非负矩阵分解算法综述

李 乐[1,2], 章毓晋[1,2]

(1. 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)

摘 要: 本文介绍了非负矩阵分解(Non-negative Matrix Factorization, NMF)的基本原理和性质, 将现有 NMF 算法分为了基于基本 NMF 模型的算法和基于改进 NMF 模型的算法两大类, 在此基础上较为系统地分析、总结和比较了它们的构造原则、应用特点以及存在的问题, 最后预测和分析了未来 NMF 算法研究的可能方向.

关键词: 非负矩阵分解; 多元数据描述; 特征提取

中图分类号: TP391 文献标识码: A 文章编号: 0372-2112 (2008) 04-0737-07

A Survey on Algorithms of Non-Negative Matrix Factorization

LI Le[1,2], ZHANG Yu-jin[1,2]

(1. *Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing* 100084, *China*;
2. *Department of Electronic Engineering, Tsinghua University, Beijing* 100084, *China*)

Abstract: The fundamentals and properties of non-negative matrix factorization (NMF) are introduced, and available NMF algorithms are classified into two categories: basic NMF model-based algorithms and improved NMF model-based algorithms. Based on these, the design principles, application characteristics, and existing problems of the algorithms are systematically discussed. Besides, some open problems in the development of NMF algorithms are presented and analyzed.

Key words: non-negative matrix factorization; multivariate data representation; feature extraction

1 引言

在信号处理、神经网络、模式识别、计算机视觉和图象工程的研究中, 如何构造一个能使多维观测数据被更好描述的变换方法始终是一个非常重要的问题. 通常, 一个好的变换方法应具备两个基本的特性: (1) 可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、冗余归约和独立分量分析是一些最常用的变换方法. 它们因被施加的限制不同而有着本质的区别, 然而, 它们有两个共同的特点: (1) 允许负的分解量存在 (允许有减性的描述); (2) 实现线性的维数约减. 区别于它们, 一种新的变换方法 ——非负矩阵分解(Nonnegative Matrix Factor, 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(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由对组成整体的部分的感知构成的(纯加性的)[2~6], 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数据描述的本质. 此外, 这种非负性的限制导致了相应描述在一定程度上的稀疏性[1], 稀疏性的表述已被证明是介于完全分布式的描述和单一活跃分量的描述 3 间的一种有效数据描述形式[7].

因为纯加性的和稀疏的描述能使对数据的解释变得方便(少量活跃的分量使数据的组成方式变得清晰)与合理 (许多物理信号中不可能存在负的成分)[1,8], 还因为相对稀疏性的表示方式能在一定程度上抑制由外界变化(如: 部分遮挡、光照变化和物体的旋转等) 给特征提取带来的不利影响[9], 所以 NMF 已逐渐成为信号处理、生物医学工程、模式识别、计算机视觉和图像工程等研究领域中最受欢迎的多维数据处理工具之一. 具体说, 它目前已被应用到文本分析与聚类、数字水印、人脸检测与识别、图像检索、图像复原、语言建模、声源分类、音乐信号分析与乐器识别、盲信号分离、网络安全、基因及细胞分析等的研究中.

从 NMF 被提出算起, 关于 NMF 的研究已走过了将近 7 个年头, 对现有算法的系统性综述已非常必要, 然而, 至今国内外尚无此类文章.

根据 NMF 模型中对分解结果的限制是否仅限于非负性, 本文将现有算法分为基本 NMF 算法(Basic NMF, BNMF, 基于基本 NMF 模型) 和改进 NMF 算法(Improved NMF, INMF, 基于改进 NMF 模型要依具体的期望特性对分解结果施加除非负外的其他的限制) 两大类. BNMF 在以往的(或最初的) 文献中常就称为 NMF, 这里用 BNMF命名为强调与 INMF 间的区别. 本文中, NMF 指

收稿日期: 2007-03-26; 修回日期: 2007-12-20

基金项目: 国家自然科学基金(No. 60573148)

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

BNMF和 INMF 的总和. BNMF 算法又可分为基于单个目标函数的算法和基于目标函数族的算法两类. INMF 算法也可再细分为三类: (1) 稀疏性增强的 NMF 算法; (2) 鉴别性NMF 算法; (3) 加权 NMF 算法.

2 NMF 简介

定义 对一个 M 维的随机向量 $\boldsymbol{v}$ 进行了 N 次的观测, 记这些观测为 $\boldsymbol{v}_j, j = 1, 2, \cdots, N$, 取 $\boldsymbol{V} = [V_{\cdot 1}, V_{\cdot 2}, \cdots, V_{\cdot N}]$, 其中 $V_{\cdot j} = \boldsymbol{v}_j, j = 1, 2, \cdots, N$, BNMF 要求发现非负的 $M \times L$ 的基矩阵 $\boldsymbol{W} = [W_{\cdot 1}, W_{\cdot 2}, \cdots, W_{\cdot N}]$ 和 $L \times N$ 的系数矩阵 $\boldsymbol{H} = [H_{\cdot 1}, H_{\cdot 2}, \cdots, H_{\cdot N}]$, 使 $\boldsymbol{V} \approx \boldsymbol{WH}$[1], 这也可以用向量标量积的形式更为直观地表示为 $V_{\cdot j} \approx \sum_{i=1}^{L} W_{\cdot i} H_{ij}$.

此外, BNMF 常被有盲信号分离背景的学者解释为含噪声项的产生式模型: $\boldsymbol{V} = \boldsymbol{WH} + \varepsilon$[10], ε 是 $M \times N$ 的噪声矩阵. 不同的 BNMF 算法也常可被解释为遵循了不同的 ε 分布假设下的最大似然算法.

根据需要, 可给上述模型中的 $\boldsymbol{W}$ 和 $\boldsymbol{H}$ 施加更多的限制, 构成 INMF.

由于通常设定 $L \ll \min(M, N)$, 即只用很少的基去描述大量的数据, 所以只有在 $\boldsymbol{W}$ 包含了随机变量的本质特征时, 才可能使 $\boldsymbol{V} \approx \boldsymbol{WH}$[11].

从几何的观点看, BNMF 可被认为是在正象限(广义的第一象限) 内去发现一个单形锥[12]; 从多元统计的观点看, NMF 是在非负等限制下, 在尽可能保持信息不变的情况下, 将高维的随机模式简化为低维的随机模式 $\boldsymbol{H}$, 而这种简化的基础是估计出数据中的本质结构 $\boldsymbol{W}$; 从代数的观点看, NMF 是发现数据的一种内在非负(或蕴涵更多性质的) 代数分解形式或表示方法; 从维数约减的观点看, 因为基矩阵 $\boldsymbol{W}$ 和系数矩阵 $\boldsymbol{H}$ 同时由 NMF 来确定, 系数矩阵 $\boldsymbol{H}$ 并非为数据矩阵 $\boldsymbol{V}$ 在 $\boldsymbol{W}$ 上的投影, 所以NMF 实现的是非线性的维数约减.

3 BNMF 算法

实现 NMF 的过程是一个优化求解的过程, Donoho 等从理论上分析了 BNMF 存在唯一解的条件[12], 这个条件的苛刻性告诉我们: 合理地构造一个目标函数, 以此交替地优化 $\boldsymbol{W}$ 和 $\boldsymbol{H}$ 从而得到 BNMF 的一个局部最优解才是进行 BNMF 的可行方法. 这也是目前 BNMF 算法构造的基本思想.

根据算法基于的目标函数的特点不同, BNMF 算法可分为基于单个目标函数的算法和基于目标函数族的算法两类.

3.1 基于单个目标函数的算法

Lee 和 Sueng 提出 NMF 时, 假设矩阵 $\boldsymbol{V}$ 的每个元素间是统计独立的且 V_{ij} 服从以 $[\boldsymbol{WH}]_{ij}$ 为参数的泊松分布, 以此假设下的似然函数为目标函数, 用最大似然(Maximum Likelihood, ML) 估计的思路构造了第一个NMF 算法(优化方法文中未介绍), 且称这个算法为 Poisson-ML 算法, 它具有收敛性[1].

Lee 和 Sueng 在文[11] 中为上述似然函数加了一常数项并取反得到了新的目标函数——广义 Kullback-Leibler 散度 (Generalized Kullback-Leibler Divergence, GKLD), GKLD 好于前述似然函数之处是: 当且仅当 $\boldsymbol{V} = \boldsymbol{WH}$, 它等于零. 这使对优化结果的评价变得容易, 因为直接查看所得结果对应的目标函数值大小就可知其优劣. 这样的特性也成为之后(构造 BNMF 算法时) 被使用的目标函数所共有的特性. Lee 和 Sueng 采用了类似于 EM 算法中使用的优化策略去交替式地优化 GKLD, 得到了单调下降收敛的算法[11], 且称这个算法为 GKLD-EM 算法(前半部分代表目标函数, 后半部分代表优化方法或模型, 此节中之后部分都依此法对算法命名). GKLD-EM 算法可以解释为迭代步长自学习的梯度下降算法[11], 它很好地调和了算法效率和实现简单性间的矛盾, 被普遍认为是 BNMF 的两个基准算法之一, 因此被研究和应用较多. 还需指出, GKLD-EM 算法实际也可看成对 $\boldsymbol{W}$ 和 $\boldsymbol{H}$ 交替实施了图像复原领域中著名的 EMML 算法[10, 13].

GKLD 不是距离测度, 它无对称性, 故还可用 Lee 和 Sueng 在文献[11] 中定义的 GKLD 的对偶形式为目标函数. Cichocki 等把这个目标函数称为对偶的 GKLD (Dual GKLD, DGKLD), 并利用指数梯度下降(Exponential Gradient Descent, EGD) 原则构造了可收敛的算法[10, 14], 且称这个算法为 DGKLD-EGD 算法, 它可认为是交替式的 SMART 算法[10, 14, 15]. DGKLD-EGD 算法的构造简单, 易于理解, 但其中涉及了两个学习率参数矩阵, 对它们选择得如何对算法性能会有一定的影响.

Lee 和 Sueng 还提出了利用 $\boldsymbol{V}$ 和 $\boldsymbol{WH}$ 间的欧几里德距离的平方(Square of Euclidian Distance, SED) 作为目标函数[11], SED 与假定 ε 的各元素独立同标准正态分布时的似然函数有一一对应的关系[16]. 相比 GKLD, 它更被人们熟悉, 它的数学性质也被研究得更为充分. 此时, 采用类似于 EM 算法中使用的优化策略同样可得一单调下降地收敛的算法[11], 且称这个算法为 SED-EM 算法. 与 GKLD-EM 算法类似, SED-EM 算法也很好地调和了算法效率和实现简单性间的矛盾, 它是 BNMF 的两个基准算法中另一个. SED-EM 算法是目前被研究和应用最多的 BNMF 算法. 还需指出: SED-EM 算法实际可看成对 $\boldsymbol{W}$ 和 $\boldsymbol{H}$ 交替实施了图像复原领域中著名的 ISRA 算法[10, 13]. 此外, Wild 等考虑了利用球面 K 均值聚类作为 SED-EM 算法的初始化步骤(一般情况下, NMF 算法采

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

用随机初始化. 其实, 球面 K 均值聚类应也可作为其他 BNMF 算法的初始化步骤), 这样做的好处是能使 SED-EM 算法的效率有所提高, 但这以收敛到相对不好的局部解为代价[17].

Heiler 等考虑了 SED 的展开形式, 把对 $\boldsymbol{W}$ 和 $\boldsymbol{H}$ 的优化求解归结为经典的凸二次规划(Convex Quadratic Programming, CQP) 问题, 构造了也可单调下降地收敛的算法[9], 且称这个算法为 SED-CQP 算法. SED-CQP 算法的优点是可借助成熟的凸二次规划方法(有大量现成程序), 因此算法设计中的推导极其简单(只是把问题化为凸二次规划模型即可), 非常便于理解(应是目前最好理解的NMF 算法); 它的缺点是只有利用专业的优化程序时算法的效率才可与 SED-EM 算法相当(Heiler 利用 Mosek 3.1 软件实现 SED-CQP 算法, 得到了 SED-CQP 算法和 SED-EM 算法的效率相仿的结论), 当利用非专业性的优化程序时算法效率可能远低于 SED-EM 算法(我们利用 MATLAB 自带的二次规划程序实现 SED-CQP 算法时, SED-CQP 算法的效率远低于 SED-EM 算法).

Cichocki 等利用 SED 分别为 $\boldsymbol{W}$ 和 $\boldsymbol{H}$ 的凸函数的性质, 构造了这样操作的算法: (1) 求得无非负约束下的解析解(定点(fixed point, FD) 算法的特点); (2) 通过做非线性投影(Nonlinear Projection, NP) 满足非负性的要求[14], 且称此算法为 SED-FD+ NP 算法. SED-FD+ NP 算法构造思路清晰直观, 便于理解, 但它产生的迭代序列不单调下降且其收敛性无法保证, 因此算法的实施策略和终止条件的选择将对算法性能有较大影响.

陈卫刚等用可行方向(Feasible Direction, FD) 法和模拟退火(Simulated Annealing, SA) 法结合优化 SED[18], 且称这样构造出的算法为 SED-FD+ SA 算法. SED-FD+ SA 算法收敛, 且实验表明由其得到的分解结果要比由 SED-EM 算法得到的分解结果更优[18], 但 SED-FD+ SA 算法的复杂度会高于 SED-EM 算法较多(FD 法中的一维搜索步骤以及 SA 方法都非常耗时).

表 1 对上述算法的主要性能、特点做了总结, 其中也补充了对上述算法应用现状的说明.

3.2　基于目标函数族的算法

对比基于单个目标函数的算法, 基于目标函数族的算法的共同特点是: 依目标函数族中的参数或自由设定函数的不同, 他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)[10], 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.

Compass 考虑以 β 散度(β Divergence, βD) 作为目标函数, 并以步长动态调整的梯度下降(Gradient Descent with Adaptive Iterative Step, GDAIS) 原则构造了相应的算法[19], 且称这个算法为 βD-*GDAIS* 算法. βD 代表了一族目标函数, 作为特例, 当 $\beta \to 0$ 时它是 GKLD, 当 $\beta \to -1$ 时它是 Itakura-Saito 距离, 当 $\beta = 1$ 它是 SED[14,19]. βD-GDAIS 算法的收敛性无法保证, 所以算法实施策略和终止条件的确定对其性能有较大影响.

Cichocki 和 Zdunek 提出了用性质与 βD 类似的 α 散度(αDivergence, αD) 作为目标函数, αD 也包含了一族目标函数, 作为特例, 当 $\alpha = 2$ 时它是 Pearson 距离, 当 $\alpha = 0.5$ 时它是 Hellinger 距离, 当 $\alpha = -1$ 时它是 Neyman's χ^2 距离, 当 $\alpha \to 1$ 时它是 GKLD, 当 $\alpha \to 1$ 时它是 DGKLD[10,14]. 他们先是采用线性变换投影梯度(Linearly Transformed Projected Gradient, LTPG) 法去优化 αD[10], 得到的算法且称为 αD-LTPG 算法. 从算法的收敛性上说, αD-LTPG 算法优于 βD-GDAIS 算法但不如 GKLD-EM 算法、SED-EM 算法 和 SED-CQP 算法, 因为它收敛[10, 14], 但由它产生的迭代序列并非单调下降. 为了提高算法效率, Zdunek 和 Cichocki 还考虑了用拟牛顿(Quasi-Newton, QN) 法和非线性投影(NP) 技术组合对 αD 优化[20], 得到的算法且称为 αD- QN+ NP 算法. αD- QN+ NP 算法的理论效率高于 αD-LTPG 算法, 但实现起来复杂, 且收敛性无法保证(但由拟牛顿法的特点考虑, 经过一定的迭代步数后, 由其应能得到一个较令人满意解), 实施策略和终止条件的设定对其最终性能有一定影响.

Dhillon 等把比 αD 和 βD 更为一般性的 Bregman 散度(Bregman Divergence, BD) 作为了目标函数, 说 BD 更具一般性是因为在它的表达式中包含了一个可自由设定的函数 $\Psi(\cdot)$, 而不是某个可调参数[21]. 利用类似于 GKLD-EM 算法和 SED-EM 算法的构造思路, 得到的算法[21] 且称为 BD-EM 算法. Dhillon 等也考虑了与前述 BD 在形式上对偶的目标函数, 可称之为对偶的 BD(Dual BD, DBD). 此时, 利用基于 KKT 条件的优化策略得到的算法[21] 且称为 DBD-KKT 算法. 类似于 GKLD-EM 算法和 SED-EM 算法, BD-EM 算法收敛且由其得到的迭代序列单调下降. 一些实验表明由 DBD-KKT 算法得到的迭代序列也是单调下降的, 但这以及它的收敛性在数学上都很难被严格证明[21]. 然而, 根据 DBD-KKT 算法的构造原则可定性地说, 经过一定的迭代步数后, 由其可得到一个较令人满意解. 此外, BD-EM 算法和 DBD-KKT 算法因每步迭代涉及的运算较多而复杂度高于 GKLD-EM 算法和 SED-EM 算法(它们两者间相比, DBD-KKT 算法的复杂度要低些[21]). 还应指出, $\Psi(\cdot)$ 的具体形式并未确定, 需使用者根据具体情况来确定, 这对使用者提出了非常高的要求.

为便于与基于单个目标函数的算法比较, 对上述算法的主要性能、特点的总结以及对上述算法应用现状的说明也置于表 1 中.

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

表 1 对现有 BNMF 算法的总结与分析

算法类别	算法	实现难度	收敛性	对使用者要求	应用情况	主要特点评述	算法定位
基于单个目标函数的算法	PoissonML[1]	低	收敛	低	较好	严格的最大似然估计操作	先驱算法
	GKLD-EM[11]	低	收敛	低	好	简单性和效率间的较好折中	基准算法
	DGKLD-ED[10,14]	较低	收敛	较低	应用少	设计思路简单直接,便于理解	无公认定位
	SED-EM[11]	低	收敛	低	最好	简单性和效率间的较好折中	基准算法
	SED-CQP[9]	低	收敛	低	应用少	非常便于理解和实现	无公认定位
	SED-FD+ NP[14]	低	可能震荡	低	应用少	设计思路简单直接,便于理解	无公认定位
	SED-FD+ SA[18]	中	收敛	较低	应用少	可能得到很好的结果,但速度较慢	无公认定位
基于目标函数族的算法	βD-GDAIS[14,19]	低	可能震荡	较高	应用少	适用性变广,应用难度增加	无公认定位
	αD-LTPG[10]	低	可能震荡	较高	应用少	适用性变广,应用难度增加	无公认定位
	αD-QN+ NP[20]	较高	可能震荡	较高	应用少	适用性变广,应用难度增加	无公认定位
	BD-EM[21]	较高	收敛	高	应用少	适用性变广,应用难度明显增加	无公认定位
	DBD-KKT[21]	较高	可能震荡	高	应用少	适用性变广,应用难度明显增加	无公认定位

4 INMF 算法

上节讨论的 BNMF 算法对分解结果仅施加了非负性的限制,为使分解结果能满足更多的被期望性质,新的限制常被添加到原有的 BNMF 模型中,这样形成了大量的 INMF 算法. INMF 算法大体可分为三类:(1)稀疏性增强的 NMF 算法;(2)鉴别式 NMF 算法;(3)加权 NMF 算法.

4.1 稀疏性增强的 NMF 算法

BNMF 能无监督地导致相对稀疏的或局部化的描述,这是它最主要的特点或优点之一,但有时它导致的描述的稀疏程度并不令人满意,因此稀疏性增强的 NMF 算法被广泛研究,它是 INMF 算法研究中最为活跃的方面.

Cichocki 等把稀疏性罚项和 SED 结合作为目标函数,对 SED-EM 算法的迭代规则做修改,并引入非线性投影技术保证非负性,得到了稀疏性增强的 SED-EM 算法[14];他们同样是把稀疏性罚项和 SED 结合作为目标函数,更为直接地(与原算法的推导步骤完全一致)改进了 SED-FD+ NP 算法[14]. Compass 的 βD-GDAIS 算法也可直接被修改,去寻求更为稀疏的描述[10,14,18],其思路类似于 Cichocki 等修改 SED-EM 算法的思路.以上这些工作的共同特点是:它们都只是对原有 NMF 算法的迭代规则做了一些简单的修改,这种修改上的简单性的代价是得到的改进算法均可能发散(无论原 NMF 算法是否收敛).

Cichocki 等还以对 GKLD-EM 算法的每步迭代结果均增加了一次指数运算的方式去构造稀疏性增强的 GKLD-EM 算法[14],这里指数运算的作用是使小于 1 的数更小,大于 1 的数更大,从而使整体上变得更为稀疏(当然,对 SED-EM 算法也可做类似的修改).这是一种启发式的改进思路,其实现非常简单,但改进后的算法可能失去收敛性.

Li 等构造了局部 NMF(Local NMF, LNMF)算法[22],基于原有的 GKLD-EM 算法,他们的关键改进是为 $\boldsymbol{W}$ 引入了列正交的限制,这能使 $\boldsymbol{W}$ 更为稀疏化(在 $\boldsymbol{W}$ 非负的前提下,列正交一定导致稀疏),但 $\boldsymbol{H}$ 变得非常不稀疏是 W 更为稀疏化的代价[23,24].此外,有实验表明:LNMF 算法对数据的描述力较差[7],这应与其为使对 $\boldsymbol{H}$ 的优化形式简单而采取了较多近似表达有关. Xu 和黄钢石等提出了受限 NMF (constrained NMF, CNMF)算法[25,26],CNMF 算法与 LNMF 算法的区别仅在于对罚项的加权策略不一样.

Hoyer 把稀疏编码和 NMF 结合,用 SED 和由 1 范数定义的稀疏性罚项组合为目标函数,对 $\boldsymbol{W}$ 采用梯度投影法优化,对 $\boldsymbol{H}$ 采用类似 EM 算法的方式优化,构造了非负稀疏编码(Nonnegative Sparse Coding, NSC)算法[27]. Liu 提出了稀疏 NMF(Sparse NMF, SNMF)算法[28],他们的思路与 Hoyer 非常类似,区别仅在于:(1)他们选择的目标函数是以 GKLD 和由 1 范数定义的稀疏性罚项组合而成的;(2)他们对 $\boldsymbol{W}$ 和 $\boldsymbol{H}$ 都采用了类似 EM 算法的方式优化,得到的算法形式与 GKLD-EM 算法和 SED-EM 算法类似.

Hoyer 对自己和 Liu 的工作做进一步的发展,提出了可精确控制稀疏性的算法——NMF with sparseness constraints(NMFSC)[8],这个算法以 SED 为目标函数,它最具特点也是最具创新性的地方是以非线性投影实现对稀疏性的精确控制,但应指出:对 $\boldsymbol{W}$ 和 $\boldsymbol{H}$ 同时施加较高的稀疏性限制时,NMFSC 对数据的描述力常常很差[23]. Stadlthanner 等证明了 Hoyer 提出的非线性投影的唯一性[29],但需说明:当被投影点与期望投影区域距离很远时,这个非线性投影法可能遭遇较严重的数值问题.

Heiler 等对 NMFSC 又做了发展,他们仍以 SED 为目标函数,把一个稀疏性区间对应的可行域表示为不同二阶锥间的集合运算的结果,并以此和非负性作为约束条件,在二阶锥规划的框架下构造了两个可把稀疏

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

性控制在一定范围内的算法[9], 这两个算法中一个效率较高但可能震荡, 另一个效率稍低但能保证收敛.

当NMF 模型由 $V \approx WH$ 变为 $V \approx WSH$, 因为 WSH 被期望为 V 的很好近似, 所以如果 S 是非常平滑的, 那么 W和H 都应是非常稀疏. Pascual-Montano 等基于这样的分析利用 SED-EM 算法的框架构造了非平滑 NMF (Non-smooth NMF, NSNMF) 算法[23]. 实验表明, NSNMF 算法很好地调和了分解结果的稀疏性和表述力间的矛盾[23].

Liu 等的工作不同于前人, 他们理论证明并实验验证了通过对保留的维数及归一化过程的合理设定可使由 GKLD-EM 算法得到的解更为稀疏[30].

表 2 总结了上述算法, 并补充强调了各算法进行稀疏性增强处理的对象.

表 2　对现有稀疏性增强的 NMF 算法的总结与分析

典型算法	稀疏性增强对象	稀疏性增强的方式	主要特点
Cichock 等[14]和 Compass 等[19]的算法	W、H 或同时 W 和 H	施加稀疏性限制	直接修改原算法, 修改简单但得到的算法可能震荡
Cichock 等的算法[19]	W、H 或同时 W 和 H	使每步迭代结果中小于 1 的数更小, 大于 1 的数更大	直接修改原算法, 修改非常简单但得到的算法可能震荡
LNMF[22]和 CNMF[25,26]	W	施加正交性限制	W 变得非常稀疏; H 变得不稀疏; WH 的表述力不够
NSC[27]和 SNMF[28]	H	施加用 1 范数表述的稀疏性限制	NMF 和稀疏编码的有机结合
NMFSC[8]	W、H 或同时 W 和 H	构造能精确控制稀疏性的非线性投影方法	稀疏性被精确控制
Heiler 的算法[9]	W、H 或同时 W 和 H	把稀疏性约束转化为二阶锥规划模型的约束条件, 用二阶锥规划模型求解	稀疏性可被控制在一定范围内
NSNMF[23]	同时 W 和 H	把 NMF 模型化转为 $V \approx WSH$, 并给 S 较高的平滑度	很好地调和了分解结果稀疏性和表述力间的矛盾
Liu 等的算法[30]	H	合理设定归一化过程以及保留的维数	思维独特, 不改变 NMF 算法本身

4.2　鉴别性 NMF 算法

BNMF 的结果不带有任何类别信息, 因此当用于分类问题时, 加入鉴别性的 NMF 更受关注. Wang 等把 Fisher 鉴别信息(类内散度和类间散度的差) 作为罚项加入到 GKLD 中构成目标函数, 仿效 GKLD-EM 算法的构造方式构造了 Fisher NMF(FNMF) 算法[31]. FNMF 在对数据进行非负分解的同时, 最大化了类内散度和类间散度的差从而使分解结果中蕴涵了较多的类别信息. 此外, 实验表明 FNMF 导致的分解结果也是非常稀疏化的[31]. Xue 等的工作与 Wang 等的工作非常类似, 因为这两项工作基于同样的构造思想和目标函数, Xue 等的工作有新意的地方是由不同的推导方式得到了不同迭代形式的算法[32]. Buciu 等把 LNMF 和 FNMF 结合构造了稀疏性更强的类监督 NMF 算法[33]. 需指出上述三种算法均不能保证收敛.

Heiler 等只考虑了类内聚集性, 把满足类内聚集性限制的可行域表示成为了一个二阶锥, 以 SED 为目标函数, 在二阶锥规划的框架下实现了有类监督能力的算法[9]. Heiler 等只考虑类内聚集性而没有考虑类间散开性是为了能利用二阶锥规划模型, 但这并非不合理, 因为算法本身要追求对输入数据的表示性, 通常好的表示性和小的类内差异一定意味着大的类间差异. Heiler 的算法一定能够收敛, 但其计算复杂度要高于前述的 3 种算法.

还应指出, 目前这类方法都存在推广性不强的问题, 用于实际识别问题时, 一测试数据属于的类不知道, 所以对应到它的类内集聚性信息和类间散开性信息也都不可知, 此测试数据对应的系数不可由与训练样本对应系数一样的方式得到[32]. 虽有替代方法[32], 但无论如何, 这时测试数据对应的系数体现的类别特征相对较少. 这将极大地削弱鉴别性 NMF 算法用于实际识别问题时的效果.

4.3　加权 NMF 算法

任何方法被提出后, 总会伴随出现一些加权式的改进措施. 加权 NMF 是 INMF 算法中的另一类. 现有的三个加权 NMF 算法均可归结为 Blondel 等提出的数学框架(对 GKLD-EM 算法和 SED-EM 算法的加权改进) [34], 但他们的加权目的各异.

Blondel 等的加权目的是使数据中的重要区域被更好地描述[34]. 他们的加权方式是: (1) 定义数据中的重要区域, (2) 在优化过程中给重要区域上的重建错误以更多的权重.

Guillamet 等的加权目的是在训练样本不均衡的时减少基的冗余[35,36]. 实验表明, 如果训练样本是不均衡的(即: 存在某类样本远多于另一类的情况) , 那么得到的基就会更多地体现样本数量占优的类的特性[35]. Guillamet 等的加权方式是: (1) 统计各类样本在训练集中占的比重; (2) 优化过程中对每类样本的重建错误按其所占比重的倒数加权.

Wang 等的加权目的是消除训练样本的不确定性对 NMF 结果的影响[37]. 他们的加权方式是: (1) 估计每个训练样本的每个元素的不确定度; (2) 优化过程中对每个

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

样本的每个元素的重建错误按其不确定度的倒数加权.

5 结束语

本文综述的文献主要来自 IEEE、Elsevier、SpringerLink 和万方数据库, 也有部分来自如 Nature、Journal of Machine Learning Research 以及 BMC Bioinformatics 等的国际著名期刊, 其中涉及的算法覆盖了 NMF 算法研究的主要成果.

在科学分类的基础上, 本文对这些算法的构造原则(目标函数和优化方法)、应用特点及存在的问题进行了系统的分析、总结和比较. 作者希望这能为了解 NMF 算法研究的现状和开展相关工作提供有益的参考.

客观地讲, 目前 NMF 算法研究虽取得了一些成果, 但尚处于起步阶段, 一些非常有意义的问题还有待被探讨和解决, 这些问题包括: (1) 如何构造高效的 NMF 算法; (2) NMF 算法的评价与比较准测(包含评价比较原则和标准测试数据两方面的内容); (3) 提高鉴别性 NMF 算法推广性的有效方法; (4) 有新特性的 INMF 算法; (5) 如何确定应保留基的个数 L; (6) NMF 的初始化问题. 表 3 详细地阐述了研究上述问题的意义所在.

表 3 对未来 NMF 算法研究方向的分析

研究方向	研究意义
如何构造高效的 NMF 算法	具有高效的算法是 NMF 被大量实用的前提. 目前, NMF 算法效率均不高, 用于中等规模问题时, 一般就至少需要几百乃至上千步的迭代才可接近收敛.
NMF 算法的评价与比较准测	越来越多的算法已被提出, 如何客观界定一个算法的优劣已显非常必要. 有统一的评价和比较准则, 才能客观地分析现有算法的性能, 这即为它们的应用提供参考, 又为发展新算法指明方向. 此外, 一套完善规范的评价和比较准则的出现也是一个研究领域由起步走向初步成熟的标志和必然需要.
提高鉴别性 NMF 算法推广性的有效方法	将鉴别信息加入到 NMF 中是使 NMF 更适于识别问题的一个很好的研究思路, 但目前这类方法的推广性较差, 其实用效果无法保证.
有新特性的 INMF 算法	结合应用问题的特点, 对 BNMF 模型加入除本文综述的 3 类约束外的新约束有可能构造出更符合应用要求的 INMF 算法.
如何确定应保留基的个数 L	目前通常是先人为设定一个 L 值, 然后再根据得到的 NMF 结果是否符合要求(如: 某一重建误差值)来调整它, 并重新做实验验证新调整的 L 值是否正确, 所以确定一个满意 L 值常需多次实验. 如能根据要求较准确地预测 L 值, 将给实际使用带来很大的方便.
NMF 的初始化问题	Wild 等指出用结构化的初始值可提高 BNMF 算法的效率[17], 但这需借助球面 K-means 来完成, 且以收敛到稍差的局部解为代价. 可否为 BNMF 找到更简单有效且不影响解的质量的初始化手段呢? 可否为各类 INMF 算法也构造合理的初始值提高其效率呢?

参考文献:

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

[2] S E Palmer. Hierarchical structure in perceptual representation [J]. Cogn Psychol, 1977, 9(3): 441-474.

[3] E Wachsmuth, M W Oram, D I Perrett. Recognition of objects and their component parts: Responses of single units in the temporal cortex of the macaque[J]. Cereb Corte, 1994, 4(5): 509-522.

[4] N K Logothetis, D L Sheinberg. Visual object recognition[J]. Annu Rev Neurosci, 1996, 19(1): 577-621.

[5] I Biederman. Recognition-by-components: A theory of human image understanding[J]. Psychol Rev, 1987, 94(2): 115-147.

[6] S Ullman. High-Level Vision: Object Recognition and Visual Cognition[M]. Cambridge: MIT Press, 1996.

[7] D J Field. What is the goal of sensory coding[J]? Neural Computation, 1994, 6(4): 559-601.

[8] P O Hoyer. Non-negative matrix factorization with sparseness constraints[J]. J of Mach Learning Res, 2004, 5(9): 1457-1469.

[9] M Heiler, C Schnorr. Learning sparse representations by non-negative matrix factorization and sequential cone programming [J]. J of Mach. Learning Res, 2006, 7(7): 1385-1407.

[10] A Cichocki, R Zdunek, S Amari. Csiszar's divergences for non-negative matrix factorization: Family of new algorithms [J]. Lecture Notes in Computer Science, 2006, 3889: 32-39.

[11] D D Lee, H S Seung. Algorithms for non-negative matrix faetorization[A]. Adv in Neur Inform Proc Syst[C]. Cambridge: MIT Press, 2000. 556-562.

[12] D Donoho, V Stodden. When does non-negative matrix factorization give a correct decomposition into parts[A]. Adv in Neur Inform Proc Syst[C]. Cambridge: MIT Press, 2004. 1141-1148.

[13] C L Byrnes. Accelerating the EMML algorithm and related iterative algorithms by rescaled block-iterative methods[J]. IEEE Trans Image Processing, 1998, 7(1): 100-109.

[14] A Cichocki, S Amari, R Zdunek, et al. Extended SMART algorithms for non-negative matrix factorization [J]. Lecture Notes in Artificial Intelligence, 2006, 4029: 548-562.

[15] C Byrne. Choosing parameters in block-iterative or ordered subset reconstruction algorithms[J]. IEEE Transactions on

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

Image Processing, 2005, 14(3): 321– 327.

[16] W X Liu, N Zheng, Q B You. Non-negative matrix factorization and its applications in pattern recognition [J]. Chinese Science Bulletin, 2006, 51(1): 7– 18.

[17] S Wild, J Curry, A Dougherty. Improving non-negative matrix factorizations through structured initialization [J]. Pattern Recognition, 2004, 37(11): 2217– 2232.

[18] 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[J]. 电子学报, 2003, 31(z1): 1290– 1293.
Chen Weigang, Qi Feihu. Learning NMF representation using a hybrid method combining feasible direction algorithm and simulated annealing [J]. Acta Electronica Sinica, 2003, 31(z1): 1290– 1293. (in Chinese)

[19] R Kompass. A generalized divergence measure for nonnegative matrix factorization [A]. Proc of Neuroinfomatics Workshop [C]. Torun, Poland, 2005.

[20] R Zdunek, A Cichocki. Non-negative matrix factorization with quasi-Newton optimization[J]. Lecture Notes in Artificial Intelligence, 2006, 4029: 870– 879.

[21] I Dhillon, S Sra. Generalized nonnegative matrix approximations with Bregman divergences[A]. Adv in Neur Inform Proc Syst[C]. Cambridge: MIT Press, 2005. 283– 290.

[22] S Z Li, X W Hou, H J Zhang, et al. Learning spatially localized, parts-based representation[A]. Proc of Comp Vision and Pattern Recog [C]. Los Alamitos, California, USA, 2001. I. 207– 212.

[23] X R Chen, L Gu, S Z Li, et al. Learning representative local features for face detection[A]. Proc of Comp Vision and Pattern Recog[C]. Los Alamitos, California, USA, 2001. I. 1126 – 1131.

[24] A Pascual-Montano, J M Carzzo, K Lochi, et al. Non-smooth nonnegative matrix factorization (nsNMF) [J]. IEEE Trans Pattern Anal Machine Intell, 2006, 28(3): 403– 415.

[25] B W Xu, J J Lu, G S Huang. A constrained non-negative matrix factorization in information retrieval[A]. Proc of the 2003 IEEE Int Conf on Information Reuse and Integration[C]. Nevada, USA, 2003. 273– 277.

[26] 黄钢石, 张亚飞, 陆建江, 等. 一种受限非负矩阵分解方法[J]. 东南大学学报(自然科学版), 2004, 34(2): 189– 193.
Huang Gangshi, Zhang Yafei, Lu Jianjiang, et al. Constrained factorization method for non-negative matrix[J]. J of Southeast University(Natural Science Edition), 2004, 34(2): 189– 193. (in Chinese)

[27] P O Hoyer. Non-negative sparse coding[A]. Proc of IEEE Worksop on Neural Networks for Signal Processing[C]. Martigny, Switzerland, 2002. 557– 565.

[28] W Liu, N Zheng, X Lu. Non-negative matrix factorization for visual coding [A]. Proc of IEEE Int Conf on Acoustics, Speech and Signal Processing[C]. Hong Kong, 2003. III. 293 – 296.

[29] K Stadlthanner, F J Theis, C G Puntonet, et al. Extended sparse nonnegative matrix factorization[J]. Lecture notes in computer science, 2005, 3512: 249– 256.

[30] W X Liu, N Zheng. Learning sparse features for classification by mixture models[J]. Pattern Recognition Letters, 2004, 25(2): 155– 161.

[31] Y Wang, Y Jia, C Hu, et al. Fisher non-negative matrix factorization for learning local features[A]. Proc of Asian Conf on Comp Vision[C]. Jeju Island, Korea, 2004.

[32] Y Xue, C S Tong, W S Chen, et al. A modified non-negative matrix factorization algorithm for face recognition[A]. Proc of the 18th int conf on Pattern Recognition[C]. Hong Kong, 2006. III. 495– 498.

[33] I Buciu, I Pitas. A new sparse image representation algorithm applied to facial expression recognition[A]. Proc of IEEE Workshop on Machine Learning for Signal Processing[C]. Sao Luis, Brazil, 2004. 539– 548.

[34] V Blondel, N D Ho, P V Dooren. Algorithms for weighted non-negative matrix factorization[Z/OL]. http://www.inma.ucl.ac.be/publi/303209.pdf, 2007-3-15.

[35] D Guillamet, M Bressan, J Vitria. A weighted non-negative matrix factorization for local representation[A]. Proc of Comp Vision and Pattern Recog [C]. Los Alamitos, California, USA, 2001. I. 942– 947.

[36] D Guillamet, J Vitri , B Schiele. Introducing a weighted non-negative matrix factorization for image classification[J]. Pattern Recognition Letters, 2003, 24(14): 2447– 2454.

[37] G L Wang, A V Kossenkov, M F Ochs. LS-NMF: A modified non-negative matrix factorization algorithm utilizing uncertainty estimates[J]. BMC Bioinformatics, 2006, 7: 175.

作者简介:



李　乐　男, 1979 年生, 讲师, 博士研究生, 研究方向: 非负矩阵分解、非负矩阵集分解及它们在图象工程中的应用.
E-mail: lile05@mails.tsinghua.edu.cn



章毓晋　男, 1954 年生, 教授, 博士生导师, 研究方向: 图象工程及相关学科.
E-mail: zhang-yj@mail.tsinghua.edu.cn

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net