

基于线性投影结构的非负矩阵分解

李乐^{1,2,3} 章毓晋^{1,2}

摘要 非负矩阵分解 (Non-negative matrix factorization, NMF) 是一个近年来非常流行的非负数据处理方法, 它常用于维数约减、特征提取和数据挖掘等. NMF 定义中采用的数学模型基于非线性投影结构构造, 这决定了 NMF 降维需借助计算量很大的迭代操作来实现. 此外, 由此模型提取的 NMF 特征常不稀疏, 这与 NMF 的设计期望相差甚远. 为一并解决上述两个问题, 本文提出了一个新的模型 — 基于线性投影结构的 NMF (Linear projection-based NMF, LPBNMF), 并构造了一个单调的 LPBNMF 算法. 从数学的角度看, LPBNMF 可理解为实现 NMF 的一种特殊方式. LPBNMF 降维通过线性变换来完成, 它所采用的数学模型的自身结构特点决定了由其得到的特征一定非常稀疏. 大量的比较实验表明, LPBNMF 的降维效率显著高于 NMF, LPBNMF 特征明显比 NMF 特征更稀疏和局部化. 最后, 基于 AR 人脸数据库的实验揭示, LPBNMF 特征比 NMF、LDA 以及 PCA 等特征更适合于用最近邻分类法处理有遮挡人脸识别问题.

关键词 非负矩阵分解, 基于线性投影结构的非负矩阵分解, 特征提取, 数据描述, 降维效率, 稀疏特征, 有遮挡人脸识别

DOI 10.3724/SP.J.1004.2010.00023

Linear Projection-based Non-negative Matrix Factorization

LI Le^{1,2,3} ZHANG Yu-Jin^{1,2}

Abstract Non-negative matrix factorization (NMF) is a newly popular method for non-negative dimensionality reduction, feature extraction, data mining, etc. The mathematical model in NMF definition is based on nonlinear projection, therefore dimension reduction by NMF is implemented by iterative updates which lead to high computational load. Additionally, NMF features extracted by this model are usually not very sparse, and this fails to meet the expectation of designing NMF. To simultaneously resolve the above two problems, this paper proposes a new model, linear projection-based NMF (LPBNMF), and designs an monotonic algorithm for it. From mathematical point of view, LPBNMF is a special mode for implementing NMF, which linearly implements dimension reduction. The high sparseness of LPBNMF features is assured by the inherent characteristics of its mathematic model. The comparison experiments validate that dimension reduction by LPBNMF is much more efficient than that by NMF, and that LPBNMF features are much more sparse and localized than NMF ones. Finally, experiments based on AR face database indicate that LPBNMF features are more suitable for nearest neighbor classification-based occluded face recognition than NMF, LDA, and PCA ones.

Key words Non-negative matrix factorization (NMF), linear projection-based NMF (LPBNMF), feature extraction, data representation, efficiency of dimensionality reduction, sparse feature, occluded face recognition

非负矩阵分解 (Non-negative matrix factorization, NMF)^[1] 由 Lee 和 Seung 于 1999 年在 *Nature* 上提出, 它使分解后的所有分量均为非负值 (寻求纯加性的描述). NMF 的心理学和生理学构造依据是对整体的感知由对组成整体的部分的感知构成 (感知是纯加性的生理机能)^[2-4], 这也符合直观的理解: 整体是由部分组成的^[1], 因此它在某种

意义上抓住了智能化数据描述的本质.

Lee 和 Seung 提出 NMF 时, 在泊松产生模型的假设下用最大似然估计的思路构造了第一个 NMF 算法^[1]. 随后, Lee 和 Seung 采用类似于 EM (Expectation maximization) 算法中使用的优化策略对广义 Kullback-Leibler 散度和欧几里德距离的平方 (它们用来度量被处理数据和 NMF 结果间的差异, 本文后面提到的被优化函数也均作此用) 分别做交替优化得到了两个迄今最为经典和使用最广的单调算法^[5]. Wild 等考虑了利用球面 K 均值聚类作为上述基于欧几里德距离平方算法的初始化步骤, 这样做的好处是能使算法效率有所提高, 但这以收敛到相对不好的局部解为代价^[6]. Cichocki 等构造了利用指数梯度下降原则交替优化对偶的广义 Kullback-Leibler 散度的算法^[7]. Heiler 等考虑了欧几里德距离平方的展开形式, 把对 NMF 的优化求解归结为一组交替进行的经典凸二次规划问题, 构造了单调下降的算法^[8]. Cichocki 等利用了欧几里

收稿日期 2008-09-17 录用日期 2009-05-06
Manuscript received September 17, 2008; in revised form May 6, 2009

国家自然科学基金 (60872084) 资助
Supported by National Natural Science Foundation of China (60872084)

1. 清华大学清华信息科学与技术国家实验室 北京 100084 2. 清华大学电子工程系 北京 100084 3. 中国电子信息产业发展研究院 北京 100048

1. Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084 2. Department of Electronic Engineering, Tsinghua University, Beijing 100084 3. China Center for Information Industry Development, Beijing 100048

德距离平方的局部凸特性,构造了以“先求得非负约束下的解析解(定点算法的特点),再通过做非线性投影满足非负性要求”的方式完成交替优化的算法^[7],此算法速度较快但可能发生震荡.陈卫刚等构造了用可行方向法和模拟退火法结合来交替优化欧几里德距离平方的算法^[9],它(理论)复杂度较高,因为可行方向法中的一维搜索步骤以及模拟退火法都比较耗时.Li等构造了利用抛物线原理交替优化欧几里德距离平方的单调定点算法,此算法中的每次操作都使此刻对应优化问题的目标函数仅由这一次操作达到最小值,因此效率远远超过其他算法^[10].

以上的 NMF 算法均是基于单一目标函数构造的,还有另一类 NMF 算法,它们是基于一组目标函数(目标函数族)构造的,这包括 Compass 基于 β 散度的算法^[11]、Cichocki 和 Zdunek 等基于 α 散度的算法^[12-13]以及 Dhillon 等基于 Bregman 散度和对偶 Bregman 散度的算法^[14].这类算法的共同特点是:依目标函数族中的参数或自由设定函数的不同,它们与一系列潜在产生式模型(对应不同的概率分布假设)下的似然函数一一对应,因此它们的适用范围更广(特别是对于盲分离问题^[12]),但与此同时,这些参数和自由设定函数的选择也对使用者提出了较高的要求.

有关上述算法的更细致研究评述以及 NMF 算法研究的其他成果可参见综述文章《非负矩阵分解算法综述》^[15].

目前, NMF 已应用到文本分析^[1]与聚类^[16]、数字水印^[17-18]、人脸检测^[19]与识别^[20]、图像检索^[21-22]、图像复原^[23]、语言建模^[24-25]、声源分类^[26]、音乐信号分析^[27]与乐器识别^[28]、盲信号分离^[12]、网络安全^[29]、基因及细胞分析^[30-31]等方面的研究中.

NMF 定义中采用的数学模型基于非线性投影结构构造,这决定了 NMF 降维需借助计算量较大的迭代操作来实现,因此,其处理高维数据(常见于计算机视觉、图像工程和机器学习等研究中)时常遇到效率瓶颈.此外,由此模型提取的 NMF 特征常不稀疏,这与 NMF 的设计期望(提取稀疏和局部化的特征^[1])相差甚远.目前的研究结果显示, NMF 还只是一个数据依赖的稀疏特征提取方法,这限制了其应用范围,增加了应用难度.

为一并解决以上两个问题,本文提出了一个新的模型——基于线性投影结构的 NMF (Linear projection-based NMF, LPBNMF),构造了一个单调的 LPBNMF 算法.从数学的角度看, LPBNMF 是实现 NMF 的一种特殊方式.

本文以下各节安排为:第 1 节介绍 NMF,结合示例指出其存在的问题;第 2 节定义 LPBNMF,

分析 LPBNMF 的性质;第 3 节构造 LPBNMF 算法,证明此算法的稳定性和收敛性;第 4 节分析 LPBNMF 的一个相关(相关指两方法同以线性变换实现降维)方法——投影非负矩阵分解(Projective NMF, P-NMF),分析 LPBNMF 与 P-NMF 在构造目的、模型、算法以及算法特性上的本质区别,指出 LPBNMF 相比 P-NMF 的优越性;第 5 节用大量的比较实验验证 LPBNMF 比 NMF 降维效率高且特征稀疏,揭示基于 LPBNMF 的方法较常用方法更适于解决有遮挡人脸识别问题;第 6 节总结全文.

1 NMF 及其存在的问题

对一个 M 维的正随机向量 \mathbf{v} 进行 N 次观测,记这些观测为 $\mathbf{v}_j, j = 1, 2, \dots, N$, 取 $V = [\mathbf{V}_{\cdot 1}, \mathbf{V}_{\cdot 2}, \dots, \mathbf{V}_{\cdot N}]$, 其中, $\mathbf{V}_{\cdot j} = \mathbf{v}_j, j = 1, 2, \dots, N$.

定义 1. NMF 是通过解

$$\min_{0 \leq W, H} f_V(W, H) \quad (1)$$

($f_V(W, H)$ 刻画 V 与 WH 间差异性) 来求得非负的 $M \times L$ 维基矩阵 W 和非负的 $L \times N$ 维系数矩阵 H , 使^[1, 8]

$$V \approx WH \quad (2)$$

研究结果显示,所有可使用的 f_V 均不是 (W, H) 的凸函数,所以对于计算机视觉、模式识别以及图像工程等研究领域中所处理的大规模问题,求解 NMF 的全局最优解并不现实^[1, 5, 8],当前的全部算法均是通过交替地优化 W 和 H 得到相应 NMF 问题的一个局部解.

W 蕴含了 NMF 学习结果的全部内容,因为 W 定了之后依据 $f_V(W, H)$, H 也就定了.这里, H 和 W 间不是“ H 等于 V 向 W 做线性投影的结果”的关系,而是“ H 等于

$$\min_{0 \leq H} f_V(W, H) \quad (3)$$

的结果”的关系,这表明 NMF 定义中采用的模型是基于非线性投影结构构造的.

式(3)也蕴含了在 NMF 训练完成后(即得到式(3)的解 W 后)进行 NMF 降维的优化模型

$$\min_{0 \leq \mathbf{h}_j} f_{\mathbf{v}_j}(W, \mathbf{h}_j), j > N \quad (4)$$

其中, $\mathbf{v}_j, j > N$ 是随机向量 \mathbf{v} 的一个新观测(非训练用观测), \mathbf{h}_j 是它的 NMF 投影系数(降维结果).无论 $f_{\mathbf{v}}$ 为何种距离度量或散度,解式(4)均需要用迭代的方式来完成,所以 NMF 降维计算量很大,这使其处理高维数据时常面临效率瓶颈.举例来说,取 $f_{\mathbf{v}}$ 为欧几里德距离的平方,此时 NMF 优化问题为

$$\min_{0 \leq W, H} \|V - WH\|_F^2 \quad (5)$$

Lee 和 Seung 基于辅助函数法提出可通过交替地执行

$$H \leftarrow H \otimes (W^T V) \oslash (W^T W H) \quad (6)$$

$$W \leftarrow W \otimes (V H^T) \oslash (W H H^T) \quad (7)$$

(\otimes 和 \oslash 分别代表 Hardmard 乘和除) 直到收敛来解式 (5)^[5] (这个算法非常常用, 实现简单, 效率尚可, 已被公认为基准算法, 本文的全部 NMF 实验均借助此算法完成, 本文要说明的 NMF 缺陷是 NMF 模型造成的, 与某个具体算法无关). 由此算法学习出 W 后 (执行式 (6) 是为了辅助式 (7) 求 W), 对于随机向量 \mathbf{v} 的一个新观测 $\mathbf{v}_j, j > N$ 的 NMF 降维是解

$$\min_{0 \leq \mathbf{h}_j} \|\mathbf{v}_j - W \mathbf{h}_j\|_F^2 \quad (8)$$

这通过执行

$$\mathbf{h}_j \leftarrow \mathbf{h}_j \otimes (W^T \mathbf{v}_j) \oslash (W^T W \mathbf{h}_j) \quad (9)$$

直到收敛来实现 (式 (9) 等价于式 (6), 是式 (6) 的向量表达形式). 用 ORL 库^[32] 中的 200 幅图像 (任取每人的 5 幅图) 进行 NMF 训练, 表 1 记录了利用这时的 NMF 训练结果把 ORL 库里某幅非训练用图像 (原始维数是 10 324) 降到 5 个不同维度的计算耗时 (降维耗时基于配置为 Pentium 3.06 G CPU, 1.5 G 内存的 PC 记录). 用 NMF 做应用研究时, 一次实验中常需对成百上千幅这样的图像完成降维, 表 1 揭示的降维效率显然无法令人满意.



(a) 基于 ORL 库的全部 400 幅图

(a) Based on all 400 images of ORL database



(b) 基于 AR 库中 118 人的每人 8 幅无遮挡和无表情变化图


(b) Based on 118 objects' 944 images (8 images of each object) without occlusions and expressional variations in AR database

图 1 NMF 特征 ($L = 100$)

Fig. 1 NMF features ($L = 100$)

表 1 NMF 降维耗时示例

Table 1 Exhibitions for computational load of dimensionality reduction by NMF

被处理 图像	降到 k 维时的计算耗时 (s)				
	$k = 40$	$k = 80$	$k = 120$	$k = 160$	$k = 160$
	2.7124	12.6781	18.5714	29.3181	49.3811

Lee 和 Seung 提出 NMF 时, 利用基于 MIT 人脸数据库的实验结果说明通过 NMF 可以学习出稀疏和局部化的特征, 但他们未在理论上对此予以证明^[1]. Donoho 等定义了可分离因子联结族 (Separable factorial articulation family, SFAF), 证明了如被处理数据集是一个 SFAF, 对其进行 NMF 可以得到稀疏和局部化的特征^[33]. 除此之外, 目前尚无其他理论研究结论. 但 SFAF 是非常理想化的, 实际数据集几乎不可能满足其定义. 因此, 推断用 NMF 处理实际数据集所得特征可能不稀疏, 这被图 1 所示基于 ORL 和 AR 数据库^[34] 的实验结果所支持. 所以, NMF 还只是一种数据依赖的稀疏特征提取方法, 这制约了其应用范围, 增加了应用它的难度.

2 LPBNMF 及其性质

为一并解决以上两个问题, 本节提出一个新的模型 — 基于线性投影结构的 NMF (Linear projection-based NMF, LPBNMF).

定义 2. LPBNMF 通过解

$$\min_{0 \leq Q, W} \|V - WQV\|_F^2 \quad (10)$$

来求得非负的 $M \times L$ 维基矩阵 W 和非负 $L \times M$

维变换矩阵 Q , 使

$$V \approx WQV \quad (11)$$

从数学的角度看, LPBNMF 是进行 NMF 的一种特殊方式, LPBNMF 定义与取 f_V 为欧几里德距离平方且取

$$H = QV \quad (12)$$

时的 NMF 定义完全一致.

对于正随机向量 \mathbf{v} 的一个观测 \mathbf{v}_j 的 LPBNMF 降维通过

$$\mathbf{h}_j = Q\mathbf{v}_j \quad (13)$$

完成, 所以其效率要远高于 NMF 降维.

具体说, LPBNMF 降维的计算量为 ML 次乘运算和 $L(M-1)$ 次加运算, 而式 (9) 表示的 NMF 降维计算量 (其中, 计量 $W^T W \mathbf{h}_i$ 的计算量时, 计量的是先算 $W \mathbf{h}_i$ 、再算 $W^T (W \mathbf{h}_i)$ 的计算量, 这样做比按顺序计算 $W^T W \mathbf{h}_i$ 的运算量低) 为 $k(3M+1)L$ 次乘运算、 $k(3ML-2L-M)$ 次加运算以及 KL 次除运算 (k 是迭代的步数). 显然, 式 (9) 表示的 NMF 降维中, 一步迭代的计算量已远超过 LPBNMF 降维的全部计算量.

通过交替优化 W 和 Q 解式 (10) 时, 当固定 W 求 Q , Q 有趋向于其无约束最优解 $(W^T W)^{-1} W^T$ 的趋势, 这是因为: $\|V - WQV\|_F^2$ 是 Q 的凸函数 (之后的引理 4 对此有证明), Q 越趋向于 $(W^T W)^{-1} W^T$, $\|V - WQV\|_F^2$ 越小. 同理, 当固定 Q 求 W , W 同样也有趋向于其无约束最优解 $Q^T (QQ^T)^{-1}$ 的趋势. 因为 $(W^T W)^{-1} W^T$ 和 $Q^T (QQ^T)^{-1}$ 在一般情况下并不是非负的, $\forall W > 0, \forall Q > 0$, 只有 W 和 Q 正交, 才能保证 $(W^T W)^{-1} W^T$ 和 $Q^T (QQ^T)^{-1}$ 非负. 也就是说, 在非负的约束下, W 和 Q 越趋向于正交, 它们越可以优化到 $Q^T (QQ^T)^{-1}$ 和 $(W^T W)^{-1} W^T$. 那么, 在对式 (10) 中问题的交替优化求解过程中, W 和 Q 总是呈现出正交化的趋势, 从而使 Q 趋向于 $(W^T W)^{-1} W^T$, 使 W 趋向于 $Q^T (QQ^T)^{-1}$, 最终使 $\|V - WQV\|_F^2$ 尽可能小. 事实上可以认为, 式 (10) 的优化模型中有一潜在罚项来约束 W 和 Q 趋向正交. 正因为此, 用以往任何 NMF 算法几乎不可能得到 LPBNMF 解.

在非负的约束下, 正交即意味着稀疏. 因为对于一组非负正交向量, 每一元素位置上只能有一个向量的元素为非零值, 例如, 假定 A 是一个 10×5 的非负正交阵, 则 A 的一个可能形式是

$$A = \begin{bmatrix} a_1 & a_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_3 & a_4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a_5 & a_6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & a_7 & a_8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_9 & a_{10} \end{bmatrix}^T \quad (14)$$

式中, $a_1, a_2, \dots, a_{10} \in \mathbf{R}_+$.

W 的列代表 NMF 特征, 所以 LPBNMF 特征一定是非常稀疏的, 且 L 越大, W 列的稀疏性越强 (L 越大, 每列中零元素越多).

由 LPBNMF 的目标函数 (式 (10)) 知, LPBNMF 是追求表示性的. 由 $\mathbf{v}_j \approx \sum_{i=1}^L \mathbf{W}_{\cdot i} [Q\mathbf{v}_j]_i$ 以及 W 列的稀疏性和趋于正交性知, $\forall i \in \{1, 2, \dots, L\}$, \mathbf{v}_j 中与 $\mathbf{W}_{\cdot i}$ 非零元素对应的位置基本只由 $\mathbf{W}_{\cdot i}$ 和 $[Q\mathbf{v}_j]_i$ 来表示. 那么, LPBNMF 特征 ($\mathbf{W}_{\cdot i}, \forall i \in \{1, 2, \dots, L\}$) 中非零元素应主要分布于被处理数据 \mathbf{v}_j 的某一局部区域 (局部区域的元素间有较强的相关性, 变化相对会较平滑) 上. 否则, 很难做到 $\forall i \in \{1, 2, \dots, L\}, \forall j \in 1, 2, \dots, N, \mathbf{W}_{\cdot i} [Q\mathbf{v}_j]_i$ 中的非零元素能很好地逼近 \mathbf{v}_j 中对应位置的元素. 举例来说, 如用 LPBNMF 去描述人脸, 则 LPBNMF 特征应是分别刻画如眼、鼻、嘴、面部、额头等局部区域的, 假若 LPBNMF 特征 $\mathbf{W}_{\cdot i}$ 同时刻画了眼睛和面部, 则无论如何, 很难做到 $\forall j \in 1, 2, \dots, N, \mathbf{W}_{\cdot i} [Q\mathbf{v}_j]_i$ 都能很好地描述被处理人脸图像的眼睛和面部. 因为不同人脸图像这两部分间的差异程度不一样 (比如说, 不同人的肤色不同, 面部的反光度不同, 所以眼睛和面部的灰度差异就不同). 综上所述, LPBNMF 特征不但应是稀疏的, 而且应是局部化的.

3 LPBNMF 算法

基于式 (10) 的优化模型可构造如下 2 个定理, 它们是构造 LPBNMF 算法的基础.

定理 1. $\forall i \in \{1, 2, \dots, L\}$, 记

$$E_{ab} = [QV V^T Q^T]_{ab} (1 - \delta_{ai} \delta_{ib}) \quad \forall a, b \in \{1, 2, \dots, L\} \quad (15)$$

固定 Q 和 $\{\mathbf{W}_{\cdot 1}, \mathbf{W}_{\cdot 2}, \dots, \mathbf{W}_{\cdot L}\} / \mathbf{W}_{\cdot i}$, 并保证 $\|[QV]_{\cdot i}\|_1 > 0$, 在 $\mathbf{W}_{\cdot i} \geq \epsilon$ (ϵ 是一个很小的正数) 的约束下, 式 (10) 中的目标函数在

$$\mathbf{W}_{\cdot i} = \max \left\{ \frac{[V V^T Q^T]_{\cdot i} - W \mathbf{E}_{\cdot i}}{[Q V V^T Q^T]_{ii}}, \epsilon \right\} \quad (16)$$

($\max\{\cdot, \epsilon\}$ 使向量或矩阵中小于 ϵ 的元素等于 ϵ , 大于 ϵ 的元素保持不变) 处达到最小值.

此定理的证明可参见文献 [10] 中对定理 1 的证明 (只需用 QV 替代其中的 H 即可).

定义 3. 如果 $G(\mathbf{h}, \mathbf{g}) \geq F(\mathbf{h})$, 当且仅当 $\mathbf{g} = \mathbf{h}$ 时有 $G(\mathbf{g}, \mathbf{h}) = F(\mathbf{h})$, 则 $G(\mathbf{g}, \mathbf{h})$ 称为 $F(\mathbf{h})$ 的辅助函数^[5].

引理 1. 如果 $G(\mathbf{g}, \mathbf{h})$ 为 $F(\mathbf{h})$ 的辅助函数, 且 \mathbf{g} 固定后, $G(\mathbf{g}, \mathbf{h})$ 是 \mathbf{h} 的严格凸函数, 则依照

$$\mathbf{h}^{t+1} = \arg \min_{\mathbf{h}} G(\mathbf{h}, \mathbf{h}^t) \quad (17)$$

更新 \mathbf{h} 可使 $F(\mathbf{h})$ 严格单调下降 (\mathbf{h}^t 和 \mathbf{h}^{t+1} 分别为 \mathbf{h} 的当前值和更新后的值).

证明. $F(\mathbf{h}^{t+1}) \leq G(\mathbf{h}^{t+1}, \mathbf{h}^t) < G(\mathbf{h}^t, \mathbf{h}^t) = F(\mathbf{h}^t)$. \square

定理 2. 当 $V > 0$, 固定 $W > 0$, 如果 Q 的当前值 $\geq \epsilon$, 则在 $Q \geq \epsilon$ 的约束下, 通过

$$Q \leftarrow \max\{Q \otimes \sqrt{W^T V V^T \circledast W^T W Q V V^T}, \epsilon\} \quad (18)$$

可使式 (10) 中的目标函数值下降.

证明. 因为

$$\begin{aligned} & \|V - WQV\|_F^2 = \\ & \text{tr}\{(V - WQV)(V - WQV)^T\} = \\ & \text{tr}\{VV^T - 2WQVV^T + WQVV^T Q^T W^T\} = \\ & \text{tr}\{VV^T - 2Q^T W^T V V^T + W^T W Q V V^T Q^T\} \end{aligned} \quad (19)$$

固定 W , Q 的目标函数可简化为

$$\begin{aligned} F(Q) = & \\ & \text{tr}\{-2Q^T W^T V V^T + W^T W Q V V^T Q^T\} = \\ & -2 \sum_{i=1}^L \sum_{j=1}^M Q_{ij} [W^T V V^T]_{ij} + \\ & \sum_{i=1}^L \sum_{j=1}^M Q_{ij} \sum_{k=1}^L \sum_{l=1}^M [W^T W]_{ik} Q_{kl} [V V^T]_{lj} \end{aligned} \quad (20)$$

仿照文献 [35] 中的设计思路, 构造

$$\begin{aligned} Z(Q, P) = & \sum_{i=1}^L \sum_{j=1}^M G(Q_{ij}) = \\ & -2 \sum_{i=1}^L \sum_{j=1}^M [W^T V V^T]_{ij} P_{ij} (1 + \log \frac{Q_{ij}}{P_{ij}}) + \\ & \sum_{i=1}^L \sum_{j=1}^M \frac{Q_{ij}^2}{P_{ij}} \sum_{k=1}^L \sum_{l=1}^M [W^T W]_{ik} P_{kl} [V V^T]_{lj} \end{aligned} \quad (21)$$

显然, $Z(Q, Q) = F(Q)$. $\forall i, j$, 取 $x_{ij} = Q_{ij}/P_{ij}$, 根据不等式 $x \geq 1 + \log(x)$ ^[36], 有

$$Q_{ij} \geq P_{ij} (1 + \log \frac{Q_{ij}}{P_{ij}}) \quad (22)$$

此外, 按文献 [35] 中的分析, 当 $P > 0$ 时, 有

$$\begin{aligned} & \sum_{i=1}^L \sum_{j=1}^M \frac{Q_{ij}^2}{P_{ij}} \sum_{k=1}^L \sum_{l=1}^M [W^T W]_{ik} P_{kl} [V V^T]_{lj} - \\ & \sum_{i=1}^L \sum_{j=1}^M Q_{ij} \sum_{k=1}^L \sum_{l=1}^M [W^T W]_{ik} Q_{kl} [V V^T]_{lj} = \\ & \frac{1}{2} \sum_{i=1}^L \sum_{j=1}^M \frac{Q_{ij}^2}{P_{ij}} \sum_{k=1}^L \sum_{l=1}^M [W^T W]_{ik} P_{kl} [V V^T]_{lj} + \\ & \frac{1}{2} \sum_{k=1}^L \sum_{l=1}^M \frac{Q_{kl}^2}{P_{kl}} \sum_{k=i}^L \sum_{j=1}^M [W^T W]_{ki} P_{ij} [V V^T]_{jl} - \\ & \sum_{i=1}^L \sum_{j=1}^M Q_{ij} \sum_{k=1}^L \sum_{l=1}^M [W^T W]_{ik} Q_{kl} [V V^T]_{lj} = \\ & \sum_{i=1}^L \sum_{j=1}^M \sum_{k=1}^L \sum_{l=1}^M \frac{1}{2} x_{ij}^2 P_{ij} [W^T W]_{ik} P_{kl} [V V^T]_{lj} + \\ & \sum_{i=1}^L \sum_{j=1}^M \sum_{k=1}^L \sum_{l=1}^M \frac{1}{2} x_{kl}^2 P_{kl} [W^T W]_{ki} P_{ij} [V V^T]_{jl} - \\ & \sum_{i=1}^L \sum_{j=1}^M \sum_{k=1}^L \sum_{l=1}^M x_{ij} x_{kl} P_{ij} [W^T W]_{ik} P_{kl} [V V^T]_{lj} = \\ & \sum_{i=1}^L \sum_{j=1}^M \sum_{k=1}^L \sum_{l=1}^M P_{ij} [W^T W]_{ik} P_{kl} [V V^T]_{lj} \times \\ & \left(\frac{1}{2} x_{ij}^2 + \frac{1}{2} x_{kl}^2 - x_{ij} x_{kl} \right) \geq 0 \end{aligned} \quad (23)$$

所以, $Z(Q, P) \geq F(Q)$. 那么, $Z(Q, P)$ 是 $F(Q)$ 的辅助函数.

固定 $P > 0$, 因为 $V > 0$ 且 $W > 0$, $Z(Q, P)$ 是 Q 的严格凸函数. 因为 $G(Q_{11}), G(Q_{21}), \dots, G(Q_{L1}), G(Q_{12}), G(Q_{22}), \dots, G(Q_{L2}), \dots, G(Q_{1M}), G(Q_{2M}), \dots, G(Q_{LM})$ 间无关联, $Z(Q, P)$ 在满足

$$\left[\frac{\partial Z(Q, P)}{\partial Q} \right]_{ij} = \left[\frac{\partial G(Q_{ij})}{\partial Q_{ij}} \right] =$$

$$-2[W^T V V^T]_{ij} \frac{P_{ij}}{Q_{ij}} + 2 \frac{Q_{ij}}{P_{ij}} [W^T W P V V^T]_{ij} = 0$$

$$\forall i \in \{1, 2, \dots, L\}, j \in \{1, 2, \dots, M\} \quad (24)$$

的点处达到无约束下最小值, 易知此点为

$$P \otimes \sqrt{W^T V V^T \otimes W^T W P V V^T} \quad (25)$$

$\forall i \in \{1, 2, \dots, L\}, j \in \{1, 2, \dots, M\}$, $G(Q_{ij})$ 在 $Q_{ij} > 0$ 时是单调增函数. 所以在 $Q \geq \epsilon$ 的约束下, $Z(Q, P)$ 在

$$\max\{P \otimes \sqrt{W^T V V^T \otimes W^T W P V V^T}, \epsilon\} \quad (26)$$

处达到最小值. 取 P 为 Q 的当前值, 则式 (26) 就是式 (18) 的右半部分. 根据引理 1 知, 依据式 (18) 更新 Q , 可使 (10) 式中的目标函数值下降. \square

给 W 和 Q 均取定大于 ϵ 的初值, 对 $\mathbf{W}_{\cdot 1}, \mathbf{W}_{\cdot 2}, \dots, \mathbf{W}_{\cdot L}$ 和 Q 分别根据定理 1 和定理 2 做交替优化, 可使式 (10) 中的目标函数值严格单调下降. 上述操作可表述为如下所示的 LPBNMF 算法.

LPBNMF 算法

输入. $V > 0, L$

输出. W, Q

步骤 1. 随机初始化 $W(0) \geq \epsilon$ 和 $Q(0) \geq \epsilon, k \leftarrow 0$

步骤 2. 重复如下操作直到 $(\|V - W(k)Q(k)V\|_F - \|V - W(k+1)Q(k+1)V\|_F) / \|V\|_F < \text{终止条件}$

步骤 3.

$$D(k) = Q(k) V V^T Q(k)^T, F(k) = V [V^T Q(k)^T]$$

步骤 4. i 从 1 到 L 执行

步骤 5. $E_{ab}(k) = D_{ab}(1 - \delta_{ai}\delta_{ib}), \forall a, b \in \{1, 2, \dots, L\}$

步骤 6. $\mathbf{W}_{\cdot i}(k+1) \leftarrow \max\{\{\mathbf{F}_{\cdot i} - [\mathbf{W}_{\cdot 1}(k+1), \dots, \mathbf{W}_{\cdot i-1}(k+1), \mathbf{W}_{\cdot i}(k), \dots, \mathbf{W}_{\cdot L}(k)] \mathbf{E}_{\cdot i}(k)\} / D_{ii}(k), \epsilon\}$

步骤 7.

$$Q(k+1) \leftarrow \max\{Q(k) \otimes$$

$$\sqrt{W(k+1)^T V V^T \otimes W(k+1)^T W(k+1) Q(k) V V^T}, \epsilon\}$$

步骤 8. $k \leftarrow k+1$.

值得指出, 定理 1 和定理 2 中分别取约束条件 $W \geq \epsilon$ 和 $Q \geq \epsilon$, 而不是 $W \geq 0$ 和 $Q \geq 0$, 不但避免了在 LPBNMF 算法的步骤 6 和步骤 7 中出现数学上无定义情况 (除以“0”) 的可能性, 而且几乎不影响算法的下降速度 (因为 ϵ 取一个足够小的正数).

为证明 LPBNMF 算法的稳定性和收敛性, 需如下 4 个引理.

引理 2. $\forall A \in \mathbf{R}^{M \times L}, \|A\|_{\text{Abs-sum}} = \sum_{i=1}^M \sum_{j=1}^L |A_{ij}|$ 是 A 的范数.

证明. $\forall A \in \mathbf{R}^{M \times L}, \forall C \in \mathbf{R}^{L \times N}$

$$\|AC\|_{\text{Abs-sum}} =$$

$$\left\| \left[\sum_{i=1}^L \mathbf{A}_{\cdot i} C_{i1}, \sum_{i=1}^L \mathbf{A}_{\cdot i} C_{i2}, \dots, \sum_{i=1}^L \mathbf{A}_{\cdot i} C_{iN} \right] \right\|_{\text{Abs-sum}}$$

$$= \sum_{j=1}^N \left\| \sum_{i=1}^L \mathbf{A}_{\cdot i} C_{ij} \right\|_1 \leq \sum_{j=1}^N \sum_{i=1}^L \|\mathbf{A}_{\cdot i}\|_1 |C_{ij}| =$$

$$\sum_{i=1}^L \|\mathbf{A}_{\cdot i}\|_1 \|\mathbf{C}_{\cdot i}\|_1 \leq \sum_{i=1}^L \|\mathbf{A}_{\cdot i}\|_1 \sum_{i=1}^L \|\mathbf{C}_{\cdot i}\|_1 =$$

(根据 Holder 不等式可得)

$$\|A\|_{\text{Abs-sum}} \|C\|_{\text{Abs-sum}} \quad (\text{相合性}) \quad (27)$$

$\forall A, B \in \mathbf{R}^{M \times L}, \forall \alpha \in \mathbf{R}, \|A\|_{\text{Abs-sum}} \geq 0$ (非负性), $\|\alpha A\|_{\text{Abs-sum}} = |\alpha| \|A\|_{\text{Abs-sum}}$ (齐次性) 和 $\|A + B\|_{\text{Abs-sum}} \leq \|A\|_{\text{Abs-sum}} + \|B\|_{\text{Abs-sum}}$ (三角不等式) 是显然成立的. \square

引理 3. 依 LPBNMF 算法规则, 序列 $\{(W(k), Q(k))\}$ 中存在收敛子列.

证明. $\|W(k)Q(k)V\|_{\text{Abs-sum}} = \sum_{i=1}^L \|\mathbf{W}_{\cdot i}(k)\|_1 \times \|[Q(k)V]_{\cdot i}\|_1 \geq M\epsilon \|Q(k)V\|_{\text{Abs-sum}}$. 如果 $\{Q(k)\}$ 无界, 因 $V > 0$, 则 $\{\|W(k)Q(k) \times V\|_{\text{Abs-sum}}\}$ 无界. 因为 $\|W(k)Q(k)V\|_{\text{Abs-sum}} - \|V\|_{\text{Abs-sum}} \leq \|V - W(k)Q(k)V\|_{\text{Abs-sum}}$, 所以 $\{\|V - W(k)Q(k)V\|_{\text{Abs-sum}}\}$ 无界. 依 LPBNMF 算法规则, $\{\|V - W(k)Q(k)V\|_F\}$ 单调下降且有下界, 所以它有界. 根据引理 2, $\|\cdot\|_{\text{Abs-sum}}$ 是范数, 在有限维空间中, 它与 $\|\cdot\|_F$ 等价, 所以 $\|V - W(k)Q(k)V\|_{\text{Abs-sum}}$ 有界, 这与假设 $\{Q(k)\}$ 无界导出的结论相悖, 所以 $\{Q(k)\}$ 有界.

$\|W(k_l)Q(k_l)V\|_{\text{Abs-sum}} \geq \|W(k_l)\|_{\text{Abs-sum}} N\epsilon$, 同理, $\{W(k)\}$ 有界. 所以, $\{(W(k), Q(k))\}$ 中存在收敛子列. \square

需指出, LPBNMF 算法中规定 $Q \geq \epsilon$, 而不是 $Q \geq 0$ (定理 2 中取约束条件 $Q \geq \epsilon$, 而不是 $Q \geq 0$), 与 $W \geq \epsilon$ 约束一起为 $\{(W(k), Q(k))\}$ 有界性的证明提供了便利. 这不但几乎不影响算法的收敛速度 (因 ϵ 设定为一个足够小的正数), 而且为从数学上保证算法的稳定性 (定理 3) 创造了基础.

引理 4. $\|V - WQV\|_F^2$ 是 (W, Q) 的连续函数.

证明.

$$\lim_{\Delta W \rightarrow 0, \Delta Q \rightarrow 0} \|V - (W + \Delta W)(Q + \Delta Q)V\|_F^2 -$$

$$\begin{aligned}
& \|V - WQV\|_F^2 = \\
& \lim_{\Delta W \rightarrow 0, \Delta Q \rightarrow 0} \text{tr}\{[V - (W + \Delta W)(Q + \Delta Q)V] \times \\
& [V - (W + \Delta W)(Q + \Delta Q)V]^T - \\
& (V - WQV)(V - WQV)^T\} = \\
& \lim_{\Delta W \rightarrow 0, \Delta Q \rightarrow 0} \text{tr}\{2(-V + WQV) \times \\
& (\Delta WQV + W\Delta QV + \Delta W\Delta QV)^T + \\
& (\Delta WQV + W\Delta QV + \Delta W\Delta QV) \times \\
& (\Delta WQV + W\Delta QV + \Delta W\Delta QV)^T\} = 0 \quad (28)
\end{aligned}$$

依多元函数连续性定义^[37], 命题得证. \square

引理 5. 取 $V > 0$, $\forall i \in \{1, 2, \dots, L\}$, 当固定 $Q > 0$ 和 $\{\mathbf{W}_{\cdot 1}, \mathbf{W}_{\cdot 2}, \dots, \mathbf{W}_{\cdot L}\}/\mathbf{W}_{\cdot i}$, $\|V - WQV\|_F^2$ 是 $\mathbf{W}_{\cdot i}$ 的严格凸函数; $\forall j \in \{1, 2, \dots, L\}$, 当固定 $W > 0$ 和 $\{\mathbf{Q}_{\cdot 1}, \mathbf{Q}_{\cdot 2}, \dots, \mathbf{Q}_{\cdot L}\}/\mathbf{Q}_{\cdot j}$, W 列满秩, $\|V - WQV\|_F^2$ 是 $\mathbf{Q}_{\cdot j}$ 的严格凸函数.

证明. 因为

$$\begin{aligned}
& \|V - WQV\|_F^2 = \\
& \text{tr}\{VV^T - 2WQVV^T + WQVV^TQ^TW^T\} = \\
& \text{tr}\{VV^T\} - 2 \sum_{d=1}^M \mathbf{w}_d [QVV^T]_{\cdot d} + \\
& \sum_{d=1}^M \mathbf{w}_d QVV^TQ^T\mathbf{w}_d^T = \\
& \text{tr}\{VV^T - 2 \sum_{d=1}^M \sum_{h=1}^L W_{dh} [QVV^T]_{hd} + \\
& \sum_{d=1}^M \sum_{k=1}^L \sum_{l=1}^L W_{dk} [QVV^TQ^T]_{kl} W_{dl}\} \quad (29)
\end{aligned}$$

$\forall i \in \{1, 2, \dots, L\}$, 固定 $Q > 0$ 和 $\{\mathbf{W}_{\cdot 1}, \mathbf{W}_{\cdot 2}, \dots, \mathbf{W}_{\cdot L}\}/\mathbf{W}_{\cdot i}$, $\mathbf{W}_{\cdot i}$ 的目标函数简化为

$$\begin{aligned}
f(\mathbf{W}_{\cdot i}) &= \mathbf{W}_{\cdot i}^T \{[QVV^TQ^T]_{ii} I\} \mathbf{W}_{\cdot i} + 2 \sum_{d=1}^M W_{di} \times \\
& \left(\sum_{k=1, k \neq i}^L W_{dk} [QVV^TQ^T]_{ki} - [QVV^T]_{id} \right) \quad (30)
\end{aligned}$$

因为 $V > 0$ 且 $Q > 0$, $[QVV^TQ^T]_{ii} I$ 是正定的. $f(\mathbf{W}_{\cdot i})$ 是二次函数, $[QVV^TQ^T]_{ii} I$ 是 $f(\mathbf{W}_{\cdot i})$ 的

Hesse 矩阵, 根据文献 [38] 中定理 1.4.16, $f(\mathbf{W}_{\cdot i})$ 是 $\mathbf{W}_{\cdot i}$ 的严格凸函数.

由式 (19) 继续推导, 有

$$\begin{aligned}
& \|V - WQV\|_F^2 = \\
& \text{tr}\{VV^T - 2Q^TW^TVV^T + VV^TQ^TW^TWQ\} = \\
& \text{tr}\{VV^T\} - 2\text{tr}\{VV^TWQ\} + \\
& \text{tr} \left\{ \begin{bmatrix} [VV^T]_{11} & [VV^T]_{12} & \cdots & [VV^T]_{1M} \\ [VV^T]_{21} & [VV^T]_{22} & \cdots & [VV^T]_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ [VV^T]_{M1} & [VV^T]_{M2} & \cdots & [VV^T]_{MM} \end{bmatrix} \right. \\
& \left. \begin{bmatrix} \mathbf{Q}_{\cdot 1}^T W^T W \mathbf{Q}_{\cdot 1} & \mathbf{Q}_{\cdot 1}^T W^T W \mathbf{Q}_{\cdot 2} & \cdots & \mathbf{Q}_{\cdot 1}^T W^T W \mathbf{Q}_{\cdot M} \\ \mathbf{Q}_{\cdot 2}^T W^T W \mathbf{Q}_{\cdot 1} & \mathbf{Q}_{\cdot 2}^T W^T W \mathbf{Q}_{\cdot 2} & \cdots & \mathbf{Q}_{\cdot 2}^T W^T W \mathbf{Q}_{\cdot M} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Q}_{\cdot M}^T W^T W \mathbf{Q}_{\cdot 1} & \mathbf{Q}_{\cdot M}^T W^T W \mathbf{Q}_{\cdot 2} & \cdots & \mathbf{Q}_{\cdot M}^T W^T W \mathbf{Q}_{\cdot M} \end{bmatrix} \right\} \\
& = \text{tr}\{VV^T\} - 2 \sum_{d=1}^M [VV^TW]_{d \cdot} \mathbf{Q}_{\cdot d} + \\
& \sum_{c=1}^M \sum_{d=1}^M [VV^T]_{cd} \mathbf{Q}_{\cdot d}^T W^T W \mathbf{Q}_{\cdot c}, \quad \forall j \in \{1, 2, \dots, L\} \quad (31)
\end{aligned}$$

固定 W 和 $\{\mathbf{Q}_{\cdot 1}, \mathbf{Q}_{\cdot 2}, \dots, \mathbf{Q}_{\cdot L}\}/\mathbf{Q}_{\cdot j}$, $\mathbf{Q}_{\cdot j}$ 的目标函数简化为

$$\begin{aligned}
f(\mathbf{Q}_{\cdot j}) &= [VV^T]_{jj} \mathbf{Q}_{\cdot j}^T W^T W \mathbf{Q}_{\cdot j} - 2([VV^TW]_{j \cdot} - \\
& \sum_{d=1, d \neq j}^M [VV^T]_{jd} \mathbf{Q}_{\cdot d}^T W^T W) \mathbf{Q}_{\cdot j} \quad (32)
\end{aligned}$$

因为 $V > 0$ 且 $W > 0$ 列满秩, $[VV^T]_{jj} W^T W$ 是正定的. $f(\mathbf{Q}_{\cdot j})$ 是二次函数, $[VV^T]_{jj} W^T W$ 是 $f(\mathbf{Q}_{\cdot j})$ 的 Hesse 矩阵, 根据文献 [38] 中定理 1.4.16, $f(\mathbf{Q}_{\cdot j})$ 是 $\mathbf{Q}_{\cdot j}$ 的严格凸函数. \square

定理 3 (LPBNMF 算法的稳定性定理). 依据 LPBNMF 算法规则, $\forall b \in \mathbf{Z}^+$, $\|(W(k+b), Q(k+b)) - (W(k), Q(k))\| \rightarrow 0$ (\mathbf{Z}^+ 为正整数集).

证明. 假设 $\|(W(k+1), Q(k+1)) - (W(k), Q(k))\| \rightarrow 0$ 不成立, 则存在 $\{(W(k), Q(k))\}$ 的子列 $\{(W(k_l), Q(k_l))\}$, 使 $\|(W(k_l+1), Q(k_l+1)) - (W(k_l), Q(k_l))\| > \delta > 0$. 根据引理 3, 存在 $\{(W(k_l), Q(k_l))\}$ 的收敛子列 $\{(W(k_{l_h}), Q(k_{l_h}))\}$. 显然, $\|(W(k_{l_h}+1), Q(k_{l_h}+1)) - (W(k_{l_h}), Q(k_{l_h}))\|$

$> \delta > 0$.

记 $\{(W(k_{l_h}), Q(k_{l_h}))\} \rightarrow (\bar{W}, \bar{Q})$, 根据引理 4, $\|V - WQV\|_F^2$ 是 (W, Q) 的连续函数, $\{\|V - W(k_{l_h})Q(k_{l_h})V\|_F^2\} \rightarrow \|V - \bar{W}\bar{Q}V\|_F^2$. 依 LPBNMF 算法的迭代规则, $\{\|V - W(k)Q(k)V\|_F^2\}$ 是收敛的, 它一定与 $\{\|V - W(k_{l_h})Q(k_{l_h})V\|_F^2\}$ 收敛到同一点.

根据 LPBNMF 算法规则, $\{(\mathbf{W}_{\cdot 1}(k_{l_h} + 1), \mathbf{W}_{\cdot 2}(k_{l_h}), \dots, \mathbf{W}_{\cdot L}(k_{l_h}), Q(k_{l_h}))\}$ 可以认为是 $\{(W(k), Q(k))\}$ 的子列, 所以 $\{\|V - [(\mathbf{W}_{\cdot 1}(k_{l_h} + 1), \mathbf{W}_{\cdot 2}(k_{l_h}), \dots, \mathbf{W}_{\cdot L}(k_{l_h}))Q(k_{l_h})V]\|_F^2\} \rightarrow \|V - \bar{W}\bar{Q}V\|_F^2$.

根据引理 3, $\{\mathbf{W}_{\cdot 1}(k_{l_h} + 1)\}$ 中含有收敛子列, 任取其中收敛的子列 $\{\mathbf{W}_{\cdot 1}(k_{l_{h_q}} + 1)\}$. 根据引理 5, 固定 $\mathbf{W}_{\cdot 2}, \dots, \mathbf{W}_{\cdot L}$ 和 $Q > 0$ 后, $\|V - WQV\|_F^2$ 关于 $\mathbf{W}_{\cdot 1}$ 是严格凸的, 又因为 $\{(\mathbf{W}_{\cdot 2}(k_{l_{h_q}}), \dots, \mathbf{W}_{\cdot L}(k_{l_{h_q}}), Q(k_{l_{h_q}}))\} \rightarrow (\bar{\mathbf{W}}_{\cdot 2}, \dots, \bar{\mathbf{W}}_{\cdot L}, \bar{Q})$, 所以 $\{(\mathbf{W}_{\cdot 1}(k_{l_{h_q}} + 1), \mathbf{W}_{\cdot 2}(k_{l_{h_q}}), \dots, \mathbf{W}_{\cdot L}(k_{l_{h_q}}), Q(k_{l_{h_q}}))\} \rightarrow (\bar{\mathbf{W}}_{\cdot 1}, \bar{\mathbf{W}}_{\cdot 2}, \dots, \bar{\mathbf{W}}_{\cdot L}, \bar{Q})$, 那么 $\{\mathbf{W}_{\cdot 1}(k_{l_{h_q}} + 1)\} \rightarrow \bar{\mathbf{W}}_{\cdot 1}$. 因为收敛子列 $\{\mathbf{W}_{\cdot 1}(k_{l_{h_q}} + 1)\}$ 为任取的, 所以 $\{\mathbf{W}_{\cdot 1}(k_{l_h} + 1)\} \rightarrow \bar{\mathbf{W}}_{\cdot 1}$. 那么, $\{(\mathbf{W}_{\cdot 1}(k_{l_h} + 1), \mathbf{W}_{\cdot 2}(k_{l_h}), \dots, \mathbf{W}_{\cdot L}(k_{l_h}), Q(k_{l_h}))\} \rightarrow (\bar{\mathbf{W}}_{\cdot 1}, \bar{\mathbf{W}}_{\cdot 2}, \dots, \bar{\mathbf{W}}_{\cdot L}, \bar{Q})$.

同理, $\{(\mathbf{W}_{\cdot 1}(k_{l_h} + 1), \mathbf{W}_{\cdot 2}(k_{l_h} + 1), \dots, \mathbf{W}_{\cdot L}(k_{l_h}), Q(k_{l_h}))\} \rightarrow (\bar{\mathbf{W}}_{\cdot 1}, \bar{\mathbf{W}}_{\cdot 2}, \dots, \bar{\mathbf{W}}_{\cdot L}, \bar{Q})$. 依此继续, 显然有 $\{W(k_{l_h} + 1), Q(k_{l_h})\} \rightarrow (\bar{W}, \bar{Q})$.

W 是 LPBNMF 的基, 列满秩. 同理, $\{W(k_{l_h} + 1), Q(k_{l_h} + 1)\} \rightarrow (\bar{W}, \bar{Q})$. 所以, $\|(W(k_{l_h} + 1), Q(k_{l_h} + 1)) - (W(k_{l_h}), Q(k_{l_h}))\| \rightarrow 0$, 这与假设 $\|(W(k_{l_h} + 1), Q(k_{l_h} + 1)) - (W(k_{l_h}), Q(k_{l_h}))\| \rightarrow 0$ 不成立推出的结论相悖, 所以 $\|(W(k + 1), Q(k + 1)) - (W(k), Q(k))\| \rightarrow 0$.

$\|(W(k + 2), Q(k + 2)) - (W(k), Q(k))\| \leq \|(W(k + 2), Q(k + 2)) - (W(k + 1), Q(k + 1))\| + \|(W(k + 1), Q(k + 1)) - (W(k), Q(k))\| \rightarrow 0$. 根据数学归纳法, $\forall b \in \mathbf{Z}^+$, $\|(W(k + b), Q(k + b)) - (W(k), Q(k))\| \rightarrow 0$. \square

根据定理 3, $\forall \epsilon_1 > 0, \forall b \in \mathbf{Z}^+, \exists M(\epsilon_1, b) \in \mathbf{Z}^+, \|(W(k + b), Q(k + b)) - (W(k), Q(k))\| < \epsilon_1$, 只要 $k > M(\epsilon_1, b)$; 或者 $\forall M \in \mathbf{Z}^+, \forall b \in \mathbf{Z}^+, \exists \epsilon_2(M, b) > 0, \|(W(k + b), Q(k + b)) - (W(k), Q(k))\| < \epsilon_2(M, b)$, 只要 $k > M$. 也就是说, 经过足够步的迭代后, 再做 b 步迭代, LPBNMF 算法的输出不会被显著地改变; 或者说, 经过一定步数的迭代后, 再做 b 步迭代导致的 LPBNMF 算法输出变化可被控制在一定范围内.

定理 4 (LPBNMF 算法的收敛性定理). 取

$V > 0$, 定义解集合 $\Omega = \{\min_{\epsilon \leq W, \epsilon \leq Q} \|V - WQV\|_F^2 \text{ 的稳定点}\}$, LPBNMF 算法关于 Ω 收敛.

证明. 根据引理 3, $\{(W(k), Q(k))\}$ 中存在收敛子列 $\{(W(k_i), Q(k_i))\}$, 令 $\{(W(k_i), Q(k_i))\} \rightarrow (\bar{W}, \bar{Q})$. 根据引理 5, 固定 $\forall i, \{\mathbf{W}_{\cdot 1}, \mathbf{W}_{\cdot 2}, \dots, \mathbf{W}_{\cdot L}\}/\mathbf{W}_{\cdot i}$ 和 $Q > 0$ 后, $\|V - WQV\|_F^2$ 关于 $\mathbf{W}_{\cdot i}$ 是凸的, 且 LPBNMF 算法单调下降, 所以 $\forall \mathbf{W}_{\cdot i} \geq \epsilon, \|V - \bar{W}\bar{Q}V\|_F^2 \leq \|V - [\bar{\mathbf{W}}_{\cdot 1}, \dots, \bar{\mathbf{W}}_{\cdot i-1}, \mathbf{W}_{\cdot i}, \bar{\mathbf{W}}_{\cdot i+1}, \dots, \bar{\mathbf{W}}_{\cdot L}]QV\|_F^2$. 那么,

$$\forall \mathbf{W}_{\cdot i} \geq \epsilon, \left[\frac{\partial(\|V - WQV\|_F^2)}{\partial \mathbf{W}_{\cdot i}} \Big|_{(W, Q) = (\bar{W}, \bar{Q})} \right]^T \times (\mathbf{W}_{\cdot i} - \bar{\mathbf{W}}_{\cdot i}) \geq 0, \forall i \in \{1, 2, \dots, L\} \quad (33)$$

因为此时从 $\bar{\mathbf{W}}_{\cdot i}$ 点出发在 $\{\mathbf{W}_{\cdot i} | \mathbf{W}_{\cdot i} \geq \epsilon\}$ 区域内找不到下降方向. 同理,

$$\forall Q_{\cdot j} \geq 0, \left[\frac{\partial(\|V - WQV\|_F^2)}{\partial Q_{\cdot j}} \Big|_{(W, Q) = (\bar{W}, \bar{Q})} \right]^T \times (Q_{\cdot j} - \bar{Q}_{\cdot j}) \geq 0, \forall j \in \{1, 2, \dots, M\} \quad (34)$$

式 (33) 和 (34) 合在一起构成了 (\bar{W}, \bar{Q}) 为 $\min_{\epsilon \leq W, \epsilon \leq Q} \|V - WQV\|_F^2$ 的稳定点的充分条件^[39]. 定义解集合 $\Omega = \{\min_{\epsilon \leq W, \epsilon \leq Q} \|V - WQV\|_F^2 \text{ 的稳定点}\}$. $\{(W(k_i), Q(k_i))\}$ 为任取的, 所以 $\{(W(k), Q(k))\}$ 的所有极限点均在 Ω 中, 根据算法收敛的定义^[38, 40], LPBNMF 算法关于 Ω 收敛. \square

4 相关方法讨论和分析

截断奇异值分解 (Truncated singular value decomposition, tSVD) 是一种由 SVD 重建结合负分量归零处理来实现对数据非负重建描述的方法, 由其重建的数据矩阵的秩重建前不能确定 (只能重建后再判断) 和描述准确性差是它的两个主要缺点^[41]. 为构造无上述缺陷的非负投影 (重建) 方法, Yuan 等提出了投影 NMF (Projective NMF, P-NMF)^[41]. P-NMF 也是通过线性变换实现降维, 这是它与 LPBNMF 相关的地方, 但比较前面对 P-NMF 设计目的的介绍和第 2 节对 LPBNMF 设计目的说明可知, 这两种方法的设计目的迥异, 除此之外, 它们在模型、算法设计方式以及算法性质上也存在明显差异, 以下具体讨论.

定义 4. 对于 $M \times N$ 维数据矩阵 V , P-NMF 通过解 ($\|\cdot\|$ 代表矩阵范数或散度)^[41]

$$\min_{0 \leq W} \|V - WW^T V\| \quad (35)$$

求得非负的 $M \times L$ 维矩阵 W , 使得

$$V \approx WW^T V \quad (36)$$

比较定义 2 和定义 4 可知, LPBNMF 和 P-NMF 使用的数据表示模型不同. Yuan 等称数据重建 $\hat{V} = PV$ 时使用的矩阵 P 为投影矩阵, 并称其采用的投影矩阵 $P = WW^T$ 是对称式的^[41]; 然而, LPBNMF 采用的投影矩阵 $P = WQ$ 是非对称式的 (沿用 Yuan 等使用的术语). 应指出, 投影矩阵构造上的不同决定了 LPBNMF 模型性质比 P-NMF 模型性质好, 这是因为: 一方面, 在优化中, LPBNMF 每次处理的优化问题 (优化 W 或 Q 中的一个) 比 P-NMF 处理的优化问题 (优化 W 但它以 WW^T 的形式出现在目标函数中) 简单, 这有利于 (像定理 1 和定理 2 那样) 利用模型性质构造简单实用的专门算法; 另一方面, LPBNMF 的优化问题具有局部凸性质 (见引理 5), 局部凸特性利于算法性质的分析和证明 (见定理 3 和定理 4), 然而, 无论取 $\|\cdot\|$ 为哪种形式的范数或散度, P-NMF 优化模型均不具有局部凸特性 (举例来说, 当选取最常用的欧几里德距离平方来度量 V 与 $WW^T V$ 间的差异, 即取 $\|\cdot\|$ 为 $\|\cdot\|_F^2$, $\|V - WW^T V\|_F^2$ 是 W 的四次函数, 它显然不是局部凸的).

Yuan 等按迭代步长自适应选取的梯度下降原则构造了两个 P-NMF 算法^[41], 这两个算法的迭代特性相同, 所以下边重点介绍和分析其中的一个 — 取 $\|\cdot\|$ 为欧几里德距离平方时的 P-NMF 算法.

取 $\|\cdot\|$ 为欧几里德距离平方时, P-NMF 的优化模型为

$$\min_{0 \leq W} \frac{1}{2} \|V - WW^T V\|_F^2 \quad (37)$$

采用常规的梯度法可得 W 的优化规则

$$W_{ij} \leftarrow W_{ij} - \eta_{ij} \frac{\partial \frac{1}{2} \|V - WW^T V\|_F^2}{\partial W_{ij}} \\ \frac{\partial \frac{1}{2} \|V - WW^T V\|_F^2}{\partial W_{ij}} = -2[VV^T W]_{ij} + \\ [WW^T VV^T W]_{ij} + [V^T W W^T W]_{ij}, \quad \forall i, j \quad (38)$$

为使每步迭代后的 W 值均非负, Yuan 等提出了按^[41]

$$\eta_{ij} = \frac{W_{ij}}{[WW^T VV^T W]_{ij} + [V^T W W^T W]_{ij}}, \quad \forall i, j \quad (39)$$

取迭代步长, 将式 (39) 带入式 (38) 可得 W 的优化规则 (注意文献 [41] 的推导中丢掉了式 (40) 中的系

数 “2”)

$$W_{ij} \leftarrow \frac{2W_{ij}[VV^T W]_{ij}}{[WW^T VV^T W]_{ij} + [V^T W W^T W]_{ij}}, \quad \forall i, j \quad (40)$$

比较定理 1、定理 2 以及式 (40) 的构造过程可知, LPBNMF 算法和 P-NMF 算法的构造方式本质不同. P-NMF 算法设计迭代步长表达式 (式 (39)) 的目的是为了得到乘性更新迭代规则 (式 (40)), 它可保证每步迭代结果均非负. 但众所周知, 使用梯度下降法时, 迭代步长的选取是非常关键的, 步长选取不好会造成算法震荡, 所以按上述方式给迭代步长设定一个固定的表达式 (也就是说, 当迭代初值给定, 每步迭代的步长也就固定了) 是无法保证 P-NMF 算法单调下降的, 此时算法的稳定性和收敛性更是无法保证 (文献 [41] 中未对 P-NMF 算法的这两个性质进行讨论). 图 2 记录了取 $L = 80$ 时, 用 P-NMF 算法处理 ORL 库 400 幅图像的迭代曲线, 它显示此时算法是剧烈震荡的, 这与上述分析是一致的. LPBNMF 算法对 $W_{\cdot 1}, W_{\cdot 2}, \dots, W_{\cdot L}$ 和 Q 分别根据定理 1 和定理 2 做交替优化, 定理 1 和定理 2 的性质在理论上保证了 LPBNMF 算法一定单调下降, 这如图 3 中记录的取 $L = 80$ 时 LPBNMF 算法处理 ORL 库 400 幅图像的迭代曲线所示. 此外, 根据 LPBNMF 目标函数的局部凸特性 (引理 5) 及连续性 (引理 3)、LPBNMF 算法的有界性 (引理 2) 和单调性, LPBNMF 算法的稳定性 (定理 3) 和收敛性 (定理 4) 在数学上可证明.

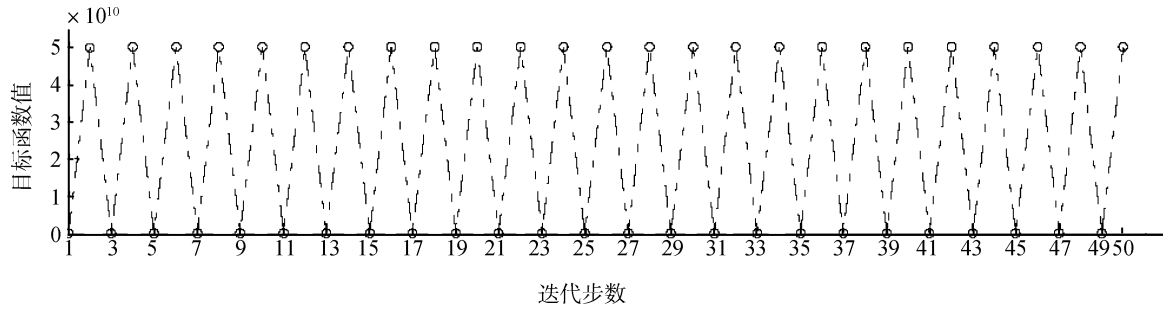
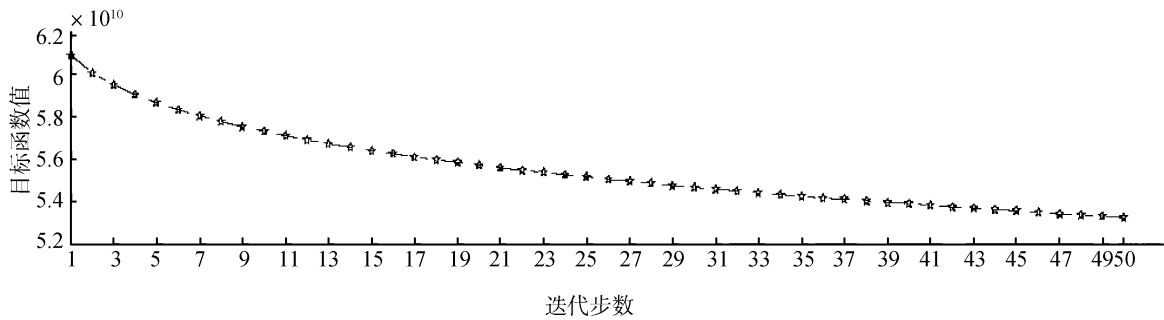
综上所述, P-NMF 和 LPBNMF 均采用了线性变换来实现数据降维, 但它们的设计出发点不同, 它们基于的数学模型也不同, LPBNMF 模型的性质比 P-NMF 模型好. LPBNMF 算法和 P-NMF 算法的构造方式本质不同, P-NMF 算法不具单调性、稳定性和收敛性, 实用性不好, 而 LPBNMF 算法单调、稳定、收敛, 便于实际应用.

5 实验

本节用比较实验直观地验证 LPBNMF 的性能, 实验涉及降维效率、特征稀疏性以及在有遮挡人脸识别中的应用等三个方面.

5.1 降维效率

LPBNMF 比 NMF 的降维效率高是显然的, 此处利用实验说明 LPBNMF 降维较 NMF 降维的效率提高程度. 用 ORL 库中的 200 幅图像 (每人任取的 5 幅图) 分别进行 NMF 和 LPBNMF 训练, 表 2 (见第 33 页) 记录了利用这时的 NMF 和 LPBNMF

图 2 P-NMF 算法处理 ORL 库 400 幅图像的迭代曲线 ($L = 80$)Fig. 2 The iterative curve of processing all 400 images of ORL database by P-NMF algorithm ($L = 80$)图 3 LPBNMF 算法处理 ORL 库 400 幅图像的迭代曲线 ($L = 80$)Fig. 3 The iterative curve of processing all 400 images of ORL database by LPBNMF algorithm ($L = 80$)

训练结果将 ORL 库中 10 幅非训练用图像降到 10 个不同维度的计算耗时 (降维耗时基于配置为 Pentium 3.06 G CPU, 1.5 G 内存的 PC 记录).

观察表 2 最后一行的“平均效率提升”知: 此时 LPBNMF 的平均降维效率至少比 NMF 高 145 倍, 保留维数越多, 两者间效率差异越大. 降维效率提升倍数基本随保留维数的增加而显著增加, 当保留 200 维时, LPBNMF 平均降维效率较 NMF 提升可达 2967 倍.

5.2 特征稀疏性

图 4 和图 5 分别基于 ORL 库的全部 400 幅图和 AR 数据库 118 人的每人 8 幅图 (每人 8 幅无遮挡、无表情变化图, 其中正面光照下 2 幅, 其他光照条件下 6 幅) 记录了取 L 为 64、100 和 144 时做 LPBNMF 和 NMF 得到的特征图像. 图 4 和图 5 均显示, 无论 L 取值大小, NMF 特征都是非常全局化的, 而 LPBNMF 特征都是非常稀疏和局部化的, 这验证了第 2 节的理论分析结论.

为定量分析 LPBNMF 和 NMF 特征的稀疏性, 可引入稀疏度的定义. 单个特征的稀疏度用

$$sparseness(\mathbf{W}_{\cdot i}) = \frac{\sqrt{M} - \frac{\|\mathbf{W}_{\cdot i}\|_1}{\|\mathbf{W}_{\cdot i}\|_2}}{\sqrt{M} - 1}, \quad \forall \mathbf{W}_{\cdot i} \in \mathbf{R}^M \quad (41)$$

来度量^[42]. 如果 $\mathbf{W}_{\cdot i}$ 中仅有一个元素非零 (最稀疏的情况), 则 $sparseness(\mathbf{W}_{\cdot i})$ 为 1; 如果 $\mathbf{W}_{\cdot i}$ 中所有元素的值均相等 (最不稀疏/最平滑的情况), 则 $sparseness(\mathbf{W}_{\cdot i})$ 为 0. 全部特征的稀疏度用











$$sparsenessW = \frac{\left\{ \sum_{i=1}^L sparseness(\mathbf{W}_{\cdot i}) \right\}}{L} \quad (42)$$

来度量. 图 4 和图 5 的图题中给出了不同 L 取值下的 LPBNMF 和 NMF 特征 (全部特征) 的稀疏度, 由其可从定量角度再次验证 LPBNMF 特征远比 NMF 特征稀疏. 此外, 观察不同 L 取值下 LPBNMF 特征稀疏度易知, 随 L 取值增大, LPBNMF 特征稀疏度也伴随提高, 这验证了第 2 节的理论分析结论.

图 6 和图 7 分别基于 ORL 库的全部 400 幅图和 AR 数据库 118 人的每人 8 幅图 (每人 8 幅无遮

表 2 LPBNMF 与 NMF 的降维效率比较

Table 2 Comparison of computational loads of dimensionality reduction between LPBNMF and NMF

被处理 图像	方法	降到 k 维时的计算耗时 (s)									
		$k = 20$	$k = 40$	$k = 60$	$k = 80$	$k = 100$	$k = 120$	$k = 140$	$k = 160$	$k = 180$	$k = 200$
	LPBNMF	0.005	0.0068	0.0072	0.0075	0.0085	0.0119	0.0125	0.0136	0.0148	0.017
	NMF	0.8904	2.7124	7.9432	12.6781	14.2605	18.5714	25.7311	29.3181	39.5891	49.3811
	LPBNMF	0.0057	0.0065	0.0066	0.0079	0.0083	0.012	0.0133	0.0135	0.0147	0.0158
	NMF	1.0201	2.8617	7.9911	11.8712	13.872	18.2179	23.6676	30.6902	38.0799	43.9134
	LPBNMF	0.0056	0.0071	0.0073	0.0078	0.0086	0.0121	0.0132	0.0138	0.0137	0.0159
	NMF	0.7034	2.4151	6.0356	9.1215	11.1235	15.1315	22.2251	30.4177	33.7389	45.2733
	LPBNMF	0.008	0.0064	0.007	0.0072	0.0086	0.0117	0.0133	0.0135	0.0131	0.0156
	NMF	0.8415	2.3334	7.5893	9.7237	12.7782	17.9596	24.8947	35.1049	33.6579	48.2156
	LPBNMF	0.0057	0.0067	0.007	0.0074	0.0087	0.0113	0.0127	0.0133	0.0139	0.0161
	NMF	0.8222	2.3532	7.9201	9.788	13.9182	17.6892	26.4126	34.2119	39.4444	49.6071
	LPBNMF	0.0056	0.0067	0.007	0.007	0.0087	0.0116	0.0128	0.0132	0.0116	0.0158
	NMF	0.8019	2.5696	7.1564	10.8508	11.6862	16.3642	26.1722	32.2577	40.1346	51.5443
	LPBNMF	0.0056	0.0064	0.0064	0.0068	0.0087	0.0116	0.013	0.0134	0.0142	0.0157
	NMF	0.8488	2.809	7.294	12.4735	12.1533	21.6051	29.4304	29.9614	39.3462	48.5454
	LPBNMF	0.0056	0.0067	0.0067	0.0069	0.0088	0.0115	0.0126	0.0133	0.0143	0.0155
	NMF	0.8745	2.5462	8.6122	13.2749	11.3076	16.3672	24.551	28.7593	35.2532	45.3644
	LPBNMF	0.0056	0.0065	0.007	0.0071	0.0088	0.0115	0.0126	0.0134	0.0144	0.0156
	NMF	0.6875	2.6951	5.0178	8.505	10.7513	21.1786	21.2424	25.5285	37.2458	41.6976
	LPBNMF	0.0056	0.0065	0.0073	0.0078	0.0088	0.0116	0.0123	0.0132	0.0149	0.0157
	NMF	0.9712	3.6955	5.4385	9.6921	12.5069	19.0481	25.8863	37.0015	41.6306	48.3634
平均降 维耗时	LPBNMF	0.0058	0.0066	0.0069	0.0073	0.0087	0.0117	0.0128	0.0134	0.014	0.0159
	NMF	0.8461	2.6991	7.0684	10.7979	12.5348	18.2311	22.3801	31.3251	37.8121	47.1906
平均效率提升 (倍)		145	408	1023	1478	1428	1555	1747	2336	2700	2967

挡、无表情变化图, 其中正面光照下 2 幅, 其他光照条件下 6 幅) 记录了取 L 为 64、100 和 144 时 LPBNMF 特征和 NMF 特征正交性随迭代步数增加的变化情况。

$\{\mathbf{W}_{.1}, \mathbf{W}_{.2}, \dots, \mathbf{W}_{.L}\}$ 的正交性用

$$\frac{\sum_{i=1}^L \sum_{j=1}^L [W^T W]_{ij} - \text{tr}\{W^T W\}}{\text{tr}\{W^T W\}} \quad (43)$$

来度量. 式 (43) 基于这样的思路构造: 如果 $\{\mathbf{W}_{.1}, \mathbf{W}_{.2}, \dots, \mathbf{W}_{.L}\}$ 是正交的, 则 $W^T W$ 的非对角元素均为 0; 那么, 当 $\{\mathbf{W}_{.1}, \mathbf{W}_{.2}, \dots, \mathbf{W}_{.L}\}$ 不是严格正交的, 则 $W^T W$ 的非对角元素和占对角元素和的比例越小, $\{\mathbf{W}_{.1}, \mathbf{W}_{.2}, \dots, \mathbf{W}_{.L}\}$ 越趋向于正交. 第 2 节讨论了 LPBNMF 特征远比 NMF 特征稀疏和局部化的原因在于: 做 LPBNMF 时 W 有趋向正交的趋势, 而做 NMF 时无这种趋势, 图 6 和图 7 对此予以了验证.

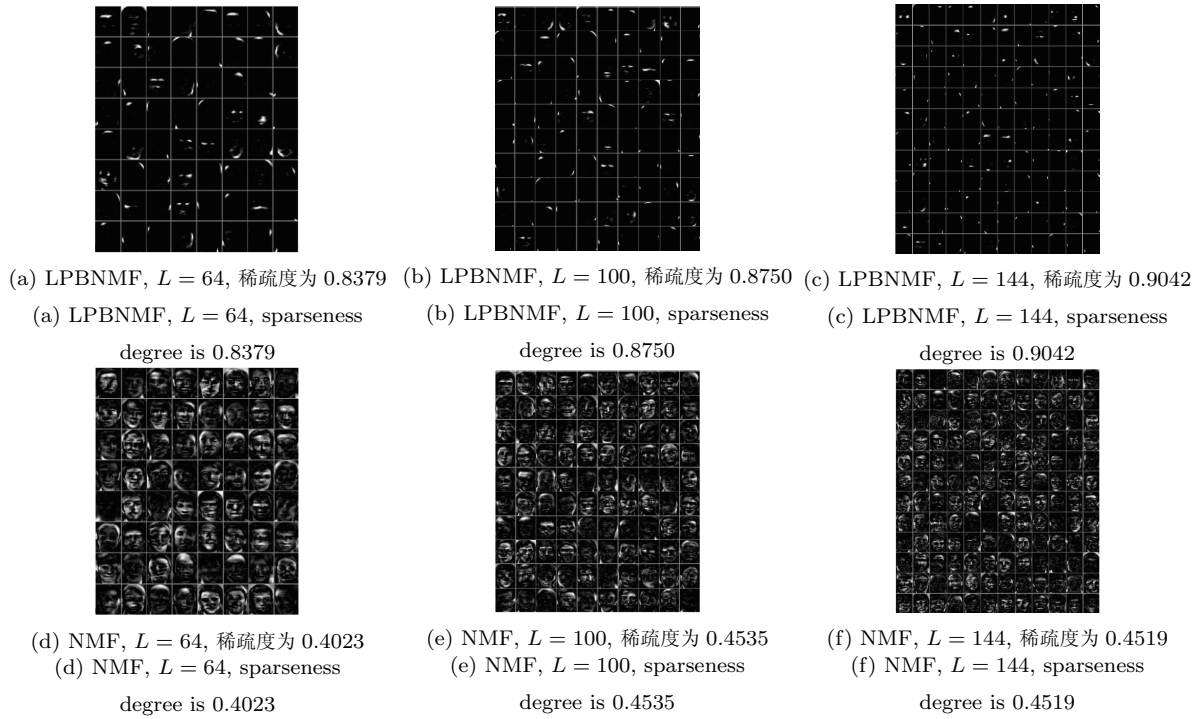


图 4 LPBNMF 和 NMF 的特征稀疏性比较 (基于 ORL 库的全部 400 幅图)
 Fig. 4 Comparison of the sparseness of features between LPBNMF and NMF
 (based on all 400 images of ORL database)

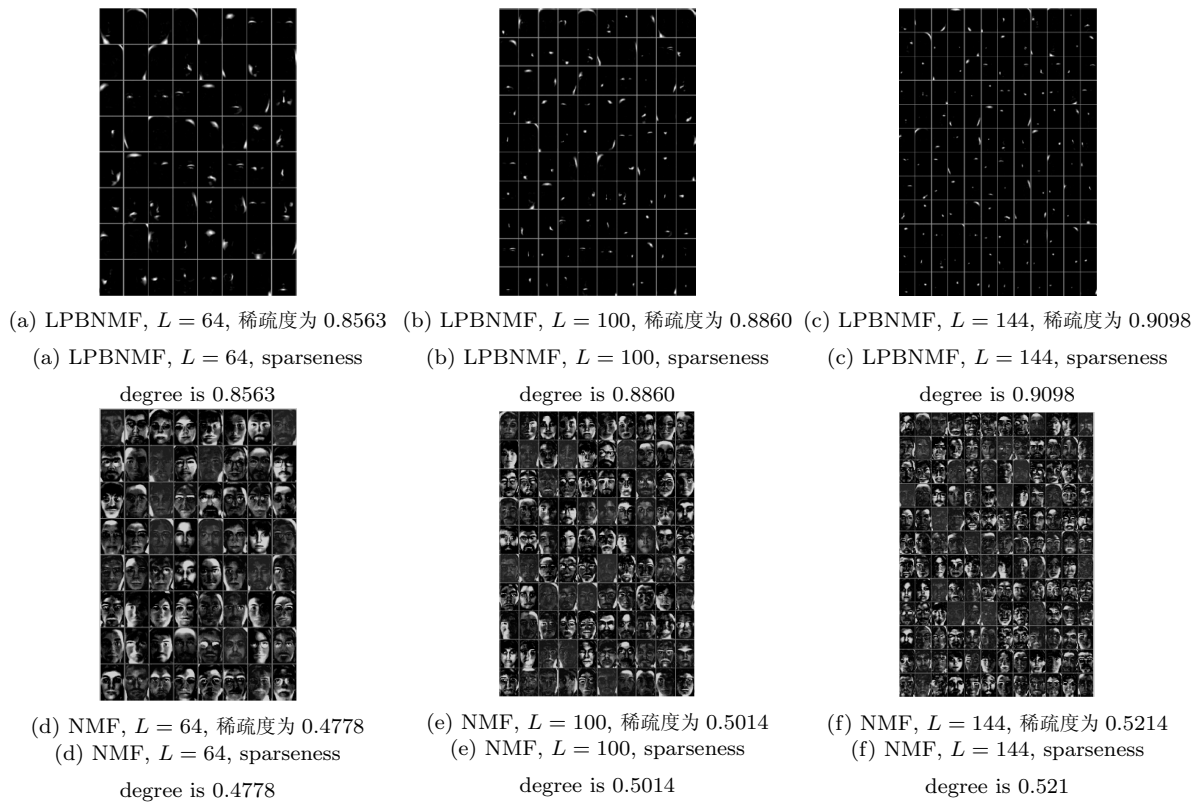


图 5 LPBNMF 和 NMF 的特征稀疏性比较 (基于 AR 库中 118 人的每人 8 幅无遮挡和表情变化图)
 Fig. 5 Comparison of the sparseness of features between LPBNMF and NMF (based on 118 objects' 944 images
 (8 images of each object) without occlusions and expressional variations in AR database)

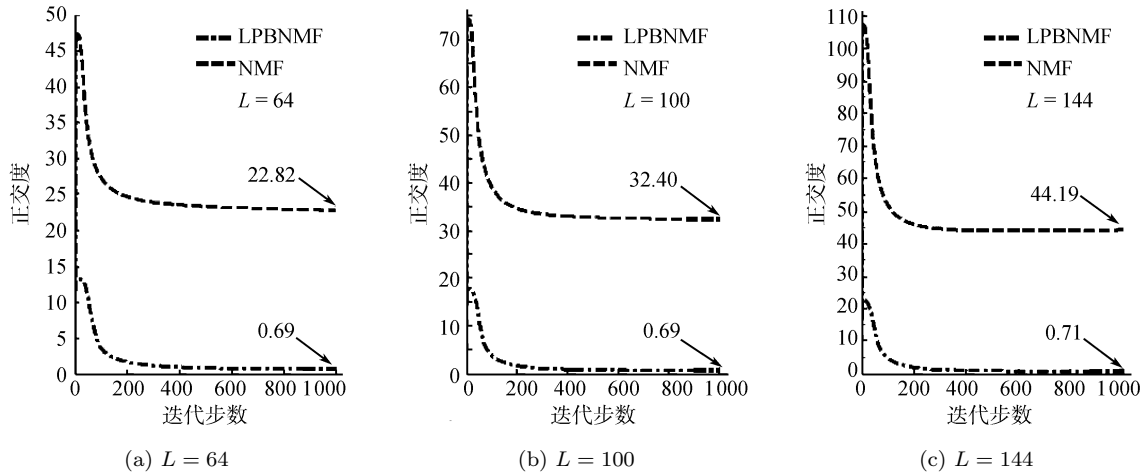


图 6 迭代过程中 LPBNMF 和 NMF 特征正交性的变化规律 (基于 ORL 库的全部 400 幅图)
Fig. 6 Orthogonal degree of features in the iterative implementation of LPBNMF and NMF (based on all 400 images of ORL database)

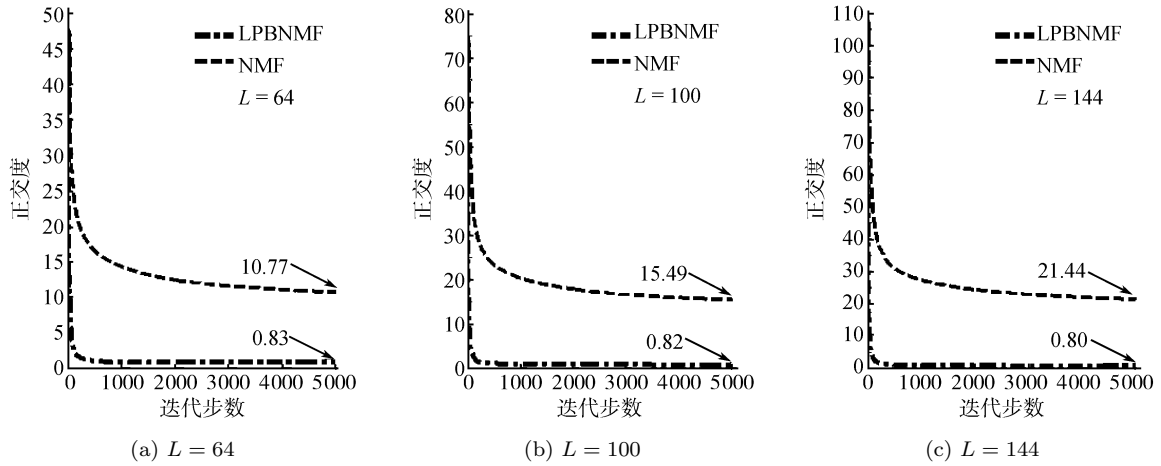


图 7 迭代过程中 LPBNMF 和 NMF 特征正交性的变化规律 (基于 AR 库 118 人每人 8 幅无遮挡和表情变化图)
Fig. 7 Orthogonal degree of features in the iterative implementation of LPBNMF and NMF (based on 118 objects' 944 images (8 images of each object) without occlusions and expressional variations in AR database)



图 8 AR 数据库人脸图像示例
Fig. 8 Examples of face images in AR database

5.3 有遮挡人脸识别

利用稀疏和局部化的特征可以有效抑制由遮挡给特征提取造成的不利影响, 因为这时部分遮挡只会影响有限的特征投影系数值. 那么, 可推知 LPB-NMF 特征将非常适合于处理有遮挡人脸识别问题, 本小节利用基于 AR 数据库的实验对此进行验证.

AR 数据库中包含 76 名男性和 60 名女性的人脸数据, 其中 118 人含有如图 8 所示的 26 幅图像. 图 8 中图像分上下两组, 它们在摄取时间上间隔为两周, 每组中从左到右 13 幅图像的摄取特征分别为正常表情 + 正常光照、微笑 + 正常光照、生气 + 正常光照、震惊 + 正常光照、正常表情 + 右侧强光、正常表情 + 左侧强光、正常表情 + 左右侧强光、眼部遮挡 + 正常光照、眼部遮挡 + 右侧强光、眼部遮挡 + 左侧强光、嘴部遮挡 + 正常光照、嘴部遮挡 + 右侧强光以及嘴部遮挡 + 左侧强光.

本小节利用这 118 人的图像进行了实验. 这 118 人中, 每人两组眼部被遮挡图像共 118×6 幅作为一组测试图像, 每人两组嘴部被遮挡图像共 118×6 幅作为另一组测试图像. 因每人的眼部或嘴部被遮挡的 6 幅图像中均蕴含左右侧的光照变化以及拍摄时间变化, 为突出测试遮挡情况下的人脸识别性能, 减弱光照变化和拍摄时间变化给识别造成的影响, 所以使训练集也包含这两种变化因素. 那么, 这 118 人的每人两组正常表情 + 正常光照、正常表情 + 右侧强光和正常表情 + 左侧强光图像共 118×6 幅作为训练图像.

为进一步减弱光照变化给识别造成的影响, 对全部训练图和测试图进行分块光照和对比度补偿, 分块光照和对比度补偿这样进行: 1) 将人脸图像等分为 9 块; 2) 设定被补偿后图像的均值 em 和标准差 $estd$; 3) 逐块求原图像块的均值 m 和标准差 std ; 4) 逐块依照 $I_{ij}^k \leftarrow em + (I_{ij}^k - m) \times estd/std$ 逐个像素调整灰度值, 得到进行补偿后的图像 (用 I_{ij}^k 表示人脸图像的第 k 块的第 i 行第 j 列的位置的像素值, $k \in \{1, 2, \dots, 9\}$). 采用分块光照和对比度补偿, 而不采用对整幅图像的光照和对比度补偿 (或直方图均衡化), 是为了避免遮挡物与未被遮挡人脸间的灰度剧烈差异给“削弱光照变化”预处理带来的负面影响.

为通过对比的方式直观说明 LPB-NMF 特征在处理有遮挡人脸识别问题上的性能, NMF 特征、LDA (Fisherface) 特征^[43] 以及 PCA (Eigenface) 特征^[44] 在这里作为 LPB-NMF 特征的对比对象.

人脸识别时的特征投影系数匹配采用最近邻法, 考虑 6 个分类器: 基于城市距离的最近邻分类器 (简记为 L1)^[45]、基于欧几里德距离的最近邻分类器

(简记为 L2)^[45]、基于夹角角度的最近邻分类器 (简记为 θ)^[45] 以及标准化了的 L1、L2 和 θ , 分别记为 NL1、NL2 和 $N\theta$. 上述标准化的最近邻匹配指, 在常规最近邻匹配前先将特征投影系数各维均基于 2-范数做归一化处理的最邻匹配. 例如, 假设 $\{h_j, j = 1, 2, \dots, N\}$ 为训练图像的特征投影系数, 归一化训练图像特征投影系数 $\{Nh_j, j = 1, 2, \dots, N\}$ 按

$$Nh_{ij} \leftarrow \frac{h_{ij}}{\sqrt{\sum_{j=1}^N h_{ij}^2}}, \quad \forall i, j \quad (44)$$

求取, 对于某测试图像的特征投影系数 h^{test} , 其归一化结果 Nh^{test} 为

$$Nh_i^{test} \leftarrow \frac{h_i^{test}}{\sqrt{\sum_{j=1}^N h_{ij}^2}}, \quad \forall i \quad (45)$$

基于上述的实验设置, 图 9 和图 10 分别给出了 LPB-NMF、NMF、LDA (Fisherface) 和 PCA (Eigenface) 匹配最优近邻分类器时的眼部遮挡人脸识别性能和嘴部遮挡人脸识别性能. 基于 LPB-NMF、NMF 和 PCA 的实验依保留维数为 20, 25, \dots , 200 共做 37 组; 根据 LDA 自身的特点, 最大保留维数为训练样本类数减 1^[41], 即 117 维, 基于 LDA 的实验依保留维数为 21, 24, \dots , 117 共做 33 组. 图 9 中, 为突出不同方法最优识别性能间的差异 (避免因显示的范围过大而使不同方法最优识别性能间的差异在图上不好辨识), 只显示识别率高于 60% 的实验结果. 图 9 和图 10 显示, 无论遮挡出现在眼部还是在嘴部, 基于 LPB-NMF 的识别效果均远好于基于 NMF、LDA 以及 PCA 的识别效果. 具体来说, 眼部遮挡时, 基于 LPB-NMF 方法的最优识别率至少比其他方法的最优识别率高 4 个百分点; 嘴部遮挡时, 基于 LPB-NMF 方法的最优识别率至少比其他方法的最优识别率高 7 个百分点. 图 8 显示, 实验图像的嘴部遮挡范围要远大于眼部遮挡范围, 因此图 7 和图 8 的实验结果还说明: 一方面, 无论对哪一种方法来说, 都是遮挡越严重识别起来越难; 另一方面, 遮挡越厉害, 基于 LPB-NMF 的方法较其他方法的优越性越突出 (最优识别率间的差异越大).

6 结论

降维效率低和特征的稀疏性差是用 NMF 做数据处理时常遇到的两个问题, 这限制了 NMF 的应用范围, 制约了 NMF 应用研究的发展. 为一并解决

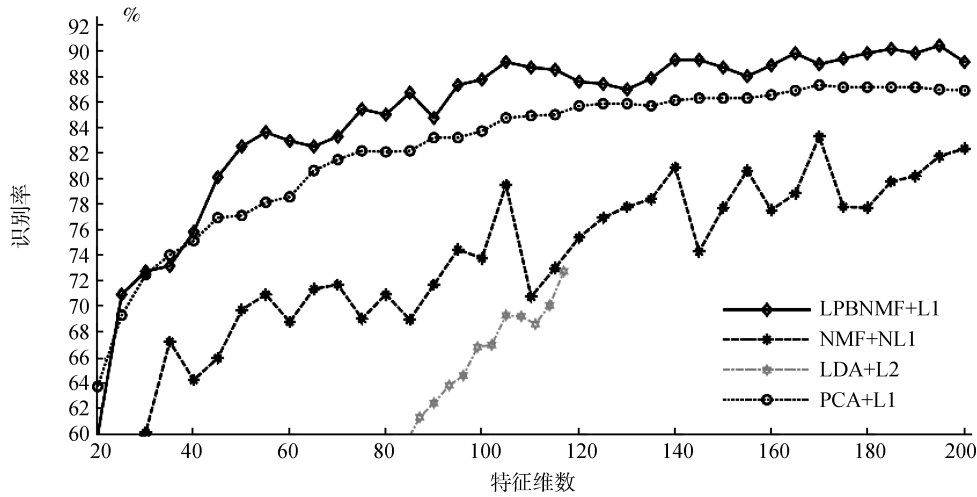


图9 LPBNMF、NMF、LDA 和 PCA 匹配最优近邻分类器时的眼部遮挡人脸识别性能
Fig. 9 The best nearest neighbor classification-based recognition of faces with an occlusion in eyes while using LPBNMF, NMF, LDA and PCA

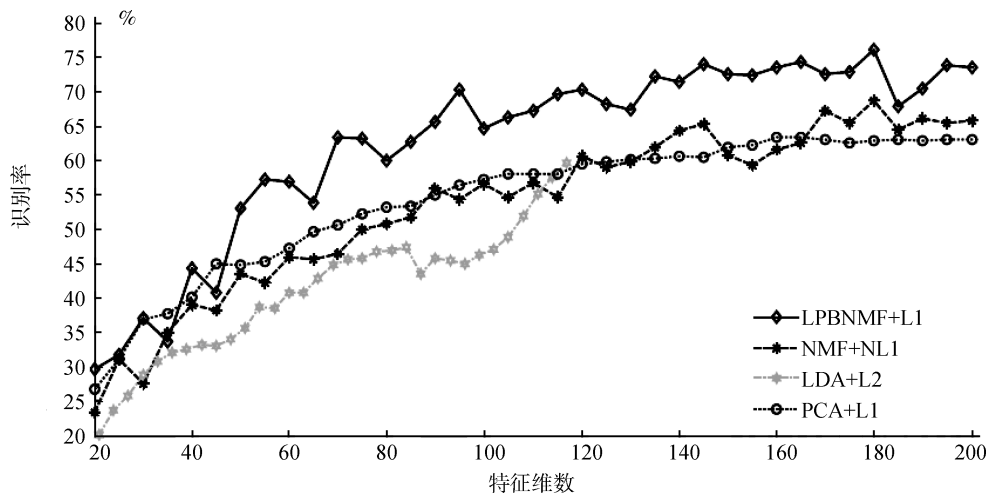


图10 LPBNMF、NMF、LDA 和 PCA 匹配最优近邻分类器时的嘴部遮挡人脸识别性能
Fig. 10 The best nearest neighbor classification-based recognition of faces with an occlusion in mouth while using LPBNMF, NMF, LDA, and PCA

这两个问题, 本文提出了一个新的模型 — LPBNMF, 构造了一个单调的 LPBNMF 算法, 证明了此算法的稳定性和收敛性。

理论分析和实验结果均显示: 1) LPBNMF 的降维效率远高于 NMF; 2) LPBNMF 特征明显比 NMF 特征更稀疏和局部化; 3) LPBNMF 特征比 NMF 特征、LDA 特征以及 PCA 特征更适合于用最近邻分类法处理有遮挡人脸识别问题。

References

- 1 Lee D D, Seung H S. Learning the parts of objects by non-negative matrix factorization. *Nature*, 1999, **401**(6755): 788–791
- 2 Palmer S E. Hierarchical structure in perceptual representation. *Cognitive Psychology*, 1977, **9**(4): 441–474
- 3 Wachsmuth E, Oram M W, Perrett D I. Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque. *Cerebral Cortex*, 1994, **4**(5): 509–522
- 4 Logothetis N K, Sheinberg D L. Visual object recognition. *Annual Review of Neuroscience*, 1996, **19**(1): 577–621
- 5 Lee D D, Seung H S. Algorithms for non-negative matrix factorization. In: *Proceedings of the Conference on Neural Information Processing Systems*. Cambridge, USA: MIT Press, 2000. 556–562
- 6 Wild S, Curry J, Dougherty A. Improving non-negative matrix factorizations through structured initialization. *Pattern Recognition*, 2004, **37**(11): 2217–2232

- 7 Cichocki A, Amari S, Zdunek R, Kompass R, Hori G, He Z H. Extended SMART algorithms for non-negative matrix factorization. In: Proceedings of the 8th International Conference on Artificial Intelligence and Soft Computing. Zakopane, Poland: Springer, 2006. 548–562
- 8 Heiler M, Schnorr C. Learning sparse representations by non-negative matrix factorization and sequential cone programming. *Journal of Machine Learning Research*, 2006, **7**(7): 1385–1407
- 9 Chen Wei-Gang, Qi Fei-Hu. Learning NMF representation using a hybrid method combining feasible direction algorithm and simulated annealing. *Acta Electronica Sinica*, 2003, **31**(z1): 2190–2193
(陈卫刚, 戚飞虎. 可行方向算法与模拟退火结合的 NMF 特征提取方法. 电子学报, 2003, **31** (z1): 2190–2193)
- 10 Li L, Zhang Y J. FastNMF: highly efficient monotonic fixed-point non-negative matrix factorization algorithm with good applicability. *Journal of Electronic Imaging*, 2009, **18**(3): 033004
- 11 Kompass R. A generalized divergence measure for nonnegative matrix factorization. *Neural Computation*, 2007, **19**(3): 780–791
- 12 Cichocki A, Zdunek R, Amari S. Csiszar's divergences for non-negative matrix factorization: family of new algorithms. In: Proceedings of the 6th International Conference on Independent Component Analysis and Blind Signal Separation. Charleston, USA: Springer, 2006. 32–39
- 13 Zdunek R, Cichocki A. Non-negative matrix factorization with quasi-Newton optimization. Proceedings of the International Conference on Artificial Intelligence and Soft Computing. Zakopane, Poland: Springer, 2006. 870–879
- 14 Dhillon I S, Sra S. Generalized nonnegative matrix approximations with Bregman divergences. In: Proceedings of the Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2005. 283–290
- 15 Li Le, Zhang Yu-Jin. Survey on algorithms of non-negative matrix factorization. *Acta Electronica Sinica*, 2008, **36**(4): 737–743
(李乐, 章毓晋. 非负矩阵分解算法综述. 电子学报, 2008, **36**(4): 737–743)
- 16 Shahnaz F, Berry M W, Pauca V P, Plemmons R J. Document clustering using nonnegative matrix factorization. *Information Processing and Management*, 2006, **42**(2): 373–386
- 17 Liu W X, Zheng N N, You Q B. Non-negative matrix factorization and its applications in pattern recognition. *Chinese Science Bulletin*, 2006, **51**(1): 7–18
- 18 Ouhsain M, Hamza A B. Image watermarking scheme using nonnegative matrix factorization and wavelet transform. *Expert Systems with Applications*, 2009, **36**(2): 2123–2129
- 19 Wang Feng. Research of Face Detection Technology Based on Skin Color and Non-negative Matrix Factorization [Master dissertation], Huaqiao University, China, 2006
(王丰. 基于肤色和非负矩阵分解算法的人脸检测技术研究 [硕士学位论文], 华侨大学, 中国, 2006)
- 20 Guillamet D, Vitrià J. Non-negative matrix factorization for face recognition. In: Proceedings of the 5th Catalanian Conference on AI: Topics in Artificial Intelligence. Castellon, Spain: Springer, 2002. 336–344
- 21 Liang D, Yang J, Chang Y C. Relevance feedback based on non-negative matrix factorisation for image retrieval. *IEEE Proceedings on Vision, Image and Signal Processing*, 2006, **153**(4): 436–444
- 22 Liang Dong, Yang Jie, Lu Jin-Jun, Chang Yu-Chou. The latent semantic image retrieval based on non-negative matrix factorization. *Journal of Shanghai Jiaotong University*, 2006, **40**(5): 787–791
(梁栋, 杨杰, 卢进军, 常宇畴. 基于非负矩阵分解的隐含语义图像检索. 上海交通大学学报, 2006, **40**(5): 787–791)
- 23 Kopriva I, Nuzillard D. Non-negative matrix factorization approach to blind image deconvolution. In: Proceedings of the 6th International Conference on Independent Component Analysis and Blind Signal Separation. Charleston, USA: Springer, 2006. 966–973
- 24 Stouten V, Demuyneck K, van Hamme H. Automatically learning the units of speech by non-negative matrix factorisation. In: Proceedings of European Conference on Speech Communication and Technology. Antwerp, Belgium: Springer, 2007. 1937–1940
- 25 Stouten V, Demuyneck K, van Hamme H. Discovering phone patterns in spoken utterances by non-negative matrix factorization. *IEEE Signal Processing Letters*, 2008, **15**(1): 131–134
- 26 Cho Y C, Choi S, Bang S Y. Non-negative component parts of sound for classification. In: Proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology. Washington D. C., USA: IEEE, 2003. 633–636
- 27 Holzapfel A, Stylianou Y. Musical genre classification using nonnegative matrix factorization-based features. *IEEE Transactions on Audio, Speech, and Language Processing*, 2008, **16**(2): 424–434
- 28 Benetos E, Kotti M, Kotropoulos C. Musical instrument classification using non-negative matrix factorization algorithms and subset feature selection. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. Washington D. C., USA: IEEE, 2006. 14–19
- 29 Guan X H, Wang W, Zhang X L. Fast intrusion detection based on a non-negative matrix factorization model. *Journal of Network and Computer Applications*, 2009, **32**(1): 31–44
- 30 Frigyesi A, Høglund M. Non-negative matrix factorization for the analysis of complex gene expression data: identification of clinically relevant tumor subtypes. *Cancer Informatics*, 2008, **6**: 275–292
- 31 Kim H, Park H. Cancer class discovery using non-negative matrix factorization based on alternating non-negativity-constrained least squares. In: Proceedings of the 3rd International Symposium on Bioinformatics Research and Applications. Atlanta, USA: Springer, 2007. 477–487
- 32 AT and T Research Laboratory. ORL face database [Online], available: <http://www.cl.cam.ac.uk/Research/DTG/archive/facedatabase.html>, May 8, 2009

- 33 Donoho D, Stodden V. When does non-negative matrix factorization give a correct decomposition into parts? In: Proceedings of the Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2004. 1141–1148
- 34 Martinez A, Benavente R. The AR Face Database, Technical Report #24, Purdue University, USA, 1998
- 35 Ding C, Li T, Peng W, Park H. Orthogonal nonnegative matrix tri-factorization for clustering. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Philadelphia, USA: ACM, 2006. 126–135
- 36 Writing Group of Handbook of Mathematics. *Handbook of Mathematics*. Beijing: Higher Education Press, 1979 (《数学手册》编写组. 数学手册. 北京: 高等教育出版社, 1979)
- 37 Xu Sen-Lin, Xue Chun-Hua. *Mathematical Analysis (Volume 2)*. Beijing: Tsinghua University Press, 2006 (徐森林, 薛春华. 数学分析 (第二册). 北京: 清华大学出版社, 2006)
- 38 Chen Bao-Lin. *Theories and Algorithms of Optimization*. Beijing: Tsinghua University Press, 1989 (陈宝林. 最优化理论与算法. 北京: 清华大学出版社, 1989)
- 39 Bertsekas D P. *Nonlinear Programming (Second Edition)*. Massachusetts: Athena Scientific, 1999
- 40 Luenberger D G. *Introduction to Linear and Nonlinear Programming*. Massachusetts: Addison-Wesley Publication Company, 1973
- 41 Yuan Z J, Oja E. Projective nonnegative matrix factorization for image compression and feature extraction. In: Proceedings of the 14th Scandinavian Conference on Image Analysis. Joensuu, Finland: Springer, 2005. 333–342
- 42 Hoyer P O. Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, 2004, **5**(9): 1457–1469
- 43 Martinez A M, Kak A C. PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, **23**(2): 228–233
- 44 Turk M, Petland A. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 1991, **3**(1): 71–86
- 45 Bian Zhao-Qi, Zhang Xue-Gong. *Pattern Recognition*. Beijing: Tsinghua University Press, 2000 (边肇琪, 张学工. 模式识别. 北京: 清华大学出版社, 2000)



李乐 讲师, 博士. 主要研究方向为非负矩阵分解、非负矩阵集分解及它们在图像工程中的应用. 本文通信作者.

E-mail: lile05@mails.tsinghua.edu.cn
(LI Le Lecturer and Ph.D. at Tsinghua University. His research interest covers nonnegative matrix factorization, nonnegative matrix-set factorization and their applications on image engineering. Corresponding author of this paper.)



章毓晋 教授. 主要研究方向为图像工程及相关学科.

E-mail: zhangyj@ee.tsinghua.edu.cn
(ZHANG Yu-Jin Professor at Tsinghua University. His research interest covers image engineering and related subjects.)