Local Coordinate Projective Non-negative Matrix Factorization

Qing Liao, Qian Zhang
Dept. of Computer Science & Engineering
The Hong Kong University of Science & Technology
Hong Kong, China
Email: {qnature,qianzh}@cse.ust.hk

Xiang Zhang, Naiyang Guan
College of Computer
National University of Defense Technology
Chang Sha, China
Email: zhangxiang_43@aliyun.com, ny_guan@nudt.edu.cn

***Abstract*—Non-negative matrix factorization (NMF) decomposes a group of non-negative examples into both lower-rank factors including the basis and coefficients. It still suffers from the following deficiencies: 1) it does not always ensure the decomposed factors to be sparse theoretically, and 2) the learned basis often stays away from original examples, and thus lacks enough representative capacity. This paper proposes a local coordinate projective NMF (LCPNMF) to overcome the above deficiencies. Particularly, LCPNMF induces sparse coefficients by relaxing the original PNMF model meanwhile encouraging the basis to be close to original examples with the local coordinate constraint. Benefitting from both strategies, LCPNMF can significantly boost the representation ability of the PNMF. Then, we developed the multiplicative update rule to optimize LCPNMF and theoretically proved its convergence. Experimental results on three popular frontal face image datasets verify the effectiveness of LCPNMF comparing to the representative methods.**

***Keywords*-Local coordinate factorization, Non-negative matrix factorization**

I. Introduction

Subspace learning intrinsically reveals the structure of data and enhances the subsequent processing [1][2]. Clustering analysis is an unsupervised learning problem which assigns samples to different groups based on the similarities among data samples. Since it can automatically organize the samples into meaningful clusters, it is a key topic in many fields, such as text mining [3] and bioinformatics [4].

Since non-negative matrix factorization (NMF) can uncover the underlying data structure and then reduce data redundancy, it has become a popular way for clustering. NMF decomposes original data into both learned lower-rank matrices including the basis and coefficients. It aims at uncovering the underlying structure of dataset and facilitating subsequent data analysis. Recent works also claimed that NMF was equivalent to several well-known clustering methods such as a relaxed K-means [5] and spectral graph cuts [6]. In clustering tasks, the learned basis can be regarded as the cluster centroids, while the coefficients are viewed as the confidence scores assigned to specific categories. Thus, sparseness of the coefficients plays an important role in clustering tasks. However, the sparseness of NMF is always not good enough and it is difficult to control the sparsity of coefficients.

To induce sparseness constraint on coefficients, Hoyer *et al.* [7] proposed the sparse NMF method (SNMF) which minimizes the summation of all entries of both factors via the L_1-norm regularization. Yuan *et al.*[8] proposed the projective NMF method (PNMF) which introduces the orthogonal constraint over the learned basis to learn spatially localized, parts-based representations of visual patterns. However, it often produces incorrect cluster centroids. Recently, Cai *et al.*[9] proposed a locally consistent concept factorization method which simultaneously considers the geometrical structure and sparsity of the data. Chen *et al.* [10] proposed the non-negative local coordinate factorization (NLCF) to guarantee its sparseness via the local coordinate constraint. However, NLCF always induces the trivial basis. To overcome this shortcoming, Liu *et al.* [11] developed the local coordinate concept factorization (LCF) which incorporates the local coordinate constraint into concept factorization (CF, [12]) to learn the effective basis and sparse coefficients, but it still does not achieve satisfactory results.

In this paper, we propose a local coordinate projective NMF method (LCPNMF) which integrates the local coordinate constraint into PNMF to induce sparsity over coefficients learned by NMF. Particularly, we introduce the auxiliary basis to relax the original PNMF. This brings two advantages: 1) it incorporates the local coordinate constraint over the basis, and 2) PNMF explicitly learns the basis to tradeoff both the representative power of the basis and the sparseness of coefficients. Inspired by [10] and [11], LCPNMF incorporates the local coordinate constraint over the basis and the coefficients to further boost the representative capacity of the basis. Benefitting from both strategies, LCPNMF can significantly improve the representative ability of NMF. We developed a multiplicative update rule to optimize LCPNMF and theoretically proved its convergence. Experiments of clustering on three facial image datasets including UMIST [13], ORL [14], and FERET [15] datasets suggest the effectiveness of LCPNMF.

II. Related Works

A. Non-negative Matrix Factorization (NMF)

Non-negative matrix factorization was popularized by Lee and Seung [16] due to their simple and efficient multiplicative rule. Recently, NMF has also been successfully applied for clustering due to the effectiveness of its representation. Given a non-negative data matrix $X \in R_{+}^{m\times n}$, where m and n denote the number of data samples and features, respectively. NMF aims to find non-negative matrices $W \in R^{m\times k}$ and $H \in R^{k\times n}$, i.e.,

$$\min_{W,H\geq 0} \|X - WH\|_F^2, \quad (1)$$

where both W and H denote the basis and coefficients, respectively, and $k \ll \{m, n\}$ that stands for the number of clusters. Besides, $\| \bullet \|_F$ denotes the Frobenius norm. Traditional NMF cannot guarantee the decomposition results to be sparse in theory, and thus it still does not work well in clustering task.

B. Non-negative Local Coordinate Factorization (NLCF)

To enhance sparsity of the coefficients, Chen *et al.* [10] proposed the NLCF which incorporates the local coordinate constraint into NMF. The local coordinate constraint imposes the basis to be close to data points and thus induces the coefficients to be as sparse as possible. Thus, the model of NLCF can be written as follows:

$$\min_{W,H\geq 0} \|X - WH\|_F^2 + \mu \sum_{i=1}^{N}\sum_{k=1}^{K} |h_{ki}| \, \|\mathbf{w}_k - \mathbf{x}_i\|_2^2, \quad (2)$$

where $\mu \geq 0$ is a regularization parameter which tradeoff the sparseness of the coefficients via the so-called local coordinate constraint. The local coordinate constraint

978-1-5090-0287-0/15 $31.00 © 2015 IEEE
DOI 10.1109/ICMLA.2015.47

preserves the local structure by giving a heavy penalty if $\mathbf{x}_i$ is far away from the anchor point $\mathbf{w}_k$ while its coordinate h_{ki} with respect to $\mathbf{w}_k$ is large. By minimizing (2), there is only few coefficients h_{ki} are nonzero which can guarantee the sparsity of the matrix factors. However, it often obtains the trivial solution.

C. Local Coordinate Concept Factorization (LCF)

To address the above issue, Liu *et al.* [11] proposed the LCF method which also incorporates the local coordinate constraint into the concept factorization (CF, [12]). The objective of LCF has the following form:

$$\|X - XWH\|_F^2 + \lambda \sum_{i=1}^{N} \sum_{k=1}^{K} |h_{ki}| \| \sum_{j=1}^{N} w_{jk}\mathbf{x}_j - \mathbf{x}_i\|_2^2, \qquad (3)$$

where λ is a regularization parameter. The first term of (3) is the concept factorization while the second corresponds to the local coordinate constraint. Since LCF requires that the learned basis vector to be close to the original data points, each data point can be approximated by a linear combination of as few basis vectors as possible. But LCF still obtain unsatisfactory performance in clustering tasks due to the weak representation capacity of the implicit cluster centroids.

III. Local Coordinate Projective NMF

As addressed above, the local coordinate constraint not only guarantees the coefficients to be sparse but also induces correct cluster centroids. To enhance clustering performance of NMF, this section proposes a local coordinate projective NMF (LCPNMF) which incorporates the local coordinate constraint into projective non-negative matrix factorization.

A. The Proposed Model

The original projective non-negative matrix factorization (PNMF, [8]) can learn sparse coefficients for clustering tasks, because it can always guarantee the decomposed results to be sparse. Given a data matrix $V = [\mathbf{v}_1, \ldots, \mathbf{v}_n]^T \in R^{n\times m}$, where n denotes the number of samples and m denotes the feature dimension. PNMF learns the coefficients $H \in R^{n\times r}$ to represent original samples, i.e.,

$$\min_{H\geq 0} \left\| V - HH^TV \right\|_F^2, \qquad (4)$$

where r denotes the number of clusters. Although PNMF learns the sparse coefficients, its cluster centroids may be wrong suffering from the incorrect coefficients. This is because PNMF implicitly learns the basis. According to [17] , the objective function (4) is non-trivial to analyze the convergence in theory because (4) contains a fourth-order term. To overcome the above deficiency, we can rewrite (4) as:

$$\min_{H\geq 0} \|V - HW\|_F^2 \,, s.t., W = H^TV \qquad (5)$$

where W denotes the auxiliary basis. The objective (5) not only preserves the orthogonal constraint over H of PNMF but also shares the identical objective function with the basic non-negative matrix factorization. Thus, objective (5) can inherit the merits of both NMF and PNMF. Then we recast (5) as:

$$\min_{H,W\geq 0} \frac{1}{2}\|V - HW\|_F^2 + \frac{\alpha}{2}\left\| W - H^TV \right\|_F^2, \qquad (6)$$

where α denotes the positive regularization parameter which trade-off the orthogonality of the coefficients. To improve representative power of the basis, we can incorporate the local coordinate constraint [9] into (6) to induce sparse

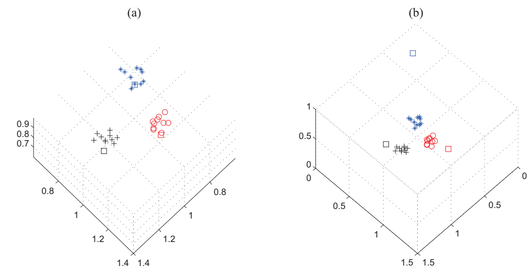

Figure 1: The basis learned by (a) LCPNMF and (b) NMF, respectively. Three colors with different shapes denote three categories while the squares denote the basis, i.e., the cluster centroids.

coefficients to be close to true cluster indicators. Thus, we obtain the following objective of LCPNMF:

$$\begin{aligned} &\min_{H,W\geq 0} \tfrac{1}{2}\|V - HW\|_F^2 + \tfrac{\alpha}{2}\|W - H^TV\|_F^2 \\ &+\tfrac{\beta}{2}\sum_{i=1}^{n}\sum_{j=1}^{r}|H^{ij}|\|V^i - W^j\|_2^2, \end{aligned} \qquad (7)$$

where β trades off the local coordinate regularization and denotes the i-th row and j-th column element of coefficients H, W^j and V^i, signifying the i-th and j-th row vector of W and V, respectively. To clearly illustrate the effectiveness of LCPNMF, we randomly generate three categorical 3D samples from three uniform distributions with the mean $[1.1, 0.9, 0.9]$, $[0.75, 0.75, 0.75]$ and $[0.95, 1.15, 0.85]$, respectively. Figure 1 (a) and (b) shows that three squares denote the cluster centroids learned by LCPNMF and NMF, respectively. It also implies that LCPNMF can enhance the representative power of the cluster centroids compared with NMF.

B. Optimization Algorithm

The model (7) is not jointly convex over W and H, and thus it is impossible to obtain the global solution. Fortunately, it is convex with respect to W with H fixed, and vice versa. Thus we developed a multiplicative update rule (MUR) to optimize LCPNMF by alternatively updating them. We optimize (7) with respect to one variable with the other fixed as follows:

Computation of W: Given H , we can yield the derivative of (7) with respect to W,

$$\Delta_W = H^THW + \alpha W + \beta FW - (1 + \alpha + \beta)H^TV, \qquad (8)$$

By setting the derivative of W to zero, we can obtain:

$$W = W \otimes \frac{(1 + \alpha + \beta)H^TV}{H^THW + \alpha W + \beta FW}, \qquad (9)$$

where $\otimes$ denotes the element-wise product operator, F is diagonal matrix whose entries are row sums of H.

Computation of H: Given W , we can yield the derivative of (7) with respect to H,

$$\Delta_H = HWW^T + \alpha VV^TH + \frac{\beta}{2}(A + B) - (1 + \alpha + \beta)VW^T, \qquad (10)$$

By setting the derivative of H to zero, we obtain:

$$H = H \otimes \frac{(1 + \alpha + \beta)VW^T}{HWW^T + \alpha VV^TH + \beta/2(A + B)}, \qquad (11)$$

where $A = [a, ..., a]$ wherein $a = diag(VV^T)$, and $B = [b, ..., b]$, wherein $b = diag(WW^T)$. For completeness, we summarize the optimization procedure of LCPNMF into **Algorithm** 1.

Algorithm 1 MUR for LCPNMF

Input: Examples $V \in R^{n \times m}$, the number of cluster r, penalty parameter α and parameter β
Output: W and H

1: Initialize W, H with random values.
2: **repeat**
3: Update W via (9).
4: Update H via (11).
5: **until** {Convergence.}

To reduce the time overhead, **Algorithm** 1 utilizes the objective relative error as the stopping criterion; in addition, set the maximum objective relative error to 10^{-7} in our experiments. The time cost of **Algorithm** 1 lies in the line 3 and line 4, respectively. The line 3 takes $O(mnr + rm + r^2n + r^2m + nr)$ and the line 4 takes $O(mnr + rn + rm + n^2r + r^2n + r^2m)$. The total time complexity of **Algorithm** 1 is $O(mnr+rm+r^2n+r^2m+nr+n^2r)$.

C. Theoretical Analysis

This subsection employs the auxiliary function technique to prove the convergence of LCPNMF. We can establish the following theorem:

Theorem 1. *The objective function* (7) *is non-increasing under the following multiplicative update rules in* (9) *and* (11).

Proof: According to (7), we can obtain the objective with respect to W as follows:

$$\begin{aligned} &J(W) = \tfrac{1}{2}Tr(-2VW^TH^T + HWW^TH^T) \\ &+\tfrac{\alpha}{2}Tr(WW^T - 2WV^TH) \\ &+\tfrac{\beta}{2}Tr(\sum_{i=1}^{n}(V^i)^T 1^T \Lambda_i 1 V^i - 2W^TH^TV + W^TFW), \end{aligned} \tag{12}$$

where Λ^i denotes the diagonal matrix whose diagonal elements are the i-th row vector values of H.

By (12), we define the auxiliary function of $J(W)$ as

$$\begin{aligned} &G(W,W') = -(1+\alpha)Tr(WV^TH) + \tfrac{1}{2}\sum_{ij}\frac{(H^THW')_{ij}}{W'_{ij}}W_{ij}^2 \\ &+\tfrac{\alpha}{2}Tr(WW^T) - \beta Tr(W^TH^TV) + \tfrac{\beta}{2}\sum_{ij}\frac{(FW')_{ij}}{W'_{ij}}W_{ij}^2. \end{aligned} \tag{13}$$

Obviously, objective (13)

$$G(W,W') \geq J(W) = G(W,W). \tag{14}$$

We can obtain the derivative of (14) as follows:

$$\begin{aligned} &\frac{\partial G(W,W')}{\partial W_{ij}} = -((1+\alpha)H^TV)_{ij} + \frac{(H^THW')_{ij}}{W'_{ij}}W_{ij} \\ &+\alpha W_{ij} - \beta(H^TV)_{ij} + \beta\frac{(FW')_{ij}}{W'_{ij}}W_{ij}. \end{aligned} \tag{15}$$

Based on (15), we have

$$W_{ij} = W'_{ij}\frac{((1+\alpha+\beta)H^TV)_{ij}}{(H^THW' + \alpha W' + \beta FW')_{ij}}. \tag{16}$$

By simple algebra, the formula (9) can be deduced from (16). Likewise, we also obtain the update rule (11) for H.

Moreover, according to (14),(16) and (11), we have

$$J(W^{t+1}, H^{t+1}) \leq J(W^{t+1}, H^t) \leq J(W^t, H^t). \tag{17}$$

Based on (17), these update rules always guarantee the objective function monotonically decreases. Thus, this completes the proof. ■

Figure 2: The image instances of (a) UMIST, (b) ORL and (c) FERET datasets, respectively.

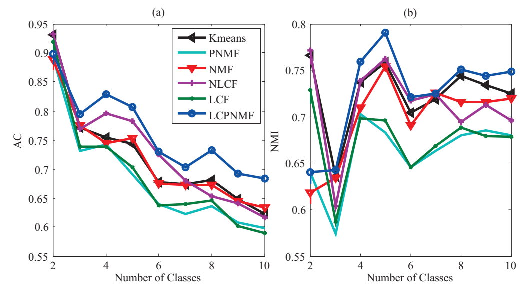

Figure 3: Average accuracy (AC) and normalized mutual information (NMI) versus different numbers of classes on UMIST dataset.

IV. Experiments

We verify the effectiveness by comparing the clustering performance of LCPNMF to the representative methods including NMF [16], PNMF [8], LCF [11], CF [12] and K-means on the UMIST [13], ORL [14] and FERET [15] datasets. We randomly selected $r = \{2, \ldots, 10\}$ individuals from these datasets and utilize LCPNMF and the compared methods to cluster images. Figure 2 shows some image instances. All selected images are normalized into the range from 0 to 1. LCPNMF learns the sparse coefficients and thus selects the index of the maximum coefficient values of each sample as clustering identity. Then, each experiment is conducted ten times and we reported the average results to evaluate the compared methods based on two popular clustering evaluation metrics including average accuracy (AC) and the normalized mutual information (NMI), respectively.

A. UMIST Dataset

The UMIST dataset [13] contains 300 images of 20 people with different races, genders, and appearances, covering a range of poses from profile to frontal views. All images are cropped to 40×40-pixel grayscale images and reshaped into a 1600-dimensional vector. For K-means, NMF, PNMF, they involve no parameter settings. We select the parameters $\alpha = 0.01$ and $\beta = 0.1$ for LCPNMF, $\lambda = 0.5$ for LCF, and $\mu = 0.1$ for NLCF on this dataset. Figure 3 reports the clustering accuracy and normalized mutual information (NMI) of the compared methods on UMIST dataset. This also implies that LCPNMF is superior to the representative methods in quantities.

B. ORL Dataset

The ORL database [14] contains 400 face images with 40 distinct subjects and each subject has 10 images taken at different times, lighting conditions, facial expressions, and facial details (glasses/no glasses) against a dark homogeneous background. All images are cropped to 32×32-pixel grayscale images and reshaped into a 1024-dimensional vector. For K-means, NMF, PNMF, they involve no parameter

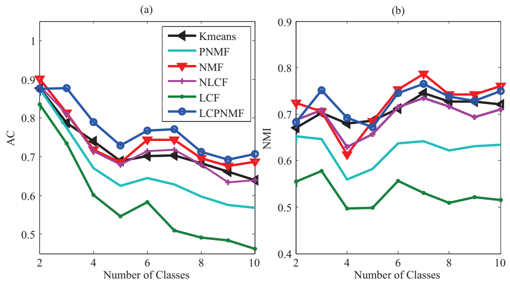

Figure 4: Average accuracy (AC) and normalized mutual information (NMI) versus different numbers of classes on ORL dataset.

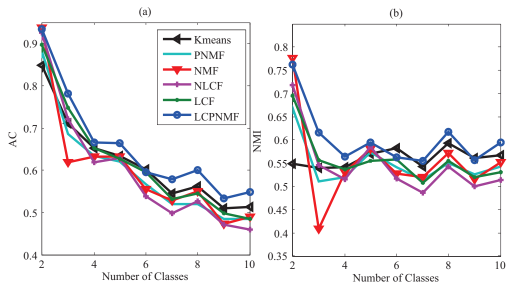

Figure 5: Average accuracy (AC) and normalized mutual information (NMI) versus different numbers of classes on FERET dataset.

settings. For NLCF, we select the best parameter $\mu = 0.1$ on this dataset. Likewise, we set $\lambda = 0.5$ for LCF. Note that all the results come from the average of ten trails. They distinguish the results reported in [11], where the best result is chosen for each trail. Thus, it is no strange that there exist different performance curves. Besides, we also set the parameters $\alpha = 0.01$, and $\beta = 0.1$ for LCPNMF. Figure 4 reports the average clustering accuracy and normalized mutual information (NMI) of the compared methods on ORL dataset. It also shows that LCPNMF is superior to the representative methods in quantities.

C. FERET Dataset

The FERET dataset[15] contains 700 face images collected from 100 individuals, each individuals contains 7 images. All images are cropped to 40×40-pixel grayscale images and reshaped into a 1600-dimensional vector. Similar to the above subsection, we set the parameter $\alpha = 0.01$ and $\beta = 0.1$ for LCPNMF, $\mu = 0.1$ for NLCF, and $\lambda = 0.5$ for LCF, respectively, on this dataset. The resting compared methods involve no parameter settings. Figure 5 shows the clustering accuracy and normalized mutual information of all the methods on FERET dataset. The results imply that LCPNMF outperforms the compared methods under different class numbers.

V. Conclusion

This paper proposes a local coordinate projective non-negative matrix factorization (LCPNMF) which incorporates the local coordinate constraint into PNMF to boost clustering performance. Particularly, LCPNMF can explicitly learn the basis, i.e., the cluster centroids, to avoid the overfitting problem meanwhile maintaining the orthogonal constraint over the coefficients. Moreover, it incorporates the local coordinate constraint over both the basis and coefficients to further enhance the representative power of the basis. Experimental results of image clustering on three popular face image datasets show the effectiveness of LCPNMF in quantities.

VI. Acknowledgement

This work was partially supported by the National Natural Science Foundation of China (under grant No. 61502515).

References

[1] D. Tao, L. X, W. X, and J. M. S, "General tensor discriminant analysis and gabor features for gait recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1700–1715, 2007.

[2] ——, "Geometric mean for subspace selection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 260–274, 2009.

[3] W. Xu, X. Liu, and Y. Gong, "Document clustering based on non-negative matrix factorization," in *SIGIR*. ACM, 2003, pp. 267–273.

[4] S. E. Palmer, "Hierarchical structure in perceptual representation," *Cognitive psychology*, vol. 9, no. 4, pp. 441–474, 1977.

[5] C. H. Ding, X. He, and H. D. Simon, "On the equivalence of nonnegative matrix factorization and spectral clustering." in *SDM*, vol. 5. SIAM, 2005, pp. 606–610.

[6] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[7] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *The Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, 2004.

[8] Z. Yuan and E. Oja, "Projective nonnegative matrix factorization for image compression and feature extraction," in *SCIA*. Springer, 2005, pp. 333–342.

[9] D. Cai, X. He, and J. Han, "Locally consistent concept factorization for document clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 6, pp. 902–913, 2011.

[10] Y. Chen, J. Zhang, D. Cai, W. Liu, and X. He, "Nonnegative local coordinate factorization for image representation," *IEEE Transactions on Image Processing*, vol. 22, no. 3, pp. 969–979, 2013.

[11] H. Liu, Z. Yang, and Z. Wu, "Locality-constrained concept factorization," in *IJCAI*, vol. 22, no. 1. Citeseer, 2011, p. 1378.

[12] W. Xu and Y. Gong, "Document clustering by concept factorization," in *SIGIR*. ACM, 2004, pp. 202–209.

[13] D. B. Graham and N. M. Allinson, "Characterising virtual eigensignature for general face recognition," in *Face Recognition*, pp. 446–456.

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

[15] P. J. Phillips, H. Moon, S. Rizvi, P. J. Rauss *et al.*, "The feret evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

[16] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[17] X. Zhang, N. Guan, L. Lan, D. Tao, and Z. Luo, "Box-constrained projective nonnegative matrix factorization via augmented lagrangian method," in *IJCNN*. IEEE, 2014, pp. 1900–1906.