

# 基于局部相似性学习的鲁棒非负矩阵分解

侯兴荣, 彭冲

(青岛大学计算机科学技术学院, 青岛 266071)

**摘要:** 现有的非负矩阵分解方法往往聚焦于数据全局结构信息的学习, 在很多情况下忽略了对数据局部信息的学习, 而局部学习的方法也通常局限于流行学习, 存在一些缺陷。为解决这一问题, 提出了一种基于数据局部相似性学习的鲁棒非负矩阵分解算法 (Robust nonnegative matrix factorization with local similarity learning, RLS-NMF)。采用一种新的数据局部相似性学习方法, 它与流形方法存在显著区别, 能够同时学习数据的全局结构信息, 从而能挖掘数据类内相似和类间相离的性质。同时, 考虑到现实应用中的数据存在异常值和噪声, 该算法还使用  $l_{2,1}$  范数拟合特征残差, 过滤冗余的噪声信息, 保证了算法的鲁棒性。多个基准数据集上的实验结果显示了该算法的最优性能, 进一步证明了该算法的有效性。

**关键词:** 非负矩阵分解; 聚类; 全局结构; 局部相似性; 鲁棒性

**中图分类号:** TP391      **文献标志码:** A

## Robust Nonnegative Matrix Factorization with Local Similarity Learning

HOU Xingrong, PENG Chong

(College of Computer Science and Technology, Qingdao University, Qingdao 266071, China)

**Abstract:** The existing nonnegative matrix factorization methods mainly focus on learning global structure of the data, while ignoring the learning of local information. Meanwhile, for those methods that attempt to exploit local similarity, the manifold learning is often adopted, which suffers some issues. To solve this problem, a new method named the robust nonnegative matrix factorization with local similarity learning (RLS-NMF) is proposed. In this paper, a new local similarity learning method is adopted, which is starkly different from the widely used manifold learning. Moreover, the new method can simultaneously learn the global structural information of the data, and thus exploit the intra-class similarity and the inter-class separability of the data. To address the issues of outliers and noise effects in real word applications, the  $l_{2,1}$  norm is used to fit the residuals to filter the redundant noise information, ensuring the robustness of the algorithm. Extensive experimental results show the superior performance of the proposed method on several benchmark datasets, further demonstrating its effectiveness.

**Key words:** nonnegative matrix factorization; clustering; global structure; local similarity; robustness

## 引言

高维数据已无处不在,信息检索作为数据挖掘和文本聚类的重要任务之一<sup>[1]</sup>,通常需要处理高维数据。因此,在矩阵补全<sup>[2]</sup>、信号处理<sup>[3]</sup>和文本挖掘<sup>[4]</sup>等需要信息检索的学习任务中,寻求一种数据表示方法来对高维数据进行处理显得尤为重要。为此,研究者们提出了很多数据表示方法,其中矩阵分解作为一种高维数据的表示方法已经被广泛应用在数据挖掘、计算机视觉和机器学习等领域。

矩阵分解的原理是用两个数据矩阵的乘积来近似原始数据。基于矩阵分解的方法已经很流行<sup>[5]</sup>,除此之外还有一些其他的方法,例如主成分分析、奇异值分解、低秩表示和非负矩阵分解。潜在语义分析(Latent semantic index, LSI)正是利用了奇异值分解的原理,它选取最大的 $k$ 个奇异值,一些小奇异值所对应的信息就会被消除,从而过滤掉不必要的信息,提高文本挖掘任务(文本检索、聚类等)的性能。但考虑到奇异值分解(Singular value decomposition, SVD)计算量大和解释混合符号数据困难的缺点,非负矩阵分解的方法<sup>[6]</sup>又被提出。对于图像和文档等非负数据,非负矩阵分解被用来寻找两个非负的因子矩阵近似此类数据,这也导致了基于部分的表达。而且有心理学和生理学的证据表明<sup>[7]</sup>,人脑中存在基于部分的表征。非负矩阵分解就是依赖于这种表征来学习部分的人脸和文本的语义特征。

现有的非负矩阵分解算法可以分为4类,包括基本非负矩阵分解(Nonnegative matrix factorization, NMF)、约束非负矩阵分解<sup>[8]</sup>、结构化非负矩阵分解和广义非负矩阵分解。文献[9]中有相当全面的综述,其中正交非负矩阵分解(Orthogonal NMF, ONMF)<sup>[10]</sup>对系数矩阵施加了正交约束,加强了聚类解释;加权非负矩阵分解(Weighted NMF, WNMF)<sup>[11]</sup>对不同元素的相对重要性赋予不同的权重;半非负矩阵分解(Semi-NMF)<sup>[12]</sup>仅对系数矩阵施加非负性约束。上面提出的几种非负矩阵分解方法仅仅学习了线性分布的数据,为了学习非线性分布的数据又提出了如下几种方法:流形上的非负矩阵分解<sup>[13]</sup>考虑了非线性分布的数据。核非负矩阵分解<sup>[14]</sup>将原始模型扩展到了希尔伯特空间,使模型能够处理非线性可分的数据。文献[15]中已经证明,在无监督学习中保留局部几何数据结构很重要,主要体现在:(1)将高维数据转化为低维空间数据时,保留其内部信息非常重要;(2)全局结构不足以充分体现数据的特征信息;(3)数据的局部几何结构可以看做是变换矩阵的正则化,从而避免过拟合。目前已有的考虑数据局部相似性信息的方法<sup>[13]</sup>通常需要借助流形技术,通过样本的最相似近邻来考虑数据的局部几何结构,而这种方法通常强调低维表达的类内相似性而忽视类间相异性。而RLS-NMF通过成对数据样本间的相似性来学习数据的局部结构信息,该方法使用数据的全局和局部相似性信息来学习同时具有类内相似性和类间相异性的低维表示矩阵。此外,原始的非负矩阵分解算法通过优化误差的平方来求解目标函数,这就导致了最后的结果容易受异常值的影响,RLS-NMF算法在原始非负矩阵分解的基础上使用了 $l_{2,1}$ 范数替代Frobenius范数,保证了方法的鲁棒性<sup>[16]</sup>。该方法的主要贡献有:(1)通过成对数据间的相似性来学习数据的局部几何结构,使学习到的系数矩阵能揭示原始数据的更多结构信息;(2)将 $l_{2,1}$ 范数和局部相似性的学习集成到一个框架中,在学习数据局部结构的同时还保证了模型的鲁棒性;(3)将正交约束与局部相似性的学习结合,使局部相似性和聚类相互增强。

## 1 提出方法

给定非负数据 $X = [x_1, x_2, \dots, x_n] \in \mathbb{R}^{m \times n}$ , NMF将原始数据 $X$ 分解为两个非负矩阵 $H \in \mathbb{R}^{m \times k}$ 和 $G \in \mathbb{R}^{n \times k}$ 的乘积,优化问题如下

$$\min_{H \geq 0, G \geq 0} \|X - HG^T\|_F^2 \quad (1)$$

式中:原始数据 $X$ 的每一列分别对应一组 $H$ 矩阵列向量的加权和,每一组权重系数对应 $G$ 矩阵的每一行,因此 $H$ 被称为基矩阵, $G$ 被称为系数矩阵, $m$ 表示数据维数, $n$ 表示样本大小, $k \ll n$ 保证了用一个低

维系数矩阵来表示原始数据矩阵。凸非负矩阵分解(Convex NMF, CNMF)在NMF的基础上将基矩阵限制为数据 $X$ 的每一列的凸组合,即 $H=XW$ 。目标函数式(1)就转化为对以下问题的优化

$$\min_{W, G \geq 0} \|X - XWG^T\|_F^2 \quad (2)$$

CNMF将 $H$ 的列解释为数据点的加权和,而 $W_{ij}$ 就是 $X_i$ 对于 $H_j$ 的权重。现实应用中的大部分数据,尤其是高维数据,都存在一些冗余的特征。受特征选择的影响,对于冗余的特征, $X$ 的近似值不需要很好的拟合这些特征。因此,RLS-NMF用近似残差的行稀疏性来放松对这些特征的拟合,用 $l_{2,1}$ 范数替换Frobenius范数,得到如下正则化公式

$$\min_{W, G \geq 0} \|X - XWG^T\|_{2,1} = \sum_{i=1}^m \left\| [X - XWG^T]_{(i)} \right\|_2 \quad (3)$$

式中: $\|\cdot\|_2$ 为 $l_2$ 范数, $A_{(i)}$ 表示矩阵 $A$ 的第 $i$ 行, $\|A\|_{2,1}$ 是对矩阵 $A$ 每一行的 $l_2$ 范数求和。 $l_{2,1}$ 范数作为一种行稀疏范数,还可用于解释特征噪声和异常值,从而近似的忽略异常或噪声特征造成的损失。在优化问题(3)中,总和最小就要求每一行的 $l_2$ 范数最小,即每一个行向量中的零元素尽可能多,从而实现行稀疏,提取出重要特征。

值得注意的是,式(3)与子空间聚类密切相关<sup>[17]</sup>,而在子空间聚类中通常将高维数据投影到低维子空间中,要寻找这样的子空间需要建立自我表达假设。这个自我表达假设指的是 $X=XZ$ ,其中 $Z \in \mathbb{R}^{n \times n}$ 是一个表示矩阵。为了考虑数据的局部结构,一个自然的假设就是,如果两个数据 $x_i$ 和 $x_j$ 是相近的,那么它们之间的相似性 $Z_{ij}$ 应该大,否则, $Z_{ij}$ 应该小。该假设会导致如下的最优化问题

$$\min_Z \sum_{ij} \|x_i - x_j\|_2^2 Z_{ij} \Leftrightarrow \min_Z \text{tr}(Z^T D) \quad (4)$$

式中: $D \in \mathbb{R}^{n \times n}$ , $D$ 的每一个元素 $D_{ij}$ 可以表示为 $D_{ij} = \|x_i - x_j\|_2^2$ ,或者写成矩阵的形式 $D = I_n I_n^T \text{diag}(X^T X) + \text{diag}(X^T X) I_n I_n^T - 2X^T X$ ,其中 $I_n$ 表示长度为 $n$ 的全1列向量。

式(4)的优化强制 $Z_{ij}$ 反映数据的两两相似度信息。注意到 $W$ 和 $G$ 非负,受自我表达的启发,可以取 $WG^T$ 作为局部相似性矩阵 $Z$ ,即 $Z_{ij} = (WG^T)_{ij} = W_{(i)} G_{(j)}^T$ 。 $W_{(i)}$ 可以看成数据 $x_i$ 在基向量上的得分向量, $G_{(j)}$ 可以看成第 $j$ 个样本相对于新基的系数向量。如果数据 $x_i$ 和 $x_j$ 是相近的或者属于同一个聚类,那么 $W_{(i)}$ 和 $G_{(j)}$ 也具有较高的相似性,反之亦然,所以 $Z = WG^T$ 是有意义的。然后将式(4)加入式(3),得到

$$\min_{W, G \geq 0} \|X - XWG^T\|_{2,1} + \lambda \text{tr}(Z^T D) \quad (5)$$

对于式(5),如果采用增广拉格朗日的优化方法,每次迭代都需要求解SVD,这是低效的,所以参考Peng等<sup>[18]</sup>提出的引入松弛变量 $E$ 的方法,式(5)可以进一步写为

$$\min_{W, G, E} \frac{1}{2} \|X - XWG^T - E\|_F^2 + \gamma \|E\|_{2,1} + \lambda \text{tr}(W^T D G) \quad \text{s.t. } W \geq 0, G \geq 0, G^T G = I \quad (6)$$

式中: $\lambda, \gamma$ 为平衡参数,目标函数(6)的第1项学习了数据的全局结构,第3项则是利用数据的局部结构与成对相似性之间的关系来考虑数据的局部相似性。CNMF作为一种通过流形结构来学习局部相似性的方法,对于相似度高的样本,它通过图拉普拉斯来最小化这些样本对应的表示矩阵间的差距,这就提高了表示矩阵的类内相似度。图拉普拉斯保证了数据和系数向量之间的平滑度。与图拉普拉斯不同,RLS-NMF的第3项加强了数据和得分向量以及系数向量之间的平滑度,通过得分向量和系数向量来考虑数据的局部结构。此外,施加 $G$ 的正交性约束,不仅保证了解的唯一性,还直接将该方法与聚类

联系起来。受 Wang 等<sup>[19]</sup>的启发,区别于上文中直接使用  $X$  构造  $D$ ,在本文中  $D$  基于  $X - E$  进行构造,并在优化过程中随着  $E$  的更新来进行调整。

## 2 优化求解

### 2.1 优化 $W$

固定  $G$ ,关于  $W$  的子问题为

$$\min_{W \geq 0} \frac{1}{2} \text{tr}(-2W^T X^T (X - E)G + W^T X^T X W G^T G) + \lambda \text{tr}(W^T D G) \quad (7)$$

根据不等式约束  $W \geq 0$ ,引进拉格朗日乘子  $\Psi = [\varphi_{ij}]$ ,得到拉格朗日函数

$$\ell_W = \frac{1}{2} \text{tr}(-2W^T X^T (X - E)G + W^T X^T X W G^T G) + \lambda \text{tr}(W^T D G) + \text{tr}(\Psi W^T) \quad (8)$$

对  $\ell_W$  求偏导

$$\frac{\partial \ell_W}{\partial W} = -X^T (X - E)G + X^T X W G^T G + \lambda D G + \Psi \quad (9)$$

根据互补松弛条件得到

$$(-X^T (X - E)G + X^T X W G^T G + \lambda D G)_{ik} W_{ik} = \varphi_{ik} W_{ik} = 0 \quad (10)$$

式(10)提供了极限解满足的固定点条件,并且式(10)可以进一步写为

$$(-X^T (X - E)G + X^T X W G^T G + \lambda D G)_{ik} W_{ik}^2 = 0 \quad (11)$$

从而得到  $W$  的更新规则

$$W_{ik} \leftarrow W_{ik} \sqrt{\frac{(X^T (X - E)G)_{ik}}{(X^T X W G^T G + \lambda D G)_{ik}}} \quad (12)$$

因为式(10)和式(11)的解都是  $W_{ik} = 0$  或者  $(-X^T (X - E)G + X^T X W G^T G + \lambda D G)_{ik} = 0$ ,所以式(10)和式(11)是等价的,证明了  $W$  更新规则的合理性。用辅助函数的方法<sup>[20]</sup>还可以证明更新规则(12)的收敛性和合理性,附录 A 给出了相关的定义、引理和证明过程。

### 2.2 优化 $G$

固定  $W$ ,得到关于  $G$  的最优化问题为

$$\min_{G \geq 0} \frac{1}{2} \text{tr}(-2W^T X^T (X - E)G + W^T X^T X W G^T G) + \lambda \text{tr}(W^T D G) \quad (13)$$

引入大小为  $k \times k$  的对称拉格朗日乘子  $\Theta$ ,得到拉格朗日函数

$$\begin{aligned} \ell_G = & \frac{1}{2} \text{tr}(-2(X - E)^T X W G^T + G W^T X^T X W G^T) + \lambda \text{tr}(W^T D G) + \frac{1}{2} \text{tr}(\Theta (G^T G - \Lambda)) = \\ & \frac{1}{2} \text{tr}(-2(X - E)^T X W G^T + 2\lambda D W G^T) + \frac{1}{2} \text{tr}(G W^T X^T X W G^T + G \Theta G^T) - \xi = \\ & \frac{1}{2} \text{tr}(-2(X - E)^T X W G^T + 2\lambda D W G^T) + \frac{1}{2} \text{tr}(G (W^T X^T X W + \Theta)^+ G^T - \\ & G (W^T X^T X W + \Theta)^- G^T) - \xi \end{aligned} \quad (14)$$

式中:定义  $\xi = \frac{1}{2} \text{tr}(\Theta \Lambda)$ ,  $\Lambda$  为非负对角矩阵,保证了  $G$  的正交性<sup>[10]</sup>,  $M^+$  和  $M^-$  是两个非负矩阵且满足  $M^+ - M^- = M$ 。对  $\ell_G$  求导可得

$$\frac{\partial \ell_G}{\partial G} = -(X - E)^T XW + GW^T X^T XW + \lambda DW + G\Theta \quad (15)$$

根据互补松弛条件可得

$$\left( -(X - E)^T XW + GW^T X^T XW + \lambda DW + G\Theta \right)_{ik} G_{ik} = 0 \quad (16)$$

式(16)等价于

$$\left( -(X - E)^T XW + GW^T X^T XW + \lambda DW + G\Theta \right)_{ik} G_{ik}^2 = 0 \quad (17)$$

又因为  $W^T X^T XW + \Theta = (W^T X^T XW + \Theta)^+ - (W^T X^T XW + \Theta)^-$ , 最终得到更新规则为

$$G_{ik} \leftarrow G_{ik} \sqrt{\frac{\left( (X - E)^T XW + \lambda GG^T DW \right)_{ik}}{\left( \lambda DW + GG^T (X - E)^T XW \right)_{ik}}} \quad (18)$$

同样可以用辅助函数的方法证明目标函数(13)在更新规则(18)下的收敛性,证明过程展示在附录B中,结果也进一步验证了收敛解的合理性。

### 2.3 优化 $E$

当  $W$  和  $G$  固定时,类似于文献[18],关于  $E$  的优化可以改写为

$$\min_E \frac{1}{2} \|X - XWG^T - E\|_F^2 + \gamma \|E\|_{2,1} \quad (19)$$

解决式(19)需要引入一个定理。

**定理 1** 对给定矩阵  $W = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n] \in \mathbb{R}^{m \times n}$  和一个正标量参数  $\lambda$ ,  $X^*$  是如下公式的最优解:

$\min_X \frac{1}{2} \|W - X\|_F^2 + \lambda \|X\|_{2,1}$ , 那么  $X^*$  的每一列可以表示为

$$X_i^* = \begin{cases} \frac{\|\mathbf{w}_i\|_2 - \lambda}{\|\mathbf{w}_i\|_2} \mathbf{w}_i & \|\mathbf{w}_i\|_2 > \lambda \\ 0 & \text{其他} \end{cases} \quad (20)$$

根据定理1,得到式(19)的更新规则为

$$E_{(i)} = \begin{cases} \frac{\|Q_{(i)}\|_2 - \gamma}{\|Q_{(i)}\|_2} Q_{(i)} & \|Q_{(i)}\|_2 > \gamma \\ 0 & \text{其他} \end{cases} \quad (21)$$

式中  $Q = X - XWG^T$ 。

在更新规则(12)和(18)下,对于固定的  $E^{(t)}$ ,关于  $W$  和  $G$  的子问题  $f(W, G|E)$  在变量序列  $\{W^{(t+1,s)}, G^{(t+1,s)}|E^{(t)}\}$  下可以得到  $f(W^{(t+1,s)}, G^{(t+1,s)}|E^{(t)}) \geq f(W^{(t+1,s+1)}, G^{(t+1,s+1)}|E^{(t)})$ , 即序列  $f(W^{(t+1,s)}, G^{(t+1,s)}|E^{(t)})$  是单调递减且有界的。

**算法 1** 基于局部相似性学习的鲁棒非负矩阵分解算法

输入:参数  $\lambda, \gamma$ , 最大迭代次数  $t_{\max}$  和  $s_{\max}$ , 收敛条件  $\epsilon$

- (1) 初始化:  $W^{(0)}, G^{(0)}, E^{(0)}, t = 1$ 。
- (2) Repeat(外循环)
- (3)  $W^{(t,0)} = W^{(t-1)}, G^{(t,0)} = G^{(t-1)}, s = 1$

- (4) Repeat(内循环)
  - (5) 根据式(12)更新  $\mathbf{W}^{(t,s)}$ , 根据式(18)更新  $\mathbf{G}^{(t,s)}$ ;
  - (6) 检查收敛条件:  $s \geq s_{\max}$  或者两次迭代的目标函数值之差小于  $\epsilon$ ;
  - (7)  $s = s + 1$ ;
  - (8) End(内循环)
  - (9) 更新  $\mathbf{W}^{(t)} = \mathbf{W}^{(t,s)}$ ,  $\mathbf{G}^{(t)} = \mathbf{G}^{(t,s)}$ ,  $\mathbf{E}^{(t)}$ ;
  - (10) 检查收敛条件:  $t \geq t_{\max}$  或者两次迭代的目标函数值之差小于  $\epsilon$ ;
  - (11)  $t = t + 1$ ;
  - (12) End(外循环)
- 输出:  $\mathbf{W}^* = \mathbf{W}^{(t)}$ ,  $\mathbf{G}^* = \mathbf{G}^{(t)}$ ,  $\mathbf{E}^* = \mathbf{E}^{(t)}$ 。

### 3 基于部分的表示

**定理 2** 当  $\mathbf{W}$  和  $\mathbf{G}$  初始化为非负矩阵,  $\mathbf{E}$  初始化为零矩阵, 矩阵序列  $\{\mathbf{W}^{(t)}, \mathbf{G}^{(t)}, (\mathbf{X} - \mathbf{E})^{(t)}\}$  在更新规则(12, 18, 21)下是非负的, 因此, RLS-NMF 能够学习基于部分的表示。

对于非负矩阵分解, 需要限制基和系数矩阵为非负来获得基于部分的表示。定理 2 说明了在更新规则(12, 18, 21)下目标函数(6)确保能学习基于部分的表示, 附录 C 中给出了定理 2 的证明过程。

## 4 实验部分

本节将展示大量的实验结果来说明 RLS-NMF 的性能, 并使用准确率和纯度作为评估指标<sup>[21]</sup>, 另外还详细介绍了数据集、对比方法、聚类性能以及其他性能<sup>[22]</sup>。

### 4.1 基准数据集

实验中分别使用了 Jaffe、Semeion、PIX、COIL20 和 COIL100 数据集。其中, Jaffe 数据集共包含 213 张图像, 由 10 个日本女学生作出的 7 种面部表情的照片组成, 每张图片的大小为  $26 \times 26$ ; PIX 数据集包含来自 10 个物体的 100 张大小为  $100 \times 100$  的灰度图像; Semeion 数据集收集了由 80 个人手写的数字 0—9 的 1 593 张图片, 每张图片被重构为  $16 \times 16$  大小; COIL20 数据集包含了 20 个物体分别水平旋转  $360^\circ$ , 每隔  $5^\circ$  拍摄的一张照片, 每个物体旋转  $360^\circ$  后拍摄 72 张照片, 所以数据集一共包含了 1 440 张大小为  $128 \times 128$  的图像; 大数据物品图像数据集 COIL100 是 COIL20 数据集的扩展, 包含 100 张物体的照片。

### 4.2 比较算法

为了体现 RLS-NMF 的性能, 将本文方法与几种比较先进的非负矩阵分解方法进行了比较。NMF<sup>[6]</sup>对基矩阵和系数矩阵施加了非负约束, 采用标准的乘法更新规则进行求解; Semi-NMF<sup>[12]</sup>只对系数矩阵施加了非负约束, 比 NMF 更通用; WNMF<sup>[11]</sup>是为了处理不完备的数据提出的一种方法, 它在原始非负矩阵分解的基础上增加了权重矩阵; ONMF<sup>[10]</sup>除了对分解的因子矩阵施加了非负约束外还对系数矩阵施加了正交约束; CNMF<sup>[12]</sup>将基矩阵限制为数据列的凸组合, 使得基矩阵可以用某些数据点的加权和表示; 基于核的图正则化正交非负矩阵分解 (Graph regularized kernel-based orthogonal NMF, KOGNMF)<sup>[23]</sup>在正交非负矩阵分解的基础上增加了图正则化项, 考虑了非线性分布的数据; 截断柯西非负矩阵分解 (Truncated Cauchy NMF, TCNMF)<sup>[23]</sup>通过截断大的误差来处理异常值, 提高了非负矩阵分解算法的鲁棒性。RLS-NMF 相对于以上这些方法增加了对局部结构的学习, 参数  $\lambda$  和  $\gamma$  的选择范围均设置为  $\{0.001, 0.01, 0.1, 1, 10, 100, 1\ 000\}$ 。

### 4.3 聚类性能分析

本小节通过在 Jaffe、Semeion、PIX、COIL20 和 COIL100 数据集上进行实验来比较算法。为了更直观地说明这些数据集,图 1 展示了 Jaffe、Semeion、PIX 和 COIL20 数据集的部分示例。

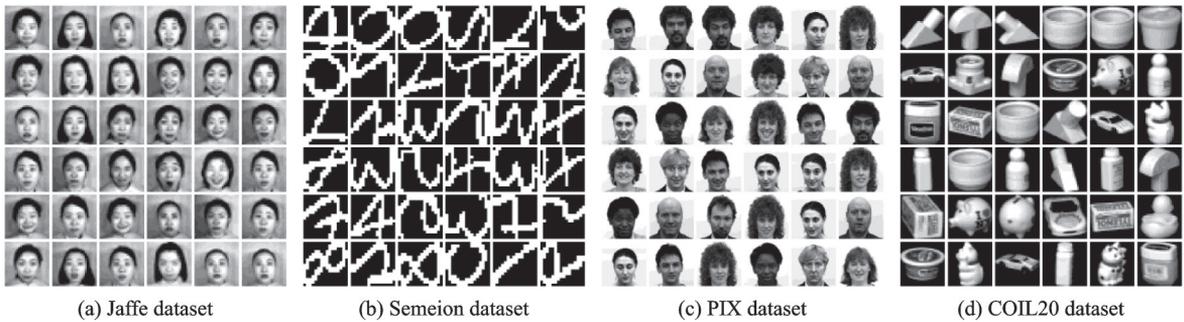


图 1 Jaffe、Semeion、PIX 和 COIL20 数据集上的示例

Fig.1 Examples selected from Jaffe, Semeion, PIX and COIL20 datasets

实验通过聚类性能来评估算法,对于每种方法均使用 K-means 进行聚类。实验设置了 200 次迭代并分别在 Jaffe、Semeion、PIX 和 COIL20 数据集上随机选择了  $\bar{N}$  类来进行实验。从有  $N$  类的数据集选取  $\bar{N}$  类,一共有  $\bar{N}! / ((\bar{N} - N)! N!)$  种选择。从中随机选取 10 组计算平均值,最后选出所有参数组合下的最好的一组平均值作为最终结果,其他方法的实验结果也以同样方法得出。在 Jaffe、Semeion、PIX 和 COIL20 数据集上的实验结果分别展示在表 1~4 中。表 1~4 中每个表又分为两个子表来分别记录两个评估指标的最优值。从结果可以看出,在所有的数据集中,RLS-NMF 几乎在所有的子集中具有

表 1 各个方法在 Jaffe 数据集上的聚类性能

Table 1 Clustering performance of methods on Jaffe dataset

N	Accuracy								%
	NMF	CNMF	ONMF	WNMF	Semi-NMF	TCNMF	KOGNMF	RLS-NMF	
2	99.75	99.75	94.60	99.75	99.75	99.75	100	100	
4	95.92	80.37	88.58	95.92	95.32	95.44	99.07	99.07	
6	91.41	76.26	88.35	89.47	91.01	88.29	97.80	97.80	
8	90.75	69.44	69.62	92.05	<b>94.17</b>	91.46	82.22	91.30	
10	92.90	69.95	80.28	90.61	91.55	92.96	86.85	<b>96.24</b>	
Average	94.15	79.15	84.29	93.56	94.36	93.58	93.19	<b>96.88</b>	
N	Pur								%
	NMF	CNMF	ONMF	WNMF	Semi-NMF	TCNMF	KOGNMF	RLS-NMF	
2	99.75	99.75	94.60	99.75	99.75	99.75	100	100	
4	95.92	83.29	91.64	95.92	95.32	95.44	99.07	99.07	
6	91.41	78.30	89.94	89.47	91.01	88.29	97.80	97.80	
8	90.75	70.44	74.63	92.23	<b>94.52</b>	91.46	85.69	92.12	
10	93.90	74.18	81.22	90.61	91.55	92.96	87.79	<b>96.24</b>	
Average	94.35	81.19	86.41	93.60	94.43	93.58	94.07	<b>97.05</b>	

最好的表现。并且在每个方法所有子集的平均性能中 RLS-NMF 也基本都是最优的,在准确率和纯度方面分别比次优的方法提高 1%~2% 和 1%~3%。这些实验结果证明  $l_{2,1}$  范数和局部相似性的学习集成在一个非负矩阵分解模型中,使非负矩阵分解更具竞争力和稳定性。

表 2 各个方法在 Semeion 数据集上的聚类性能

Table 2 Clustering performance of methods on Semeion dataset

%

N	Accuracy							
	NMF	CNMF	ONMF	WNMF	Semi-NMF	TCNMF	KOGNMF	RLS-NMF
2	87.60	88.18	87.10	87.57	87.44	88.27	<b>88.86</b>	<b>88.86</b>
4	72.40	70.32	68.48	71.95	72.44	67.67	<b>75.32</b>	<b>75.32</b>
6	61.55	57.50	60.64	58.25	64.38	56.75	64.40	<b>65.47</b>
8	59.52	53.52	55.63	59.63	63.38	55.14	63.47	<b>65.11</b>
10	51.41	45.20	50.78	55.56	59.01	52.98	48.34	<b>61.96</b>
Average	66.50	62.94	64.53	66.59	69.33	64.16	68.08	<b>71.34</b>
N	Pur							
	NMF	CNMF	ONMF	WNMF	Semi-NMF	TCNMF	KOGNMF	RLS-NMF
2	87.60	88.18	87.10	87.57	87.44	88.27	<b>88.86</b>	<b>88.86</b>
4	72.81	71.91	70.28	72.33	72.69	69.32	<b>76.51</b>	<b>76.51</b>
6	63.53	61.03	62.96	60.91	65.30	60.37	66.89	<b>67.77</b>
8	61.05	55.72	57.75	60.58	64.61	57.22	65.40	<b>67.21</b>
10	54.11	45.20	52.92	56.56	60.95	55.37	50.66	<b>61.96</b>
Average	67.82	64.41	66.20	67.59	70.20	66.11	69.66	<b>72.46</b>

表 3 各个方法在 PIX 数据集上的聚类性能

Table 3 Clustering performance of methods on PIX dataset

%

N	Accuracy							
	NMF	CNMF	ONMF	WNMF	Semi-NMF	TCNMF	KOGNMF	RLS-NMF
2	94.00	<b>96.50</b>	89.00	94.00	94.00	94.00	94.50	95.00
4	95.00	88.00	85.25	92.75	93.75	92.50	<b>96.25</b>	<b>96.25</b>
6	88.17	77.50	77.67	85.00	86.50	86.33	89.17	<b>90.50</b>
8	82.62	78.50	80.88	80.88	82.37	87.13	85.00	<b>88.13</b>
10	82.00	80.00	69.00	74.00	<b>88.00</b>	83.00	78.00	83.00
Average	88.36	84.10	80.36	85.33	88.92	88.59	88.58	<b>90.58</b>
N	Pur							
	NMF	CNMF	ONMF	WNMF	Semi-NMF	TCNMF	KOGNMF	RLS-NMF
2	94.00	<b>96.50</b>	89.00	94.00	94.00	94.00	94.50	95.00
4	95.00	89.00	87.00	92.75	93.75	92.50	<b>96.25</b>	<b>96.25</b>
6	89.50	79.00	81.50	87.33	88.50	88.83	90.50	<b>91.17</b>
8	85.38	81.12	81.75	83.88	84.63	88.38	85.38	<b>89.13</b>
10	85.00	83.00	76.00	79.00	<b>88.00</b>	86.00	80.00	87.00
Average	89.78	85.72	83.05	87.39	89.78	89.94	89.33	<b>91.71</b>

表4 各个方法在 COIL20数据集上的聚类性能  
**Table 4 Clustering performance of methods on COIL20 dataset**

N	Accuracy							
	NMF	CNMF	ONMF	WNMF	Semi-NMF	TCNMF	KOGNMF	RLS-NMF
4	74.72	73.19	63.85	72.36	70.73	71.67	74.13	<b>76.88</b>
8	70.45	66.70	67.34	66.82	66.35	67.60	72.47	<b>75.61</b>
12	65.80	58.16	61.19	67.07	66.25	<b>70.68</b>	65.98	70.00
16	<b>68.00</b>	57.83	58.43	63.77	66.56	65.61	63.88	65.30
20	63.96	52.64	57.01	66.88	60.63	<b>69.17</b>	51.81	59.17
Average	68.59	61.70	61.56	67.38	66.10	68.95	65.65	<b>69.39</b>

N	Pur							
	NMF	CNMF	ONMF	WNMF	Semi-NMF	TCNMF	KOGNMF	RLS-NMF
4	74.97	73.23	66.94	72.74	71.94	72.40	75.59	<b>78.51</b>
8	72.80	69.11	70.02	69.93	69.25	71.02	74.57	<b>77.29</b>
12	67.91	60.51	65.16	68.96	68.22	71.76	68.04	<b>71.77</b>
16	<b>68.85</b>	60.10	61.96	65.52	68.37	67.32	66.26	67.44
20	65.56	56.46	59.38	67.99	61.81	<b>70.35</b>	54.93	61.60
Average	70.02	63.88	64.69	69.03	67.92	70.57	67.88	<b>71.32</b>

在 Jaffe 数据集上,RLS-NMF 在绝大部分子集中表现最优,尤其相对于比较有竞争力的 KOGNMF 方法,在  $N=8$  和  $N=10$  时,RLS-NMF 在准确率方面比 KOGNMF 都提高了 9%。在准确率和纯度方面,RLS-NMF 的平均性能分别比次优方法提高 2% 和 3%。相对于具有鲁棒性的方法 TCNMF,RLS-NMF 的准确率的平均性能比 TCNMF 提高 3%。

在 Semeion 数据集上,RLS-NMF 在所有子集上都达到了最优的表现。在小的子集上,KOGNMF 的性能和 RLS-NMF 的性能表现相同,TCNMF 的表现也仅次于 RLS-NMF,但在比较大的子集上,TCNMF 和 KOGNMF 的准确率明显低于 RLS-NMF。例如,在  $N=2$  时,RLS-NMF 在准确率上仅比 TCNMF 高不到 1%,但在  $N=6$  时,RLS-NMF 在准确率上比 TCNMF 提高 8%,在  $N=10$  时,RLS-NMF 比 KOGNMF 提高 13%。RLS-NMF 在除了  $N=2$  的所有子集上都比 TCNMF 提高 5% 以上。

在 PIX 数据集上,CNMF 在  $N=2$  时有最优的表现,但随着  $N$  增大,RLS-NMF 具有较高的性能,CNMF 的性能反而降低了。在小的子集上,KOGNMF 和 NMF 都与 RLS-NMF 性能相近,但随着子集变大,这两种方法远不如 RLS-NMF。在准确率和纯度方面,RLS-NMF 在所有子集上的平均性能都比 NMF 提高 2%。RLS-NMF 虽然没有在所有子集上取得最优表现,但在 12 种情况的 8 种里取得了最优表现,而且在所有子集的平均性能上,RLS-NMF 的表现依然是最优的。

在 COIL20 数据集上,NMF 和 TCNMF 在个别子集上达到了最优表现,但它们在数据集的所有子集上的平均性能依然比 RLS-NMF 低。而且在 COIL20 数据集的 12 个统计结果中,RLS-NMF 在 7 个统计结果中达到了最优表现,NMF 和 TCNMF 达到最优表现的情况远少于 RLS-NMF。TCNMF 作为比较有竞争力的方法,虽然在 COIL20 数据集上的表现与 RLS-NMF 的表现相当,但是它在 Jaffe、Semeion 和 PIX 数据集上的表现远不如 RLS-NMF。这也说明了 RLS-NMF 方法在多个数据集上的稳定性。

为了说明 RLS-NMF 在大规模数据集上同样有效,实验引进 COIL100 数据集,关于该数据集的部分示例如图 2 所示。

COIL100数据集相对于前面4个数据集,最大的子集数达到了100。在COIL100数据集上进行同样的实验,结果如表5所示。可以看出,RLS-NMF在所有比较方法中几乎都达到了最优表现,在准确率和纯度方面都能较次优的方法提高1%。在COIL100数据集的12个统计结果中,RLS-NMF在9种情况下达到了最优表现,而NMF和KOGNMF



图2 COIL100数据集上的示例

Fig.2 Examples selected from COIL100 dataset

仅分别在1个和2个情况下达到了最好的结果。由观察可发现,在一些小规模数据集上表现好的方法在COIL100数据集上表现不一定好。例如,KOGNMF在Semeion和PIX数据集上的性能仅次于RLS-NMF,在个别子集上甚至与RLS-NMF表现相当,而在COIL100数据集的大分子集上表现不如RLS-NMF。这说明局部几何结构的学习在大规模数据集上也能提高聚类性能, $l_{2,1}$ 范数的使用也使RLS-NMF更具稳定性,进一步证明了RLS-NMF的有效性。

表5 各个方法在COIL100数据集上的聚类性能

Table 5 Clustering performance of methods on COIL100 dataset

%

N	Accuracy							
	NMF	CNMF	ONMF	WNMF	Semi-NMF	TCNMF	KOGNMF	RLS-NMF
20	59.96	55.21	59.21	61.33	58.33	60.58	61.54	<b>62.13</b>
40	54.31	43.23	55.10	52.71	48.98	55.75	55.38	<b>56.17</b>
60	<b>52.33</b>	37.08	50.54	49.57	43.61	51.24	50.86	52.03
80	45.80	32.89	48.26	44.92	37.21	44.85	47.58	<b>48.76</b>
100	43.58	35.00	45.33	46.67	33.25	45.17	<b>48.08</b>	46.92
Average	51.20	40.68	51.69	51.04	44.28	51.52	52.69	<b>53.20</b>
N	Pur							
	NMF	CNMF	ONMF	WNMF	Semi-NMF	TCNMF	KOGNMF	RLS-NMF
20	63.38	58.21	62.33	65.00	61.54	63.75	64.21	<b>65.25</b>
40	58.44	46.40	59.23	57.10	53.31	59.42	58.67	<b>60.25</b>
60	55.61	40.74	54.57	53.47	47.40	55.01	54.51	<b>55.86</b>
80	49.47	36.34	52.04	48.65	40.55	49.21	50.84	<b>52.70</b>
100	47.00	38.33	50.33	48.42	35.83	48.42	<b>51.25</b>	50.92
Average	54.78	44.00	55.70	54.53	47.73	55.16	55.90	<b>57.00</b>

总之,RLS-NMF在所有数据集的绝大部分子集中表现都是最好的。像TCNMF和KOGNMF等比较有竞争力的方法,并没有在所有的数据集中表现突出,而且在每个数据集的大多数情况下,RLS-NMF的表现也都是最优的,实验结果进一步证实了其聚类性能的高效性。在固定参数 $\lambda=0.001, \gamma=0.001$ 的条件下,RLS-NMF在Semeion数据集上3个类和COIL20数据集上6个类的聚类效果如图3所示。

#### 4.4 收敛性分析

本节通过实验说明了RLS-NMF的收敛性,实验选取了每个数据集的全部数据,并固定参数 $\lambda=0.001, \gamma=0.001$ 。实验分别观察了内部循环目标函数值 $f(W^{(t,s)}, G^{(t,s)})E^{(t-1)}$ 、外部循环目标函数值

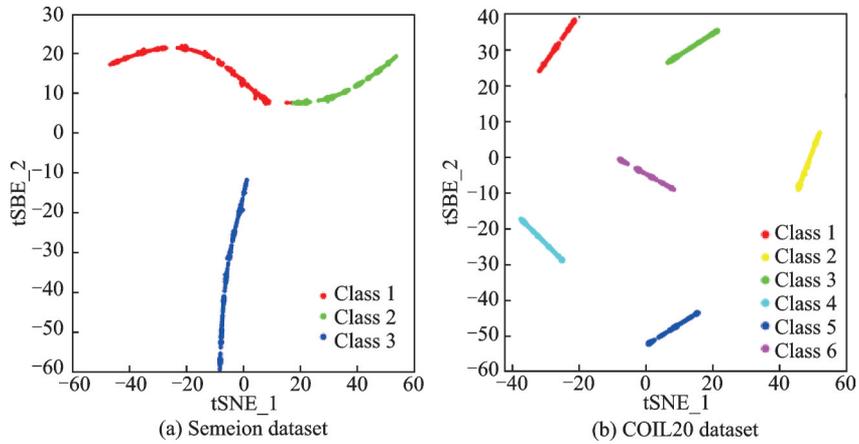


图3 RLS-NMF在Semeion和COIL20数据集上的聚类效果图

Fig.3 Clustering results of RLS-NMF on Semeion and COIL20 datasets

$\mathcal{F}(W^{(t)}, G^{(t)}, E^{(t)})$ 、序列  $\{W^{(t)}\}$  和序列  $\{G^{(t)}\}$  的收敛情况, Semeion 和 COIL20 数据集上的实验结果如图 4~6 所示, 其中, 图 4 和图 5 分别为内部循环目标函数值和外部循环目标函数值随迭代次数的变化。

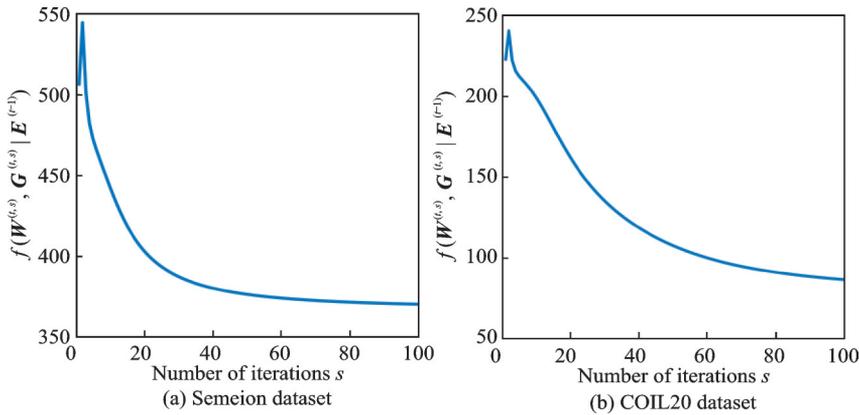


图4 Semeion 和 COIL20 数据集上的目标函数值在内部迭代中的收敛性

Fig.4 Convergence of objective function values on Semeion and COIL20 datasets in inner iteration

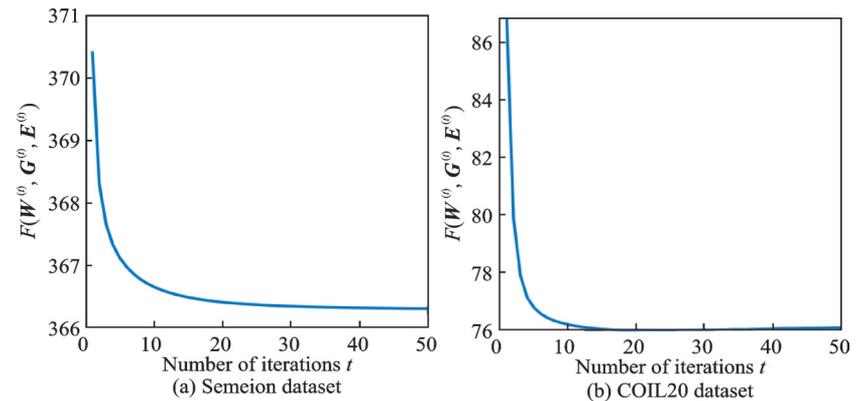


图5 Semeion 和 COIL20 数据集上的目标函数值在外部迭代中的收敛性

Fig.5 Convergence of objective function values on Semeion and COIL20 datasets in outer iteration

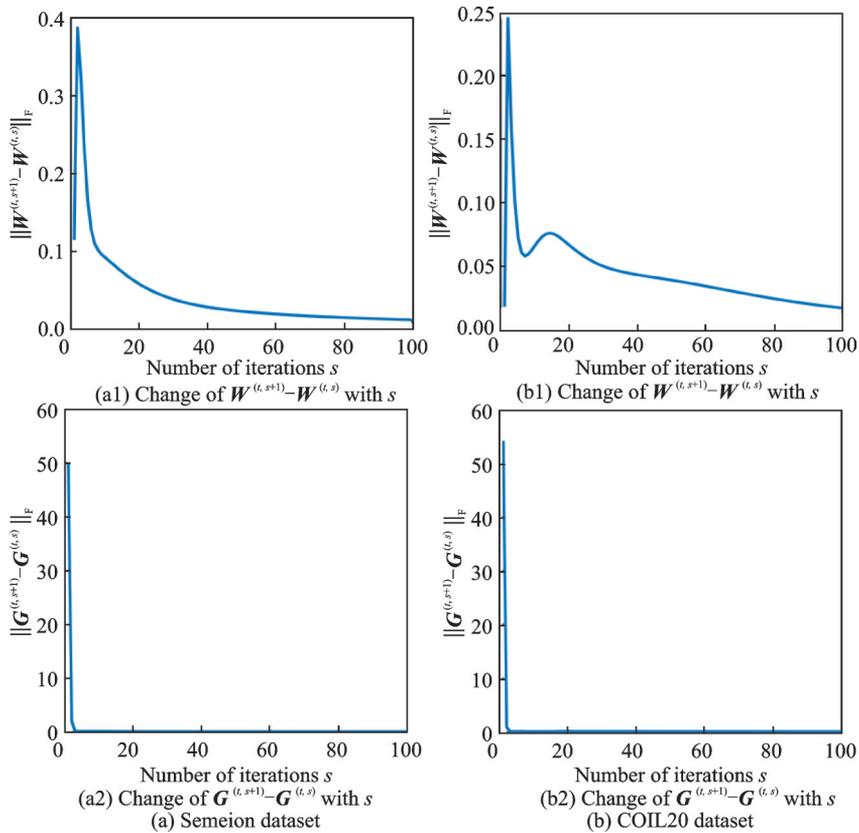


图6 变量序列 $\{W^{(t)}\}$ 和 $\{G^{(t)}\}$ 在Semeion和COIL20数据集上的收敛性

Fig.6 Convergence of variable sequences  $\{W^{(t)}\}$  and  $\{G^{(t)}\}$  on Semeion and COIL20 datasets

在内部迭代中,图4仅展示了当 $t=1$ 时目标函数值随迭代次数的变化。可以看出,对于固定的 $E$ ,目标函数值随迭代次数的增大逐渐减小。从图5可以看出RLS-NMF在很少的数次内收敛。

为了说明变量序列 $\{W^{(t)}\}$ 和 $\{G^{(t)}\}$ 的收敛性,图6展示了对于固定的 $E$ ,两次连续的 $W$ 和 $G$ 的差值随着迭代次数的变化差异。通过观察可以发现,随着迭代次数的增加,两次连续的 $W$ 和 $G$ 的变化差异越来越小,并在非常少的迭代次数内收敛到0。这些实验结果进一步验证了RLS-NMF的收敛性。

#### 4.5 参数敏感度

对于无监督学习方法,对参数变化的敏感性是评估算法性能的一个重要标准。为了观察不同参数组合下算法的性能,在每个数据集上选取了 $N$ 值最小的类,展示了在每种参数组合下的聚类性能,其中参数 $\lambda$ 和 $\gamma$ 都是在 $\{0.001, 0.01, 0.1, 0, 1, 10, 100, 1000\}$ 范围内变化。图7展示了RLS-NMF在Semeion和COIL20数据集上在49种参数组合下的性能表现。由观察可发现,在准确率和纯度方面,算法性能随参数的变化浮动范围较小,说明了该算法对参数变化的不敏感性,在其他数据集上也有同样的结论,所以RLS-NMF在较大的参数变化范围内获得良好的性能是有意义的。

## 5 结束语

本文提出了一种新的基于局部相似性学习的非负矩阵分解方法。该方法在保证鲁棒性的基础上学习了数据的全局结构和局部结构信息,能够挖掘数据类内相似和类间相离的性质,使学习到的基矩

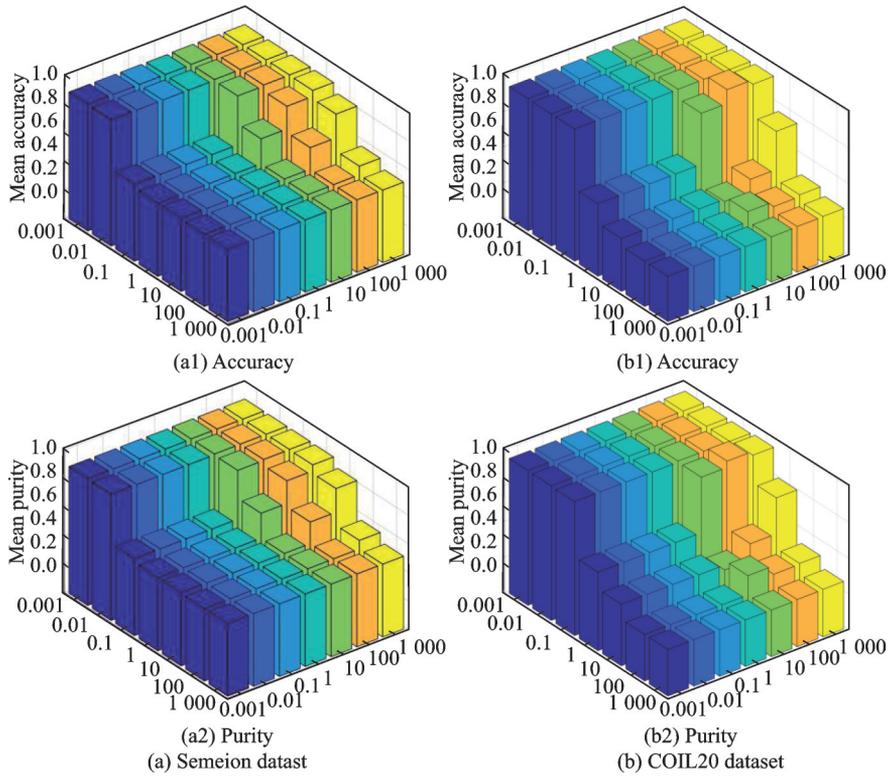


图7 Semeion 和 COIL20 数据集上聚类的准确率和纯度在  $\lambda$  和  $\gamma$  的不同组合下的性能变化

Fig.7 Performance variation of accuracy and purity on Semeion and COIL20 datasets under different combinations of  $\lambda$  and  $\gamma$

阵和系数矩阵能够保留更多原始数据的信息。另外,给系数矩阵施加的正交性约束与局部相似性的学习相互增强了解的唯一性,并提供了聚类的解释。从大量的实验中也可以看出RLS-NMF在准确率和纯度等方面的性能表现优于目前大部分方法,说明该方法具有一定的应用价值。该方法对于非线性分布数据的应用有待于进一步研究。

附录 A

定义 1 对于任意的  $H$  和  $H'$ , 如果满足

$$J(H, H') \geq L(H), J(H, H) = L(H) \tag{A1}$$

那么  $J(H, H')$  就是函数  $L(H)$  的辅助函数。

引理 1 如果  $J(H, H')$  是  $L(H)$  的辅助函数, 那么  $\{H^{(t)}\}$  存在如下更新规则

$$H^{(t+1)} = \arg \min_H J(H, H^{(t)}) \tag{A2}$$

使得目标值  $\{L(H^{(t)})\}$  单调递减, 即  $L(H^{(t)}) = J(H^{(t)}, H^{(t)}) \geq J(H^{(t+1)}, H^{(t)}) \geq L(H^{(t+1)})$ 。

引理 2 对于矩阵  $\Gamma \in \mathbb{R}_+^{n \times n}, \Omega \in \mathbb{R}_+^{k \times k}, S \in \mathbb{R}_+^{n \times k}$  和  $S' \in \mathbb{R}_+^{n \times k}$ , 并且  $\Gamma$  和  $\Omega$  是对称的, 那么以下不等式成立<sup>[12]</sup>。

$$\sum_{i=1}^n \sum_{s=1}^k \frac{(\Gamma S' \Omega)_{is} S_{is}^2}{S_{ik}'} \geq \text{Tr}(S^T \Gamma S \Omega) \tag{A3}$$

证明 固定  $G$ , 得到关于  $W$  的目标函数

$$P(W) = \text{tr} \left( -W^T X^T (X - E) G + \frac{1}{2} W^T X^T X W G^T G \right) + \lambda \text{tr}(W^T D G)$$

根据引理2和不等式  $a \leq (a^2 + b^2)/2b, a \geq 0, b \geq 0$ , 得到  $P(W)$  中正数项的上界

$$\text{tr}(W^T DG) = \sum_{ik} (DG)_{ik} W_{ik} \leq \sum_{ik} (DG)_{ik} \frac{W_{ik}^2 + W_{ik}'^2}{2W_{ik}'}, \quad (\text{A4})$$

$$\text{tr}(W^T X^T X W G^T G) \leq \sum_{ik} \frac{(X^T X W' G^T G)_{ik} W_{ik}^2}{W_{ik}'}, \quad (\text{A5})$$

根据不等式  $a \geq 1 + \lg a, a \geq 0$ , 得到  $P(W)$  中负数项的下界

$$\text{tr}(W^T X^T (X - E) G) = \sum_{ik} (X^T (X - E) G)_{ik} W_{ik} \geq \sum_{ik} (X^T (X - E) G)_{ik} W_{ik}' \left(1 + \lg \frac{W_{ik}}{W_{ik}'}\right) \quad (\text{A6})$$

结合这些边界, 最终得到  $P(W)$  的辅助函数  $P(W, W')$  为

$$P(W, W') = - \sum_{ik} (X^T (X - E) G)_{ik} W_{ik}' \left(1 + \lg \frac{W_{ik}}{W_{ik}'}\right) + \frac{1}{2} \sum_{ik} \frac{(X^T X W' G^T G)_{ik} W_{ik}^2}{W_{ik}'} + \lambda \sum_{ik} (DG)_{ik} \frac{W_{ik}^2 + W_{ik}'^2}{2W_{ik}'} \quad (\text{A7})$$

接下来证明更新规则(12)是辅助函数(A7)的最优解, 为此, 先要证明辅助函数(A7)是凸的。辅助函数  $P(W, W')$  的一阶导数为

$$\frac{\partial P(W, W')}{\partial W_{ik}} = - \frac{(X^T (X - E) G)_{ik} W_{ik}'}{W_{ik}} + \frac{(X^T X W' G^T G)_{ik} W_{ik}}{W_{ik}'} + \lambda \frac{(DG)_{ik} W_{ik}}{W_{ik}'} \quad (\text{A8})$$

对辅助函数求二阶导数得到

$$\frac{\partial^2 P(W, W')}{\partial W_{ik} \partial W_{jl}} = \delta_{ij} \delta_{kl} \left( \frac{(X^T (X - E) G)_{ik} W_{ik}'}{W_{ik}^2} + \frac{(X^T X W' G^T G)_{ik} + \lambda (DG)_{ik}}{W_{ik}'} \right) \quad (\text{A9})$$

其中  $\delta$  是  $P(W, W')$  的海森矩阵, 并且满足如果  $i = j, \delta_{ij} = 1$ ; 否则  $\delta_{ij} = 0$ 。  $\delta$  的对角线上是非零元素, 非对角线上是零元素, 因此它是正定的。这就说明  $P(W, W')$  是凸的, 令  $P(W, W')$  的一阶导数等于 0, 就能得到  $P(W, W')$  的全局最优解, 即

$$\frac{(X^T (X - E) G)_{ik} W_{ik}'}{W_{ik}} = \frac{(X^T X W' G^T G)_{ik} W_{ik}}{W_{ik}'} + \lambda \frac{(DG)_{ik} W_{ik}}{W_{ik}'} \quad (\text{A10})$$

式(A10)可进一步化简为

$$W_{ik} = W_{ik}' \sqrt{\frac{(X^T (X - E) G)_{ik}}{(X^T X W' G^T G)_{ik} + \lambda (DG)_{ik}}} \quad (\text{A11})$$

定义  $W^{(t, s+1)} = W, W^{(t, s)} = W'$ , 可以看出更新规则(A10)本质上遵循式(A2), 根据引理1可以证明目标函数(7)是递减的。

## 附录 B

**证明** 首先通过不等式  $a \leq (a^2 + b^2)/2b, a \geq 0, b \geq 0$  找到  $\ell_G$  中正数项的上界

$$\text{tr}(G^T DW) = \sum_{ik} (DW)_{ik} G_{ik} \leq \sum_{ik} (DW)_{ik} \frac{G_{ik}^2 + G_{ik}'^2}{2G_{ik}'} \quad (\text{B1})$$

$\Gamma$  和  $S$  为单位矩阵, 根据引理2得到如下正数项的上界

$$\text{tr}\left(G(W^T X^T X W + \Theta)^+ G^T\right) \leq \sum_{ik} \frac{\left(G(W^T X^T X W + \Theta)^+\right)_{ik} G_{ik}^2}{G_{ik}'} \quad (\text{B2})$$

然后, 通过不等式  $a \geq 1 + \lg a, a \geq 0$ , 得到负数项的下界

$$\text{tr}\left(G^T (X - E)^T X W\right) \geq \sum_{ik} \left((X - E)^T X W\right)_{ik} G_{ik}' \left(1 + \lg \frac{G_{ik}}{G_{ik}'}\right) \quad (\text{B3})$$

$$\text{tr}\left(G(W^T X^T X W + \Theta)^+ G^T\right) \geq \sum_{ikl} (W^T X^T X W + \Theta)^-_{kl} G'_{ik} G'_{il} \left(1 + \lg \frac{G_{ik} G_{il}}{G'_{ik} G'_{il}}\right) \quad (\text{B4})$$

因此,结合这些边界得到  $\ell_G$  的辅助函数

$$\begin{aligned} J(G, G') = & -\sum_{ik} \left((X-E)^T X W\right)_{ik} G'_{ik} \left(1 + \lg \frac{G_{ik}}{G'_{ik}}\right) + \lambda \sum_{ik} (D W)_{ik} \frac{G_{ik}^2 + G'^2_{ik}}{2G'_{ik}} + \\ & \frac{1}{2} \sum_{ik} \frac{\left(G'(W^T X^T X W + \Theta)^+\right)_{ik} G_{ik}^2}{G'_{ik}} - \frac{1}{2} \sum_{ikl} (W^T X^T X W + \Theta)^-_{kl} G'_{ik} G'_{il} \left(1 + \lg \frac{G_{ik} G_{il}}{G'_{ik} G'_{il}}\right) \end{aligned} \quad (\text{B5})$$

对辅助函数(B5)求导得到

$$\begin{aligned} \frac{\partial J(G, G')}{\partial G_{ik}} = & -\frac{\left((X-E)^T X W\right)_{ik} G'_{ik}}{G_{ik}} + \lambda \frac{(D W)_{ik}}{G'_{ik}} G_{ik} + \\ & \frac{\left(G'(W^T X^T X W + \Theta)^+\right)_{ik} G_{ik}^2}{G'_{ik}} - \frac{\left(G'(W^T X^T X W + \Theta)^-\right)_{ik} G'_{ik}}{G_{ik}} \end{aligned} \quad (\text{B6})$$

进一步对式(B6)求二阶导数,得到其海森矩阵为

$$\begin{aligned} \frac{\partial^2 J(G, G')}{\partial G_{ik} \partial G_{jl}} = & \delta_{ij} \delta_{kl} \left( \frac{\left((X-E)^T X W\right)_{ik} G'_{ik}}{G_{ik}^2} + \lambda \frac{(D W)_{ik}}{G'_{ik}} + \right. \\ & \left. \frac{\left(G'(W^T X^T X W + \Theta)^+\right)_{ik} G_{ik}^2}{G'_{ik}} + \frac{\left(G'(W^T X^T X W + \Theta)^-\right)_{ik} G'_{ik}}{G_{ik}^2} \right) \end{aligned} \quad (\text{B7})$$

根据海森矩阵可以看出  $J(G, G')$  关于  $G$  是凸的,所以其全局最优值可以通过一阶导数等于零求得,从而得到  $G$  的更新规则为

$$G_{ik} \leftarrow G'_{ik} \frac{\left(\left((X-E)^T X W\right)_{ik} + \left(G'(W^T X^T X W + \Theta)^-\right)_{ik}\right)}{\lambda (D W)_{ik} + \left(G'(W^T X^T X W + \Theta)^+\right)_{ik}} \quad (\text{B8})$$

设  $G^{(t,s+1)} = G$  和  $G^{(t,s)} = G'$ , 根据引理1可以看出  $\ell_G$  在更新规则(B8)下递减。为了确定式(B8)中的乘子  $\Theta$ , 通过  $\ell_G$  的一阶最优化条件得出

$$\begin{aligned} G^T \left( -(X-E)^T X W + G W^T X^T X W + \lambda D W + G \Theta \right) = \\ -G^T (X-E)^T X W + G^T G W^T X^T X W + \lambda G^T D W + G^T G \Theta = \\ -G^T (X-E)^T X W + W^T X^T X W + \lambda G^T D W + \Theta = 0 \end{aligned} \quad (\text{B9})$$

因此,就有  $W^T X^T X W + \Theta = G^T (X-E)^T X W - \lambda G^T D W$ , 又因为  $M^+ - M^- = M$ , 所以  $(W^T X^T X W + \Theta)^+ = G^T (X-E)^T X W$  和  $(W^T X^T X W + \Theta)^- = \lambda G^T D W$ , 代入式(B8), 得到更新规则(18)。

## 附录 C

**证明** 初始化值  $W^{(0)}$ ,  $G^{(0)}$  和  $E^{(0)}$  是非负的, 所以当  $t=0$  时, 能够保证  $\left\{W^{(t)}, G^{(t)}, (X-E)^{(t)}\right\}$  是非负的。接下来将要证明对于任意的  $t \geq 0$ , 在已知  $W^{(t)} \geq 0$ ,  $G^{(t)} \geq 0$  和  $(X-E)^{(t)} \geq 0$  的情况下,  $W^{(t+1)} \geq 0$ ,  $G^{(t+1)} \geq 0$  和  $(X-E)^{(t+1)} \geq 0$  也成立。假设在第  $t$  轮迭代中矩阵  $W^{(t)}$ ,  $G^{(t)}$  和  $(X-E)^{(t)}$  的元素都是非负的, 那么在代数更新规则(12)和(18)下,  $W^{(t+1)} \geq 0$ ,  $G^{(t+1)} \geq 0$ 。根据式(21),  $[E^{(t+1)}]_{(i)}$  有两种情况, 一种是等于0, 另一种

等于  $\frac{\left\| \left[ (X - XWG^T)^{(t+1)} \right]_{(i)} \right\|_2 - \gamma}{\left\| \left[ (X - XWG^T)^{(t+1)} \right]_{(i)} \right\|_2} \left[ (X - XWG^T)^{(t+1)} \right]_{(i)}$ 。如果  $\left[ E^{(t+1)} \right]_{(i)} = 0$ , 那么  $\left[ X - E^{(t+1)} \right]_{(i)} = X_{(i)}$  一定是非负的。如果  $\left[ E^{(t+1)} \right]_{(i)} = \frac{\left\| \left[ (X - XWG^T)^{(t+1)} \right]_{(i)} \right\|_2 - \gamma}{\left\| \left[ (X - XWG^T)^{(t+1)} \right]_{(i)} \right\|_2} \left[ (X - XWG^T)^{(t+1)} \right]_{(i)}$ , 令  $Q^{(t+1)} = (X - XWG^T)^{(t+1)}$ , 那么

就会得到

$$\begin{aligned} \left[ X - E^{(t+1)} \right]_{(i)} &= X_{(i)} - \left[ Q^{(t+1)} \right]_{(i)} + \gamma \frac{\left[ Q^{(t+1)} \right]_{(i)}}{\left\| \left[ Q^{(t+1)} \right]_{(i)} \right\|_2} = \left[ (XWG^T)^{(t+1)} \right]_{(i)} + \gamma \frac{\left[ X - (XWG^T)^{(t+1)} \right]_{(i)}}{\left\| \left[ Q^{(t+1)} \right]_{(i)} \right\|_2} = \\ &= \gamma \frac{X_{(i)}}{\left\| \left[ Q^{(t+1)} \right]_{(i)} \right\|_2} + \left[ (XWG^T)^{(t+1)} \right]_{(i)} \frac{\left\| \left[ Q^{(t+1)} \right]_{(i)} \right\|_2 - \gamma}{\left\| \left[ Q^{(t+1)} \right]_{(i)} \right\|_2} \end{aligned} \quad (C1)$$

当  $\left\| \left[ Q^{(t+1)} \right]_{(i)} \right\|_2 > \gamma$  时, 式(C1)的最右边项一定是非负的。因此, 通过数学归纳, 序列  $\{W^{(t)}, G^{(t)}, (X - E)^{(t)}\}$  是非负的。

#### 参考文献:

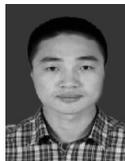
- [1] ZHU Xiaofeng, ZHANG Shichao, LI Yonggang, et al. Low-rank sparse subspace for spectral clustering[J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 31(8): 1532-1543.
- [2] ABIANTUN R, PRABHU U, SAVVIDES M. Sparse feature extraction for pose-tolerant face recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(10): 2061-2073.
- [3] SUN Chenhao, WANG Yonghua, WAN Pin, et al. A cooperative spectrum sensing algorithm based on principal component analysis and K-medoids clustering[C]//Proceedings of 2018 33rd Youth Academic Annual Conference of Chinese Association of Automation. Nanjing, China: IEEE, 2018: 835-839.
- [4] DENG Ping, LI Tianrui, WANG Hongjun, et al. Tri-regularized nonnegative matrix tri-factorization for co-clustering[J]. Knowledge-Based Systems, 2021, 226: 107101.
- [5] WANG Qi, CHEN Mulin, NIE Feiping, et al. Detecting coherent groups in crowd scenes by multiview clustering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(1): 46-58.
- [6] LEE D, SEUNG H. Learning the parts of objects by non-negative matrix factorization[J]. Nature, 1999, 401(6755): 788-791.
- [7] LOGOTHETIS N, SHEINBERG D. Visual object recognition, annual review of neuroscience[J]. 1996, 19(1): 577-621.
- [8] PENG Siyuan, SER W, CHEN Badong, et al. Robust orthogonal nonnegative matrix tri-factorization for data representation [J]. Knowledge-Based Systems, 2020, 201/202: 106054.
- [9] WANG Yuxiong, ZHANG Yujin. Nonnegative matrix factorization: A comprehensive review[J]. IEEE Transactions on Knowledge and Data Engineering, 2013, 25(6): 1336-1353.
- [10] DING C, LI T, PENG W, et al. Orthogonal nonnegative matrix t-factorizations for clustering[C]//Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM, 2006: 126-135.
- [11] KIM Y, CHOI S. Weighted nonnegative matrix factorization[C]//Proceedings of 2009 IEEE International Conference on Acoustics, Speech and Signal Processing. Taipei, China: IEEE, 2009: 1541-1544.
- [12] DING C, LI T, JORDAN M. Convex and semi-nonnegative matrix factorizations[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(1): 45-55.
- [13] HUANG Jin, NIE Feiping, HUANG Heng, et al. Robust manifold nonnegative matrix factorization[J]. ACM Transactions on

- Knowledge Discovery from Data (TKDD), 2014, 8(3): 1-21.
- [14] PENG Chong, ZHANG Zhilu, KANG Zhao, et al. Nonnegative matrix factorization with local similarity learning[J]. Information Sciences, 2021, 562: 325-346.
- [15] LIU Xinwang, WANG Lei, ZHANG Jian, et al. Global and local structure preservation for feature selection[J]. IEEE Transactions on Neural Networks and Learning Systems, 2014, 25(6): 1083-1095.
- [16] 周长宇, 姚明海, 李劲松. 基于L21范式的多图正则化非负矩阵分解方法[J]. 计算机应用与软件, 2021, 38(4): 271-275, 310. ZHOU Changyu, YAO Minghai, LI Jinsong. Multi-graph regularized non-negative matrix factorization method based on the L21 norm[J]. Computer Applications and Software, 2021, 38(4): 271-275, 310.
- [17] PENG Chong, ZHANG Qing, KANG Zhao, et al. Kernel two-dimensional ridge regression for subspace clustering[J]. Pattern Recognition, 2021, 113: 107749.
- [18] PENG Chong, KANG Zhao, HU Yunhong, et al. Nonnegative matrix factorization with integrated graph and feature learning [J]. ACM Transactions on Intelligent Systems and Technology, 2017, 8(3): 1-29.
- [19] SONG Qinbao, NI Jingjie, WANG Guangtao. A fast clustering-based feature subset selection algorithm for high-dimensional data[J]. IEEE Transactions on Knowledge and Data Engineering, 2013, 25(1): 1-14.
- [20] LEE D, SEBASTIAN H. Algorithms for non-negative matrix factorization[C]//Proceedings of the 13th International Conference on Neural Information Processing Systems. Cambridge, MA, USA: Massachusetts Institute of Technology Press, 2000: 535-541.
- [21] 徐晓华, 方威, 何萍, 等. 峰值点非负矩阵分解聚类算法[J]. 南京航空航天大学学报, 2021, 53(5): 772-779. XU Xiaohua, FANG Wei, HE Ping, et al. Peaks non-negative matrix factorization[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2021, 53(5): 772-779.
- [22] PENG Chong, KANG Zhao, CAI Shuting, et al. Integrate and conquer: Double-sided two-dimensional k-means via integrating of projection and manifold construction[J]. ACM Transactions on Intelligent Systems and Technology, 2018, 9(5): 57.
- [23] TOLIĆ D, ANTULOV-FANTULIN N, KOPRIVA I. A nonlinear orthogonal non-negative matrix factorization approach to subspace clustering[J]. Pattern Recognition, 2018, 82: 40-55.
- [24] GUAN Naiyang, LIU Tongliang, ZHANG Yangmuzi, et al. Truncated Cauchy non-negative matrix factorization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(1): 246-259.

## 作者简介:



侯兴荣(1997-),女,硕士研究生,研究方向:机器学习, E-mail: houxr1997s@163.com。



彭冲(1991-),通信作者,男,博士,副教授,研究方向:机器学习和数据挖掘, E-mail:pchong1991@163.com。

(编辑:王静)