

属性自表达的低秩无监督属性选择算法

郑 威,文国秋*,何 威,胡荣耀,赵树之

(广西师范大学 广西多源信息挖掘与安全重点实验室,广西 桂林 541004)

摘 要:针对现有无监督属性约简方法只单一使用子空间学习或属性选择的方法,并且忽略数据之间的内在相关性,本文提出一种新的属性选择方法。首先提出一个属性自表达损失函数加上一个稀疏正则化($l_{2,1}$ -范数)实现无监督学习与属性选择。然后嵌入子空间学习方法,并使用低秩约束和图正则化项考虑数据的全局结构和局部结构。经聚类实验验证,该算法较对比算法能取得更好的效果。

关键词:低秩约束;属性选择;子空间学习;属性约简;无监督

中图分类号:TP181 **文献标志码:**A **文章编号:**1001-6600(2018)01-0061-09

引用格式:郑威,文国秋,何威,等.属性自表达的低秩无监督属性选择算法[J].广西师范大学学报(自然科学版),2018,36(1):61-69.

ZHENG Wei, WEN Guoqiu, HE Wei, et al. Low-rank unsupervised feature selection based on self-representation[J]. Journal of Guangxi Normal University(Natural Science Edition), 2018, 36(1): 61-69.

Low-rank Unsupervised Feature Selection Based on Self-representation

ZHENG Wei, WEN Guoqiu*, HE Wei, HU Rongyao, ZHAO Shuzhi

(Guangxi Key Lab of Multi-source Information Mining and Security, Guangxi Normal University, Guilin Guangxi 541004, China)

Abstract: A new method of feature selection is proposed to remedy several drawbacks with the existing unsupervised feature reduction method which only utilize the single subspace approach or feature selection method and ignores the inherent correlation within data. Specifically, a novel feature self-representation loss function is proposed to conduct unsupervised learning and feature selection by combining a sparse regularization ($l_{2,1}$ -norm). And then, subspace learning method and low rank constraint with graph regularization are embedded into the model, to take into account of the global structure and local structure of the data, respectively. Experimental results show that the presented algorithm can achieve better results than some selected comparison algorithms.

Keywords: low-rank constraint; feature selection; subspace learning; dimensionality reduction

大数据的实际应用,例如文本分类和图像检索等,经常遇到高维数据^[1-2]。高维度的属性增加数据计算难度和储存成本。此外,高维数据中往往存在噪声或冗余属性,会干扰学习模型的建立。因此,属性约

收稿日期:2017-06-20

基金项目:国家自然科学基金(61573270);国家“973”计划项目(2013CB329404);中国博士后科学基金(2015M570837);广西自然科学基金(2015GXNSFCB139011);广西研究生教育创新计划项目(XYCSZ2017064, YCSW2017065)

通信联系人:文国秋(1987—),女,广西桂林人,广西师范大学讲师。E-mail:zwgxnu@163.com

筒在高维大数据处理中受到广泛关注^[3]。

属性约简能降低数据处理时间, 获得有泛化能力并且紧实的学习模型^[4-5]。当前属性约简技术可分为子空间学习和属性选择两类^[6-7]。子空间学习是通过转换矩阵将原始数据集投影到低维子空间, 以去除冗余和噪声。常见的子空间学习方法包括主成分分析(principal component analysis, PCA)^[8]、线性判别分析(linear discriminant analysis, LDA)^[9]和局部保持投影(locality preserving projections, LPP)^[10]等。属性选择方法在原始数据集中选择一个最能表达原始特征子集, 同时消除不相关或不重要的属性。常见的属性选择方法包括 t 检验^[11]和稀疏回归^[12]等。在实际应用中, 属性选择具有解释性, 而子空间学习比属性选择效果好。然而, 通过子空间学习方法得到的属性空间不是原始的, 这导致了子空间学习的结果难以解释^[13]。为此, 本文结合属性选择和子空间学习来提升属性约简的效果和可解释性。

现实应用中常因为某些原因导致标签缺失或者获取样本标签需要较高的代价, 这使无监督属性选择受到很多关注。属性自表达通过属性间的近似表达, 使用数据矩阵代替数据标签作为响应矩阵, 从而解决无标签的机器学习问题^[14]。现实数据中噪声和冗余属性的存在通常增加矩阵的秩, 常用的解决方法是对数据进行低秩约束以去除噪声与冗余, 即只保留数据的内在重要属性, 从而增强回归和聚类的效果。

本文提出一种基于属性自表达的低秩无监督属性选择算法(low-rank unsupervised feature selection based on self-representation, LFSR), 同时考虑数据的两个方面: 低秩约束和 $l_{2,1}$ -范数稀疏正则化。与之前的属性选择和子空间学习方法不同, 本文将子空间学习嵌入到属性选择的框架中, 实现可解释的结果和稳定的模型。另外, 通过低秩结构^[15]和图拉普拉斯^[16](graph Laplacian)正则化使子空间学习同时考虑数据的全局结构和局部结构。最后通过属性自表达实现无监督学习。

1 相关理论

1.1 拉普拉斯矩阵

假设数据集形成一个无向图 $G=(V, E)$, 其中 V 为顶点的集合, 我们定义 S 为顶点之间的权重矩阵。拉普拉斯矩阵被定义为:

$$L = D_S - S. \quad (1)$$

矩阵 D_S 是一个对角矩阵, 其元素定义为:

$$d_i = \sum_{j=1}^n s_{ij}, i \in (1, 2, \dots, n). \quad (2)$$

图拉普拉斯是以样本为顶点, 样本间相似度作为边权重描述数据点之间相似关系, 能够在低维空间中保持每个样本的最近邻域, 从而保持数据的局部拓扑结构。

1.2 低秩约束

假定有样本集 $X \in \mathbf{R}^{n \times d}$, 响应矩阵 $Y \in \mathbf{R}^{n \times c}$, 回归系数矩阵 $W \in \mathbf{R}^{d \times c}$ 。为考虑不同响应之间的相关性, 对矩阵 W 进行低秩约束:

$$\min_W \|Y - XW\|_F^2, \text{ s.t. }, s = \text{rank}(W) \leq \min(d, c). \quad (3)$$

加入低秩结构的回归模型变为:

$$\min_{A, B} \|Y - XAB\|_F^2, \quad (4)$$

其中 $W = A \times B$, ($A \in \mathbf{R}^{d \times s}$, $B \in \mathbf{R}^{s \times c}$)。

使用低秩约束本质上是使用 LDA 进行子空间学习, 而 LDA 是一种全局结构学习方法, 可以在降低数据维度的情况下尽量保持全局结构信息。

1.3 属性自表达

假定样本集 $X \in \mathbf{R}^{n \times d}$, 列向量 $f_i \in \mathbf{R}^{1 \times n}$ 表示样本集 X 中第 i 列属性, 属性自表达就是对 X 中的每一个属性 f_i 都可以使用所有属性的线性组合表示: $f_i = \sum_{j=1}^d f_j w_{ji} + e_i$ 。对全部属性可写为矩阵形式:

$$X = XW + E, \quad (5)$$

其中 $W = [w_{ji}] \in \mathbf{R}^{d \times d}$ 表示系数矩阵, $E = [e_i] \in \mathbf{R}^{n \times d}$ 表示残差项。

无监督学习因没有类标签导致难以确认响应矩阵 Y 。而属性自表达通过属性间的自表达特性,使用数据矩阵 X 作为响应矩阵,建立新的自表达损失函数:

$$\min_W l(X - XW). \quad (6)$$

2 算法描述与优化

2.1 算法描述

通常使用经典的最小二乘损失函数来描述标签与属性之间的线性关系:

$$\min_W l(Y - XW - eb), \quad (7)$$

其中: $W \in \mathbf{R}^{d \times c}$ 是属性权重矩阵,用来对 X 进行属性筛选, $e \in \mathbf{R}^{n \times 1}$ 为元素全是 1 的列向量, $b \in \mathbf{R}^{1 \times c}$ 为偏差项。

为提升学习模型的效果,引入 $l_{2,1}$ -范数正则化项。 $l_{2,1}$ -范数可以导致行稀疏,能够有效地去除冗余和离群数据并避免在属性选择过程中发生维度灾难问题^[17]。 W 的 $l_{2,1}$ -范数被定义为 $\|W\|_{2,1} = \sum_i \sqrt{\sum_j w_{ij}^2}$ 。使用 $l_{2,1}$ -范数作为正则化项惩罚系数矩阵 W ,可以得到:

$$\min_{W,b} \|Y - XW - eb\|_F^2 + \beta \|W\|_{2,1}, \quad (8)$$

其中 β 为调节行稀疏正则化控制参数。

本文使用图拉普拉斯正则化和低秩约束保持数据的局部和全局的拓扑结构来提高属性选择的性能^[18]。对于给定的两个样本 x_i 和 x_j ,它们投影到子空间后的距离被它们在原始空间的距离所约束:

$$\min_W \sum_{i,j} (x_i W - x_j W)^2 s_{i,j}. \quad (9)$$

式(9)通过热核函数 $H(x_i, x_j) = \exp\left[-\frac{\|x_i - x_j\|^2}{\sigma}\right]$ ($\sigma \in \mathbf{R}^+$) 来定义样本之间的相似性 $s_{i,j} \in \mathbf{R}^{n \times n}$ 。

由式(1)和式(2)经过数学变换得到:

$$\begin{aligned} \sum_{i,j} (x_i W - x_j W)^2 s_{i,j} &= \\ \sum_i (W^T x_i^T d_{i,i} x_i W) - \sum_{i,j} (W^T x_i^T s_{i,j} x_j W) &= \\ \text{tr}(W^T X^T D X W) - \text{tr}(W^T X^T S X W) &= \\ \text{tr}(W^T X^T L X W). \end{aligned} \quad (10)$$

式(10)通过嵌入图拉普拉斯正则化来保持数据样本映射到低维空间后的近邻关系结构:

$$\min_{W,b} \|Y - XW - eb\|_F^2 + \alpha \text{tr}(W^T X^T L X W) + \beta \|W\|_{2,1}. \quad (11)$$

噪声和冗余的存在会增加数据矩阵的秩,影响数据处理的效果,引入低秩约束方法可以有效避免这类问题,由式(4)得到:

$$\min_{A,B,b} \|Y - XAB - eb\|_F^2 + \alpha \text{tr}(B^T A^T X^T L X A B) + \beta \|AB\|_{2,1}, \text{ s.t., } \text{rank}(AB) \leq \min(d, c). \quad (12)$$

针对无监督学习响应矩阵难以确认的特点,本文使用属性自表达方法,通过式(6)得到最终的目标函数:

$$\min_{A,B,b} \|X - XAB - eb\|_F^2 + \alpha \text{tr}(B^T A^T X^T L X A B) + \beta \|AB\|_{2,1}, \text{ s.t., } \text{rank}(AB) \leq \min(d, c). \quad (13)$$

本文提出的 LFSR 算法采用属性自表达方法并且在考虑数据的全局结构和局部结构的情况下进行无监督属性选择。本算法的优点有:

1) 不同于以往无监督学习中通过计算样本间相似性或者流形结构来建立响应矩阵,属性自表达方法使用数据矩阵代替响应矩阵。使用 $l_{2,1}$ -范数惩罚属性权重矩阵 W 可以去除冗余和无关属性,使目标函数能够自主选择重要属性。

2) 图拉普拉斯正则化建立包含数据邻近信息的图来保持数据的局部结构。低秩约束是建立在子空

间学习方法线性判别分析(LDA)上的,这不仅有效地移除噪声与冗余,而且考虑了所有样本间的关系(全局结构)。本算法结合了数据的全局结构与局部结构,比以往的子空间学习有更好的结果。

3) 属性选择具有可解释性和子空间学习具有稳定性已经被证明^[19],所以融入属性选择和子空间学习的框架可以得到最佳的属性约简效果。

本文的算法伪代码见算法 1。

算法 1: LFSR 算法。

输入: 训练样本 $X \in \mathbf{R}^{n \times d}$, 控制参数 α, β 。

输出: 聚类分析结果。

1. 初始化关系矩阵 S , 通过式(1) 计算拉普拉斯矩阵 L ;
2. 根据所选择的模型(13), 调用算法 2 求解全局最优解, 得到系数矩阵 $W^* = A^* B^*$;
3. 利用最优解 W 对原始数据集 X 进行属性选择, 得到的属性集作为新的属性集;
4. 对新的属性集构成的样本采用 K -means 聚类, 求出预测结果;
5. 对预测结果进行评价。

2.2 算法优化

本文算法目标函数包含 3 个变量, 优化的时候采用固定 2 个变量优化另一个变量的方法。

1) 固定 B 和 A , 更新 b 。

因为固定 B 和 A , 所以式(13) 中不含变量 b 的项可以看作常量, 可以重写式(13) 为:

$$\min_b \|X - XAB - eb\|_F^2. \quad (14)$$

对式(14) 求 b 的导数, 并使结果为 0, 可得:

$$e^T eb = e^T X - e^T XAB. \quad (15)$$

经过简单的变换得到:

$$b = \frac{1}{n}(e^T X - e^T XAB). \quad (16)$$

2) 固定 b 和 A , 更新 B 。

设 $H = I_n - \frac{1}{n}ee^T$, 其中 $I_n \in \mathbf{R}^{n \times n}$ 是单位矩阵, 而且有 $H = HH, H = H^T$ 。将式(16) 代入式(13):

$$\min_B \text{tr}(X^T H - B^T A^T X^T H)(HX - HXAB) + \alpha \text{tr}(B^T A^T X^T HLHXAB) + \beta \text{tr}(B^T A^T DAB). \quad (17)$$

对式(17) 求 B 的导数, 并使结果为 0, 可得:

$$B = (A^T(X^T HX + \alpha X^T HLHX + \beta D)A)^{-1} A^T X^T HX, \quad (18)$$

其中 $D \in \mathbf{R}^{d \times d}$ 是对角矩阵, D 中的每一个元素:

$$d_{i,i} = \frac{1}{2 \|w^i\|_2}, i = 1, 2, \dots, d, \quad (19)$$

其中 w^i 表示矩阵 $W^* = A^* B^*$ 的第 i 行, A^* 和 B^* 各自代表 A 和 B 的最优解, 而 W^* 是矩阵 W 的最优解。

3) 固定 B 和 b , 更新 A 。

将式(19) 代入式(13), 可以得到:

$$\max_A \text{tr}((A^T(X^T HX + \alpha X^T HLHX + \beta D)A)^{-1} A^T X^T HX X^T HX A). \quad (20)$$

令 $S_a = X^T HX + \alpha X^T HLHX + \beta D, S_b = X^T HX X^T HX$ 。式(20) 实际上是 LDA 的求解问题, 所以可通过特征根分解选择 r 个最大特征值和与其对应的 r 个最大特征向量来解决这个问题。算法 2 列出了式(13) 的优化过程。

算法 2: 优化求解式(13)。

输入: $X \in \mathbf{R}^{n \times d}$ 和参数 α, β , 低秩参数 r 。

输出: $A \in \mathbf{R}^{d \times r}$ 和 $B \in \mathbf{R}^{r \times c}$ 。

1. 初始化 $t = 1$;
2. 通过随机初始化矩阵 $A^{(t)}$ 和 $B^{(t)}$, 得到初始化矩阵 $W^{(t)} = A^{(t)} B^{(t)}$;

3. 初始化 $D^{(t)} = I \in \mathbf{R}^{d \times d}$;
- 重复
4. 根据式(16) 更新 $b^{(t+1)}$;
5. 根据式(18) 更新 $B^{(t+1)}$;
6. 根据式(20) 通过特征分解(SVD) 计算 $A^{(t+1)}$;
7. 更新对角矩阵 $D^{(t+1)} \in \mathbf{R}^{d \times d}$, 第 i 个对角元素由式(19) 计算得出, 其中: $w^i = [A^{(t+1)} B^{(t+1)}]^i$;
8. 更新 $t = t + 1$;
9. 直到公式(13) 收敛;
10. 输出最优解。

2.3 优化算法收敛性证明

式(13) 的正则化项是凸且非平滑的, 需要优化 A 和 B 两个变量, 本文采用一种简单有效的算法来优化该问题, 下面给出算法 1 收敛性证明的步骤。

从算法 2 的第 4 步到第 7 步, 能够得到第 t 次迭代对应的结果:

$$\langle A^{(t+1)}, B^{(t+1)} \rangle = \arg \min_{A, B} \|HX - HXAB\|_F^2 + \beta \text{tr}(B^T A^T D^{(t)} AB) + \alpha \text{tr}(B^T A^T X^T LXAB). \quad (21)$$

由式(21) 可得:

$$\begin{aligned} & \|HX - HXA^{(t+1)} B^{(t+1)}\|_F^2 + \beta \text{tr}(B^{(t+1)T} A^{(t+1)T} D^{(t)} A^{(t+1)} B^{(t+1)}) + \alpha \text{tr}(B^{(t+1)T} A^{(t+1)T} X^T LX A^{(t+1)} B^{(t+1)}) \leq \\ & \|HX - HXA^{(t)} B^{(t)}\|_F^2 + \beta \text{tr}(B^{(t)T} A^{(t)T} D^{(t)} A^{(t)} B^{(t)}) + \alpha \text{tr}(B^{(t)T} A^{(t)T} X^T LX A^{(t)} B^{(t)}). \end{aligned} \quad (22)$$

其中, 要注意 $W^{(t)} = A^{(t)} B^{(t)}$, $W^{(t+1)} = A^{(t+1)} B^{(t+1)}$ 。将式(19) 得到的对角矩阵 D 代入不等式(22) 可以得到:

$$\begin{aligned} & \|HX - HXW^{(t+1)}\|_F^2 + \beta \sum_{i=1}^m \frac{\|w^{i(t+1)}\|_2^2}{2 \|w^{i(t)}\|_2^2} + \alpha \text{tr}(W^{(t+1)T} X^T LX W^{(t+1)}) \leq \\ & \|HX - HXW^{(t)}\|_F^2 + \beta \sum_{i=1}^m \frac{\|w^{i(t)}\|_2^2}{2 \|w^{i(t)}\|_2^2} + \alpha \text{tr}(W^{(t)T} X^T LX W^{(t)}), \end{aligned} \quad (23)$$

其中 $w^{i(t)}$ 和 $w^{i(t+1)}$ 分别是 $W^{(t)}$ 和 $W^{(t+1)}$ 的第 i 行, 对于每个 i , 可以得到不等式:

$$\|w^{i(t+1)}\|_2 - \frac{\|w^{i(t+1)}\|_2^2}{2 \|w^{i(t)}\|_2^2} \leq \|w^{i(t)}\|_2 - \frac{\|w^{i(t)}\|_2^2}{2 \|w^{i(t)}\|_2^2}. \quad (24)$$

对不等式(24) 乘上控制参数 β 并且累加后可以得到:

$$\beta \sum_{i=1}^m \|w^{i(t+1)}\|_2 - \frac{\|w^{i(t+1)}\|_2^2}{2 \|w^{i(t)}\|_2^2} \leq \beta \sum_{i=1}^m \|w^{i(t)}\|_2 - \frac{\|w^{i(t)}\|_2^2}{2 \|w^{i(t)}\|_2^2}. \quad (25)$$

最后, 结合不等式(23) 和(25) 便可以得到:

$$\begin{aligned} & \|HX - HXW^{(t+1)}\|_F^2 + \beta \sum_{i=1}^m \|w^{i(t+1)}\|_2^2 + \alpha \text{tr}(W^{(t+1)T} X^T LX W^{(t+1)}) \leq \\ & \|HX - HXW^{(t)}\|_F^2 + \beta \sum_{i=1}^m \|w^{i(t)}\|_2^2 + \alpha \text{tr}(W^{(t)T} X^T LX W^{(t)}). \end{aligned} \quad (26)$$

综上所述, 再结合式(16) 和 $H = I_n - \frac{1}{n} ee^T$ 可知:

$$\begin{aligned} & \|X - XW^{(t+1)} - eb\|_F^2 + \beta \|W^{(t+1)}\|_{2,1} + \alpha \text{tr}(W^{(t+1)T} X^T LX W^{(t+1)}) \leq \\ & \|X - XW^{(t)} - eb\|_F^2 + \beta \|W^{(t)}\|_{2,1} + \alpha \text{tr}(W^{(t)T} X^T LX W^{(t)}). \end{aligned} \quad (27)$$

不等式(27) 表明算法 2 在每次的迭代中, 目标函数都是单调递减的, 所以本文提出的算法能够最终达到全局收敛的结果。

3 实验分析

为了验证本文提出属性选择方法 LFSR 在聚类应用的效率, 将属性约简后的数据使用 K -means 方法

进行聚类分析。

3.1 实验设置与评价标准

本文使用 6 个数据集对提出的属性选择算法进行性能测试,其中数据集 COIL20 来源于哥伦比亚大学图像库 (<http://featureselection.asu.edu/datasets.php>),UMIST 来自谢尔菲德大学面部数据集^[20]。USPS 来自美国邮政数据 (<http://statweb.stanford.edu/~tibs/ElemStatLearn/data.html>),ISOLET、ECOLI、MOVEMENT 均来自 UCI 数据集 (<http://archive.ics.uci.edu/ml/>)。

表 1 数据集信息统计

Tab.1 Dataset information statistics

数据集	样本数	属性数	类数
COIL20	1 440	1 024	20
ISOLET	1 560	617	26
ECOLI	336	343	8
USPS	1 000	256	10
MOVEMENT	360	90	15
UMIST	575	644	20

所有实验均在 Win10 系统下运行,使用 MATLAB2014a 软件进行编程。本文选择主成分分析 (principal component analysis, PCA)^[21]、局部保留投影 (locality preserving projection, LPP)^[10]、正则化自表达 (regularized self representation, RSR)^[14]、结构化最优图属性选择 (structured optimal graph feature selection, SOGFS)^[22] 以及未做属性选择的 K -means 作为对比算法来与本文提出的算法进行比较。本文将分别采用准确率 (accuracy)、互信息 (mutual information) 作为聚类效果的评价标准。

3.2 实验结果与分析

为证明本文提出的优化算法的有效性和高效性,在图 1 中展示出本文算法在数据集上的收敛性。从图 1 中可以明显看出,随着迭代次数的增加,算法能够快速收敛到全局最优解。图 2 中展示了参数 α 和 β 对目标函数(13)的敏感性评估。其中,参数 α 和 β 的取值范围为 $\{0.01, 0.1, 1, 10, 100, 1\ 000\}$,并且使用准确率作为度量。从图 2 可以看出,本文提出的方法在不同参数下得到不同的聚类准确率。这说明本文提出的目标函数是对参数敏感的,通过对参数的调节可以得到更好的属性选择结果。

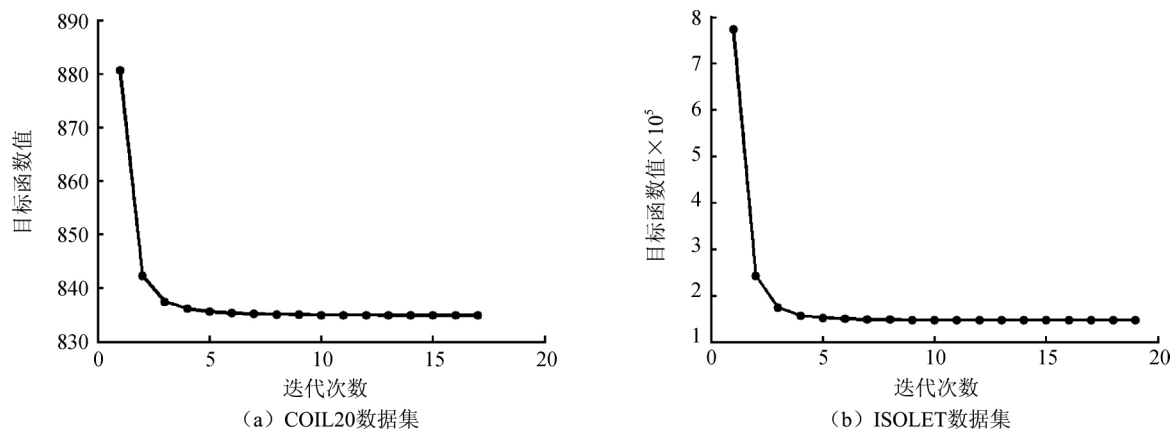


图 1 不同迭代次数下目标函数收敛性

Fig.1 Numbers of iterations under the objective function convergence diagram

表 2、3 中分别列出了每个算法在不同数据集上的准确率、互信息,最优结果用粗体标出。其中互信息是度量 2 个变量之间的相关性,能够表现预测值与真实值之间的相关性。如表 2 所示,算法 LFSR 在各个数据集上的准确率均为最高,与其他对比算法比较,最少超过 7.83 个百分点。其中,比算法 PCA 的准确率高出 17.53 个百分点,比算法 LPP 的准确率高出 15.44 个百分点,比算法 RSR 的准确率高出 12 个百分点,

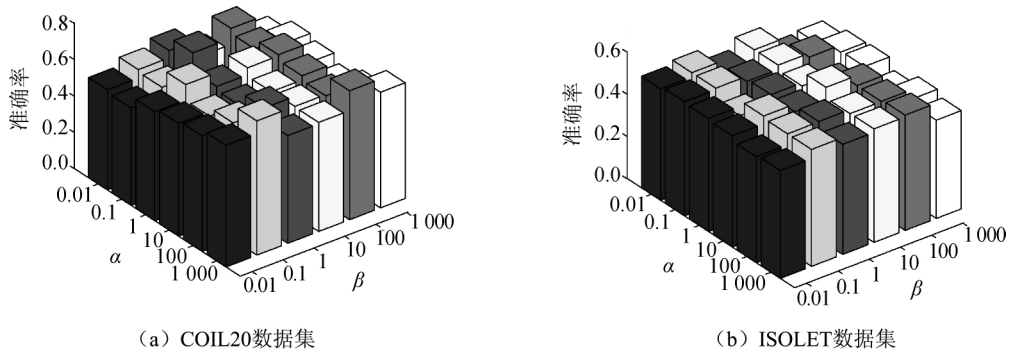


图2 不同参数下准确率的直方图

Fig.2 Accuracy of the different parameters under the histogram

比算法 SOGFS 的准确率高出 18.11 个百分点,尤其比没有进行属性选择的 K -means 算法准确率要高出 28.35 个百分点。在表 3 中可以看到,算法 LFSR 在各个数据集上互信息平均值最高,对比其他算法最高超过 11.03 个百分点,最低超过 6.3 个百分点,比没有进行属性选择的 K -means 算法高出 14.89 个百分点。原因在于算法 LFSR 在属性约简上不只单方面考虑子空间学习,而且在其基础上结合了稀疏属性选择方法,同时加入低秩约束和图学习,充分考虑了数据的整体与局部信息,能够更有效地选择主要的类判别属性和去除噪声属性,所以可以显著提高属性选择的性能。

表 2 准确率统计结果

Tab.2 Accuracy statistics

数据集	K -means	PCA	LPP	RSR	SOGFS	LFSR
COIL20	54.65	56.18	56.39	61.67	56.32	75.28
ECOLI	45.24	61.90	53.27	63.10	55.06	77.08
ISOLET	53.83	61.22	63.40	65.96	61.35	70.02
MOVEMENT	42.72	50.28	60.89	62.78	63.89	71.72
UMIST	41.39	43.13	63.13	47.30	42.61	73.22
USPS	38.06	68.10	56.30	73.20	58.10	78.70
平均值	45.98	56.80	58.89	62.33	56.22	74.33

表 3 互信息统计结果

Tab.3 Mutual information statistics

数据集	K -means	PCA	LPP	RSR	SOGFS	LFSR
COIL20	69.48	70.77	73.78	61.67	72.37	82.64
ECOLI	43.27	52.13	50.30	50.93	49.44	61.75
ISOLET	63.49	70.14	71.96	71.60	72.15	73.04
MOVEMENT	52.05	56.89	57.60	54.99	55.22	58.35
UMIST	64.61	66.04	76.51	65.68	63.18	85.77
USPS	51.32	61.60	65.76	63.10	55.03	68.05
平均值	57.37	62.92	65.96	61.32	61.23	72.26

不同数据集存在不同的数据分布以及干扰因素。实验结果表明,本文提出的 LFSR 算法在每个数据集上都表现出了更好的聚类效果,相对所有对比算法,具有最好的鲁棒性。同时,与单一属性选择算法对比,本文算法在保留提取后的重要属性的同时,利用子空间学习方法对筛选属性作进一步调整,保证数据自身的全局和局部结构,因此具有更好的解释性。

4 结束语

本文提出无监督属性选择算法 LFSR 用于聚类分析,同时将低秩约束、子空间学习、属性选择和图拉普拉斯整合到一个框架中。本算法综合了稀疏学习和子空间学习的优点,并弥补了基于稀疏学习的属性选择算法的不足。经实验验证,本文算法能够在聚类效果和稳定性上取得显著提高。在未来工作中,尝试将本文提出算法用于半监督属性选择方面。

参 考 文 献:

- [1] ZHU Xiaofeng, HUANG Zi, SHEN Hengtao, et al. Dimensionality reduction by mixed kernel canonical correlation analysis[J]. *Pattern Recognition*, 2012, 45(8): 3003-3016. DOI: 10.1016/j.patcog.2012.02.007.
- [2] ZHU Xiaofeng, ZHANG Shichao, JIN Zhi, et al. Missing value estimation for mixed-attribute data sets[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2011, 23(1): 110-121. DOI: 10.1109/TKDE.2010.99.
- [3] ZHANG Shichao, LI Xuelong, ZONG Ming, et al. Learning k for KNN classification[J]. *ACM Transactions on Intelligent Systems and Technology*, 2017, 8(3): 43. DOI: 10.1145/2990508.
- [4] ZHU Xiaofeng, LI Xuelong, ZHANG Shichao. Block-row sparse multiview multilabel learning for image classification [J]. *IEEE Transactions on Cybernetics*, 2016, 46(2): 450-461. DOI: 10.1109/TCYB.2015.2403356.
- [5] ZHU Xiaofeng, LI Xuelong, ZHANG Shichao, et al. Robust joint graph sparse coding for unsupervised spectral feature selection[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, 28(6): 1263-1275. DOI: 10.1109/TNNLS.2016.2521602.
- [6] ZHU Xiaofeng, HUANG Zi, CHENG Hong, et al. Sparse hashing for fast multimedia search[J]. *ACM Transactions on Information Systems*, 2013, 31(2): 9. DOI: 10.1145/2457465.2457469.
- [7] ZHU Xiaofeng, HUANG Zi, YANG Yang, et al. Self-taught dimensionality reduction on the high-dimensional small-sized data[J]. *Pattern Recognition*, 2013, 46(1): 215-229. DOI: 10.1016/j.patcog.2012.07.018.
- [8] PYATYKH S, HESSER J, ZHENG Lei. Image noise level estimation by principal component analysis[J]. *IEEE Transactions on Image Processing*, 2013, 22(2): 687-699. DOI: 10.1109/TIP.2012.2221728.
- [9] FAN Zizhu, XU Yong, ZHANG D. Local linear discriminant analysis framework using sample neighbors[J]. *IEEE Transactions on Neural Networks*, 2011, 22(7): 1119-1132. DOI: 10.1109/TNN.2011.2152852.
- [10] HE Xiaofei, NIYOQI P. Locality preserving projections[M]//THRUN S, SAUL L K, SCHÖLKOPF B. *Advances in Neural Information Processing Systems 16*. Cambridge, MA: MIT Press, 2004: 153-160.
- [11] KONIETSCHKE F, PAULY M. Bootstrapping and permuting paired t -test type statistics [J]. *Statistics and Computing*, 2014, 24(3): 283-296. DOI: 10.1007/s11222-012-9370-4.
- [12] LIIMATAINEN K, HEIKKILÄ R, YLI-HARJA O, et al. Sparse logistic regression and polynomial modelling for detection of artificial drainage networks[J]. *Remote Sensing Letters*, 2015, 6(4): 311-320. DOI: 10.1080/2150704X.2015.1031919.
- [13] GU Quanquan, LI Zhenhui, HAN Jiawei. Joint feature selection and subspace learning[C]// *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*. Menlo Park, CA: AAAI Press, 2011: 1294-1299. DOI: 10.5591/978-1-57735-516-8/IJCAI11-219.
- [14] ZHU Pengfei, ZUO Wangmeng, ZHANG Lei, et al. Unsupervised feature selection by regularized self-representation [J]. *Pattern Recognition*, 2015, 48(2): 438-446. DOI: 10.1016/j.patcog.2014.08.006.
- [15] CAI Xiao, DING C, NIE Feiping, et al. On the equivalent of low-rank linear regressions and linear discriminant analysis based regressions [C]// *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM Press, 2013: 1124-1132. DOI: 10.1145/2487575.2487701.
- [16] MERRIS R. Laplacian matrices of graphs: a survey[J]. *Linear Algebra and Its Applications*, 1994, 197/198: 143-176. DOI: 10.1016/0024-3795(94)90486-3.

- [17] ZHU Xiaofeng, SUK H, WANG Li, et al. A novel relational regularization feature selection method for joint regression and classification in AD diagnosis[J]. Medical Image Analysis, 2017, 38: 205-214. DOI: 10.1016/j.media.2015.10.008.
- [18] ZHU Xiaofeng, ZHANG Shichao, ZHANG Jilian, et al. Cost-sensitive imputing missing values with ordering[C]// Proceedings of the 22nd National Conference on Artificial Intelligence: Volume 2. Menlo Park, CA: AAAI Press, 2007: 1922-1923.
- [19] ZHU Xiaofeng, ZHANG Lei, HUANG Zi. A sparse embedding and least variance encoding approach to hashing[J]. IEEE Transactions on Image Processing, 2014, 23(9): 3737-3750. DOI: 10.1109/TIP.2014.2332764.
- [20] GRAHAM D B, ALLINSON N M. Characterizing virtual eigensignatures for general-purpose face recognition[M]// WECHSLER H, PHILLIPS P J, BRUCE V, et al. Face Recognition: From Theory to Applications. Berlin: Springer, 1998: 446-456. DOI: 10.1007/978-3-642-72201-1_25.
- [21] WOLD S, ESBENSEN K, GELADI P. Principal component analysis[J]. Chemometrics and Intelligent Laboratory Systems, 1987, 2(1/2/3): 37-52. DOI: 10.1016/0169-7439(87)80084-9.
- [22] NIE Feiping, ZHU Wei, LI Xuelong. Unsupervised feature selection with structured graph optimization[C]// Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. Menlo Park, CA: AAAI Press, 2016: 1302-1308.

(责任编辑 黄 勇)