The Ninth International Symposium on Operations Research and Its Applications (ISORA'10)
Chengdu-Jiuzhaigou, China, August 19–23, 2010

Survey on the Variations and Applications of Nonnegative Matrix Factorization

Zhong-Yuan Zhang[1] Jie Zhang[2]

[1]School of Statistics, Central University of Finance and Economics, Beijing 100080
[2]School of Mathematical Sciences, Dalian University of Technology, Dalian

Abstract Nonnegative Matrix Factorization (NMF) has been proved to be valuable in many applications of data mining, especially in unsupervised learning. This paper will briefly review the variations and applications of NMF. All of these applications are based on the ability of NMF to extract the hidden patterns or trends behind the observed samples automatically.

Keywords Nonnegative Matrix Factorization; Variations of NMF; Image Processing; Clustering; Co-clustering

1 Introduction

Nonnegative Matrix Factorization has been proved to be valuable in many fields of data mining, especially in unsupervised learning. In this paper, we will briefly review its variations and applications in image processing, data clustering, semi-supervised clustering, bi-clustering (co-clustering) and financial data mining. Note that we cannot cover all the interesting works on NMF, but generally speaking, the special point on NMF is its ability to recover the hidden patterns or trends behind the observed data automatically, which makes it suitable for image processing, feature extraction, dimensional reduction and unsupervised learning. The preliminary theoretical analysis concerning this ability, in other words, the relations between NMF and some other unsupervised learning models have been studied in ref. [4, 5].

The rest of the paper is organized as follows: Sect. 2 surveys a variety of variations of NMF, Sect. 3 surveys the applications of NMF, and Sect. 4 concludes.

2 Variations of NMF

In this section, we will briefly review the variations that are rooted from NMF and proposed from different perspectives. We summarize the different models in Table 1 but omit the details here due to space limitation. One can find more in the corresponding references.

3 Applications of NMF

3.1 Image Processing ([18]):

Nonnegative Matrix Factorization can trace its history back to 1970's, but has attracted lots of attention due to the research of Lee & Seung ([18, 19]). In their works, the model

Table 1: Summary of different models based on NMF. Each row lists a variant and its associated constraints. $\pm$ means that the matrix in the corresponding column may have mixed signs, $+$ means that the matrix is nonnegative, $0-1$ means that the elements in the matrix can only be zero or one and I denotes identity matrix.

Models	Cost Function	X	F	S	G
NMF1 [18, 19]	Least Squares Error	$+$	$+$	I	$+$
NMF2 [18, 19]	K-L Divergence	$+$	$+$	I	$+$
Semi-NMF [7]	Least Squares Error	$\pm$	$\pm$	I	$+$
Convex-NMF [7]	Least Squares Error	$\pm$	$\pm$, $F_{:,j}$ is the convex combination of $\{X_{:,j}; j=1,\cdots,n\}$	I	$+$
Tri-NMF [6]	Least Squares Error	$+$	$+, F^TF=I$	$+$	$+, G^TG=I$
Symmetric-NMF [21]	Least Squares Error	$+$, symmetric	$+, F=G$	$+$	$+$
K-means [4]	Least Squares Error	$+$, symmetric	$+, F=G$	I	$+, G^TG=I$
PLSI [5, 11]	K-L Divergence	$\sum_{i,j} X_{ij}=1$	$\sum_i F_{ik}=1, i=1,\cdots m$	Diagonal, $\sum_k S_{kk}=1$	$\sum_j G_{jk}=1, j=1,\cdots n$
LNMF [20, 9]	K-L Divergence with penalty terms[a]	$+$	$+$	I	$+$
NNSC [12]	Least Squares Error with penalty terms[b]	$+$	$+$	I	$+$
SNMF1 [22]	K-L Divergence with penalty terms[c]	$+$	$+$	I	$+$
SNMF2 [10]	Least Squares Error with penalty terms[d]	$+$	$+$	I	$+$
SNMF3 [26]	Least Squares Error with penalty terms[e]	$+$	$+$	I	$+$
NMFSC [13]	Least Squares Error	$+$	$+, \mathrm{Sp}(F_{:,j})=S_F{}^f, j=1,\cdots,k$	I	$+, \mathrm{Sp}(G_{:,j})=S_G{}^g, j=1,\cdots,k$
NMF/L [16]	Least Squares Error with penalty terms[h]	$+$	$+$	I	$+$
NMF/R [16]	Least Squares Error with penalty terms[i]	$+$	$+$	I	$+$
nsNMF [25]	K-L Divergence	$+$	$+$	$S=(1-\theta)I+\frac{\theta}{k}II^T$	$+$
CUR [24]	Least Squares Error	$+$	$+^j$	$+$	$+^k$
BMF [32]	Least Squares Error	$0-1$	$0-1$	I	$0-1$

[a] $\sum_{i,j}(X_{ij}\log\frac{X_{ij}}{(FG^T)_{ij}}-X_{ij}+(FG^T)_{ij})+\alpha\sum_{i,j}(F^TF)_{ij}-\beta\sum_i(G^TG)_{ii}$.

[b] $\|X-FG^T\|_F^2+\lambda\sum_{i,j}G_{ij}$

[c] $\sum_{i,j}[X_{ij}\log\frac{X_{ij}}{(FG^T)_{ij}}-X_{ij}+(FG^T)_{ij}]+\lambda\sum_{i,j}G_{ij}$.

[d] $\|X-FG^T\|_F^2+\lambda\sum_{i,j}G_{ij}^2$

[e] $\|X-FG^T\|_F^2+\alpha\sum_{i,j}F_{ij}^2+\beta\sum_{i,j}G_{ij}^2$

[f] $\mathrm{Sp}(a)=(\sqrt{n}-\sum|a_j|/\sqrt{\sum a_j^2})/\sqrt{n-1}$, S_F is some constant.

[g] S_G is some constant.

[h] $\|X-FG^T\|_F^2+\eta\|G\|_F^2+\beta\sum_i\|F_{i,:}\|_1^2$.

[i] $\|X-FG^T\|_F^2+\eta\|F\|_F^2+\beta\sum_i\|G_{i,:}\|_1^2$.

[j] Columns of F are composed of a small number of columns in X

[k] Columns of G are composed of a small number of rows in X

was applied to image processing successfully. Hence we review the applications of NMF on this aspect firstly.

In image processing, the data can be represented as an $n \times m$ nonnegative matrix X, each column of which is an image described by n nonnegative pixel values. Then NMF model can find two factor matrices F and G such that $X \approx FG^T$. F is the so-called basis matrix since each column can be regarded as a part of the whole such as nose, ear or eye, etc. for facial image data. G is the coding matrix and each row is the weights by which the corresponding image is reconstructed as the linear combination of the columns of F.

In summary, NMF can discover the common basis hidden behind the observations and the way how the images are reconstructed by the basis. Indeed, the psychological and physiological researches have shown evidence for part-based representation in the brain, which is also the foundation of some computational theories ([18]). But further researches have also shown that the standard NMF model does not necessarily give the correct part-of-whole representations ([20, 14]), hence many efforts have been done to improve the sparseness of NMF in order to identify more localized features that are building parts for the whole representation.

3.2 Clustering ([1, 23, 31]):

One of the most interesting and successful applications of NMF is to cluster data such as text, image or biology data, i.e. discovering patterns automatically from data. Given a nonnegative $n \times m$ matrix X, each column of which is a sample and described by n features, NMF can be applied to find two factor matrices F and G such that $X \approx FG^T$, where F is $n \times r$ and G is $m \times r$, and r is the cluster number. Columns of F can be regarded as the cluster centroids while G is the cluster membership indicator matrix. In other words, the sample i is of cluster k if G_{ik} is the largest value of the row $G_{i,:}$.

The good performance of NMF in clustering has been validated in several different fields including bioinformatics (tumor sample clustering based on microarray data, [1]), community structure detection of the complex network ([23]) and text clustering ([27, 30, 31]).

3.3 Semi-supervised Clustering ([2]):

In many cases, some background information concerning the pairwise relations of some samples are known and we can add them into the clustering model in order to guide the clustering process. The resulting constrained problem is called semi-supervised clustering. Specifically, the following two types of pairwise relations are often considered:

- *Must-link* specifies that two samples should have the same cluster label;
- *Cannot-link* specifies that two samples should not have the same cluster label.

Then, One can establish two nonnegative matrices $W_{\text{reward}} = \{w_{ij} :\ \text{sample i and sample } j \text{ is in the same class}\}$ and $W_{\text{penalty}} = \{w_{ij} :\ \text{sample } i \text{ and sample } j \text{ is not in the same class}\}$ based on the above information, and the similarity matrix W of the samples can then be replaced by $W - W_{\text{reward}} + W_{\text{penalty}}$ (note that it is still a symmetric matrix). Finally, NMF is applied:

$$\min_{S \geqslant 0, G \geqslant 0} \|(A - W_{\text{reward}} + W_{\text{penalty}}) - GSG^T\|_F^2,$$

where the symmetric matrix G is the cluster membership indicator, i.e., sample j is of cluster i if the element G_{ij} is the largest value of the column $G_{:,j}$ (and is also the largest value of the row $G_{j,:}$). Theoretical analysis and practical applications have been contributed by ref [2]. We summarize the main theoretical results but omit the details here.

Theorem 1.
Orthogonal Semi-Supervised NMF clustering is equivalent to Semi-Supervised Kernel K-means ([17]).

Theorem 2.
Orthogonal Semi-Supervised NMF clustering is equivalent to Semi-Supervised Spectral clustering with Normalized Cuts ([15]).

3.4 Bi-clustering (co-clustering)

Bi-clustering was recently introduced by Cheng & Church ([3]) for gene expression data analysis. In practice, many genes are only active in some conditions or classes and remain silent under other cases. Such gene-class structures, which are very important to understand the pathology, can not be discovered using the traditional clustering algorithms. Hence it is very necessary to develop bi-clustering models/algorithms to identify the local structures. Bi-clustering models/algorithms are different from the traditional clustering methodologies which assign the samples into specific classes based on the genes' expression levels across *ALL* the samples, they try to cluster the rows (features) and the columns (samples) of a matrix simultaneously.

In other words, the idea of bi-clustering is to characterize each sample by a subset of genes and to define each gene in a similar way. As a consequence, bi-clustering algorithms can select the groups of genes that show similar expression behaviors in a subset of samples that belong to some specific classes such as some tumor types, thus identify the local structures of the microarray matrix data [3, 28]. Binary Matrix Factorization (BMF) was presented for solving bi-clustering problem: the input binary gene-sample matrix X[1] is decomposed into two binary matrices F and G such that $X \approx FG^T$. The binary matrices F and G can explicitly designate the cluster memberships for genes and samples. Hence BMF offers a framework for simultaneously clustering the genes and samples.

An example is given here to demonstrate the biclustering capability of BMF. Given the original data matrix

$$X = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

One can see two biclusters, one in the upper-right corner, and one in lower-left corner. Our BMF model gives

[1] [32] has discussed the details on how to discretize the microarray data into a binary matrix

$$F = (F_{:,1}, F_{:,2}) = \begin{pmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}; \quad G = (G_{:,1}, G_{:,2}) = \begin{pmatrix} 0 & \boxed{1 \quad 1} \\ 0 & 0 \quad 0 \end{pmatrix}$$

The two discovered biclusters are recovered in a clean way:

$$FG^T = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

3.5 Financial Data Mining

3.5.1 Underlying Trends in Stock Market ([8]):

In the stock market, it has been observed that the stock price fluctuations does not behave independently of each other but are mainly dominated by several underlying and unobserved factors. Hence try to identify the underlying trends from the stock market data is an interesting problem, which can be solved by NMF. Given an $n \times m$ nonnegative matrix X, columns of which is the records of the stock prices during n time points, NMF can be applied to find two nonnegative factors F and G such that $X \approx FG^T$, where columns of F are the underlying components. Note that identifying the common factors that drive the prices is somewhat similar to blind source separation (BSS) in signal processing. Furthermore, G can be used to identify the cluster labels of the stocks (see Sect. 3.2) and the most interesting result is that the stocks of the same sector is not necessarily assigned into the same cluster and vice versa, which is of potential use to guide diversified portfolio, in other words, investors should diversify their money into not only different sectors, but also different clusters.

3.5.2 Discriminant Features Extraction in Financial Distress Data ([29]):

Building appropriate financial distress prediction model based on the extracted discriminative features is more and more important under the background of financial crisis. Ref [29] presents a new prediction model which is indeed a combination of K-means, NMF and support vector machine (SVM). The basic idea is to train a SVM classifier in the reduced dimensional space which is spanned by the discriminative features extracted by NMF, the algorithm of which is initialized by K-means. The details can be found in ref [29].

4 Conclusions and Future Works

We have given a short survey on the variations and applications of NMF. Besides the good performance of unsupervised learning, another advantage of NMF is its flexibility regarding the choices of its objective functions and the algorithms employed to solve it. These divergence functions and algorithms may lead to different numerical results, but there is still lack of systematic analysis of the relationships among the objective divergence functions, the algorithms and the applications, which is an interesting problem.

Acknowledgements

This work is supported by the Foundation of Academic Discipline Program at Central University of Finance and Economics and the National Natural Science Foundation of China under grant No. 60873205.

References

[1] J. P. Brunet, P. Tamayo, T. R. Golub, and J. P. Mesirov. Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci U S A*, 101(12):4164–4169, March 2004.

[2] Yanhua Chen, Manjeet Rege, Ming Dong, and Jing Hua. Incorporating user provided constraints into document clustering. In *ICDM '07: Proceedings of the 2007 Seventh IEEE International Conference on Data Mining*, pages 103–112, Washington, DC, USA, 2007. IEEE Computer Society.

[3] Yizong Cheng and George M. Church. Biclustering of expression data. In *Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology*, pages 93–103. AAAI Press, 2000.

[4] C. Ding, X. He, and H. D. Simon. On the equivalence of nonnegative matrix factorization and spectral clustering. In *SIAM Data Mining Conf*, 2005.

[5] Chris Ding, Tao Li, and Wei Peng. On the equivalence between non-negative matrix factorization and probabilistic latent semantic indexing. *Comput. Stat. Data Anal.*, 52(8):3913–3927, 2008.

[6] Chris Ding, Tao Li, Wei Peng, and Haesun Park. Orthogonal nonnegative matrix t-factorizations for clustering. In *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 126–135, New York, NY, USA, 2006. ACM.

[7] Chris H. Q. Ding, Tao Li, and Michael I. Jordan. Convex and semi-nonnegative matrix factorizations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(1):45–55, 2010.

[8] Konstantinos Drakakis, Scott Rickard, Ruairi de Frein, and Andrzej Cichocki. Analysis of financial data using non-negative matrix factorization. *International Mathematical Forum*, 3(38):1853–1870, 2008.

[9] Tao Feng, Stan Li, Heung-Yeung Shum, and HongJiang Zhang. Local non-negative matrix factorization as a visual representation. In *Development and Learning, International Conference on*, 2002.

[10] Yuan Gao and George Church. Improving molecular cancer class discovery through sparse non-negative matrix factorization. *Bioinformatics*, 21(21):3970–3975, 2005.

[11] Eric Gaussier and Cyril Goutte. Relation between plsa and nmf and implications. In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 601–602, New York, NY, USA, 2005. ACM.

[12] P. O. Hoyer. Non-negative sparse coding. In *Neural Networks for Signal Processing, 2002. Proceedings of the 2002 12th IEEE Workshop on*, pages 557–565, 2002.

[13] Patrik O. Hoyer. Non-negative matrix factorization with sparseness constraints. *J. Mach. Learn. Res.*, 5:1457–1469, 2004.

[14] Patrik O. Hoyer. Non-negative matrix factorization with sparseness constraints, Aug 2004.

[15] Xiang Ji and Wei Xu. Document clustering with prior knowledge. In *SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 405–412, New York, NY, USA, 2006. ACM.

[16] Hyunsoo Kim and Haesun Park. Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics*, 23(12):1495–1502, 2007.

[17] Brian Kulis, Sugato Basu, Inderjit Dhillon, and Raymond Mooney. Semi-supervised graph clustering: a kernel approach. In *ICML '05: Proceedings of the 22nd international conference on Machine learning*, pages 457–464, New York, NY, USA, 2005. ACM.

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

[19] Daniel D. Lee and Sebastian H. Seung. Algorithms for non-negative matrix factorization. In *Annual Conference on Neural Information Processing Systems*, pages 556–562, 2000.

[20] S. Z. Li, Xin W. Hou, Hong J. Zhang, and Qian S. Cheng. Learning spatially localized, parts-based representation. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, 2001.

[21] Tao Li and Chris Ding. The relationships among various nonnegative matrix factorization methods for clustering. In *ICDM '06: Proceedings of the Sixth International Conference on Data Mining*, pages 362–371, Washington, DC, USA, 2006. IEEE Computer Society.

[22] W. Liu, N. Zheng, and X. Lu. Non-negative matrix factorization for visual coding. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'2003)*.

[23] Bo Long, Xiaoyun Wu, Zhongfei Zhang, and Philip S. Yu. Community learning by graph approximation. In *Proceedings of the 8th IEEE International Conference on Data Mining*. ICDM 2007, 2007.

[24] Michael W. Mahoney and Petros Drineas. Cur matrix decompositions for improved data analysis. *Proceedings of the National Academy of Sciences*, 106(3):697–702, January 2009.

[25] A. Pascual-Montano, J. M. Carazo, K. Kochi, D. Lehmann, and R. D. Pascual-Marqui. Nonsmooth nonnegative matrix factorization (nsnmf). *IEEE transactions on Pattern Analysis and Machine Intelligence*, 28(3):403–415, March 2006.

[26] V. Paul Pauca, J. Piper, and Robert J. Plemmons. Nonnegative matrix factorization for spectral data analysis. *Linear Algebra and its Applications*, 416(1):29–47, 2006.

[27] V. Paul Pauca, Farial Shahnaz, Michael W. Berry, and Robert J. Plemmons. Text mining using non-negative matrix factorizations. In *Proceedings of the Fourth SIAM International Conference on Data Mining*, 2004.

[28] Amela Prelić, Stefan Bleuler, Philip Zimmermann, Anja Wille, Peter Bühlmann, Wilhelm Gruissem, Lars Hennig, Lothar Thiele, and Eckart Zitzler. A systematic comparison and evaluation of biclustering methods for gene expression data. *Bioinformatics*, 22(9):1122–1129, 2006.

[29] Bernardete Ribeiro, Catarina Silva, Armando Vieira, and João Carvalho das Neves. Extracting discriminative features using non-negative matrix factorization in financial distress data. In *ICANNGA*, pages 537–547, 2009.

[30] Farial Shahnaz, Michael W. Berry, Pauca, and Robert J. Plemmons. Document clustering using nonnegative matrix factorization. *Information Processing & Management*, 42(2):373–386, March 2006.

[31] Wei Xu, Xin Liu, and Yihong Gong. Document clustering based on non-negative matrix factorization. In *SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 267–273, New York, NY, USA, 2003. ACM Press.

[32] Zhong-Yuan Zhang, Tao Li, Chris Ding, Xian-Wen Ren, and Xiang-Sun Zhang. Binary matrix factorization for analyzing gene expression data. *Data Min. Knowl. Discov.*, 20(1):28–52, 2010.