

基于相关熵和流形正则化的图像聚类

时照群, 刘兆伟*, 刘惊雷*

(烟台大学计算机与控制工程学院, 烟台, 264005)

摘要:近年来,聚类作为机器学习、数据挖掘等领域的基本问题受到广泛的关注及研究,然而数据中普遍存在的噪声和异常值严重影响聚类结果. 提出一个基于相关熵和流形正则化的聚类框架 CRNMF (Correntropy and Manifold Regularization Non-Negative Matrix Factorization). 首先,采用基于相关熵的非负矩阵分解(Non-Negative Matrix Factorization, NMF)作为损失函数来抑制非高斯噪声和异常值的影响;其次,充分考虑数据的结构信息,采用流形正则化学习数据的局部结构,并通过 $l_{2,1}$ -范数对非负矩阵进行稀疏约束;最后,利用半二次优化技术(Half-Quadratic Optimization Technique, HQ)进行优化,并分析了收敛性和计算复杂度. 在五个图像数据集上进行测试,实验结果表明,提出的框架在图像聚类任务中具有较好的有效性和鲁棒性.

关键词:非负矩阵分解,相关熵,流形正则化,半二次优化技术,图像聚类

中图分类号:TP391

文献标志码:A

Image clustering based on correntropy and manifold regularization

Shi Zhaoqun, Liu Zhaowei*, Liu Jinglei*

(School of Computer and Control Engineering, Yantai University, Yantai, 264005, China)

Abstract: As a basic problem in machine learning, data mining and other fields, clustering has received extensive attention and research in recent years. However, the widespread noise and outliers of data affect the clustering results seriously. Therefore, a clustering framework based on correntropy and manifold regularization (CRNMF) is proposed. First, the Non-Negative Matrix Factorization (NMF) based on correntropy is used as the loss function to suppress the effects of non-Gaussian noise and outliers. Second, the structure information of the data is fully considered, the manifold regularization is used to learn the local structure of the data, and the sparse constraint is applied to the non-negative matrix by $l_{2,1}$ -norm. Finally, the Half-Quadratic Optimization Technique (HQ) is used to optimize the algorithm, and the convergence and computational complexity are analyzed. Experimental results show that the proposed CRNMF is effective and robust on five image datasets.

Key words: Non-Negative Matrix Factorization (NMF), correntropy, manifold regularization, Half-Quadratic Optimization Technique (HQ), image clustering

聚类的目的是将相似度高的数据点聚为一类,将相似度低的数据点聚为不同类. 由于聚类在机器学习、数据挖掘、计算机视觉等领域受到广泛关注,如何降低噪声和异常值对聚类结果的影

响已成为一个重要课题. 几十年来,大量的聚类算法被提出,例如 k -means^[1]、谱聚类(Spectral Clustering, SC)^[2]和非负矩阵分解(Non-Negative Matrix Factorization, NMF)^[3]等. k -means 既是基

基金项目:国家自然科学基金(62072391, 62172351),山东省自然科学基金(ZR2020MF148)

收稿日期:2022-01-11

* 通讯联系人, E-mail: lzw@ytu.edu.cn, jinglei_liu@sina.com

于划分的聚类算法,也是基于矩阵分解的聚类算法,虽然实现速度快,但无法处理非线性数据.谱聚类可以用来处理非线性数据,但其计算成本较高.大多数情况下,现实世界聚类任务中遇到的数据维数非常高,传统的聚类算法及其变种^[4-6]因为噪声和冗余的影响,处理高维数据的性能不佳,因此,发现高维数据的低维表示尤为重要.由于 NMF 对数据表示良好的可解释性和明确的物理意义,已成为最流行的维度还原算法之一^[7-8].本文主要研究 NMF 在聚类中的应用.

传统矩阵分解有低秩表示(Latent Low-Rank Representation, LRR)^[9]、主成分分析(Principal Components Analysis, PCA)^[10]和奇异值分解(Singular Value Decomposition, SVD)^[11]等.与它们不同, NMF 的目标是学习两个低维非负矩阵,即分解后的两个非负矩阵的乘积能够逼近原始数据矩阵.由于施加在低维子空间中分解矩阵上的非负约束条件且只允许学习矩阵的加法组合, NMF 只学习基于部分的数据表示^[12-13].更重要的是, NMF 符合大脑中人类认知过程的直观概念,即结合部分形成整体,这已在心理学和生理学中得到研究^[14-15].因此, NMF 可以被认为是一种优秀的降维算法,并已成功应用于包括文本分析和图像处理在内的各种实际应用中.一般地,大多数传统 NMF 及其扩展算法中的矩阵分解的目标函数或优化准则是基于欧式距离或 KL (Kullback-Leibler)散度,因为它们具有简单、良好的数学性质和对高斯噪声不敏感等优点.然而在许多实际应用中,数据通常包含不同类型的非高斯噪声和异常值,这种情况下,传统 NMF 算法及其扩展的性能很差.在 Lee and Seung^[12]之后,受到上述原始 NMF 优越特性的激励, NMF 受欢迎的程度日益提高.随后,为了解决不同的问题,提出了许多以 NMF 为代表的变体和扩展.但是,如果不考虑数据的流形结构,传统 NMF 算法的性能依旧很差.因此,保持数据流形结构的重要性在机器学习、数据挖掘和模式识别的研究领域引起了相当大的关注^[16-18].例如,图正则化非负矩阵分解(Graph Regularized Non-Negative Matrix Factorization, GNMF)充分利用数据的固有几何结构,通过拉普拉斯正则化将数据空间建模为嵌入

在环境空间中的流形. GNMF 同时考虑了原始空间中数据点的线性和非线性关系,与只考虑欧氏距离的普通 NMF 区别很大^[19-20],因此 GNMF 比普通 NMF 更适合聚类问题.在此基础上, Guan et al^[21]充分考虑数据的几何结构和不同类别的判别信息,提出流形正则化判别非负矩阵分解算法(Manifold Regularized Discriminative Non-Negative Matrix Factorization, MD-NMF). Shang et al^[22]提出一种考虑样本流形和特征流形的图对偶正则化算法(Graph Dual Regularization Non-Negative Matrix Factorization, DNMF),用于处理共聚类问题. Zeng et al^[23]提出超图正则化算法(Hyper-Graph Regularized Non-Negative Matrix Factorization, HNMF)来更好地反映数据中固有的几何结构. Li et al^[24]同时考虑标签信息和局部流形正则化,提出基于图的判别非负矩阵分解算法(Graph-Based Discriminative Non-Negative Matrix Factorization, GDNMF),通过构造 p -近邻图对数据的局部几何结构进行编码,并在 NMF 中考虑标签信息,增强了聚类表示的区分能力.

以上算法都使用欧式距离来最小化原始数据矩阵和重构矩阵之间的距离,由于欧式距离的可解释性和可处理性,已被广泛应用于基于 NMF 的算法中.然而,基于欧式距离的 NMF 对非高斯噪声和异常值敏感,进行聚类任务时性能显著降低,因此,相关熵被提出用于处理基于 NMF 的任务. Wang et al^[25]将相关熵引入 GNMF,提出一种用于图像聚类的非负矩阵分解算法(Correntropy Induced Metric Based Graph Regularized Non-Negative Matrix Factorization, CGNMF),在考虑数据流形结构的同时增强算法的鲁棒性. Du et al^[26]提出一种基于相关熵诱导度量的鲁棒算法(Robust Non-Negative Matrix Factorization Method Based on the Correntropy Induced Metric, CIM-NMF),通过结合关于离群点的结构知识导出用于处理离群点的基于行的 CIM-NMF. Li et al^[27]通过最大化用于非可视化图像聚类的相关熵准则,发展图正则化的非负矩阵分解算法(Graph Regularized Non-Negative Matrix Factorization by Maximizing Correntropy, GRNMF). Zhou et al^[28]提出一种无监督的特征选择算法,称为基于相关熵的全局和局部

保持鲁棒稀疏子空间学习算法(Global and Local Structure Preserving Robust Sparse Subspace Learning, GLoRSS). 谢林江和尹东^[29]提出基于最大相关熵准则的鲁棒度量学习算法,将相关熵引入度量学习来更好地实现分类. Yu et al^[30]提出用于癌症聚类的超图正则化非负矩阵分解算法(Correntropy-Based Hypergraph Regularized Non-Negative Matrix Factorization, CHNMF),用相关熵代替欧氏距离来提高鲁棒性并引入超图正则化项,在癌症数据集上效果较好.

同时,引入范数对矩阵进行稀疏约束来发展一些稳健的NMF算法. Peng et al^[31]提出一种基于相关熵的鲁棒非负矩阵分解算法(l_1 -Norm Non-Negative Matrix Factorization Based on Maximum Correntropy Criterion, l_1 -CNMF),将传统的NMF算法与 l_1 -范数和最大相关熵准则结合,对系数矩阵进行稀疏约束,在图像聚类任务中取得了较好的效果. Zhang et al^[32]提出基于 $l_{2,1}$ -范数的矩阵分解来获得抗噪声和离群点的鲁棒解. Chen et al^[33]提出基于流形正则化和稀疏学习的鲁棒非负矩阵分解算法(Robust Non-Negative Matrix Factorization via Jointly Manifold Regularization and Sparse Learning, MS-RNMF),在NMF中引入 $l_{2,1}$ -范数约束,同时对基矩阵和系数矩阵进行稀疏约束.

然而,传统的基于NMF的聚类算法仍然具

有以下问题:(1)容易受非高斯噪声和异常值的影响而产生较大的误差;(2)无法充分利用数据的结构信息;(3)没有利用矩阵的稀疏性,导致计算复杂,优化时间长.

受上述工作启发,为了充分保持数据的结构信息并继承传统NMF算法的优点,本文提出基于相关熵和流形正则化的图像聚类算法(CRNMF). 在提出的框架中引入相关熵代替欧式距离来减轻非高斯噪声的影响. 另外,充分考虑数据的结构信息并通过 $l_{2,1}$ -范数进行稀疏约束,减轻异常值的影响. CRNMF算法的流程如图1所示.

本文的主要贡献如下:

(1)针对图像聚类问题,提出一个基于相关熵和流形正则化的联合优化框架CRNMF,将相关熵、稀疏约束和流形正则整合到一个统一的目标函数中.

(2)通过相关熵代替欧式距离,通过流形正则化保存数据的局部结构,并引入 $l_{2,1}$ -范数对非负矩阵进行稀疏约束,增强了算法的抗非高斯噪声能力和局部平滑性.

(3)利用半二次优化技术(Half-Quadratic Optimization Technique, HQ)对CRNMF进行优化,将非凸目标函数转化为乘法形式的二次函数,导出算法的更新规则,同时在理论上证明了算法的收敛性,并分析了计算复杂度.

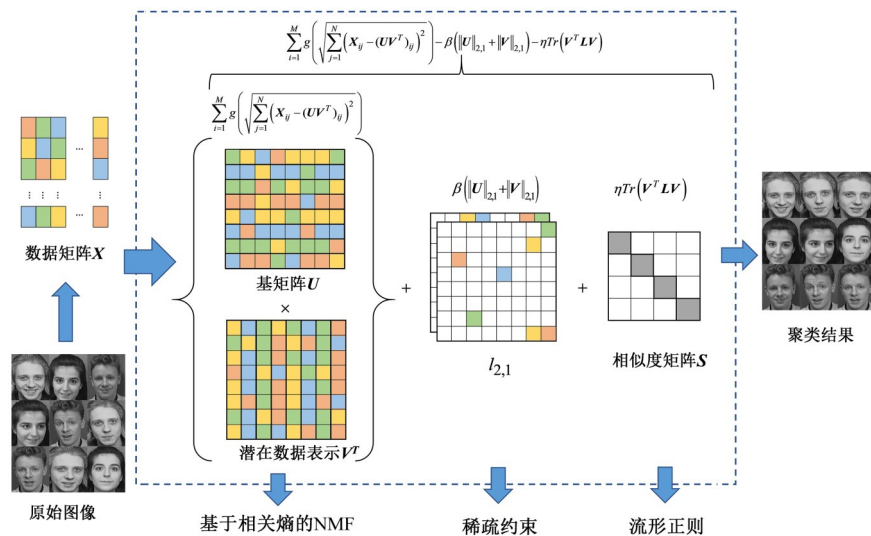


图1 CRNMF 框架

Fig. 1 The framework of CRNMF

(4)在五个图像数据集上进行实验,实验结果表明,提出的算法在图像聚类任务中有较好的有效性和鲁棒性.

1 相关工作

首先,本文工作使用的符号如表 1 所示.

表 1 相关符号

Table 1 Related notations

符号	描述
$X \in \mathbb{R}^{M \times N}$	数据矩阵
U	基矩阵
V	潜在数据表示矩阵
L	拉普拉斯矩阵
S	相似度矩阵
D	度矩阵
H, M_1, M_2	对角矩阵
σ	核带宽参数
ρ	图最近邻数
β	稀疏参数
η	正则化参数
λ	权衡参数
ϱ	变量标量
M	特征维度
N	样本数量
K	嵌入维度
$E[\cdot]$	期望算子
$\kappa(\cdot, \cdot)$	核函数
$[\cdot]^T$	转置运算符
$Tr(\cdot)$	矩阵的迹
$\ \cdot\ _F$	Frobenius 范数
$\ \cdot\ _{2,1}$	$l_{2,1}$ 范数

1.1 非负矩阵分解 (NMF) NMF 的主要思想是将非负数据矩阵近似分解为两个低维非负矩阵,其乘积尽可能逼近原数据矩阵以找到原始数据矩阵的恰当低维表示.

假设 $X = [x_1, x_2, \dots, x_N] \in \mathbb{R}_{\geq 0}^{M \times N}$ 为原始数据矩阵, X 的每一列表示包含 M 个元素的样本向量, NMF^[3] 可以表示为:

$$X \approx UV^T \quad (1)$$

其中, $K < \min(N, M)$ 为正整数, $U = [u_{ik}] \in \mathbb{R}_{\geq 0}^{M \times K}$

和 $V = [v_{jk}] \in \mathbb{R}_{\geq 0}^{N \times K}$ 分别表示基矩阵和系数矩阵.

多数情况下,传统的 NMF 通常使用欧式距离作为度量,表示为以下优化问题:

$$\mathcal{O}_{SED} = \sum_{j=1}^N \sum_{i=1}^M \left(X_{ij} - (UV^T)_{ij} \right)^2 = \|X - UV^T\|_F^2 \quad (2)$$

其中, $\|\cdot\|_F$ 为 Frobenius 范数, $[\cdot]^T$ 为转置运算符.

虽然 U 或 V 在式 (2) 中是非凸的,但不是同时非凸. 因此,传统 NMF 的乘法更新规则如下:

$$U \leftarrow U \frac{XV}{UV^TV} \quad (3)$$

$$V \leftarrow V \frac{X^TU}{VU^TV} \quad (4)$$

实际上,通常 $K \ll M$ 并且 $K \ll N$. 因此, NMF 基本上为寻找原始数据矩阵的压缩近似. 近似值按列查看如下所示:

$$x_j \approx \sum_{k=1}^K u_k v_{jk} \quad (5)$$

其中, u_k 为 U 的第 k 列向量. 因此,数据向量 x_j 由 U 的列的线性组合近似,由 V 的分量加权.

与其他矩阵分解算法最大的不同是, NMF 对 U 和 V 的非负约束只允许碱基之间的加性组合,因此 NMF 是基于部分的表示. 由于基于部分表示的优点, NMF 已经被应用于许多现实世界的任务,如聚类任务、人脸分析、基因表达分析等^[19].

1.2 流形正则化 在许多现实世界的应用中,数据点常位于一个低维的流形上,因此流形学习引起了巨大的关注. 通常,邻近的数据点可能有相似的特征并被分类或划分到同一组中. 假设邻近的数据点在新的数据空间中也是相近的,称为“流形假设”^[34]. 将“流形假设”引入非负矩阵分解中,即局部流形学习,以期望小邻域内的邻近数据点可以共享内在的几何结构.

为了捕捉数据的内在几何结构使用基于图的流形方法,使用热核函数构建最近邻图来度量数据点之间的相似度. 若两个数据点在原始空间中相似则在目标空间中也相似. 给定具有 N 个数据点的数据集 $X = \{x_1, x_2, \dots, x_N\}$, 根据其中任意两点间的相似度建立相似度矩阵 $S \in \mathbb{R}^{N \times N}$, 定义:

$$S = \begin{cases} e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}, & i \text{ 与 } j \text{ 相邻} \\ 0, & \text{否则} \end{cases} \quad (6)$$

其中, λ 是一个权衡参数.

为了保留局部信息, 结合流形正则化和局部相似度使数据点达到无差异^[33]. v_i 和 v_j 是与数据点 x_i, x_j 关联的特征向量. 通过 $\frac{1}{2} \min \sum_{i,j} \|v_i - v_j\|^2$ 测量数据点 v_i 和 v_j 之间的微小差异. 因此, 局部相似度得分越高则数据点之间的差异越小, 定义如下:

$$\mathcal{O}_{\text{data}} = \sum_{i,j=1}^N d(v_i, v_j) \times S = \text{Tr}(V^T D V) - \text{Tr}(V^T S V) = \text{Tr}(V^T L V) \quad (7)$$

其中, $\text{Tr}(\cdot)$ 表示矩阵的迹, 度矩阵 $D = \sum_{j=1}^N S$, 并且 $L \in \mathbb{R}^{N \times N}$ 为拉普拉斯矩阵, $L = D - S$.

1.3 相关熵 由于对非高斯噪声和异常值的鲁棒性, 相关熵已被广泛应用. 作为一个非线性和局部相似性度量函数, 相关熵直接与相似随机变量 A 和 B 在联合空间邻域中的概率有关, 定义如下^[28, 35-36]:

$$V(A, B) = E[\kappa(A, B)] = \int \kappa(a, b) dF_{AB}(a, b) \quad (8)$$

其中, $E[\cdot]$ 和 $\kappa(\cdot, \cdot)$ 分别是期望算子和核函数, $F_{AB}(a, b)$ 表示 (A, B) 的联合分布函数. 核函数 $\kappa(\cdot, \cdot)$ 定义如下:

$$\kappa_\sigma(a, b) = g(a, b) = e^{-\frac{(a-b)^2}{2\sigma^2}} \quad (9)$$

其中, $\sigma > 0$ 是控制高斯核的所有属性的核带宽参数. 由于联合分布函数 $F_{AB}(a, b)$ 的值几乎不可能精确估计, 因此, 通常使用有限数量的数据样本 $\{(a_n, b_n)\}_{n=1}^N$ 将相关熵表示为:

$$\hat{V}_{N,\sigma}(A, B) = \frac{1}{N} \sum_{n=1}^N g(a_n - b_n) \quad (10)$$

其中, $\hat{V}_{N,\sigma}$ 代表相关熵的估计量.

从两个不同的方面进行分析来说明相关熵的鲁棒性. 首先, 可以利用高斯核的泰勒级数展开重写式(8):

$$V(A, B) = \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n \sigma^{2n} n!} \iint \|a - b\|^{2n} dF_{AB}(a, b) \quad (11)$$

从统计学角度看, 相关熵可以捕提高阶矩(即所有的偶数阶矩), 这有助于抑制非高斯噪声和异常值的负面影响. 其次, 每个数据样本对相关熵的贡献在很大程度上取决于沿联合空间中 $a = b$ 的高斯核函数. 此外, 如果样本点相距很近, 则相

关熵表现为 l_2 -范数距离; 如果样本点相距较远, 则相关熵表现为 l_1 -范数距离; 如果样本点相距非常远, 则相关熵表现为 l_0 -范数. 因此, 基于上述分析, 相关熵对非高斯噪声和异常值不敏感.

1.4 半二次优化技术(HQ) 通常, 基于信息论学习 (Information Theoretic Learning, ITL) 方法的目标函数是非线性的且非凸的, 难以直接进行优化. 幸运的是, HQ 常被用来解决基于 ITL 的优化问题^[30-31, 36]. 假设 $\phi(e)$ 是向量 $e \in \mathbb{R}^M$ 的潜在成本函数, 定义如下:

$$\phi(e) = \sum_{i=1}^M \phi(e_i) \quad (12)$$

其中, e_i 是向量 e 的第 i 个元素. 基于最小化的 HQ, 通过引入辅助变量 q_i 来固定 e_i , 方法如下:

$$\phi(e_i) = \min_{q_i} \{Q(e_i, q_i) + \varphi(q_i)\} \quad (13)$$

其中, q_i 仅取决于 $\phi(\cdot)$, $Q(e_i, q_i)$ 为 HQ 函数, $\varphi(\cdot)$ 为 $\phi(\cdot)$ 的双势函数. 本文考虑 $Q(e_i, q_i)$ 的乘法形式, 定义如下:

$$Q(e_i, q_i) = \frac{1}{2} q_i e_i^2 \quad (14)$$

假设 $Q(e, q) = \sum_{i=1}^M Q(e_i, q_i)$, 其中 $q = [q_1, \dots, q_M]$. 结合式(13)和式(14), 可以得到式(13)的向量形式如下:

$$\phi(e) = \min_q \left\{ Q(e, q) + \sum_{i=1}^M \varphi(q_i) \right\} \quad (15)$$

2 CRNMF 算法

2.1 目标函数 如前所述, 为了克服传统 NMF 算法的局限性, 学习有区别的数据表示, 充分利用数据的结构信息, 增强矩阵的稀疏性, 本文提出基于相关熵和流形正则化的 CRNMF 框架. CRNMF 用相关熵代替欧式距离, 通过流形正则化学习局部结构, 并引入 $l_{2,1}$ -范数对非负矩阵进行约束, 有效减轻了非高斯噪声和异常值的影响, 具有良好的有效性和鲁棒性.

CRNMF 目标函数如下:

$$\max_{U, V} \mathcal{O} = \sum_{i=1}^M g \left(\sqrt{\sum_{j=1}^N \left(X_{ij} - (UV^T)_{ij} \right)^2} \right) - \beta \left(\|U\|_{2,1} + \|V\|_{2,1} \right) - \eta \text{Tr}(V^T L V) \quad (16)$$

其中,第一项为用相关熵代替欧式距离,第二项是对聚类中心 U 和聚类指示器 V 进行稀疏约束,第三项是流形正则, β 为稀疏参数, η 为正则化参数.

2.2 算法优化 显然,上述目标函数(16)的优化问题很难直接解决,因为目标函数(16)中的 U 和 V 是非二次和非凸的. 近年来, HQ 已成功用于解决 ITL 中的优化问题^[37], 因此, 本文使用基于共轭函数理论的 HQ 对 CRNMF 进行优化. 首先将非凸目标函数转化为乘法形式的二次函数, 然后利用乘法更新规则对转化后的二次函数进行求解. 基于凸共轭函数的性质, 有以下命题:

命题 1 存在 $g(x)$ 的凸共轭函数 $\varphi(\cdot)$, 使:

$$g(x) = \max_{\varrho} \left(\varrho \frac{\|x\|^2}{\sigma^2} - \varphi(\varrho) \right) \quad (17)$$

其中, ϱ 是一个标量变量, 对于一个固定的 x , ϱ 的最大值为 $\varrho = -g(x)$.

显然, CRNMF 的优化问题(16)可以通过使用命题 1 进行优化:

$$\max_{U, V, \varrho} \hat{\mathcal{O}} = \sum_{i=1}^M \frac{\varrho_i}{\sigma^2} \left(\sum_{j=1}^N \left(X_{ij} - (UV^T)_{ij} \right)^2 - \varphi(\varrho_i) \right) - \beta(\|U\|_{2,1} + \|V\|_{2,1}) - \eta \text{Tr}(V^T L V) \quad (18)$$

其中, $\hat{\mathcal{O}}$ 表示 \mathcal{O} 的扩充目标函数, 并且 $\varrho = [\varrho_1, \dots, \varrho_M]^T$ 为辅助向量. 当固定 U 和 V , 并使 $\hat{\mathcal{O}}$ 最大时, 辅助参数 ϱ_i 的计算式如下:

$$\varrho_i = -g \left(\sqrt{\sum_{j=1}^N \left(X_{ij} - (UV^T)_{ij} \right)^2} \right) = -e^{-\frac{\sum_{j=1}^N \left(X_{ij} - (UV^T)_{ij} \right)^2}{2\sigma^2}} \quad (19)$$

由于应用的多样性, 没有一个统一的策略来选择最佳的核带宽参数 σ . 因此, σ 通常是根据经验确定. 本文中核带宽参数 σ 的计算式如下:

$$\sigma = \sqrt{\frac{1}{2M} \sum_{i=1}^M \sum_{j=1}^N \left(X_{ij} - (UV^T)_{ij} \right)^2} \quad (20)$$

类似地, 固定 ϱ 和 σ 时可以将式(18)改写为:

$$\max_{U, V} \hat{\mathcal{O}} = \sum_{i=1}^M \varrho_i \sum_{j=1}^N \left(X_{ij} - (UV^T)_{ij} \right)^2 - \beta(\|U\|_{2,1} + \|V\|_{2,1}) - \eta \text{Tr}(V^T L V) \quad (21)$$

式(21)相当于以下优化问题:

$$\begin{aligned} \min_{U, V} \hat{\mathcal{O}}_1 = & - \sum_{i=1}^M \varrho_i \sum_{j=1}^N \left(X_{ij} - (UV^T)_{ij} \right)^2 + \\ & \beta(\|U\|_{2,1} + \|V\|_{2,1}) + \eta \text{Tr}(V^T L V) = \\ & \text{Tr}(X^T H X - 2VU^T H X + VU^T H U V^T) + \\ & \beta \text{Tr}(U M_1 U^T + V M_2 V^T) + \eta \text{Tr}(V^T L V) \end{aligned} \quad (22)$$

其中, H, M_1, M_2 为正对角矩阵.

$$(H)_{ii} = -\varrho_i = e^{-\frac{\sum_{j=1}^N \left(X_{ij} - (UV^T)_{ij} \right)^2}{2\sigma^2}} \quad (23)$$

$$(M_1)_{ii} = \frac{1}{\|U_i\|}, i = 1, 2, \dots, M \quad (24)$$

$$(M_2)_{jj} = \frac{1}{\|V_j\|}, j = 1, 2, \dots, N \quad (25)$$

假设 $\Phi = [\Phi_{ik}] \in \mathbb{R}_{\geq 0}^{M \times K}$ 和 $\Psi = [\Psi_{jk}] \in \mathbb{R}_{\geq 0}^{N \times K}$ 分别是 U 和 V 的拉格朗日乘子, 对于约束 $U \geq 0$ 和 $V \geq 0$, 有以下等价目标函数:

$$\begin{aligned} \mathcal{L} = & \text{Tr}(X^T H X - 2VU^T H X + VU^T H U V^T) + \\ & \beta \text{Tr}(U M_1 U^T + V M_2 V^T) + \eta \text{Tr}(V^T L V) + \\ & \text{Tr}(\Phi U^T) + \text{Tr}(\Psi V^T) \end{aligned} \quad (26)$$

更新 U : 固定 V 和 H 更新 U .

因为:

$$\frac{\partial \mathcal{L}}{\partial U} = -2H X V + 2H U V^T V + 2\beta M_1 U + \Phi \quad (27)$$

使用 KKT 条件 $\Phi_{ik} u_{ik} = 0$, 可得以下更新规则:

$$u_{ik} \leftarrow u_{ik} \frac{(H X V)_{ik}}{(H U V^T V + \beta M_1 U)_{ik}} \quad (28)$$

更新 V : 固定 U 和 H 更新 V .

因为:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial V} = & -2X^T H U + 2V U^T H U + \\ & 2\beta M_2 V + 2\eta L V + \Psi \end{aligned} \quad (29)$$

使用 KKT 条件 $\Psi_{jk} v_{jk} = 0$, 可得以下更新规则:

$$v_{jk} \leftarrow v_{jk} \frac{(X^T H U + \eta S V)_{jk}}{(V U^T H U + \beta M_2 V + \eta D V)_{jk}} \quad (30)$$

完整的 CRNMF 算法总结在算法 1 中.

算法 1 CRNMF 算法

输入: 数据矩阵 X , 图最近邻数 p , 参数 β, η , 嵌入维度 K , 最大迭代次数 T .

输出: 基矩阵 U 和潜在数据表示矩阵 V .

1. 随机初始化 U 和 V ;

2. 根据式(6)计算相似度矩阵 S ;
3. 重复;
4. 根据式(24)和式(25)分别计算对角矩阵 M_1, M_2 ;
5. 根据式(20)更新 σ ;
6. 根据式(23)更新 H ;
7. 根据式(28)更新 U ;
8. 根据式(30)更新 V ;
9. 直到收敛.

3 算法分析

3.1 收敛性分析 不失一般性,使用辅助函数法证明提出算法的收敛性,并有以下定理:

定理 1 CRNMF 中的目标函数 \hat{C}_1 在更新规则(28)和(30)下不递增.

在证明定理 1 之前,首先给出辅助函数和引理 1 的定义如下:

定义 1 $G(v, v')$ 是 $F(v)$ 的辅助函数,满足下列条件:

$$G(v, v') \geq F(v), G(v, v) = F(v) \quad (31)$$

引理 1 如果 G 是 F 的辅助函数,那么 F 在从 t 迭代到 $t+1$ 的以下更新步骤下不递增:

$$v^{t+1} = \underset{v}{\operatorname{argmin}} G(v, v') \quad (32)$$

证明

$$F(v^{t+1}) \leq G(v^{t+1}, v') \leq G(v', v') = F(v')$$

显然,如果使用引理 1 和更新规则(28)和(30)证明最小化目标函数 \hat{C}_1 的辅助函数 G ,则可以得出结论,在更新规则(28)和(30)下, \hat{C}_1 不递增. 假设 u_{ik} 和 v_{jk} 是矩阵 U 和 V 中的任意项,并且将 F_{ik} 和 F_{jk} 定义为分别只涉及 u_{ik} 项和 v_{jk} 项的部分,则相对于 U 和 V 的一阶和二阶偏导数为:

$$F'_{ik} = \left(\frac{\partial \hat{C}_1}{\partial U} \right)_{ik} = (-2HXV + 2HUV^TV)_{ik} + 2\beta(M_1U)_{ik} \quad (33)$$

$$F''_{ik} = \left(\frac{\partial^2 \hat{C}_1}{\partial U^2} \right)_{ik} = 2H_{ii}(V^TV)_{kk} + 2\beta(M_1)_{ii} \quad (34)$$

$$F'_{jk} = \left(\frac{\partial \hat{C}_1}{\partial V} \right)_{jk} = (-2X^THU + 2VU^THU)_{jk} + 2\beta(M_2V)_{jk} + 2\eta(LV)_{jk} \quad (35)$$

$$F''_{jk} = \left(\frac{\partial^2 \hat{C}_1}{\partial V^2} \right)_{jk} = 2(U^THU)_{kk} + 2\beta(M_2)_{jj} + 2\eta(L)_{jj} \quad (36)$$

本质上,更新规则是元素式的. 因此,需要证明 F_{ik} 和 F_{jk} 在规则(28)和(30)下都是不递增的.

引理 2 $G(u, u'_{ik})$ 和 $G(v, v'_{jk})$ 是 F_{ik} 和 F_{jk} 的辅助函数,定义如下:

$$G(u, u'_{ik}) = F_{ik}(u'_{ik}) + F'_{ik}(u - u'_{ik}) + \frac{(HUV^TV + \beta M_1U)_{ik}}{u'_{ik}}(u - u'_{ik})^2 \quad (37)$$

$$G(v, v'_{jk}) = F_{jk}(v'_{jk}) + F'_{jk}(v - v'_{jk}) + \frac{(VU^THU + \beta M_2V + \eta DV)_{jk}}{v'_{jk}}(v - v'_{jk})^2 \quad (38)$$

证明 显然, $G(u, u) = F_{ik}(u)$ 和 $G(v, v) = F_{jk}(v)$, 接下来需要证明 $G(u, u'_{ik}) \geq F_{ik}(u)$ 和 $G(v, v'_{jk}) \geq F_{jk}(v)$. 结合式(33)和式(34), 将 $F_{ik}(u)$ 的泰勒级数展开, 得到以下等式:

$$F_{ik}(u) = F_{ik}(u'_{ik}) + F'_{ik}(u'_{ik})(u - u'_{ik}) + \frac{1}{2}F''_{ik}(u'_{ik})(u - u'_{ik})^2 = F_{ik}(u'_{ik}) + F'_{ik}(u'_{ik})(u - u'_{ik}) + (u - u'_{ik})^2 \left(H_{ii}(V^TV)_{kk} + \beta(M_1)_{ii} \right) \quad (39)$$

为了保证 $G(u, u'_{ik}) \geq F_{ik}(u)$, 结合式(37)和式(39), 得到以下不等式:

$$\frac{(HUV^TV + \beta M_1U)_{ik}}{u'_{ik}} \geq H_{ii}(V^TV)_{kk} + \beta(M_1)_{ii} \quad (40)$$

因为:

$$(HUV^TV)_{ik} = \sum_{l=1}^K u'_{il} H_{ii}(V^TV)_{lk} \geq u'_{ik} H_{ii}(V^TV)_{kk} \quad (41)$$

式(40)得证. 所以, $G(u, u'_{ik}) \geq F_{ik}(u)$.

结合式(35)和式(36), 将 $F_{jk}(v)$ 的泰勒级数展开, 得到以下等式:

$$F_{jk}(v) = F_{jk}(v'_{jk}) + F'_{jk}(v'_{jk})(v - v'_{jk}) + \frac{1}{2}F''_{jk}(v'_{jk})(v - v'_{jk})^2 = F_{jk}(v'_{jk}) + F'_{jk}(v'_{jk})(v - v'_{jk}) + (v - v'_{jk})^2 \left((U^THU)_{kk} + \beta(M_2)_{jj} + \eta L_{jj} \right) \quad (42)$$

为了保证 $G(v, v'_{jk}) \geq F_{jk}(v)$, 结合式(38)和

式(42),得到以下不等式:

$$\frac{(VU^THU + \beta M_2V + \eta DV)_{jk}}{v'_{jk}} \geq (U^THU)_{kk} + \beta(M_2)_{jj} + \eta L_{jj} \quad (43)$$

由于 $\beta > 0, \eta > 0$,

$$(VU^THU)_{jk} = \sum_{l=1}^K v'_{jl}(U^THU)_{lk} \geq v'_{jk}(U^THU)_{kk} \quad (44)$$

并且

$$\begin{aligned} (\eta DV)_{jk} &= \eta \sum_{l=1}^N D_{jl} v'_{lk} \geq \eta D_{jj} v'_{jk} \\ \eta(D-S)_{jj} v'_{jk} &= \eta L_{jj} v'_{jk} \end{aligned} \quad (45)$$

式(43)得证. 所以, $G(v, v'_{jk}) \geq F_{jk}(v)$.

综上,引理2得证.

接下来,证明定理1.

证明 对于目标函数 \hat{C}_1 ,将 $G(u, u'_{ik})$ 替换为式(37),将 $G(v, v'_{jk})$ 替换为式(38),导出以下更新规则:

$$u_{ik}^{t+1} = u'_{ik} - u'_{ik} \frac{F'_{ik}(u'_{ik})}{2(HUV^TV + \beta M_1U)_{ik}} = \frac{(HXV)_{ik}}{(HUV^TV + \beta M_1U)_{ik}} \quad (46)$$

$$v_{jk}^{t+1} = v'_{jk} - v'_{jk} \frac{F'_{jk}(v'_{jk})}{2(VU^THU + \beta M_2V + \eta DV)_{jk}} = \frac{(X^THU + \eta SV)_{jk}}{(VU^THU + \beta M_2V + \eta DV)_{jk}} \quad (47)$$

3.2 鲁棒性分析 本文通过HQ对CRNMF算法进行优化,将非凸优化问题转化为迭代加权的NMF问题.由命题1可知,当固定矩阵 U 和 V 时,数据矩阵 X 中每个样本的每一项都被赋予一个权重.权重的计算如下:

$$\varrho_i = -g\left(\sqrt{\sum_{j=1}^N (X_{ij} - (UV^T)_{ij})^2}\right) = -e^{-\frac{\sum_{j=1}^N (X_{ij} - (UV^T)_{ij})^2}{2\sigma^2}} \quad (48)$$

根据式(48)可知,当数据受损严重时将被分配较小的权重,因而对目标函数的贡献较小.通过引入相关熵和 $l_{2,1}$ -范数,CRNMF 能够有效地减轻噪声对聚类结果的影响.因此,CRNMF 算法对噪声和异常值不敏感.

3.3 计算复杂度分析 计算复杂度对于评估算法的效率非常重要.分析每次更新迭代的CRNMF计算成本,并使用最常用的符号 O 来表示所提出算法的复杂度.首先,根据式(6)可知,CRNMF需要 $O(MN^2)$ 构造 p -近邻图来计算相似度矩阵, M 为特征维度, N 为样本数量.此外,根据式(20)、式(23)、式(28)和式(30)更新 σ, H, U 和 V 的计算复杂度为 $O(TMNK)$, K 为嵌入维度.假设提出的算法在 T 次迭代后停止,则CRNMF的整体计算复杂度为 $O(MN^2 + TMNK)$.

4 实验

4.1 实验设置

4.1.1 实验环境 所有算法均用 Matlab R2018a 实现,在一台 2.40 GHz Intel Core CPU 16.0 GB RAM 和 Windows 10 操作系统的计算机上进行.

4.1.2 数据集 实验使用五个图像数据集,分别为 CMU PIE, ORL, UMIST, YaleB 和 COIL20.用于聚类任务的数据集的详细特征如表2所示,图2至图6给出了每个数据集的图像样本.对于每个图像数据集,将数据格式化为具有相同维数和样本数大小的数据矩阵,并将每个样本(即数据矩阵的每一列)归一化为欧几里得长度.

表2 实验中使用的数据集描述

Table 2 Description of datasets used in experiments

数据集	样本数量	维度	类
CMU PIE	2856	1024	68
ORL	400	1024	40
UMIST	575	1024	20
YaleB	2414	1024	38
COIL20	1440	1024	20

CMU PIE:包含68个人的2856幅图像,每个人有13种不同的姿态、4种不同表情和43种不同的光照条件.

ORL:剑桥大学提出的数据集,包含40个人的400幅图像.每个人有10张面部表情图像.

UMIST:包含20个人的575幅图像,每个主题由不同种族、性别或外貌的图像组成,并包含从侧面到正面的一系列姿势.

YaleB:包含38个人的2414幅正面人脸图像

(大约每人64幅图像),所有图像都已损坏。

COIL20:哥伦比亚大学提出的数据集,包含20个对象的1440幅图像,每个对象都由72幅不同角度的图像组成。



图2 CMU PIE数据集样本

Fig. 2 Samples of CMU PIE dataset



图3 ORL数据集样本

Fig. 3 Samples of ORL dataset

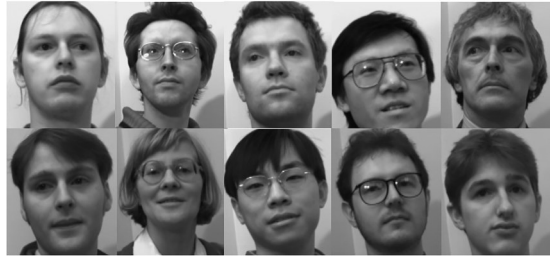


图4 UMIST数据集样本

Fig. 4 Samples of UMIST dataset



图5 YaleB数据集样本

Fig. 5 Samples of YaleB dataset

4.1.3 评价指标 为了衡量聚类性能,采用三个常用的评价指标,即聚类精度(ACC)、归一化互信息(NMI)和纯度(PUR)。



图6 COIL20数据集样本

Fig. 6 Samples of COIL20 dataset

聚类精度(ACC)定义如下:

$$ACC = \frac{\sum_{n=1}^N \delta(r_n, \text{map}(c_n))}{N} \quad (49)$$

其中, N 是数据集的总数, r_n 和 c_n 分别是数据集提供的标签和聚类标签. $\delta(h, z)$ 表示 δ 函数, 当 $h = z$ 时 δ 函数返回 1, 否则返回 0. $\text{map}(\cdot)$ 是通过使用 Kuhn-Munkres 算法^[38] 将每个聚类标签映射到来自数据集的等价标签的映射函数。

归一化互信息(NMI)定义如下:

$$NMI = \frac{\sum_{i=1}^N \sum_{j=1}^N n_{i,j} \lg \frac{n_{i,j}}{n_i n_j}}{\sqrt{\left(\sum_{i=1}^N n_i \lg \frac{n_i}{n} \right) \left(\sum_{j=1}^N \hat{n}_j \lg \frac{\hat{n}_j}{n} \right)}} \quad (50)$$

其中, n_i 和 \hat{n}_j 分别表示第 i 个簇和第 j 个类的大小, $n_{i,j}$ 表示它们的交集中的数据量, N 是簇的数量。

纯度(PUR)定义如下:

$$PUR = \frac{1}{n} \sum_{i=1}^N \max(n'_i) \quad (51)$$

其中, n'_i 是簇中属于第 i 类数据点的数量。

具体地, 聚类精度(ACC)计算聚类中的数据点来自同一类的百分比; 归一化互信息(NMI)衡量不同簇的相似性; 纯度(PUR)衡量聚类中的数据点来自一个类的程度. 所有评价指标范围是 0 到 1, 值越大, 聚类性能越好。

4.1.4 比较算法 为了证明提出的 CRNMF 算法在聚类任务中的有效性和鲁棒性, 将 CRNMF 与七种相关聚类算法进行比较, 它们是 k -means, NMF, GNMF^[19], CGNMF^[25], GDNMF^[24], GLoRSS^[28] 和 l_1 -CNMF^[31]. 在这些比较算法中, k -means 是基线。

(1) k -means 算法是应用最广泛的聚类算法之一, 常被用于聚类算法的基线。

(2) 非负矩阵分解(NMF)将原始数据矩阵分

解为两个非负矩阵,实现了高维数据的低维表示,以更好地应用于聚类任务.

(3)图正则化非负矩阵分解(GNMF)通过结合拉普拉斯正则化来考虑流形数据上的固有几何结构.

(4)基于相关熵诱导度量的图正则化非负矩阵分解算法(CGNMF)将相关熵引入 GNMF,在保持数据几何结构的同时增强算法的抗噪声能力.

(5)基于图的带标签信息的判别非负矩阵分解(GDNMF)是在考虑了数据的局部几何结构的同时,在 NMF 中引入了标签信息.

(6)GLoRSS 是一种基于最大相关熵准则的鲁棒子空间学习算法.

(7) l_1 -CNMF 是基于最大相关熵准则的 l_1 -范数非负矩阵分解算法,该算法用相关熵提高鲁棒性并且对系数矩阵进行 l_1 -范数约束.

4.1.5 其他设置 实验需要预先设置四个主要参数,分别是图最近邻数 p 、稀疏参数 β 、正则化参数 η 和嵌入维度 K .在不提及其他方面的情况下,

将参数 p, β 和 η 分别设置为 $(5, 0.1, 100)$, 嵌入维度 K 设置为簇的数量.热核函数涉及的权衡参数 λ 设置为 1, CRNMF 中相关熵使用的核带宽参数 σ 由式(20)计算. GNMF 算法的正则化参数设置为 100. CGNMF 算法的正则化参数设置为 10^{-6} . GDNMF 算法的正则化参数设置为 1, 标签约束参数在 $\{1, 5, 10, 100\}$ 中选取. GLoRSS 算法的局部结构保持项参数设置为 1, 稀疏参数在 $\{0.001, 0.01, 0.1, 1, 10, 100\}$ 中选取. l_1 -CNMF 算法的稀疏参数设置为 0.1. 与直接应用于原始数据的 k -means 不同,其他算法的聚类过程包括两个主要步骤:(1)获得原始数据矩阵的潜在低维表示矩阵;(2)对学习的表示矩阵运用 k -means 算法.根据经验,最大迭代次数设置为 300 次,并对所有算法进行 20 次具有不同初始值的相关蒙特卡罗方法.对于每个数据集,计算出平均结果作为所有算法的最终聚类结果,并突出显示最佳结果.

4.2 聚类结果 在五个图像数据集上的实验结果如表 3 所示,表中粗体字为结果最优. 可以看

表 3 本文算法和七个对比算法在五个数据集上的聚类结果:精度、归一化互信息和纯度

Table 3 Clustering results of ACC, NMI and PUR by our algorithm and other algorithms on five image datasets

Dataset	ACC							
	k -means	NMF	GNMF	CGNMF	GDNMF	GLoRSS	l_1 -CNMF	CRNMF
CMU PIE	24.23%	51.32%	76.22%	76.36%	76.07%	72.35%	73.89%	78.30%
ORL	49.02%	58.66%	60.15%	60.69%	56.07%	56.31%	61.62%	61.73%
UMIST	40.05%	49.34%	53.70%	54.67%	54.47%	51.31%	54.35%	63.13%
YaleB	07.25%	30.47%	31.33%	34.47%	33.39%	33.07%	33.97%	35.42%
COIL20	59.72%	65.77%	71.84%	75.46%	79.83%	73.12%	76.33%	81.67%
Dataset	NMI							
	k -means	NMF	GNMF	CGNMF	GDNMF	GLoRSS	l_1 -CNMF	CRNMF
CMU PIE	53.65%	73.26%	88.31%	88.57%	88.13%	85.14%	87.25%	92.52%
ORL	69.23%	50.14%	73.38%	74.76%	73.46%	73.76%	73.88%	74.93%
UMIST	58.04%	46.92%	70.84%	74.61%	71.07%	70.91%	74.13%	76.75%
YaleB	08.18%	38.12%	42.70%	42.19%	42.08%	41.29%	42.78%	53.06%
COIL20	70.05%	71.98%	79.53%	85.99%	88.33%	82.30%	87.06%	90.03%
Dataset	PUR							
	k -means	NMF	GNMF	CGNMF	GDNMF	GLoRSS	l_1 -CNMF	CRNMF
CMU PIE	30.62%	59.49%	80.67%	86.25%	87.46%	84.33%	86.70%	86.79%
ORL	59.20%	60.96%	64.22%	66.15%	61.97%	62.23%	65.41%	66.18%
UMIST	45.73%	54.73%	66.83%	67.84%	73.04%	64.19%	67.68%	67.65%
YaleB	12.21%	32.15%	38.91%	39.98%	39.99%	38.34%	39.57%	40.49%
COIL20	68.59%	74.57%	73.97%	84.77%	85.83%	83.31%	85.43%	86.39%

出,提出的CRNMF算法基本在所有图像数据集上都获得了最佳的聚类性能,在大多数情况下,CRNMF在精度(ACC)、归一化互信息(NMI)和纯度(PUR)上优于其他算法.例如,在五个图像数据集上,CRNMF的聚类性能比GNMF有明显提高,可以证明相关熵的有效性.在CMU PIE数据集上,CRNMF的聚类精度(ACC)、归一化互信息(NMI)和纯度(PUR)比CGNMF分别提高1.94%,3.95%和0.54%,证明增加稀疏约束在一定程度上可以提高聚类结果.

在所有数据集上,CRNMF的聚类结果均优于 l_1 -CNMF,证明利用 $l_{2,1}$ -范数对矩阵约束的有效性.与传统的聚类算法相比,CRNMF在大多数情况下优于这些比较算法的主要原因可以概括如下:该算法在目标函数中使用相关熵代替欧式距离,可以减少非高斯噪声污染对数据集的负面影响;CRNMF还充分考虑了数据的局部结构,并通过 $l_{2,1}$ -范数对矩阵进行稀疏约束,在保持局部平滑的同时充分利用了矩阵的稀疏性,因此具有更好的聚类效果.此外, k -means的聚类结果比其他算法差得多,表明挖掘数据潜在的低维表示可以提高聚类任务的性能.

同时,为了验证CRNMF各部分的有效性,分别将 β 和 η 设为0.当 $\beta=0$ 时,CRNMF仅包含相关熵项和流形正则项,此时在CMU PIE数据集上 $ACC=75.56\%$,可以证明稀疏约束的重要性.当 $\eta=0$ 时,CRNMF仅包含相关熵项和稀疏约束项,此时在CMU PIE数据集上 $ACC=19.64\%$,因此,流形正则项在提高聚类性能方面发挥极大的作用.

CRNMF在COIL20数据集上的T-SNE可视化结果如图7和图8所示,由图可见,CRNMF可以将相似的数据聚集到同一类中.

4.3 鲁棒性 为了验证CRNMF算法的鲁棒性,在CMU PIE数据集上进行实验.具体地,将噪声密度为0.01的椒盐噪声加入CMU PIE数据集. CMU PIE数据集加入椒盐噪声后的图像如图9所示. CMU PIE数据集加入椒盐噪声后聚类精度的变化如图10所示,其中, $n\%$ 表示所有图像的噪声百分比, $n=0,10,20,30,40,50$, n 越大,噪声

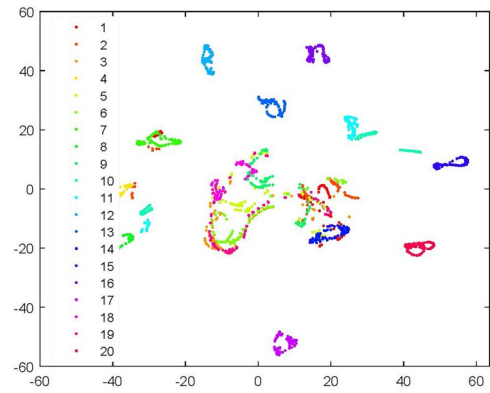


图7 COIL20数据集T-SNE可视化

Fig. 7 T-SNE visualization of COIL20 dataset

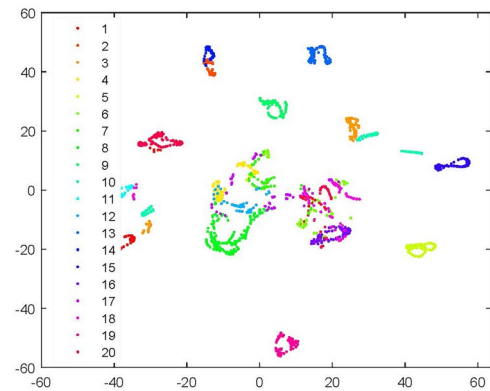


图8 CRNMF在COIL20数据集上的T-SNE可视化

Fig. 8 T-SNE visualization of CRNMF on COIL20 dataset



图9 加入椒盐噪声的CMU PIE数据集

Fig. 9 CMU PIE dataset with salt and pepper noise

水平越高.由图可见,随着噪声水平的增加,提出的CRNMF算法在大多数情况下具有最佳的聚类精度,表明相关熵对非高斯噪声和异常值不敏感,所以处理非高斯噪声和异常值是有效的.与实验中的其他算法相比,CRNMF算法在具有明显异常值的CMU PIE数据集上鲁棒性更强.

4.4 参数选择 在COIL20, CMU PIE和UMIST数据集上研究参数 p , β 和 η 的选择对聚

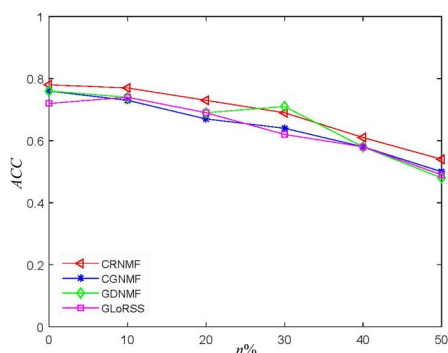


图 10 CMU PIE 数据集上椒盐噪声影响下的聚类精度变化

Fig. 10 Variation of ACC under the influence of salt and pepper noise on CMU PIE dataset

类性能的影响. 参数 p, β 和 η 分别选自 $\{3, 5, 7, 9, 11\}$, $\{0.01, 0.1, 1, 10, 100\}$ 和 $\{10, 50, 100, 500, 1000\}$. 图 11 至图 13 显示了参数 p, β 和 η 对聚类精度的影响. 由图可见, 对于参数 $p, p=5$ 时聚类精度最高, p 太大或太小都会对 CRNMF 产生较大的影响, 因此 $p=5$ 是最好的选择. 在范围 $\{0.01, 0.1, 1, 10, 100\}$ 中 β 比较稳定, 当 $\beta=0.1$ 时在三个数据集上性能最佳, 但 β 不能太大或太小, 否则会出现过拟合的问题. 同时, 当 $\eta=100$ 时, CRNMF 在三个数据集上的表现最好.

4.5 收敛性 为了验证 CRNMF 的收敛性, 在 COIL20 数据集上进行实验. 实验结果如图 14 所示, x 轴表示迭代次数, y 轴表示 CRNMF 的目标函数值 \hat{O}_1 . 由图可见, 对于 COIL20 数据集, 在乘法更新规则 (28) 和 (30) 下, \hat{O}_1 的值单调递减, 并

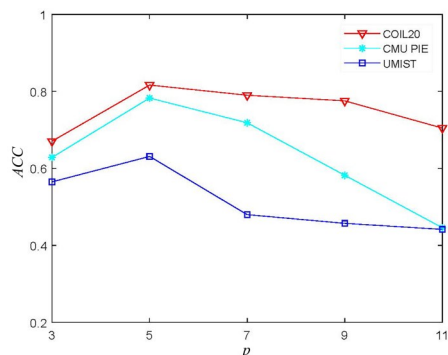


图 11 在 COIL20, CMU PIE 和 UMIST 数据集上 CRNMF 随参数 p 的聚类精度变化

Fig. 11 ACC of CRNMF with the change of parameter p on COIL20, CMU PIE and UMIST datasets

且, CRNMF 算法收敛速度快, 通常在 200 次迭代以内收敛, 证明了半二次优化技术的有效性.

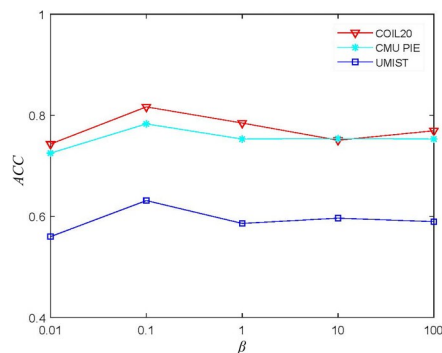


图 12 在 COIL20, CMU PIE 和 UMIST 数据集上 CRNMF 随参数 β 的聚类精度变化

Fig. 12 ACC of CRNMF with the change of parameter β on COIL20, CMU PIE and UMIST datasets

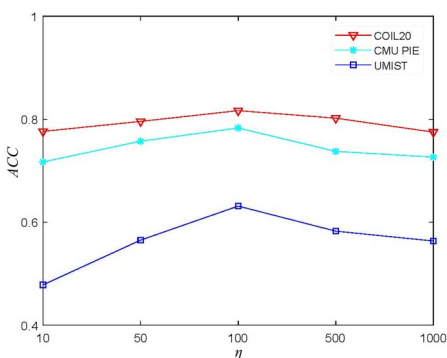


图 13 在 COIL20, CMU PIE 和 UMIST 数据集上 CRNMF 随参数 η 的聚类精度变化

Fig. 13 ACC of CRNMF with the change of parameter η on COIL20, CMU PIE and UMIST datasets

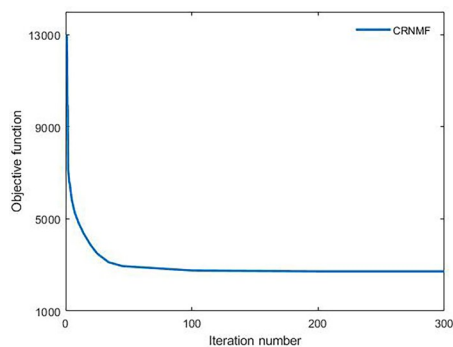


图 14 CRNMF 在 COIL20 数据集上的收敛曲线

Fig. 14 Convergence curves of CSNMF on COIL20 dataset

5 结论和展望

本文提出一种基于相关熵和流形正则化的图像聚类算法(CRNMF),将相关熵、稀疏约束和流形正则集成为一个统一框架.采用相关熵代替传统NMF中的欧氏距离,在充分学习数据的局部结构的同时对非负矩阵进行稀疏约束,有效抑制非高斯噪声和异常值的影响.最后,采用半二次优化技术(HQ)导出CRNMF的乘法更新规则,证明了算法的正确性和收敛性.在五个图像数据集上进行实验,并与经典的聚类算法进行比较,证明CRNMF的聚类精度(*ACC*)、归一化互信息(*NMI*)和纯度(*PUR*)都有所提高.

尽管CRNMF在聚类任务中取得了较好的效果,但仅限于处理非高斯噪声.为了处理更加复杂的噪声,未来的工作将进一步提高CRNMF的有效性和鲁棒性.首先,拓展所提出的算法,增加双图调节或超图约束;其次,充分利用数据中的监督信息,研究算法在半监督学习中的性能;最后,将所提出的算法应用到多视图聚类任务中.

参考文献

- [1] Hartigan J A, Wong M A. Algorithm AS 136: A k -means clustering algorithm. *Applied Statistics*, 1979, 28(1):100—108.
- [2] Von Luxburg U. A tutorial on spectral clustering. *Statistics and Computing*, 2007, 17(4):395—416.
- [3] Lee D D, Seung H S. Algorithms for non-negative matrix factorization//*Proceedings of the 13th International Conference on Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 2000(13):535—541.
- [4] Jain A K. Data clustering: 50 years beyond k -means. *Pattern Recognition Letters*, 2010, 31(8):651—666.
- [5] Kriegel H P, Kröger P, Zimek A. Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering. *ACM Transactions on Knowledge Discovery from Data*, 2009, 3(1):1.
- [6] Chakraborty S, Paul D, Das S, et al. Entropy regularized power k -means clustering. 2020, arXiv, 2001.03452.
- [7] Wang Y X, Zhang Y J. Nonnegative matrix factorization: A comprehensive review. *IEEE Transactions on Knowledge and Data Engineering*, 2013, 25(6):1336—1353.
- [8] Tolić D, Antulov-Fantulin N, Kopriva I. A nonlinear orthogonal non-negative matrix factorization approach to subspace clustering. *Pattern Recognition*, 2018(82):40—55.
- [9] Liu G C, Lin Z C, Yan S C, et al. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(1):171—184.
- [10] Jolliffe I T. Principal component analysis. *Journal of Marketing Research*, 2002, 87(4):513.
- [11] Höcker A, Kartvelishvili V. SVD approach to data unfolding. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1996, 372(3):469—481.
- [12] Lee D D, Seung H S. Learning the parts of objects by non-negative matrix factorization. *Nature*, 1999, 401(6755):788—791.
- [13] Li S Z, Hou X W, Zhang H J, et al. Learning spatially localized, parts-based representation//*Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Kauai, HI, USA: IEEE, 2001:5.
- [14] Ullman S. High-level vision: Object-recognition and visual cognition. *Optical Engineering*, 1997, 36(11):231—256.
- [15] Logothetis N K, Sheinberg D L. Visual object recognition. *Annual Review of Neuroscience*, 1996 (19):577—621.
- [16] Liu X W, Wang L, Zhang J, et al. Global and local structure preservation for feature selection. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, 25(6):1083—1095.
- [17] Chen J, Ma Z M, Liu Y. Local coordinates alignment with global preservation for dimensionality reduction. *IEEE Transactions on Neural Networks and Learning Systems*, 2013, 24(1):106—117.
- [18] Luong K, Nayak R. Learning inter- and intra-manifolds for matrix factorization-based multi-aspect data clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2020, DOI: 10.1109/TKDE.2020.3022072.

- [19] Cai D, He X F, Han J W, et al. Graph regularized nonnegative matrix factorization for data representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(8): 1548–1560.
- [20] He P, Xu X H, Ding J, et al. Low-rank nonnegative matrix factorization on stiefel manifold. *Information Sciences*, 2020(514): 131–148.
- [21] Guan N Y, Tao D C, Luo Z G, et al. Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent. *IEEE Transactions on Image Processing*, 2011, 20(7): 2030–2048.
- [22] Shang F H, Jiao L C, Wang F. Graph dual regularization non-negative matrix factorization for co-clustering. *Pattern Recognition*, 2012, 45(6): 2237–2250.
- [23] Zeng K, Yu J, Li C H, et al. Image clustering by hyper-graph regularized non-negative matrix factorization. *Neurocomputing*, 2014(138): 209–217.
- [24] Li H R, Zhang J S, Shi G, et al. Graph-based discriminative nonnegative matrix factorization with label information. *Neurocomputing*, 2017(266): 91–100.
- [25] Wang Y Y, Wu S Y, Mao B, et al. Correntropy induced metric based graph regularized non-negative matrix factorization. *Neurocomputing*, 2016(204): 172–182.
- [26] Du L, Xuan L, Shen Y D. Robust nonnegative matrix factorization via half-quadratic minimization//2012 IEEE 12th International Conference on Data Mining. Brussels, Belgium: IEEE, 2013: 201–210.
- [27] Li L, Yang J J, Zhao K L, et al. Graph regularized non-negative matrix factorization by maximizing correntropy. *Journal of Computers*, 2014, 9(11): 2570–2579.
- [28] Zhou N, Xu Y Y, Cheng H, et al. Maximum correntropy criterion-based sparse subspace learning for unsupervised feature selection. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019, 29(2): 404–417.
- [29] 谢林江, 尹东. 基于最大相关熵准则的鲁棒度量学习算法. *计算机系统应用*, 2018, 27(10): 146–153.
- (Xie L J, Yin D. Robust metric learning based on maximum correntropy criterion. *Computer Systems and Applications*, 2018, 27(10): 146–153.)
- [30] Yu N, Wu M J, Liu J X, et al. Correntropy-based hypergraph regularized NMF for clustering and feature selection on multi-cancer integrated data. *IEEE Transactions on Cybernetics*, 2020, 51(8): 3952–3963.
- [31] Peng S Y, Ser W, Lin Z P, et al. Robust sparse nonnegative matrix factorization based on maximum correntropy criterion//2018 IEEE International Symposium on Circuits and Systems. Florence, Italy: IEEE, 2018: 1–5.
- [32] Zhang L F, Zhang Q, Du B, et al. Robust manifold matrix factorization for joint clustering and feature extraction//31st AAAI Conference on Artificial Intelligence. San Francisco, CA, USA: Machine Learning Applications, 2017: 1662–1668.
- [33] Chen G F, Xu C, Wang J Y, et al. Robust non-negative matrix factorization for link prediction in complex networks using manifold regularization and sparse learning. *Physica A: Statistical Mechanics and its Applications*, 2020, 539: 122882, DOI: 10.1016/j.physa.2019.122882.
- [34] Belkin M, Niyogi P, Sindhvani V. Manifold regularization: A geometric framework for learning from Labeled and Unlabeled examples. *The Journal of Machine Learning Research*, 2006, 7(1): 2399–2434.
- [35] Peng S Y, Chen B D, Sun L, et al. Constrained maximum correntropy adaptive filtering. *Signal Processing*, 2017, 140: 116–126.
- [36] Peng S Y, Ser W, Chen B D, et al. Robust semi-supervised nonnegative matrix factorization for image clustering. *Pattern Recognition*, 2021(111): 107683, DOI: 10.1016/j.patcog.2020.107683.
- [37] He R, Hu B G, Yuan X T, et al. Robust recognition via information theoretic learning. New York: Springer, 2014.
- [38] Lovász L, Plummer M D. Matching theory. Amsterdam, The Netherlands: North-Holland, 1986: 544.

(责任编辑 杨可盛)