



自适应图正则化的低秩非负矩阵分解算法

余沁茹, 卢桂馥, 李华

引用本文:

余沁茹, 卢桂馥, 李华. 自适应图正则化的低秩非负矩阵分解算法[J]. 智能系统学报, 2022, 17(2): 325–332.

YU Qinru, LU Guifu, LI Hua. Nonnegative low-rank matrix factorization with adaptive graph neighbors[J]. *CAAI Transactions on Intelligent Systems*, 2022, 17(2): 325–332.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202102007>

您可能感兴趣的其他文章

图正则化稀疏判别非负矩阵分解

Graph-regularized, sparse discriminant, non-negative matrix factorization

智能系统学报. 2019, 14(6): 1217–1224 <https://dx.doi.org/10.11992/tis.201811021>

一种深度自监督聚类集成算法

A deep self-supervised clustering ensemble algorithm

智能系统学报. 2020, 15(6): 1113–1120 <https://dx.doi.org/10.11992/tis.202006050>

低秩分块矩阵的核近似

Kernel approximation of a low-rank block matrix

智能系统学报. 2019, 14(6): 1209–1216 <https://dx.doi.org/10.11992/tis.201904058>

局部自适应输入控制的随机游走抠图

Random-walk matting with local adaptive input control

智能系统学报. 2019, 14(5): 1007–1016 <https://dx.doi.org/10.11992/tis.201809014>

结合度量融合和地标表示的自编码谱聚类算法

An autoencoder-based spectral clustering algorithm combined with metric fusion and landmark representation

智能系统学报. 2020, 15(4): 687–696 <https://dx.doi.org/10.11992/tis.201911039>

结合谱聚类的标记分布学习

Label distribution learning based on spectral clustering

智能系统学报. 2019, 14(5): 966–973 <https://dx.doi.org/10.11992/tis.201809019>

微信公众平台



关注微信公众号, 获取更多资讯信息

DOI: 10.11992/tis.202102007

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1538.TP.20211012.2052.010.html>

自适应图正则化的低秩非负矩阵分解算法

余沁茹, 卢桂馥, 李华

(安徽工程大学 计算机与信息学院, 安徽 芜湖 241009)

摘要: 图正则化 (nonnegative matrix factorization, NMF) 算法 (graph regularization nonnegative matrix factorization, GNMF) 仍存在一些不足之处: GNMF 算法并没有考虑数据的低秩结构; 在 GNMF 算法中, 其拉普拉斯图是使用 K 近邻 (K nearest neighbor, KNN) 方法预先定义的, 而 KNN 方法无法总是获得最优图解, 从而使得 GNMF 算法的性能不能达到最优。为此, 本文提出了一种自适应图正则化的非负矩阵分解算法 (nonnegative low-rank matrix factorization with adaptive graph neighbors, NLMFAN)。一方面, 通过引入低秩约束, 使得 NLMFAN 可以获得原始数据集的有效低秩结构; 另一方面, 设计了一种通过自适应求解相似度矩阵的方法来进行图的构建, 即图的构造和矩阵分解的结果被融入一个整体的框架中, 使得图中节点的相似性是自动从数据中学习得到的。此外, 本文还给出了一种求解 NLMFAN 的有效算法。在多种数据集上的实验验证了本文所提出的算法的有效性。

关键词: 聚类; 特征提取; 降维; 流形学习; 非负矩阵分解; 低秩约束; 图正则化; 自适应聚类

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1673-4785(2022)02-0325-08

中文引用格式: 余沁茹, 卢桂馥, 李华. 自适应图正则化的低秩非负矩阵分解算法 [J]. 智能系统学报, 2022, 17(2): 325-332.

英文引用格式: YU Qinru, LU Guifu, LI Hua. Nonnegative low-rank matrix factorization with adaptive graph neighbors[J]. CAAI transactions on intelligent systems, 2022, 17(2): 325-332.

Nonnegative low-rank matrix factorization with adaptive graph neighbors

YU Qinru, LU Guifu, LI Hua

(School of Computer and Information, Anhui Polytechnic University, Wuhu 241009, China)

Abstract: The existing graph regularization nonnegative matrix factorization (GNMF) method still has some shortcomings: The GNMF algorithm does not consider the low-rank structure of data. In the GNMF algorithm, the Laplacian graph uses the K-nearest neighbor (KNN) method, and the KNN method cannot always obtain the optimal diagram, which makes the performance of the GNMF algorithm not optimal. For this reason, we propose an algorithm called nonnegative low-rank matrix factorization with adaptive graph neighbors (NLMFAN). On the one hand, by introducing low-rank constraints, NLMFAN can obtain the effective low-rank structure of the original dataset. On the other hand, a method for adaptively solving the similarity matrix is designed to construct the graph. This implies that the structure of the graph and the results of the matrix decomposition are integrated into an integrated framework so that the similarity of the nodes in the graph is automatically learned from the data. In addition, an effective algorithm for solving NLMFAN is given, and experiments on a variety of datasets verify the effectiveness of the algorithm.

Keywords: cluster; feature extraction; dimensionality reduction; manifold learning; nonnegative matrix factorization; low-rank constrain; graph regularization; adaptive clustering

聚类是计算机视觉中一项受到广泛关注且充满挑战的任务。近些年来, 图聚类因为相较于其

他方法表现出了更好的性能而被广泛应用于图像分类、图像检索等方面^[1-2]。在信息检索、计算机视觉和模式识别的许多问题中, 由于样本数据的维数很高, 使得从样本直接学习并聚类的方法是往往不可行的^[3]。因此, 人们希望找到两个或

收稿日期: 2021-02-07. 网络出版日期: 2021-10-13.

基金项目: 国家自然科学基金项目 (61976005, 61772277); 安徽省自然科学基金项目 (1908085MF183).

通信作者: 卢桂馥. E-mail: luguifu_jsj@163.com.

多个低维矩阵,使它们的乘积可以很好地近似于原始矩阵,越来越多的矩阵分解技术如 LU 分解、QR 分解、矢量量化和奇异值分解(singular value decomposition, SVD)、非负矩阵分解(non-negative matrix factorization, NMF)等由此被提出^[4-6]。在面部识别和文档聚类任务中, NMF 已被证明优于 SVD 等其他矩阵分解技术^[7]。NMF 的目的是找到两个非负矩阵,它们的乘积可以很好地近似原始矩阵。由于 NMF 更新规则仅允许加法运算,矩阵分解的非负约束使得学习到的成分矩阵成为基于局部的表示,且是学习对象各部分的最佳选择。为了提高 NMF 的性能,学界已提出了许多 NMF 的拓展算法,例如 $L_{2,1}$ 范数非负矩阵分解^[8]、结构不相干的低秩矩阵分解^[9]等。

近年来,有些研究者提出了基于流形学习的算法,例如局部线性嵌入^[10]、拉普拉斯特征图^[11]等,这些研究表明,高维数据其本质分布于低维子空间。研究人员发现,在 NMF 算法中引入流形学习方法获得数据的流形结构,有助于提升 NMF 的性能^[12]。Cai 等^[13-14]在 NMF 的基础上结合流形学习技术提出了图非负矩阵分解算法(graph non-negative matrix factorization, GNMF)。Zhou 等在 GNMF 的基础上为 NMF 添加了额外的约束^[15],进一步提出了局部学习正则化 NMF(local learning regularized NMF, LLNMF)^[16]。Li 等^[17]提出图正则化非负矩阵分解(graph nonnegative low-rank matrix factorization, GNLMF),该方法使得图正则化时可以获得原始图像数据的低秩结构。Du 等^[18]提出图嵌入正则化投影非负矩阵分解方法(graph embedding regularized projection nonnegative matrix factorization for face image feature extraction, GEPNMF),通过引入图嵌入正则化项,学习的子空间可以保留数据的局部几何结构,同时提升了算法判别力。Yin 等^[19]提出一种拉普拉斯正则化低秩表示及其应用算法(Laplacian regularized low-rank representation and its applications, LLRR),该算法通过对图形的正则化,既能表示数据的全局低维结构,又能捕获数据中固有的非线性几何信息。Wang 等^[20]提出一种降维局部约束图优化算法(locality constrained graph optimization for dimensionality reduction, LC-GODR),该算法将图优化和投影矩阵学习结合到了一个框架中。由于图不是预先构造并保持不变的,使得其在降维过程中的图形可以自适应更新。Meng 等^[21]提出的具有稀疏和正交约束的对偶图正则化非负矩阵分解(dual-graph regularized non-negative matrix factoriz-

ation with sparse and orthogonal constraints, SODNMF),此方法能同时考虑数据空间和特征空间的流形结构。受到 Nie 等^[22]的启发, Huang 等^[23]提出了局部自适应结构的正则化非负矩阵分解算法(regularized nonnegative matrix factorization with adaptive local structure learning, NMFAN)。NMFAN 算法使用自适应学习的方法来利用数据局部结构中流形信息,但是, NMFAN 算法并没有利用数据的有效低秩结构信息。

以上这些算法通过构建拉普拉斯图来利用数据的局部流形结构信息,其所使用的图的极大地影响着算法的性能。研究人员往往利用 K 近邻(K nearest neighbor, KNN)来构建图。然而,一方面, KNN 方法构造的图有可能会破坏部分原始数据的局部连通性;另一方面,其图的构造与矩阵分解的结果无关,从而使得相关算法的性能不能达到最优。此外,这些算法往往没有考虑数据的低秩结构,而数据的有效信息往往隐藏在其低秩结构中。为此,本文提出了一种自适应图正则化的非负矩阵分解算法(nonnegative low-rank matrix factorization with adaptive graph neighbors, NLMFAN)。一方面, NLMFAN 通过引入低秩约束来获得原始数据集的潜藏的有效信息,以此提升现有算法性能;另一方面, NLMFAN 将图的构造和矩阵分解的结果融入一个整体的框架中,自动从数据中学习得到图中节点的相似性,优化了算法精度。文中同时给出了一种求解 NLMFAN 的有效算法,并在多种数据集上进行了实验验证本文所提出的算法的有效性。

1 相关工作

1.1 非负矩阵分解(NMF)

NMF 是一种广泛使用的矩阵分解算法。它试图将一个非负矩阵分解为两个阶乘矩阵,其乘积是其自身的最佳近似值。给定一个数据矩阵 $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_n] \in \mathbf{R}^{M \times N}$ 的每一列都是样本矢量, NMF 的目的是找到两个非负矩阵 $\mathbf{U} = [\mathbf{u}_{ik}] \in \mathbf{R}^{M \times K}$ 和 $\mathbf{V} = [\mathbf{v}_{jk}] \in \mathbf{R}^{N \times K}$,使得它们的乘积可以很好地近似于原始矩阵 \mathbf{X} :

$$\mathbf{X} \approx \mathbf{U}\mathbf{V}^T \quad (1)$$

即标准 NMF 的目标是搜索两个非负矩阵 \mathbf{U} 和 \mathbf{V} 以优化以下目标

$$\begin{cases} \min_{\mathbf{U}, \mathbf{V}} \|\mathbf{X} - \mathbf{U}\mathbf{V}^T\|_F^2 \\ \text{s.t. } \mathbf{U} \geq 0, \mathbf{V} \geq 0 \end{cases} \quad (2)$$

其中 $\|\cdot\|_F$ 表示矩阵的 F 范数,同时其存在以下更新规则的表示形式:

$$u_{ij} \leftarrow u_{ij} \frac{(UV)_{ij}}{(UV^T V)_{ij}}, v_{ij} \leftarrow v_{ij} \frac{(X^T U)_{ij}}{(V U^T U)_{ij}} \quad (3)$$

1.2 图正则化非负矩阵分解 (GNMF)

为了能在 NMF 算法中保留其内在结构的同时探索数据的局部几何结构,近几年将流形学习与 NMF 相结合的算法研究成为热点。Cai 等^[13]通过将图正则化项融入到标准 NMF 算法框架中提出了图正则化非负矩阵分解 (GNMF),其目标函数如下:

$$\begin{cases} \min_{U,V} \frac{1}{2} \|X - UV\|_F^2 + \lambda \cdot \text{tr}(V L_{\text{ap}} V^T) \\ \text{s.t. } U, V \geq 0 \end{cases} \quad (4)$$

式中: $\text{tr}(V L_{\text{ap}} V^T)$ 是图正则项; L_{ap} 是图拉普拉斯矩阵, $\lambda < 0$ 是图正则化参数。文献 [13] 中给出了 GNMF 算法的更新规则:

$$\begin{cases} u_{il} \leftarrow u_{il} \frac{(X V^T)_{il}}{(U V V^T)_{il}} \\ v_{ij} \leftarrow v_{ij} \frac{(U^T X + \lambda V S_{\text{ap}})_{ij}}{(U^T U V + \lambda V D)_{ij}} \end{cases} \quad (5)$$

其中, D_{ap} 是对角矩阵, $L_{\text{ap}} = D_{\text{ap}} - S_{\text{ap}}$ 。

2 自适应低秩非负矩阵分解算法

研究表明,原始图像数据通常嵌入位于高维欧式空间中的非线性低维流形上,且其有效信息通常隐藏在低秩结构中。另外,在图正则化相关算法中所使用的图一般是预先定义的,其图的构造与算法的其他部分往往互相独立。因此,这些算法使用的图并不一定是最优的。为此,在本小节提出一种自适应图正则化的非负矩阵分解算法 NLMFAN,该算法同时兼顾了原始图像数据的有效低秩结构和自适应图构造。

2.1 NLMFAN 的算法模型

设输入的原始数据集为 X ,先给出本文 NLMFAN 算法的求解形式如下:

$$\begin{cases} \min \|X - L - E\|_F^2 + \alpha \left(\sum_{j=1}^n (\|x_i - x_j\|_2^2 s_{ij} + \gamma s_{ij}^2) \right) + \\ \lambda \cdot \text{tr}(V^T L_S V) \\ \text{s.t. } L = UV^T, L, U, V \geq 0, \text{rank}(L) \leq r, \text{card}(E) \leq e, \\ s_i^T \mathbf{1} = 1, 0 \leq s_i \end{cases} \quad (6)$$

目标函数的第 1 项为在 $\text{rank}(L) \leq r, \text{card}(E) \leq e$ 条件的约束下通过原始数据集 X 求其低秩形式 L ,再对得出的结果使用 NMF 分解。其中矩阵 X 为原始数据集, L 代表矩阵的低秩部分, E 代表其稀疏部分, U, V 由 L 非负矩阵分解得出。 r 为 L 的秩范围, e 为 E 设置的稀疏范围, $\text{rank}(L)$ 表示矩阵 L 的秩, $\text{card}(E)$ 表示矩阵 E 的非零条目

数, $\mathbf{1}$ 表示一个所有元素都为 1 的列向量。

目标函数的第 2 项为自适应正则项,使用其完成自适应相似矩阵的构造,其中 α 均为正则化参数。此处我们直接基于数据点间的欧几里得距离构造邻域矩阵,并设距离越小则成为最近邻的可能性越大。原数据集 X 中的任意一数据点 x_i ,都有所有数据点 $\{x_1, x_2, \dots, x_n\}$ 可以作为近邻与 x_i 连接,连接的概率为 s_{ij} 。 $s_i \in \mathbf{R}^{n \times 1}$ 表示向量,其第 j 个元素为 s_{ij} 。在此基础上,可以得到相似度矩阵构建的求解函数为

$$\min_{s_i^T \mathbf{1} = 1, 0 \leq s_i \leq 1} \sum_{j=1}^n \|x_i - x_j\|_2^2 s_{ij} \quad (7)$$

上述相似度矩阵求解的问题存在简化的可能,即我们仅基于欧氏距离求解距离最近的点,并设其与 x_i 的连通概率为 1,其他的点概率都为 0。此时,对于式 (7) 中所有的数据点 x_i ,除其之外的所有同数据集的点都存在相同的连通概率即 $\frac{1}{n}$ 。此时,式 (7) 可写作:

$$\min_{s_i^T \mathbf{1} = 1, 0 \leq s_i \leq 1} \sum_{j=1}^n s_{ij}^2 \quad (8)$$

结合式 (7) 和 (8) 可得目标函数中的第 2 项,即自适应正则项为

$$\min_{s_i^T \mathbf{1} = 1, 0 \leq s_i \leq 1} \sum_{j=1}^n (\|x_i - x_j\|_2^2 s_{ij} + \gamma s_{ij}^2) \quad (9)$$

其中 γ 是正则化参数。通过求解式 (9) 获得的矩阵 $S \in \mathbf{R}^{n \times n}$,可被视为相似度矩阵^[22]。当我们假设每个点可以被表示为一个向量 v_i 时,则有:

$$\begin{aligned} \frac{1}{2} \sum_{i,j=1}^n \|v_i - v_j\|_2^2 &= \sum_{i,j=1}^n v_i s_{ij} v_j^T - \sum_{i,j=1}^n v_i s_{ij} v_i^T = \\ &= \sum_{i,j=1}^n v_i (D_s)_{ij} v_j^T - \sum_{i,j=1}^n v_i s_{ij} v_i^T = \end{aligned} \quad (10)$$

$$\text{tr}(V^T D_s V) - \text{tr}(V^T W_s V) = \text{tr}(V^T L_s V)$$

式中: L_s 是 S 的拉普拉斯矩阵, $L_s = D_s - W_s$; $D_s \in \mathbf{R}^{k \times k}$ 是 $\sum_j (s_{ij} + s_{ji})/2$ 的对角阵。

目标函数的第 3 项是局部图拉普拉斯约束函数,用来衡量数据低维表示的平滑度,其中 $\text{tr}(\cdot)$ 表示矩阵的秩。

2.2 NLMFAN 的算法求解

设计了一种有效的更新算法来求解 NLMFAN,该算法通过固定其他变量迭代更新一个变量的值来优化目标,此过程重复直到收敛。

2.2.1 固定 U, V 更新 S

观察式 (6),不难发现在固定变量 U, V 的情况下更新 S 的最小化,式 (6) 可等价于解决式 (11):

$$\begin{cases} \min_S \sum_{i,j}^n (\|x_i - x_j\|_2^2 s_{ij} + \gamma s_{ij}^2) + \beta \cdot \text{tr}(\mathbf{V}^T \mathbf{L}_S \mathbf{V}) \\ \text{s.t. } \mathbf{s}_i^T \mathbf{1} = 1, 0 \leq s_i \leq 1 \end{cases} \quad (11)$$

其中 $\beta = \frac{\lambda}{\alpha}$ 。由式(6)结合式(11)有:

$$\begin{cases} \min_s \sum_{i,j=1}^n \left(\|x_i - x_j\|_2^2 s_{ij} + \gamma s_{ij}^2 + \frac{1}{2} \beta \|v_i - v_j\|_2^2 s_{ij} \right) \\ \text{s.t. } \mathbf{s}_i^T \mathbf{1} = 1, 0 \leq s_i \leq 1 \end{cases} \quad (12)$$

本文可以对每个 i 解决式(12), 此时式(12)可等价于:

$$\begin{cases} \min_s \sum_{i,j=1}^n \left(\|x_i - x_j\|_2^2 + \frac{1}{2} \beta \|v_i - v_j\|_2^2 s_{ij} + \gamma s_{ij}^2 \right) \\ \text{s.t. } \mathbf{s}_i^T \mathbf{1} = 1, 0 \leq s_i \leq 1 \end{cases} \quad (13)$$

令 $d_{ij}^x = \|x_i - x_j\|_2^2$, $d_{ij}^v = \frac{1}{2} \|v_i - v_j\|_2^2$, 且 $\mathbf{d}_i \in \mathbf{R}^{m \times 1}$, 第 j 个元素为 $d_{ij} = d_{ij}^x + \beta d_{ij}^v$, 将其代入式(13)可得:

$$\min_{s_i^T \mathbf{1} = 1, 0 \leq s_i \leq 1} \left\| s_i + \frac{1}{2\gamma} \mathbf{d}_i \right\|_2^2 \quad (14)$$

设 ζ 和 $\eta \geq 0$ 为拉格朗日乘子, 可以写出式(11)的拉格朗日函数:

$$R(s_i, \eta, \zeta_i) = \frac{1}{2} \left\| s_i + \frac{1}{2\gamma_i} \mathbf{d}_i \right\|_2^2 - \eta (\mathbf{s}_i^T \mathbf{1} - 1) - \zeta_i^T s_i \quad (15)$$

根据 KKT 条件^[23], 满足 $s_{ij} \geq 0$, 最优解可以定义为

$$s_{ij} = \left(-\frac{1}{2\gamma_i} d_{ij}^x + \eta \right) \quad (16)$$

其中, $\{d_{i1}^x, d_{i2}^x, \dots, d_{in}^x\}$ 由小到大表示。如果最优 s_i 只有 k 个非零元素, 则根据式(16)可知 $s_{ik} > 0$ 且 $s_{i,k+1} = 0$ 。因此, 有:

$$\begin{cases} -\frac{1}{2\gamma_i} d_{ik}^x + \eta > 0 \\ -\frac{1}{2\gamma_i} d_{i,k+1}^x + \eta \leq 0 \end{cases} \quad (17)$$

由式(17)和 $\mathbf{s}_i^T \mathbf{1} = 1$ 的约束, 有:

$$\sum_{j=1}^k \left(-\frac{1}{2\gamma_i} d_{ij}^x + \eta \right) = 1 \Rightarrow \eta = \frac{1}{k} + \frac{1}{2k\gamma_i} \sum_{j=1}^k d_{ij}^x \quad (18)$$

结合式(17)和(18), 可以得到:

$$\frac{k}{2} d_{ik}^x - \frac{1}{2} \sum_{j=1}^k d_{ij}^x > \gamma_i \leq \frac{k}{2} d_{i,k+1}^x - \frac{1}{2} \sum_{j=1}^k d_{ij}^x \quad (19)$$

为了获得具有 k 个非零值的最优 s_i , 可以将 γ_i 设置为

$$\gamma_i = \frac{k}{2} d_{i,k+1}^x - \frac{1}{2} \sum_{j=1}^k d_{ij}^x \quad (20)$$

为了便于计算, 可以将整体 γ 设置为 $\gamma_1, \gamma_2, \dots, \gamma_n$ 的均值, 即

$$\gamma = \frac{1}{n} \sum_{i=1}^n \left(\frac{k}{2} d_{i,k+1}^x - \frac{1}{2} \sum_{j=1}^k d_{ij}^x \right) \quad (21)$$

通过取 γ_i 的平均值, 所有 s_i 的平均非零元素

应为 k 。我们不直接搜索正则化参数 γ , 而是搜索近邻数 k 。因为 k 是一个整数并且其值是有限的 (即 $0 \leq k \leq n$), 所以参数搜索会更加容易。

2.2.2 固定 \mathbf{S} 、 \mathbf{V} 更新 \mathbf{U}

观察式(6), 不难发现在固定变量 \mathbf{S} 、 \mathbf{V} 的情况下更新 \mathbf{U} 的最小化, 式(6)可等价于解决下式:

$$\begin{aligned} & \min_{\mathbf{L}, \mathbf{E}} \|\mathbf{X} - \mathbf{L} - \mathbf{E}\|_F^2 \\ & \text{s.t. } \mathbf{L} = \mathbf{U}\mathbf{V}^T, \mathbf{L}, \mathbf{U}, \mathbf{V} \geq 0, \text{rank}(\mathbf{L}) \leq r, \text{card}(\mathbf{E}) \leq e \end{aligned} \quad (22)$$

在优先求解 \mathbf{L} 时, 求解式(22)可转换为先求解式(23)再求解 \mathbf{U} 的过程:

$$\begin{aligned} & \min_{\mathbf{L}, \mathbf{E}} \|\mathbf{X} - \mathbf{L} - \mathbf{E}\|_F^2 \\ & \text{s.t. } \mathbf{L} \geq 0, \text{rank}(\mathbf{L}) \leq r, \text{card}(\mathbf{E}) \leq e \end{aligned} \quad (23)$$

将式(23)可拆两个子问题来解决。当固定 \mathbf{L} 或 \mathbf{E} 时, 有:

$$\begin{cases} \mathbf{L}_t = \arg \min_{\text{rank}(\mathbf{L}) \leq r} \|\mathbf{X} - \mathbf{L} - \mathbf{E}_{t-1}\|_F^2 \\ \mathbf{E}_t = \arg \min_{\text{card}(\mathbf{E}) \leq e} \|\mathbf{X} - \mathbf{L}_t - \mathbf{E}\|_F^2 \end{cases} \quad (24)$$

可见式(24)对于 \mathbf{L} 或 \mathbf{E} 都是凸的, 因此可以固定一个变量并更新另一个, 直到收敛。通过迭代 $\mathbf{X} - \mathbf{E}_{t-1}$ 将奇异值赋给 \mathbf{L}_t , 在 $\mathbf{X} - \mathbf{L}_t$ 上进行相同操作得到 \mathbf{E}_t , 由此得到两个子问题的解。

但是, 在每次迭代中求 $\mathbf{X} - \mathbf{E}_{t-1}$ 的 SVD 都很耗时。本文采用文献[24]中提出的双边随机投影(bilateral random projection, BRP)来解决式(24)。

对于给定的矩阵 $\mathbf{X} \in \mathbf{R}^{m \times n}$, 其 r 个 BRPs 如下:

$$\mathbf{Y}_1 = \mathbf{X} \mathbf{A}_1 \quad (25)$$

$$\mathbf{Y}_2 = \mathbf{X}^T \mathbf{A}_2 \quad (26)$$

式中: $\mathbf{A}_1 \in \mathbf{R}^{n \times r}$, $\mathbf{A}_2 \in \mathbf{R}^{m \times r}$ 是随机矩阵, 那么 \mathbf{X} 的秩 r 逼近为

$$\mathbf{L} = (\tilde{\mathbf{L}})^{\frac{1}{2q+1}} = \mathbf{Q}_1 \left[\mathbf{R}_1 (\mathbf{A}_2^T \mathbf{Y}_1)^{-1} \mathbf{R}_2^T \right]^{\frac{1}{2q+1}} \mathbf{Q}_2^T \quad (27)$$

由此可计算得出 \mathbf{L} 。在此基础上, 对式(28)求解:

$$\begin{cases} \min_t \|\mathbf{L} - \mathbf{U}\mathbf{V}^T\|_F^2 \\ \text{s.t. } \mathbf{U} \geq 0 \end{cases} \quad (28)$$

可以看出式(28)为式(1)中的标准 NMF 式。因此, 对于等式中的 \mathbf{U} , 我们有式(3)中完全相同的迭代更新规则:

$$u_{ij} \leftarrow u_{ij} \frac{(\mathbf{X}\mathbf{V})_{ij}}{(\mathbf{U}\mathbf{V}^T\mathbf{V})_{ij}} \quad (29)$$

2.2.3 固定 \mathbf{S} 、 \mathbf{U} 更新 \mathbf{V}

观察式(6), 不难发现在固定变量 \mathbf{S} 、 \mathbf{U} 的情况下更新 \mathbf{V} 的最小化, 式(6)可等价于解决下式:

$$\begin{aligned} & \min_{\mathbf{L}, \mathbf{E}} \|\mathbf{X} - \mathbf{L} - \mathbf{E}\|_F^2 + \lambda \cdot \text{tr}(\mathbf{V}^T \mathbf{L}_S \mathbf{V}) \\ & \text{s.t. } \mathbf{L} = \mathbf{U}\mathbf{V}, \mathbf{L}, \mathbf{U}, \mathbf{V} \geq 0, \text{rank}(\mathbf{L}) \leq r, \text{card}(\mathbf{E}) \leq e \end{aligned} \quad (30)$$

参考 2.2.2 节中方法可得一确定 \mathbf{L} 。由此, 化

式(30)为

$$\min_{L, E} \|L - UV^T\|_F^2 + \lambda \cdot \text{tr}(V^T L_S V) \quad (31)$$

s.t. $V \geq 0$

为了约束 $V \geq 0$, 令 $\xi \in \mathbf{R}^{n \times c}$ 表示对应的拉格朗日乘数, 则可以将拉格朗日函数定义为

$$R(V) = \|L - UV^T\|_F^2 + \lambda \cdot \text{tr}(V^T L_S V) - \text{tr}(\xi V^T) \quad (32)$$

设 $\frac{\partial R(V)}{\partial V} = 0$, 可得:

$$\xi = -2L^T U + 2VU^T U + 2\lambda L_S V \quad (33)$$

根据 KKT 条件^[23] 满足 $\xi_{ij} v_{ij} = 0$, 有:

$$(-2L^T U + 2VU^T U + 2\lambda L_S V)_{ij} v_{ij} = 0 \quad (34)$$

即

$$(-L^T U + VU^T U + \lambda D_S V)_{ij} v_{ij} = 0 \quad (35)$$

则式(35)遵从如下迭代规则:

$$v_{ij} \leftarrow v_{ij} \frac{(L^T U + \lambda W_S V)_{ij}}{(VU^T U + \lambda D_S V)_{ij}} \quad (36)$$

至此, 可以给出 NLMFAN 的算法求解步骤。

算法 1 自适应图正则化的低秩非负矩阵分解算法

输入 数据矩阵 $X \in \mathbf{R}_+^{m \times n}$ 的初始值, 簇号 c , 参数 k, λ, β ;

输出 聚类指标矩阵 $V \in \mathbf{R}^{m \times c}$ 。

1) 对 X 进行 BRP 操作, 求出 Y_1 和 Y_2 ;

2) 对 Y_1 和 Y_2 进行 QR 分解, 计算出矩阵 X 的低秩结构 L ;

3) 通过对式(7)的最优解来初始化 S ;

4) 重复更新

$$u_{ij} \leftarrow u_{ij} \frac{(XV)_{ij}}{(UV^T V)_{ij}}, v_{ij} \leftarrow v_{ij} \frac{(L^T U + \lambda W_S V)_{ij}}{(VU^T U + \lambda D_S V)_{ij}}$$

对于每个 i , 通过求解式(15)更新 S 的第 i 行。其中 $d_i \in \mathbf{R}^{n \times 1}$, 第 j 个元素为 $d_{ij} = \|x_i - x_j\|_2^2 + \beta \|v_i - v_j\|_2^2$; 直到收敛;

5) 求得聚类指标矩阵 V 。

2.3 NLMFAN 算法收敛性分析

定理 1 对于 $U \geq 0, V \geq 0$, 式(6)中的目标在式(14)、(21)、(36)中的更新规则下不增加, 因此收敛。

证明 显然, 式(14)可以用第 2.2 节中描述的闭式解来解决。因此, 我们只需要证明式(21)和式(36)在每次迭代中的更新规则下目标值是不增加的。此外, 因为式(6)中目标的第 2 项与 U 无关, 第 1 项在计算 L 时不涉及 U 值的更新, 且计算得出 L 后的操作符合标准 NMF 分解, 所以 NLMFAN 中的 U 更新规则与原始 NMF 完全相同。因此, 可以使用 NMF 收敛的证明方式来证明在等式中更新规则下目标没有增加。有关

详细信息, 可参考文献[17]。

现在, 只需要证明在式(36)中的更新规则下我们的目标不会增加即可, 而式(32)中更新规则其实等价于文献[25]中式(26)、(36)的收敛性证明可参考文献[25]附录中证明过程, 此处不列出。

3 现实数据集实验评估

3.1 实验数据集

为了评估本文提出方法的性能, 我们在实际基准数据集进行了实验。

本文实验用到的数据集包括 CLUTO 数据工具及 UCI 机器学习数据集。其中, CLUTO 是一个软件包, 用于对低维和高维数据集进行聚类, 并分析各种聚类的特征。CLUTO 的数据工具包中包含的信息检索、客户购买交易、网络、地理信息系统、科学和生物学等数据集非常适于聚类测试。UCI 机器学习数据集是加州大学欧文分校提出的用于机器学习的数据库, 该数据库目前共有 559 个数据集, 其数目还在不断增加。UCI 数据集是一个常用的标准测试数据集。

选取 CLUTO 数据工具中的 4 个数据集 (Cacmcisi、Hitech、K1a、K1b) 与 UCI 机器学习数据集 4 个数据集 (Abalone、Krvs、Wdbc、Vote) 进行实验。表 1 给出了各数据集及其特征。

表 1 数据集及特征
Table 1 Description of data sets

数据集	样本量	特征量	类	数据集	样本量	特征量	类
Cacmcisi	4663	348	2	Abalone	2282	8	4
Hitech	2301	2216	6	Krvs	3196	36	2
K1a	2340	1326	6	Vote	435	16	2
K1b	2340	1326	6	Wdbc	569	30	2

3.2 实验步骤

为了进一步评估所提出方法的性能, 将 NLMFAN 与一些经典算法和最新算法进行了对比实验。主成分分析 (principal component analysis, PCA): 经典的降维方法^[26]。NMF: 标准 NMF 算法^[6]。LLNMF(2009): 使用局部结构学习的 NMF 算法^[16]。GNMF(2011): 图正则化的 NMF^[13-14]。NMFAN(2020): 具有局部自适应结构的正则化 NMF 算法^[25]。

本文实验条件为 Intel(R) Core(TM) i7-1065G7 1.50 GHz CPU, 16 GB DDR3 内存, Matlab2019b。在后续的性能评估中, 以准确率 (accuracy, ACC)、标准互信息率 (normalized mutual information, NMI) 两个指标作为评价算法性能好坏的标准,

在8个基准数据集上测试算法。实验中所用算法的参数已结合原论文根据需要调节至最优。

在每个数据集上都进行了算法精度与收敛速度测试,并从8个数据集中选取了4个用于测试

λ 参数对聚类性能的影响。

3.3 CLUTO 数据集实验

CLUTO 的4个文档数据集上进行了每组10次实验。实验结果均值如表2所示。

表2 各算法在 CLUTO 数据集上聚类结果比较
Table 2 ACC and NMI of clustering on CLUTO database

数据集	PCA		NMF		LLNMF		GNMF		NMFAN		NLMFAN	
	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
Cacmcisi	68.67	68.69	90.46	90.46	74.20	74.20	92.09	92.09	92.22	92.22	92.44	92.44
Hitech	26.25	26.42	22.73	27.03	22.16	26.99	24.12	26.99	22.90	26.42	22.51	26.60
K1a	29.40	59.57	27.95	61.88	30.21	62.09	27.05	61.50	29.74	62.14	30.86	62.22
K1b	29.40	59.57	28.12	62.31	29.27	59.36	31.20	62.01	32.26	62.35	33.70	62.61
均值	38.43	53.56	43.32	60.42	38.96	55.66	43.62	60.65	44.28	60.78	44.88	60.97

从表2中可以看出: NMF 相关算法的聚类结果还是优于传统聚类方法。同时,在 NMF 算法中, NLMFAN 算法显示出了优于其他几种方法的聚类效果。尤其是与改进前的 NMFAN 算法相比, NLMFAN 算法在每个数据集上的精度及平均精度均有所提高。

图1展现了在 CLUTO 的4个数据集上 NLMFAN 算法的收敛情况。图中目标函数值即最小误差值。可以看出,本文方法在数据集上均可快速收敛,迭代次数均在50次内。

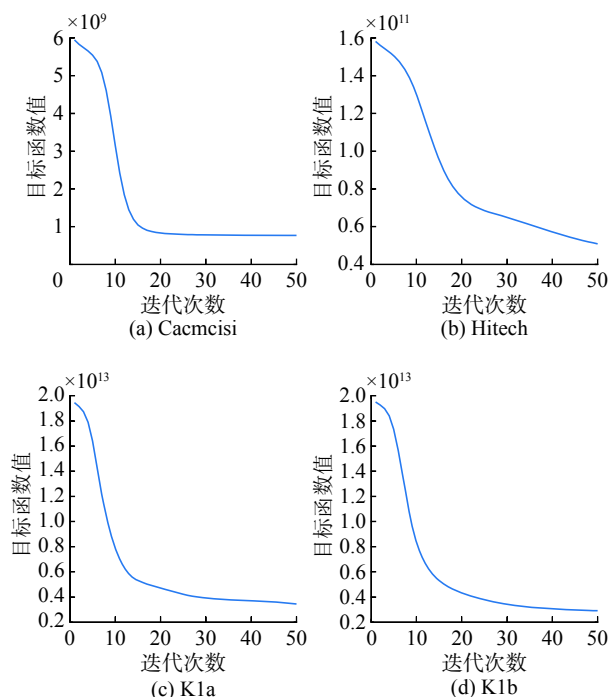


图1 NLMFAN 在 CLUTO 数据集上的收敛曲线

Fig. 1 Convergence curves of NLMFAN on CLUTO database

3.4 UCI 数据集实验

本节使用了与上节相同的几种算法在 UCI 的4个文档数据集上进行了每组十次实验。实验结果如表3所示。

从表3中可以看出:在 UCI 数据集上, NLMFAN 算法的聚类效果基本优于其他算法,同时其在 Abalone 数据集上的实验数据较原有的其他算法有 5.89%~0.57% 的提升。图2呈现了在 UCI 的4个数据集上 NLMFAN 算法的收敛情况。可以看出, NLMFAN 算法在几个数据集上均有收敛,且速度较快。

3.5 参数选择实验

本节对算法中 λ 参数的选择进行讨论。本文从 CLUTO 中选取了 Hitech、K1b, 从 UCI 中选取了 Abalone、Krvs, 总计4个数据集进行实验。

从图3可以看出:

1) 除了 NMF 以外, GNMF、NMFAN、NLMFAN 都受到 λ 值的影响。总体来说,当 λ 的值取在 10 以内时 NLMFAN 可以获得较好的聚类结果,当 λ 取到 10 时可以获得所有算法平均最优结果。

2) 除了在 Hitech 数据集上有明显的波动外, GNMF 在其他数据集上都没有明显波动变化,说明其受 λ 的影响不大。而根据提出 GNMF 的文献[13]中对其的实验结果表明, GNMF 的结果会在固定 λ 值时,根据其选取的最近邻数 k 值的变化而变化。这是因为 GNMF 的精度主要依赖于其通过 KNN 方法构建的拉普拉斯图,且其图构造与 NMF 相互独立的计算模式使得该特征的表现尤为明显。而 NMFAN 与 NLMFAN 依赖的则是自适应选择最近邻后生成相似度矩阵的结果,

即式(6)中包含基于相似度矩阵 S 的拉普拉斯矩阵 L_s 的项,该项即为受到 λ 影响的图正则项。

表3 各算法在UCI数据集上聚类结果比较

Table 3 ACC and NMI of clustering on UCI database

%

数据集	PCA		NMF		LLNMF		GNMF		NMFAN		NLMFAN	
	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
Abalone	30.24	30.28	34.22	34.22	33.17	34.14	35.50	37.25	35.58	36.11	36.15	36.33
Krvs	52.19	52.22	52.22	52.22	57.48	57.48	54.82	54.82	52.47	52.47	52.57	52.57
Vote	61.61	61.61	85.98	85.98	64.60	64.60	86.44	86.44	87.36	87.36	87.59	87.59
Wdbc	62.92	62.92	84.53	84.53	51.14	62.74	85.41	85.41	85.41	85.41	85.46	85.46
均值	62.27	62.27	85.26	85.26	57.87	63.67	85.93	85.93	86.39	86.39	86.53	86.53

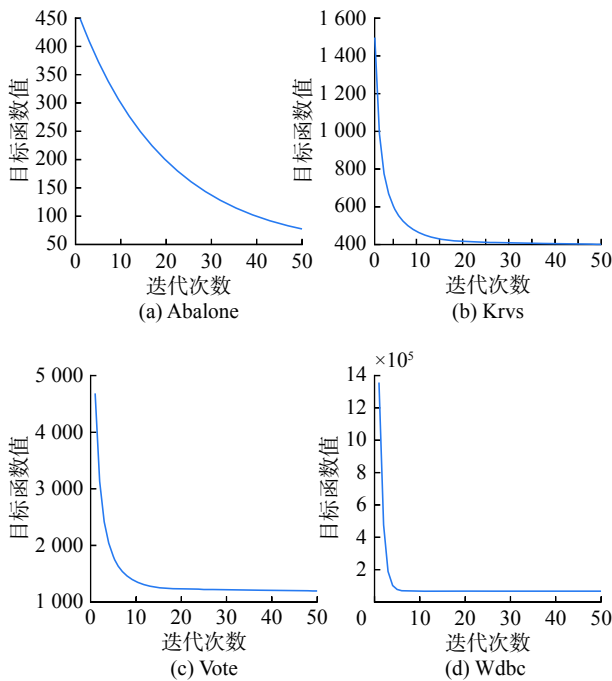
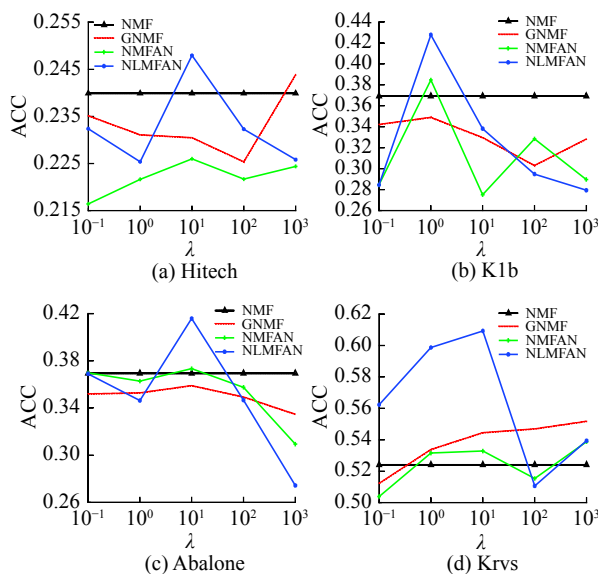


图2 NLMFAN在UCI数据集上的收敛曲线

Fig. 2 Convergence curve of NLMFAN on UCI database

图3 各算法中 λ 参数对性能的影响Fig. 3 Influence of λ parameter on different algorithms

4 结束语

在本文中,提出了一种自适应图正则化的非负矩阵分解算法(NLMFAN)。

1)提出了一种新的具有低秩特性的自适应邻域GNMF模型,同时也设计了一种求解NLMFAN的高效求解算法;

2)NLMFAN方法可以在低秩约束条件下同时执行局部结构学习和矩阵分解过程,即可以根据分解的结果自适应地学习局部流形结构,并重新构造合适的图以保留精炼的局部结构;

3)传统的基于图的方法通常基于固定的数据图对数据进行聚类,容易受到预先构造的不准确的图的影响。NLMFAN通过引入自适应邻居(adaptive neighbors, ANs)正则项对迭代中的相似度矩阵进行修正,从而减少分配ANs对数据点之间相似性带来的改变。

然而在许多聚类应用中,数据中的噪声是通过结构或组聚类来分布的,本文算法中并未考虑空间联系。在未来的工作中,我们将把结构化稀疏性作为约束来提升NLMFAN的性能。

参考文献:

- [1] ZHAO K, HAIQI P, STEVEN C H H, et al. Robust graph learning from noisy data[J]. *IEEE transactions on cybernetics*, 2019, 50(5): 1833–1843.
- [2] CHEN C F, WEI C P, WANG Y C F. Low-rank matrix recovery with structural incoherence for robust face recognition[C]//Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA, 2012: 2618–2625.
- [3] DUDA R O, HART P E, STORK D G. Pattern classification[M]. 2nd ed. New York: Wiley-Interscience, 2000.
- [4] MAKHOUL J, ROUCOS S, GISH H. Vector quantization in speech coding[J]. *Proceedings of the IEEE*, 1985, 73(11): 1551–1588.
- [5] TURK M, PENTLAND A. Eigenfaces for recognition[J]. *Journal of cognitive neuroscience*, 1991, 13(6): 556–562.

- [6] LEE D D, SEUNG H S. Algorithms for non-negative matrix factorization[J]. *Nips*, 2000, 13(6): 556–562.
- [7] YANG C F, YE Mao, ZHAO Jing. Document clustering based on nonnegative sparse matrix factorization[C]//Proceedings of the First International Conference on Natural Computation. Changsha, China, 2005: 557–563.
- [8] LIU Guangcan, LIN Zhouchen, YAN Shuicheng, et al. Robust recovery of subspace structures by low-rank representation[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2013, 35(1): 171–184.
- [9] LU Yuwu, YUAN Chun, ZHU Wenwu, et al. Structurally incoherent low-rank nonnegative matrix factorization for image classification[J]. *IEEE transactions on image processing*, 2018, 27(11): 5248–5260.
- [10] ROWEIS S T, SAUL L K. Nonlinear dimensionality reduction by locally linear embedding[J]. *Science*, 2000, 290(5500): 2323–2326.
- [11] BELKIN M, NIYOGI P. Laplacian eigenmaps and spectral techniques for embedding and clustering[C]//Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic. Vancouver, British Columbia, Canada, 2001: 585–591.
- [12] BELKIN M, NIYOGI P, SINDHWANI V. Manifold regularization: a geometric framework for learning from labeled and unlabeled examples[J]. *Journal of machine learning research*, 2006, 7(11): 2399–2434.
- [13] CAI Deng, HE Xiaofei, HAN Jiawei, et al. Graph regularized nonnegative matrix factorization for data representation[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2011, 33(8): 1548–1560.
- [14] CAI Deng, HE Xiaofei, WU Xiaoyun, et al. Non-negative matrix factorization on manifold[C]//Proceedings of 2008 Eighth IEEE International Conference on Data Mining. Pisa, Italy, 2008.
- [15] BOTTOU L, VAPNIK V. Local learning algorithm[J]. *Neural computation*, 1992, 4(6): 888–900.
- [16] GU Quanquan, ZHOU Jie. Local learning regularized nonnegative matrix factorization[C]//Proceedings of the 21st International Joint Conference on Artificial Intelligence. Pasadena, USA, 2009: 1046–1051.
- [17] LI Xuelong, CUI Guosheng, DONG Yongsheng. Graph regularized non-negative low-rank matrix factorization for image clustering[J]. *IEEE transactions on cybernetics*, 2017, 47(11): 3840–3853.
- [18] DU H, HU Q, ZHANG X, et al. Image feature extraction via graph embedding regularized projective non-negative matrix factorization[J]. *Communications in computer & information science*, 2014, 483(1): 196–209.
- [19] YIN Ming, GAO Junbin, LIN Zhouchen. Laplacian regularized low-rank representation and its applications[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2016, 38(3): 504–517.
- [20] WANG Jianzhong, ZHAO Rui, WANG Yang, et al. Locality constrained graph optimization for dimensionality reduction[J]. *Neurocomputing*, 2017, 245: 55–67.
- [21] MENG Yang, SHANG Ronghua, JIAO Licheng, et al. Dual-graph regularized non-negative matrix factorization with sparse and orthogonal constraints[J]. *Engineering applications of artificial intelligence*, 2018, 69: 24–35.
- [22] NIE Feiping, WANG Xiaoqian, HUANG Heng. Clustering and projected clustering with adaptive neighbors [C]//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA, 2014.
- [23] LEMARÉCHAL C, BOYD S, VANDENBERGHE L. convex optimization, Cambridge university press, 2004 hardback, 65 US\$, ISBN 0 521 83378 7[J]. *European journal of operational research*, 2006, 170(1): 326–327.
- [24] ZHOU Tianyi, TAO Dacheng. GoDec: randomized low-rank & sparse matrix decomposition in noisy case[C]//Proceedings of the 28th International Conference on Machine Learning. Bellevue, USA, 2011: 33–240.
- [25] HUANG Shudong, XU Zenglin, KANG Zhao, et al. Regularized nonnegative matrix factorization with adaptive local structure learning[J]. *Neurocomputing*, 2020, 382(11): 196–209.
- [26] WHITLARK D, DUNTEMAN G H. Principal components analysis[J]. *Journal of marketing research*, 1990, 27(2): 243.

作者简介:



余沁茹, 硕士研究生, 主要研究方向为图像处理与计算机视觉。



卢桂馥, 教授, 主要研究方向为计算机图形学及图像处理。发表学术论文 49 篇。



李华, 硕士研究生, 主要研究方向为图像处理与计算机视觉。