

非负矩阵集分解

李 乐 章毓晋

(清华信息科学与技术国家实验室 北京 100084)

(清华大学电子工程系 北京 100084)

摘 要: 非负矩阵分解(Nonnegative Matrix Factorization, NMF)是一种新近被提出的方法,它以非线性的方式实现对非负多元数据的纯加性、局部化、线性和低维描述。NMF 可使数据中的潜在结构、特征或模式变得清晰,因此它作为一种有效的特征提取手段已被成功应用在许多领域的研究中。但是, NMF 的处理对象本质上是向量,用 NMF 处理数据矩阵集时要先将被处理矩阵集中的矩阵逐一矢量化,这常使对应的学习问题成为典型的小样本问题,从而使 NMF 结果的描述力不强、推广性差。为克服这两个问题,并保留 NMF 的好的特性,该文提出了非负矩阵集分解(Nonnegative Matrix-Set Factorization, NMSF),不同于 NMF 处理数据矩阵的矢量化结果, NMSF 直接处理数据矩阵本身。理论分析显示:处理数据矩阵集时, NMSF 会比 NMF 描述力强、推广性好。为了说明 NMSF 如何实现,也为了能对 NMSF 的性能做实验验证,构造了 NMSF 实现方式之一的基于双线性型的 NMSF(Bilinear Form-Based NMSF, BFBNMSF)算法。BFBNMSF 和 NMF 的比较实验结果支持了理论分析的结论。需要指出,更佳描述力和更好的推广性意味着 NMSF 比 NMF 更善于抓住数据矩阵的本质特征。

关键词: 图像描述; 特征提取; 非负矩阵集分解; 非负矩阵分解; 多元数据描述

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2009)02-0255-06

Nonnegative Matrix-Set Factorization

Li Le Zhang Yu-Jin

(Tsinghua National Laboratory for Information Science and Technology, Beijing 100084, China)

(Department of Electronic Engineering, Tsinghua University, Beijing 100084, China)

Abstract: Nonnegative Matrix Factorization (NMF) is a recently developed technique for nonlinearly finding purely additive, parts-based, linear, and low-dimension representations of nonnegative multivariate data to consequently reveal the latent structure, feature or pattern in the data. Although NMF has been successfully applied to several research fields, it is confronted with two main problems (unsatisfactory accuracy and bad generality) while the processed is a matrix-set, because the processed objects of NMF are intrinsically vectors and the necessary vectorization for every matrix in the processed matrix-set often make corresponding NMF learning to be typical small-sample learning. In this paper, Nonnegative Matrix-Set Factorization (NMSF) is conceived to overcome the problems and to retain NMF's good properties. As opposed to NMF, NMSF directly processes original data matrices rather than vectorization results of them. Theoretical analysis shows that while processing a data matrix-set, NMSF should be more accurate and has better generality than NMF. To show how to implement NMSF, and to validate NMSF's properties by experiments, Bilinear Form-Based NMSF (BFBNMSF) algorithm, as an implementation mode of NMSF, is formulated. Results of comparison experiments between BFBNMSF and NMF stably support the theoretical analysis. It is worth noting that higher accuracy and better generality actually means that NMSF is better at extracting essential features of data matrices than NMF.

Key words: Image representation; Feature extraction; Nonnegative Matrix-Set Factorization (NMSF); Nonnegative Matrix Factorization (NMF); Multivariate data representation

1 引言

在信号处理、神经网络、模式识别、计算机视觉以及图

像工程的研究中,如何构造一个能使多维观测数据被更好地描述出来的变换方法始终是一个有基础的重要性的问题。通常,一个好的变换方法应具备两个基本特性:(1)可以使数据的某种潜在结构变得清晰;(2)能使数据的维数得到一定程度的约减。

主分量分析、线性鉴别分析、投影寻踪、因子分析、冗

2007-08-03 收到, 2008-04-29 改回

国家自然科学基金(60573148)和教育部高等学校博士学科点专项科研基金(SRFDP-20060003102)资助课题

余归约和独立分量分析是一些最常用的数据变换方法。这些方法都可理解为在一定的限制下对数据进行变换或分解,不同的方法间因施加于其上的限制条件不同而有本质区别,但它们却有两个共同的特点:(1)允许负的分解量存在(即:允许减性的描述);(2)实现线性的维数约减。不同于它们,一种新的方法——非负矩阵分解(Nonnegative Matrix Factorization, NMF)^[1]由 Lee 和 Seung 在《Nature》上提出,它使分解后的所有分量均为非负值(即要求纯加性的描述),并且同时实现非线性的维数约减。

NMF 的心理学和生理学构造依据是对整体的感知是基于对组成整体的部分的感知构成的(感知的过程是纯加性的过程)^[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生物或智能数据描述的本质。此外,纯加性(或非负性)限制的引入导致了 NMF 分解结果在一定程度上的稀疏性^[1],稀疏的描述是处于完全分布式的描述和单一活跃分量的描述间的一种有效数据描述形式^[5]。

作为一种有效的、新兴的特征提取方法,近年来 NMF 在计算机视觉、信号处理、模式识别和图像工程等领域均获得了大量的应用。其原因可归结为:(1)稀疏的描述使对数据的解释变得方便(仅有少量分量活跃使数据的组成方式变得直观清晰)^[1,6];(2)纯加性的描述使对数据的分析显得合理(许多物理信号中不可能存在负的组成成分)^[1,6];(3)一般说来,相对稀疏的表示方式能在一定程度上抑制由外界变化(如:拍摄图像时可能遇到的物体被部分遮挡、光照变化和物体旋转等)给特征提取带来的不利影响^[7]。

在很多应用中,数据样本常以矩阵的形式呈现(如:灰度图像),一般的做法是先将这些数据矩阵先向量化,然后再放到 NMF 的模型中去处理^[1,6,7],这样做的结果是用于学习的样本量可能远远小于被学习参数的维数。这是统计学习中典型的小样本问题,它会导致:(1)NMF 对数据的描述能力不强(描述的准确度不够);(2)NMF 结果的推广性差。(图 1,图 2 直观诠释了上述问题,本文第 2 节会对此做详细的说明)。

为解决上述问题,并保留 NMF 的好的特性,本文提出非负矩阵集分解(Nonnegative Matrix-Set Factorization, NMSF)。不同于 NMF, NMSF 直接处理数据矩阵集中矩阵。

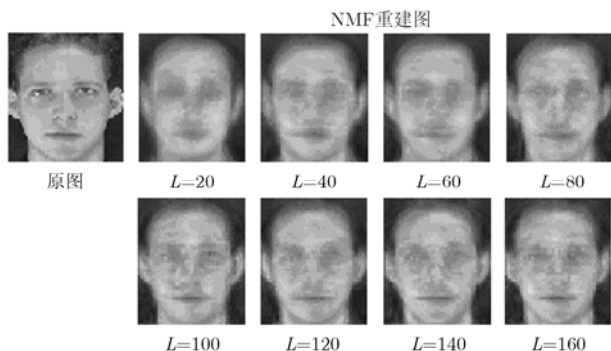


图 1 NMF 的描述力(以训练人脸像重建为例)



图 2 NMF 的推广性(以非训练人脸像重建为例)

NMSF 对数据矩阵集的描述力会强于 NMF,由它得出结果的推广性也会好于 NMF 结果的推广性。为了说明 NMSF 如何实现,也为了能对 NMSF 的性能做实验验证, NMSF 的一种实现形式——基于双线性型的非负矩阵集分解(Bilinear Form-Based NMSF, BFBNMSF)算法在文中将被构造。BFBNMSF 和 NMF 间的比较实验将被进行,比较结果将被用于验证对 NMSF 性质的理论分析结论。

本文余下各部分这样安排:第 2 节介绍 NMF 及其存在的问题;第 3 节提出并分析 NMSF;第 4 节构造 BFBNMSF 算法(作为 NMSF 的实现举例);第 5 节进行 BFBNMSF 和 NMF 间的比较实验,以此验证对 NMSF 的理论分析结论;第 6 节总结全文。

2 NMF 及其存在的问题

对一个 M 维的随机向量 \mathbf{v} 进行了 N 次的观测,记这些观测为 $\mathbf{v}_j, j=1,2,\dots,N$,取 $\mathbf{V}=[\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_N]$,其中 $\mathbf{V}_j = \mathbf{v}_j, j=1,2,\dots,N$ 。

定义 1 NMF 是发现非负的 $M \times L$ 的基矩阵 \mathbf{W} 和 $L \times N$ 的系数矩阵 \mathbf{H} 使^[1]

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

由于通常设定 $L \ll \min(M, N)$,所以只有在 \mathbf{W} 包含了随机变量 \mathbf{v} 的本质特征时,才可能使 $\mathbf{V} \approx \mathbf{WH}$ (即:用少量的数据可描述大量的数据)^[8]。

NMF 的实质是:在加性描述的限制下,在尽可能保持信息完整的情况下,将高维的随机模式 \mathbf{V} 简化为低维的随机模式 \mathbf{H} ,这种简化的基础是估计出数据中的本质结构 \mathbf{W} 。从代数的观点看, $\mathbf{V}_j \approx \sum_i \mathbf{W}_i H_{ij}$, \mathbf{W} 的列是基, \mathbf{H} 依 \mathbf{W} 的存在而存在。从机器学习的角度看, \mathbf{W} 包含了随机向量 \mathbf{v} 的某些本质特性,它除了要用于描述训练数据外还要用于描述非训练数据,且 \mathbf{W} 定了后依据一定的 \mathbf{V} 和 \mathbf{WH} 间差异度量准则, \mathbf{H} 也就定了,所以 \mathbf{W} 蕴涵了 NMF 学习结果的全部内容,它是学习过程中被学习的唯一参数。此外,无论依据哪种 \mathbf{V} 和 \mathbf{WH} 间差异度量准则, \mathbf{H} 均为 \mathbf{V} 在 \mathbf{W} 上

的非线性投影结果(或者说,对于随机向量 \mathbf{v} 的任意一次实现 \mathbf{v}_i 有 $\mathbf{v}_i \approx \mathbf{W}\mathbf{h}_i$, \mathbf{h}_i 为 \mathbf{v}_i 对 \mathbf{W} 做非线性投影的结果, \mathbf{v}_i 可为训练数据, 也可为非训练数据), 所以 NMF 体现了一种非线性的数据维数约减思路。

从计算的角度讲, NMF 可归结为对如下的优化问题求解:

$$\min_{\mathbf{W}, \mathbf{H}} f_V(\mathbf{W}, \mathbf{H}) \quad (2)$$

\mathbf{V} 为要被分解的矩阵, $f_V(\mathbf{W}, \mathbf{H})$ 刻画了 \mathbf{V} 与 $\mathbf{W}\mathbf{H}$ 间的差异性, f_V 可以是任何形式的距离度量或散度。已有的研究结果显示, 现有的 f_V 均不是 (\mathbf{W}, \mathbf{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^[7], 可行的办法是通过交替地优化 \mathbf{W} 和 \mathbf{H} 得到相应 NMF 问题的一个局部解^[1,7,8]。这里应强调, 交替优化过程中, 求 \mathbf{H} 只是为了辅助求 \mathbf{W} , 求 \mathbf{W} 是 NMF 的根本目的。

如果观测数据不是向量, 而是 $m \times n$ 维的矩阵 \mathbf{A}^k , $k = 1, 2, \dots, N$, 通常的做法是将它们矢量化为 $\text{vec}(\mathbf{A}^k)$, $k = 1, 2, \dots, N$, 再对 $[\text{vec}(\mathbf{A}^1), \text{vec}(\mathbf{A}^2), \dots, \text{vec}(\mathbf{A}^N)]$ 做 NMF 得到相应的 \mathbf{W} 和 \mathbf{H} ^[1,6,7], 这是因为 NMF 的处理对象本质上是向量。如前所述, \mathbf{W} 是 NMF 学习过程的被学习参数; 学习 \mathbf{W} 的过程是一个用 N 个样本去学习 $mn \times K$ 维的过程, 这常成为典型的小样本学习过程, 被学习对象的维数和样本量间差距巨大, 两者比为 $mn \times K / N$, 往往达数百甚至数千。根据统计学习理论, 小样本问题通常导致训练结果的精度不够、推广性差。因此, 处理数据矩阵集时, NMF 对数据矩阵集的描述能力常常不强, NMF 结果的推广性也往往很差, 图 1 和图 2 直观地揭示了这两个问题。图 1 显示了将 ORL 数据库^[9]中 400 幅图像进行 NMF 后, 利用得到的分解结果重建人脸图像的一个示例(L 是保留基的个数), 显然: 重建图像的质量虽然随基的个数增多而逐渐变好, 但即使保留 160 个基(此时压缩率仅为 2.41)重建效果也不能令人满意, 这说明 NMF 对数据矩阵集的描述能力不强。在图 2 中, 左边 3 幅从上到下分别取自 YALE 数据库^[10]、PIE 数据库^[11]和 UMIST 数据库^[12], 右边 3 幅从上到下依次为用构造图 1 时得到的那 160 个 NMF 基对左边 3 幅图进行重建的结果, 这里的 3 幅重建图的质量远差于图 1 中用这 160 个 NMF 基得到的重建图(图 1 中右下角那幅图)的质量, 这说明: NMF 对训练图的描述能力远远好于对非训练图的描述能力, 即: 此时 NMF 的推广性很差。需说明, 图 1 和图 2 是基于 Lee 和 Sueng 的(以欧几里德距离的平方为被分解矩阵和分解结果间差异度量的)基准算法^[8]得到的(文中后续实验结果也都基于这个算法得到), 当用其他算法时也一定会得到类似的结果, 因为 NMF 处理数据矩阵集时遇到的问题源于 NMF 本身, 不是依赖于具体算法。

3 NMSF

取 NMF 之长, 补 NMF 之短, 本文提出 NMSF, 它直

接处理数据矩阵集中的矩阵, 且仍对分解结果施加非负性的限制。

对一个 $m \times n$ 维的随机矩阵 \mathbf{A} 进行了 N 次的观测, 记这些观测为矩阵集 $\{\mathbf{A}^k, k = 1, 2, \dots, N\}$ 。

定义 2 NMSF 是发现非负的 $m \times l_1$ 的矩阵集 $\{\mathbf{B}_1^k, k = 1, 2, \dots, N\}$, 非负的 $l_1 \times l_2$ 的矩阵集 $\{\mathbf{B}_2^k, k = 1, 2, \dots, N\}$, ..., 非负的 $l_{K-2} \times l_{K-1}$ 的矩阵集 $\{\mathbf{B}_{K-1}^k, k = 1, 2, \dots, N\}$ 以及非负的 $l_{K-1} \times n$ 矩阵集 $\{\mathbf{B}_K^k, k = 1, 2, \dots, N\}$ 使

$$\mathbf{A}^k \approx \mathbf{B}_1^k \mathbf{B}_2^k \cdots \mathbf{B}_{K-1}^k \mathbf{B}_K^k, k = 1, 2, \dots, N \quad (3)$$

这里定义的是一个一般性的操作框架, K 的取值要依具体模型而定, 而且规定 $\exists c_1, c_2, \dots, c_{K-1} \in \{1, 2, \dots, K\}$ 使 $\mathbf{B}_{c_q}^1 = \mathbf{B}_{c_q}^2 = \cdots = \mathbf{B}_{c_q}^k = \mathbf{B}_{c_q}, q = 1, 2, \dots, K-1$, $c_q, q = 1, 2, \dots, K-1$ 的取值也要依具体的模型而定。

通常 $l_1, l_2, \dots, l_{K-1} \ll \min(m, n)$, 且 K 不很大, 那么只有 $\mathbf{B}_{c_q}, q = 1, 2, \dots, K-1$ 包含了随机矩阵 \mathbf{A} 的本质特征时, 才可能使 $\mathbf{A}^k \approx \mathbf{B}_1^k \mathbf{B}_2^k \cdots \mathbf{B}_{K-1}^k \mathbf{B}_K^k, k = 1, 2, \dots, N$ (即: 用很少量的数据可描述大量的数据)。

从整体上看, NMSF 把训练用的全部数据矩阵看成集合做整体处理, 提取出共同的描述矩阵 $\mathbf{B}_{c_q}, q = 1, 2, \dots, K-1$; 从局部上看, NMSF 对矩阵集中每一个矩阵都进行一次新形式的 NMF ($\mathbf{A}^k \approx \mathbf{B}_1^k \mathbf{B}_2^k \cdots \mathbf{B}_{K-1}^k \mathbf{B}_K^k$)。

NMSF 和 NMF 间有本质的区别, 它们是两类完全不同的方法, 因为 NMF 的处理对象本质上是向量(对于随机向量 \mathbf{v} 的任意一次实现 \mathbf{v}_i 有 $\mathbf{v}_i \approx \mathbf{W}\mathbf{h}_i$), 而 NMSF 的处理对象本质上是矩阵(对于随机矩阵 \mathbf{A} 的任意一次实现 \mathbf{A}^i , 有 $\mathbf{A}^k \approx \mathbf{B}_1^k \mathbf{B}_2^k \cdots \mathbf{B}_{K-1}^k \mathbf{B}_K^k$); NMSF 和 NMF 的性质会有相近之处, 因为它们都对分解结果施加了非负的限制。

此外, NMSF 继承了 NMF 将高维的随机模式 ($\{\mathbf{A}^k, k = 1, 2, \dots, N\}$) 简化为低维的随机模式 ($\{\mathbf{B}_{c_k}^k, c_k = \{1, 2, \dots, K\} / \{c_1, c_2, \dots, c_{K-1}\}, k = 1, 2, \dots, N\}$) 和非线性数据维数约减的特点。

从机器学习的角度看, $\mathbf{B}_{c_q}, q = 1, 2, \dots, K-1$ 包含了随机矩阵 \mathbf{A} 的某些本质特性, 它们除了要用于描述训练数据外还要被用于描述非训练数据, 且 $\mathbf{B}_{c_q}, q = 1, 2, \dots, K-1$ 定了后依据一定的 \mathbf{A}^k 和 $\mathbf{B}_1^k \mathbf{B}_2^k, \dots, \mathbf{B}_{K-1}^k \mathbf{B}_K^k$ 间差异度量准则 $\mathbf{B}_{c_k}^k, c_k = \{1, 2, \dots, K\} / \{c_1, c_2, \dots, c_{K-1}\}$ 也就定了, 所以 $\mathbf{B}_{c_q}, q = 1, 2, \dots, K-1$ 蕴涵了 NMSF 学习结果的全部内容, 它们是学习过程中被学习的参数。

从认知学的角度分析, NMSF 继承了 NMF 的构造特点, 考虑到了感知的特性(非负性), NMSF 模型也要求所有分解量均为非负值, $\mathbf{B}_{c_q}, q = 1, 2, \dots, K-1$ 代表了特征, $\mathbf{B}_{c_k}^k, c_k = \{1, 2, \dots, K\} / \{c_1, c_2, \dots, c_{K-1}\}$ 代表了因感知 \mathbf{A}^k 而对 $\mathbf{B}_{c_q}, q = 1, 2, \dots, K-1$ 的响应强度。

从计算角度讲, NMSF 可归结为对如下的优化问题的求解:

$$\min_{\mathbf{B}_1^k, \mathbf{B}_2^k, \dots, \mathbf{B}_{K-1}^k, \mathbf{B}_K^k, k=1, 2, \dots, N} \sum_{k=1}^N f_{A^k}(\mathbf{B}_1^k, \mathbf{B}_2^k, \dots, \mathbf{B}_{K-1}^k, \mathbf{B}_K^k) \quad (4)$$

\mathbf{A}^k 为要被分解的矩阵, $f_{\mathbf{A}^k}(\mathbf{B}_1^k, \mathbf{B}_2^k, \dots, \mathbf{B}_{K-1}^k, \mathbf{B}_K^k)$ 刻画了 \mathbf{A}^k 与 $\mathbf{B}_1^k \mathbf{B}_2^k, \dots, \mathbf{B}_{K-1}^k \mathbf{B}_K^k$ 间的差异性, $f_{\mathbf{A}^k}$ 可以是任何的距离度量或散度。学习 $\mathbf{B}_{c_q}, q=1, 2, \dots, K-1$ 的过程是用 N 个样本学习 $m \times l_1 + l_1 \times l_2 + \dots + l_{K-2} \times l_{K-1} + l_{K-1} \times n - \text{size}(\mathbf{B}_{c_K}^k)$ 维参数的过程 ($\text{size}(\mathbf{B}_{c_K}^k)$ 指 $\mathbf{B}_{c_K}^k$ 的行维与列维的乘积), 只要 K 不很大, N 不非常小, 用于学习的样本量和被学习对象的维数间差距就不会很大, 这里的学习问题就不再是小样本学习问题。并且, 无论 N 为多少, 只要 K 不非常大, 此时样本量和被学习参数的维数之比一定远小于做 NMF 时的样本量和被学习参数的维数之比。所以, 在处理数据矩阵集时, NMSF 会比 NMF 描述准确、推广性好。

4 NMSF 实现举例——BFBNMSF 算法

上节定义的 NMSF 为解决 NMF 处理数据矩阵集时遇到的问题提供了一般性的操作框架, 在这个框架下, 可以实现多个算法。本节实现了一个——BFBNMSF 算法, 以把上述的 NMSF 定义具体化, 并基于此用实验来验证上述理论分析的正确性。

NMSF 的实现包括两个方面的问题: (1)代数模型; (2)优化算法。

4.1 BFBNMSF 模型

对一个 $m \times n$ 维的随机矩阵 \mathbf{A} 进行了 N 次观测, 记这些观测为矩阵集 $\{\mathbf{A}^k, k=1, 2, \dots, N\}$ 。依 NMSF 的定义, 此时取 $K=3, c_1=1, c_2=3, c_3=2, \mathbf{B}_1^k = \mathbf{L}, k=1, 2, \dots, N, \mathbf{B}_2^k = \mathbf{D}^k, k=1, 2, \dots, N$ 和 $\mathbf{B}_3^k = \mathbf{R}, k=1, 2, \dots, N$, 有如下的 BFBNMSF 定义。

定义 3 BFBNMSF 是发现非负的 $m \times l_1$ 描述矩阵 \mathbf{L} , 非负的 $l_2 \times n$ 描述矩阵 \mathbf{R} 以及非负的 $l_1 \times l_2$ 的双线性型矩阵集 $\{\mathbf{D}^k, k=1, 2, \dots, N\}$ 使:

$$\mathbf{A}^k \approx \mathbf{L} \mathbf{D}^k \mathbf{R}, k=1, 2, \dots, N \quad (5)$$

因为式(5)右边的表达式与双线性型对应的矩阵表达^[13]一致, 所以把式(5)表示的 NMSF 模型命名为 BFBNMSF。

需要强调, BFBNMSF 是 NMSF 的一种形式, 因此 BFBNMSF 具有 NMSF 的通性, 它应比 NMF 描述力强、推广性好。 \mathbf{L} 和 \mathbf{R} 代表了整个矩阵集的共同特性, \mathbf{D}^k 依 \mathbf{L} 和 \mathbf{R} 而确定, \mathbf{L} 和 \mathbf{R} 是 BFBNMSF 训练过程中被学习的参数, 它们蕴涵了 BFBNMSF 学习结果的全部。

4.2 BFBNMSF 算法

取式(4)中的 $f_{\mathbf{A}^k}$ 为欧几里德距离的平方, 那么, BFBNMSF 可归结为如下的优化问题:

$$\min_{0 \leq \mathbf{L}, \mathbf{R}, \mathbf{D}^k, k=1, 2, \dots, N} \sum_{k=1}^N \left\| \mathbf{A}^k - (\mathbf{L} + \mathbf{EPS}_L) \cdot (\mathbf{D}^k + \mathbf{EPS}_{D^k}) (\mathbf{R} + \mathbf{EPS}_R) \right\|_F^2 \quad (6)$$

\mathbf{EPS}_X 是与 \mathbf{X} 同大小的矩阵或向量, 它中的所有元素均为 eps (一个极小数)。引入 $\mathbf{EPS}_L, \mathbf{EPS}_{D^k}, k=1, 2, \dots, N$ 和 \mathbf{EPS}_R 是为了避免算法推导和执行过程中出现数学上无定义的情况(除以零)。

依上述优化模型, 对 \mathbf{L}, \mathbf{R} 以及 $\mathbf{D}_1^1, \mathbf{D}_2^1, \dots, \mathbf{D}_{l_2}^1, \mathbf{D}_1^2, \mathbf{D}_2^2, \dots, \mathbf{D}_{l_2}^2, \dots, \mathbf{D}_1^N, \mathbf{D}_2^N, \dots, \mathbf{D}_{l_2}^N$ 采用交替优化的策略, 可推得了下文所示的单调下降地收敛的 BFBNMSF 算法(为控制篇幅, 也因本文重点是提出并分析 NMSF, 而非构造 BFBNMSF 算法, 所以推导过程暂被省略)。其中, $\mathbf{H}^k = (\mathbf{D}^k + \mathbf{EPS}_{D^k})(\mathbf{R} + \mathbf{EPS}_R), \mathbf{Q}^k = (\mathbf{L} + \mathbf{EPS}_L)(\mathbf{D}^k + \mathbf{EPS}_{D^k}), \mathbf{S} = (\mathbf{L} + \mathbf{EPS}_L)^T (\mathbf{L} + \mathbf{EPS}_L), \mathbf{T} = (\mathbf{R} + \mathbf{EPS}_R)(\mathbf{R} + \mathbf{EPS}_R)^T, \mathbf{U}^k = (\mathbf{L} + \mathbf{EPS}_L)^T \mathbf{A}^k (\mathbf{R} + \mathbf{EPS}_R)^T, \mathbf{V}_{\cdot c}^k = \mathbf{U}_{\cdot c}^k - \sum_{j=1, j \neq c}^{l_2} \mathbf{S} \cdot (\mathbf{D}_j^k + \mathbf{EPS}_{D_j^k}) T_{jc}; \cdot \times$ 和 $\cdot \div$ 分别代表 Hardward 乘和除(矩阵对应元素间的乘和除)。

BFBNMSF 算法步骤

输入 $\{\mathbf{A}^k, k=1, 2, \dots, N\}, l_1, l_2$

(1)随机初始化 $\mathbf{L}, \mathbf{R}, \{\mathbf{D}^k, k=1, 2, \dots, N\} > 0$;

(2)重复如下操作, 直到收敛;

(3) 执行 $\mathbf{L} \leftarrow \mathbf{L} \cdot \times \sum_{k=1}^N \mathbf{A}^k \mathbf{H}^{kT} \cdot \div \left\{ [\mathbf{L} + \mathbf{EPS}_L] \cdot \sum_{k=1}^N \mathbf{H}^k \mathbf{H}^{kT} \right\}$;

(4) 执行 $\mathbf{R} \leftarrow \mathbf{R} \cdot \times \sum_{k=1}^N \mathbf{Q}^{kT} \mathbf{A}^k \cdot \div \left\{ \sum_{k=1}^N \mathbf{Q}^{kT} \mathbf{Q}^k [\mathbf{R} + \mathbf{EPS}_R] \right\}$;

(5) k 从 1 到 N 做;

(6) c 从 1 到 l_2 做;

(7)如果 $\forall i \in \{1, 2, \dots, l_1\}, V_{ic}^k \geq 0$, 执行 $\mathbf{D}_{\cdot c}^k \leftarrow \mathbf{V}_{\cdot c}^k \cdot \times \mathbf{D}_{\cdot c}^k \cdot \div \{ST_{cc}[D_{\cdot c}^k + \mathbf{EPS}_{D_{\cdot c}^k}]\}$;

(8)否则, q 从 1 到 l_1 做;

(9)如果 $\left\{ V_{qc}^k - \sum_{j=1, j \neq q}^{l_1} [ST_{cc}]_{qj} (D_{jc}^k + \text{eps}) \right\} / [ST_{cc}]_{qq} \geq \text{eps}$, 执行 $\mathbf{D}_{qc}^k \leftarrow \text{eps} + \left\{ V_{qc}^k - \sum_{j=1, j \neq q}^{l_1} [ST_{cc}]_{qj} (D_{jc}^k + \text{eps}) \right\} / [ST_{cc}]_{qq}$;

(10)否则, $\mathbf{D}_{qc}^k \leftarrow 0$ 。

输出 $\mathbf{L}, \mathbf{R}, \{\mathbf{D}^k, k=1, 2, \dots, N\}$ 。

5 BFBNMSF 和 NMF 的比较实验

本节将进行 BFBNMSF 和 NMF 的比较实验, 比较内容为描述力和推广性。进行比较实验的目的是验证对 NMSF 的理论分析结论(NMSF 比 NMF 描述力强、推广性好)。

5.1 描述力

比较 BFBNMSF 和 NMF 间在描述力上的差异, 需设定统一的参照指标, 这个指标为压缩率最合理。假设被分解数据矩阵集为 $\{\mathbf{A}^k, k=1, 2, \dots, N\}$, \mathbf{A}^k 是 $m \times n$ 维的, 根据压缩率的定义^[14], 做 BFBNMSF 时的压缩率为 $mnN / [(ml_1 + nl_2) + l_1 l_2 N]$, 做 NMF 时的压缩率为 $mnN / (mL + LN)$ 。

基于 ORL 人脸数据库的 400 幅图像, 两种被比较方法的描述力(用被处理后的数据与原始数据间的相对误差说明)分别在不同压缩率(压缩率约为 2 的 1 次到 7 次幂)下被计算, 图 3 勾画了这两种方法的 \log_2 压缩率——平均相对误差曲线。图 3 的纵坐标为平均相对误差, 这是指: 为了客观, 图

上有关 BFBNMSF 和 NMF 的每个数据点都是相同实验设置下 10 次随机初始化得到的 10 个解算得的相对误差的平均值。BFBNMSF 的曲线在图 3 中始终列于下方且远离 NMF 的曲线, 这说明: 同压缩率下, BFBNMSF 比 NMF 对矩阵集描述准确得多。图 4 对图 3 所记录的比较结果给出了图示, 这些图示直观地支持了上述的结论。

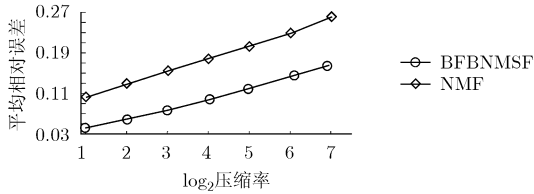


图 3 BFBNMSF 和 NMF 间的描述力比较(基于 ORL 人脸数据库)

5.2 推广性

本小节比较 BFBNMSF 和 NMF 间在推广性上的差异, 统一的参照指标是描述系数数量。相同描述系数数量下, BFBNMSF 对训练数据的描述力通常要差于 NMF(因为在相

同描述系数数量下, 做 BFBNMSF 时的压缩率常远高于做 NMF 时的压缩率), 如果此时 BFBNMSF 对测试数据的描述力好于 NMF, 那么说明 BFBNMSF 对训练数据的描述力与对测试数据的描述力间差异小于 NMF 对训练数据的描述力与对测试数据的描述力间差异, 即 BFBNMSF 比 NMF 推广性好。

图 5 以 ORL 库的 400 幅图为训练数据, 以 YALE 库、PIE 库和 UMIST 库的图像为测试数据, 给出了对这 2 种方法的推广性进行比较实验的示例。这时, 在同保留 225 个描述系数的情况下, 做 BFBNMSF 时的压缩率(44.3)远高于做 NMF 时的压缩率(1.71), BFBNMSF 对训练数据的描述力差于 NMF(见图 5 中“训练结果示例”下的重建图及其对应的重建相对误差); 但 BFBNMSF 对 YALE 库、PIE 库和 UMIST 库的图像的描述力远远好于 NMF(见图 5 中“推广性测试结果示例”下的重建图及其对应的重建相对误差)。由此得到的结论是: 训练数据为数据矩阵集时, BFBNMSF 结果的推广性比 NMF 结果的推广性要好得多。

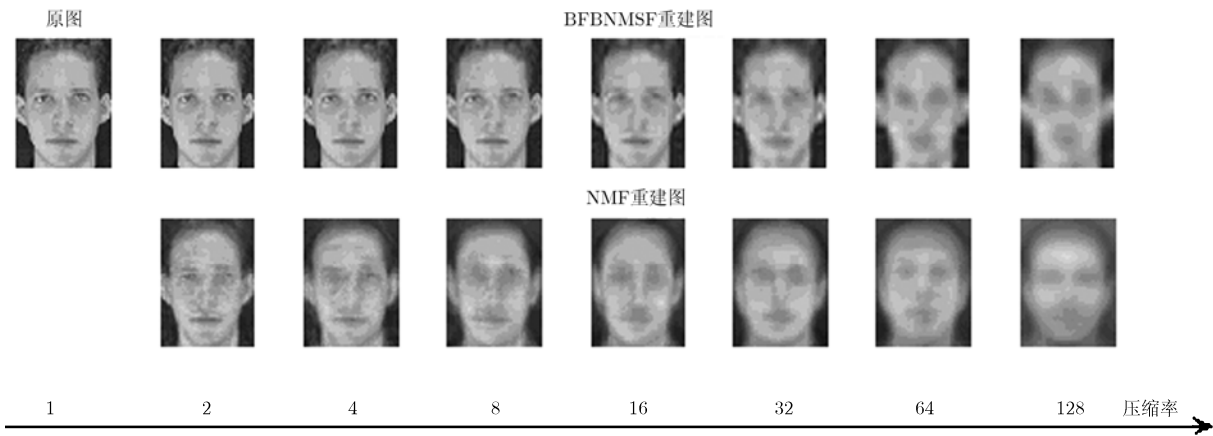


图 4 BFBNMSF 和 NMF 间的描述力比较示例(示例中原图取自 ORL 人脸数据库)

方法		训练结果示例			推广性测试结果示例									
		取自 ORL 库			取自 YALE 库			取自 PIE 库			取自 UMIST 库			
BFBNMSF	原图													
	重建													
	误差	0.101	0.125	0.120	0.117	0.093	0.101	0.105	0.107	0.121	0.152	0.141	0.115	
NMF	$L=225$ /压缩率=1.71													
	重建													
	误差	0.092	0.106	0.103	0.268	0.268	0.244	0.180	0.196	0.173	0.249	0.227	0.233	

注: 1. 训练数据为 ORL 库的全部 400 幅图像; 2. 压缩率指对训练数据的压缩率; 3. 误差指重建时的相对误差

图 5 BFBNMSF 和 NMF 间的推广性比较(训练数据为 ORL 人脸数据库)

6 结束语

非负矩阵分解(NMF)是一种新近被提出的有效的非负多元数据描述方法。NMF 的处理对象本质上是向量,为使数据矩阵集能适于用 NMF 处理,矩阵集中的矩阵要被逐一矢量化。这常使对应的 NMF 学习问题成为典型的小样本问题,从而使 NMF 结果的描述力不强、推广性差,最终大大抑止了在处理数据矩阵集时 NMF 的可用性。取 NMF 之长,补 NMF 之短,本文提出了非负矩阵集分解(NMSF),理论分析表明:处理数据矩阵集时,NMSF 会比 NMF 描述力强、推广性好。为了说明 NMSF 如何实现,也为了能对 NMSF 的性能做实验验证,NMSF 实现方式之一的基于双线性型的非负矩阵集分解(BFBNMSF)算法被构造。BFBNMSF 和 NMF 的比较实验结果坚实地支持了对 NMSF 的理论分析结论。需要指出,更强的描述力和更好的推广性意味着由 NMSF 提取出的特征要比由 NMF 提取出的特征更为本质。

本文中提出的 NMSF 为解决 NMF 处理数据矩阵集时遇到的问题提供了一个一般性的操作框架,不同的 NMSF 实现形式可被采用,本文实现的 BFBNMSF 算法是这些实现形式中一种。理论上,不同实现形式的 NMSF 将均能解决 NMF 处理数据矩阵集时遇到的问题,除此以外,它们将蕴涵不同的代数学、认知学和统计学解释,将有不同的特性。实现各种各样的蕴涵某种或某些期望特性的 NMSF 算法将是进一步的研究目标。

参考文献

- [1] Lee D D and Seung H S. Learning the parts of objects by non-negative matrix factorization. *Nature*, 1999, 401(6755): 788-791.
- [2] Wachsmuth E, Oram M W, and Perrett D I. Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque. *Cereb. Corte.*, 1994, 4(5): 509-522.
- [3] Biederman I. Recognition-by-components: A theory of human image understanding. *Psychol. Rev.* 1987, 94(2): 115-147.
- [4] Palmer S E. Hierarchical structure in perceptual representation. *Cogn. Psychol.*, 1977, 9(3): 441-474.
- [5] Field D J. What is the goal of sensory coding? *Neural Computation*, 1994, 6(4): 559-601.
- [6] Hoyer P O. Non-negative matrix factorization with sparseness constraints. *J. of Mach. Learning Res.*, 2004, 5(9): 1457-1469.
- [7] Heiler M and Schnorr C. Learning Sparse Representations by Non-Negative Matrix Factorization and Sequential Cone Program-ming. *J. of Mach. Learning Res.*, 2005, 7(7): 1385-1407.
- [8] Lee D D and Seung H S. Algorithms for non-negative matrix factorization. In *Adv. in Neur. Inform. Proc. Syst. (NIPS)*, Cambridge: MIT Press, 2000: 556-562.
- [9] ORL face database. <http://www.cl.cam.ac.uk/Research/DTG/attarchive/facedatabase.html>, 1992.
- [10] YALE face database. <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>, 1997.
- [11] Sim T, Baker S, and Bsat M. The CMU pose, illumination, and expression database. *IEEE Trans. on Pattern Anal. Machine Intell.*, 2003, 25(12): 1615-1618.
- [12] UMIST face database. <http://images.ee.umist.ac.uk/danny/database.html>, 1997.
- [13] 章贤科 许普华. 高等代数学. 第2版, 北京: 清华大学出版社, 2004, 第8章.
- [14] 章毓晋. 图像工程(上册——图像处理). 第2版, 北京: 清华大学出版社, 2006, 第11章.

李乐: 男, 1979年生, 讲师, 博士生, 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。
章毓晋: 男, 1954年生, 教授, 博士生导师, 研究方向为图像工程及相关学科。