

2013 IEEE INTERNATIONAL WORKSHOP ON MACHINE LEARNING FOR SIGNAL PROCESSING, SEPT. 22–25, 2013, SOUTHAMPTON, UK

AN ENHANCED INITIALIZATION METHOD FOR NON-NEGATIVE MATRIX FACTORIZATION

Liyun Gong¹, Asoke K. Nandi^{2,3}

¹Department of Electrical Engineering and Electronics, Liverpool University, Liverpool, Merseyside, L69 3BX, UK;

²Department of Electronic and Computer Engineering, Brunel University, Uxbridge, Middlesex, UB8 3PH, UK;

³Department of Mathematical Information Technology, University of Jyväskylä, Jyväskylä, Finland; l.gong@liv.ac.uk; asoke.nandi@brunel.ac.uk;

ABSTRACT

Non-negative matrix factorization (NMF) is a dimensionality reduction tool, and has been applied to many areas such as bioinformatics, face image classification, etc. However, it often converges to some local optima because of its random initial NMF factors (W and H matrices). To solve this problem, some researchers have paid much attention to the NMF initialization problem. In this paper, we first apply the k-means clustering to initialize the factor W , and then we calculate the initial factor H using four different initialization methods (three standard and one new). The experiments were carried out on the eight real datasets and the results showed that the proposed method (EIn-NMF) achieved less error and faster convergence compared with both random initialization based NMF and the three standard methods for k-means based NMF.

Index Terms— k-means clustering; nonnegative matrix factorization.

1. INTRODUCTION

Non-negative matrix factorization (NMF), proposed by Lee and Seung [1, 2], is a new dimensionality reduction tool, and has been applied to many areas such as bioinformatics, face image classification, and so on. For example, in bioinformatics, Pascual-Montano et al. proposed a versatile tool called bioNMF based on NMF to cluster and bicluster gene expression data [3]. In [4], NMF was used for recognizing protein sequence patterns. In [5, 6], clustering results of gene expression data obtained by NMF were compared with hierarchical clustering and self-organizing

Asoke K. Nandi would like to thank TEKES for their award of the Finland Distinguished Professorship.

maps. Researchers have also proposed several different algorithms based on the traditional NMF to make improvements such as Least squares-NMF [7], Weighted-NMF [8], Local-NMF [9], and so on. However, since NMF is a nonconvex algorithm, it often converges to some local optima because of its random initial NMF factors (W and H matrices). In order to solve this problem, some researchers have paid much attention to the NMF initialization problem. Wild proposed the initialization method based on spherical k-means clustering to initialize W matrix and apply Nonnegative Least Square (NNLS) to calculate H matrix [10]. Zheng et al. proposed PCA-based initialization method and, after obtaining W and H , all the negative elements in these two matrices are changed to zero [11]. For fuzzy c-means clustering (FCM) initialization in [11], the initial matrix W is obtained by taking the final cluster centroids after FCM algorithm and the initial matrix H is determined by the membership degrees of FCM in which the largest membership degree of each data point is set to one and the others to zero. This method was followed by Rezaei et al. who used the membership degrees of FCM to determine the initial H matrix directly [12]. Zhao et al. used the absolute value for all elements in the initial matrices W and H after PCA initialization [13] in contrast to the method in [11]. With these initialization methods, enhanced convergence rates as well as better accuracy were achieved. In this paper, we propose a method to calculate the nonnegative initial matrix H efficiently in order to enhance the performance of NMF algorithm based on k-means clustering initialization. We first apply k-means clustering to initialize the W matrix [10] and use the proposed method to initialize the H matrix (EIn-NMF). These initial matrices are then passed through NMF algorithm and the result will be compared with both random initialization based NMF and the three standard methods for

k-means based NMF. The experiments were carried out on the eight real datasets from [14] and the results showed that the proposed method (EIn-NMF) achieved faster convergence while maintaining the main data structure.

The rest of the paper is organized as follows. In Section 2 we first review the basic knowledge of non-negative matrix factorization method (NMF) and its k-means clustering initialization, then review the random initialization based NMF and three standard methods for k-means based NMF, and describe the proposed initialization method (EIn-NMF) used in this paper. The experimental results based on these five initialization methods are evaluated and analysed in Section 3. Finally, conclusion is drawn in Section 4.

Contribution of this paper: we propose a new initialization method (EIn-NMF) for k-means based NMF to improve its performance. We also compare four standard initialization methods which have not been compared and contrasted in the previous literature.

2. METHODOLOGY

2.1. NMF

We briefly review the basic idea of NMF. Given a non-negative matrix X with m rows and n columns, the NMF algorithm seeks to find non-negative factors W and H such that

$$X \approx WH \quad (1)$$

where W is an $m \times k$ matrix and H is a $k \times n$ matrix. Each column of W is considered as the basis vectors while each column of H contains the encoding coefficient. All the elements in W and H represent non-negative values.

Many algorithms have been proposed to obtain W and H [11]. In this paper, we use the multiplication update rule to minimize an objective function (the root-mean-squared residual D). The formulae are given as follows.

$$D = \sqrt{\frac{\sum_{j=1}^n \sum_{i=1}^m (X_{ij} - (WH)_{ij})^2}{m \times n}}$$

$$H \leftarrow H \frac{W^T X}{W^T W H} \quad (2)$$

$$W \leftarrow W \frac{X H^T}{W H H^T}$$

NMF is a nonconvex programming in the iteration process, thus it may lead to different solutions with the different initial values of W and H . In this paper, we focus on using k-means clustering to initialize the factor W , and then we calculate the initial factor H using four different initialization methods (three standard and one new). The

results are then compared with the random initialization method. The details of these methods are described below.

2.2. K-means based NMF

In this section, we briefly summarize the process of the k-means based NMF algorithm and the details are described in Table 1.

Table1: The description of the k-means based NMF algorithm.

STEP 1: Given the $(m \times n)$ nonnegative matrix X and its pseudo-inverse A , apply k-means clustering on the matrix X to generate the cluster centroids c_i ($i=1, 2, \dots, k$), where k is the reduced dimensionality which can be set to any value.

STEP 2: The basis matrix of NMF W is then constructed by inverting the cluster centroids c_i (i.e. c_i^T).

STEP 3: Obtain the initial matrix H which is calculated by the different methods described below (three standard and one new).

STEP 4: Apply NMF algorithm to A with the initial matrices W and H obtained above and obtain the final results.

2.3. The random initialization and three standard methods for k-means based NMF

Here we describe the random initialization based NMF (H1) and the three standard initialization methods for k-means based NMF (H2, H3 and H4).

- H1: This is the common method for initializing both the two NMF factors W and H randomly.
- H2: The initial basis matrix W is constructed by using the cluster centroids obtained from k-means clustering and the initial matrix H is selected randomly.
- H3: The initial basis matrix W is constructed by using the cluster centroids obtained from k-means clustering. The initial matrix H is denoted by $H = W^T X$ and then we get the absolute value for all elements in H in order to satisfy the initial constraint of NMF.
- H4: The initial basis matrix W is constructed by using the cluster centroids obtained from k-means clustering. The initial matrix H is denoted by $H = W^T X$ and then we set all the negative elements in H to zero in order to satisfy the initial constraint of NMF.

2.4. The proposed methods (EIn-NMF) for k-means based NMF

Here we propose to use the k-means clustering initialization of W incorporated with our proposed initialization of H (H5) to improve the performance of k-means based NMF. The details of this method are shown below.

- Proposed H5: The initial basis matrix W is constructed by using the cluster centroids obtained from k-means clustering.

We calculate the membership degrees of each data point by using some results from k-means clustering. The details are as follows.

$$h_{kq} = 1 / \left[\sum_{k'=1}^k \left(\frac{d(x_q, c_{k'})}{d(x_q, c_k)} \right)^{\frac{2}{1-m}} \right] \quad (3)$$

where $d(\cdot)$ represents the Euclidean distance between the two points, x_q represents the q th data point and c_k represents the k th cluster centroid. Also, m is the fuzzification parameter which is set to 2 in this paper. The initial matrix H is then obtained by using the membership degrees above. This proposed enhanced initialization method for NMF is termed as EIn-NMF.

3. EXPERIMENT RESULTS AND ANALYSIS

3.1. Datasets and preprocessing

Eight datasets used in this paper are from [14]. Based on the similar preprocessing method as in [14], we removed the genes (columns) which vary little across samples (rows) to reduce the computational complexity. Table 2 shows some parameters for preprocessing. ‘N’ represents no operation performed and ‘inf’ represents the infinite value. ‘Ratio’, ‘Diff’ and ‘Std’ are the ratio, difference and standard values between maximum and minimum values of a gene (column) across samples (rows). We first set all the expression levels within the defined range. For each dataset, elements below the minimum value of the range were assigned this minimum value, and those exceeding the maximum value of the range were assigned this maximum value. Then some genes varying little across samples were removed according to the three parameters (Ratio, Difference and Standard), that is, we remove the columns which have the smaller values on Ratio, Difference and Standard. After preprocessing, the properties of these dataset are presented in Table 3. For simplicity, we use the dataset no. shown in Table 3 instead of its gene name for the following discussion.

Table 2: The parameters for the preprocessing

Name	Range	Ratio	Diff	Std.
9_Tumors	[100,16000]	N	N	N
Brain_Tumors1	[20,16000]	3	500	N
Brain_Tumors2	[20,16000]	3	500	N
Leukemia1	[20,16000]	3	500	N
Lung_Cancer	[0,inf]	N	N	50
SRBCT	N	N	N	N
Prostate_Tumor	[20,16000]	5	N	N
DLBCL	[20,16000]	3	500	N

Table 3: The properties of the final datasets

Dataset no.	Name	Rows	Columns	cluster
1	9_Tumors	60	5726	9
2	Brain_Tumors1	90	4922	5
3	Brain_Tumors2	50	1013	4
4	Leukemia1	72	3621	3
5	Lung_Cancer	203	3312	5
6	SRBCT	83	2308	4
7	Prostate_Tumor	102	3722	2
8	DLBCL	77	4204	2

3.2. Results

The experiments were carried out by using above eight datasets, we applied five different initialization methods (H1-H5) to improve the performance of NMF. In order to avoid the influence of randomness, each initialization method was run 20 times and the total number of iterations for each run of NMF was set to 500 in this paper. The rank (dimensionality) k for each dataset is set to the number of cluster of the corresponding dataset which is shown in Table 3.

Figure 1-8 shows the average log2 value of root-mean-squared residual D (RMS Residual D) from these 20 runs of five different initialization methods with the increasing iteration number. As mentioned in section 2.1, D is the error and the lower D log value suggests the better performance for NMF. It is seen from the eight figures that all the D log values decrease fast in the early iterations of NMF and become stable at the end. The proposed initialization method (H5) always gets the lower D log values compared with the other four methods on the eight datasets in the short-term. It shows that the D log values in the early iterations obtained by the five initialization methods always satisfy the inequality $D(H1) > D(H2) > D(H3) = D(H4) > D(H5)$. The method H1 which initializes both the two NMF factors W and H randomly has the worst performance of NMF as we expect. The reason is that the random initialization has nothing to do with the initial factors W

and H while the other four initialization methods already works for NMF algorithm with predefined factor W (H2) or both the two factors (H3, H4 and H5). So the D log value for random initialization is the largest. The methods H3, H4 and H5 adds the initiation process of the factor H , so their D log values are smaller than H2. The methods H2, H3 and H4 start at the similar D log value, however, H3 and H4 need far fewer iterations to converge than H2 on these eight datasets. From figure 1-8, we summarize that the proposed method H5 achieves less error and faster convergence at the beginning of the NMF iterations compared with the other four standard initialization methods.

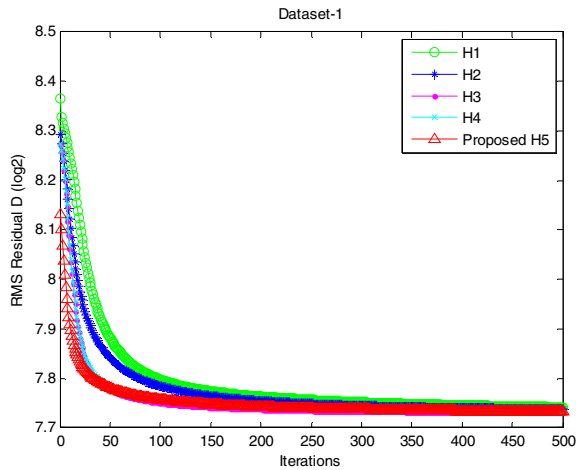


Figure 1: The average D log values of each of the five initialization methods (Dataset 1).

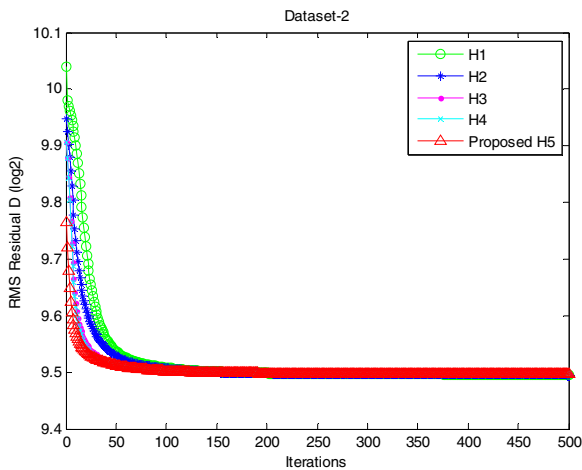


Figure 2: The average D log values of each of the five initialization methods (Dataset 2).

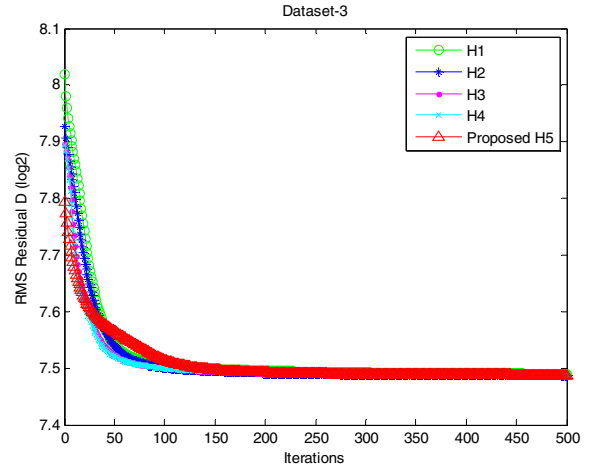


Figure 3: The average D log values of each of the five initialization methods (Dataset 3).

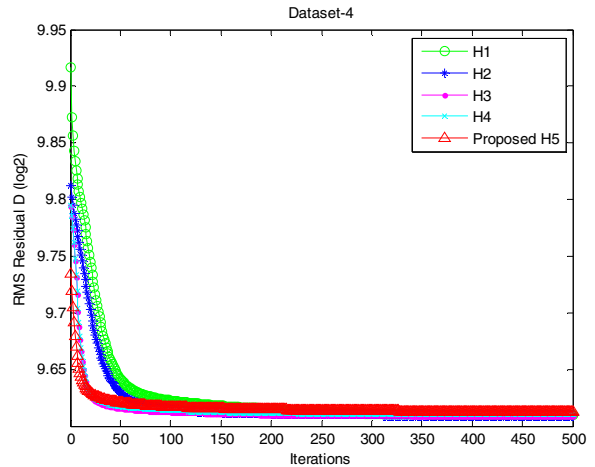


Figure 4: The average D log values of each of the five initialization methods (Dataset 4).

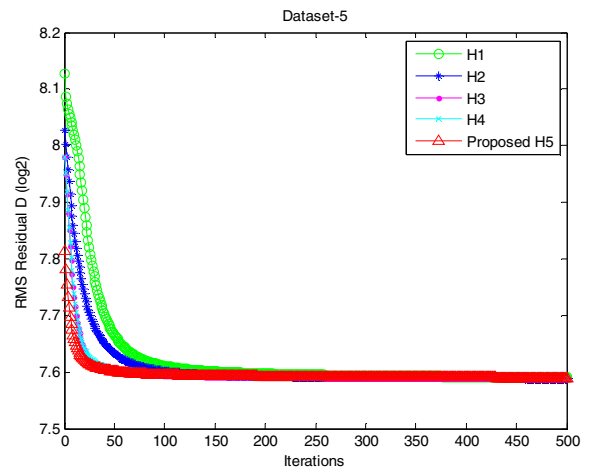


Figure 5: The average D log values of each of the five initialization methods (Dataset 5).

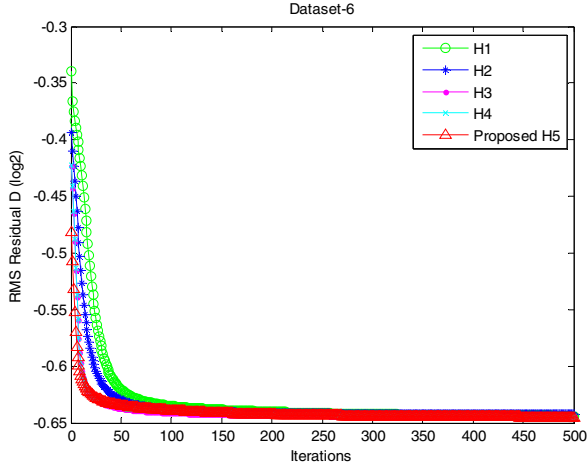


Figure 6: The average D log values of each of the five initialization methods (Dataset 6).

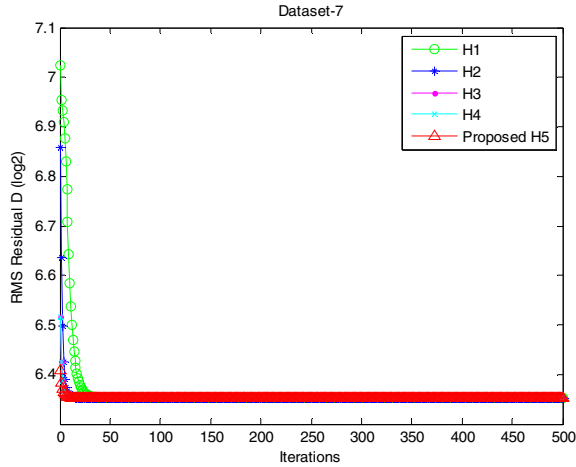


Figure 7: The average D log values of each of the five initialization methods (Dataset 7).

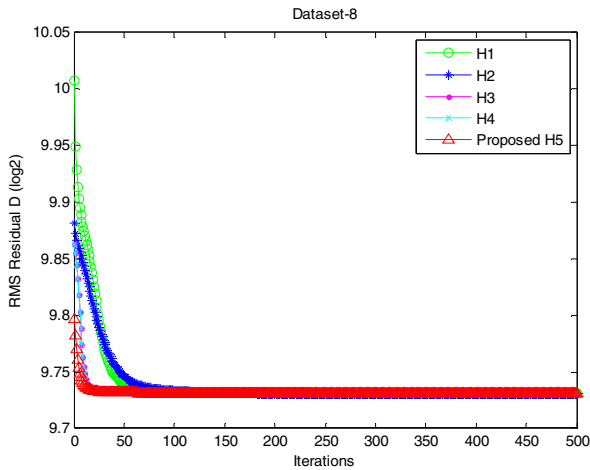


Figure 8: The average D log values of each of the five initialization methods (Dataset 8).

We have seen from the previous result that in the short-term, our proposed method (EIn-NMF) results in the better performance of NMF than the other four standard initialization methods. This property provides an idea to seek to benefit from the short-term behavior for our proposed method (EIn-NMF). We have used the eight real datasets which already have the predefined rank k (the cluster number). However, the cluster number is unknown for most of the datasets in reality. For this reason, we analyze the RMS Residual D under the different dimensionality for each dataset. Here we calculate the NMF algorithm by using one iteration, because we expect to investigate the properties of the five methods at the start point of the NMF algorithm with the different dimensionality. The values of dimensionality for each dataset are set to $\{5, 10, 15, \dots, [q], q\}$ where q is the row number and $[q]$ means the maximum integer which can be divided by 5 and smaller than q . In order to avoid the influence of the randomness, each dimensionality value under one initialization method was run 20 times and the number of iterations for each run of NMF was set to 1. The two figures below (figure 9 and 10) show the average RMS residual D values from these 20 runs of each of the five different initialization methods with the increasing number of dimensionality.

In these two figures, it can be easily seen that the four standard initialization methods (H1-H4) are almost stay at the higher D values while the proposed method H5 keeps decreasing with the increasing dimensionality. For Dataset 1-3, the last values of dimensionality are 60, 90 and 50 separately which are the same as their row numbers. This means there is no change on the datasets after NMF algorithm. Only H1 and the proposed H5 can recognize this at the early iteration of NMF which have the zero D values at these dimensionality values. For Dataset 4-8 which their row numbers, the proposed H5 can still achieve the relative low D values while the four standard methods cannot. So we conclude the proposed H5 always outperforms the other four standard initialization methods.

4. CONCLUSIONS

NMF algorithm often converges to some local optima because of its random initial NMF factors (W and H matrices). To solve this problem, some researchers have paid much attention to the NMF initialization problem. In this paper, we have proposed EIn-NMF, an initialization method for the factor H for k-means based on NMF. Altogether, we also have compared our method with the other four different standard initialization methods—random initialization based NMF and three standard initialization methods for k-means based NMF. The experiments were carried out on the eight real datasets from [14] and we assessed the NMF performance by the root-mean-squared

residual D. the results demonstrate that the proposed initialization method, EIn-NMF, gets better performance of NMF compared with the other four standard initialization methods.

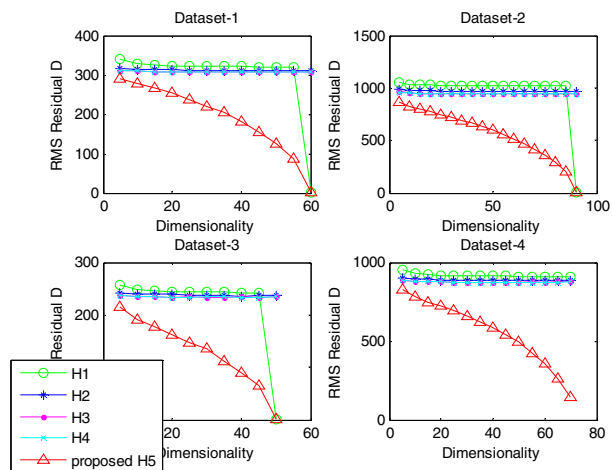


Figure 9: The average D values of each of the five initialization methods with the increasing dimensionality number on Dataset 1 - 4.

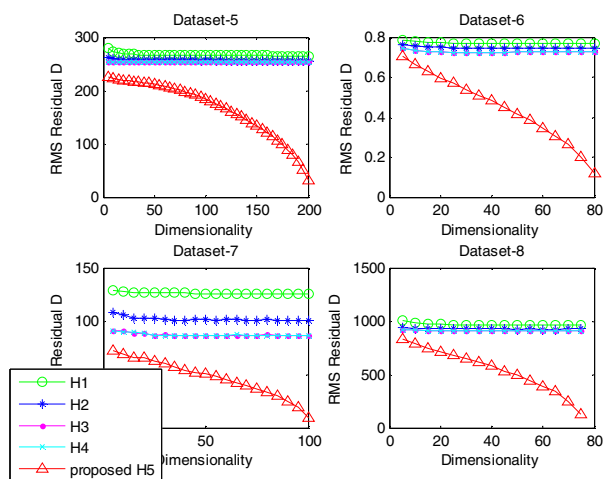


Figure 10: The average D values of each of the five initialization methods with the increasing dimensionality number on Dataset 5 - 8.

5. REFERENCES

[1] D D Lee and H S Seung, "Learning the parts of objects by nonnegative matrix factorization", *Nature*, vol. 401, pp. 788-791, 1999.

[2] D D Lee and H S Seung, "Algorithm for non-negative matrix factorization", *Advances in Neural Information Processing Systems*, vol. 13, pp. 556-562, 2001.

[3] A Pascual-Montano, P Carmona-Saez, M Chagoyen, F Tirado, J M Carazo, and R D Pascual-Marqui, "bioNMF: a versatile tool for nonnegative matrix factorization in biology", *BMC Bioinformatics*, vol. 7(1), pp. 366-374, 2006.

[4] A Heger and L Holm, "Sensitive pattern discovery with fuzzy alignments of distantly related proteins", *Bioinformatics*, vol. 19(90001), pp. 130-137, 2003.

[5] J P Brunet, P Tamayo, T R Golub, and J P Mesirov, "Metagenes and molecular pattern discovery using matrix factorization", *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, vol. 101(12), pp. 4164-4173, 2004.

[6] Y Gao and G Church, "Improving molecular cancer class discovery through sparse non-negative matrix factorization", *Bioinformatics*, vol. 21, pp. 3970-3975, 2005.

[7] G Wang, A V Kossenkov, and M F Ochs, "LS-NMF: a modified nonnegative matrix factorization algorithm utilizing uncertainty estimates", *BMC Bioinformatics*, vol. 7, pp. 175-271, 2006.

[8] D Guillamet, J Vitria, and B Scheile, "Introducing a weighted nonnegative matrix factorization for image classification", *Pattern Recognition Letters*, vol. 24, pp. 2447-2454, 2003.

[9] S Z Li, X Hou, H Zhang, and Q Cheng, "Learning spatially localized parts-based representation", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 207-212, 2001.

[10] S Wild, "Seeding non-negative matrix factorizations with the spherical k-means clustering", *Master of Science Thesis, University of Colorado*, 2003.

[11] Z Zheng, J Yang, and Y Zhu, "Initialization enhancer for non-negative matrix factorization", *Engineering Applications of Artificial Intelligence*, vol. 20, pp. 101-110, 2007.

[12] M Rezaei, R Boostani, and M Rezaei, "An efficient initialization method for nonnegative matrix factorization", *Journal of Applied Sciences*, vol. 11, pp. 354-359, 2011.

[13] L Zhao, G Zhuang, and X Xu, "Facial expression recognition based on PCA and NMF", *Proceedings of the 7th World Congress on Intelligent Control and Automation*, pp. 6826-6829, 2008.

[14] W Liu, K Yuan, and D Ye, "Reducing microarray data via nonnegative matrix factorization for visualization and clustering analysis", *Journal of Biomedical Informatics*, vol. 41, pp. 602-606, 2008.