

稀疏诱导流形正则化凸非负矩阵分解算法

邱飞岳^{1,2}, 陈博文², 陈铁明², 章国道²

(1. 浙江工业大学教育科学与技术学院, 浙江 杭州 310023; 2. 浙江工业大学计算机科学与技术学院, 浙江 杭州 310023)

摘 要: 针对非负矩阵分解方法在有噪声的真实数据中获得特征的有效性, 提出了一种稀疏诱导的流形正则化凸非负矩阵分解算法。所提算法在流形正则化的基础上, 向低维子空间的基矩阵添加基于 $L_{2,1}$ 范数的稀疏约束, 构建了乘法更新规则, 分析在该规则下算法的收敛性, 并设计了在低维子空间上不同噪声环境下的聚类实验。 K 均值聚类实验结果表明, 稀疏约束降低了噪声特征在学习中的表达能力, 所提算法在不同程度上优于同类 8 种算法, 对噪声有更强的稳健性。

关键词: 非负矩阵分解; 流形正则化; 稀疏约束; K 均值聚类

中图分类号: TP181

文献标识码: A

doi: 10.11959/j.issn.1000-436x.2020064

Sparsity induced convex nonnegative matrix factorization algorithm with manifold regularization

QIU Feiyue^{1,2}, CHEN Bowen², CHEN Tieming², ZHANG Guodao²

1. College of Education Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

2. College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

Abstract: To address problems that the effectiveness of feature learned from real noisy data by classical nonnegative matrix factorization method, a novel sparsity induced manifold regularized convex nonnegative matrix factorization algorithm (SGCNMF) was proposed. Based on manifold regularization, the $L_{2,1}$ norm was introduced to the basis matrix of low dimensional subspace as sparse constraint. The multiplicative update rules were given and the convergence of the algorithm was analyzed. Clustering experiment was designed to verify the effectiveness of learned features within various of noisy environments. The empirical study based on K -means clustering shows that the sparse constraint reduces the representation of noisy features and the new method is better than the 8 similar algorithms with stronger robustness to a variable extent.

Key words: nonnegative matrix factorization, manifold regularization, sparse constraint, K -means clustering

1 引言

在一些研究和应用领域里, 经常遇到输入的数据维数过于庞大, 导致计算复杂度过高和模型过拟合^[1]。为了保证模型的简洁性和准确性, 寻求一种尽可能简单地描述数据的方法至关重要。矩阵的乘法分解是解决这一类问题的常用方法, 常见的分解

方式有主成分分析 (PCA, principal components analysis)、奇异值分解 (SVD, singular value decomposition)、矢量量化 (VQ, vector quantization) 和非负矩阵分解 (NMF, nonnegative matrix factorization) 等。1999 年, Lee 等^[2]提出了对所有元素都是非负的矩阵的乘法分解方法 NMF, 将数据矩阵分解为基矩阵和编码矩阵的乘积。由于得到的矩

收稿日期: 2019-12-04; 修回日期: 2020-03-04

基金项目: 浙江省自然科学基金资助项目 (No.LQ20F020014); 浙江省重点研发计划基金资助项目 (No.2018C01080); 国家自然科学基金资助项目 (No.61472366, No.61379077)

Foundation Items: The Natural Science Foundation of Zhejiang Province(No.LQ20F020014), The Key Technologies Research and Development Program of Zhejiang Province(No.2018C01080), The National Natural Science Foundation of China(No.61472366, No.61379077)

阵均为非负矩阵, 即基之间只允许做加法运算, 不允许做减法运算, 因此, NMF 保留了数据的局部特征。同时, 新的隐式表征空间的维度远小于原始空间的维度, 这些特性赋予了其可以从输入的数据中学习嵌入其中的低维子空间的能力。NMF 被广泛应用于降维与特征提取领域, 如网络结构检测^[3-5]、高光谱解混^[6-7]、人脸识别^[8]和推荐系统^[9]等。

流形学习认为低维子空间是镶嵌在高维特征空间中的低维流形^[10], 通过恢复流形结构可以学习常被传统方法所忽略的样本空间的结构信息。常见的流形学习思想有局部线性嵌入 (LLE, Locally linear embedding)、拉普拉斯图 (LAPPE, Laplacian eigenmap) 和等距特征映射 (ISOMAP, isometric feature mapping)^[11]等。Cai 等^[12]将基于拉普拉斯图的流形正则化项引入 NMF, 在非负矩阵分解的过程中学习了样本空间的隐式结构信息, 提出了图正则化非负矩阵分解 (GNMF, graph regularized NMF)。Huang 等^[13]提出 RGNMF (robust GNMF), 把原始数据空间表示为特征空间与噪声空间的和, 在添加图结构约束的同时向噪声矩阵添加 L_1 范数, 在原始数据存在噪声和野值点时仍然有良好的稳健性。Li 等^[14]提出图正则化低秩非负矩阵分解 (GNLMF, graph regularized non-negative low-rank matrix factorization), 先对数据进行降噪处理, 再进行基于流形学习的非负矩阵分解。Wu 等^[15]提出 MNMFL_{2,1} (manifold NMF with $L_{2,1}$ norm), 在 GNMF 的基础上为矩阵分解的误差添加 $L_{2,1}$ 稀疏约束, 提高了模型的泛化能力和对噪声的稳健性。GNMF 中流形结构 (关联图) 旨在表达样本之间的近邻关系, Zeng 等^[16]基于超图理论构建超图流形结构, 提出超图正则化非负矩阵分解 (HNMF, hyper-graph regularized NMF)。超图中的近邻关系不再限制顶点数量, 可以将关联图结构理解为超图结构的二阶特例, 超图相比关联图可以保留更多的近邻关系。多数流形学习方法预先定义图结构, 在算法迭代过程中图结构保持不变, 因此失去了对空间结构信息变化的敏感性。为了解决这个问题, 一些算法添加了动态图约束, 以在迭代过程中保证图结构对目标空间的持续影响力。Wang 等^[17]提出 RALSE (robust adaptive low-rank and sparse embedding for feature representation) 框架, 集合了矩阵的稀疏低秩表示与添加了 L_1 范数的图结构, 实现了特征的有效学习与空间结构的实时保留。Yi 等^[18]将正交约束、近邻样本间稀

疏表示关系相结合, 提出了带有局部关系保持约束的动态图结构的非负矩阵分解 (NMFLCAG, NMF with locality constrained adaptive graph)。Lu 等^[19]将局部线性嵌入模型与稀疏的动态图约束结合, 在动态学习结构信息的同时保证了相似度矩阵的稀疏性。Yin 等^[20]通过学习自适应的低秩图亲和矩阵来进行子空间聚类, 其中图正则化项的各项参数不再预先指定, 而是在更新过程中自适应更新。

NMF 和 GNMF 都只适用于非负约束下的矩阵分解, 即输入数据不允许出现负值, 不能满足实际应用场的需求。为了解决这个问题, Ding 等^[21]提出了凸非负矩阵分解 (CNMF, convex NMF), 将 NMF 适用范围从非负矩阵扩大至混合符号矩阵。Wang 等^[22]在矩阵分解的基础上对非负基矩阵引入正交约束, 提出基于矩阵分解的特征提取算法 (MFFS, feature selection via matrix factorization)。一些方法将混合符号矩阵分解与流形正则化相结合, 如 Hu 等^[23]提出带有流形正则化项的凸非负矩阵分解 (GCNMF, CNMF with manifold regularization), 聚类实验表明, 其相较于没有学习数据空间结构信息的方法有很大的性能提升。Cui 等^[24]将子空间学习与凸非负矩阵分解相结合提出 SCCNMF (subspace clustering guided convex NMF), 利用样本间的线性表示关系进行流形结构的动态保持。Li 等^[25]考虑相同类标签的图像拥有类似的图像表示, 在 GCNMF 的基础上添加了对成对约束, 提出了 PGCNMF (pairwise constrained graph regularized CNMF)。

传统的特征选择往往忽略了特征之间的关系, 而基于稀疏约束的方法可以发掘特征间的潜在关联, 同时可以增强模型的泛化能力和稳健性^[26]。因此, 各种稀疏化方法被引入非负矩阵分解方法。Kong 等^[27]提出了 $L_{2,1}$ -NMF, 引入 $L_{2,1}$ 范数对编码矩阵做泛化处理。Hoyer 等^[28]提出带有稀疏约束的非负矩阵分解 (NMFSC, NMF with sparseness constraint), 其对基矩阵、编码矩阵添加稀疏约束, 提高了分解结果的稀疏性。张旭等^[29]采用 $L_{2,1}$ 范数增强了图正则化非负矩阵分解在基因表达谱上的噪声稳健性。本文在 GCNMF 的基础上引入 $L_{2,1}$ 范数, 提出一种新的稀疏诱导流形正则化凸非负矩阵分解 (SGCNMF, sparsity induced GCNMF) 算法。该算法不仅学习了数据空间隐式流形结构信息, 同时使数据矩阵不再受限于非负约束, 将非负矩阵分解扩展到了混合符号矩阵分解。

另外, SGCNMF 对模型进行了稀疏化处理, 提高了其泛化能力和噪声稳健性。在一组图像数据集上的聚类实验结果验证了 SGCNMF 的有效性。

2 相关工作

2.1 凸非负矩阵分解

NMF 将非负数据矩阵分解为 2 个非负矩阵的乘积。给定一个矩阵 $\mathbf{X} = [x_1, x_2, \dots, x_n] \in \mathbb{R}_+^{m \times n}$, 找到 2 个非负矩阵 $\mathbf{U} = [u_{ik}] \in \mathbb{R}_+^{m \times k}$ 和 $\mathbf{V} = [v_{jk}] \in \mathbb{R}_+^{n \times k}$, 使

$$\mathbf{X} \approx \mathbf{UV}^T \quad \text{s.t. } \mathbf{U} \geq \mathbf{0}, \mathbf{V} \geq \mathbf{0} \quad (1)$$

其中, \mathbf{U} 为基矩阵, \mathbf{V} 为编码矩阵。原数据矩阵与分解后矩阵乘积的误差越小, 则对原始数据的拟合越好。为了最小化分解误差, Lee 等^[30]提出了乘法更新规则算法。

混合符号矩阵是指含有正数、零和负数的矩阵, 经典 NMF 只解决含非负元素矩阵的分解问题。Ding 等^[21]提出了 CNMF 并证明其收敛性。与 NMF 将矩阵分解成 2 个非负矩阵的乘积不同, CNMF 是一种矩阵的三分解方法^[31], 其允许数据矩阵 \mathbf{X} 中出现负元素, 分解为式(2)所示形式。

$$\mathbf{X}_{\pm} = \mathbf{X}_{\pm} \mathbf{U}_{\pm} \mathbf{V}_{\pm}^T \quad (2)$$

本文采用欧氏距离描述分解误差, CNMF 的代价函数 O_1 定义为

$$O_1 = \|\mathbf{X}_{\pm} - \mathbf{X}_{\pm} \mathbf{U}_{\pm} \mathbf{V}_{\pm}^T\|^2 \quad (3)$$

与 NMF 的乘法更新规则^[30]类似, CNMF 有乘法更新规则^[21], 如式(4)所示。

$$\begin{aligned} u_{ik} &\leftarrow u_{ik} \sqrt{\frac{\left((\mathbf{X}^T \mathbf{X})^+ \mathbf{V} + (\mathbf{X}^T \mathbf{X})^- \mathbf{UV}^T \mathbf{V} \right)_{ik}}{\left((\mathbf{X}^T \mathbf{X})^- \mathbf{V} + (\mathbf{X}^T \mathbf{X})^+ \mathbf{UV}^T \mathbf{V} \right)_{ik}}} \\ v_{jk} &\leftarrow v_{jk} \sqrt{\frac{\left((\mathbf{X}^T \mathbf{X})^+ \mathbf{U} + \mathbf{VU}^T (\mathbf{X}^T \mathbf{X})^- \mathbf{U} \right)_{jk}}{\left((\mathbf{X}^T \mathbf{X})^- \mathbf{U} + \mathbf{VU}^T (\mathbf{X}^T \mathbf{X})^+ \mathbf{U} \right)_{jk}}} \end{aligned} \quad (4)$$

2.2 图正则化非负矩阵分解 GNMF

局部不变性假设^[32]认为, 在高维空间里局部近邻样本映射到低维空间时对应的样本仍为近邻的。为了保持数据空间的几何结构, 对于 2 个原本近邻的点 x_i 与 x_j , 用新的基表示的两点 v_i 和 v_j 也应该相近。基于此, GNMF 首先构建一个邻接图 G , 并连接 $\mathbf{X} = [x_1, x_2, \dots, x_n]$ 中近邻的点。然后, 构建权重矩阵 \mathbf{W} 用于量化点之间的近邻程度。构建 \mathbf{W} 有 3 种方式: 0-1 加

权、热核加权和点积加权。本文以 0-1 加权为例。

$$W_{i,j} = \begin{cases} 0, & x_i \text{ 与 } x_j \text{ 无连接} \\ 1, & x_i \text{ 与 } x_j \text{ 有连接} \end{cases} \quad (5)$$

为了保持 v_i 与 v_j 两点间的近邻关系, 定义图正则化项如式(6)所示。

$$\begin{aligned} R(\mathbf{V}) &= \frac{1}{2} \sum_{i,j} (\|v_i - v_j\|^2) W_{ij} = \\ \text{Tr}(\mathbf{V}^T \mathbf{D} \mathbf{V}) - \text{Tr}(\mathbf{V}^T \mathbf{W} \mathbf{V}) &= \text{Tr}(\mathbf{V}^T \mathbf{L} \mathbf{V}) \end{aligned} \quad (6)$$

其中, $\text{Tr}(\bullet)$ 表示矩阵的迹; \mathbf{D} 表示一个对称矩阵, 其对角元素 D_{ii} 为矩阵 \mathbf{W} 的第 i 行的元素和 (或者对应第 i 列的元素和); \mathbf{L} 为描述空间结构特征的拉普拉斯矩阵, 且 $\mathbf{L} = \mathbf{D} - \mathbf{W}$ 。将图正则化项作为惩罚项, 构建目标函数 O_2 , 如式(7)所示。

$$\begin{aligned} O_2(\mathbf{U}, \mathbf{V}) &= \|\mathbf{X} - \mathbf{UV}^T\|^2 + \lambda \text{Tr}(\mathbf{V}^T \mathbf{L} \mathbf{V}) \\ \text{s.t. } \mathbf{U}_{ij} &\geq 0, \mathbf{V}_{ij} \geq 0 \end{aligned} \quad (7)$$

λ 是非负的正则化参数, 当 $\lambda=0$ 时 GNMF 退化成 NMF。使式(7)最小化, 乘法更新规则为

$$\begin{aligned} u_{ik} &\leftarrow u_{ik} \frac{(\mathbf{XV})_{ik}}{(\mathbf{UV}^T \mathbf{V})_{ik}} \\ v_{jk} &\leftarrow v_{jk} \frac{(\mathbf{X}^T \mathbf{U} + \lambda \mathbf{WV})_{jk}}{(\mathbf{VU}^T \mathbf{U} + \lambda \mathbf{DV})_{jk}} \end{aligned} \quad (8)$$

Cai 等^[12]证明了在乘法更新规则式(8)下, 目标函数 O_2 可以保证收敛性。

2.3 流形正则化凸非负矩阵分解

CNMF 可以看作 K -means 松弛编码空间的正交约束后的版本^[21], CNMF 在分解过程中丢失了数据的空间结构信息。Hu 等^[23]将 GNMF 与 CNMF 相结合, 提出 GCNMF, 在松弛了 NMF 的非负约束的同时, 通过最小化流形正则化项来保留数据空间的流形结构特征。与 GNMF 类似, GCNMF 构建了邻接图 G 、权重矩阵 \mathbf{W} 和拉普拉斯矩阵 \mathbf{L} , 并在代价函数 O_1 中添加图正则化项 $R(\mathbf{V})$ 。GCNMF 的目标函数 O^* 为

$$\begin{aligned} O^*(\mathbf{U}, \mathbf{V}) &= O_1(\mathbf{U}, \mathbf{V}) + R(\mathbf{V}) = \|\mathbf{X} - \mathbf{XUV}^T\|^2 + \lambda \text{Tr}(\mathbf{V}^T \mathbf{L} \mathbf{V}) \\ \text{s.t. } \mathbf{U}_{ij} &\geq 0, \mathbf{V}_{ij} \geq 0 \end{aligned} \quad (9)$$

2.4 稀疏约束非负矩阵分解

基于稀疏约束的方法可以发掘特征空间的潜在关联。稀疏约束通常以对目标约束项添加 $L_{p,q}$ 范数的形式存在, 是一种常见的提高矩阵分解性能的

方式。在矩阵分解中施加稀疏约束有多种方式，施加不同的约束项也有不同的侧重点^[28]。例如，MNMF $L_{2,1}$ 在全局分解误差上添加 $L_{2,1}$ 范数稀疏约束使分解误差更加平滑^[15]；RGNMF向数据中的噪声添加 L_1 范数，以提高对噪声特征的屏蔽能力^[13]；LMFAGR(low-rank matrix factorization with adaptive graph regularize)对局部线性嵌入的权重向量添加 L_1 范数稀疏约束以提取更加有判别能力的特征^[19]。在现有的多种范数中， $L_{2,1}$ 范数对数据野值点的平衡更加平滑，对异常数据更加稳健，对处理噪声有良好的效果，是特征提取中常用的一种稀疏方式^[33]。对于矩阵 $\mathbf{X}_{m \times n}$ ，其 $L_{2,1}$ 范数定义如式(10)所示。

$$\|\mathbf{X}\|_{2,1} = \sum_{i=1}^m \|\mathbf{X}_{i,:}\|_2 = \sum_{i=1}^m \sqrt{\sum_{j=1}^n X_{i,j}^2} \quad (10)$$

其中， $\|\mathbf{X}_{i,:}\|_2$ 是矩阵第 i 行的 L_2 范数， $L_{2,1}$ 范数相当于矩阵 \mathbf{X} 每一行的 L_2 范数的和。将 $\|\mathbf{X}\|_{2,1}$ 作为正则化项加入目标函数，在优化目标函数时，使用最小化 $L_{2,1}$ 范数能保证矩阵行列均稀疏，即得到矩阵整体的稀疏性，使模型的局部特征特异性降低，既可以降低噪声对算法的影响，又可以提高模型的泛化能力。

3 SGCNMF 算法描述与求解

现实中的原始数据往往夹杂着噪声，如图像中不同强度的光照、语音数据中的噪音等，这些噪声对特征学习造成了一定的干扰。在非负矩阵分解中，基矩阵是新的特征子空间的基，保留了原始特征空间中的局部特征，数据中的噪声特征同时被部分学习并保留到基矩阵中成为“噪声特征”。考虑到 $L_{2,1}$ 范数对野值点具有抑制和平衡作用，还可以提高模型的泛化能力，本文在GCNMF的基础上引入 $L_{2,1}$ 范数，通过对基矩阵施加稀疏约束的方式，增强模型的泛化能力和抗噪声能力。

3.1 算法描述

对于非负矩阵分解，基矩阵越稀疏，局部特征贡献度越低。式(9)中 \mathbf{U} 为保留了噪声特征的基矩阵，采用基于 $L_{2,1}$ 范数的稀疏约束，提出目标函数 O ，如式(11)所示。

$$\begin{aligned} O(\mathbf{U}, \mathbf{V}) &= \|\mathbf{X} - \mathbf{XUV}^T\|^2 + \alpha \text{Tr}(\mathbf{V}^T \mathbf{LV}) + \\ &\beta \|\mathbf{U}\|_{2,1} = \text{Tr}((\mathbf{X} - \mathbf{XUV}^T)(\mathbf{X} - \mathbf{XUV}^T)^T) + \\ &\alpha \text{Tr}(\mathbf{V}^T \mathbf{LV}) + \beta \text{Tr}(\mathbf{EUU}^T) \\ \text{s.t. } &U_{ij} \geq 0, V_{ij} \geq 0 \end{aligned} \quad (11)$$

其中， α 、 β 为正则化参数， α 控制流形正则化项在矩阵分解过程中的影响力， β 控制基矩阵的稀疏程度。 \mathbf{E} 为对称矩阵，且其对角元素为

$$E_{ii} = \frac{1}{\sqrt{\sum_{j=1}^n U_{i,j}^2}} \quad (12)$$

下面将讨论目标函数 O 的更新规则和在此规则下目标函数收敛的理论分析。

3.2 算法求解

令 θ_{ik} 、 ϕ_{jk} 分别为约束 $u_{ik} \geq 0$ 、 $v_{jk} \geq 0$ 下的拉格朗日乘子，构建矩阵 $\Theta = [\theta_{ik}]$ 、 $\Phi = [\phi_{jk}]$ 。对式(11)构造拉格朗日函数 G 。

$$\begin{aligned} G(\mathbf{U}, \mathbf{V}) &= \text{Tr}(\mathbf{X}\mathbf{X}^T) - 2\text{Tr}(\mathbf{XUV}^T\mathbf{X}^T) + \\ &\text{Tr}(\mathbf{XUV}^T\mathbf{VU}^T\mathbf{X}^T) + \alpha \text{Tr}(\mathbf{V}^T \mathbf{LV}) + \\ &\beta \text{Tr}(\mathbf{EUU}^T) + \text{Tr}(\Theta\mathbf{U}^T) + \text{Tr}(\Phi\mathbf{V}^T) \end{aligned} \quad (13)$$

G 对 \mathbf{U} 、 \mathbf{V} 分别求偏导并令其偏导数为0，有

$$\frac{\partial G(\mathbf{U}, \mathbf{V})}{\partial \mathbf{U}} = -2\mathbf{X}^T \mathbf{XV} + 2\mathbf{X}^T \mathbf{XUV}^T \mathbf{V} + 2\beta \mathbf{EU} + \Theta = 0 \quad (14)$$

$$\frac{\partial G(\mathbf{U}, \mathbf{V})}{\partial \mathbf{V}} = -2\mathbf{X}^T \mathbf{XU} + 2\mathbf{VU}^T \mathbf{X}^T \mathbf{XU} + 2\alpha \mathbf{LV} + \Phi = 0 \quad (15)$$

对式(14)和式(15)应用KKT(Karush-Kuhn-Tucker)条件，即

$$\theta_{ik} u_{ik} = 0, \phi_{jk} v_{jk} = 0 \quad (16)$$

将式(16)代入式(14)和式(15)可得

$$-(\mathbf{X}^T \mathbf{XV})_{ik} u_{ik} + (\mathbf{X}^T \mathbf{XUV}^T \mathbf{V})_{ik} u_{ik} + \beta (\mathbf{EU})_{ik} u_{ik} = 0 \quad (17)$$

$$-(\mathbf{X}^T \mathbf{XU})_{jk} v_{jk} + (\mathbf{VU}^T \mathbf{X}^T \mathbf{XU})_{jk} v_{jk} + \alpha (\mathbf{LV})_{jk} v_{jk} = 0 \quad (18)$$

定义2个非负矩阵 $(\mathbf{X}\mathbf{X}^T)^+$ 和 $(\mathbf{X}\mathbf{X}^T)^-$ ，如式(19)所示。

$$\begin{aligned} (\mathbf{X}^T \mathbf{X})^+ &= \frac{|\mathbf{X}^T \mathbf{X}| + \mathbf{X}^T \mathbf{X}}{2} \\ (\mathbf{X}^T \mathbf{X})^- &= \frac{|\mathbf{X}^T \mathbf{X}| - \mathbf{X}^T \mathbf{X}}{2} \end{aligned} \quad (19)$$

则存在式(20)所示的关系。

$$\mathbf{X}^T \mathbf{X} = (\mathbf{X}^T \mathbf{X})^+ - (\mathbf{X}^T \mathbf{X})^- \quad (20)$$

将式(20)代入式(17)和式(18)，可得

$$\begin{aligned} &((\mathbf{X}^T \mathbf{X})^- \mathbf{V})_{ik} u_{ik} + ((\mathbf{X}^T \mathbf{X})^+ \mathbf{UV}^T \mathbf{V})_{ik} u_{ik} + \beta (\mathbf{EU})_{ik} u_{ik} = \\ &((\mathbf{X}^T \mathbf{X})^+ \mathbf{V})_{ik} u_{ik} + ((\mathbf{X}^T \mathbf{X})^- \mathbf{UV}^T \mathbf{V})_{ik} u_{ik} \end{aligned} \quad (21)$$

$$\begin{aligned} & ((X^T X)^- U)_{jk} v_{jk} + (V U^T (X^T X)^+ U)_{jk} v_{jk} + \\ & \alpha (L V)_{jk} v_{jk} = ((X^T X)^+ U)_{jk} v_{jk} + \\ & (V U^T (X^T X)^- U)_{jk} v_{jk} \end{aligned} \quad (22)$$

解得乘法更新规则, 如式(23)所示。

$$\begin{aligned} u_{ik} & \leftarrow u_{ik} \frac{((X^T X)^+ V + (X^T X)^- U V^T V)_{ik}}{((X^T X)^- V + (X^T X)^+ U V^T V + \beta E U)_{ik}} \\ v_{jk} & \leftarrow v_{jk} \frac{((X^T X)^+ U + V U^T (X^T X)^- U + \alpha W V)_{jk}}{((X^T X)^- U + V U^T (X^T X)^+ U + \alpha D V)_{jk}} \end{aligned} \quad (23)$$

定理 1 式(11)所示的目标函数在式(23)所示的更新规则下是单调非增的。

证明 见附录 A。

在求出 U 、 V 的更新规则后, SGCNMF 如算法 1 所述。

算法 1 稀疏诱导流形正则化凸非负矩阵分解算法

输入 数据矩阵 $X \in \mathbb{R}^{n \times t}$, 聚类中心数 k

输出 低维矩阵 $V \in \mathbb{R}^{t \times k}$

1) 对数据矩阵进行预处理, 如添加噪声 (如算法 2 所示);

2) 随机数初始化基矩阵 $U \in \mathbb{R}_+^{t \times k}$ 和编码矩阵 $V \in \mathbb{R}_+^{t \times k}$;

3) do

4)
$$U \leftarrow U \otimes \frac{(X^T X)^+ V + (X^T X)^- U V^T V}{(X^T X)^- V + (X^T X)^+ U V^T V + \beta E U}$$

5)
$$V \leftarrow V \otimes \frac{(X^T X)^+ U + V U^T (X^T X)^- U + \alpha W V}{(X^T X)^- U + V U^T (X^T X)^+ U + \alpha D V}$$

其中, \otimes 表示矩阵的哈达玛积;

6) while 目标函数式(11)收敛。

4 聚类实验及结果分析

非负矩阵分解在图像聚类上具有广泛应用, 因此本节设计了图像聚类实验来测试所提算法

SGCNMF 的有效性。本文对所提 SGCNMF 算法与非负矩阵分解算法, 即 GNLMF^[14]、HNMF^[16]、NMFLCAG^[18]、CNMF^[21]、MFFS^[22]、GCNMF^[23]、SCCNMF^[24] 进行对比实验。当 SGCNMF 的流形正则化项参数设为 0 时, SGCNMF 退化为带有稀疏约束的凸非负矩阵分解算法 (SCNMF, CNMF with sparse constraint)。为了验证流形学习和稀疏约束的效力, 本文设置了一组 SCNMF 对照实验, 分别验证流形正则化项和稀疏约束对矩阵分解性能的影响。

4.1 数据准备

本文选取了图像聚类常用的 3 个公开数据集, 即 Grimace 数据集、Faces95 数据集和 USPS 数据集, 并且设计了随机遮挡算法与 Grimace 数据集和 Faces95 数据集合成相应的有噪声数据集。

Grimace 数据集和 Faces95 数据集是由英国埃塞克斯大学创建的面部数据库。Faces95 数据集包括 72 个人、每人 20 张, 共 1 440 幅面部图像; Grimace 数据集包括 18 个人、每人 20 张, 共 360 幅面部图像。2 个数据库的每幅图像均为 64 像素×64 像素, 表示为一个 4 096 维的列向量。数据集的所有元素为非负值。

USPS (United States postal service) 数据集是美国邮政扫描快件上的手写体数据集, 其中包括手写体字母和数字数据集。其手写数字数据集包括 0~9 的手写体, 共 7 291 张训练图像和 2 007 张测试图像, 每张图像为 16 像素×16 像素的 256 灰度图^[34]。本文采用其训练图像数据集, 每类数字选取前 200 张图像组成聚类数据集进行聚类实验。该数据集为混合符号矩阵, 同时存在正数、负数和零^[23], 因此不对其进行 HNMF、GNLMF 和 NMFLCAG 聚类实验。

4.2 噪声处理

为了测试噪声的类型与强度对算法模型的影响, 本文设置了 3 种噪声类型, 即黑色椒盐噪声、白色野值椒盐噪声和黑色块状噪声的对比实验。不同噪声类型如图 1 所示。针对噪声强度实验, 对于椒盐噪声, 设



(a) 黑色椒盐噪声

(b) 白色野值椒盐噪声

(c) 黑色块状噪声

图 1 不同噪声类型示例

置噪声强度数值组为[0.001,0.01,0.1,0.2,0.3], 每个噪声强度数值代表噪声点占总元素数的比例; 对于块状噪声, 设置噪声宽度数值组为[2,4,6,8,10], 每个噪声强度数值代表噪声块边长。具体噪声算法如算法 2 所述。

算法 2 随机噪声算法

输入 数据矩阵 $V \in \mathbb{R}^{m \times n}$, 噪声率 ρ , 噪声类型 T

输出 有噪声图像矩阵 $V' \in \mathbb{R}^{m \times n}$

- 1) 归一化处理数据矩阵 V ;
- 2) 如果 T 为黑色椒盐噪声;
- 3) 对数据矩阵 V 中每个数据点生成 0~1 的随机数 r ;
- 4) 若 $r < \rho$, 则置该点值为 0 ;
- 5) 重复步骤 3)和步骤 4), 直到遍历完 V 中所有数据;
- 6) 如果 T 为白色野值噪声;
- 7) 对每个数据点生成 0~1 的随机数 r ;
- 8) 若 $r < \rho$, 则置该点值为 $255r$;
- 9) 重复步骤 7)和步骤 8)直到遍历完 V 中所有数据;
- 10) 如果 T 为黑色块状噪声;
- 11) 对数据矩阵 V 中每一个样本还原成方形原始图像 I , 其边长为 s ;
- 12) 随机生成一个小于 $\text{index} = s - \rho$ 的数值;
- 13) 将图像矩阵 I 中(index, index)至(index + ρ , index + ρ)的对角矩形区域所有元素置为 0;
- 14) 重复步骤 11)~步骤 13), 直到遍历完 V 中所有数据。

4.3 性能评价指标

本实验采用 2 个常用聚类评价指标, 即准确度 (AC, accuracy) 和归一化互信息 (NMI, normalized mutual information), 来验证 SGCNMF 的准确性和稳健性。AC 和 NMI 的定义如下。

AC 用于衡量样本获得正确的类标签的比率。对给定 n 个样本的数据集, 令其通过聚类方法获得的类标签集为 $L = \{l_1, l_2, \dots, l_n\}$, 真实的类标签集为 $R = \{r_1, r_2, \dots, r_n\}$ 。则该次聚类结果的 AC 定义为

$$AC = \frac{\sum_{i=1}^n \Gamma(g_i, \text{map}(l_i))}{n} \quad (24)$$

其中, $\text{map}(\bullet)$ 表示基于 Kuhn-Munkres 方法的一个最佳映射函数, 它将聚类方法获得的类标签与真实的类标签做最佳匹配^[35]; $\Gamma(\bullet)$ 是一个二值函数, 当聚类标签映射后与真实标签相同时则为 1, 否则

为 0。AC 值越高则说明聚类准确度越高。

互信息 (MI, mutual information) 常用来度量 2 个聚类集合间的相似度。给定数据集的真实类别信息集合 C 和算法聚类出来的类别信息集合 C' , 则 $\text{MI}(C, C')$ 定义为

$$\text{MI}(C, C') = \sum_{c_i \in C, c'_j \in C'} p(c_i, c'_j) \log \frac{p(c_i, c'_j)}{p(c_i)p(c'_j)} \quad (25)$$

其中, $p(c_i)$ 和 $p(c'_j)$ 代表从数据集中任意抽取一个样本分别属于类别 c_i 和 c'_j 的概率, $p(c_i, c'_j)$ 代表抽取的样本既属于 C 又属于 C' 的联合概率。将 MI 归一化后, 定义 NMI 为

$$\overline{\text{MI}}(C, C') = \frac{\text{MI}(C, C')}{\max(H(C), H(C'))} \quad (26)$$

其中, $H(C)$ 和 $H(C')$ 分别表示 C 和 C' 的熵。NMI 值越接近于 1, 则 2 个聚类结果相似度越高; 反之, 则相似度越低。

4.4 参数选择

SGCNMF 的 2 个核心参数为流形正则化项参数 α 和稀疏系数 β 。本节在 3 个数据集上进行了实验以确定 α 和 β 的取值范围。实验发现, α 取值在 $10 \sim 10^2$ 时效果良好且结果较稳定。若稀疏系数 β 太大, 分解出来的基图像过于稀疏, 图像的特征不能得到很好的保留; 若稀疏系数 β 太小, 则抗噪能力不足 ($\beta=0$ 时 SGCNMF 退化成 GCNMF)。 β 取值在 $10 \sim 10^3$ 时效果良好且结果较稳定。实验选择流形正则化参数 α 为 10, 稀疏系数 β 为 100。

经实验测试发现, 各个算法在迭代 50 次以内已基本收敛, 因此将聚类实验中每次迭代的次数设为 100 次, 收敛情况详见 4.6 节收敛性分析。

4.5 聚类实验

对每个数据集都均匀地选取了多个聚类中心进行聚类实验。为了消除实验结果的随机性, 设计实验过程如算法 3 所述。

算法 3 聚类实验算法

输入 数据矩阵 V , 待测试聚类中心数 K

输出 各算法在数据集上聚类结果的平均值和标准差

1) 从数据集中随机选择 K 类数据, 并选取 K 值作为编码矩阵的聚类数 (进行噪声类型、强度对算法性能的影响实验时, K 值固定为数据集最大样本类别数);

2) 对于给定的 K 值, 运行相应的算法获取低

维的特征空间;

3) 在新的低维空间上执行 20 次随机初始化 K 个聚类中心的 K -means 聚类。对聚类结果进行 AC 和 NMI 评估, 每次运行的最佳值作为该次运行的结果;

4) 计算 20 次聚类结果的平均值和标准差;

5) 重复 20 次步骤 1)~步骤 4)并记录平均值和标准差。

对 Grimace 数据集和 Faces95 数据集添加噪声强度为 0.2、噪声值为 255 的白色野值椒盐噪声, 各算法在数据集上的结果如表 1~表 4 所示。对 USPS

数据集进行噪声处理, 各算法的实验结果如表 5~表 6 所示。表 1~表 6 中数值都为百分数。

为了验证不同噪声类型和强度对模型性能的影响, 按照算法 2 所述流程对 Grimace 数据集和 Faces95 数据集添加 3 种不同强度的噪声并进行实验, 实验结果如图 2 和图 3 所示。

从实验结果可以得到以下结论。

1) 基于流形学习的算法在实验数据集上聚类准确度和归一化互信息均明显优于非流形学习算法, 证明了特征空间的隐式结构信息在聚类应用中的重要性。

表 1 聚类算法在噪声 Grimace 数据集上的准确度对比

K	CNMF	SGCNMF($\alpha=0$)	MFFS	HNMF	GNLMF	NMFLCAG	SCCNMF	GCNMF	SGCNMF
4	48.12±7.53	50.75±8.13	48.22±7.53	89.88±12.56	85.38±12.78	69.56±10.81	75.63±9.72	86.37±10.47	89.12±8.31
6	43.71±8.07	44.25±6.24	43.94±4.81	80.58±11.99	80.75±11.68	55.58±8.71	68.58±6.95	79.79±11.89	83.15±10.51
8	40.94±7.84	39.06±6.46	41.04±5.78	74.44±9.15	74.13±11.87	51.73±2.10	63.25±6.40	73.44±13.91	76.35±13.59
10	38.38±8.29	36.90±2.26	38.80±6.35	69.9±7.89	69.5±11.17	44.50±7.31	61.00±7.78	67.51±16.00	70.07±16.23
12	36.35±8.53	37.21±3.28	37.09±6.63	59.33±5.87	63.25±6.68	41.29±6.75	58.63±5.47	63.12±16.86	65.60±17.14
14	34.52±8.85	35.43±2.29	35.71±6.81	60.86±5.84	62.86±6.21	35.20±4.16	55.71±3.03	59.46±17.49	61.62±18.06
16	32.90±9.12	32.87±2.60	34.57±6.90	54.47±3.19	56.97±2.95	33.16±3.77	50.88±2.64	56.24±18.02	58.18±18.73
平均值	39.27±8.32	39.50±4.47	39.91±6.40	69.92±8.07	70.41±9.05	39.90±6.23	61.95±6.00	69.42±14.95	72.01±14.65

表 2 聚类算法在噪声 Grimace 数据集上的归一化互信息对比

K	CNMF	SGCNMF($\alpha=0$)	MFFS	HNMF	GNLMF	NMFLCAG	SCCNMF	GCNMF	SGCNMF
4	20.71±10.08	26.25±9.70	27.79±1.85	83.67±11.07	77.96±10.50	54.24±7.96	64.43±14.98	77.37±13.07	81.01±11.99
6	21.72±8.91	21.61±9.47	28.41±1.93	73.91±10.49	78.70±7.41	52.94±6.15	60.24±8.76	72.32±12.76	76.33±11.85
8	22.80±7.71	21.49±7.99	29.35±2.19	70.57±7.25	68.19±10.78	44.10±8.92	55.97±7.79	68.07±12.69	71.41±12.68
10	23.22±6.94	22.04±5.28	30.14±2.41	69.89±5.17	71.81±7.92	38.74±5.49	53.29±8.23	64.33±12.93	66.97±13.57
12	23.58±6.42	29.06±5.41	30.96±2.75	59.61±4.53	68.93±4.04	34.90±5.73	50.03±5.30	61.94±12.57	64.21±13.43
14	23.82±5.97	26.34±2.69	31.72±3.05	58.82±3.89	59.58±3.94	29.81±2.53	48.51±4.50	59.99±12.32	61.86±13.37
16	23.91±5.61	24.94±3.35	32.43±3.33	58.95±2.42	57.16±2.71	27.41±2.10	40.80±3.77	58.20±12.24	59.82±13.36
平均值	22.82±7.38	24.53±6.27	30.11±2.50	67.92±6.40	68.90±6.76	40.31±5.55	53.32±7.62	66.03±12.65	68.80±12.89

表 3 聚类算法在噪声 Faces95 数据集上的准确度对比

K	CNMF	SGCNMF($\alpha=0$)	MFFS	HNMF	GNLMF	NMFLCAG	SCCNMF	GCNMF	SGCNMF
4	41.37±3.78	50.50±8.13	47.56±1.96	64.75±12.06	65.88±14.74	56.64±7.81	56.13±4.42	61.00±8.67	63.13±10.97
6	38.79±4.63	44.75±9.93	43.53±4.45	54.33±9.65	53.67±9.88	49.58±3.26	50.83±4.71	55.85±10.32	57.75±11.38
8	35.94±5.81	37.94±6.66	40.92±5.24	51.31±9.66	50.44±7.38	38.69±10.12	52.69±11.82	51.78±10.88	53.22±11.94
10	33.82±6.35	35.15±4.25	38.78±5.88	47.40±9.93	49.50±7.71	36.75±8.24	47.25±3.18	48.21±11.44	49.23±12.58
12	32.02±6.79	32.17±3.26	37.08±6.28	44.40±5.24	43.96±7.13	35.24±4.61	39.00±8.25	45.05±12.14	45.98±13.09
14	30.36±7.26	29.54±3.62	35.74±6.48	39.75±4.77	39.50±7.95	28.21±10.57	37.79±3.54	42.37±12.62	43.17±13.53
16	29.09±7.44	28.22±2.79	34.58±6.64	42.91±4.88	42.88±7.02	29.46±3.53	35.59±2.66	40.12±12.96	40.87±13.77
平均值	34.48±6.01	36.90±5.52	39.74±5.28	49.26±8.03	49.40±8.83	39.22±6.16	45.61±5.51	49.20±11.29	50.48±12.47

表 4 聚类算法在噪声 Faces95 数据集上的归一化互信息对比

K	CNMF	SGCNMF($\alpha=0$)	MFFS	HNMF	GNLMF	NMFLCAG	SCCNMF	GCNMF	SGCNMF
4	12.42±4.56	24.07±11.22	16.73±1.77	54.39±9.83	50.63±17.79	38.91±10.24	37.76±7.58	42.57±12.13	45.98±14.18
6	15.20±6.56	15.03±6.47	18.06±2.18	42.36±11.51	41.81±11.00	32.54±7.96	34.62±4.31	42.02±11.33	44.54±12.57
8	16.10±6.21	16.82±5.84	19.21±2.55	40.02±10.05	38.50±10.01	26.87±5.39	42.32±12.19	40.63±11.08	42.48±11.88
10	17.14±6.13	19.3±5.41	20.16±2.83	38.92±9.47	40.90±7.89	21.61±8.52	39.14±1.47	39.49±10.14	40.91±10.85
12	17.76±5.80	18.92±2.65	20.91±2.97	37.94±4.78	36.57±8.75	24.31±1.79	36.55±13.07	38.33±9.65	39.52±10.32
14	18.22±5.50	20.39±3.73	21.70±3.25	34.10±5.13	34.17±9.09	20.78±6.97	34.89±3.43	37.20±9.22	38.27±9.90
16	18.75±5.32	20.84±1.77	22.41±3.48	39.82±4.08	39.50±8.26	22.74±4.10	38.71±1.06	36.57±8.80	37.52±9.46
平均值	16.51±5.73	19.34±5.30	19.88±3.13	41.08±7.84	40.30±10.40	26.82±6.42	37.71±6.16	39.54±10.34	41.32±11.31

表 5 聚类算法在 USPS 数据集上的准确度对比

K	CNMF	SGCNMF	MFFS	SCCNMF	GCNMF	SGCNMF
2	92.48±6.24	91.25±11.57	94.59±4.81	92.62±10.98	98.83±1.50	98.83±1.50
3	86.67±7.60	87.91±7.44	84.79±10.35	91.63±8.29	96.08±4.11	96.27±3.96
4	77.94±9.25	73.00±12.84	73.03±12.41	84.92±10.91	89.51±8.35	89.56±8.20
5	72.72±9.40	66.77±8.05	67.30±9.13	82.26±7.29	89.62±10.23	90.57±10.66
6	59.09±7.64	65.82±6.66	65.13±8.21	77.01±8.36	85.80±7.75	88.50±7.33
7	58.20±6.77	59.42±6.37	59.91±6.92	70.47±6.89	82.04±5.52	84.36±6.15
8	54.38±3.60	59.62±7.81	58.37±4.84	72.51±11.43	77.97±9.43	80.33±8.97
9	49.58±3.37	53.27±6.77	56.72±5.11	66.60±5.11	74.67±7.58	76.79±7.26
10	47.5	54.26	51.76	63.34	73.9	77.35
平均值	66.51±6.73	67.92±8.44	61.75±7.77	77.93±8.66	85.38±6.81	86.95±6.75

表 6 聚类算法在 USPS 数据集上的归一化互信息对比

K	CNMF	SGCNMF	MFFS	SCCNMF	GCNMF	SGCNMF
2	66.82±19.03	66.96±28.02	74.13±17.45	81.70±11.92	92.11±8.68	92.11±8.68
3	65.30±11.71	66.03±14.29	62.56±15.14	81.34±9.54	87.08±10.07	87.44±9.74
4	61.93±9.68	55.60±14.43	55.42±12.51	72.88±7.31	80.40±8.76	80.40±8.67
5	55.96±9.30	49.93±6.87	52.14±8.42	71.70±9.16	82.72±8.89	84.35±9.57
6	47.65±7.70	53.05±6.25	53.21±7.22	69.25±7.18	79.55±5.53	82.17±4.86
7	48.36±5.40	49.00±8.12	51.12±5.38	59.50±6.56	75.01±5.39	78.04±5.31
8	47.40±4.85	50.60±6.16	50.24±3.78	68.13±10.70	75.38±6.22	76.50±6.02
9	43.93±3.80	45.01±5.14	49.49±3.9	65.00±3.42	73.05±5.65	75.88±3.67
10	42.61	47.81	47.02	64.22	76.59	77.93
平均值	53.33±8.93	53.77±11.16	51.23±6.87	70.41±8.22	80.21±7.40	81.65±7.07

2) 在不同的噪声环境下, SGCNMF 的效果均优于其他同类凸非负矩阵分解方法, 证明了本文提的增加基矩阵稀疏约束的有效性。稀疏约束平滑了数据中野值点对分解结果的影响, 同时使基矩阵所表达的局部特征更加稀疏, 提高了模型的泛化能力。但是, 在黑色椒盐噪声和黑色块状噪声中其效果往往不如 GNLMF, 在白色野值噪声中 SGCNMF 的效果优于其他算法。证明了 GNLMF 的去噪算法的有效性,

以及 $L_{2,1}$ 范数对数据中的野值点的平滑作用。

3) 在将 SGCNMF 方法的图正则化参数设置为 0 时, SGCNMF 退化为 SCNMF。与 CNMF 相比, SCNMF 的性能虽有一定的提升, 但是不够稳定。尽管基矩阵的稀疏约束给 GCNMF 带来了稳健的性能提升, 聚类结果显示独立的稀疏约束还不足以改进算法的性能, 需要与流形正则化项等其他约束项结合才能达到更好的效果。

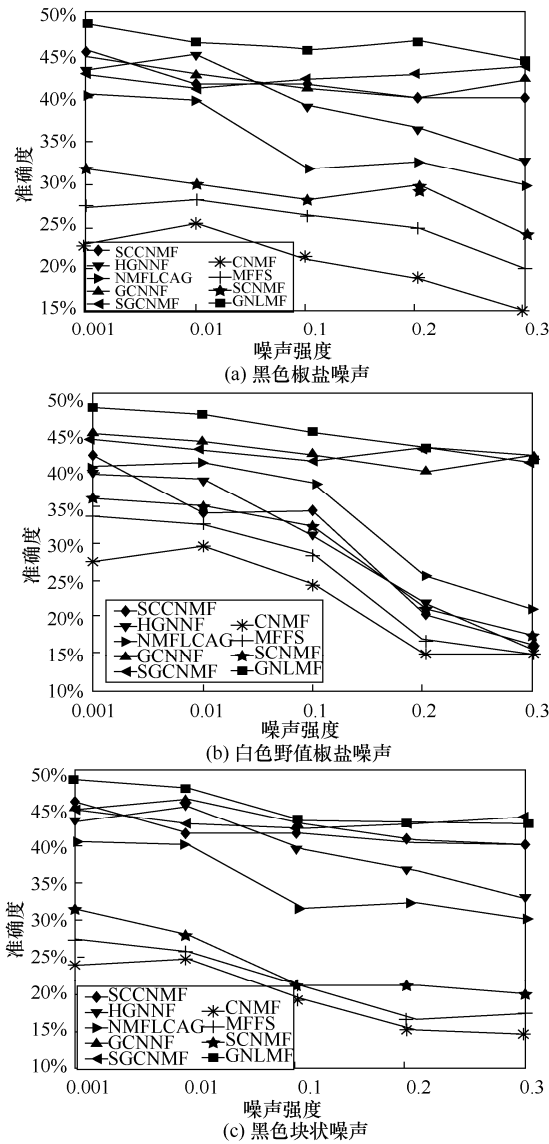


图 2 Faces95 数据集在不同噪声强度下的准确度

4) 非负矩阵分解 ($X \approx UV^T$) 着重于非负矩阵的局部特征提取与线性加权, 而 SGCNMF 拓展非负矩阵到混合符号矩阵的分解, 并引入了带有噪声的数据矩阵作为分解结果的一部分, 即 $X \approx XUV^T$ 这种三分解方式, 这也使 SGCNMF 很难取得比 HNMF 和 GNLMF 明显优异的效果。但是, SGCNMF 分解混合符号矩阵的能力是这 2 种方法所不具备的, 因此在实际使用中也有着更大的应用范围。

5) 对于噪声类型来说, 黑色噪声相较于白色野值椒盐噪声对算法的性能影响更小。这可能是由于在归一化数据后数据接近 0, 即黑色噪声对原始数据的改变幅度小于白色野值椒盐噪声, 这意味着无论是在计算样本间距 (即流形结构) 还是计算分解误差时影响程度均小于白色野值椒盐噪声。

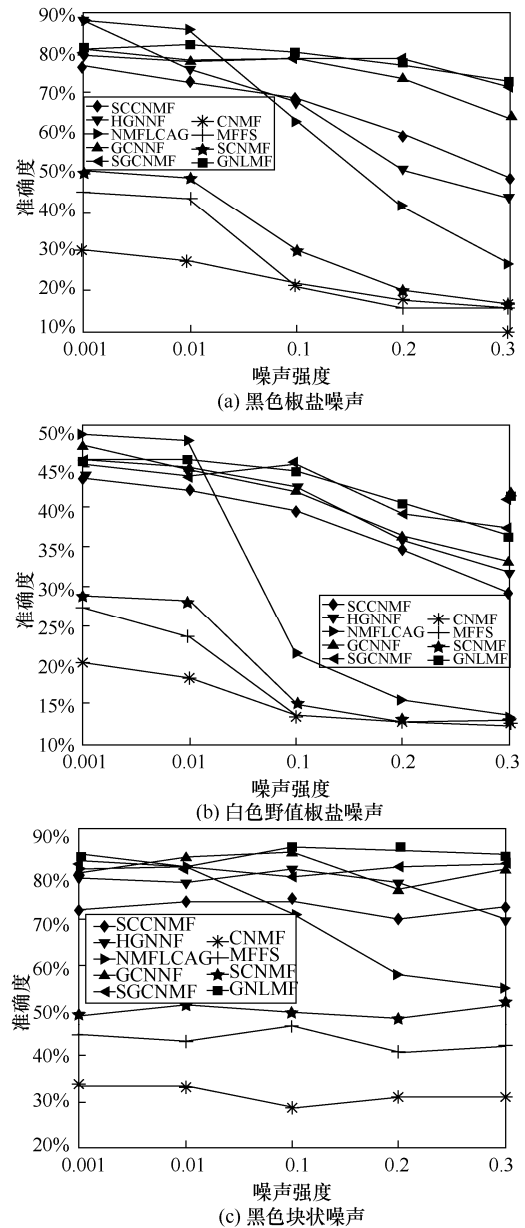


图 3 Grimace 数据集在不同噪声强度下的准确度

6) 对于噪声强度来说, 随着噪声强度的增加, 各个方法的准确度和归一化互信息一般均有不同程度的下降, 这在白色野值椒盐噪声上表现尤为明显。在噪声强度分别为 0.001 和 0.01 时, 其效果相差往往不大, 这说明了流形非负矩阵分解方法本身就带有一定的稀疏性^[23], 在噪声强度过小时并不会过多影响算法的性能。随着噪声强度的增加, 基于流形学习的方法相较于非流形学习方法仍然能保持较大的性能优势。值得注意的是, NMFLCAG 在低噪声环境下效果优异, 但是其性能在噪声强度达到 0.1 时出现大幅度下降, 远低于其他流形学习算法, 这意味着该方法的动态图结构对噪声的稳健性不够强。

4.6 收敛性分析

本文在附录 A 给出在式(23)所示的更新规则条件下，式(11)所示的目标函数式(11)的收敛性理论证明。在聚类实验过程中 SGCNMF 在 3 个数据集的收敛结果如图 4 所示。图中纵轴表示算法的分解误差值，横轴表示迭代次数。从图 4 中明显可以看出，SGCNMF 在 3 个数据集上都很快收敛，在迭代 20 次以内就几乎逼近极限值。

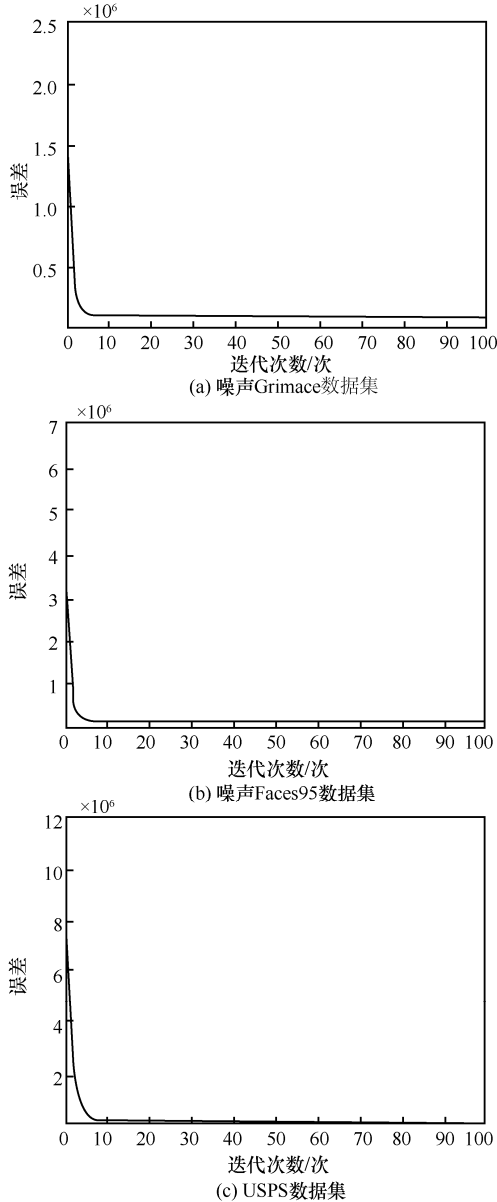


图4 SGCNMF 收敛情况

5 结束语

本文结合流形学习和凸非负矩阵分解，提出了一种稀疏诱导的流形正则化的凸非负矩阵分

解算法 SGCNMF。所提算法在松弛非负矩阵分解的非负约束的同时，通过流形学习，保留了数据空间的结构特征，向基矩阵添加了稀疏约束，给出了乘法更新规则，证明了在该规则下目标函数的收敛性，增强了算法的聚类准确性和对噪声的稳健性。实验结果表明，添加了基矩阵稀疏约束的流形正则化凸非负矩阵分解在聚类问题上，与没有稀疏约束的方法相比聚类准确度得到了提升，并且在数据存在大量噪声时具有更强的抗干扰性。

附录 A 定理 1 的证明

目标函数 O (如式(11)所示) 在 U, V 上并不同时具有凸的性质，但固定一个变量，求解另一个变量时目标函数是收敛的。下面证明目标函数 O 在式(23)所示的乘法更新规则的收敛性。

定义 1 当 $Q(u, u')$ 满足下列 2 个性质时，定义 Q 为函数 T 的辅助函数，具有以下性质。

$$1) Q(u, u') \geq T(u)$$

$$2) Q(u, u') = T(u)$$

引理 1 如果 Q 是 T 的辅助函数，则 T 在式(27)约束条件下迭代时，值不会增大。

$$u^{(t+1)} = \arg \min_u Q(u, u^{(t)}) \quad (27)$$

证明 在约束条件式(27)下，第 t 次迭代时 $Q(u^{(t+1)}, u^{(t)})$ 取得最小值，即

$$Q(u^{(t+1)}, u^{(t)}) \leq Q(u^{(t)}, u^{(t)}) \quad (28)$$

由定义 1 中辅助函数的性质 1) 可得到

$$T(u^{(t+1)}) \leq Q(u^{(t+1)}, u^{(t)}) \quad (29)$$

联立辅助函数的性质 2) 和式(28)、式(29)，可得

$$T(u^{(t+1)}) \leq T(u^{(t)}) \quad (30)$$

即在式(27)所示的约束条件下 T 是非增函数，引理 1 证毕。

要证明式(11)收敛，即证明在式(23)乘法更新规则下的每一次迭代，目标函数 O 的值都不会增大，则在有限次迭代后， O 的值一定可以取到极小值。在 GCNMF 中已经说明了式(9)中 O^* 的收敛性^[23]，且 $\|U\|_{2,1}$ 范数只与 U 相关，即只需证明在式(23)中 U 的迭代规则下 O 的值不会增大即可。

函数 T 表示 O 中仅与 U 相关的部分， T 对 U 求偏导，有

$$T' = \frac{\partial O}{\partial U} = -2X^T XV + 2X^T XUV^T V + 2\beta EU \quad (31)$$

$$T'' = 2X^T XV^T V + 2\beta E \quad (32)$$

引理 2 函数 $Q(u, u^{(t)})$ 是仅与 U 相关的函数 T 的辅助函数，且有

$$Q(u, u^{(t)}) = T(u^{(t)}) + T'(u^{(t)})(u - u^{(t)}) + \left((X^T X)^{-1} V + (X^T X)^+ UV^T V + \beta EU \right) (u - u^{(t)})^2 \quad (33)$$

证明 证明 Q 为 T 的辅助函数, 即证明 Q 符合定义 1 中辅助函数的 2 个性质。很明显, 将 $u^{(t)} = u$ 代入 $Q(u, u^{(t)})$, 有 $Q(u, u) = T(u)$, 即再证明有 $Q(u, u^{(t)}) \geq T(u)$, 则引理 2 得证。

对函数 T 做泰勒展开, 有

$$T(u) = T(u^{(t)}) + T'(u^{(t)})(u - u^{(t)}) + (X^T X V^T V + \beta E)(u - u^{(t)})^2 \quad (34)$$

此时只需证明

$$(X^T X)^{-1} V + (X^T X)^+ UV^T V + \beta EU - u^{(t)}(X^T X V^T V + \beta E) > 0 \quad (35)$$

GCNMF^[23]中已说明在没有稀疏项 $\|U\|_{2,1}$ 时目标函数的

收敛性, 即有

$$(X^T X)^{-1} V + (X^T X)^+ UV^T V - u^{(t)}(X^T X V^T V) > 0 \quad (36)$$

联立式(34)和式(36), 有

$$Q(u, u^{(t)}) \geq T(u) \quad (37)$$

引理 2 证毕。

将式(33)代入式(27), 则在以下更新规则下, Q 是函数 T 的辅助函数, T 是非增函数。

$$u^{(t+1)} = u^{(t)} - u^{(t)} \frac{T'(u^{(t)})_{ik}}{2((X^T X)^{-1} V + (X^T X)^+ UV^T V + \beta EU)_{ik}} = u^{(t)} \frac{\left((X^T X)^+ V + (X^T X)^- UV^T V \right)_{ik}}{\left((X^T X)^{-1} V + (X^T X)^+ UV^T V + \beta EU \right)_{ik}}$$

引理 3 式(11)中的目标函数 O 在式(23)所示更新规则下是非增的。当且仅当 U 和 V 在某一个固定点时, 目标函数 O 在这些更新规则下是稳定的。

证明 在第 t 次迭代时, 对于给定的 v^t , 根据引理 2 和式(23)所示的更新规则, 得到式(38)所示的不等式

$$O(u^{(t+1)}, v^{(t)}, E^{(t)}) \leq O(u^{(t)}, v^{(t)}, E^{(t)}) \quad (38)$$

给定 $u^{(t+1)}$, 根据引理 2 可得

$$O(u^{(t+1)}, v^{(t+1)}, E^{(t)}) \leq O(u^{(t+1)}, v^{(t)}, E^{(t)}) \quad (39)$$

综合式(38)和式(39), 可得

$$O(u^{(t+1)}, v^{(t+1)}, E^{(t)}) \leq O(u^{(t)}, v^{(t)}, E^{(t)}) \quad (40)$$

因此, 有

$$\sum_{r=1}^l \frac{\|u_r^{t+1}\|_2^2}{2\|u_r^t\|_2^2} \leq \sum_{r=1}^n \frac{\|u_r^t\|_2^2}{2\|u_r^t\|_2^2} \quad (41)$$

由 Nie^[32]中引理 1, 有

$$\|u_r^{t+1}\|_2 - \frac{\|u_r^{t+1}\|_2^2}{2\|u_r^t\|_2^2} \leq \|u_r^t\|_2 - \frac{\|u_r^t\|_2^2}{2\|u_r^t\|_2^2} \quad (42)$$

即

$$\sum_{r=1}^l \left(\|u_r^{t+1}\|_2 - \frac{\|u_r^{t+1}\|_2^2}{2\|u_r^t\|_2^2} \right) \leq \sum_{r=1}^l \left(\|u_r^t\|_2 - \frac{\|u_r^t\|_2^2}{2\|u_r^t\|_2^2} \right) \quad (43)$$

结合式(41)和式(43), 可得

$$\sum_{r=1}^l \|u_r^{t+1}\|_2 \leq \sum_{r=1}^l \|u_r^t\|_2 \quad (44)$$

即

$$\|U^{t+1}\|_{2,1} \leq \|U^t\|_{2,1} \quad (45)$$

因此, 有

$$O(u^{(t+1)}, v^{(t+1)}, E^{(t+1)}) \leq O(u^{(t)}, v^{(t)}, E^{(t)}) \quad (46)$$

引理 3 证毕。 T 在式(23)乘法更新准则下是非增函数。

SGCNMF 收敛性证毕。

参考文献:

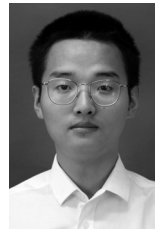
- [1] DUDA R O, HART P E, STORK D G. Pattern classification, seconded[M]. New York: John Wiley & Sons, 2001.
- [2] LEE D D, SEUNG H S. Learning the parts of objects by non-negative matrix factorization[J]. Nature, 1999, 401(6755): 788-791.
- [3] WU W H, KWONG S, ZHOU Y, et al. Nonnegative matrix factorization with mixed hypergraph regularization for community detection[J]. Information Sciences, 2018, 435(4): 263-281.
- [4] CHEN G F, XU C, WANG J Y, et al. Graph regularization weighted nonnegative matrix factorization for link prediction in weighted complex network[J]. Neurocomputing, 2019, 369(12): 50-60.
- [5] 常振超, 陈鸿昶, 黄瑞阳, 等. 基于非负矩阵分解的半监督动态社团检测[J]. 通信学报, 2016, 37(2): 132-143.
CHANG Z C, CHEN H C, HUANG R Y, et al. Semi-supervised dynamic community detection based on non-negative matrix factorization[J]. Journal on Communications, 2016, 37(2): 132-143.
- [6] LIU R, DU B, ZHANG L P. Hyperspectral unmixing via double abundance characteristics constraints based NMF[J]. Remote Sensing, 2016, 8(6): 1-23.
- [7] WANG H, YANG W J, GUAN N Y. Cauchy sparse NMF with manifold regularization: a robust method for hyperspectral unmixing[J]. Knowledge-Based Systems, 2019, 184(104898): 1-16.
- [8] CHEN W S, LIU J, PAN B, et al. Face recognition using nonnegative matrix factorization with fractional power inner product kernel[J]. Neurocomputing, 2019, 348(7): 40-53.
- [9] YANG Z, CHEN W T, HUANG J. Enhancing recommendation on extremely sparse data with blocks-coupled non-negative matrix factorization[J]. Neurocomputing, 2018, 278(2): 126-133.
- [10] LI S Z, HOU X W, ZHANG H J, et al. Learning spatially localized, parts-based representation[C] // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR). Piscataway: IEEE Press. 2001:1-6.
- [11] ZHAN S, WU J, HAN N, et al. Unsupervised feature extraction by

- low-rank and sparsity preserving embedding[J]. *Neural Networks*, 2019, 109(1): 56-66.
- [12] CAI D, HE X F, HAN J H, et al. Graph regularized nonnegative matrix factorization for data representation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(8):1548-1560.
- [13] HUANG S D, WANG H J, LI T, et al. Robust graph regularized non-negative matrix factorization for clustering[J]. *Data Mining and Knowledge Discovery*, 2018, 32(2): 483-503.
- [14] LI X L, CUI G, DONG Y. Graph regularized non-negative low-rank matrix factorization for image clustering[J]. *IEEE Transactions on Cybernetics*, 2016, 47(11): 3840-3853.
- [15] WU B L, WANG E Y, ZHU Z, et al. Manifold NMF with L-21 norm for clustering[J]. *Neurocomputing*, 2018, 273(1): 78-88.
- [16] ZENG K, YU J, LI C, et al. Image clustering by hyper-graph regularized non-negative matrix factorization[J]. *Neurocomputing*, 2014, 138(8): 209-217.
- [17] WANG L, ZHANG Z, LIU G C, et al. Robust adaptive low-rank and sparse embedding for feature representation[C]// *International Conference on Pattern Recognition (ICPR)*. 2018: 800-805.
- [18] YI Y G, WANG J Z, ZHOU W, et al. Non-negative matrix factorization with locality constrained adaptive graph[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, 30(2): 427-441.
- [19] LU G F, WANG Y, ZOU J. Low-rank matrix factorization with adaptive graph regularizer[J]. *IEEE Transactions on Image Processing*, 2016, 25(5): 2196-2205.
- [20] YIN M, XIE S L, WU Z Z, et al. Subspace clustering via learning an adaptive low-rank graph[J]. *IEEE Transactions on Image Processing*, 2018, 27(8): 3716-3728.
- [21] DING C H Q, LI T, JORDAN M I. Convex and semi-nonnegative matrix factorizations[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 32(1):45-55.
- [22] WANG S P, PEDRYCZ W, ZHU Q, et al. Subspace learning for unsupervised feature selection via matrix factorization[J]. *Pattern Recognition*, 2015, 48(1): 10-19.
- [23] HU W J, CHOI K S, WANG P L, et al. Convex non-negative matrix factorization with manifold regularization[J]. *Neural Networks*, 2015, 63(3):94-103.
- [24] CUI G S, LI X L, DONG Y S. Subspace clustering guided convex non-negative matrix factorization[J]. *Neurocomputing*, 2018, 292(5): 38-48.
- [25] LI G P, ZHANG X Y, ZHENG S Y, et al. Semi-supervised convex non-negative matrix factorizations with graph regularized for image representation[J]. *Neurocomputing*, 2017, 237(5):1-11.
- [26] LI Z C, LIU J, TANG J H, et al. Robust structured subspace learning for data representation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(10):2085-2098.
- [27] KONG D G, DING C, HUANG H. Robust non-negative matrix factorization using L21-norm[C]// *20th ACM Conference on Information and Knowledge Management (CIKM)*. New York: ACM Press, 2011: 673-682.
- [28] HOYER P O. Non-negative matrix factorization with sparseness constraints[J]. *Journal of Machine Learning Research*, 2004, 5(1): 1457-1469.
- [29] 张旭, 陈志奎, 高静. 基于图正则化和 $L_{1/2}$ 稀疏约束的非负矩阵分解算法[J]. *小型微型计算机系统*, 2018, 39(11):130-134.
- ZHANG X, CHEN Z K, GAO J. Nonnegative matrix factorization via graph regularization and $L_{1/2}$ sparse constraints[J]. *Journal of Chinese Computer Systems*, 2018, 39(11):130-134.
- [30] LEE D D, SEUNG H S. Algorithms for non-negative matrix factorization[C]// *Advances in Neural Information Processing Systems (NIPS)*. 2001: 556-562.
- [31] WANG Y X, ZHANG Y J. Nonnegative matrix factorization: a comprehensive review[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2012, 25(6): 1336-1353.
- [32] LUXBURG U V. A tutorial on spectral clustering[J]. *Statistics and Computing*, 2007, 17(4):395-416.
- [33] NIE F P, HUANG H, CAI X, et al. Efficient and robust feature selection via joint $\ell_{2,1}$ -norms minimization[C]// *Advances in Neural Information Processing Systems (NIPS)*. 2010: 1813-1821.
- [34] HULL J J. A database for handwritten text recognition research[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1994, 16(5): 550-554.
- [35] CAI D, HE X F, HAN J H. Document clustering using locality preserving indexing[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2005, 17(12):1624-1637.

[作者简介]



邱飞岳（1965—），男，浙江诸暨人，博士，浙江工业大学教授、博士生导师，主要研究方向为智能计算、机器学习、虚拟现实等。



陈博文（1996—），男，安徽合肥人，浙江工业大学硕士生，主要研究方向为机器学习与虚拟现实等。

陈铁明（1978—），男，浙江诸暨人，博士，浙江工业大学教授、博士生导师，主要研究方向为网络空间安全与大数据智能分析等。

章国道（1988—），男，浙江衢州人，浙江工业大学博士生，主要研究方向为数据挖掘。