IMPROVING NON-NEGATIVE MATRIX FACTORIZATION VIA RANKING ITS BASES

Sheng Huang[†,§], *Mohamed Elhoseiny*[§], *Ahmed Elgammal*[§], *Dan Yang*[†]

[†]College of Computer Science at Chongqing University, Chongqing, PRC
[§]Department of Computer Science at Rutgers University, NJ, USA
{huangsheng, dyang}@cqu.edu.cn[†], {elgammal, m.elhoseiny}@cs.rutgers.edu[§]

ABSTRACT

As a considerable technique in image processing and computer vision, Nonnegative Matrix Factorization (NMF) generates its bases by iteratively multiplicative update with two initial random nonnegative matrices W and H, that leads to the randomness of the bases selection. For this reason, the potentials of NMF algorithms are not completely exploited. To address this issue, we present a novel framework which uses the feature selection techniques to evaluate and rank the bases of the NMF algorithms to enhance the NMF algorithms. We adopted the well known Fisher criterion and Least Reconstruction Error criterion, which is proposed by us, as two instances to show how that works successfully under our framework. Moreover, in order to avoid the hard combinatorial optimization issue in ranking procedure, a de-correlation constraint can be optionally imposed to the NMF algorithms for giving a better approximation to the global optimum of the NMF projections. We evaluate our works in face recognition, object recognition and image reconstruction on ORL and ETH-80 databases and the results demonstrate the enhancement of the state-of-the-art NMF under our framework.

***Index Terms*—** Nonnegative Matrix Factorization, Fisher Score, Object Recognition, Face Recognition, Image Reconstruction

1. INTRODUCTION

Non-negative Matrix Factorization (NMF) [1, 2] is a classical linear multivariate analysis technique, and has been shown recently to be useful for many applications in computer vision, pattern recognition, multimedia and image processing. In contrast with other multivariate analysis techniques, its main advantage is that it can provide an intuitive visual interpretation of each basis, since the linear combination of the basis can be only additive. Various impressive NMF algorithms have been proposed for addressing different issues in recent years, e.g. [3, 4, 5, 6, 7, 8]. A comprehensive survey of NMF algorithms is recently presented in [9].

This work has been supported by the Fundamental Research Funds for the Central Universities (No. CDJXS11181162 and CDJZR12098801)

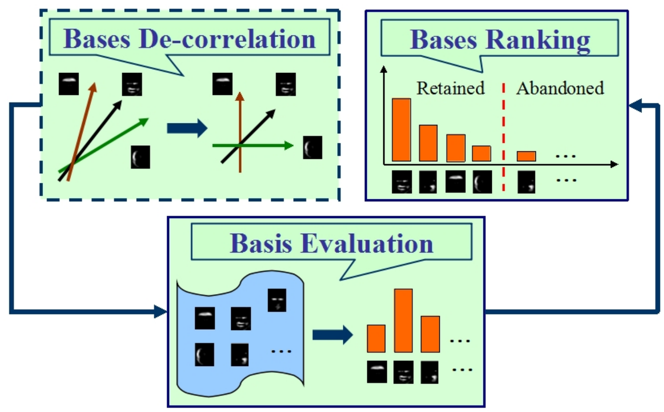


Fig. 1. The overview of the framework.

Although the NMF algorithms have obtained many remarkable successes, their potentials are not completely exploited, since the NMF algorithms are solved by iterative multiplicative update with two initial random nonnegative matrices W and H, in which the bases are randomly ranked. However, researchers always choose the first (or the last) n bases to yield the final projections in particular applications. Thus, the selected n NMF projections are ususally not corresponding to the n best bases.

Generally speaking, there are two ways for addressing this problem. The first solution is to use other optimization tools to replace the iterative multiplicative update for factorizing the non-negative matrix, e.g. [10, 11]. The second solution is to utilize some statistical analysis techniques to estimate an optimal initialization of factor matrices H and W [12]. However, both of these two approaches cannot guarantee the global optimum and are computationally complex.

In this paper, we present a framework to generally and systematically improve the performances of NMF algorithms by ranking their bases inspired by the existing feature selection techniques. To the best of our knowledge, there are no prior work that suggested this solution. So we actually open a new path towards the solution of this problem. The proposed method is general and applicable to all available NMF algorithms, and does no conflict with the previous two strategies, which means it can be readily combined with them.

The rest of paper is organized as follows: Section 2 presents the involved NMF algorithms; Section 3 describes our methodology. Experiments are presented in Section 4, and conclusion is summarized in Section 5.

2. INVOLVED NMF ALGORITHMS

This section summarize two NMF algorithms that are the targets for our improvement. Let a set of n training images be given as $l \times n$ matrix $X = [x_1, \cdots, x_n]$ where x_i is the ith column of matrix X and denotes the ith vectorized training image. A $l \times m$ matrix $W = [w_1, \cdots, w_m]$ denotes a set of $m \leq l$ basis vectors and its corresponding coefficients (loadings) are denoted as a $m \times n$ matrix $H = [h_1, \cdots, h_n]$ where $x_i \approx Wh_i$. Hence, the training image matrix can be approximately factorized as $X \approx WH$, which represents the image reconstruction process using the bases and loadings. Its reverse process can be done as $h_i = W^{-}x_i$.

2.1. Non-negative Matrix Factorization

Non-negative Matrix Factorization (NMF) [1, 2] imposes the non-negativity constraints, $W, H \geq 0$, to ensure that all entries of W and H are non-negative. Consequently, NMF only allows non-subtractive combinations. There are two cost functions that can be defined to find an approximate factorization $X \approx WH$. The first is based on the Euclidean distance and the second is based on divergence. In this paper, we only introduce the Euclidean distance based version and the divergence based version can be referenced from their original papers. So, the NMF problem can be finally formulated as a following optimization problem:

$$\hat{W} = \underset{W,H}{\operatorname{argmin}} ||X - WH||^2, \quad s.t \quad W, H \geq 0 \tag{1}$$

It can be solved by using multiplicative update rules. Furthermore, an optional constraint $\sum_i w_{ij} = 1$ is always imposed for stabilizing the computation.

2.2. Graph Regularized NMF

Graph Regularized Non-negative Matrix Factorization (GRNMF) [3] imposes an additional graph regularizer, which encodes the local manifold structures information, to the standard NMF. GRNMF constructs an affinity weight matrix Q to weight the Euclidean distance of each two represented samples, and such regularizer is denoted as follows

$$\begin{aligned} \mathcal{R} &= \frac{1}{2}\sum_{i,j=1}^{n} ||h_i - h_j||^2 Q_{ij} \qquad (2)\\ &= Tr(HDH^T) - Tr(HQH^T) = Tr(HLH^T), \end{aligned}$$

where Q_{ij}, which is the i,jth entry of matrix Q, denotes the weight of the distance between the ith and the jth samples, $Tr(\cdot)$ denotes the trace of a matrix and matrix D is a diagonal matrix whose entries are column (or row, since Q is symmetric) sums of W, $D_{ii} = \sum_j Q_{ij}$, $L = D - Q$, which is called graph Laplacian. By minimizing $\mathcal{R}$, the projections can ensure that if samples x_i and x_j are close then their projected samples h_i and h_j are close as well. Combining this regularizer with the original NMF objective function leads to the object function of GRNMF:

$$\hat{W} = \underset{W,H}{\operatorname{argmin}} ||X - WH||^2 + \lambda Tr(HLH^T), \; s.t. \; W, H \geq 0$$

where the regularization parameter $\lambda > 0$ controls the smoothness of the local manifold structures preservation.

3. METHODOLOGY

The idea of our framework is very simple. To a specific task, we evaluate each basis and generate a score which can indicate its related ability. After that, we rank the bases based on the scores. In such case, there should be two basic steps in our framework. One is the basis evaluation and the other is the bases ranking. However, in the most of time, the bases of the NMF algorithms exist the correlation. In other words, the combination of the top n bases may be not the real optimal NMF projections. Searching such global optimum is a typical combinatorial optimization problem. Actually, the global optimum indeed can be achieved by a exhaustive enumeration method. But, it is very time consuming. Instead of it, we approximate the global optimum via de-correlating the bases. An important reason why we can adopt this way is due to the fact that the local representation requires the parts (the bases) to be distinct from each other [4, 5, 13]. In other words, the NMF bases should be naturally independent with each other. Thus, a bases de-correlation procedure can be **optionally** applied before the basis evaluation and the bases ranking to optimize the framework.

3.1. Bases De-correlation

By introducing the Lagrangian multiplier, an additional uncorrelated constraint is imposed to the NMF algorithms for de-correlating the bases. This step is an optional step, since this step will modify the algorithm itself. For example, if we need to keep the original structures of the bases, such step can be ignored.

To achieve de-correlation, each two bases should meet the following conditions: $w_i^T w_j = 0$ when $i \neq j$ and $w_i^T w_i = p$ where p denotes a positive number. However, typically, the bases are normalized to 1, so p is set to 1. Thus, the uncorrelated constraint is equivalent to an orthogonality constraint. We can integrate all the bases together to get a final holistic uncorrelated constraint: $W^T W = I$ where matrix I is an identity matrix.

According to the above analysis, the objective functions of Uncorrelated Non-negative Matrix Factorization (UNMF) and Uncorrelated Graph Regularized Non-negative Matrix

Factorization (UGRNMF) are respectively written as

$$\begin{aligned} J_1 &= ||X - WH||^2 + \gamma ||I - W^T W||^2, & (3) \\ & \quad s.t. \quad W, H \geq 0 \\ J_2 &= J_1 + \lambda Tr(HLH^T), \quad s.t. \quad W, H \geq 0 & (4) \end{aligned}$$

where parameter $\gamma \geq 0$ controls the de-correlation of the bases. Let θ and ϕ be the Lagrange multipliers for constraints $W_{ij} \geq 0$ and $H_{ij} \geq 0$ respectively.

Following the solution procedure of NMF and GRNMF, the update rules of UNMF and UGRNMF can be obtained via using Karush-Kuhn-Tucker conditions. The multiplicative update rules of UNMF with respect to W and H are

$$w_{ik} \leftarrow w_{ik} \frac{(XH^T + 2\gamma W)_{ik}}{(WHH^T + 2\gamma WW^T W)_{ik}} \quad (5)$$

$$h_{jk} \leftarrow h_{jk} \frac{(W^T X)_{jk}}{(W^T WH)_{jk}} \quad (6)$$

and the multiplicative update rules of UGRNMF as follows

$$w_{ik} \leftarrow w_{ik} \frac{(XH^T + 2\gamma W)_{ik}}{(WHH^T + 2\gamma WW^T W)_{ik}} \quad (7)$$

$$h_{jk} \leftarrow h_{jk} \frac{(W^T X + \lambda HW)_{jk}}{(W^T WH + \lambda HD)_{jk}} \quad (8)$$

3.2. Basis Evaluation

Basis evaluation is the core of our method and determines the measure of how well a problem can be solved. In this step, the original training data is used as prior knowledge for evaluation. In this section, the classification issue and reconstruction issue is taken as two instances to specify how our framework works. Bases ranking can be seen as a selection step that aims to select *meaningful* bases, which can benefit the solution of the given task. This is very close to the feature selection task. So we can seek the solution of basis evaluation from the studies of feature selection. In this paper, we adopt the well known feature selection technique, Fisher Score (Fisher Criterion), to evaluate the discriminant ability of basis and propose the Least Reconstruction Error (LRE) Criterion to evaluate the reconstruction ability of basis.

3.2.1. Fisher Criterion

Fisher criterion [14, 15, 16, 17] measures the scattering of classes by calculating the ratio of the trace of the between-class scatter matrix to the trace of the within-class scatter matrix along the direction of basis. Since the projection is known and the samples projected by each basis are all scalars, the between-class scatter matrix and the within-class scatter matrix are exactly the variance of the means of different classes and the sum of the variances of the homogenous samples respectively. Consequently, the Fisher criterion evaluating function for basis w can be written as

$$\mathcal{F}(w) = \frac{\sum_{c \in C} n_c \cdot \sigma(w^T X_c)}{\sigma(w^T M)} \quad (9)$$

where $\sigma(x)$ is the variance of x, X_c denotes the matrix constructed by the samples belonging to the class c and n_c indicates its sample number. M is the matrix whose ith column is the mean of the samples belonging class i.

Finally, we can grade each basis with a score, which indicates its discrimination ability. For example, smaller score indicates stronger discrimination ability.

3.2.2. Least Reconstruction Error Criterion

For a reconstruction task, we measure the reconstruction ability of basis by simply computing its reconstruction error to the training data. And we name this evaluation criterion *Least Reconstruction Error Criterion (LRE)*. The basic procedure is to use a learned NMF basis w and its corresponding learned coefficient h to reconstruct the data and then measure the Euclidean distance between training data and the reconstructed data directly for getting the reconstruction error. The reconstruction error computing function is presented as follows:

$$\mathcal{L}(w) = ||X - wh||^2 \quad (10)$$

where h is a row of matrix H corresponding to w of matrix W. A smaller reconstruction error means the basis has a better reconstruction ability and carries more important information along its direction.

3.3. Re-ranking Bases By Evaluation Results

After basis evaluation, each basis gets a score and then we can re-rank the bases. We select the m top ranked bases to yield the final projections. The detail of basis ranking is described in Algorithm 1.

Algorithm 1 Re-ranking the Bases

Require:
The training data X;
The sample class labels L;
The original $l \times n$ NMF projections $W = [w_1, \cdots, w_n]$;
Ensure:
The output re-ranked $l \times n$ projections W_r;
1: Define a temporary array F to store the evaluation scores.
2: **for** each $i \in [1, n]$ **do**
3: Calculate the evaluation score f of the ith basis by Equation 9 or Equation 10 (different evaluations are suitable to different issues) with parameters of w_i, X and L;
4: Put the evaluation score f into the ith entry of array F;
5: **end for**
6: Ascendingly sort the evaluation score array F, $[F, index] = SORT(F)$ where $index$ indicates the new index of array after the sorting;
7: Re-rank the bases W basing on $index$, $W_r = W(index)$;
8: **return** W_r;

4. EXPERIMENTS

This section presents several results that shows the potential of our framework. ORL face database and ETH-80 Object database are employed for face recognition, object recognition and Image reconstruction. Nonnegative Matrix Factorization [2] and Graph Regularized Non-negative Matrix Factorization (GRNMF) [3], are selected as the target methods for improving.

4.1. Face Recognition using ORL Database

The ORL database contains 400 images from 40 subjects [18]. Each subject has ten images acquired at different times. We resize each face image to 32×32 pixels to the face recognition issue while keep the original size to the image reconstruction task. We apply the five-fold, three-fold and two-fold cross-validations to evaluate the recognition performances of the NMF algorithms. Fisher Criterion is adopted for basis evaluation. The initial dimension of NMF projections is fixed to the number of testing samples and we construct the Graph Laplacian of GRNMF in a **supervised** way that only puts the edge between two homogenous samples. The parameter γ, which is used to control the de-correlations of the bases, is empirically assigned as 1 and 0.1 to NMF and GRNMF respectively.

Methods	Schemes-Recognition Accuracy (ARA±STD)		
	Five-fold	Three-fold	Two-fold
NMF	88.25±4.11	87.50±5.83	81.25±1.06
Re-ranked	90.00±3.54	89.44±5.55	82.75±1.96
De-correlated	91.75±3.78	90.56±1.73	82.75±1.96
Combined	93.00±3.78	91.39±2.68	84.25±1.77
Improvement	**4.75**	**3.89**	**3.00**
GRNMF	81.25±4.15	80.28±2.10	76.00±0.71
Re-ranked	85.50±3.38	81.39±1.27	78.00±0.71
De-correlated	92.50±3.19	90.56±5.02	84.50±1.41
Combined	94.50±2.74	91.94±3.76	86.75±0.35
Improvement	**13.25**	**11.66**	**10.75**

Table 1. Recognition performance (%) on ORL database

4.2. Object Recognition using ETH-80 Database

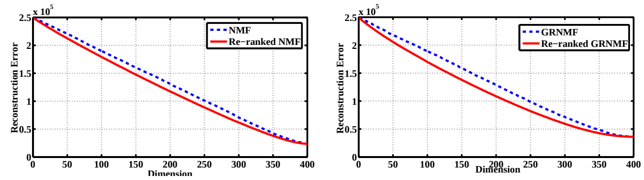
The ETH-80 object database contains 80 objects from 8 categories [19]. Each object is represented by 41 views spaced evenly over the upper viewing hemisphere. The original size of each image is 128×128 pixels. We resize them to 32×32 pixels. The experiments on ETH-80 Database follow the same experimental setting on ORL database. The ten-fold, Five-fold and two-fold cross-validation schemes are employed to evaluate the recognition performance.

Methods	Schemes-Classification Accuracy (ARA±STD)		
	Ten-fold	Five-fold	Two-fold
NMF	29.63±3.13	27.80±4.42	28.17±2.16
Ranked	31.49±3.13	28.29±5.00	29.21±1.12
De-correlated	52.44±5.64	53.35±4.84	62.50±1.64
Combined	72.38±6.39	71.95±5.78	73.81±3.92
Improvement	**42.77**	**44.15**	**45.64**
GRNMF	56.65±8.08	54.48±5.81	55.64±1.68
Ranked	57.29±9.01	54.15±4.00	56.28±1.47
De-correlated	65.49±8.98	63.87±5.65	61.77±2.59
Combined	66.19±8.69	64.36±6.22	62.31±2.37
Improvement	**9.54**	**9.88**	**6.67**

Table 2. Classification performance (%) on ETH-80 database

4.3. Image Reconstruction using ORL Database

The image reconstruction experiments are all based on the Least Reconstruction Error criterion. Figure 2 depicts the relation between the error and the retained dimension of the NMF algorithms, before and after ranking. It is clear that the ranked NMF algorithms obtain smaller reconstruction errors.



(a) (b)

Fig. 2. The reconstruction errors of NMF algorithms and its ranked version on ORL database, (a) the reconstruction errors of NMF and ranked NMF, (b) the reconstruction errors of GRNMF and ranked GRNMF.

4.4. Discussion

The following conclusions can be made from the experimental results listed in Tables 1 and 2:

1. The results show that each of the de-correlation and the ranking steps improves the results of each of the baseline algorithms. The results also show that combining the the two steps further improves the results as expected (combined improvement shown in bold). The improvement is consistent in all cases, and significant in most of the cases. For example, the combined improvement over the NMF algorithm is more than 40% ETH-80 database.
2. There exist a large different between the different improvements of different algorithms using different database. We think the large different in improvement is due to the fact that the performance of the baseline algorithms relies on the random initialization, which accidentally can get good initialization, and hence the room of improvement is limited.

5. CONCLUSION

We present a new framework for further exploiting the potential of the NMF algorithms. It utilizes the feature selection techniques to evaluate and rank the bases, which is generalizable for all NMF algorithm . In order to show how our framework works, the well known Fisher criterion and a proposed criterion named Lowest Reconstruction Error criterion are adopted to respectively enhance the discrimination and reconstruction abilities. Since the bases may exist correlation, the global optimal NMF projections searching is a hard combinatorial optimization issue. In order to avoid this, a bases de-correlation step is optionally add. We apply our framework to NMF and GRNMF for addressing the face recognition, object recognition and image reconstruction tasks. The results of experiments demonstrate its effectiveness.

6. REFERENCES

[1] Daniel D Lee and H Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, 1999.

[2] D Seung and L Lee, "Algorithms for non-negative matrix factorization," *Advances in Neural Information Processing Systems*, 2001.

[3] Deng Cai, Xiaofei He, Jiawei Han, and Thomas S Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2011.

[4] Stan Z Li, Xin Wen Hou, Hong Jiang Zhang, and Qian Sheng Cheng, "Learning spatially localized, parts-based representation," in *International Conference on Computer Vision and Pattern Recognition*, 2001.

[5] Zhao Li, Xindong Wu, and Hong Peng, "Nonnegative matrix factorization on orthogonal subspace," *Pattern Recognition Letters*, 2010.

[6] Ji-Yuan Pan and Jiang-She Zhang, "Large margin based nonnegative matrix factorization and partial least squares regression for face recognition," *Pattern Recognition Letters*, 2011.

[7] Yuan Wang and Yunde Jia, "Fisher non-negative matrix factorization for learning local features," in *Asian Conference on Computer Vision*, 2004.

[8] Taiping Zhang, Bin Fang, Yuan Yan Tang, Guanghui He, and Jing Wen, "Topology preserving non-negative matrix factorization for face recognition," *IEEE transaction on Image Processing*, 2008.

[9] Yu-Xiong Wang and Yu-Jin Zhang, "Nonnegative matrix factorization: A comprehensive review," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 6, pp. 1336–1353, 2013.

[10] Pinghua Gong and Changshui Zhang, "Efficient nonnegative matrix factorization via projected newton method," *Pattern Recognition*, 2012.

[11] Chih-Jen Lin, "Projected gradient methods for nonnegative matrix factorization," *Neural computation*, 2007.

[12] Christos Boutsidis and Efstratios Gallopoulos, "Svd based initialization: A head start for nonnegative matrix factorization," *Pattern Recognition*, 2008.

[13] Seungjin Choi, "Algorithms for orthogonal nonnegative matrix factorization," in *International Joint Conference on Neural Networks*. IEEE, 2008.

[14] Peter N. Belhumeur, P. Hespanha, and David J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 1997.

[15] Quanquan Gu, Zhenhui Li, and Jiawei Han, "Generalized fisher score for feature selection," in *The Conference on Uncertainty in Artificial Intelligence*, 2011.

[16] Zhu Manli and Aleix M Martínez, "Pruning noisy bases in discriminant analysis," *IEEE Transaction on Neural Networks*, vol. 19, no. 1, pp. 148–157, 2008.

[17] Koji Tsuda, Motoaki Kawanabe, and Klaus-Robert Müller, "Clustering with the fisher score," in *Advances in Neural Information Processing Systems*, 2002.

[18] F. S. Samaria, F. S. Samaria, A.C. Harter, and Old Addenbrooke, "Parameterisation of a stochastic model for human face identification," 1994.

[19] Bastian Leibe and Bernt Schiele, "Analyzing appearance and contour based methods for object categorization," in *International Conference on Computer Vision and Pattern Recognition*, 2003.