) Clustering result of NMF (b) Clustering result of GNMF

Figure 2: Clustering result of NMF and GNMF

Local Centroids Structured Non-negative Matrix Factorization

In this section, we will propose our novel non-negative matrix factorization, that is Local Centroids Structured Non-negative Matrix Factorization.

Given a non-negative dataset $X = [x_1, x_2, \cdots, x_n] \in \Re^{d\times n}$ whose distribution may have a complicated geometry structure, and assume it has c clusters. Existing NMF methods fail to tackle such a dataset since their centroids cannot capture the local manifold structure. Instead of employing single centroid for each cluster, we introduce multiple local centroids for individual cluster such that the manifold structure can be captured by the local centroids. A direct implementation is as follows:

$$\min_{F\geq 0, G\geq 0} ||X - FG^T||_F^2 \,, \tag{3}$$

where $F \in \Re^{d\times k}$ and $G \in \Re^{n\times k}$. Here, k is a hyperparameter, and we set $k = mc$ where $m > 1$ is an integer so that each cluster will have multiple centroids. This is the difference between Eq. (3) and Eq. (1) whose $k = c$. With

multiple centroids, the new basis F will have more powerful ability to catpure the manifold information than existing NMF methods. Moreover, existing methods usually incorporate manifold information by the *regularization* term, which is not enough to capture it. Here, we incorporate manifold information into the *cost* function, which is the first work to do so as far as we know.

However, there are two weaknesses in this model. First, although the basis F incorporates the local manifold structure, yet the new representation G ignores the local manifold information. As a result, the clustering result is suboptimal. Second, due to the multiple centroids, we cannot directly get the clustering indicator from G like existing methods. We have to perform K-means on G to get the final discrete clustering indicator. However, K-means is sensitive with initialization, and it is a non-convex problem so that it may converge to a local suboptimal solution. To address these problems, we propose the following model:

$$\begin{aligned} &\min_{F\geq 0, G\geq 0} ||X - FG^T||_F^2 + \lambda R(G) \\ &s.t. ||G_{i.}||_0 \leq s, i = 1, 2, \cdots, n \end{aligned} \tag{4}$$

where $\lambda > 0$, $G_{i.}$ is the i-th row of G, $||.||_0$ denotes ℓ_0-norm and $s > 0$ is a hyper-parameter. $R(G)$ is the regularization with respect to G to obtain the clustering result directly, which will be defined later.

Here, to incorporate the local manifold information into new representation G, we restrict $||G_{i.}||_0 \leq s$ which means each data point is represented by at most s centroids. Intuitively, for a data point locating in a complicated geometry structure, if using all of centroids in this geometry structure to represent it, it may be disturbed by some far away centroids, destroying the local manifold structure. On the other hand, the nearby centroids have most of manifold information so that a data point can be represented by only a few nearby centroids.

When adopting multiple local centroids, the largest element in each row of G cannot indicate the final clustering result as the conventional NMF method. To obtain the clustering indicator directly, we construct a bipartite graph with the affinity matrix S as follows:

$$S = \begin{bmatrix} \mathbf{0} & G \\ G^T & \mathbf{0} \end{bmatrix}, \tag{5}$$

where $S \in \Re^{(n+k)\times(n+k)}$. Ideally, data points from a certain cluster will be represented by only the centroids from the same cluster. Therefore, they will constitute a connected components in this graph. By enforcing the bipartite graph with exact c connected components, then we can directly obtain the clustering indicator based on the graph partition. We show an illustration of bipartite graph in Figure 3. Data points in each group are represented by their own centroids, thus each group is a connected component.

Furthermore, to enforce a graph with exact c connected components, we need the following theorem (Mohar et al. 1991).

Theorem 1. *The number of connected components in the graph is equal to the multiplicity c of the eigenvalue 0 of its Laplacian matrix L_S.*

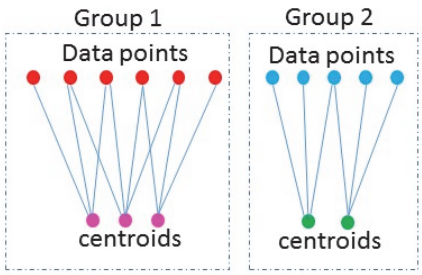


Figure 3: An illustration of bipartite graph constructed on data points and centroids.

Here, the Laplacian matrix is defined as $L_S = D - S$ where D is a diagonal matrix with diagonal elements $D_{i,i} = \sum_j S_{ij}$. Theorem 1 indicates that $rank(L_S)$ should be $(n + k) - c$ to ensure c connected components existing in our bipartite graph.

Following the proof in (Nie, Wang, and Huang 2014), $rank(L_S) = (n + k) - c$ is equivalent to enforce $\sum_{i=1}^{c} \sigma_i(L_S) = 0$ where $\sigma_i(L_S)$ is the i-th smallest eigenvalue of Laplacian matrix L_S, and furthermore,

$$\sum_{i=1}^{c} \sigma_i(L_S) = \min_{P\in\Re^{(n+k)\times c}, P^TP=I} Tr(P^T L_S P) . \tag{6}$$

Then, to directly obtain the final clustering result, we have our final model as follows:

$$\begin{aligned} &\min_{F\geq 0, G\geq 0, P} ||X - FG^T||_F^2 + \lambda Tr(P^T L_S P) \\ &s.t. P^T P = I, ||G_{i.}||_0 \leq s, i = 1, 2, \cdots, n \end{aligned} \tag{7}$$

where $\lambda > 0$ is a hyper-parameter. With the second term in the objective function, the bipartite graph is enforced to have c connected components. As a result, we can directly obtain the clustering indicator according to the bipartite graph's partition .

In a summary, our proposed method adopts multiple local centroids to capture the local manifold structure. At the same time, each data point is restricted to be represented by only a few nearby centroids to preserve the local manifold information. Furthermore, a bipartite graph is constructed to be enforced with exact c connected components, thus our method can obtain the clustering result directly from the partition of this graph.

In the next section, we will propose an efficient algorithm to solve this non-convex problem.

Optimization Algorithm

The problem defined in Eq. (7) is a non-convex problem. Here, we adopt the alternating algorithm to solve it efficiently.

Update P When fixing F and G, we have the following problem with respect to P:

$$\begin{aligned} &\min_{P\in\Re^{(n+k)\times c}} Tr(P^T L_S P) \\ &s.t. P^T P = I . \end{aligned} \tag{8}$$

The optimal solution is the eigenvectors corresponding to the c smallest eigenvalues of L_S.

Update F When fixing P and G, we have

$$\min_{F\geq 0} ||X - FG^T||_F^2 \tag{9}$$

which can be solved by multiplicative method (Lee and Seung 2001) as follows:

$$F_{ij} \leftarrow F_{ij}\frac{(XG)_{ij}}{(FG^TG)_{ij}} \tag{10}$$

Algorithm 1 Algorithm to solve Eq. (15)

Input: $F \in \Re^{d\times k}$, $x \in \Re^{d\times 1}$, $a \in \Re^{1\times k}$, $\alpha > 0$, $\eta > 0$, $s > 0$

Output: g

repeat

Step 1: compute the gradient

$$\nabla_g = 2(\alpha a^T - F^T x + F^T F g)$$

Step 2: projected gradient descent

$$g = \max(g - \eta * \nabla_g, \mathbf{0})$$

Step 3: solve ℓ_0-norm constraint

$$g(i) = \begin{cases} g(i), & g(i)\ is\ the\ largest\ \text{s}\ value\ of\ g \\ 0, & otherwise \end{cases}$$

until Converges

Update G When fixing P and F, we have

$$\begin{aligned} &\min_{G\geq 0} ||X - FG^T||_F^2 + 2\lambda Tr(P^T L_S P) \\ &s.t. ||G_{i.}||_0 \leq s, i = 1, 2, \cdots, n \end{aligned} \tag{11}$$

To solve Eq. (11), we need the following lemma, which is proved in (Von Luxburg 2007).

Lemma 1. *Given an affinity matrix $W \in \Re^{n\times n}$ and a diagonal matrix D defined as $D_{i,i} = \sum_j W_{ij}$, for its Laplacian matrix $L \in \Re^{n\times n}$ defined as $L = D - W$ and a matrix $P \in \Re^{n\times c}$, we have*

$$Tr(P^T LP) = \frac{1}{2}Tr(WQ)\,, \tag{12}$$

where $Q_{ij} = ||p^i - p^j||_2^2$, and p^i is the i-th row of matrix P.

According to Lemma 1, we have

$$\begin{aligned} Tr(P^T L_S P) &= \frac{1}{2}Tr(SQ) \\ &= \frac{1}{2}Tr(\begin{bmatrix} \mathbf{0} & G \\ G^T & \mathbf{0} \end{bmatrix}\begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}) \\ &= \frac{1}{2}Tr(\begin{bmatrix} GQ_{21} & GQ_{22} \\ G^TQ_{11} & G^TQ_{12} \end{bmatrix}) \\ &= \frac{1}{2}Tr(G^T(Q_{21}^T + Q_{12})) \\ &= \frac{1}{2}Tr(G^T A) \end{aligned} \tag{13}$$

where $A = Q_{21}^T + Q_{12}$. Therefore, Eq. (11) can be reformulated as follows,

$$\begin{aligned} &\min_G ||X - FG^T||_F^2 + 2\alpha Tr(G^T A) \\ &s.t. G \geq 0, ||G_{i.}||_0 \leq s, i = 1, 2, \cdots, n \end{aligned} \tag{14}$$

where $\alpha = \lambda/2$ is a large number to ensure there exists c connected components. It can be further decoupled to solve each column g_i of G^T separately as follows,

$$\begin{aligned} &\min_{g_i} ||x_i - Fg_i||_2^2 + 2\alpha a_i g_i \\ &s.t. g_i \geq 0, ||g_i||_0 \leq s\,, \end{aligned} \tag{15}$$

where x_i is the i-th corresponding column of X, a_i is the i-th row of A. This problem can be solved by projected gradient descent method as shown in Algorithm 1. Finally, the optimization algorithm to solve Eq. (7) is summarized in Algorithm 2.

Algorithm 2 Algorithm to solve Eq. (7)

Initialize F and G by K-means.

repeat

Update P by solving Eq. (8):

$$\min_{P^TP=I} Tr(P^T L_S P)$$

Update F by Eq. (10):

$$F_{ij} \leftarrow F_{ij}\frac{(XG)_{ij}}{(FG^TG)_{ij}}$$

Update G: solve each row of G by Algorithm 1.

until Converges

Initialization Since Eq. (7) is a non-convex problem, it is important to give it a good initialization. Following (Gao et al. 2015), K-means is taken to initialize it. Specifically, one can perform K-means on data points to get k clusters where $k = mc$ so that there should be multiple local centroids. F can be initialized with these centroids, and G can be initialized by the similarity between each data point and the nearest s centroids.

Determination of λ The stop criterion is that there are exact c connected components. Therefore, it is important to tune the parameter λ. Here, we use the method proposed in (Nie, Wang, and Huang 2014) where λ can be automatically adjusted to obtain exact c connected components. More details can be found in the corresponding literature.

Complexity Analysis When updating P, the complexity is dominated by eigenvalue decomposition which is $O((n+k)^3)$. The complexity of updating F is $O(ndk)$, and it is also $O(ndk)$ for updating G. Compared with conventional NMF whose complexity is $O(ndc)$, the increased computation is mainly due to updating P.

Experiment

In this section, we will verify the correctness and effectiveness of our proposed method on both toy dataset and real world datasets.

Toy Dataset

Here, we run our proposed method on the complicated toy dataset as shown in Figure 1 where the first 201 data points belong to group 1, and the rest belong to group 2. In our experiment, we use K-means to initialize F and G, and k is set as 10. Thus, this toy dataset is clustered into 10 groups by K-means, just as shown in Figure 4. These centroids are used to initialize F. Additionally, the parameter s in Eq. (7) is set as 2, which means each data point is represented by only two nearest centroids, just as shown in Figure 5. G is then initialized with the similarity between data points and their two nearest centroids. Apparently, there exists some connections between these two groups. After running our proposed method with the above initialization, it can be clustered perfectly.

To verify the result, we show the image of GG^T in Figure 6. Note that GG^T denotes the similarity matrix between data points. In Figure 6(a), there exist non-zero connections between the first 201 data points and the remained 201 data points. In Figure 6(b), there is no connection between two groups which means that our method can cluster this complicated toy dataset correctly.

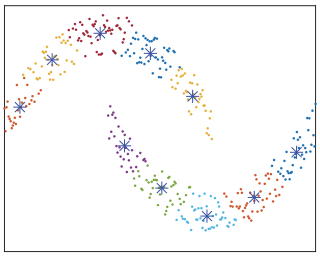


Figure 4: 10 clusters obtained by K-means, and asteroids denote centroids.

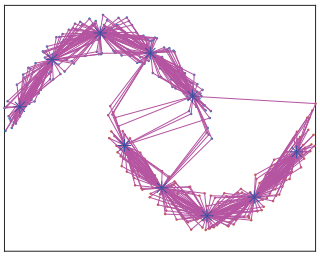
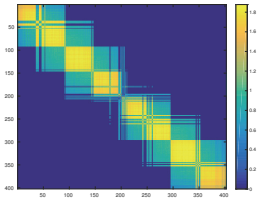


Figure 5: Each data point is represented by its two nearest centroids.

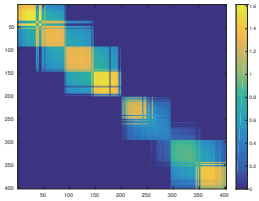
Real World Dataset

To further verify the effectiveness of our proposed method, we evaluate it on four real world benchmark datasets.

Data Description The details about four benchmark dataset are described as follows.



(a) The initialized GG^T



(b) GG^T of our method

Figure 6: Comparison of the similarity matrix GG^T obtained by initialization and our proposed method.

Table 1: Description of Benchmark Datasets

Dataset	#Instance	#Dimensionality	#Classes
ORL	400	644	40
UMIST	575	644	20
PIE	680	1024	68
COIL20	1440	1024	20

- ORL (Samaria and Harter 1994) is a face recognition benchmark dataset. It consits of 40 subjects, and each subject has 10 images taken under various conditions. The task is to cluster images from the same subject together. Here, each image is scaled to 28×23.
- UMIST (Graham and Allinson 1998) is also a face recognition benchmark dataset. There are 20 different subjects. Each subject is taken images from different views, obtaining 575 images totally. Each image in our experiment is rescaled to 28×23.
- PIE (Sim, Baker, and Bsat 2002) is another face recognition benchmark dataset, which contains 68 subjects. Each subject is taken 42 images under different conditions. Here, we randomly select 10 images for each subject, and each images is rescaled to 32×32.
- COIL20 (Nene et al. 1996) is an object recognition benchmark dataset. There are totally 20 objects, and each object has 72 images. Each image is resized to 32×32 in our experiment.

In our experiment, each data point is normalized to have an unit length. We summarize these datasets in Table 1.

Experiment Setup To evaluate the performance of our proposed method, we compare it with five state-of-the-art methods as follows. With in line mode this is typeset as $\lim_{x \to 2} f(x) = 5$

- NMF (Lee and Seung 2001) is defined in Eq. (1). In our experiment, we set k as the number of classes.
- Orthogonal NMF (ONMF) (Ding et al. 2006) is defined as $\min_{F,G \geq 0, G^T G = I} ||X - FG^T||_F^2$, where $F \in \Re^{d \times k}, G \in \Re^{n \times k}$. Theoretically, ONMF has better clustering performance than conventional NMF. We also set k as the number of classes.

Table 2: Clustering Result on Benchmark Datasets

Dataset	Metric	K-Means	NMF	ONMF	$\ell_{2,1}$-NMF	GNMF	Our
ORL	ACC	0.6885	0.6693	0.7140	0.6935	0.7275	**0.7608**
	NMI	0.8337	0.8131	0.8404	0.8228	0.8504	**0.8891**
UMIST	ACC	0.4670	0.4402	0.4798	0.4485	0.6546	**0.6708**
	NMI	0.5982	0.5654	0.6048	0.5745	0.7520	**0.8020**
PIE	ACC	0.3000	0.4059	0.3544	0.4191	0.3544	**0.4583**
	NMI	0.5860	0.6670	0.6388	**0.6810**	0.6401	0.6428
COIL20	ACC	0.6327	0.5256	0.6344	0.6353	0.7530	**0.7783**
	NMI	0.7365	0.6601	0.7244	0.7198	0.8750	**0.8995**

- $\ell_{2,1}$-norm based robust NMF ($\ell_{2,1}$-NMF) (Kong, Ding, and Huang 2011) is defined as $\min_{F\geq 0, G\geq 0} ||X - FG^T||_{2,1}$, where $F \in \Re^{d\times k}, G \in \Re^{n\times k}$. Here, k is set as the number of classes. This method is robust to noises and outliers theoretically.
- Graph regularized NMF (GNMF) (Cai et al. 2011) is defined in Eq. (2). k is also set as the number of classes in our experiment.

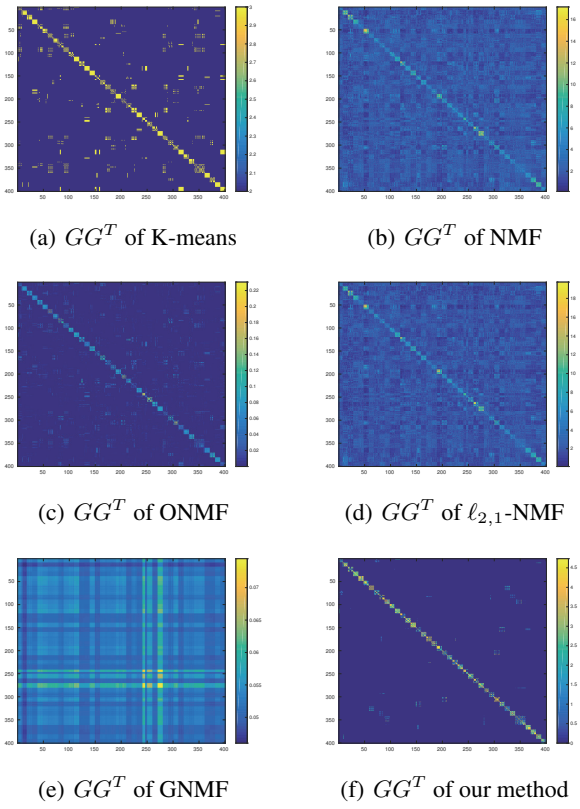
Other than these NMF methods, we also compare it with K-means. For all the above NMF methods except GNMF, the clustering result is directly indicated by the largest element in each row of G. GNMF needs K-means as the post-processing step to get clustering indicator since its G has no clear structure. We run all the methods for 10 times and report the average clustering accuracy (ACC) and normalized mutual information (NMI) in Table 2.

Here, we set the number of centroids k in our method around 80%-90% of the number of data points in each cluster, and each data point is restricted to be represented by 3-5 nearby centroids. This setting is consistent with our intuition and verified by our experiment on the toy dataset. Only with a lot of local centroids can a complicated geometry structure be characterized correctly. Furthermore, for a data point lying a complicated structure, only nearby centroids can provide its manifold information and the far away centroids provide considerably few information. Therefore, the setting of our method is reasonable and further verified by Table 2. From Table 2, we can find that our method and GNMF have better performance than the others due to incorporating manifold information, and our method is better than GNMF due to the multiple local centroids.

To further verify the performance of our method, we plot GG^T of different methods running on the ORL dataset in Figure 7. Apparently, our method has much clearer block diagonal structure than others which means our method has better clustering result.

Conclusion

In this paper, we propose a novel local centroids structured non-negative matrix factorization. This method can successfully handle the dataset with a complicated geometry structure. Specifically, our method use multiple local centroids to capture the local manifold structure. At the same time, the new representation preserves local manifold information by enforcing each data point to be represented by only a few nearby centroids. Furthermore, a novel bipartite graph is adopted to ensure there exist exact c connected components so that our method can get clustering result directly. The experimental results on both toy dataset and real world datasets have shown the success of our proposed method.



(a) GG^T of K-means (b) GG^T of NMF

(c) GG^T of ONMF (d) GG^T of $\ell_{2,1}$-NMF

(e) GG^T of GNMF (f) GG^T of our method

Figure 7: Comparison of the similarity matrix GG^T obtained by different methods running on the ORL dataset.

References

Cai, D.; He, X.; Han, J.; and Huang, T. S. 2011. Graph regularized nonnegative matrix factorization for data represen-

tation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33(8):1548–1560.

Ding, C.; Li, T.; Peng, W.; and Park, H. 2006. Orthogonal nonnegative matrix t-factorizations for clustering. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 126–135. ACM.

Ding, C. H.; He, X.; and Simon, H. D. 2005. On the equivalence of nonnegative matrix factorization and spectral clustering. In *SDM*, volume 5, 606–610. SIAM.

Ding, C. H.; Li, T.; and Jordan, M. I. 2010. Convex and semi-nonnegative matrix factorizations. *IEEE transactions on pattern analysis and machine intelligence* 32(1):45–55.

Ding, C.; Li, T.; and Peng, W. 2006. Nonnegative matrix factorization and probabilistic latent semantic indexing: Equivalence chi-square statistic, and a hybrid method. In *Proceedings of the national conference on artificial intelligence*, volume 21, 342. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.

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

Graham, D. B., and Allinson, N. M. 1998. Characterising virtual eigensignatures for general purpose face recognition. In *Face Recognition*. Springer. 446–456.

Huang, J.; Nie, F.; Huang, H.; and Ding, C. 2014. Robust manifold non-negative matrix factorization. *ACM Transactions on Knowledge Discovery from Data* 8(3):11:1–11:21.

Kong, D.; Ding, C.; and Huang, H. 2011. Robust nonnegative matrix factorization using l21-norm. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, 673–682. ACM.

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

Lee, D. D., and Seung, H. S. 2001. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, 556–562.

Mohar, B.; Alavi, Y.; Chartrand, G.; and Oellermann, O. 1991. The laplacian spectrum of graphs. *Graph theory, combinatorics, and applications* 2(871-898):12.

Nene, S. A.; Nayar, S. K.; Murase, H.; et al. 1996. Columbia object image library (coil-20). Technical report, Technical report CUCS-005-96.

Nie, F.; Wang, X.; and Huang, H. 2014. Clustering and projected clustering with adaptive neighbors. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 977–986. ACM.

Samaria, F. S., and Harter, A. C. 1994. Parameterisation of a stochastic model for human face identification. In *Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on*, 138–142. IEEE.

Sim, T.; Baker, S.; and Bsat, M. 2002. The cmu pose, illumination, and expression (pie) database. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, 46–51. IEEE.

Von Luxburg, U. 2007. A tutorial on spectral clustering. *Statistics and computing* 17(4):395–416.

Wang, H.; Nie, F.; and Huang, H. 2015. Large-scale cross-language web page classification via dual knowledge transfer using fast nonnegative matrix tri-factorization. *ACM Transactions on Knowledge Discovery from Data (TKDD)*.

Wang, D.; Nie, F.; and Huang, H. 2016. Fast robust nonnegative matrix factorization for large-scale human action data clustering. In *IJCAI*.