

## 非负矩阵分解的一个约束稀疏算法

李臣明,张师明,李昌利\*

(河海大学 计算机与信息学院,江苏 南京 211100)

**摘要:**针对非负矩阵分解中系数矩阵不够稀疏的问题,提出一个新的约束非负矩阵分解算法。在经典非负矩阵分解的优化函数中施加稀疏性约束,并对分解系数矩阵施加最小相关约束,与此同时对基矩阵施加2-范数约束,在保证非负约束和分解精度的基础上,使分解后得到的矩阵尽可能稀疏,这样可以更加节省存储空间,分解结果更优。对比实验表明,提出的算法具有更好的稀疏性,且实验误差更小。

**关键词:**非负矩阵分解(NMF);稀疏性;最小相关系数;2-范数

**中图分类号:**TP391.41

**文献标志码:**A

### A Constrained Sparse Algorithm for Nonnegative Matrix Factorization

LI Chenming, ZHANG Shiming, LI Changli\*

(College of Computer and Info. Eng., Hohai Univ., Nanjing 211100, China)

**Abstract:** Aiming at the lack of sparseness of factorization matrix in the nonnegative matrix factorization (NMF) algorithm, a new constrained NMF algorithm was proposed. A sparseness constraint was added to the original nonnegative matrix factorization (NMF) algorithm, and the minimum correlation constraint was imposed on the coefficient matrices and the 2-norm constraint was imposed on the basis matrix at the same time, which can ensure the non-negative constraint and accurate decomposition, and can also make the decomposed matrix sparse as far as possible, saving more storage space. Comparison with the experiments showed that the propose algorithm has the better sparseness and smaller error than both the original NMF algorithm and the SNMF algorithm.

**Key words:** non-negative matrix factorization (NMF); sparseness; the least correlated component constraints; 2-norm

随着移动多媒体应用的推广和普及,各种多媒体数据(如图像数据等)得到爆发式地增长,如何在海量数据中进行快速分析、搜索并获取少量感兴趣数据成为现阶段需要解决的问题。近年来,盲源分离(blind source separation, BSS)成为信号处理领域的一个研究热点<sup>[1]</sup>。其基本思想是在未知源信号和传输信道参数的情况下,仅从观测到的混合信号中恢复出源信号。

1999年Lee与Seung<sup>[2]</sup>在著名期刊Nature上发表了文章,首次提出一种带有非负约束的矩阵分解方法,即非负矩阵分解(non-negative matrix factorization, NMF)。该方法要求给定源数据满足非负性,即其物理意义是数据的非负性要满足其物理信号的

真实性,在实际应用过程中,非负数据广泛的存在,且分解后的结果具有明确的物理含义,如灰度图像的元素值为0~255(或0~1)、电表数据一直保持非负等<sup>[3]</sup>。NMF的非负限定符合直观上的理解:整体是由部分组成的,因此它在某种意义上把握了数据描述的本质;非负性的限制导致了相应描述在一定程度上稀疏性<sup>[4]</sup>。稀疏性描述将在一定程度上将抑制外界环境干扰,如光照的变化、图像部分遮挡或物体的旋转等。非负矩阵分解具有的这些优势,逐渐成为一种盲源分离的新的有效手段,该方法成为近年来研究的热点,有望在信号处理、海量数据分析和数据挖掘等领域取得突破性进展。

目前,很多利用增加新的限制条件到已有的

收稿日期:2014-06-11

基金项目:河海大学中央高校基本科研业务费项目(2013B32514);国家自然科学基金资助项目(61101211)

作者简介:李臣明(1969—),男,副教授,博士。研究方向:复杂系统建模仿真、智能信息处理及嵌入式系统技术。

\*通信联系人

NMF 模型中<sup>[5]</sup>,以便使得分解结果能够满足更多性质的要求,例如加入稀疏性<sup>[3,6-7]</sup>、平滑性<sup>[8]</sup>、正交性<sup>[9]</sup>、对称性<sup>[10]</sup>和聚类分析<sup>[11]</sup>等条件限制,从而得到分解结果更加合理。也有对 NMF 系数矩阵施加相关性约束以实现盲源分离,放宽了对基矩阵独立性的要求<sup>[7]</sup>,但其要求基矩阵稀疏。而实际的基矩阵信号信道参数常常是非稀疏的,该方法具有一定的局限性。作者提出一种基于非负矩阵分解的新的稀疏算法,该算法对系数矩阵施加最小相关系数约束和对基矩阵施加向量 2 - 范数约束,并通过梯度下降法进行迭代求解。

## 1 非负矩阵分解问题算法

给定非负矩阵  $\mathbf{V} \in \mathbb{R}^{m \times n}$ , NMF 旨在求解 2 个非负子矩阵  $\mathbf{W} \in \mathbb{R}^{m \times r}$  和  $\mathbf{H} \in \mathbb{R}^{r \times n}$  使得

$$\mathbf{V} \approx \mathbf{WH} \quad (1)$$

其中,  $\mathbf{W}$  为基矩阵,  $\mathbf{H}$  为系数矩阵。一般情况下  $r \ll \min(m, n)$ , 达到降低数据存储维数, 从而降低数据存储空间。非负矩阵分解算法是一种低秩逼近算法, 较常用的是运用  $\mathbf{V}$  和  $\mathbf{WH}$  之间的欧几里德 (Euclidean) 距离平方作为目标函数来达到最佳逼近:

$$F = \frac{1}{2} \|\mathbf{V} - \mathbf{WH}\|_F^2 = \frac{1}{2} \sum_{i,j} (\mathbf{V}_{ij} - \sum_{k=1}^r \mathbf{W}_{ik} \mathbf{H}_{kj})^2 \quad (2)$$

NMF 的目标函数(2)关于  $\mathbf{W}$  和  $\mathbf{H}$  分别都是凸的, 但是联合起来并非是凸的, 即同时关于  $\mathbf{W}$  和  $\mathbf{H}$  的时候目标函数都是非凸的, 所以 NMF 求出的结果只能是局部最优解。利用梯度下降法给出一种更新规则, 通过迭代收敛求得  $\mathbf{W}$  和  $\mathbf{H}$ , 迭代规则为:

$$\begin{cases} \mathbf{W}_{ik} \leftarrow \mathbf{W}_{ik} - \eta_w \frac{\partial F}{\partial \mathbf{W}_{ik}}, \\ \mathbf{H}_{kj} \leftarrow \mathbf{H}_{kj} - \eta_H \frac{\partial F}{\partial \mathbf{H}_{kj}} \end{cases} \quad (3)$$

当学习速率按式(4)迭代取值时,

$$\begin{cases} \eta_w = \frac{\mathbf{W}_{ik}}{(\mathbf{WHH}^T)_{ik}}, \\ \eta_H = \frac{\mathbf{H}_{kj}}{(\mathbf{W}^T \mathbf{WH})_{kj}} \end{cases} \quad (4)$$

即可得到以下乘性迭代规则:

$$\begin{cases} \mathbf{W}_{ik} \leftarrow \mathbf{W}_{ik} \frac{(\mathbf{VH}^T)_{ik}}{(\mathbf{WHH}^T)_{ik}}, \\ \mathbf{H}_{kj} \leftarrow \mathbf{H}_{kj} \frac{(\mathbf{W}^T \mathbf{V})_{kj}}{(\mathbf{W}^T \mathbf{WH})_{kj}} \end{cases} \quad (5)$$

## 2 非负矩阵分解的一个约束稀疏算法

### 2.1 非负矩阵分解的稀疏约束

在数理统计理论中, 非相关源信号间的相关系数的绝对值一定小于它们混合信号间的相关系数的绝对值<sup>[4]</sup>。分离出来的信号的相关系数的绝对值越小就表明分离出来的信号间的相关性越小, 因此可以把分离信号相关系数最小作为对  $\mathbf{H}$  的一个约束条件。相关系数定义为<sup>[4]</sup>:

$$R = \frac{1}{2} \left[ \sum_{i=1}^n \ln(\mathbf{HH}^T)_{ii} - \ln|\mathbf{HH}^T| \right] \quad (6)$$

对  $i \neq j$ , 若  $\langle h_i, h_j \rangle = 0$ , 则  $R(\mathbf{H})$  取最小值。

矩阵范数是具有“长度”概念的函数, 2 - 范数  $\|\cdot\|_F$  的几何意义为欧几里得 2 维空间中矩阵对应点与原点的距离<sup>[12]</sup>。在基矩阵  $\mathbf{W}$  按列归一化的前提下, 最大化  $\|\mathbf{W}\|_F^2$  有利于结果尽量的稀疏, 故若在算法中对基矩阵  $\mathbf{W}$  的 2 - 范数的平方值作为约束, 可以使得结果更加稀疏。

提出的算法在 Lee 和 Seung 的基本非负矩阵算法的基础上, 增加了对系数矩阵  $\mathbf{H}$  进行最小相关约束和对基矩阵  $\mathbf{W}$  进行 2 - 范数约束, 得到改进的目标函数如下:

$$F(\mathbf{W}, \mathbf{H}) = \frac{1}{2} \|\mathbf{V} - \mathbf{WH}\|_F^2 - \alpha_w \|\mathbf{W}\|_F^2 + \alpha_H R(\mathbf{H}) \quad (7)$$

式中,  $R(\mathbf{H})$  为相关系数 (式(6)),  $\|\mathbf{W}\|_F^2 = \sum_{i,k} \mathbf{W}_{ik}^2$ ,  $\alpha_w$  和  $\alpha_H$  为正则化参数。

### 2.2 迭代学习规则

运用梯度下降法得到迭代规则(式(3))。对目标函数(7)采用梯度下降法得到:

$$\begin{cases} \frac{\partial F}{\partial \mathbf{W}_{ik}} = - \sum_{i,k} (\mathbf{VH}^T)_{ik} + \sum_{i,k} (\mathbf{WHH}^T)_{ik} - \alpha_w \mathbf{W}_{ik}, \\ \frac{\partial F}{\partial \mathbf{H}_{kj}} = - \sum_{k,j} (\mathbf{W}^T \mathbf{V})_{kj} + \sum_{k,j} (\mathbf{W}^T \mathbf{WH})_{kj} + \\ \alpha_H \left[ \frac{1}{(\mathbf{h}_k \mathbf{h}_k^T)_{\mathbf{H}_{kj}}} - [(\mathbf{HH}^T)^{-T} \mathbf{H}]_{kj} \right] \end{cases} \quad (8)$$

当学习速率按式(4)取值, 得到迭代公式:

$$\begin{cases} \mathbf{W}_{ik} \leftarrow \mathbf{W}_{ik} \frac{(\mathbf{VH}^T)_{ik} + \alpha_w \mathbf{W}_{ik}}{(\mathbf{WHH}^T)_{ik}}, \\ \mathbf{H}_{kj} \leftarrow \mathbf{H}_{kj} \left[ \frac{(\mathbf{W}^T \mathbf{V})_{kj}}{(\mathbf{W}^T \mathbf{WH})_{kj}} + \alpha_H \frac{[(\mathbf{HH}^T)^{-T} \mathbf{H}]_{kj}}{(\mathbf{W}^T \mathbf{WH})_{kj}} - \right. \\ \left. \alpha_H \frac{1}{(\mathbf{h}_k \mathbf{h}_k^T)_{\mathbf{H}_{kj}}} \frac{1}{(\mathbf{W}^T \mathbf{WH})_{kj}} \right] \end{cases} \quad (9)$$

写成矩阵形式为:

$$\begin{cases} \mathbf{W} \leftarrow \mathbf{W} ./ (\mathbf{W}\mathbf{H}\mathbf{H}^T) .* (\mathbf{V}\mathbf{H}^T + \alpha_w \mathbf{W}), \\ \mathbf{H} \leftarrow \mathbf{H} ./ (\mathbf{W}^T \mathbf{W}\mathbf{H}) .* \{ \mathbf{W}^T \mathbf{V} - \\ \alpha_H \left[ \frac{1}{(\mathbf{h}_k \mathbf{h}_k^T) \mathbf{H}} - (\mathbf{H}\mathbf{H}^T)^{-T} \mathbf{H} \right] \} \end{cases} \quad (10)$$

式中,“ $*$ ”表示矩阵按元素相乘。

总结算法步骤如下:

- 1) 任意选择非负归一化矩阵  $\mathbf{W}$  和  $\mathbf{H}$ ;
- 2) 按式(10)进行迭代;
- 3) 在每一步迭代后将  $\mathbf{W}$  和  $\mathbf{H}$  的负元素置零,并对  $\mathbf{W}$  按列进行归一化;
- 4) 循环进行步骤 2) ~ 3) 直至收敛,此时得到基矩阵  $\mathbf{W}$  和系数矩阵  $\mathbf{H}$ 。

### 3 实验结果

实验采用 Yale 人脸图像数据库<sup>[13]</sup>和 ORL 人脸图像数据库<sup>[14]</sup>,其中, Yale 数据库中图像像素大小均为  $100 \times 100$ ,即  $100 \times 100$  的矩阵, ORL 数据库中图像像素大小均为  $112 \times 92$ ,即  $112 \times 92$  的矩阵。图像中每个像素的灰度值为  $0 \sim 255$ ,在试验前把图像变换为双精度数据类型(double),即把所有的像素灰度值除以 255 从而变换为  $0 \sim 1$  的像素值。

图 1 为选取其中一张图片示例,其中,  $\mathbf{V}$  为原图像,  $\mathbf{WH}$  为分解后复原图像。

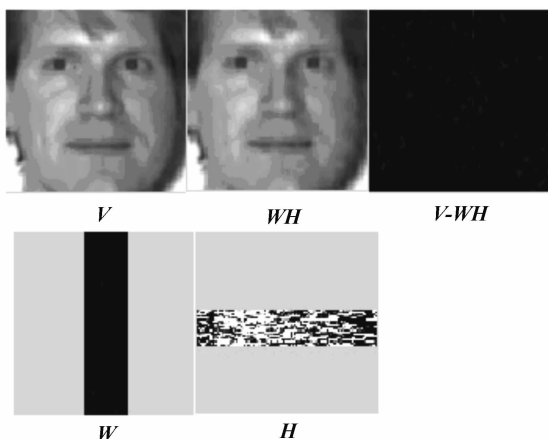


图 1 实验结果图

Fig.1 Results of the experiment

图 2 为 Yale 数据库图像实验结果。

图 3 为 ORL 数据库图像实验结果。

下面进行方法对比分析,与本实验进行对比的方法是 NMF 算法<sup>[2]</sup>和 SNMF 算法<sup>[3]</sup>,对比的指标为系数矩阵  $\mathbf{W}$  的 2-范数的平方值、系数矩阵  $\mathbf{H}$  的稀疏度( $\mathbf{H}$  所有列向量的稀疏度取平均值)、信噪比。信噪比(SNR)定义如下<sup>[5]</sup>:

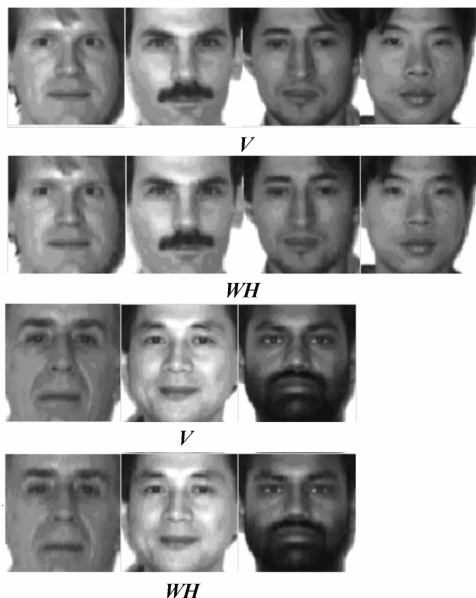


图 2 Yale 数据库实验结果

Fig.2 Results of the Yale dataset

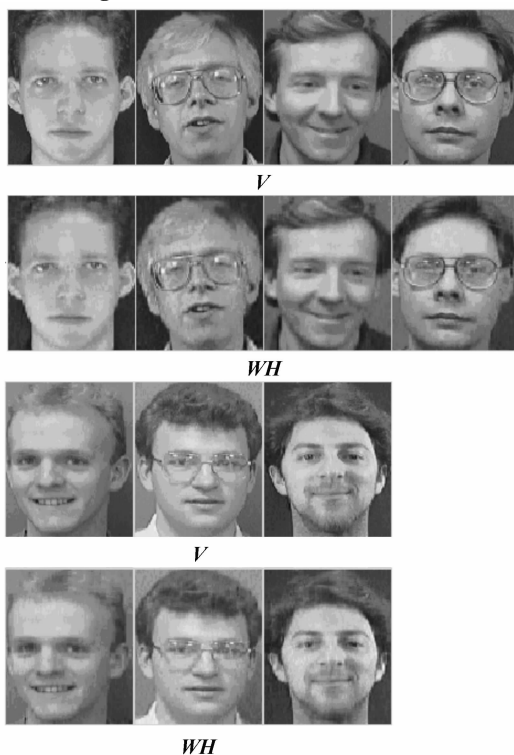


图 3 ORL 数据库实验结果

Fig.3 Results of the ORL dataset

$$SNR = 10 \lg \left[ \frac{\|\mathbf{V}\|^2}{\|\mathbf{WH} - \mathbf{V}\|^2} \right] \quad (11)$$

SNR 值越大代表分离效果越好。在实验中,选取  $\alpha_w$  为 1.5,  $\alpha_H$  为 0.2,  $r$  取 30, 最大迭代次数取 1 000。表 1 给出 3 个不同算法的 3 个性能指标,由表 1 可以看出,提出的算法获得了更好的稀疏性和更优的信噪比。

表1 不同算法的实验结果对比

Tab.1 Results comparison of different algorithms

算法	$\ W\ _2$	$H$ 稀疏度	信噪比
NMF	0.609 8	0.353 4	23.584 7
SNMF	0.493 2	0.715 6	25.352 6
提出的算法	0.383 1	0.833 1	31.490 7

## 4 结束语

提出基于梯度下降法的约束非负矩阵分解算法,迭代过程运用非负矩阵分解算法中较为常用的梯度下降法进行迭代,算法中对迭代过程添加对基矩阵向量2-范数约束以及对系数矩阵最小相关约束,从而得到提出的算法。虽然算法仍只能得到局部最优解,并没有解决全局最优解,但是通过实验结果表明,该算法在系数矩阵的稀疏度表达能力、实验结果误差上优于NMF算法和SNMF算法。

### 参考文献:

- [1] Li Changli, Cao Jiayi. Loop optimization based joint diagonalization algorithm and its application in blind source separation[J]. Journal of Sichuan University: Engineering Science Edition, 2011, 43(5): 159 - 163. [李昌利, 曹嘉毅. 基于循环优化的矩阵联合对角化算法及在盲源分离中的应用[J]. 四川大学学报: 工程科学版, 2011, 43(5): 159 - 163.]
- [2] Lee D D, Seung H S. Learning the parts of objects by non-negative matrix factorization[J]. Nature, 1999, 401(6755): 788 - 791.
- [3] Lee D D, Seung H S. Algorithms for non-negative matrix factorization[C]//Proceedings of the Advances in Neural Information Processing Systems. Vancouver, 2001, 13: 556 - 562.
- [4] Hoyer P O. Non-negative matrix factorization with sparseness constraints [J]. Journal of Machine Learning Research, 2004, 5(9): 1457 - 1469.
- [5] Bofill P, Zibulevsky M. Under determined blind source separation using sparse representations [J]. Signal Process, 2001, 81(11): 2353 - 2362.
- [6] Li Le, Zhang Yujin. A survey on algorithms of non-negative matrix factorization [J]. Acta Electronica Sinica, 2008, 36

(4): 737 - 743. [李乐, 章毓晋. 非负矩阵分解算法综述[J]. 电子学报, 2008, 36(4): 737 - 743.]

- [7] Zhang Yufei. Nonnegative matrix factorization with sparseness constraint[D]. Dalian: Dalian University of Technology, 2010. [张宇飞. 加稀疏约束的非负矩阵分解[D]. 大连: 大连理工大学, 2010.]
- [8] Zhao Zhijin, Lu Hong, Xu Chunyun, et al. An blind source separation algorithm based on constrained NMF[J]. Piezoelectrics and Acousto-optics, 2010, 21(6): 1049 - 1052. [赵知劲, 卢宏, 徐春云, 等. 基于约束 NMF 的盲源分离算法[J]. 压电与声光, 2010, 21(6): 1049 - 1052.]
- [9] Ping Shasha, Chu Leilei. A sparse algorithm for non-negative matrix factorization based on project gradient method[J]. Computer & Digital Engineering, 2012, 40(12): 20 - 22. [平沙沙, 褚蕾蕾. 基于投影梯度法的非负矩阵分解稀疏算法[J]. 计算机与数字工程, 2012, 40(12): 20 - 22.]
- [10] Huang K, Sidiropoulos N D, Swami A. NMF revisited: New uniqueness results and algorithms [C]//Proceedings of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Vancouver: ICASSP, 2013: 4524 - 4528.
- [11] Lu Xiaoqiang, Wu Hao, Yuan Yuan. Double constrained NMF for hyperspectral unmixing[J]. IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(5): 2746 - 2758.
- [12] Zhang Fuzhen. A note on the best Euclidian fit to a given distance matrix [J]. Journal of Beijing Normal University: Natural Science, 1987(4): 21 - 24. [张福振. 关于距离矩阵的最佳欧氏逼近[J]. 北京师范大学学报: 自然科学版, 1987(4): 21 - 24.]
- [13] Georghiades A, Hespanha J. Yale face database[EB/OL]. [2014 - 06 - 01]. [http://vision.ucsd.edu/datasets/yale\\_face\\_dataset\\_original/yalefaces.zip](http://vision.ucsd.edu/datasets/yale_face_dataset_original/yalefaces.zip).
- [14] Laboratories AT&T, Cambridge UK. ORL face database [EB/OL]. [2014 - 06 - 01]. [http://www.cl.cam.ac.uk/Research/DTG/attarchive:pub/data/att\\_faces.zip](http://www.cl.cam.ac.uk/Research/DTG/attarchive:pub/data/att_faces.zip).