

基于余弦相似度的稀疏非负矩阵分解算法



周 昌^{1,2} 李向利^{1,3} 李俏霖¹ 朱丹丹¹ 陈世莲¹ 蒋丽榕¹

1 桂林电子科技大学数学与计算科学学院 广西 桂林 541004

2 桂林电子科技大学广西高校数据分析与计算重点实验室 广西 桂林 541004

3 桂林电子科技大学广西密码学与信息安全重点实验室 广西 桂林 541004

(2218080688@qq.com)

摘 要 基本的非负矩阵分解应用于图像聚类时,对异常点的处理不够鲁棒,稀疏性较差。为了提高分解后的矩阵的稀疏性,在基本的非负矩阵分解算法中引入了 $L_{2,1}$ 范数,对基本的非负矩阵分解模型进行了改进,从而实现稀疏性,提升算法的性能。同时,为了降低各特征之间的关联,强化非负矩阵分解模型特征的独立性,引入了余弦相似度,提出了基于余弦相似度的稀疏非负矩阵分解算法。该算法在处理高维数据和提取特征方面具有显著优势,并且可提高算法在图像聚类中的辨别准确性。实验结果表明,所提算法在一系列评价指标上的效果优于传统的非负矩阵分解算法。

关键词: 非负矩阵分解;图像聚类; $L_{2,1}$ 范数;余弦相似度

中图法分类号 O235

Sparse Non-negative Matrix Factorization Algorithm Based on Cosine Similarity

ZHOU Chang^{1,2}, LI Xiang-li^{1,3}, LI Qiao-lin¹, ZHU Dan-dan¹, CHEN Shi-lian¹ and JIANG Li-rong¹

1 School of Mathematics and Computing Science, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China

2 Guangxi Colleges and Universities Key Laboratory of Data Analysis and Computation, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China

3 Guangxi Key Laboratory of Cryptography and Information Security, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China

Abstract When the basic non-negative matrix factorization is applied to image clustering, the processing of abnormal points is not robust enough and the sparsity is poor. In order to improve the sparsity of the factorized matrix, the $L_{2,1}$ norm is introduced into the basic non-negative matrix factorization, and the basic non-negative matrix factorization model is improved to achieve sparsity and improve the performance of the algorithm. At the same time, in order to reduce the correlation between the features and enhance the independence of the features of the non-negative matrix factorization model, the cosine similarity is introduced, and a sparse non-negative matrix factorization algorithm based on cosine similarity is proposed. The algorithm has significant advantages in high-dimensional data processing and feature extraction, and can improve the discrimination accuracy of the algorithm in image clustering. The experimental results show that the proposed new algorithm outperforms the traditional non-negative matrix factorization algorithm in a series of evaluation indicators.

Keywords Non-negative matrix factorization, Image clustering, $L_{2,1}$ norm, Cosine similarity

1 引言

非负矩阵分解 (Non-negative Matrix Factorization, NMF)^[1] 是一种基于局部特征的分解方法,通过局部特征的非负线性组合近似构成一个整体。NMF 由 Lee 等^[2] 继 Na-

ture 之后提出,其基本思想是在非负约束下对非负数据进行分解,以获取实验对象的局部特征,而其代数含义为将一个非负矩阵进一步分解为两个非负矩阵的乘积。矩阵分解方法众多,包括主成分分析^[3]、独立成分分析^[4]以及向量量化^[5]等,不同的方法对分解后的两个非负矩阵的限制也不同。

到稿日期:2019-07-17 返修日期:2019-10-18 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(11961010,71561008);广西自然科学基金(2018GXNSFAA138169);广西密码学与信息安全重点实验室研究课题(GCIS201708);广西自动检测技术与仪器重点实验室基金(YQ19111,YQ18112);广西大学生创新创业项目(201810595218)

This work was supported by the National Natural Science Foundation of China (11961010,71561008), Guangxi Natural Science Foundation (2018GXNSFAA138169), Guangxi Key Laboratory of Cryptography and Information Security (GCIS201708), Guangxi Key Laboratory of Automatic Testing Technology and Instruments (YQ19111, YQ18112) and Guangxi University Students Innovation and Entrepreneurship Project (201810595218).

通信作者:李向利(lixiangli@guet.edu.cn)

图像聚类^[6-7]是计算机视觉与模式识别领域的重要研究内容之一,其核心为聚类算法。而在利用传统聚类方法处理数据时发现,数据维数过大将不易于处理,故对高维数据进行降维已成为一种数据处理的常用手段。鉴于非负矩阵分解在进行高维数据的处理和特征提取方面有着明显的优势,其已被广泛应用于聚类算法的研究。目前,很多研究者提出了一些可行的 NMF 聚类算法。Wang 等^[8]将稀疏约束与邻近交替线性化相结合,采用邻近交替线性化算法对凸问题进行求解,并提出稀疏非负矩阵分解算法。Han 等^[9]为了提高 k-means 聚类算法在高维数据下的聚类效果,提出了一种基于稀疏约束非负矩阵分解的 k-means 算法,实现了高维数据的低维表示,并对降维后的数据进行聚类。Lu 等^[10]提出了一种带有鉴别性的非负矩阵分解方法,即非负鉴别矩阵分解算法,该算法有效地提高了分类性能。Pang 等^[11]提出了图正则化和稀疏编码的非负矩阵分解算法。文献[12-16]提出了围绕图像聚类的非负矩阵分解算法。

由于传统的非负矩阵分解应用于图像聚类时,对异常点的处理不够鲁棒,降低了图像聚类的精确率,因此需要对传统的非负矩阵分解模型进行改进。为了提高算法的稀疏性和鲁棒性、有效处理数据中的噪音值和异常值等,本文在基本的 NMF 算法中引入了 $L_{2,1}$ 矩阵范数,对基本的非负矩阵分解模型进行了改进,进而实现稀疏性,提升算法的性能。同时,为了降低各特征之间的关联,强化非矩阵分解模型特征的独立性,本文引入了余弦相似度,提出了基于余弦相似度的稀疏非负矩阵分解算法(Sparse Non-negative Matrix Factorization Algorithm Based on Cosine Similarity, SNMF-BCS),并将该算法用于图像聚类研究。

2 预备知识

2.1 非负矩阵分解

给定一个 $m \times n$ 阶的非负矩阵 $\mathbf{X}$, NMF 的目的是让系数矩阵 $\mathbf{H} \in R^{r \times n}$ 和基矩阵 $\mathbf{W} \in R^{m \times r}$ ($r \ll \min(m, n)$, $\mathbf{X}_i = \mathbf{H}\mathbf{W}_i = \sum_k \mathbf{W}_{ki} \mathbf{H}_k$) 的乘积尽可能接近原始数据矩阵。其公式如下:

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

通过以下优化问题求出 NMF^[17] 的一个解:

$$\min \frac{1}{2} \|\mathbf{X} - \mathbf{WH}\|_F^2, \text{ s. t. } \mathbf{W} \geq 0, \mathbf{H} \geq 0 \quad (2)$$

其中, $\|\cdot\|_F^2$ 是 F 范数的平方。

2.2 $L_{2,1}$ 范数的定义

$L_{2,1}$ 范数的计算公式如下:

$$\|\mathbf{X}\|_{2,1} = \sum_{i=1}^n \sqrt{\sum_{j=1}^m X_{i,j}^2} = \sum_{i=1}^n \|\mathbf{X}_{i,:}\|_2 \quad (3)$$

其中, $\mathbf{X}_{i,:}$ 表示矩阵 $\mathbf{X}$ 的第 i 行, $\mathbf{X}_{:,j}$ 表示矩阵 $\mathbf{X}$ 的第 j 列。

由 $L_{2,1}$ 范数的定义可以看出, $L_{2,1}$ 范数^[18] 的实质为以矩阵 $\mathbf{X}$ 的每一列为单位,求每一列的 L_2 范数(或 F 范数),随后再求 L_1 范数。可以看出,当 $L_{2,1}$ 范数越小,每一行的 L_2 范数越小,意味着行内 0 元素越多,从而实现行稀疏的效果,这说明了 $L_{2,1}$ 范数在 L_1 范数列稀疏的基础上还要求行稀疏。

2.3 余弦相似度

余弦相似度也称为余弦距离,它用向量空间中两个向量夹角的余弦值作为衡量两个个体间差异大小的度量。余弦值越

接近 1,表明夹角越接近 0° ,两个向量越相似,即余弦相似性。

余弦相似度可以表示为:

$$\cos \theta = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2}} \quad (4)$$

其中, $\mathbf{A}$ 和 $\mathbf{B}$ 表示属性向量, A_i 和 B_i 表示向量的分量。

相似度的范围为 -1 至 1 , -1 至 0 表示两个个体的差异性, 1 至 0 则表示两个个体的相似性^[19-20]。但余弦相似度只能分辨维间个体的差异,无法衡量每个维数值的差异,即余弦相似度对数值不敏感^[21]。因此可以将余弦相似度改进为:

$$\text{sim}(i, j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_u)(R_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_u)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_u)^2}} \quad (5)$$

其中, R_u 是第 u 个用户评级的平均值。

3 SNMF-BCS 算法

由于 $L_{2,1}$ 范数有结构化稀疏的特点,因此本文对基本的 NMF 模型进行改进,将 NMF 中的 F 范数改为 $L_{2,1}$ 范数。即:

$$\min \frac{1}{2} \|\mathbf{X} - \mathbf{WH}\|_{2,1}, \text{ s. t. } \mathbf{W} \geq 0, \mathbf{H} \geq 0 \quad (6)$$

其中, $\|\cdot\|_{2,1}$ 是 $L_{2,1}$ 范数。

输入系数矩阵 $\mathbf{H}$ 和基矩阵 $\mathbf{W}$, 计算非负矩阵 $\mathbf{X}$ 如下:

$$\mathbf{X}_i = \mathbf{H}_i \mathbf{W} = \sum_k \mathbf{H}_{ki} \mathbf{W}_k \quad (7)$$

其中, $\mathbf{W}_k$ 是第 k 个潜在特征。

由于潜在特征在 NMF 中有可能是相关的,存在着共适性的缺点,因此本文引入余弦相似度,通过每次迭代更新 $\mathbf{W}_k$ 使得潜在特征之间的余弦相似度极小, $\mathbf{W}_k$ 可以不受潜在特征的影响独立更新,达到执行改进的 NMF 以避免共适性的效果^[22]。

本文中的稀疏非负矩阵分解的目标函数为:

$$\min_{\mathbf{W} \geq 0, \mathbf{H} \geq 0} \frac{1}{2} \|\mathbf{X} - \mathbf{WH}\|_{2,1} + \frac{1}{2} \lambda \sum_{i,j} \cos(\mathbf{W}_i, \mathbf{W}_j) \quad (8)$$

其中, $\sum_{i,j} \cos(\mathbf{W}_i, \mathbf{W}_j)$ 为余弦相似度。

4 迭代更新规则

本文可以将 $\min_{\mathbf{W} \geq 0, \mathbf{H} \geq 0} \frac{1}{2} \|\mathbf{X} - \mathbf{WH}\|_{2,1} + \frac{1}{2} \lambda \sum_{i,j} \cos(\mathbf{W}_i, \mathbf{W}_j)$ 分解为两个子问题,分别固定 $\mathbf{H}$ 和 $\mathbf{W}$ 得到:

$$\min_{\mathbf{W} \geq 0, \mathbf{H} \geq 0} \frac{1}{2} \|\mathbf{X} - \mathbf{WH}\|_{2,1} + \frac{1}{2} \lambda \sum_{i,j} \cos(\mathbf{W}_i, \mathbf{W}_j) \quad (9)$$

$$\min_{\mathbf{W} \geq 0, \mathbf{H} \geq 0} \frac{1}{2} \|\mathbf{X} - \mathbf{WH}\|_{2,1} \quad (10)$$

由于式(9)中 $\|\mathbf{X} - \mathbf{WH}\|_{2,1}$ 非光滑,由文献[18]可知, $L_{2,1}$ 范数具有如下的性质:

$$\|\mathbf{X} - \mathbf{WH}^{k+1}\|_{2,1} - \|\mathbf{X} - \mathbf{WH}^k\|_{2,1} \leq \text{tr}(\mathbf{X} - \mathbf{WH}^k) \mathbf{B}(\mathbf{X} - \mathbf{WH}^k)^T \quad (11)$$

其中, $\text{tr}(\cdot)$ 表示矩阵的迹; $\mathbf{B}$ 为对角阵, $B_{ii} =$

$$\frac{1}{\|\mathbf{X}_i - \mathbf{W}^k \mathbf{H}_i^k\| + \xi}, \xi \text{ 是一个很小的正数。}$$

式(9)可转化为:

$$\min_{\mathbf{W} \geq 0} \frac{1}{2} \text{tr}(\mathbf{X} - \mathbf{WH}^{k+1}) \mathbf{D}(\mathbf{X} - \mathbf{WH}^{k+1})^T + \frac{1}{2} \lambda \sum_{i,j} \cos(\mathbf{W}_i, \mathbf{W}_j) \quad (12)$$

其中, $\mathbf{D}$ 为对角阵, $D_{ii} = \frac{1}{\|X_i - \mathbf{W}^{k+1} \mathbf{H}_i^k\| + \xi}$, ξ 是一个很小的正数。

式(10)可转化为:

$$\min_{H \geq 0} \frac{1}{2} \text{tr}(\mathbf{X} - \mathbf{W}^k \mathbf{H}) \mathbf{B} (\mathbf{X} - \mathbf{W}^k \mathbf{H})^T \quad (13)$$

由于 $\|W_i\| = \|W_j\| = 1$, 因此:

$$\sum_{i,j}^r \cos(W_i, W_j) = r^2 - \text{tr}(\mathbf{W}(\mathbf{r}\mathbf{I} - \mathbf{I}_{r \times r}) \mathbf{W}^T) \quad (14)$$

对其求导得:

$$\begin{aligned} \frac{\partial (\sum_{i,j}^r \cos(W_i, W_j))}{\partial \mathbf{W}} &= \frac{\partial (r^2 - \text{tr}(\mathbf{W}(\mathbf{r}\mathbf{I} - \mathbf{I}_{r \times r}) \mathbf{W}^T))}{\partial \mathbf{W}} \\ &= -2\mathbf{r}\mathbf{W} + 2\mathbf{W}\mathbf{I}_{r \times r} \end{aligned} \quad (15)$$

其中, r 为 $\mathbf{W}$ 的列数, $\mathbf{I}$ 为单位矩阵, $\mathbf{I}_{r \times r}$ 为 $r \times r$ 阶的全 1 矩阵。

令:

$$\begin{aligned} L_{\mathbf{W}} &= \frac{1}{2} \text{tr}(\mathbf{X} - \mathbf{W}\mathbf{H}^{k+1}) \mathbf{D} (\mathbf{X} - \mathbf{W}\mathbf{H}^{k+1})^T + \frac{1}{2} \lambda r^2 - \\ &\quad \frac{1}{2} \lambda \text{tr}(\mathbf{W}(\mathbf{r}\mathbf{I} - \mathbf{I}_{r \times r}) \mathbf{W}^T) \end{aligned} \quad (16)$$

则

$$\begin{aligned} \frac{\partial L_{\mathbf{W}}}{\partial \mathbf{W}_{i,j}} &= -\mathbf{X}\mathbf{D}(\mathbf{H}^{k+1})^T + \mathbf{W}\mathbf{H}^{k+1} \mathbf{D}(\mathbf{H}^{k+1})^T + \\ &\quad \lambda \mathbf{W}\mathbf{I}_{r \times r} - \lambda \mathbf{r}\mathbf{W} \end{aligned} \quad (17)$$

令:

$$L_{\mathbf{H}} = \frac{1}{2} \text{tr}(\mathbf{X} - \mathbf{W}^k \mathbf{H}) \mathbf{B} (\mathbf{X} - \mathbf{W}^k \mathbf{H})^T \quad (18)$$

则

$$\frac{\partial L_{\mathbf{H}}}{\partial \mathbf{H}_{i,j}} = -(\mathbf{W}^k)^T \mathbf{X}\mathbf{B} + (\mathbf{W}^k)^T \mathbf{W}^k \mathbf{H}\mathbf{B} \quad (19)$$

将式(17)、式(19)左右两端分别乘以 W_{ij} 和 H_{ij} , 得:

$$\begin{cases} -\mathbf{X}\mathbf{D}(\mathbf{H}^{k+1})^T \mathbf{W}_{ij} + \mathbf{W}\mathbf{H}^{k+1} \mathbf{D}(\mathbf{H}^{k+1})^T \mathbf{W}_{ij} + \\ (\lambda \mathbf{W}\mathbf{I}_{r \times r} - \lambda \mathbf{r}\mathbf{W}) \mathbf{W}_{ij} = 0 \\ -(\mathbf{W}^k)^T \mathbf{X}\mathbf{B}\mathbf{H}_{ij} + (\mathbf{W}^k)^T \mathbf{W}^k \mathbf{H}\mathbf{B}\mathbf{H}_{ij} = 0 \end{cases} \quad (20)$$

由 KKT 条件得:

$$\begin{cases} \mathbf{W} \geq 0, \mathbf{H} \geq 0 \\ -\mathbf{X}\mathbf{D}(\mathbf{H}^{k+1})^T + \mathbf{W}\mathbf{H}^{k+1} \mathbf{D}(\mathbf{H}^{k+1})^T + \lambda \mathbf{W}\mathbf{I}_{r \times r} - \lambda \mathbf{r}\mathbf{W} \geq 0 \\ -(\mathbf{W}^k)^T \mathbf{X}\mathbf{B} + (\mathbf{W}^k)^T \mathbf{W}^k \mathbf{H}\mathbf{B} \geq 0 \\ -\mathbf{X}\mathbf{D}(\mathbf{H}^{k+1})^T \mathbf{W}_{ij} + \mathbf{W}\mathbf{H}^{k+1} \mathbf{D}(\mathbf{H}^{k+1})^T \mathbf{W}_{ij} + \\ (\lambda \mathbf{W}\mathbf{I}_{r \times r} - \lambda \mathbf{r}\mathbf{W}) \mathbf{W}_{ij} = 0 \\ -(\mathbf{W}^k)^T \mathbf{X}\mathbf{B}\mathbf{H}_{ij} + (\mathbf{W}^k)^T \mathbf{W}^k \mathbf{H}\mathbf{B}\mathbf{H}_{ij} = 0 \end{cases} \quad (21)$$

将式(20)中的等式两端写成非负形式:

$$\mathbf{W}\mathbf{H}^{k+1} \mathbf{D}(\mathbf{H}^{k+1})^T \mathbf{W}_{ij} + \lambda \mathbf{W}\mathbf{I}_{r \times r} \mathbf{W}_{ij} = \mathbf{X}\mathbf{D}(\mathbf{H}^{k+1})^T \mathbf{W}_{ij} + \lambda \mathbf{r}\mathbf{W}\mathbf{W}_{ij} \quad (22)$$

$$(\mathbf{W}^k)^T \mathbf{W}^k \mathbf{H}\mathbf{B}\mathbf{H}_{ij} = (\mathbf{W}^k)^T \mathbf{X}\mathbf{B}\mathbf{H}_{ij} \quad (23)$$

从而得到迭代公式:

$$W_{ij} = W_{ij} \cdot \frac{(\mathbf{X}\mathbf{D}(\mathbf{H}^{k+1})^T + \lambda \mathbf{r}\mathbf{W}\mathbf{W}_{ij})}{(\mathbf{W}\mathbf{H}^{k+1} \mathbf{D}(\mathbf{H}^{k+1})^T + \lambda \mathbf{W}\mathbf{I}_{r \times r})} \quad (24)$$

$$H_{ij} = H_{ij} \cdot \frac{(\mathbf{W}^k)^T \mathbf{X}\mathbf{B}}{(\mathbf{W}^k)^T \mathbf{W}^k \mathbf{H}\mathbf{B}} \quad (25)$$

5 收敛性证明

类似于文献[2], 本文对 SNMF-BCS 算法的收敛性做以下证明。

定义 1 设 $G(\mathbf{h}, \mathbf{h}^k)$ 为 $F(\mathbf{h})$ 的辅助函数, 满足:

$$\begin{aligned} (1) G(\mathbf{h}, \mathbf{h}^k) &\geq F(\mathbf{h}) \\ (2) G(\mathbf{h}, \mathbf{h}) &= F(\mathbf{h}) \end{aligned} \quad (26)$$

引理 1 如果 $G(\mathbf{h}, \mathbf{h}^k)$ 是 $F(\mathbf{h})$ 的辅助函数, 最小化与 $\mathbf{h}$ 有关的函数 $G(\mathbf{h}, \mathbf{h}^k)$ 等同于最小化 $F(\mathbf{h})$ 。代价函数 $F(\mathbf{h})$ 在如下更新规则下是单调递减函数:

$$\mathbf{h}^{k+1} = \arg \min_{\mathbf{h}} G(\mathbf{h}, \mathbf{h}^k) \quad (27)$$

引理 2 设 $\mathbf{R}(\mathbf{h}^k)$ 是对角阵, 则有:

$$\mathbf{R}_{a,b}(\mathbf{h}^k) = \frac{\delta_{a,b} (\mathbf{W}^T \mathbf{W} \mathbf{h}^k \mathbf{B})_a}{h_a^k}, 0 \leq a, b < r \quad (28)$$

则式(29)是式(30)的辅助函数。

$$\begin{aligned} G(\mathbf{h}, \mathbf{h}^k) &= F(\mathbf{h}^k) + (\mathbf{h} - \mathbf{h}^k)^T \nabla F(\mathbf{h}^k) + \\ &\quad \frac{1}{2} (\mathbf{h} - \mathbf{h}^k)^T \mathbf{R}(\mathbf{h}^k) (\mathbf{h} - \mathbf{h}^k) \end{aligned} \quad (29)$$

$$F(\mathbf{h}) = \frac{1}{2} \text{tr}(\mathbf{x} - \mathbf{W}^k \mathbf{h}) \mathbf{B} (\mathbf{x} - \mathbf{W}^k \mathbf{h})^T \quad (30)$$

其中, $\mathbf{B}$ 为对角阵, $B_{ii} = \frac{1}{\|x^k - \mathbf{W}^k \mathbf{h}^k\| + \xi}$, ξ 是一个很小的正数。

$$\delta_{a,b} = \begin{cases} 1, & a=b \\ 0, & a \neq b \end{cases}$$

证明: 根据定义 1, 证明 $G(\mathbf{h}, \mathbf{h}) = F(\mathbf{h})$ 和 $G(\mathbf{h}, \mathbf{h}^k) \geq F(\mathbf{h})$, 有:

$$\begin{aligned} G(\mathbf{h}, \mathbf{h}) &= F(\mathbf{h}) + \underbrace{(\mathbf{h} - \mathbf{h})^T \nabla F(\mathbf{h})}_{=0} + \frac{1}{2} \underbrace{(\mathbf{h} - \mathbf{h})^T \mathbf{R}(\mathbf{h})}_{=0} (\mathbf{h} - \mathbf{h}) \\ &= F(\mathbf{h}) \end{aligned} \quad (31)$$

因此, 定义 1 的条件(2)成立。

现证明条件(1)成立, 函数 F 在点 $\mathbf{h}^k$ 的二阶泰勒展开式为:

$$\begin{aligned} F(\mathbf{h}) &= F(\mathbf{h}^k) + (\mathbf{h} - \mathbf{h}^k)^T \nabla F(\mathbf{h}^k) + \frac{1}{2} (\mathbf{h} - \mathbf{h}^k)^T \mathbf{W}^T \mathbf{W} \mathbf{B} \\ &\quad (\mathbf{h} - \mathbf{h}^k) \end{aligned} \quad (32)$$

由式(29)减去式(32)得:

$$\begin{aligned} G(\mathbf{h}, \mathbf{h}^k) - F(\mathbf{h}) &= (\mathbf{h} - \mathbf{h}^k)^T \mathbf{R}(\mathbf{h}^k) (\mathbf{h} - \mathbf{h}^k) - \\ &\quad (\mathbf{h} - \mathbf{h}^k)^T \mathbf{W}^T \mathbf{W} \mathbf{B} (\mathbf{h} - \mathbf{h}^k) \\ &= (\mathbf{h} - \mathbf{h}^k)^T (\mathbf{R}(\mathbf{h}^k) - \mathbf{W}^T \mathbf{W} \mathbf{B}) (\mathbf{h} - \mathbf{h}^k) \\ &\geq 0 \end{aligned} \quad (33)$$

若想使式(33)成立, 矩阵 $\mathbf{R}(\mathbf{h}^k) - \mathbf{W}^T \mathbf{W} \mathbf{B}$ 要为半正定, 定义为:

$$M_{a,b} = h_a^k (\mathbf{R}(\mathbf{h}^k) - \mathbf{W}^T \mathbf{W} \mathbf{B})_{a,b} h_b^k \quad (34)$$

$\mathbf{R}(\mathbf{h}^k) - \mathbf{W}^T \mathbf{W} \mathbf{B}$ 是半正定的当且仅当 $\mathbf{M}$ 是半正定的。

$$\begin{aligned} \mathbf{x}^T \mathbf{M} \mathbf{x} &= \sum_{0 \leq a, b < r} x_a M_{a,b} x_b \\ &= \sum_{0 \leq a, b < r} x_a h_a^k (\mathbf{R}(\mathbf{h}^k) - \mathbf{W}^T \mathbf{W} \mathbf{B})_{a,b} h_b^k x_b \\ &= \sum_{0 \leq a, b < r} x_a h_a^k R_{a,b}(\mathbf{h}^k) h_b^k x_b - x_a h_a^k (\mathbf{W}^T \mathbf{W} \mathbf{B})_{a,b} h_b^k x_b \\ &= \sum_{0 \leq a, b < r} x_a \delta_{a,b} \underbrace{(\mathbf{W}^T \mathbf{W} \mathbf{h}^k \mathbf{B})_a}_{=0 \text{ for } a \neq b} h_b^k x_b - x_a h_a^k (\mathbf{W}^T \mathbf{W} \mathbf{B})_{a,b} h_b^k x_b \\ &= \sum_{0 \leq a, b < r} x_a^2 h_a^k \underbrace{(\mathbf{W}^T \mathbf{W} \mathbf{h}^k \mathbf{B})_a}_{=0 \text{ for } a \neq b} h_b^k x_b - x_a h_a^k (\mathbf{W}^T \mathbf{W} \mathbf{B})_{a,b} h_b^k x_b \\ &= \sum_{0 \leq b < r} (\mathbf{W}^T \mathbf{W} \mathbf{B})_{a,b} h_b^k x_b^2 \end{aligned}$$

$$\begin{aligned}
&= \sum_{0 \leq a, b < r} h_a^k h_b^k (\mathbf{W}^T \mathbf{W} \mathbf{B})_{a,b} (x_a^2 - x_a x_b) \\
&= \frac{1}{2} \sum_{0 \leq a, b < r} h_a^k h_b^k (\mathbf{W}^T \mathbf{W} \mathbf{B})_{a,b} \left(\frac{1}{2} x_a^2 + \frac{1}{2} x_b^2 - x_a x_b \right) \\
&= \frac{1}{2} \sum_{0 \leq a, b < r} h_a^k h_b^k (\mathbf{W}^T \mathbf{W} \mathbf{B})_{a,b} (x_a - x_b)^2 \geq 0 \quad (35)
\end{aligned}$$

对于 $0 \leq a, b < r$, 得到 $x_a^2 = \frac{1}{2} x_a^2 + \frac{1}{2} x_a^2$, 当 $a = b$ 时, $x_a^2 =$

$\frac{1}{2} x_a^2 + \frac{1}{2} x_a^2$; 当 $a \neq b$ 时, 由于 $\mathbf{W}^T \mathbf{W} \mathbf{B}$ 对称, 因此有 $h_a^k h_b^k$

$(\mathbf{W}^T \mathbf{W} \mathbf{B})_{a,b} = h_b^k h_a^k (\mathbf{W}^T \mathbf{W} \mathbf{B})_{b,a}$ 。

对于 $0 \leq a, b < r$, 有 $h_a^k \geq 0, (\mathbf{W}^T \mathbf{W} \mathbf{B})_{a,b} \geq 0$ 和 $(x_a - x_b)^2 \geq 0$, 因此 $\mathbf{M}$ 是半正定矩阵。那么 $\mathbf{R}(\mathbf{h}^k) - \mathbf{W}^T \mathbf{W} \mathbf{B}$ 是半正定的, $(\mathbf{h} - \mathbf{h}^k)^T (\mathbf{R}(\mathbf{h}^k) - \mathbf{W}^T \mathbf{W} \mathbf{B}) (\mathbf{h} - \mathbf{h}^k) \geq 0$, 得到 $G(\mathbf{h}, \mathbf{h}^k) \geq F(\mathbf{h})$ 。因此 $G(\mathbf{h}, \mathbf{h}^k)$ 是 $F(\mathbf{h})$ 的辅助函数。

5.1 最小化辅助函数

根据定理 1, 最小化与 h 有关的辅助函数 $G(\mathbf{h}, \mathbf{h}^k)$ 得到更新规则 h^{k+1} 。对式(29)求偏导可得:

$$\begin{aligned}
\frac{\partial G(\mathbf{h}, \mathbf{h}^k)}{\partial \mathbf{h}} &= \nabla F(\mathbf{h}^k) + \mathbf{R}(\mathbf{h}^k) (\mathbf{h} - \mathbf{h}^k) = 0 \\
&\Rightarrow \mathbf{R}(\mathbf{h}^k) (\mathbf{h} - \mathbf{h}^k) = -\nabla F(\mathbf{h}^k) \\
&\Rightarrow \mathbf{h} = \mathbf{h}^k - \mathbf{R}(\mathbf{h}^k)^{-1} \nabla F(\mathbf{h}^k) \quad (36)
\end{aligned}$$

5.2 更新规则

由式(27)可得:

$$\mathbf{h}^{k+1} = \arg \min_{\mathbf{h}} G(\mathbf{h}, \mathbf{h}^k) = \mathbf{h}^k - \mathbf{R}(\mathbf{h}^k)^{-1} \nabla F(\mathbf{h}^k) \quad (37)$$

由于 $\mathbf{R}$ 是对角阵, 因此:

$$\mathbf{R}_{a,b}^{-1}(\mathbf{h}^k) = \delta_{a,b} \frac{h_a^k}{(\mathbf{W}^T \mathbf{W} \mathbf{B} \mathbf{h}^k)_a} \quad (38)$$

函数 $F(h_a^k)$ 的偏导数为:

$$\frac{\partial F}{\partial h_a^k} = (-\mathbf{W}^T \mathbf{B} \mathbf{x})_a + (\mathbf{W}^T \mathbf{W} \mathbf{B} \mathbf{h}^k)_a \quad (39)$$

$\mathbf{W}_a$ 是矩阵 $\mathbf{W}$ 的第 a 列, 把式(38)代入式(37), 并且只表示向量 $\mathbf{h}^k$ 的第 a 个元素, 可得:

$$\begin{aligned}
h_a^{k+1} &= h_a^k - \delta_{a,b} \frac{h_a^k}{(\mathbf{W}^T \mathbf{W} \mathbf{B} \mathbf{h}^k)_a} \nabla_a F(\mathbf{h}^k) \\
h_a^{k+1} &= h_a^k - \frac{h_a^k}{(\mathbf{W}^T \mathbf{W} \mathbf{B} \mathbf{h}^k)_a} \cdot \frac{\partial F}{\partial h_a^k} \\
&= h_a^k \left(1 - \frac{(\mathbf{W}^T \mathbf{W} \mathbf{B} \mathbf{h}^k)_a - (\mathbf{W}^T \mathbf{x} \mathbf{B})_a}{(\mathbf{W}^T \mathbf{W} \mathbf{B} \mathbf{h}^k)_a} \right) \\
&= h_a^k \frac{(\mathbf{W}^T \mathbf{x} \mathbf{B})_a}{(\mathbf{W}^T \mathbf{W} \mathbf{B} \mathbf{h}^k)_a} \quad (40)
\end{aligned}$$

式(40)为矩阵 $\mathbf{H}$ 中的第 k 列第 a 个元素的更新规则, 定理 1 中的更新规则已被证明, 同理可获得矩阵 $\mathbf{W}$ 的更新规则。

6 实验结果与分析

传统的 NMF 算法将原始数据分 $\mathbf{X}$ 解为 $\mathbf{W}$ 和 $\mathbf{H}$, $\mathbf{W}$ 为基矩阵, $\mathbf{H}$ 为系数矩阵。本文对 $\mathbf{W}$ 进行了余弦相似度处理, 并与 F 范数-NMF、 $L_{2,1}$ 范数-NMF 进行了对比, 使用图像聚类中常用的准确率 (Accuracy, ACC)^[23] 和归一化互信息 (Normalized Mutual Information, NMI)^[23] 来评价结果。ACC 是指对于给定的测试数据集, 其正确分类的样本数与总样本数之比。NMI 常用于度量两个聚类结果的相似程度。它们的计算公式分别如下:

$$\text{ACC} = \frac{1}{N} \sum_{i=1}^N \delta(C_i, \text{map}(C_i')) \quad (41)$$

$$\text{NMI} = 2 \frac{MI(C, C')}{\max(H(C), H(C_i'))} \quad (42)$$

其中, N 是数据样本总数, $\delta(x, y) = \begin{cases} 1, & x=y \\ 0, & \text{其他} \end{cases}$, $\text{map}(\cdot)$ 是从 C_i 到 C_i' 的最佳映射函数。 $H(C)$ 和 $H(C_i')$ 分别表示 C 和 C_i' 的信息熵, MI 为两个标签集合的互信息。

6.1 数据集

本文提出的模型中存在一个参数 λ , 为了确定其取值, 在 Yale_32×32, Yale_64×64, WarpAR10P, orlraws10P, Pi×raws10P, ORL_32×32, ORL_64×64, COIL20 等数据集上进行实验。

Yale 数据集是由 15 人的 165 张图像构成, 其中图像存在光照、表情和姿态等变化。Yale_32×32 中每幅图像为 32×32 维, Yale_64×64 中每幅图像为 64×64 维。WarpAR10P 数据集包含 130 个样例, 2 400 个特征, 10 个类别。ORL 数据集包含 40 人, 每个人有 10 幅图像, 其中部分图像存在姿态、表情和面部饰物的变化。ORL_32×32 中每幅图像为 32×32 维, ORL_64×64 中每幅图像为 64×64 维。COIL20 数据集包含不同角度的 20 个物体的灰度人脸图像, 实验中每个物体包含 72 幅图像。此外, 还有 orlraws10P 数据集、Pixraws10P 数据集。

6.2 实验环境

本文提出的 SNMF-BCS 算法用 MATLAB R2016a 实现, 运行于 IntelCorei5-6300HQ CPU@ 2.30 GHz 处理器的个人电脑, 运行环境为 Windows 10。

6.3 实验结果

针对本文的所有测试数据, 经过多次实验测试发现, 在 ACC 指标下 λ 的取值范围为 0~20。ACC 随参数 λ 的具体变化情况如图 1—图 8 所示。

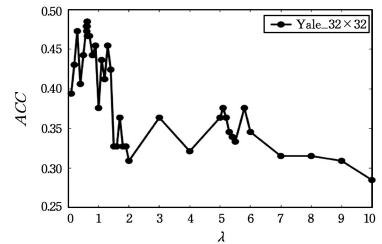


图 1 SNMF-BCS 算法在 Yale_32×32 数据集集中的 ACC 随参数 λ 的变化情况

Fig. 1 ACC of SNMF-BCS algorithm on Yale_32×32 data set changes as parameter λ changes

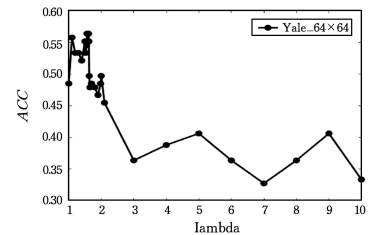


图 2 SNMF-BCS 算法在 Yale_64×64 数据集集中的 ACC 随参数 λ 的变化情况

Fig. 2 ACC of SNMF-BCS algorithm on Yale_64×64 data set changes as parameter λ changes

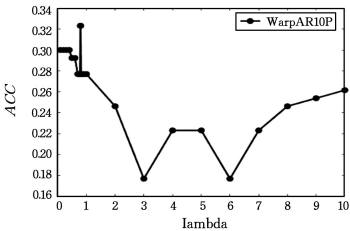


图 3 SNMF-BCS 算法在 WarpAR10P 数据集中的 ACC 随参数 λ 的变化情况

Fig. 3 ACC of SNMF-BCS algorithm on WarpAR10P data set changes as parameter λ changes

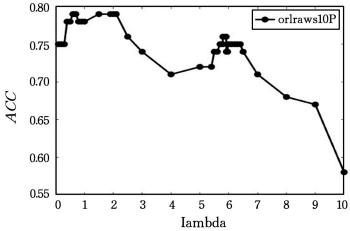


图 4 SNMF-BCS 算法在 orlraws10P 数据集中的 ACC 随参数 λ 的变化情况

Fig. 4 ACC of SNMF-BCS algorithm on orlraws10P data set changes as parameter λ changes

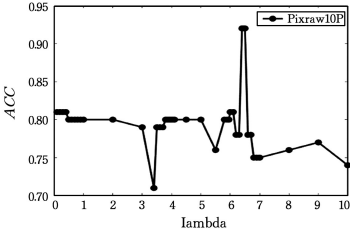


图 5 SNMF-BCS 算法在 Pixraw10P 数据集中的 ACC 随参数 λ 的变化情况

Fig. 5 ACC of SNMF-BCS algorithm on Pixraw10P data set changes as parameter λ changes

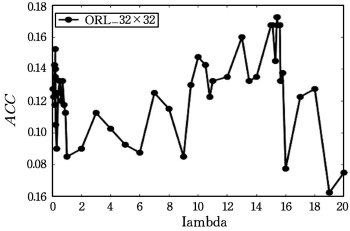


图 6 SNMF-BCS 算法在 ORL_32x32 数据集中的 ACC 随参数 λ 的变化情况

Fig. 6 ACC of SNMF-BCS algorithm on ORL_32x32 data set changes as parameter λ changes

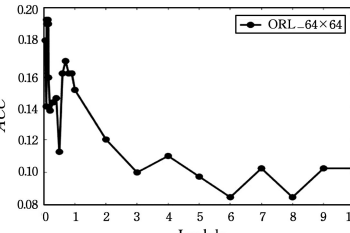


图 7 SNMF-BCS 算法在 ORL_64x64 数据集中的 ACC 随参数 λ 的变化情况

Fig. 7 ACC of SNMF-BCS algorithm on ORL_64x64 data set changes as parameter λ changes

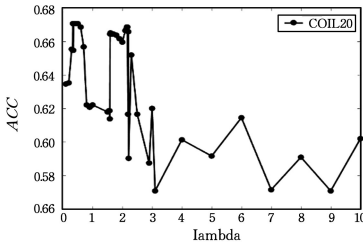


图 8 SNMF-BCS 算法在 COIL20 数据集中的 ACC 随参数 λ 的变化情况

Fig. 8 ACC of SNMF-BCS algorithm on COIL20 data set changes as parameter λ changes

由图 1—图 8 可知,当 λ 在数据集 Yale_32x32, Yale_64x64, WarpAR10P, orlraws10P, Pixraws10P, ORL_32x32, ORL_64x64, COIL20 上分别取 0.622~0.623, 1.58~1.59, 0.79~0.8, 2.6, 4~6.5, 15.4, 0.12, 0.34 时,数据集聚类效果中的 ACC 较高。

实验中各算法在不同数据集上的 ACC 和 NMI 值如表 1 所列。

表 1 各算法在数据集上的比较
Table 1 Comparison of algorithms on data sets

数据集	算法	ACC	NMI
Yale_32x32	$L_{2,1}$ 范数-NMF	0.41818	0.47759
	F 范数-NMF	0.43636	0.44841
	SNMF-BCS	0.48485	0.51030
Yale_64x64	$L_{2,1}$ 范数-NMF	0.48485	0.53411
	F 范数-NMF	0.38788	0.47437
	SNMF-BCS	0.56364	0.59357
WarpAR10P	$L_{2,1}$ 范数-NMF	0.30000	0.28792
	F 范数-NMF	0.26154	0.29462
	SNMF-BCS	0.32308	0.31457
orlraws10P	$L_{2,1}$ 范数-NMF	0.7500	0.79110
	F 范数-NMF	0.7600	0.77191
	SNMF-BCS	0.7900	0.79583
Pixraws10P	$L_{2,1}$ 范数-NMF	0.8200	0.87795
	F 范数-NMF	0.8100	0.86340
	SNMF-BCS	0.9200	0.90799
ORL_32x32	$L_{2,1}$ 范数-NMF	0.1900	0.67721
	F 范数-NMF	0.1250	0.68517
	SNMF-BCS	0.17250	0.70263
ORL_64x64	$L_{2,1}$ 范数-NMF	0.1600	0.69839
	F 范数-NMF	0.18250	0.71097
	SNMF-BCS	0.19250	0.72009
COIL20	$L_{2,1}$ 范数-NMF	0.63472	0.74130
	F 范数-NMF	0.56875	0.69829
	SNMF-BCS	0.67083	0.74484

由表 1 可以看出,在不同的数据集中,SNMF-BCS 算法的聚类性能 ACC 和 NMI 值均高于 $L_{2,1}$ 范数-NMF 和 F 范数-NMF,只有在数据集 ORL_32x32 中 SNMF-BCS 算法的 ACC 值为 0.1725,接近于 $L_{2,1}$ 范数-NMF 算法的 ACC 值 0.19。因此,综合来看,SNMF-BCS 算法相对于 $L_{2,1}$ 范数-NMF 和 F 范数-NMF 有着更好的聚类效果。

在 Yale 数据集和 ORL 数据集中,无论是高维还是低维,SNMF-BCS 算法的 NMI 值均高于 $L_{2,1}$ 范数-NMF 算法和 F 范数-NMF 算法的 NMI 值,而 SNMF-BCS 算法的 ACC 值仅在 ORL 低维数据中的表现接近于 $L_{2,1}$ 范数-NMF。因此,综合来看,SNMF-BCS 算法在处理高维和低维数据时比 $L_{2,1}$ 范

数-NMF、 F 范数-NMF 的效果更好。

$L_{2,1}$ 范数-NMF 算法在处理 Yale 数据集时,聚类性能在高维中的表现均优于低维;在处理 ORL 数据集时,ACC 在高维数据中的值低于低维,但其 NMI 值高于低维数据的 NMI 值。因此,可以看出 $L_{2,1}$ 范数-NMF 在处理高维和低维数据时效果并不稳定。

F 范数-NMF 在处理 Yale 数据集时,高维数据的 ACC 值低于低维数据的 ACC 值,但 NMI 值高于低维数据的 NMI 值;在处理 ORL 数据集时,高维中的聚类性能均优于低维。因此,可以看出 F 范数-NMF 在处理高维和低维数据时效果并不稳定。

SNMF-BCS 算法在处理 Yale 和 ORL 时,聚类效果在高维中的表现均优于低维。因此,SNMF-BCS 算法在处理高维和低维数据时效果较稳定,并且在处理高维数据方面更具优势。

结束语 本文提出了一种基于余弦相似度的稀疏非负矩阵分解算法,即利用 $L_{2,1}$ 范数对原标准矩阵分解(NMF)中的 F 范数进行替换,并结合语义空间利用余弦相似度的特点,来提高算法的稀疏性。在人脸数据集实验中与原标准矩阵分解算法、原标准 $L_{2,1}$ 范数以及 F 范数分解算法的结果相比,本文算法均显示出更好的聚类性能,从而验证了本文所提出的基于余弦相似度的稀疏非负矩阵分解算法的有效性和可行性。本文提出的算法只考虑到了余弦相似度,下一步将推广到与图正则相结合。

参 考 文 献

- [1] PAATERO P, TAPPER U. Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values [J]. *Environmetrics*, 2010, 5(2): 111-126.
- [2] LEE D D, SEBASTIAN S H. Learning the parts of objects by nonnegative matrix factorization [J]. *Nature*, 1999, 401: 788 - 791.
- [3] JOLLIFFE I T. *Principal component analysis* [M]. Springer-Verlag, 2005.
- [4] DRAPER B A, BEAK K, BARTLETT M S, et al. Recognizing faces with pca and ica [J]. *Computer Vision \ Image Understanding*, 2003, 91(1): 115-137.
- [5] GERSHO A, GRAY R, GERSHO A, et al. *Vector quantization and signal compression* [J]. Springer International, 1992, 159(1): 407-485.
- [6] LIAO Q, ZHANG Q. Local coordinate based graph-regularized NMF for image representation[J]. *Signal Processing*, 2016, 124: 103-114.
- [7] EGGERT J, WERSING H, KORNER E. Transformation-invariant representation and NMF [C] // 2004 IEEE International Joint Conference on Neural Networks, 2004. IEEE, 2004.
- [8] WANG J, YANG D. Sparse non-negative matrix factorization algorithm based on proximal alternating linearized minimization [J]. *Computer Engineering*, 2019, 45(2): 220-225, 232.
- [9] HAN S Q, JIA R. K-Means clustering algorithm based on non-negative matrix factorization with sparseness constraints [J]. *Journal of Data Acquisition and Processing*, 2017, 32(6): 1216-1222.
- [10] LU Y W. Research on manifold reression and non-negative matrix factorization in image classification[D]. Harbin: Harbin Institute of Technology, 2015.
- [11] PANG M. Matrix factorization and its application ressearch on image classification and clustering[D]. Dalian: Dalian University of Technology, 2015.
- [12] TAO X L, YU L, WANG X Y. One method based on non-negative matrix factorization and fuzzy C means for image clustering [J]. *Information Technology and Network Security*, 2019, 38(3): 44-48.
- [13] XI S N. Research on image recognition based on semi-supervised non-negative matrix Factorization[D]. Xi'an: Xidian University, 2017.
- [14] DU S Q, SHI Y Q, WANG W N. Manifold regularized-based discriminant concept factorization[J]. *Journal of Shandong University*, 2013, 48(5): 63-69.
- [15] HE M J. Research on the Multi-view clustering algorithms based on NMF [D]. Chengdu: Southwest Jiaotong University, 2017.
- [16] JIANG X Y. Research on image classification algorithm based on non-negative matrix factorization[D]. Jinzhou: Liaoning University of Technology, 2017.
- [17] LEE D D, SEUNG H S. Learning the parts of objects by non-negative matrix factorization[J]. *Nature*, 2001, 401(6755): 788-791.
- [18] KONG D, DING C, HUANG H. Robust nonnegative matrix factorization using L21-norm [C] // *Proceedings of CIKM*. 2011: 673-682.
- [19] TAN P N, STEINBACH M, KUMAR V. *Introduction to data mining* [M]. Toronto: Addison-Wesley Longman, 2005: 20-23.
- [20] OCHIAI A. Zoogeographical Studies on the Soleoid Fishes Found in Japan and its Neighbouring Regions-III [J]. *Nippon Suisan Gakkaishi*, 1957, 22(9): 531-535.
- [21] HUANG H, LIU Y J, XIE Q. An Expert Finding Method for CQA Service Based on "User-Tag" Heterogeneous Network [J/OL]. *Computer Engineering*, 2019, 1-6. <https://doi.org/10.19678/j.issn.1000-3428.0053734>.
- [22] HUANG W C, ZHAO Y, XIONG L Y. Sparse non-negative Matrix factorization algorithm for independent feature learning[J]. *Application Research of Computers*, 2019, 37(4): 1-5.
- [23] YU J L, LI X L, ZHAO P F. Sparsenon-negative matrix factorization based on kernel and hypergraph regularization[J]. *Journal of Computer Applications*, 2019, 39(3): 742-749.



ZHOU Chang, born in 1998, postgraduate. His main research interests include image clustering and so on.



LI Xiang-li, born in 1977, Ph.D, professor. Her main research interests include image clustering, non-negative matrix factorization, and optimization.