【电子与信息科学 / Electronics and Information Science】

自适应稀疏表示引导的无监督降维

岳 琴¹, 魏 巍^{1,2}, 冯 凯^{1,2}, 崔军彪¹

1) 山西大学计算机与信息技术学院, 山西太原 030006;

2) 山西大学计算智能与中文信息处理教育部重点实验室, 山西太原 030006

摘 要: 挖掘并保持数据分布信息是无监督降维的核心问题, 为解决传统无监督降维方法大多数只考虑数据分布的局部信息或者全局信息, 数据分布信息在低维空间难以保持的缺点, 提出一种同时考虑数据分布的全局和局部信息的自适应稀疏表示引导的无监督降维 (adaptive sparse representation guided unsupervised dimensionality reduction, ASR_UDR) 方法. 用稀疏表示挖掘高维空间数据分布的全局信息, 通过约束投影后的数据保持图上的平滑性, 挖掘数据分布的局部信息, 并将这两个过程统一到一个框架中, 使之相辅相成, 实现数据分布信息的自适应挖掘和数据降维. 在 WarpAR10P、USPS、MultiB、DLBCLA 和 DLB-CLB 数据集上的实验结果表明, 与已有的同类无监督降维方法相比, 所提方法在显著减少数据维数的同时, 可更好地提升后续学习算法的性能.

关键词: 人工智能; 稀疏表示; 无监督学习; 维度约简; 机器学习; 数据挖掘

中图分类号: TP181

文献标志码: A

doi: 10.3724/SP.J.1249.2020.04425

Adaptive sparse representation guided unsupervised dimensionality reduction

YUE Qin¹, WEI Wei^{1,2}, FENG Kai^{1,2}, and CUI Junbiao¹

1) School of Computer and Information Technology, Shanxi University, Taiyuan 030006, Shanxi Province, P. R. China

2) Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan 030006, Shanxi Province, P. R. China

Abstract: How to mine and preserve data distribution information is the core problem of unsupervised dimensionality reduction. Most of the traditional unsupervised dimensionality reduction methods only consider the local information or global information of data distribution, and the data distribution information is difficult to be preserved in the low dimensional space. To solve this problem, we propose an adaptive sparse representation guided unsupervised dimensionality reduction method to consider the global and local information of the data distribution simultaneously. In this method, the sparse representation is used to mine the global information of high-dimensional data distribution, and the graph smoothness is preserved to mine the local information of data distribution by constraining the projected data during the projection process, in which the graph is represented by the sparse representation coefficient matrix. These two processes are integrated into a framework in order to achieve the mutual guidance of mining information of data distribution and unsupervised dimensionality reduction. The experimental results on the data sets WarpAR10P,

Received: 2019-04-08; **Revised:** 2020-01-11; **Accepted:** 2020-02-08

Foundation: National Natural Science Foundation of China (61772323, 61502286); Scientific and Technological Innovation Programs of Higher Education Institutions of Shanxi Province (2016111); Science and Technology Research Foundation for Youths of Shanxi Province (201701D221099)

Corresponding author: Professor WEI Wei. E-mail: weiwei@sxu.edu.cn

Citation: YUE Qin, WEI Wei, FENG Kai, et al. Adaptive sparse representation guided unsupervised dimensionality reduction [J]. Journal of Shenzhen University Science and Engineering, 2020, 37(4): 425-432. (in Chinese)

USPS, MultiB, DLBCLA and DLBCLB show that compared with the related unsupervised dimensionality reduction methods, the proposed method effectively improves the performance of subsequent learning algorithm meanwhile significantly reducing the data dimensionality.

Key words: artificial intelligence; sparse representation; unsupervised learning; dimensionality reduction; machine learning; data mining

随着数据维数的增加,传统机器学习算法的性能及可解释性都受到了严重影响.降维是缓解“维度灾难”^[1]的一种有效手段.根据降维过程中所使用监督信息的多少,降维可被分为有监督降维^[2-4]、半监督降维^[5-7]和无监督降维.无监督降维指不利用任何监督信息的降维,它通常以保持某种数据分布信息(如几何信息和统计信息等)为准则^[8].然而,在高维场景中,如何有效挖掘数据分布信息是非常困难的.因此,相比其他两种降维方法,无监督降维更具挑战性.根据保持的数据分布信息的不同,无监督降维又可分为保持数据分布的局部信息降维和保持数据分布的全局信息降维两种.经典的保持数据分布局部信息的无监督降维方法有局部线性嵌入(locally linear embedding, LLE)^[9]和局部保持投影(locality preserving projection, LPP)^[10]等. LLE保持的是局部线性重构关系,在原始高维空间学习近邻线性重构系数,在低维空间中保持学到的线性重构关系. LPP保持的是样本间的局部相似性.由于LPP的相似性图是预先定义的,所以降维的结果很大程度上依赖于图的构建,而图构建本身是一个开放性问题,自适应图构建^[11]是解决此问题的有效方法.图优化局部保持投影(graph-optimized locality preserving projections, GOLPP)^[12]将图构建和降维任务统一到一个框架中,实现了图构建和降维相互指导,使得到的图对于具体任务是最优的.与 GOLPP不同,自适应近邻投影聚类(projected clustering with adaptive neighbors, PCAN)^[13]保持的是概率近邻图,若样本对是近邻的概率大,则该样本对降维后的距离近.经典的保持数据分布全局信息的无监督降维方法有多维缩放(multiple dimensional scaling, MDS)^[14]、等度量映射(isometric mapping, IOSMAP)^[15]和稀疏保持投影(sparsity preserving projections, SPP)^[16]等. MDS保持降维前后样本间的欧式距离不变; IOSMAP保持降维前后样本间的测地距离不变; SPP通过稀疏表示挖掘数据分布的全局信息,在低维空间中保持学到的全局线性重构关系.为挖掘更多数据分布信息, QI等^[17]提出基于自表示和邻接学习的维度约简(dimensionality

reduction via representation and affinity learning, DRRAL),采用文献[18]的模型进行邻接矩阵的学习并指导降维. HINTON等^[19]利用神经网络学习高维数据的低维表示(自编码器)提出基于神经网络的维度约简方法.为学到更鲁棒的特征, ABAVISANI等^[20]提出基于自动编码模型的稀疏深度编码.

在实际应用中,高维数据的潜在分布往往是复杂的,单独采用全局或局部的方法很难完全捕获数据的分布信息^[21].为同时挖掘数据分布的全局信息和局部信息,本研究提出一种自适应挖掘数据分布的局部和全局信息的无监督降维(adaptive sparse representation guided unsupervised dimensionality reduction, ASR_UDR)方法,用稀疏表示挖掘数据的类簇结构,再将稀疏表示的系数约束为凸组合,可直观地刻画样本间的概率近邻关系,并在投影后的低维数据中保持该关系,最后将上述两个过程统一到一个框架中,让二者相互指导,实现数据分布信息的自适应挖掘与数据降维.

1 基础知识

1.1 符号描述

给定数据集 $X = (x_i) \in \mathbf{R}^{D \times n}$, x_i 为第 i 个样本的第 j 个特征. 降维后的数据集为 $Y = (y_i) \in \mathbf{R}^{d \times n}$, $d < D$, 即将 D 维数据降为 d 维数据. $Y = P^T X$, 其中, $P \in \mathbf{R}^{D \times d}$ 是高维数据到低维数据的投影矩阵, y_i 为降维后的第 i 个样本.

1.2 稀疏表示与稀疏子空间聚类

稀疏表示指用所有样本的线性组合重构第 i 个样本,在重构系数 s_i 上加 l_0 范数正则项,要求系数是稀疏的.在线性子空间中, s_i 中的非零元素表示与样本 x_i 在同一子空间中的样本在重构该样本时的贡献,表达式为

$$\min_{s_i} \|s_i\|_0, \quad \text{s. t. } x_i = Xs_i \quad (1)$$

式(1)的求解是一个非确定性多项式困难(nondeterministic polynomially hard, NP-hard)问题,

一般用 l_1 范数近似 l_0 范数, 则式(1)重写为

$$\min_{s_i} \|s_i\|_1, \quad \text{s. t. } x_i = Xs_i \quad (2)$$

将稀疏表示的重构系数用作邻接矩阵系数, 此时当子空间是相互独立时, 邻接矩阵 S 呈稀疏块对角结构, 每一块对应一个子空间.

稀疏子空间聚类假设子空间与数据的类簇结构是一一对应的, 因此可用稀疏表示发现子空间, 从而发现数据的簇结构, 得到聚类结果. 使用稀疏表示得到 S 后, 用 $A = |S| + |S|^T$ 定义邻接图, 得到拉普拉斯矩阵 $L = D - A$. 其中, D 为度矩阵, 其元素 $d_{ii} = \sum_{j=1}^n a_{ij}$, a_{ij} 为 A 的矩阵元素. 然后, 用 k -means 算法对 L 的前 c 个最小特征值对应的特征向量进行聚类.

1.3 局部保持投影

局部保持投影目的是挖掘高维空间中的数据邻域信息并在低维数据空间得以保持^[10]. 其中, 高维数据的邻域信息是通过邻接矩阵刻画的. 算法步骤为:

1) 构建邻接矩阵 $S = (s_{ij}) \in \mathbf{R}^{n \times n}$. 若两个样本 x_i 与 x_j 相近, 则节点 i 和节点 j 之间用边相连, 且不同样本间边的权重 s_{ij} 不同. 具体构建方法为

$$s_{ij} = \begin{cases} e^{-\frac{\|x_i - x_j\|^2}{t}}, & x_j \in N_k(x_i) \text{ 或 } x_i \in N_k(x_j) \\ 0, & \text{其他} \end{cases} \quad (3)$$

其中, $N_k(x_i)$ 为 x_i 的 k 近邻集合.

2) 特征映射. 在投影后的低维空间中若要保持高维空间中数据的邻域信息, 则目标函数为

$$\sum_{i=1}^n \sum_{j=1}^n \|x_i - x_j\|_2^2 s_{ij} \quad (4)$$

最小化后等价于

$$\min_P P^T X L X^T P \quad (5)$$

其中, $L = D - S$, D 为度矩阵, 其元素 $d_{ii} = \sum_{j=1}^n s_{ij}$, D 提供了一种天然的度量, d_{ii} 越大, 对应的 y_i 越重要. 因此, 添加约束条件 $y^T D y = 1 \Rightarrow P^T X D X^T P = 1$ 后, 局部保持投影最终的优化问题为

$$\min_P P^T X L X^T P, \quad \text{s. t. } P^T X D X^T P = 1 \quad (6)$$

2 稀疏表示引导的无监督降维

2.1 模型构建

降维目的是在保持数据分布信息的前提下减少

数据维度: 一方面, 将稀疏子空间聚类思想用于维度约简, 用子空间学习挖掘到的数据分布的全局信息指导降维; 另一方面, 在降维中引入局部保持投影, 并将子空间学习和降维统一到一个框架中, 完成自适应地挖掘数据分布的全局信息和局部信息与降维相互指导的过程. 目标函数为

$$J(P, S) = \sum_{i=1}^n (\|x_i - Xs_i\|_2^2 + \alpha \|s_i\|_1) + \beta \sum_{i=1}^n \sum_{j=1}^n (\|P^T x_i - P^T x_j\|_2^2 s_{ij}) \quad (7)$$

其中, α 为稀疏表示中重构误差和稀疏正则项的平衡参数; β 为用于平衡全局和局部信息的参数. 目标函数中的第1项是稀疏表示的目标函数, 用来挖掘数据分布的全局信息; 第2项是用稀疏表示的系数矩阵构建邻接图, 使降维后的样本保持该图上的平滑性, 即在图上相似的样本降维后也相似. 为避免平凡解, 增加约束项 $P^T X H X^T P = I$, 约束降维后的特征与特征之间线性无关. 其中, I 为单位矩阵; $H = I - (1/n) F F^T$ 是中心化矩阵, F 为元素全为1的 $n \times 1$ 维矩阵. 为使稀疏表示的系数矩阵能更好地反映样本之间的相似性, 采用除样本以外的其余样本的凸组合重构目标样本. 凸组合系数具有天然的概率意义, 稀疏性能直观地反映数据分布的局部信息. 在降维过程中, 若样本对 (x_i, x_j) 是近邻的概率大, 则希望降维后的样本对 (y_i, y_j) 近. 同时, 降维后的样本也指导相似性矩阵的学习, 即若 (y_i, y_j) 远, 则希望该样本对的相似性小. 所以, 最终的优化模型为

$$\begin{aligned} \min_{P, S} J(P, S) &= \sum_{i=1}^n (\|x_i - Xs_i\|_2^2 + \alpha \|s_i\|_1) + \\ &\quad \beta \sum_{i=1}^n \sum_{j=1}^n (\|P^T x_i - P^T x_j\|_2^2 s_{ij}) \\ \text{s. t. } P^T X H X^T P &= I \end{aligned} \quad (8)$$

$$s_{ii} = 0, \quad i = 1, 2, \dots, n$$

$$s_{ij} \geq 0, \quad i, j = 1, 2, \dots, n$$

$$\sum_{i=1}^n s_{ij} = 1, \quad j = 1, 2, \dots, n$$

2.2 模型的求解

采用交替优化方法求解优化问题(8). 首先, 固定 S , 优化 P , 则优化问题可写为

$$\begin{aligned} \min_P J(P) &= \beta \sum_{i=1}^n \sum_{j=1}^n (\|P^T x_i - P^T x_j\|_2^2 s_{ij}) \\ \text{s. t. } P^T X H X^T P &= I \end{aligned} \quad (9)$$

式(9)可等价于

$$\begin{aligned} \min_P J(P) &= \text{tr}[(P^T X) L (P^T X)^T] \\ \text{s. t. } P^T X H X^T P &= I \end{aligned} \quad (10)$$

上述优化问题对应一个广义特征值分解问题,

$$X L X^T p = \lambda X H X^T p \quad (11)$$

其中, p 为特征值 λ 对应的特征向量.

式(11)的最优解 P^* 为由 d 个最小的广义特征值对应的特征向量组成的矩阵, 具体的构造过程可参考文献[10].

其次, 固定 P , 优化 S , 则式(8)可写为

$$\begin{aligned} \min_S J(S) &= \sum_{i=1}^n (\|x_i - X s_i\|_2^2 + \alpha \|s_i\|_1) + \\ &\quad \beta \sum_{i=1}^n \sum_{j=1}^n (\|y_i - y_j\|_2^2 s_{ij}) \\ \text{s. t. } s_{ii} &= 0, i = 1, 2, \dots, n \\ s_{ij} &\geq 0, i, j = 1, 2, \dots, n \end{aligned} \quad (12)$$

其中, $y_i = P^T x_i$ 是第 i 个样本的低维表示.

令 $W = (w_{ij}) \in \mathbf{R}^{n \times n}$, $w_{ij} = \|y_i - y_j\|_2^2$, $w_i \in \mathbf{R}^{n \times 1}$ 为 W 的第 i 列, 则优化问题(12)可写为

$$\begin{aligned} \min_S J(S) &= \sum_{i=1}^n (\|x_i - X s_i\|_2^2 + \alpha \|s_i\|_1) + \\ &\quad \beta \sum_{i=1}^n \sum_{j=1}^n w_{ij} s_{ij} \\ \text{s. t. } s_{ii} &= 0, i = 1, 2, \dots, n \\ s_{ij} &\geq 0, i, j = 1, 2, \dots, n \\ \sum_{j=1}^n s_{ji} &= 1, j = 1, 2, \dots, n \end{aligned} \quad (13)$$

对目标函数进行化简, 可得

$$\begin{aligned} J(S) &= \sum_{i=1}^n (s_i^T X^T X s_i) + \sum_{i=1}^n x_i^T x_i + \\ &\quad \sum_{i=1}^n (\alpha F^T + \beta w_i^T - 2 x_i^T X) s_i \end{aligned} \quad (14)$$

其中, F 为元素全为 1 的 $n \times 1$ 维矩阵.

由式(14)可见, S 的每列都相互独立, 可单独进行优化. 对于 S 的第 i 列 s_i , 去掉与目标函数中与 s_i 无关的常量后可得

$$\begin{aligned} \min_{s_i} J(s_i) &= (s_i^T X^T X s_i) + \\ &\quad (\alpha F^T + \beta w_i^T - 2 x_i^T X) s_i \\ \text{s. t. } s_{ii} &= 0 \\ s_{ij} &\geq 0, j = 1, 2, \dots, n \\ \sum_{j=1}^n s_{ji} &= 1 \end{aligned} \quad (15)$$

优化问题(15)是一个凸的二次规划问题, 可采用求解二次规划的算法^[22]进行求解.

2.3 算法的描述

自适应稀疏表示引导的无监督降维算法描述见图 1, 算法收敛准则设为连续两次迭代目标函数值的差的绝对值小于 1×10^{-5} .

输入: 数据矩阵 $X \in \mathbf{R}^{D \times n}$, 超参数 α 和 β , 降维后数据的维数 d
输出: 数据的低维表示 $Y \in \mathbf{R}^{d \times n}$
1) 初始化 $S = 0$;
2) 迭代优化各变量
while 不满足收敛条件
求解式(10)对应的广义特征值问题更新变量 P ;
求解式(15)对应的凸二次规划问题更新矩阵 S 的每一列;
end while
3) 计算: $Y = P^T X$
结束

图 1 自适应稀疏表示引导的无监督降维算法描述

Fig. 1 The algorithm description of adaptive sparse representation guided unsupervised dimensionality reduction

3 实验

3.1 数据集

实验共使用 5 个数据集, 基本信息见表 1. 其中, WarpAR10P 为人脸图像数据集; USPS 为手写字体数据集; MultiB、DLBCLA 和 DLBCLB 为 3 个基因数据集.

表 1 实验所用数据集基本信息

Table 1 The information of data sets used in this experiment

数据集	样本数	特征数	类别数
WarpAR10P	210	2 420	10
USPS	200	256	10
DLBCLA	141	661	3
DLBCLB	180	481	3
MultiB	32	3 355	4

3.2 实验设置

对比算法包括原始的高维数据 (baseline) 和 5

种经典的无监督降维算法 GOLPP^[12]、PCAN^[13]、DRRAL^[17]、SPP^[16]和全局和局部保持投影(global and local structure preserving projection, GLSPP)^[23]. 其中, GOLPP 和 PCAN 是只考虑数据分布局部信息的无监督降维算法; DRRAL 和 SPP 是只考虑数据分布全局信息的无监督降维算法; GLSPP 算法同时考虑了数据的全局信息和局部信息. 在实验中, 各种算法的参数设置如下:

GOLPP 算法是对 LPP^[10] 算法的改进. 在 LPP 算法中, 样本之间的相似性矩阵是预定义的, 缺乏自适应能力, 为此, GOLPP 算法将相似性矩阵的学习与降维过程统一在一个框架中, 相互指导. 该方法需初始化邻接矩阵 S , 本实验采用文献[12]的初始化方式:

$$s_{ij} = \begin{cases} e^{-\frac{\|x_i - x_j\|^2}{t}}, & x_j \in N_k(x_i) \text{ 或 } x_i \in N_k(x_j) \\ 0, & \text{其他} \end{cases} \quad (16)$$

其中, $N_k(x_i)$ 为第 i 个样本的 k 近邻, $k = 1, 2, \dots, l-1$, l 为每个类别样本的数量; t 为高斯核的参数,

本实验取 $t = \frac{1}{n} \sum_{i=1}^n \|x_i\|_2$.

PCAN 算法的主要思想是利用欧式距离指导概率近邻的学习, 从而指导降维. 在实验中, 类簇的个数设置为数据集真实的类别个数, 另外两个参数采用文献[13]中的设置.

DRRAL 是基于自表示的无监督降维方法, 算

法分两个阶段: ① 用文献[18]算法得到相似性矩阵, 超参数 $\lambda_1 = \lambda_2 = \lambda_3 = \lambda \in \{0.1, 0.2, \dots, 1.0\}$, 簇的个数设为数据集的真实类别数; ② 将得到的相似性矩阵用于指导降维.

SPP 算法是用稀疏表示学习重构系数关系, 在低维数据中保持该重构系数关系. 本实验采用文献[16]中的优化问题(16)进行稀疏重构.

GLSPP 算法利用 k -means 聚类发现数据的标签信息, 从全局角度和局部两个角度保持该信息. 其中, 用于平衡全局和局部信息的超参数 $\beta \in \{0.01, 0.1, 1, 10, 100\}$. 本研究超参数集合 $\alpha \in \{0.001, 0.005, 0.010, 0.050, 0.100\}$, $\beta \in \{0.001, 0.01, 0.1, 1, 10, 100, 1000\}$.

除 baseline 算法外, 其他算法降维后的维数 $d \in \{50, 60, \dots, 120\}$. 首先, 在所得数据的低维表示数据集上执行 k -means 聚类, 类的个数为真实类别数, 类中心采用随机方法进行初始化. 然后, 计算 k -means 聚类得到的类标签向量与真实类标签向量的聚类精度(accuracy, ACC)和归一化互信息(normalized mutual information, NMI). 为降低 k -means 聚类的随机性, 此过程重复 50 次, 再求平均值, 以此来衡量降维结果的质量.

3.3 结果及分析

表 2 和表 3 分别展示了 7 种方法在 5 个数据集上所有维度 $d \in \{50, 60, \dots, 120\}$ 中的最优 ACC 值和 NMI 值, 以及对于对应的维度.

表 2 7 种方法在 5 个数据集上的聚类精度比较¹⁾ ($\bar{x} \pm s$)
Table 2 The results of clustering accuracy on all data sets ($\bar{x} \pm s$)

数据集	baseline	GOLPP	PCAN	DRRAL	SPP	GLSPP	ASR_UDR
WarpAR10P	17.86 ± 3.10 (2 420)	22.37 ± 1.88 (70)	23.80 ± 2.64 (50)	23.94 ± 1.83 (70)	19.91 ± 1.78 (100)	21.05 ± 3.43 (50)	28.37 ± 1.23 (90)
USPS	59.14 ± 10.83 (256)	37.00 ± 4.29 (120)	43.20 ± 6.34 (90)	38.28 ± 4.04 (60)	20.04 ± 1.91 (50)	63.52 ± 10.41 (80)	71.72 ± 4.44 (90)
DLBCLA	54.21 ± 1.28 (661)	59.39 ± 1.22 (110)	58.44 ± 3.46 (110)	47.31 ± 0.85 (60)	40.53 ± 1.73 (80)	54.88 ± 4.22 (50)	62.41 ± 0 (50)
DLBCLB	61.03 ± 15.62 (481)	41.32 ± 4.79 (70)	68.24 ± 12.73 (110)	44.30 ± 7.03 (120)	36.28 ± 6.08 (110)	67.38 ± 11.40 (50)	74.17 ± 0.84 (80)
MultiB	50.75 ± 8.48 (3 355)	63.56 ± 9.57 (90)	44.56 ± 4.63 (120)	43.75 ± 0 (70)	39.31 ± 1.56 (100)	53.57 ± 7.99 (90)	66.63 ± 1.47 (110)

¹⁾ 括号内数值为对应的维数

表 3 7 种方法在 5 个数据集上的归一化互信息比较¹⁾ ($\bar{x} \pm s$)Table 3 The results of normalized mutual information on all data sets ($\bar{x} \pm s$)

%

数据集	baseline	GOLPP	PCAN	DRRAL	SPP	GLSPP	ASR_UDR
WarpAR10P	17.19 $\pm$ 4.42 (2 420)	17.62 $\pm$ 1.04 (120)	24.04 $\pm$ 2.24 (120)	22.93 $\pm$ 0.98 (60)	17.61 $\pm$ 0.88 (60)	22.66 $\pm$ 3.29 (90)	27.07 $\pm$ 0.80 (120)
USPS	66.53 $\pm$ 3.12 (256)	38.70 $\pm$ 1.25 (120)	43.25 $\pm$ 2.09 (90)	34.59 $\pm$ 1.81 (50)	12.52 $\pm$ 1.31 (50)	68.15 $\pm$ 2.59 (110)	74.35 $\pm$ 1.52 (110)
DLBCLA	20.72 $\pm$ 1.42 (661)	35.89 $\pm$ 1.61 (120)	26.59 $\pm$ 4.74 (110)	21.44 $\pm$ 0 (70)	2.93 $\pm$ 0.68 (90)	22.02 $\pm$ 4.21 (50)	35.90 $\pm$ 0 (90)
DLBCLB	40.47 $\pm$ 4.43 (481)	6.418 $\pm$ 0.76 (110)	41.22 $\pm$ 2.71 (70)	12.09 $\pm$ 0.75 (100)	0.92 $\pm$ 0 (50)	40.93 $\pm$ 4.54 (50)	43.70 $\pm$ 0.03 (80)
MultiB	43.53 $\pm$ 8.14 (3 355)	47.43 $\pm$ 9.31 (100)	39.16 $\pm$ 5.00 (50)	16.17 $\pm$ 0.83 (100)	10.54 $\pm$ 0 (50)	46.00 $\pm$ 6.36 (90)	47.44 $\pm$ 3.68 (60)

¹⁾ 括号内数值为对应的维数

由表 2 可见,与其他方法相比,本研究方法在 5 个数据集上的 ACC 值都有提升,升幅 3.02% ~ 8.20%。由表 3 可见,本研究方法的 NMI 值比其他方法在 5 个数据集上都有提升,升幅为 0.01% ~ 6.20%。这是由于本研究方法利用稀疏表示挖掘数据分布的全局信息,约束降维后的样本保持该信

息,同时约束系数矩阵是凸组合,进而有概率近邻的含义,以此挖掘数据分布的局部信息,而且将数据分布的挖掘和降维统一到一个框架中,相互指导,自适应得到数据的低维表示。

3.4 参数敏感性分析

图 2 和图 3 分别展示了 5 个数据集在不同 α 和

图 2 ASR_UDR 方法在 5 个数据集上不同参数下的聚类 ACC

Fig. 2 (Color online) Clustering ACC on five data sets with different parameters for ASR_UDR

图3 ASR_UDR方法在5个数据集上不同参数下的聚类NMI

Fig. 3 (Color online) Clustering NMI on five data sets with different parameters for ASR_UDR

β 值下, 采用 ASR_UDR 算法得到的聚类 ACC 和 NMI 值实验结果. 由图 2 可见, ASR_UDR 算法的 ACC 指标对参数 α 和 β 不是很敏感, 且当 α 相同时, 算法的 ACC 指标对 β 不敏感; 当 β 相同时, 算法对 α 较为敏感. 由图 3 可见, 相比聚类 ACC 指标, ASR_UDR 算法的 NMI 指标对 α 和 β 较敏感, 且当 α 相同时, 算法的 NMI 指标对 β 不敏感; 当 β 相同时, 算法对 α 较为敏感. 在少量参数组合下, 算法可达到较高性能.

结 语

挖掘并保持数据的分布信息是无监督降维的关键问题. 本研究用稀疏表示挖掘高维数据的子空间结构用于指导无监督降维, 同时用无监督降维的结构进一步指导稀疏子空间的学习, 二者相互提升从而自适应地挖掘数据分布的全局和局部信息并完成降维. 在大量真实高维数据集上的实验结果表明, 本研究方法可在显著降低数据维数的同时, 有效提升后续学习方法的性能.

基金项目: 国家自然科学基金资助项目 (61772323, 61502286); 山西省高等教育机构科技创新资助项目 (2016111); 山西省青年科技研究基金资助项目 (201701D221099)

作者简介: 岳琴 (1995—), 山西大学硕士研究生. 研究方向: 机器学习、数据挖掘. E-mail: 993203718@qq.com

引 文: 岳琴, 魏巍, 冯凯, 等. 自适应稀疏表示引导的无监督降维[J]. 深圳大学学报理工版, 2020, 37(4): 425-432.

参考文献 / References:

- [1] SCOTT D W. Multivariate density estimation: theory, practice, and visualization, second edition [M]. 2nd ed. Hoboken, USA: John Wiley & Sons Inc, 2015: 195-217.
- [2] FISHER R A. The use of multiple measurements in taxonomic problems [J]. Annals of Eugenics, 1936, 7(2): 179-188.
- [3] FLAMARY R, CUTURI M, COURTY N, et al. Wasserstein discriminant analysis [J]. Machine Learning, 2018, 107(12): 1923-1945.
- [4] ÖRNEK C, VURAL E. Nonlinear supervised dimensionality reduction via smooth regular embeddings [J]. Pattern Recognition, 2019, 87: 55-66.

- [5] CHEN Puhua, JIAO Licheng, LIU Fang, et al. Semi-supervised double sparse graphs based discriminant analysis for dimensionality reduction [J]. Pattern Recognition, 2017, 61: 361-378.
- [6] WANG Sheng, LU Jianfeng, GU Xingjian, et al. Semi-supervised linear discriminant analysis for dimension reduction and classification [J]. Pattern Recognition, 2016, 57(2016): 179-189.
- [7] SANODIYA R K, SAHA S, MATHEW J. Semi-supervised orthogonal discriminant analysis with relative distance; integration with a MOO approach [J]. Soft Computing, 2020, 24(3): 1599-1618.
- [8] YAN Shuicheng, XU Dong, ZHANG Benyun, et al. Graph embedding and extensions: a general framework for dimensionality reduction [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(1): 40-51.
- [9] ROWEIS S T, SAUL L K. Nonlinear dimensionality reduction by locally linear embedding [J]. Science, 2000, 290(5500): 2323-2326.
- [10] HE Xiaofei, NIYOGI P. Locality preserving projections [C] // Advances in Neural Information Processing Systems. Vancouver, Canada: The MIT Press, 2004: 153-160.
- [11] WEN Jie, XU Yong, LIU Hong. Incomplete multi-view spectral clustering with adaptive graph learning [J]. IEEE Transactions on Cybernetics. 2020, 50(4): 1418-1429.
- [12] ZHANG Limei, QIAO Lishan, CHEN Songcan. Graph-optimized locality preserving projections [J]. Pattern Recognition, 2010, 43(6): 1993-2002.
- [13] NIE Feiping, WANG Xiaoqian, HUANG Heng. Clustering and projected clustering with adaptive neighbors [C] // Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM, 2014: 977-986.
- [14] COX M A, COX T F. Multidimensional scaling [M] // Handbook of data visualization. CHEN C H. HÄROLE W, UNWIN A. Berlin: Springer, 2008: 315-347.
- [15] TENENBAUM J B, De SILVA V, LANGFORD J C. A global geometric framework for nonlinear dimensionality reduction [J]. Science, 2000, 290(5500): 2319-2323.
- [16] QIAO Lishan, CHEN Songcan, TAN Xiaoyang. Sparsity preserving projections with applications to face recognition [J]. Pattern Recognition, 2010, 43(1): 331-341.
- [17] QI Miao, LU Shuang, HUANG Xing, et al. Dimensionality reduction via representation and affinity learning [C] // The 4th International Conference on Systems and Informatics. Nanjing, China: IEEE, 2017: 1203-1208.
- [18] GUO Xiaojie. Robust subspace segmentation by simultaneously learning data representations and their affinity matrix [C] // Proceedings of the 24th International Joint Conference on Artificial Intelligence. Buenos Aires, Argentina: AAAI Press, 2015: 3547-3553.
- [19] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313(5786): 504-507.
- [20] ABAVISANI M, PATEL V M. Deep sparse representation-based classification [J]. IEEE Signal Processing Letters, 2019, 26(6): 948-952.
- [21] TAO Xinmin, WANG Ruotong, CHANG Rui, et al. Spectral clustering algorithm using density-sensitive distance measure with global and local consistencies [J]. Knowledge-Based Systems, 2019, 170: 26-42.
- [22] BOYD S, VANDENBERGHE L. Convex optimization [M]. Cambridge, UK: Cambridge University Press, 2004.
- [23] CAI Weiling. A dimension reduction algorithm preserving both global and local clustering structure [J]. Knowledge-Based Systems, 2017, 118: 191-203.

【中文责编：英 子；英文责编：木 柯】