

基于核的 $L_{2,1}$ 范数非负矩阵分解在图像聚类中的应用

余江兰¹, 李向利¹, 董晓亮²

- (1. 桂林电子科技大学数学与计算科学学院; 广西密码学与信息安全重点实验室;
广西自动检测技术与仪器重点实验室, 广西 桂林 541004)
(2. 北方民族大学数学与信息科学学院, 宁夏 银川 750021)

摘要: 本文研究了基于核技巧的 $L_{2,1}$ 范数非负矩阵分解在图像聚类中的问题. 利用基于核的稀疏鲁棒非负矩阵分解方法, 获得了算法良好的稀疏性和鲁棒性, 提高了聚类性能, 该方法也可以推广到文本聚类的应用.

关键词: 非负矩阵分解; 核技巧; $L_{2,1}$ 范数; 稀疏性; 鲁棒性

MR(2010) 主题分类号: 91C20 中图分类号: O235

文献标识码: A 文章编号: 0255-7797(2019)03-0440-15

1 前言

非负矩阵分解 (Non-negative Matrix Factorization, 简称为 NMF) [1, 2] 是继 PCA (主成分分析) [3]、ICA (独立成分分析) [4]、VQ (矢量量化) [5] 等矩阵分解方法之后提出的一种新的矩阵分解方法. 近年来, NMF 算法引起了各个领域科研人员的重视, 因为 NMF 的思想为人类处理大规模数据提出了一种新方法, 且该方法相较于一些传统的方法而言, 具有实现简便、易于存储、分解结果可解释的优点. NMF 是在一般的矩阵分解的基础上对矩阵添加了非负的限制, 其基本思想可简单概括为: 给定一个非负数据矩阵, 试图找到两个非负低秩矩阵, 使得它们的乘积能够无限逼近原始非负数据矩阵. 标准 NMF [6] 是将 Frobenius 范数 (简记为 F 范数) 作为目标函数, 计算起来很方便快捷, 但是它不能有效地处理存在于数据中的噪音值和异常值点. 因此, 急需找到一个能够改善这些问题的非负矩阵分解新版本.

此前, 源于其对事物的局部特性有很好的解释, NMF 已经被广泛应用于很多领域. 比如, 图像分析、文本数据和数据挖掘、语音处理、机器人控制、生物医学工程和化学工程, 此外, NMF 算法在环境数据处理、信号分析与复杂对象的识别方面都有着很好的应用. 扩展的 NMF [7] 也在不断适应着各种各样的目标函数, 从而能够涉入不同的数据分析问题, 包括分类、协同过滤和聚类等等.

聚类是特征学习和计算机视觉的最重要且具有挑战性的任务之一. 在过去的几十年里, 研究学者们为各种应用程序设计了许多聚类方法, 包括图像注释 [8]、图像检索 [9]、图像分类

*收稿日期: 2018-04-25 接收日期: 2018-07-06

基金项目: 国家自然科学基金 (11601012; 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 (YQ16112; YQ18112); 宁夏自然科学基金 (NZ17103); 桂林电子科技大学研究生优秀学位论文培育项目资助 (16YJPYSS22).

作者简介: 余江兰 (1993-), 女, 重庆, 硕士, 主要研究方向: 最优化理论与算法.

通讯作者: 李向利

[10]、图像分割 [11]、数据挖掘 [12] 和图像索引 [13]。然而, 对于图像聚类任务来说, 一个非常重要的步骤是找到原始数据的有效表示。为此, 不同的研究人员做了大量的工作, 他们通常强调挖掘原始数据的内在结构信息, 并使新的表示更具辨别性。传统的聚类已经比较成功的解决了低维数据的聚类问题。但是由于实际应用中数据的复杂性和冗余性, 在处理许多问题时, 现有的算法经常失效, 特别是对于高维数据和大型数据的情况。因为传统聚类方法在高维数据集中进行聚类时, 主要遇到两个问题 [14]: (1) 高维数据集中存在一些与聚类无关的维度, 从而在对数据降维处理中带来一定的挑战; (2) 高维空间中数据比低维空间中数据分布更稀疏和零散, 常存在数据点之间的距离几乎相等的情况, 而传统聚类方法是基于距离进行聚类的, 因此在高维空间中无法基于距离来构建簇。

而 NMF 作为一种处理大规模数据的矩阵分解方法, 它已逐渐成为图像、文本聚类与数据挖掘等领域最受欢迎的工具之一。实践证明, 利用 NMF 对文本、图像等高维数据进行处理时, 比传统的矩阵分解处理方法速度更快、更便捷, 被认为是对非负数据进行处理的一种有效途径, 已经引起了国内外许多科学家和研究人员的广泛关注。

近年来, 为了不断改善已有的 NMF 方法存在的问题, 很多扩展版 NMF 被提出。为了提高原始 NMF 的稀疏性, 通过将局部约束整合到标准的 NMF 中, Li 等人 [15] 提出了 LNMF (局部非负矩阵分解)。这种方法不仅可以学习对象的稀疏表示, 而且可以表明局部特征。之后, 为了确保对象的稀疏性, Hoyer 等人 [16] 将 L1- 范数约束强加在编码矩阵上, 提出了 NSC (非负稀疏编码)。虽然他们没有考虑数据的内在结构信息, 但从某种程度上说, NSC 和 LNMF 从不同角度改善了 NMF 的稀疏性。Barman 等人 [17] 提出了基于文本聚类的非负矩阵分解, 包括特征的提取和分类。杨等人 [18] 提出了非负矩阵分解的投影算法, 并将该算法人脸图像的处理中, 实验效果较好; 蔡等人试图涉及几何数据空间的信息提出了 GNMF (图正则非负矩阵分解 [7]), 它是通过构建 K - 近邻 (KNN) 图结构来编码几何结构。由于简单图中的一个边可以连接两个顶点, 所以 GNMF 只考虑两个样本之间的关系, 忽略了多个样本之间的关系。为了探索多个样本间的高阶关系, 通过创建一个超图来编码多个样本间的关系, 由于超图可以连接两个以上的顶点, 故 HNMF (超图正则化非负矩阵分解) 能找到数据的高阶关系。Zeng 等人 [19] 提出了 HNMF。虽然 GNMF 和 HNMF 运用了样本内在的流形结构, 但没有考虑带判别信息的因素。为了结合带判别力的信息, 李等人 [20] 通过合并局部流行图正则和带判别力的标签信息, 提出了 GDNMF (基于图带标签判别力的非负矩阵分解)。为了将 NMF 延伸到子空间聚类中, Dijana 等人 [21] 借助核技巧 [22, 23] 来揭示流形的非线性性质, 并将数据固有的局部几何性质都考虑在内, 提出了非线性正交非负矩阵分解, 使得聚类性能得以提升。

可以看出, 以上列举的 NMF 方法都是在不断从不同角度揭示了充分利用原始数据内在的流行几何结构, 但没有揭示其稀疏性和鲁棒性。基于此, 本文在保留原始数据的内在流行几何结构和运用核技巧来揭示流形的非线性性质的基础上, 添加了 $L_{2,1/2}$ 矩阵伪范数 [24] 作为额外的稀疏约束。由于稀疏约束能够选择有判别力的稀疏特征来改善算法的有效性, 所以它吸引了很多研究学者的极大关注。稀疏约束旨在借助一个有效的稀疏模型来实现原始数据的稀疏表示。虽然非负矩阵分解对原始的数据矩阵可以起到降维的作用, 但对于高维数据来说, 它的计算仍然是很复杂的。因此, 如果能改善基矩阵和系数矩阵的稀疏性, 那么将降低计算的复杂度。此外, 本文的目标模型中, 没有使用常用的欧氏距离作为残差项, 而是用 $L_{2,1}$ 范数 [25] 来替代标准 NMF 中的 F 范数。引入 $L_{2,1}$ 范数有三点原因 [25]: (1) 在众多的实际数据中, 都包含了很多模糊的噪音值和异常值点等, 而 $L_{2,1}$ 范数可以有效地处理原始数据中存在的这

些问题; (2) 能提供一个有效的更新规则, 进而能有效提高算法的稀疏性和鲁棒性; (3) 以它为损失函数所需的计算成本和标准的 NMF 几乎差不多, 并且能提高算法的性能.

本文的其余部分结构组织如下. 第二部分主要介绍了标准的非负矩阵分解及相关理论部分, 第三部分是本文的重要组成部分, 包括本文目标模型的构建、更新迭代规则的推导、新算法的提出以及收敛性分析, 第四部分是本文的数值实验展示实验结果分析, 第五部分是对本文的总结和概括. 本文中 A_i 矩阵 A 的第 i 列, A^i 表示 A 矩阵的第 i 行.

2 标准的非负矩阵分解

NMF 是目前国际上新的矩阵分解方法. 它已广泛应用于诸多领域, 如图像处理、生物学、文本聚类 and 语音信号处理等. 标准的 NMF 问题可描述如下.

给定非负矩阵 $X \in R_+^{m \times n}$, 旨在找到两个非负低秩矩阵 $U \in R_+^{m \times k}$ 和 $V \in R_+^{k \times n}$, 使得 $X \approx UV$, 其中 U 为基矩阵, V 为系数矩阵. 且 m, k, n 满足 $(m+n)k \ll mn$.

换句话说, 即解决极小化残差矩阵

$$O_F = \|X - UV\|_F^2 \quad \text{s.t.} \quad U \geq 0, V \geq 0, \quad (2.1)$$

其中 $\|\cdot\|_F$ 矩阵的表示 Frobenius 范数, 通过计算 F 范数的平方从而得到两个矩阵的欧式距离. Lee 等人^[2, 6] 给出了极小化问题 (2.1) 的更新迭代规则, 并证明了其收敛性, 迭代规则如下

$$u_{ij} = u_{ij} \frac{(XV^T)_{ij}}{(UVV^T)_{ij}}, \quad v_{ij} = v_{ij} \frac{(U^T X)_{ij}}{(U^T UV)_{ij}}. \quad (2.2)$$

3 基于核的 $L_{2,1}$ 范数非负矩阵分解

3.1 目标函数

在文献 [21] 中, Dijana 等人考虑一个非线性映射, 将数据点映射到高位数据空间: $x_i \rightarrow \Phi(x_i)$, 或者 $X \rightarrow \Phi(X) = (\Phi(x_1), \Phi(x_2), \dots, \Phi(x_n)) \in R^{d \times n}$, 其中 x_i 表示原始数据空间 X 的第 i 样本点. 非线性非负矩阵分解旨在找到两个非负低秩矩阵, 使得它们的乘积可以无限逼近原始矩阵的映射 $\Phi(X)$. 基于此, Dijana 等人^[21] 提出了基于核的图正则正交非负矩阵分解模型, 即

$$\min_{F, V} \|\Phi(X) - \Phi(X) FV\|_F^2 + \lambda \text{Tr}(V L V^T). \quad (3.1)$$

此模型的优势在于, 基于内核的非负光谱聚类算法, 引入了一个图正则化项, 通过图正则项来捕获非线性特征空间中固有的局部几何结构, 从而使得因子分解方法具有更强的识别能力, 可以从更高维度环境空间的子流形中提取数据点.

但 Dijana 等人却忽视了原始数据间的稀疏性和鲁棒性. 为了充分利用原始数据的内在信息和提高算法的稀疏性及鲁棒性, 基于 $L_{2,1}$ 范数能够处理原始数据中的异常点和极端值点的作用, 本文用 $L_{2,1}$ 范数对原有的 Frobenius 范数进行改进, 并引入混合 $L_{2,1/2}$ 矩阵伪范数,

结合范数算子和非负约束保证了投影矩阵的稀疏性. 构建新的目标函数如下

$$\begin{aligned} \min_{F,V} & \|\Phi(X) - \Phi(X)FV\|_{2,1} + \frac{1}{2}\lambda \text{Tr}(VLV^T) + 2\beta \|V\|_{2,\frac{1}{2}}^{\frac{1}{2}} + 2\xi \|F\|_{2,\frac{1}{2}}^{\frac{1}{2}} \\ \text{s.t.} & F, V \geq 0, \lambda, \beta, \xi \geq 0, \end{aligned} \quad (3.2)$$

其中 λ, β, ξ 是平衡因子, $L = D - W$ 为拉普拉斯矩阵, D 是一个对角矩阵, $D_{jj} = \sum_l W_{jl}$,

$$W_{jl} = \begin{cases} 1, & \text{如果 } x_j \in N_p(x_l) \text{ 或 } x_l \in N_p(x_j), \\ 0, & \text{否则,} \end{cases}$$

这里的 $N_p(x_l)$ 表示 x_l 的 p 个近邻集合^[26]. 对于 (3.2) 式, 算法只能保证其局部极小化收敛性. 其中, 该目标函数的第一项是原始数据矩阵和低秩逼近矩阵乘积间的残差项, 也是改进算法鲁棒性的重要根据; 第二项是图正则项, 为了保证其数据的几何结构和控制样本数据点间的局部几何结构的度; 第三项和第四项分别是控制对应变量的稀疏性.

3.2 更新规则

记目标函数为

$$\mathcal{O} = \|\Phi(X) - \Phi(X)FV\|_{2,1} + \frac{1}{2}\lambda \text{Tr}(VLV^T) + 2\beta \|V\|_{2,\frac{1}{2}}^{\frac{1}{2}} + 2\xi \|F\|_{2,\frac{1}{2}}^{\frac{1}{2}}, \quad (3.3)$$

其中

$$\begin{aligned} \|\Phi(X) - \Phi(X)FV\|_{2,1} &= \sum_{i=1}^n \sqrt{\sum_{j=1}^m (\Phi(X) - \Phi(X)FV)_{ji}^2} \\ &= \sum_{i=1}^n \|\Phi(X)_i - \Phi(X)FV_i\|, \end{aligned} \quad (3.4)$$

$\|V\|_{2,\frac{1}{2}} = \left(\sum_{i=1}^k \|V^i\|_2^{\frac{1}{2}}\right)^2$ 是混合 $L_{2,1/2}$ 矩阵伪范数, 结合范数算子和非负约束保证了投影矩阵的稀疏性.

令 $Y = \|V\|_{2,\frac{1}{2}}^{\frac{1}{2}}$, 则有 $Y = \sum_{i=1}^k \|V^i\|_2^{\frac{1}{2}}$. 对 V 中的任一元素 V_{po} 求偏导数有

$$\begin{aligned} \frac{\partial Y}{\partial V_{po}} &= \frac{\partial}{\partial V_{po}} \left(\sum_{i \neq p} \|V^i\|_2^{\frac{1}{2}} + \|V^p\|_2^{\frac{1}{2}} \right) \\ &= \frac{\partial}{\partial V_{po}} (V_{p1}^2 + V_{p2}^2 + \cdots + V_{po}^2 + \cdots + V_{pn}^2)^{\frac{1}{4}} \\ &= \frac{1}{4} (V_{p1}^2 + V_{p2}^2 + \cdots + V_{pk}^2)^{-\frac{3}{4}} \cdot 2V_{po} \\ &= \frac{1}{2} \|V^p\|_2^{-\frac{3}{2}} \cdot V_{po} \\ &= \frac{1}{2} \times 4 \times \frac{1}{4} \|V^p\|_2^{-\frac{3}{2}} V_{po}. \end{aligned} \quad (3.5)$$

定义 $P = \frac{1}{4} \|V^p\|_2^{-\frac{3}{2}}$, 从而得到

$$\frac{\partial Y}{\partial V} = 2PV. \quad (3.6)$$

先对目标函数分别关于 F 求偏导数, 结合 $Y = \|V\|_{2, \frac{1}{2}}^{\frac{1}{2}}$ 关于 V 的导数, 可以得到

$$\begin{aligned} \frac{\partial \mathcal{O}}{\partial F_{jk}} &= \sum_{i=1}^n \frac{1}{\sqrt{\sum_{j'} (\Phi(X) - \Phi(X) FV)_{j'i}^2}} \\ &\quad \sum_{j''=1}^d (\Phi(X) - \Phi(X) FV)_{j''i} \frac{\partial}{\partial F_{jk}} (\Phi(X) - \Phi(X) FV)_{j''i} + 4\xi(QF)_{jk} \\ &= \sum_{i=1}^n \frac{-1}{\sqrt{\sum_{j'} (\Phi(X) - \Phi(X) FV)_{j'i}^2}} (\Phi(X) - \Phi(X) FV)_{ji} (\Phi(X) V)_{ki} + 4\xi(QF)_{jk} \\ &= \sum_{i=1}^n (\Phi(X) - \Phi(X) FV)_{ji} (\Phi(X) V)_{ki} G_{ii} + 4\xi(QF)_{jk} \\ &= -(\Phi(X) - \Phi(X) FV)_{ji} G_{ii} (\Phi(X) V)_{ik}^T + 4\xi(QF)_{jk} \\ &= -\Phi(X)_{ji} G_{ii} (\Phi(X) V)_{ik}^T + (\Phi(X) FV)_{ji} G_{ii} (\Phi(X) V)_{ik}^T + 4\xi(QF)_{jk} \\ &= -(\Phi(X)^T \Phi(X) G V^T)_{jk} + (\Phi(X)^T \Phi(X) F V G V^T)_{jk} + 4\xi(QF)_{jk} \\ &= -(K G V^T)_{jk} + (K F V G V^T)_{jk} + 4\xi(QF)_{jk}, \end{aligned} \quad (3.7)$$

其中 G 和 Q 都是对角矩阵, 对角元素分别为

$$g_{ii} = \frac{1}{\|\Phi(X)_i - \Phi(X) FV_i\|}, \quad q_{ii} = \frac{1}{4 \max(\|F^i\|_2^{3/2})}.$$

为了避免它们的分母为 0 的情况, 分别添加两个足够小的正常数 ε_1 和 ε_2 在矩阵 G, Q 的定义中, 则它们的元素重新表示为

$$g_{ii} = \frac{1}{\max(\|\Phi(X)_i - \Phi(X) FV_i\|, \varepsilon_1)}, \quad q_{ii} = \frac{1}{4 \max(\|F^i\|_2^{3/2}, \varepsilon_2)},$$

这里的 K 是核矩阵^[2,3], 定义为 $K \equiv \Phi(X)^T \Phi(X)$, 其中 $\Phi(X)$ 是在非线性无限维特征空间的特征矩阵.

对 F 运用 Karush-Kuhn-Tucker 最优条件 (简记为 KKT 条件) $\frac{\partial \mathcal{O}}{\partial F_{jk}} = 0, \forall j, k$, 得到

$$-(K G V^T)_{jk} F_{jk} + (K F V G V^T)_{jk} F_{jk} + 2\xi(QF)_{jk} F_{jk} = 0, \quad \forall j, k, \quad (3.8)$$

进行简单的代数推导变换即可得到 F 的更新迭代规则为

$$F_{jk} \leftarrow F_{jk} \frac{(K G V^T)_{jk}}{(K F V G V^T + 2\xi(QF))_{jk}}. \quad (3.9)$$

现在对目标函数关于 V 求偏导数, 有

$$\begin{aligned}
 \frac{\partial \mathcal{O}}{\partial V_{ki}} &= \sum_{i'=1}^n \frac{1}{\sqrt{\sum_{j'} (\Phi(X) - \Phi(X) FV)_{j'i}^2}} \sum_{j=1}^d (\Phi(X) - \Phi(X) FV)_{ji'} \frac{\partial}{\partial V_{ki}} (\Phi(X) - \Phi(X) FV)_{ji'} \\
 &\quad + \lambda(VL)_{ki} + 4\beta(PV)_{ki} \\
 &= \sum_{i'=1}^n \frac{1}{\sqrt{\sum_{j'} (\Phi(X) - \Phi(X) FV)_{j'i}^2}} \\
 &\quad \sum_{j=1}^d (\Phi(X) - \Phi(X) FV)_{ji'} (\Phi(X) F)_{ji}^T + \lambda(VL)_{ki} + 4\beta(PV)_{ki} \\
 &= - \left(F^T \Phi(X)^T (\Phi(X) - \Phi(X) FV) \right)_{ki} G_{ii} + \lambda(VL)_{ki} + 4\beta(PV)_{ki} \\
 &= - (F^T KG)_{ki} + (F^T KFVG)_{ki} + \lambda(VL)_{ki} + 4\beta(PV)_{ki} \\
 &= - (F^T KG)_{ki} + (F^T KFVG)_{ki} + \lambda(V(D - W))_{ki} + 4\beta(PV)_{ki},
 \end{aligned} \tag{3.10}$$

其中 P 是对角矩阵, 其对角元素为

$$p_{ii} = \frac{1}{4 \max \left(\|V^i\|_2^{3/2} \right)},$$

为了避免分母为 0 的情况, 添加一个足够小的正常数 ε_3 在矩阵 Q 的定义中, 即

$$p_{ii} = \frac{1}{4 \max \left(\|V^i\|_2^{3/2}, \varepsilon_3 \right)}.$$

对 V 运用 KKT 条件 $\frac{\partial \mathcal{O}}{\partial V_{ki}} V_{ki} = 0, \forall i, k$, 则有

$$\left(-F^T KG + F^T KFVG + \lambda V(D - W) + 4\beta PV \right)_{ki} V_{ki} = 0 \tag{3.11}$$

成立. 对其进行简单的代数变换即可得到 V 的更新迭代规则为

$$V_{ki} \leftarrow V_{ki} \frac{(F^T KG + \lambda VW)_{ki}}{(F^T KFVG + \lambda VD + 4\beta PV)_{ki}} \tag{3.12}$$

关于迭代更新规则 (3.9) 和 (3.12), 有定理 1 成立.

定理 1 目标函数 (3.1) 在迭代更新规则 (3.9) 和 (3.12) 下是非增长的.

类似于文献 [21], 下面给出本文的新算法.

算法: KRSNMF

输入: $X \in R_+^{m \times n}$, $1 \leq k \leq \min\{m, n\}$, $\lambda = 0.05$, $\xi = 0.1$; $\beta = 3.5$, $\sigma = 0.22$;

输出: $U \in R_+^{d \times k}$, $V \in R_+^{k \times n}$, $F \in R_+^{n \times k}$.

步骤 1: 随机初始化 U, V, F 使得它们的取值 0 到 1 之间.

重复

步骤 2: 更新 V 根据 (3.12);

更新 F 根据 (3.9);

更新 U 根据 $U = \Phi(X)F$;

直到 满足停止条件.

3.3 收敛性分析

定义 1 如果 $Z(v, v')$ 和 $J(v)$ 之间满足 $Z(v, v') \geq J(v)$, $Z(v, v) = J(v)$, 那么 $Z(v, v')$ 是 $J(v)$ 的辅助函数.

欲证明 $Z(v, v')$ 是 $J(v)$ 的辅助函数, 首先给出引理 1.

引理 1 如果 Z 是 J 的辅助函数, 则在更新规则

$$v^{(t+1)} = \arg \min_v Z(v, v^{(t)}) \quad (3.13)$$

下, J 是非增长的, 其中右上角的 t 表示数值计算中的某一步, $t+1$ 表示 t 的下一步.

证

$$J(v^{(t+1)}) \leq Z(v^{(t+1)}, v^{(t+1)}) \leq Z(v^{(t)}, v^{(t)}) = J(v^{(t)}).$$

换句话说, 只要辅助函数 $Z(v, v')$ 达到极小值, 函数 $J(v)$ 也会达到极小值.

对于目标函数 (3.1), 考虑到矩阵 V 中的元素 V_{ab} , 记 J_{ab} 为目标函数中与 V_{ab} 相关的部分, 对其关于 V_{ab} 求偏导数, 有

$$\begin{aligned} J_{ab}' &= \left(\frac{\partial \mathcal{O}}{\partial V} \right)_{ab} \\ &= (-F^T K G + F^T K F V G + 2\lambda V (D - W) + 2\beta P V)_{ab}, \end{aligned} \quad (3.14)$$

$$\begin{aligned} J_{ab}'' &= (F^T K^T F G + 2\lambda (D - W) + 2\beta P)_{ab} \\ &= (F^T K^T F G)_{ab} + 2\lambda (D - W)_{aa} + 2\beta P_{aa}. \end{aligned} \quad (3.15)$$

引理 2 函数

$$\begin{aligned} Z(v, v_{ab}^{(t)}) &= J_{ab}(v_{ab}^{(t)}) + J_{ab}'(v_{ab}^{(t)}) (v - v_{ab}^{(t)}) \\ &\quad + \frac{(V^T F^T K^T F G)_{ab} + 2\lambda (V D)_{ab} + 2\beta (P V)_{ab}}{v_{ab}^{(t)}} (v - v_{ab}^{(t)})^2 \end{aligned} \quad (3.16)$$

是 J_{ab} 的辅助函数.

证 显然有 $Z(v, v) = J_{ab}(v)$ 成立. 要证明 $Z(v, v_{ab}^{(t)})$ 是 $J_{ab}(v)$ 的辅助函数, 只需要证明 $Z(v, v_{ab}^{(t)}) \geq J_{ab}(v)$ 即可. 因为

$$\begin{aligned} J_{ab}(v) &= J_{ab}(v_{ab}^{(t)}) + J_{ab}'(v_{ab}^{(t)})(v - v_{ab}^{(t)}) \\ &\quad + ((F^T K^T FG)_{ab} + 2\lambda(D - W)_{aa} + 2\beta P_{aa})(v - v_{ab}^{(t)})^2. \end{aligned} \quad (3.17)$$

要证明 $Z(v, v_{ab}^{(t)}) \geq J_{ab}(v)$, 等价于证明

$$\begin{aligned} &\frac{(V^T F^T K^T FG)_{ab} + 2\lambda(VD)_{aa} + 2\beta(PV)_{aa}}{v_{ab}^{(t)}} \\ &\geq (F^T K^T FG)_{ab} + 2\lambda(D - W)_{aa} + 2\beta P_{aa}. \end{aligned} \quad (3.18)$$

又因为

$$(V^T F^T K^T FG)_{ab} = \sum_{l=1}^k V_{al}^{(t)} (F^T K^T FG)_{lb} \geq V_{ab}^{(t)} (F^T K^T FG)_{bb}, \quad (3.19)$$

$$\lambda(VD)_{ab} = \lambda \sum_{j=1}^n V_{aj}^{(t)} D_{jb} \geq \lambda V_{ab}^{(t)} (D - W)_{bb} = \lambda V_{ab}^{(t)} L_{bb}, \quad (3.20)$$

$$\beta(PV)_{aa} = \beta \sum_{l=1}^k P_{al} V_{lb}^{(t)} \geq 4\beta P_{aa} V_{ab}^{(t)}. \quad (3.21)$$

所以 (3.17) 式成立, 进而有 $Z(v, v_{ab}^{(t)})$ 是 $J_{ab}(v)$ 成立.

综上所述, $Z(v, v_{ab}^{(t)})$ 是 $J_{ab}(v)$ 的辅助函数. 下面证明定理 1.

定理 1 的证明 用 (3.16) 中的 $Z(v, v_{ab}^{(t)})$ 替换 (3.13) 中的 $Z(v, v^{(t)})$, 有如下的更新规则

$$\begin{aligned} v_{ab}^{(t+1)} &= v_{ab}^{(t)} - v_{ab}^{(t)} \frac{J_{ab}'(v_{ab}^{(t)})}{(V^T F^T K^T FG)_{ab} + \lambda(VD)_{ab} + 4\beta(PV)_{ab}} \\ &= v_{ab}^{(t)} \frac{(F^T K^T FG)_{ab}}{(F^T K^T FG)_{ab} + \lambda(VD)_{ab} + 4\beta(PV)_{ab}}. \end{aligned} \quad (3.22)$$

因此 J_{ab} 在此更新迭代规则下是非增长的. 故定理 1 得证.

4 数值实验

这部分主要介绍了本文在图片聚类中的数值实验的相关内容. 聚类是在无类别标记信息的情况下将对象自动分组, 换句话说, 它旨在将具有相似性质的对象分到同一簇中, 根据不同的算法, 聚类得到的结果也不尽相同. 如图 1, 2, 3 所示. 图片相关的信息来源于 <http://www.cad.zju.edu.cn/home/dengcai/Data/data.html>. 图 4 是本文用到的 KRSNMF 算法借助幂指数核得到的聚类图.

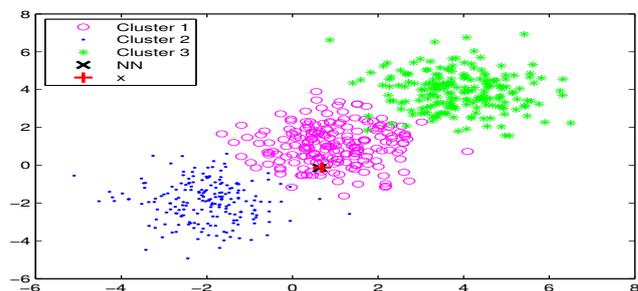


图 1: 快速近邻法聚类效果图

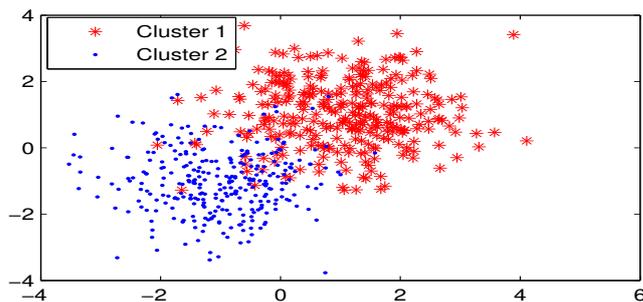


图 2: 压缩剪辑近邻算法 (Condensing) 初始样本分布图

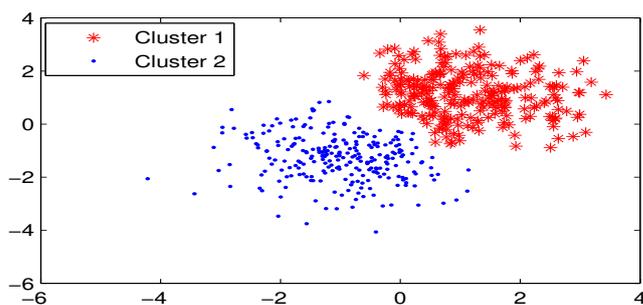


图 3: 压缩剪辑近邻算法 (Condensing) 剪辑后样本分布图

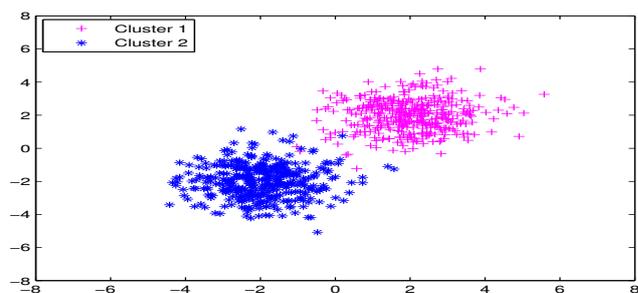


图 4: KRSNMF 算法聚类图

本文中选用的数据集都是常用的标准数据集, 即

$$\begin{aligned} & WarpAR10P, Yale_{64 \times 64}, Yale_{32 \times 32}, Yale, ORL_{64 \times 64}, \\ & orlraws10P, Pixraws10P, ORL_{32 \times 32}, COIL20. \end{aligned}$$

文中所有的数值计算都是在处理器为 Intel(R) Core(TM) i5-6500 CPU@3.20 GHz 3.19 GHz, 内存为 8.00 GB 的 64 位操作系统上进行的, 算法代码用 MATLAB R2014a 进行编写. 本文分别选取了三种核技巧在这九个数据集上进行实验, 即高斯核技巧、幂指数核技巧和拉普拉斯核技巧. 它们的计算方式分别为

$$\begin{aligned} K(x_i, x_j) &= \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \quad K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|}{\sigma}\right), \\ K(x_i, x_j) &= \exp\left(-\frac{\|x_i - x_j\|}{\sigma}\right). \end{aligned}$$

实验中的参数选取如下 $\lambda = 0.05, \xi = 0.1; \beta = 3.5, \sigma = 0.22$.

4.1 评价指标

精度 (Accuracy, ACC) 和归一化互信息 (Normalized Mutual Information, NMI) 是被广泛应用于不同聚类方法的评价聚类性能的指标. 如果 x_i 表示原始数据空间给定的数据点, l_i 是根据某一个聚类算法由数据点 x_i 计算得到的类标签, g_i 是真实的类标签. $\text{map}(\cdot)$ 是从 l_i 到 g_i 的最佳映射函数, 可以由 Hungarian 算法^[27] 计算得到. 因此聚类精度被定义如下:

$$ACC = \frac{1}{N} \sum_{i=1}^N \delta(l_i, \text{map}(g_i)),$$

这里的 N 是总数据点个数, $\delta(x, y)$ 是一个 delta 函数. 计算规则: 如果 $x = y$, 则 $\delta(x, y) = 1$; 否则 $\delta(x, y) = 0$.

如果用 C 来表示真实类标集, S 表示从某一个算法得到的类标集, 则它们的互信息 $MI(C, S)$ 定义如下:

$$MI(C, S) = \sum_{c_i \in C, s_j \in S} p(c_i, s_j) \log \frac{p(c_i, s_j)}{p(c_i) \cdot p(s_j)},$$

其中 $p(c_i)$ 和 $p(s_j)$ 分别是原始数据集中的任意样本点属于 c_i 和 s_j 的概率, $p(c_i, s_j)$ 是任意样本同时属于 c_i 和 s_j 的联合概率. 在实验中, 归一化互信息的计算方式为

$$NMI = 2 \frac{MI(C, S)}{H(C) + H(S)},$$

这里的 $H(C)$ 和 $H(S)$ 分别是 C 和 S 的信息熵, 且 NMI 的值在从 0 到 1 内. 如果两个聚类集合是同一的, 那么 $NMI = 1$; 如果两个集合是相互独立的, 则 $NMI = 0$. 也就是说, 如果两个数据点相似度越高, 越有可能被自动分到一个簇.

4.2 实验结果分析

表 1: 高斯核聚类性能表

Dataset	Algorithm	<i>ACC</i>	<i>NMI</i>
WarpAR10P	KOGNMF	0.2231	0.1456
	KRSNMF	0.2769	0.2587
Yale _{64×64}	KOGNMF	0.1939	0.2202
	KRSNMF	0.2727	0.3401
Yale _{32×32}	KOGNMF	0.2121	0.2519
	KRSNMF	0.2606	0.3407
Yale	KOGNMF	0.2121	0.2509
	KRSNMF	0.2606	0.3407
ORL _{64×64}	KOGNMF	0.0725	0.3641
	KRSNMF	0.0675	0.4445
orlraws10P	KOGNMF	0.24	0.1864
	KRSNMF	0.42	0.4623
Pixraws10P	KOGNMF	0.24	0.1960
	KRSNMF	0.47	0.4709
ORL _{32×32}	KOGNMF	0.07	0.3590
	KRSNMF	0.075	0.4492
COIL20	KOGNMF	0.0958	0.0439
	KRSNMF	0.4049	0.5045

表 2: 幂指数核聚类性能表

Dataset	Algorithm	<i>ACC</i>	<i>NMI</i>
WarpAR10P	KOGNMF	0.2077	0.1414
	KRSNMF	0.2923	0.2732
Yale _{64×64}	KOGNMF	0.2818	0.2387
	KRSNMF	0.2788	0.3413
Yale _{32×32}	KOGNMF	0.2303	0.2778
	KRSNMF	0.3879	0.4443
Yale	KOGNMF	0.2303	0.2778
	KRSNMF	0.3879	0.4443
ORL _{64×64}	KOGNMF	0.0775	0.3728
	KRSNMF	0.075	0.4447
orlraws10P	KOGNMF	0.23	0.1791
	KRSNMF	0.42	0.5201
Pixraws10P	KOGNMF	0.25	0.2143
	KRSNMF	0.47	0.5244
ORL _{32×32}	KOGNMF	0.06	0.3570
	KRSNMF	0.1325	0.5967
COIL20	KOGNMF	0.2326	0.3445
	KRSNMF	0.6118	0.7001

表 3: 拉普拉斯核聚类性能表

Dataset	Algorithm	ACC	NMI
WarpAR10P	KOGNMF	0.2385	0.1613
	KRSNMF	0.3154	0.2908
Yale _{64×64}	KOGNMF	0.1940	0.2162
	KRSNMF	0.2788	0.3402
Yale _{32×32}	KOGNMF	0.2000	0.2445
	KRSNMF	0.3758	0.4344
Yale	KOGNMF	0.2000	0.2445
	KRSNMF	0.3758	0.4344
ORL _{64×64}	KOGNMF	0.075	0.3671
	KRSNMF	0.075	0.4430
orlraws10P	KOGNMF	0.24	0.1729
	KRSNMF	0.44	0.5133
Pixraws10P	KOGNMF	0.25	0.2111
	KRSNMF	0.49	0.5464
ORL _{32×32}	KOGNMF	0.06	0.3507
	KRSNMF	0.1150	0.5647
COIL20	KOGNMF	0.1229	0.0740
	KRSNMF	0.5549	0.6429

在文献 [21] 中, Dijana 等人提出了非线性非负矩阵分解算法, 本文将此算法记为 KOGNMF. 这部分主要是 KOGNMF 算法 [21] 和 KRSNMF 算法在九个数据集上的实验结果展示及分析. 两种算法的实验结果如表 1、表 2 和表 3 所示, 它们分别展示了运用不同核技巧即高斯核技巧、幂指数核技巧和拉普拉斯核技巧的实验结果.

精度和归一化互信息是评价聚类性能好坏的常用指标. 从表 1 可以看出, 在运用高斯核技巧时, 对于不同的数据集, KRSNMF 算法计算出的 ACC 和 NMI 总是要高于 KOGNMF 的实验结果, 在 orlraws10P, Pixraws10P, ORL_{32×32} 和 COIL20 这四个数据集上的表现尤为突出. 换句话说, 对于聚类性能而言, 本文提出的 KRSNMF 算法的性能要优于已有的算法. 下面介绍幂指数核的实验结果, 幂指数核与高斯核密切相关, 只有正态的平方被忽略. 而拉普拉斯核完全等同于幂指数核, 除了对 σ 参数的变化不那么敏感. 事实上, 从高斯核、幂指数核和拉普拉斯核的计算公式来看, 三者大同小异, 都是径向基函数内核, 高斯核的 σ 的选取也同样适合幂指数核和拉普拉斯核.

从表 2 的显示结果来看, 在这九个不同的数据集上, KRSNMF 算法的性能总是优于 KOGNMF 算法, 其中在 orlraws10P, Pixraws10P, ORL_{32×32}, ORL 和 COIL20 这五个数据集上的更能体现 KRSNMF 算法的聚类性能, 这说明本文提出的 KRSNMF 算法是有效的. 最佳的实验结果体现在体现在 COIL20 数据集上, 该数据集是由 20 个不同的事物从 72 个不同的角度拍摄的照片组成的, 在此数据上, KOGNMF 算法计算出的 ACC 和 NMI 分别是 0.2326 和 0.3445, 而运用 KRSNMF 算法时的 ACC 和 NMI 分别达到了 0.6118 和 0.7001. 这充分展现了 KRSNMF 算法的有效性.

从表 3 显示的结果来看, 在九个不同的数据集上, 依然是 KRSNMF 算法的性能优于 KOGNMF 算法, 与表 2 的结果类似, 在 orlraws10P, Pixraws10P, ORL_{32×32}, ORL 和 COIL20 这五个数据集上的表现更好. 其中, 在 COIL20 上的实验结果显示, 两种算法计算出的互信息相差 0.5689 之多.

虽然三种核技巧的计算方式不同, 但以上三个表的实验结果都阐明了 KRSNMF 算法的性能更好. 从核技巧的角度而言, 选取幂函数核技巧是最有效的. 纵观三个表展示的结果来看, 本文提出的 KRSNMF 算法的性能的确要优于已有的算法. 这是因为本文在建立的目标模型中引入了对原始数据中的噪音值和异常值点有自动处理作用的 $L_{2,1}$ 范数, 并额外添加了 $L_{2,1/2}$ 矩阵伪范数作为稀疏约束, 从而使算法的稀疏性和鲁棒性得到了良好的改善.

5 结论

本文提出了一种基于核技巧的 $L_{2,1}$ 范数非负矩阵分解, 它是用 $L_{2,1}$ 范数来替代标准 NMF 中的 F 范数, 即以 $L_{2,1}$ 范数为损失函数, 并且在保留原始数据的内在流行几何结构和运用核技巧来揭示流形的非线性性质的基础上, 添加了 $L_{2,1/2}$ 矩阵伪范数作为额外的稀疏约束, 从而达到提高算法的稀疏性和鲁棒性的目的. 在九个常用的标准数据集上进行实验, 数值实验结果展示了选取幂函数核技巧是最明智的, 也验证了本文提出的 KRSNMF 算法是有效的.

参 考 文 献

- [1] Pentti Paatero, Unto Tapper. 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] Daniel D Lee, Sebastian Seung H. Learning the parts of objects by non-negative matrix factorization [J]. *Nature*, 1999, 401: 788–791.
- [3] Hans Peter Deutsch. Principle component analysis [M]. UK: Palgrave Macmillan, 2004.
- [4] Bruce A Draper, Kyungim Baek, Marian Stewart Bartlett, Ross Beveridge J. Recognizing faces with pca and ica [J]. *Computer Vision Image Understanding*, 2003, 91(1): 115–137.
- [5] Allen Gersho, Robert M Gray. Vector quantization and signal compression [M]. US: Springer, 1992.
- [6] Daniel D Lee, Sebastian Seung H. Algorithms for non-negative matrix factorization [J]. *Proc. Neural Inform. Proc. Sys.*, 2001, 13: 556–562.
- [7] Cai Deng, He Xiaofei, Han Jiawei, Huang Thomas S. Graph regularized nonnegative matrix factorization for data representation [J]. *IEEE Trans. Patt. Anal. Machine Intell.*, 2011, 33(8): 1548–1560.
- [8] TaoDapeng, Jin Lianwen, Liu Weifeng, Li Xuelong. Hessian regularized support vector machines for mobile image annotation on the cloud [J]. *IEEE Trans. Multimedia*, 2013, 15(4): 833–844.
- [9] Chen L, Xu D, Tsang I W, Li X. Spectral embedded hashing for scalable image retrieval [J]. *IEEE Trans. Cybern.*, 2014, 44(7): 1180–1190.

- [10] Biplab Banerjee, Francesca Bovolo, Avik Bhattacharya, Lorenzo Bruzzone, Subhasis Chaudhuri, Krishna Mohan B. A new self-training-based unsupervised satellite image classification technique using cluster ensemble strategy [J]. *IEEE Geosci. Remote Sensing Lett.*, 2014, 12(4): 741–745.
- [11] Liu G, Zhang Y, Wang A. Incorporating adaptive local information into fuzzy clustering for image segmentation [J]. *IEEE Trans. Image Proc. Publ. IEEE Sign. Proc. Soc.*, 2015, 24(11): 3990.
- [12] Ritendra Datta, Dhiraj Joshi, Li Jia, James Z Wang. Image retrieval: ideas, influences, and trends of the new age [J]. *Acm Comput. Surveys*, 2008, 40(2): 1–60.
- [13] Paritosh Pande, Brian E Applegate, Javier A Jo. Application of non-negative matrix factorization to multispectralflim data analysis [J]. *Biom. Optics Express*, 2012, 3(9): 2244–2262.
- [14] Qiao Jiangang. A novel probability-based clustering method [D]. Lanzhou: Lanzhou University, 2014.
- [15] Li S Z, Hou X W, Zhang H J, Cheng Q S. Learning spatially localized, parts-based representation [J]. *Computer Vision Patt. Recog.*, 2001, 1: 207–212.
- [16] Hoyer P O. Non-negative sparse coding[C]. *Neural Networks for Signal Processing, Proceedings of the 2002 IEEE Workshop on*, 2002: 557–565.
- [17] Barman P C, Nadeem Iqbal, Soo Young Lee. Non-negative matrix factorization based text mining [J]. *Feature Extract. Classi.*, 2006, 4233: 703–712.
- [18] Yang Zhirong, Yuan Zhijian, Laaksonkn Jorma. Projective non-negative matrix factorization with applications to facial image processing [J]. *Intern. J. Patt. Recog. Artif. Intell.*, 2007, 21(08): 1353–1362.
- [19] Zeng Kun, Yu Jun, Li Cuihua, You Jane, Jin Taisong. Image clustering by hyper-graph regularized non-negative matrix factorization [J]. *Neurocomputing*, 2014, 138(11): 209–217.
- [20] Li Huirong, Zhang Jiangshe, Shi Guang, Liu Junmin. Graph-based discriminative non-negative matrix factorization with label information [J]. *Neurocomputing*, 2017, (266): 91–100.
- [21] Dijana Tolic, Nino Antulovfantulin, Ivica Kopriva. A nonlinear orthogonal non-negative matrix factorization approach to subspace clustering [J]. *Pattern Recognition*, 2018, (82): 40–55.
- [22] Christopher K I Williams. Learning with kernels: support vector machines, regularization, optimization, and beyond [J]. *Publ. American Stat. Assoc.*, 2002, 98(462): 489–489.
- [23] Huang T M, Kecman V, Kopriva I. Kernel based algorithms for mining huge data sets: supervised, semi-supervised, and unsupervised learning[M]. *DBLP*, 2006.
- [24] Wan Y, Chen X, Zhang J. Global and intrinsic geometric structure embedding for unsupervised feature selection[J]. *Expert Sys. Appl.*, 2018, (93): 134–142.
- [25] Kong D, Ding C, Huang H. Robust nonnegative matrix factorization using L_{21} -norm[C]. *Proceedings of the 20th ACM International Conference on Information and Knowledge Management. ACM*, 2011: 673–682.

- [26] Ngom A, Li Y. The non-negative matrix factorization toolbox for biological data mining [J]. Source Code. Biol. Med., 2013, 8(1): 1–15.
- [27] Christos H Papadimitriou, Kenneth Steiglitz. Combinatorial optimization: algorithms and complexity [J]. IEEE Trans. Acoustics Speech Signal Proc., 1998, 32(6): 1258–1259.

KERNEL-BASED $L_{2,1}$ NORM NON-NEGATIVE MATRIX FACTORIZATION IN IMAGE CLUSTERING

YU Jiang-lan¹, LI Xiang-li¹, DONG Xiao-liang²

(1.School of Mathematics and Computing Science; Guangxi Key Laboratory of Cryptography and Information Security; Guangxi Key Laboratory of Automatic Detecting Technology and Instruments, Guilin University of Electronic Technology, Guilin 541004, China)

(2.School of Mathematics and Information Science, North Minzu Univeisity, Yinchuan 750021, China)

Abstract: The problem of norm non-negative matrix factorization with $L_{2,1}$ is studied based on kernel technique in image clustering. By kernel-based sparse robust non-negative matrix factorization method, the sparseness and robustness of the algorithm are obtained, and the clustering performance is improved. This method can also be extended to the application of text clustering.

Keywords: non-negative matrix factorization; kernel trick; $L_{2,1}$ norm; sparsity; robustness

2010 MR Subject Classification: 91C20