Character Recognition Analysis with Nonnegative Matrix Factorization

TOLGA ENSARİ
Computer Engineering
Istanbul University
Istanbul University, Computer Engineering, Avcilar, 34320, Istanbul
TURKEY
ensari@istanbul.edu.tr http://ce.istanbul.edu.tr/ensari

Abstract: - In this paper, we analyze character recognition performance of three different nonnegative matrix factorization (NMF) algorithms. These are multiplicative update (MU) rule known as standard NMF, alternating least square (NMF-ALS) and projected gradient descent (NMF-PGD). They are most preferred approaches in the literature. There are lots of application areas for NMF such as robotics, bioinformatics, vision, sound and others. We use well known MNIST digit data set to test the performance of NMF, NMF-ALS and NMF-PGD. Experimental results show that NMF-ALS is the best and the worst one is NMF-PGD for these there algorithms in the meaning of accuracy. Therefore, we suggest NMF-ALS method can be used to analyze patterns on character recognition.

Key-Words: - Nonnegative matrix factorization, character recognition, pattern recognition, multiplicative update rule, alternating least square, projected gradient descent

1 Introduction

NMF is one of the popular machine learning algorithms which has wide application area in last decade. It has been proposed by Lee and Seung in 1999 as part based learning approach [1, 2].

There are lots of NMF algorithms proposed in the literature such as probabilistic NMF, quadratic NMF, alternating least square NMF, projected gradient descent NMF, graph regularized NMF, localized NMF [5-9]. Also, more proposed structures studied by researchers in the literature. Several clustering/classification based applications give successful results especially in robotics, text classification, image classification, gene expression data and others. Because of its popularity, it has been widely studied for lots of implementation areas [5, 8 and 9].

In [5], authors propose correntropy based NMF. Experimental results on text classification and face recognition show its advantage. Accuracy values and entropy based calculations were given in this article. NMF is a feature extraction method like other unsupervised methods principal component analysis (PCA), singular value decomposition (SVD), k-means and self organizing maps (SOM). Some hybrid approaches are also tested in order to get better accuracies. Especially, several learning algorithms, k cluster values and error functions effect the performance of NMF [3, 4, 6 and 7-9].

Sparse NMF and its structure on implementations is one of the hot topic in last decade. After defining deep learning method as a next generation of neural networks, NMF composed as hybrid method. The dimensionality of data can be reduced not only with neural networks, but also with NMF and its various types. Moreover, NMF's nonnegativity constraint makes itself robust in many applications. Therefore, NMF will be one of the hot topic in machine learning research area in the future and many optimization problems can be solved with different types of NMF.

2 Nonnegative Matrix Factorization (NMF)

Nonnegative matrix factorization (NMF) can be defined approximation between A matrix and WH product. It was proposed by Lee and Seung in 1999. The authors derived and use multiplicative update rule in their study. NMF gives better accuracy results especially for part based learning [1, 2].

Here A is called nonnegative data matrix, W and H are called lower rank nonnegative matrix representations. If A has mxn dimension, let define its rank as $k \leq \min{(m,n)}$. In this case, W and H matrices' dimensions will be mxk and kxn, respectively.

The factorization process of NMF can be calculated with Eq (1).

$$A \cong W\,H \quad (1)$$

$$\text{subject to } W \geq 0 \text{ and } H \geq 0$$

where the function of $Obj(A, WH)$ is an objective function (calculation of error) in order to optimize and descent its value.

Most known objective functions used for NMF are Euclidean, Kullback-Leibler divergence and Itakura-Saito divergence. Recently proposed correntropy based objective function also has a good performance on several NMF applications [5].

2.1. NMF with MU:

In this section, we describe the multiplicative update (MU) rule which has been proposed by Lee and Seung [1, 2]. According to this update rule, W and H matrices have to be iterated for $\forall i$ and $\forall j$ shown in Eq (2) and Eq (3).

$$H_j^{k+1} = H_{bj}^{k} \frac{((W^k)^T A)_j}{((W^k)^T W^k H^k)_j}, \quad \forall\, j. \quad (2)$$

$$W_i^{k+1} = W_i^{k} \frac{(A\,(H^{k+1})^T)_i}{(W^k H^{k+1} (H^{k+1})^T)_i}, \quad \forall i. \quad (3)$$

Finalizing the iteration number generally depends on error calculation or number of maximum iteration. So, it will be ended after many times executed. Stability analysis and some conditions was studied in various articles. Therefore, it is applicable and mathematically proved, especially with Lyapunov stability theorem. When the algorithm converges to stable point, that point can be stationary or local minimum. MU does not guarantee the global convergence of NMF. Some studies in the literature can generate sufficient conditions, not necessary conditions. Some modifications on MU studied many times to accelerate the convergence and iteration for NMF applications [10-24].

2.2. NMF with ALS:

ALS algorithm was firstly proposed by Paatero and Taper [24]. ALS on MU can be stated as optimization problem of Eq (1) is convex for either W or H matrix, not convex for both W and H matrices. Therefore, values on matrices can be calculated with a simple least squares computation. All negative elements is changed with zero value. These calculations make NMF-ALS more flexible then MU. NMF-ALS also allows the iteration trend to escape to stuck in local points. The algorithm is generally very fast but depends on application area. Convergence of ALS has been also proved in the literature by many researchers [4, 10 and 20].

2.3. NMF with PGD:

Gradient descent method is one of the well known optimization algorithm to solve the engineering problems. The partial derivatives and step size parameters can control the update procedure and reach it to optimal condition. Many implementation uses a simple projections step. The updated matrices are projected to nonnegative orthant. Because of nonnegativity projection, convergence analysis will be more difficult for NMF-PGD. This algorithm is very sensitive on iterations.

3 Experimental Results

MNIST character digits data set is used for the experiments. It is one of the popular data set which used especially in machine learning area. It has been collected from many people using their handwritten digits. Some of them have been shown in Fig 1. In the initial stage, W and H matrices have random values. Generally, denominator has sufficiently small epsilon value in order to avoid zero dividing.

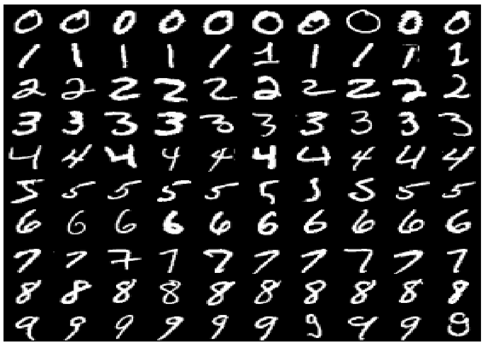

Fig 1. Some digits from MNIST data set

We use multiplicative update rule, proposed by Lee and Seung, for standard NMF. This rule can be seen below in Eq (2) and Eq (3). We use accuracy (%) metric in order to measure the performance of NMF, NMF-ALS and NMF-PGD. We mentioned and explained these algorithms in Section 2 and used Matlab program for experiments. Initial values of W and H matrices are completely random and we set 1,000 for maximum iteration number.

Table 1. Accuracy values of MNIST character data set with different k- (rank) values

Methods	Accuracy with k-values (%)							
	k=25	k=36	k=49	k=64	k=81	k=100	k=121	k=144
NMF	87	85,6	89,3	90,3	83,3	82,3	82	84
NMF-ALS	87	91,3	88,3	85,6	89,3	84	87,6	85,3
NM-PGD	86,6	83	85,3	81,6	77,6	78,3	72	67,6

Recognition results on the experiments are given in Table 1 with accuracy. Different k-values (ranks) also analyzed for this implementation for three algorithms of NMF. We get the results of accuracies only for one run of each algorithm in Matlab program.

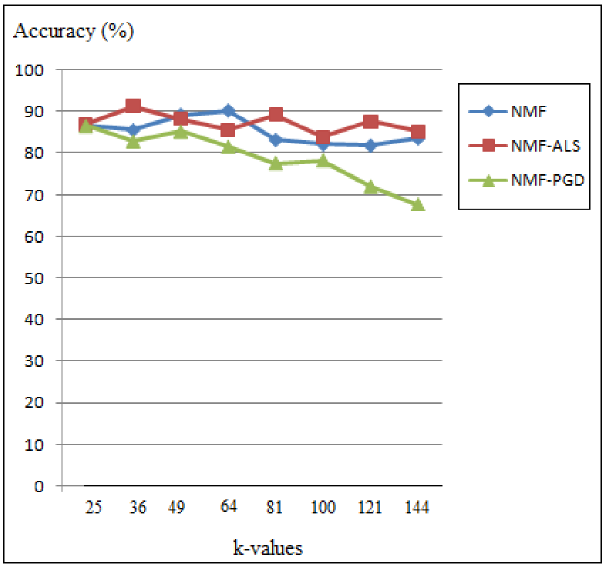

Fig 2. Accuracy graph of NMF, NMF-ALS and NMF-PGD for MNIST data set.

Results of Table 1 are graphed in Fig 2 with accuracy. According to these results, NMF-ALS reaches up to 91.3 % accuracy value as the best recognition rate. In general, we can say that NMF-ALS is the best one and NMF-PGD is the worst for character recognition application. Different k- rank values also can be preferred for several cases. On the other hand, recognition performance decreases when NMF k-lower rank values increase. Therefore, small value of k-(rank) values shall be chosen for character recognition.

Acknowledgement

This work was supported by Scientific Research Projects Coordination Unit of Istanbul University, Istanbul, Turkey with Project number: YADOP 43529.

4 Conclusion

In this study, we analyze character recognition performance of NMF, NMF-ALS and NMF-PGD. We use MNIST character recognition data set which is one of the popular in machine learning and pattern recognition area. Experimental results show that NMF-ALS is the best and NMF-PGD is the worst for three types of NMF mentioned above for character recognition. Moreover, we observe accuracy increases when factorization done with small k-lower rank values. These experiments can also be extended for different types of data sets and application areas. The best accuracy has been obtained for NMF-ALS with k=36 lower rank value.

Entropy based approaches, other learning rules and other types of NMF(s) can be implemented to get better NMF recognition performance for future studies.

References:

[1] D. D. Lee and S. Seung, Learning the Parts of Objects by Nonnegative Matrix Factorization, *Nature* 401, 1999, pp. 788-791.

[2] D. D. Lee and S. Seung, Algorithms for Nonnegative Matrix Factorization, *International Conference on Neural Information Processing*, 2000.

[3] S. Choi, “Algorithms for orthogonal nonnegative matrix factorization”, *in Proc. Int. Joint Conf. on Neural Networks*, Hong Kong, Jun. 2008, pp. 1828-1832.

[4] F. Shahnaz, M. W. Berry, V. P. Pauca, and R. J. Plemmons, “Document clustering using nonnegative matrix factorization, *Int. Journal of Information Processing and Management*, vol. 42, Issue 2, Mar. 2006, pp. 373-386.

[5] T. Ensari, J. Chorowski, and J. M. Zurada, “Correntropy-based document clustering via nonnegative matrix factorization”, *in Proc. Int. Conf. on Artificial Neural Networks*, vol. 7553, Sept. 2012, pp. 347-354.

[6] X. Li, J. Zhou, L. Tong, X. Yu, J. Guo, C. Zhao, Structured Discriminative Nonnegative Matrix Factorization for Hyperspectral

Unmixing, *IEEE International Conference on Image Processing,* 2016, pp. 1848-1852.

[7] W. Zhao, H. Ma, and N. Li, "A new non-negative matrix factorization algorithm with sparseness constraints", *in Proc. the Int. Conf. on Machine Learning and Cybernetics*, Jul. 2011, pp 1449-1452.

[8] H. Liu, Z. Wu, X. Li, D. Cai, and T. S. Huang, "Constrained nonnegative matrix factorization for image representation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, Jul. 2012.

[9] C. Fevotte and J. Idier, "Algorithms for nonnegative matrix factorization with the β-Divergence", *Neural Computation*, vol. 23, no. 9, Sept. 2011, pp. 2421-2456.

[10] M. Berry, M. Browne, A. M. Langville, V. P. Pauca, R. J. Plemmons, "Algorithms and Applications for Approximate Nonnegative Matrix Factorization", *Int. Journal of Computational Statistics for Approximate Nonnegative Matrix Factorization*, vol. 52, no. 1, Sept. 2007, pp. 155-173.

[11] C. J. Lin, "Projected Gradient Methods for Nonnegative Matrix Factorization", *Neural Computation*, *Neural Computation*, vol. 19, no. 10, Aug. 2007, pp. 2756-2779.

[12] H. Kim, H. Park, "Nonnegative Matrix Factorization Based on Alternating Nonnegativity Constrained Least Squares and Active Set Method", *SIAM Journal of Matrix Analysis and Applications*, vol. 30, no. 2, 2008, pp. 713-730.

[13] H. Liu, Z. Wu, X. Li, D. Cai, and T. S. Huang, "Constrained nonnegative matrix factorization for image representation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, Jul. 2012.

[14] D. Guillamet, J. Vitria, and B. Schiele, "Introducing a weighted non-negative matrix factorization for image classification", *Pattern Recognition Letters*, vol. 24, pp. 2447-2454, Oct. 2003.

[15] R. He, W. S. Zheng, and B. G. Hu, "Maximum correntropy criterion for robust face recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1561-1576, Aug. 2011.

[16] R. He, B. G. Hu, W. S. Zheng, and X. W. Kong, "Robust principal component analysis based on maximum correntropy criterion", *IEEE Trans. on Image Processing*, vol. 20, no. 6, pp. 1485-1494, Jun. 2011.

[17] R. Chalasani and J. H. Principe, "Self organizing maps with correntropy induced metric", *in Proc. Int. Joint Conf. on Neural Networks*, Jul. 2010, pp. 1-6.

[18] W. Zhao, H. Ma, and N. Li, "A new non-negative matrix factorization algorithm with sparseness constraints", *in Proc. the Int. Conf. on Machine Learning and Cybernetics*, Jul. 2011, pp 1449-1452.

[19] P. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, Pearson Addison Wesley, May 2005.

[20] M. W. Berry, N. Gillis, and F. Glineur, "Document classification using nonnegative matrix factorization and underapproximation", *in Proc. Int. Symp on Circuits and Systems*, May 2009, pp. 2782-2785.

[21] B. J. Shastri and M. D. Levine, "Face recognition using localized features based on non-negative sparse coding", *Machine Vision and Applications*, vol. 18, no. 2, pp. 107-122, Apr. 2007.

[22] H. J. Oh, K. M. Lee, and S. U. Lee, "Occlusion invariant face recognition using selective local non-negative matrix factorization basis images", *Image and Vision Computing*, vol. 26, Issue 11, pp. 1515-1523, Nov. 2008.

[23] J. Y. Pan and J. S. Zhang, "Large margin based nonnegative matrix factorization and partial least squares regression for face recognition", *Pattern Recognition Letters*, vol. 32, pp. 1822-1835, Oct. 2011.

[24] P. Paatero, "Least squares formulation of robust non-negative factor analysis", *Chemometrics and Intelligent Laboratory Systems 37*, pp. 23-35, May 1997.