

引文格式:

崔国盛, 李焯. 非负子空间聚类指导下的非负矩阵分解 [J]. 集成技术, 2019, 8(5): 3-12.

Cui GS, Li Y. Non-negative subspace clustering guided non-negative matrix factorization [J]. Journal of Integration Technology, 2019, 8(5): 3-12.

非负子空间聚类指导下的非负矩阵分解

崔国盛¹ 李焯^{1,2,3}

¹(中国科学院深圳先进技术研究院 深圳 518055)

²(中国科学院健康信息学重点实验室 深圳 518055)

³(健康大数据智能分析技术国家地方联合工程研究中心 深圳 518055)

摘 要 非负矩阵分解作为一种有效的数据表示方法被广泛应用于模式识别和机器学习领域。为了得到原始数据紧致有效的低维数据表示, 无监督非负矩阵分解方法在特征降维的过程中通常需要同时发掘数据内部隐含的几何结构信息。通过合理建模数据样本间的相似性关系而构建的相似度图, 通常被用来捕获数据样本的空间分布结构信息。子空间聚类可以有效发掘数据内部的子空间结构信息, 其获得的自表达系数矩阵可用于构建相似度图。该文提出了一种非负子空间聚类算法来发掘数据的子空间结构信息, 同时利用该信息指导非负矩阵分解, 从而得到原始数据有效的非负低维表示。同时, 该文还提出了一种有效的迭代求解方法来求解非负子空间聚类问题。在两个图像数据集上的聚类实验结果表明, 利用数据的子空间结构信息可以有效改善非负矩阵分解的性能。

关键词 非负子空间聚类; 非负矩阵分解; 结构信息; 图像聚类

中图分类号 TG 156 文献标志码 A doi: 10.12146/j.issn.2095-3135.20190702001

Non-Negative Subspace Clustering Guided Non-Negative Matrix Factorization

CUI Guosheng¹ LI Ye^{1,2,3}

¹(Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China)

²(Key Lab for Health Informatics, Chinese Academy of Sciences, Shenzhen 518055, China)

³(Joint Engineering Research Center for Health Big Data Intelligent Analysis Technology, Shenzhen 518055, China)

Abstract As an effective data representation method, non-negative matrix factorization has been widely used in pattern recognition and machine learning. To obtain a compact and effective data representation in data dimension reduction, unsupervised non-negative matrix factorization usually needs to discover the latent geometry structure information of the data. A similarity graph constructed by reasonably modeling similarity

收稿日期: 2019-07-02 修回日期: 2019-08-06

基金项目: 海外高层次人才创新创业专项资金项目 (KQJSCX20170731165939298)

作者简介: 崔国盛, 博士, 助理研究员, 研究方向为模式识别与机器学习; 李焯 (通讯作者), 研究员, 研究方向为健康信息学, E-mail: ye.li@siat.ac.cn.

relationships between data samples is typically used to capture spatial distribution structure information for data samples. Subspace learning can effectively explore the subspace structure information inside the data, and the obtained self-expressive coefficient matrix can be used to construct a similarity graph. In this paper, a non-negative subspace clustering algorithm is proposed to explore the subspace structure information of data which is used to guide the non-negative matrix factorization, so as to obtain the effective non-negative low-dimensional representation of the original data. At the same time, an effective iterative strategy is developed to solve the problem of non-negative subspace clustering. The results of clustering experiments on two image datasets demonstrate that utilizing the subspace structure information of data can effectively improve the performance of non-negative matrix factorization.

Keywords non-negative subspace clustering; non-negative matrix factorization; structural information; image clustering

1 引 言

随着计算机技术和传感器技术的飞速发展,人们获取的原始数据量也在快速增长。面对数据量快速增长带来的挑战,如何设计数据降维算法从而得到有效的低维数据表示成为多元数据分析中亟待解决的问题。一般而言,数据降维算法需要满足两个基本特性^[1]:首先,原始数据的维数需要被约减;其次,根据应用背景,原始数据内部的主要成分、隐含概念或潜在的变量需要被同时发掘出来。

在各个数据处理领域,原始数据的组织形式通常是矩阵(或张量)的形式。因此,从代数的角度可以将数据降维问题看作是一种矩阵分解的问题,即将一个大矩阵分解为两个小规模矩阵的乘积形式,从而减小了数据的存储复杂度以及后续数据处理的计算复杂度。如今,矩阵分解已经是一类应用非常广泛的数据表示学习技术,包括奇异值分解^[2]、主成分分析(Principal Component Analysis, PCA)^[3]、向量量化^[4]、线性判别分析^[5]、非负矩阵分解(Non-negative Matrix Factorization, NMF)^[6]等。其中,非负矩阵分解由于对因子矩阵施加了非负约束,从而可学习到原始数据基于部分的表示。Logothetis

和 Sheinberg^[7]研究表明,基于部分的数据表示与大脑的感知机制是一致的。正是由于这种特性,使得非负矩阵分解在各个领域被广泛研究和应用,包括模式识别^[8-10]、机器学习^[11-13]、数据挖掘^[14-17]、金融与经济学^[18]、生物信息学^[19]等领域。在生物信息学方面则更具体地包括基因表达分类^[20]、细胞分析、脑电图信号处理^[21]和病理诊断等。

对于无监督非负矩阵分解方法,在实现数据特征维数约减的同时,有效挖掘数据内部的潜在数据分布结构信息通常可有效改善得到的低维数据表示。受到流行学习的启发,Cai等^[22]提出了图正则非负矩阵分解(Graph Regularized Non-negative Matrix Factorization, GNMF),该方法构建一个基于数据样本点间距离的 K 近邻图来编码数据的分布结构信息,通过约束低维数据样本间保持原高维空间中样本间的邻域关系,从而使得到的低维数据表示具有更好的表征能力。虽然GNMF在一定程度上可以改善NMF的性能,但其图正则约束所依赖的 K 近邻图通常基于热核度量,这种度量对所包含的带宽参数比较敏感^[23]。另外,基于欧式距离的近邻选择方式对数据可能包含的误差干扰并不是很鲁棒。而基于自表达模型的子空间聚类算法^[24]可发掘数据的潜在子空

间结构, 其中每个数据样本点都可用整个数据集作为字典来线性表示, 对应的表达系数包含了数据样本间的相似性信息, 也编码了数据内部的子空间结构信息。为了有效利用子空间结构信息来改善非负矩阵分解算法的性能, 本文提出非负子空间聚类算法来有效发掘数据内部的子空间结构信息, 并通过在非负矩阵分解框架中引入包含子空间结构信息的图正则项来增强得到的低维数据特征的特征能力。此外, 在构建图的过程中, 采用阈值处理操作使算法对数据中可能存在的误差和噪声更鲁棒。最后, 在两个图像数据集上进行聚类实验证实了本算法的有效性。

2 非负子空间聚类指导的非负矩阵分解

本文提出非负子空间聚类算法, 并利用该算法得到的子空间表达系数构建相似度矩阵, 从而得到对应的相似度图。将该拉普拉斯图正则项纳入非负矩阵分解框架, 从而使非负矩阵分解可以得到更有效的非负低维数据表征。

2.1 非负子空间聚类

给定一组经过充分采样的数据集 $X=[X_1, \dots, X_k]=[x_1, x_2, x_3, \dots, x_n] \in \mathbb{R}^{m \times n}$ 。其中, m 表示特征维度; n 为数据样本个数。假设数据是从 k 个子空间的集合 $\{S_i\}_{i=1}^k$ 中采样得到的, 则 X_i 表示从子空间 S_i 中采样得到的 n_i 个数据样本, 且有 $n=\sum_{i=1}^k n_i$ 。子空间聚类目标即根据数据所属的潜在子空间对数据进行分割。在基于自表达的子空间聚类模型中, 假设每个数据样本点 x_i 都可用原数据集中的所有样本 $X \in \mathbb{R}^{m \times n}$ 线性表示, 即:

$$x_i = Xz_i, \text{ s.t. } z_i^i = 0 \quad (1)$$

其中, x_i 是 X 中第 i 列; z_i 是 x_i 关于自表达字典 X 的重构系数向量。约束 $z_i^i = 0$ 是为了避免得到无效解。公式(1)可以进一步写成如下矩

阵的形式。

$$X = XZ, \text{ s.t. } \text{diag}(Z) = 0 \quad (2)$$

其中, Z 是数据集中每个样本 $x_i (i=1, \dots, n)$ 以数据集本身 X 为字典对自身进行重构的重构系数矩阵。重构系数可以描述样本间的相似性, 一般可用自表达系数矩阵 Z 构建相似度矩阵, 即 $W = (|Z|^T + |Z|) / 2$ 。自表达矩阵中的元素表示样本之间的相关性, 因此, 自表达系数矩阵应该具有非负性。基于此, 本文提出的非负子空间聚类模型目标函数如下:

$$\min_{Z \geq 0} \|X - XZ\|_F^2 + \lambda \|Z\|_F^2 \quad (3)$$

其中, $\|\cdot\|_F$ 是 Frobenius 范数; $\lambda \|Z\|_F^2$ 是关于自表达矩阵的正则项, 是为了避免在求解过程中产生无效解; λ 为平衡自表达项和正则项的参数。以上最小化的问题可以通过迭代的方式进行求解。假设 $\Psi = [\psi_{ij}] \in \mathbb{R}^{n \times n}$ 是条件 $Z \geq 0$ 的拉格朗日乘子, 则上述问题的拉格朗日函数可以写为:

$$\begin{aligned} \mathcal{L}(Z) &= \|X - XZ\|_F^2 + \lambda \|Z\|_F^2 + \text{tr}(\Psi Z^T) \\ &= \text{tr}(X - XZ)(X - XZ)^T + \lambda \text{tr}(ZZ^T) + \text{tr}(\Psi Z^T) \end{aligned} \quad (4)$$

则 $\mathcal{L}(Z)$ 关于 Z 的导数为:

$$\frac{d\mathcal{L}(Z)}{dZ} = -2X^T X + 2X^T XZ + 2\lambda Z + \Psi \quad (5)$$

根据 KKT 条件^[25], 有 $\psi_{ij} z_{ij} = 0$, 则有:

$$(-2X^T X + 2X^T XZ + 2\lambda Z + \Psi)_{ij} z_{ij} = 0 \quad (6)$$

最终可得如下迭代更新规则:

$$Z = Z \odot \frac{X^T X}{X^T XZ + \lambda Z} \quad (7)$$

其中, \odot 表示对应元素相乘; 式中的除也是对应元素相除的操作。得到自表达矩阵 Z 后, 可根据 $W = Z + Z^T$ 计算相似度矩阵。另外, 为了减小不同子空间内样本的干扰, 在构建相似度矩阵的过程中, 对自表达矩阵 Z 进行了阈值处理操作^[26], 即只保留 Z 每一列中最大的 k 元素。非负子空间聚类算法的流程如表 1 所示。

表1 非负子空间聚类

Table 1 Non-negative subspace clustering

算法1 非负子空间聚类
输入: 数据集 $X=[x_1, x_2, x_3, \dots, x_n] \in \mathbb{R}^{m \times n}$, 迭代次数 M , 正则参数 λ , 阈值参数 k
输出: 自表达矩阵 Z 、相似度矩阵 W
S1: 用最小二乘回归初始化自表达矩阵 Z
S2: $m=0$
S3: 更新自表达矩阵 Z :
$Z = Z \odot \frac{X^T X}{X^T X Z + \lambda Z}$
S4: $m=m+1$, 如果 $m \leq M$, 回到 S3; 否则, 到 S5
S5: 保留 Z 每列最大的 k 个元素, 得到 Z' , 计算相似度矩阵 $W = Z + Z'^T$

2.2 自表达图正则非负矩阵分解

原始非负矩阵分解目标式如下:

$$\min_{U, V \geq 0} \|X - UV\|_F^2 \quad (8)$$

其中, $U \in \mathbb{R}^{m \times r} \geq 0$ 为基矩阵; $V \in \mathbb{R}^{r \times n} \geq 0$ 为编码矩阵。公式(8)仅对因子矩阵有非负约束, 没有探索利用数据内部潜在的分布结构信息。

为了使非负矩阵分解得到的数据低维表示可以有效利用子空间结构信息, 约束低维表示空间中的样本分布保持原始高维空间中数据的子空间结构关系。具体地, 对低维表示空间中的样本进行如下形式的约束:

$$\frac{1}{2} \sum_i^n \sum_j^n W_{ij} \|v_i - v_j\|_2^2 = \text{tr}(VLV^T) \quad (9)$$

其中, $L = D - W$; $D_{ii} = \sum_j W_{ij}$ 或 $D_{ii} = \sum_j W_{ji}$; W_{ij} 表示 x_j 与 x_i 的某种相似性度量; v_i 和 v_j 分别为编码矩阵 V 的第 i 和第 j 列, 即原数据空间中样本 x_i 和 x_j 的低维表示。如果 x_j 和 x_i 在原空间的分布是接近的, 那么样本 x_j 和 x_i 的相似度会比较大, 权重系数 W_{ij} 的值 (W_{ij} 值与 x_j 和 x_i 的相似度成正比) 也会比较大, 通过最小化 $W_{ij} \|v_i - v_j\|_2^2$ 可使对应的低维表示 v_j 和 v_i 在投影空间的分布也是接近的。当 x_j 和 x_i 在原空间的分布差异较大时, 样本 x_j 和 x_i 的相似度会比较小, W_{ij} 也会比较小。因

此, 目标式对最小化 $W_{ij} \|v_i - v_j\|_2^2$ 的作用力度会比较小。最终, 通过最小化目标公式(9)可使低维表示在局部区域保持原始空间的近邻关系。为了使低维空间中的样本保持原始空间的子空间结构信息, 公式(9)中的权重系数 W_{ij} 采用 x_i 和 x_j 之间的重构系数, 即 $z_{ij} + z_{ji}$ 。将上述正则项纳入非负矩阵分解框架, 可得如下自表达图正则非负矩阵分解目标式:

$$\min_{U, V \geq 0} \|X - UV\|_F^2 + \alpha \text{tr}(VL_z V^T) \quad (10)$$

以上目标式包含两个变量 U 和 V 是非凸的, 但是对于单独变量 U 或 V 则是凸的, 因此可通过固定其中一个变量来更新另一个变量的方式迭代求解。参考 Cai 等^[22]研究可以得到如下更新规则:

$$U = U \odot \frac{XV^T}{UVV^T} \quad (11)$$

$$V = V \odot \frac{U^T X + \alpha V W}{U^T UV + \alpha V D} \quad (12)$$

3 实验

实验部分将本文所提的算法与 K -means、PCA、NMF 和 GNMF 算法在两个图像数据集上进行了聚类比较实验。对于 PCA、NMF、GNMF 和本文提出的算法, 首先分别对原始数据进行降维; 然后, 在降维后的特征上用 K -means 算法进行聚类, 并与 K -means 算法在原始特征 (Raw Feature) 上的聚类结果进行比较。

3.1 数据集

本文提出的算法在 UMIST 和 ORL 两个人脸图像数据集上进行聚类测试。其中, UMIST 数据集包含 20 个人的 575 张人脸图像, 视角都是从侧面到正面逐渐变化, 图像大小为 28 像素 \times 23 像素; ORL 数据集包含 40 个人的 400 张脸部图像, 每个人有 10 张图像 (大小为 32 像素 \times

32 像素)。这些图像在光照、表情以及是否戴眼镜等方面具有差异性。图 1 展示了 UMIST 数据集和 ORL 数据集中的人脸样例图像。



(a) UMIST 数据集



(b) ORL 数据集

图 1 人脸样例图像

Fig. 1 Sample images

3.2 评价指标

本文对所提取的图像特征进行聚类实验, 并采用两种评价指标评价度量聚类性能, 包括精度 (Accuracy, AC) 和归一化互信息 (Normalized Mutual Information, NMI)。聚类精度的定义如下:

$$AC = \frac{\sum_{i=1}^n \delta[gnd_i, \text{map}(v_i)]}{n} \quad (13)$$

其中, gnd_i 是数据集中第 i 个样本的标签; v_i 是第 i 个样本的低维表示; $\text{map}(\cdot)$ 是最优映射函数^[27],

可以将聚类得到的标签映射为与数据集标签等价的标签; n 为数据集样本总数。 $\delta(a,b)$ 是冲激函数, 当 $a=b$ 时, 值为 1; 反之, 则为 0。 AC 数值越高表明聚类效果越好。

归一化互信息定义如下:

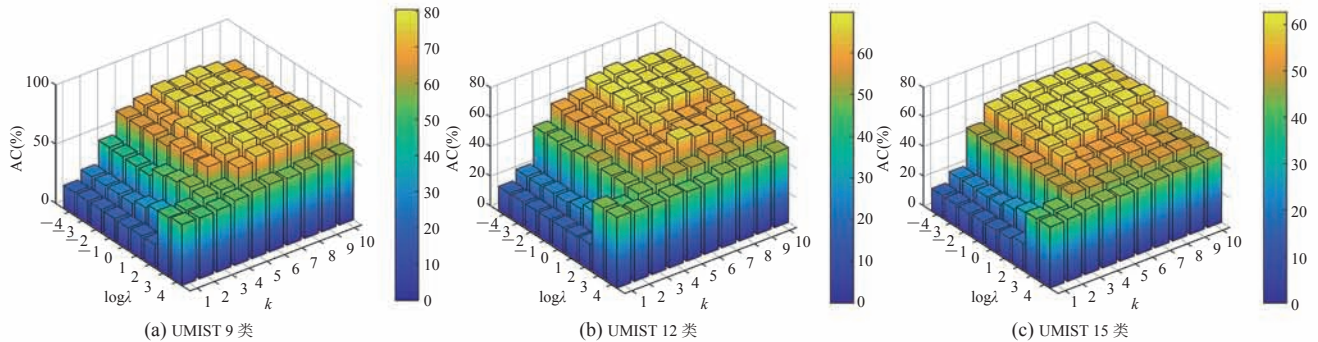
$$NMI(C, C') = \frac{MI(C, C')}{\sqrt{H(C)H(C')}} \quad (14)$$

其中, $MI(C, C')$ 表示聚类 C 和 C' 间的互信息; $H(C)$ 和 $H(C')$ 分别是聚类 C 和 C' 的熵。 $H(C)$ 和 $H(C')$ 被用于对 $MI(C, C')$ 进行归一化处理, 因此, NMI 取值范围为 $0 \sim 1$ 。如果 C 和 C' 相同, 则 $NMI(C, C')=1$; 如果 C 和 C' 独立, 则 $NMI(C, C')=0$ 。 NMI 数值越大表明聚类效果越好。

3.3 参数分析

本节在两个数据集上对算法中包含的参数 λ 和 k 进行调试, 并为算法选取合适的参数。分别从 UMIST 数据集和 ORL 数据集抽取 9、12、15 类和 15、20、25 类样本进行聚类实验, 图 2 和图 3 分别展示了算法在两个数据集不同类别个数的样本上聚类精度随参数 λ 和 k 的变化情况。其中, λ 的变化范围为 $10^{-4} \sim 10^4$; k 的变化范围为 $1 \sim 10$ 。

从图 2 可以看出, 对于 UMIST 9 类数据集, 当参数 k 大于 5、而参数 λ 小于 10^4 时, 算法性能比较稳定; 对于 UMIST 12 类数据集, 当

图 2 聚类精度在 9、12、15 类 UMIST 数据集上随着参数 λ 和 k 的变化Fig. 2 Clustering accuracy varies with parameters λ and k on UMIST datasets of 9, 12 and 15 classes

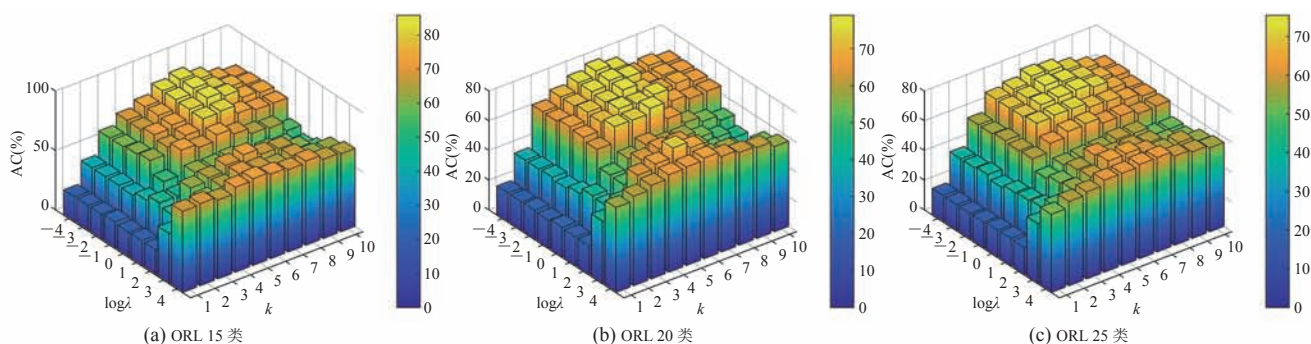


图3 聚类精度在15、20、25类ORL数据集上随着参数 λ 和 k 的变化

Fig. 3 Clustering accuracy varies with parameters λ and k on ORL datasets of 15, 20 and 25 classes

参数 k 在6~10内、而参数 λ 小于 10^0 时,算法表现比较稳定;在UMIST 15类数据集,当参数 k 在4~10内变化、参数 λ 小于 10^0 时,算法表现较为平稳。在9类、12类和15类UMIST数据集上算法取得最好结果的参数(k, λ)分别为(8, 10^{-1})、(7, 10^{-4})、(5, 10^{-1})。

由图3可见,对于ORL 15类数据集,当参数 k 取值范围为6~8、 λ 变化范围为 10^{-4} ~ 10^{-1} 时,算法表现较好;在ORL 20类数据集上,当参数 k 在5~7内取值、 λ 在 10^{-4} ~ 10^0 时,算法得到的精度比较高;对于ORL 25类数据集,当参数 k 在4~8内取值、 λ 在 10^{-4} ~ 10^{-1} 时,算法精度较高。在15类、20类和25类ORL数据集上算法取得最好结果的参数(k, λ)分别为(7, 10^{-4})和(7, 10^{-3})、(6, 10^{-4})、(6, 10^{-4})。

从图2和图3可以看出,随着数据集包含的数据类别个数的增加,在一定范围内随着参数 k 的取值变小,算法表现逐渐变好。这是因为随着数据类别增多,类间干扰变强,需要减小 k 值来抑制类间干扰;另一方面, k 值不能设置得过小,如果 k 选择过小,那么会导致算法不能很好地捕捉数据的子空间结构信息,即数据邻域结构没有被有效地捕捉。然而,对于比较大范围的 k 取值,算法性能表现相对稳定。相对来说,算法对参数 λ 的敏感性没有对参数 k 强。从UMIST数据集可以看出,对于变化范围很大的 λ 取值,

算法表现都比较平稳。但总体而言,特别是从ORL数据集上可以看出, λ 取值小于 10^0 时,算法表现比较好。

3.4 比较实验

本节将本文算法与K-means、PCA、NMF和GNMF算法进行了比较。具体地,在几种特征学习算法学习到的低维特征上进行了图像聚类实验,同时对比了在未经降维的原始特征上的聚类效果。对于GNMF算法包含的正则参数和近邻个数分别采用Cai等^[22]研究中的默认设定,分别为100和5。本文算法的参数 λ 和 k 参考3.3节的参数分析,分别设置为 10^{-4} 、5。表2和表3分别是本文算法与其他算法在UMIST数据集和ORL数据集上的比较实验结果。

从表2和3可以看出,本文提出的算法性能要优于其他比较算法。其中,GNMF也试图发掘并利用数据内部的潜在结构信息,一定程度上改

表2 不同算法在UMIST数据集上的聚类结果

UMIST dataset		
算法类型	评价指标 (%)	
	精度	归一化互信息
K-means	41.03±1.9	59.44±1.9
PCA	39.44±1.6	58.71±1.5
NMF	40.45±2.3	58.27±2.0
GNMF	62.59±4.3	77.58±2.7
本文算法	71.27±6.0	85.03±2.7

表 3 不同算法在 ORL 数据集上的聚类结果

Table 3 Clustering performance of different methods on

算法类型	ORL dataset	
	评价指标 (%)	
	精度	归一化互信息
K-means	50.30±3.0	69.64±2.2
PCA	54.40±2.3	73.06±1.0
NMF	57.22±1.8	75.54±1.3
GNMF	60.52±1.6	75.06±0.9
本文算法	69.23±3.0	83.87±1.5

善了 NMF 算法的性能。但本文算法在两个数据集上的表现都优于 GNMF。在这两个数据集上可以看出, 本文算法可以更好地发掘并利用数据内部的隐含结构信息, 从而得到更好的数据低维表示。GNMF 根据样本间的欧式距离来构建相似度图, 然而, 在高维图像像素张成的空间中数据样本过于稀疏, 利用欧式距离并不能很好地描述样本之间的相似性。本文提出的算法从子空间聚类的角度度量样本之间的相似性, 处于相同子空间的样本之间可以更好地互相表征, 同时利用样本之间的重构系数作为样本间相似性的度量, 可以更好地表征样本之间的相似性。与此同时, 算法

采用的阈值处理操作有效地抑制了不同子空间样本之间的干扰。

图 4 和图 5 分别是本文提出的算法与 GNMF 算法的聚类结果比较——分别从 UMIST、ORL 数据集抽取两个聚类进行聚类结果的图例对比展示。图 4 和图 5 中红色方框表示被错误聚类的图像。由此可以看出, 在这两个数据集的 4 个聚类中, 本文算法正确聚类的图像个数要比 GNMF 多, 而被错误聚类的图像个数要比 GNMF 少。这表明本文算法的聚类效果优于 GNMF 算法。

图 6 和图 7 分别为实验中对比的几种算法所提取的特征(包括原始数据特征)用 t -SNE 算法可视化的对比图。从特征可视化的效果来看, 本文提出的算法可以令相似的聚类尽量聚合, 相异的聚类尽量分离。由于本文提出的算法和 GNMF 算法都利用了数据的潜在分布结构信息, 因此, 相比于其他方法提取的特征更具有判别性。

4 总 结

本文提出一种非负子空间聚类指导的非负矩



图 4 本文算法和 GNMF 算法在 UMIST 数据集上的聚类结果图例对比 (两个聚类)

Fig. 4 Comparison of the clustering results of our method and GNMF on the UMIST dataset (on two clusters)



图 5 本文算法和 GNMF 算法在 ORL 数据集上的聚类结果图例对比 (两个聚类)

Fig. 5 Comparison of the clustering results of our method and GNMF on the ORL dataset (on two clusters)

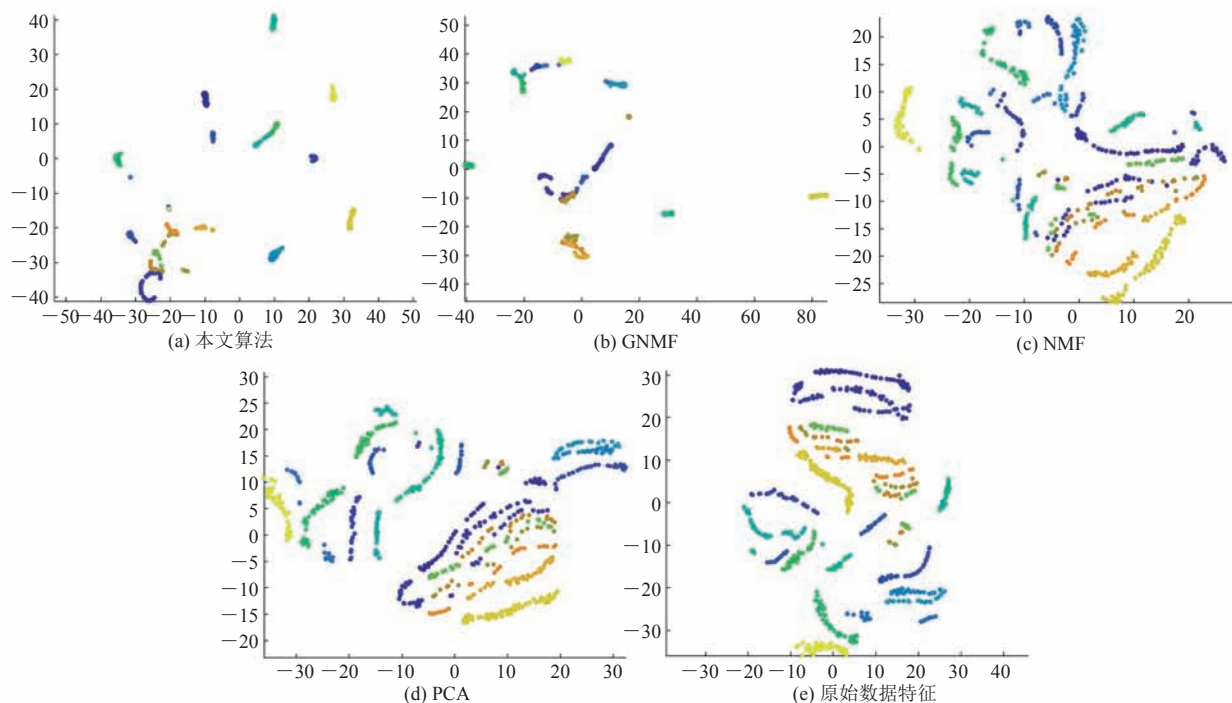


图 6 本文算法、GNMF、NMF 和 PCA 算法提取的特征与原始数据特征用 t -SNE 算法可视化的结果 (UMIST 数据集)

Fig. 6 The features extracted by our method and GNMF, NMF and PCA algorithms as well as original data feature are visualized by t -SNE algorithm (on UMIST dataset)

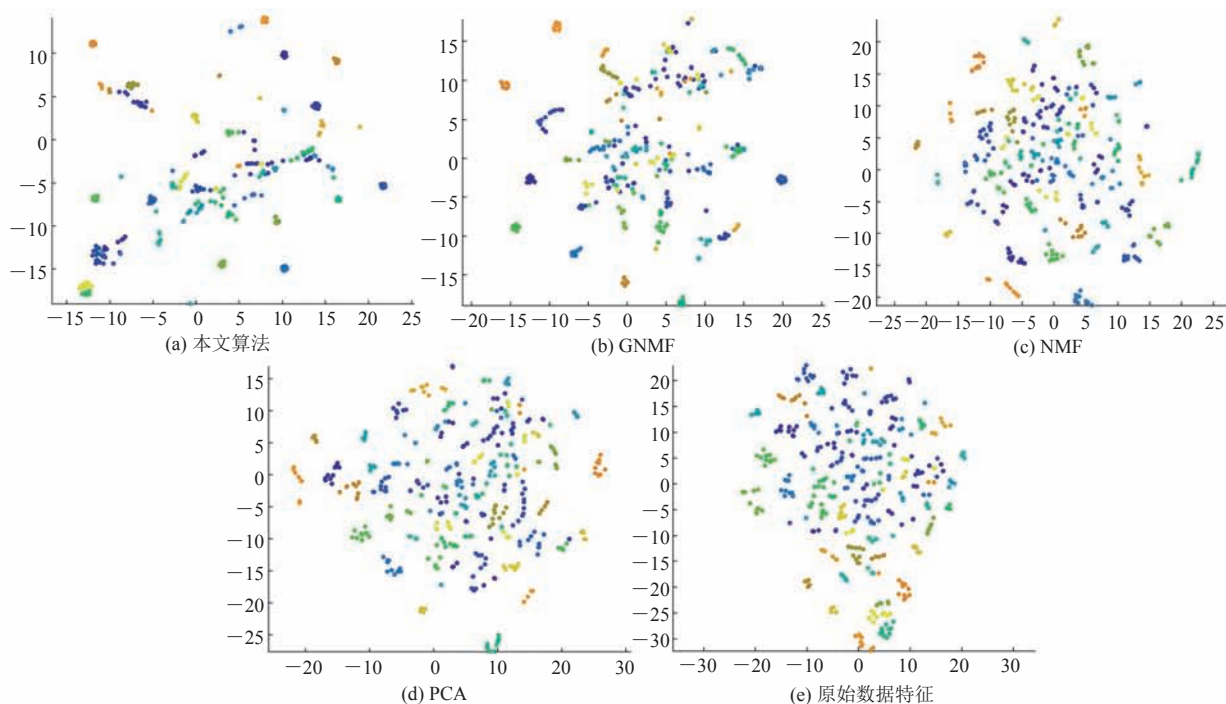


图 7 本文算法、GNMF、NMF 和 PCA 算法提取的特征与原始数据特征用 t -SNE 算法可视化的结果 (ORL 数据集)

Fig. 7 The features extracted by our method and GNMF, NMF and PCA algorithms as well as original data feature are visualized by t -SNE algorithm (on ORL dataset)

阵分解算法。该算法发掘并利用了数据内部的潜在子空间结构信息, 从而使得非负矩阵分解得到更有效的低维数据表示。另外, 还提出了基于自表达的非负子空间聚类模型, 并给出了有效的迭代求解算法。在构建基于自表达矩阵的拉普拉斯图的过程中, 采用阈值处理操作有效抑制了噪声和误差的干扰, 从而使算法更加鲁棒。最后, 在 UMIST 和 ORL 两个图像数据集上的聚类实验验证了该算法的有效性。

参 考 文 献

- [1] Wang YX, Zhang YJ. Nonnegative matrix factorization: a comprehensive review [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2013, 25(6): 1336-1353.
- [2] Turk M, Pentland A. Eigenfaces for recognition [J]. *Journal of Cognitive Neuroscience*, 1991, 3(1): 71-86.
- [3] Jolliffe I. *Principal Component Analysis* [M]. Wiley Online Library, 2005.
- [4] Makhoul J, Roucos S, Gish H. Vector quantization in speech coding [J]. *Proceedings of the IEEE*, 1985, 73(11): 1551-1588.
- [5] Izenman AJ. *Linear discriminant analysis* [M] // *Modern Multivariate Statistical Techniques*, Springer, 2013: 237-280.
- [6] Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization [J]. *Nature*, 1999, 401(6755): 788-791.
- [7] Logothetis NK, Sheinberg DL. Visual object recognition [J]. *Annual Review of Neuroscience*, 1996, 19(1): 577-621.
- [8] Zhang XC, Zong LL, Liu XY, et al. Constrained clustering with nonnegative matrix factorization [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, 27(7): 1514-1526.
- [9] Babae M, Tsoukalas S, Babae M, et al. Discriminative nonnegative matrix factorization for dimensionality reduction [J]. *Neurocomputing*, 2016, 173: 212-223.
- [10] Sun FM, Xu MX, Hu XK, et al. Graph regularized and sparse nonnegative matrix factorization with hard constraints for data representation [J]. *Neurocomputing*, 2016, 173: 233-244.
- [11] Gao HC, Nie FP, Huang H. Local centroids structured non-negative matrix factorization [C] // *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.
- [12] Zhao HD, Ding ZM, Fu Y. Multi-view clustering via deep matrix factorization [C] // *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.
- [13] Trigeorgis G, Bousmalis K, Zafeiriou S, et al. A deep matrix factorization method for learning attribute representations [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 39(3): 417-429.
- [14] Li ZC, Tang JH, He XF. Robust structured nonnegative matrix factorization for image representation [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, 29(5): 1947-1960.
- [15] Huang SD, Wang HJ, Li T, et al. Robust graph regularized nonnegative matrix factorization for clustering [J]. *Data Mining and Knowledge Discovery*, 2018, 32(2): 483-503.
- [16] Zhang LF, Zhang LP, Du B, et al. Hyperspectral image unsupervised classification by robust manifold matrix factorization [J]. *Information Sciences*, 2019, 485: 154-169.
- [17] Berry MW, Browne M, Langville AN, et al. Algorithms and applications for approximate nonnegative matrix factorization [J]. *Computational*

- Statistics and Data Analysis, 2007, 52(1): 155-173.
- [18] Konstantinos D, Scott R, Ruairi de F, et al. Analysis of financial data using non-negative matrix factorization [J]. International Mathematical Forum, 2008, 3(37): 1853-1870.
- [19] Sra S, Dhillon IS. Non-negative Matrix Approximation: Algorithms and Applications [M]. University of Texas at Austin, 2006.
- [20] Kim H, Park H. Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis [J]. Bioinformatics, 2007, 23(12): 1495-1502.
- [21] Cichocki A, Lee HY, Kim YD, et al. Non-negative matrix factorization with α -divergence [J]. Pattern Recognition Letters, 2008, 29(9): 1433-1440.
- [22] Cai D, He XF, Han JW, et al. Graph regularized nonnegative matrix factorization for data representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(8): 1548-1560.
- [23] Wang F, Zhang CS. Label propagation through linear neighborhoods [J]. IEEE Transactions on Knowledge and Data Engineering, 2007, 20(1): 55-67.
- [24] Lu CY, Min H, Zhao ZQ, et al. Robust and efficient subspace segmentation via least squares regression [C] // Proceedings of European Conference on Computer Vision, 2012: 347-360.
- [25] Boyd S, Vandenberghe L. Convex Optimization [M]. Cambridge University Press, 2004.
- [26] Peng X, Yi Z, Tang HJ. Robust subspace clustering via thresholding ridge regression [C] // Proceedings of AAAI Conference on Artificial Intelligence, 2015: 3827-3833.
- [27] Papadimitriou CH, Kenneth S. Combinatorial Optimization: Algorithms and Complexity [M]. Prentice Hall Inc., 1982.