

基于稀疏学习的鲁棒自表达属性选择算法

何 威¹ 刘星毅^{2*} 程德波¹ 胡荣耀¹

¹(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)

²(广西钦州学院 广西 钦州 535000)

摘 要 受属性选择处理高维数据表现的高效性和低秩自表达方法在子空间聚类上成功运用的启发,提出一种基于稀疏学习的自表达属性选择算法。算法首先将每个属性用其他属性线性表示得到自表达系数矩阵;然后结合稀疏学习的理论(即整合 $L_{2,1}$ -范数为稀疏正则化项惩罚目标函数)实现属性选择。在以分类准确率和方差作为评价指标下,相比其他算法,实验结果表明该算法可更高效地选择出重要属性,且显示出非常好的鲁棒性。

关键词 高维数据 属性选择 属性自表达 稀疏学习

中图分类号 TP181 文献标识码 A DOI: 10.3969/j.issn.1000-386x.2016.11.045

FEATURE SELECTION ALGORITHM WITH ROBUST SELF-REPRESENTATION BASED ON SPARSE LEARNING

He Wei¹ Liu Xingyi^{2*} Cheng Debo¹ Hu Rongyao¹

¹(Guangxi Key Lab of Multi-source Information Mining and Security, Guangxi Normal University, Guilin 541004, Guangxi, China)

²(Qinzhou University, Qinzhou 535000, Guangxi, China)

Abstract Inspired by the high efficiency of feature selection in dealing with high-dimensional data and the success application of low-rank self-representation in subspace clustering, we proposed a sparse learning-based self-represented feature selection algorithm. The algorithm first represents every feature in other feature linearity to obtain the self-representation coefficient matrix; then in combination with sparse learning theory (i. e. to integrate the $L_{2,1}$ -norm as a sparse regularisation punishment object function) it implements feature selection. With the evaluation indexes of classification accuracy and variance, and compared with the algorithms to be contrasted, experimental results indicated that the proposed algorithm could be more efficient in selecting important features and showed excellent robustness as well.

Keywords High-dimensional data Feature selection Feature self-representation Sparse learning

0 引 言

在机器学习和数据挖掘的分类任务中,往往需要处理大量的高维数据,其中无关或冗余的属性会对分类性能产生严重影响。属性选择方法能选择出小部分重要属性作为新的属性集,并能维持甚至提高分类性能^[1],因而在机器学习领域得到了广泛的应用。常见的属性选择方法有:有监督方法、无监督和半监督方法^[13]。有监督的属性选择方法通过已知的类标签,对训练结果测试,从而进行属性的重要性评估,常见的方法有 t 检验法和稀疏正则化线性回归等^[14]。无监督的属性选择,没有指导最小特征子集选择的类标签信息。无监督属性选择方法,主要使用一些评价指标,如方差^[5]、拉普拉斯算子分值^[7]或秩比率^[15]等,来评估每一个单独的特征或特征子集,然后选择最重要的前 k 个特征或最佳特征子集。这些指标可评价聚类性能、冗余、信息损失、样本相似性或流形结构等。但是,在使用这些方法对所有属性进行搜索时,计算复杂度过高,不能适用于高维大数据。半监督方法是有监督与无监督方法的结合,主要利用少量有标签样本和大量的无标签样本进行训练和分类,如:自训练算法、

多视图算法^[13]。本文研究无监督属性选择方法。

在属性选择方法中, $L_{2,1}$ -范数的组稀疏正则化已被成功应用,且在各种实际应用中表现出了良好的性能^[6]。在自然界中,自相似性是普遍存在的,即一个对象的一部分与其本身的其他部分相似。研究发现,拥有自相似性质的对象,每个属性可以近似表示为其自身属性的线性组合,称之为自表达。基于对象的自相似性的性质,自表达性质适用于大多数的高维数据,并已广泛应用于机器学习和计算机视觉领域^[11]。而在无监督的属性选择中,由于没有类标签指导,很难获得最小属性子空间^[4]。本文提出特征的自表达属性方法,即每个属性通过用其他属性来表征,更好地利用了数据的内部特征,从而解决了无监督属性选择方法中,无类标签的特点。基于以上考虑,本文利用特征的

收稿日期:2015-07-27。国家自然科学基金项目(61170131, 61263035, 61363009);国家高技术研究发展计划项目(2012AA011005);国家重点基础研究发展计划项目(2013CB329404);广西自然科学基金项目(2012GXNSFGA060004);广西高校科学技术研究重点项目(2013ZD041);广西研究生教育创新计划项目(YCSZ2015095, YCSZ2015096)。何威,硕士,主研领域:数据挖掘,机器学习。刘星毅,硕士。程德波,硕士。胡荣耀,硕士。

自我表达能力 特征矩阵代表本身寻找有代表性的特征成分,并结合稀疏学习以达到良好的效果。

1 相关理论背景及简介

稀疏学习^[2]最早应用于图形、图像视觉等领域。由于具有强大的内在理论及应用价值,所以稀疏学习得到了迅速的发展,并在模式识别与机器学习等领域得到广泛的应用。

设 $W \in R^{n \times 1}$ 是模型的参数向量 n 是样本数,其中 w_1, w_2, \dots, w_n 表示 n 个样本参数,且 $W = [w_1, w_2, \dots, w_n]$ 。本质上,数据的特点是由少量关键特征决定的。属性选择的目标是找出一个有代表性的特征子集,所有属性都能通过它来重建。鉴于稀疏学习可实现特征的自动选择,为此本文算法采用稀疏学习的方法,来自动选择重要的特征子集。在稀疏学习的基本理论中,首先通过对模型的参数向量 $w \in R^n$ 进行一种稀疏假设,再用训练样本对参数 w 进行拟合,主要实现的目标函数为:

$$\min_w (w) = f(w) + \lambda \phi(w) \quad (1)$$

其中 $f(x)$ 是损失函数, $\phi(w)$ 是正则化项, λ 调节 w 的稀疏性,且 λ 越大, w 越稀疏,反之亦然。

将稀疏学习应用于属性选择算法当中,能够将原始数据之间的系数权重作为重要的鉴别信息引入模型,通过稀疏约束来使得输入数据进行稀疏表示。这样可以去除冗余和不相关属性,同时保证重要属性能够被选择^[9]。鉴于稀疏学习的正则化因子通常选用能够转化为凸优化问题来求解,故更能保证本文提出的模型求得唯一的全局最优解^[3]。在稀疏学习中, L_0 -范数是最有效的稀疏正则因子,但因其求解为 NP 难,故很多文献均采用近似正则项 L_1 -范数来替代 L_0 -范数;而 $L_{2,1}$ -范数能导致行稀疏,已经被证明比 L_1 -范数更适合于属性选择^[18]。 $L_{2,1}$ -范数正则化因子能用于稀疏表示来获得组稀疏,因此本文采用 $L_{2,1}$ -范数作为稀疏正则化因子来对属性自表达进行行稀疏处理,以达到剔除冗余和不相关属性的目的,且能够有效地减少离群点对结果的影响。

2 算法描述和优化方法

基于低秩自表达方法在子空间聚类上的成功运用,本文利用特征的自表达能力,提出了一种简单而且十分有效的无监督属性选择方法。通过属性自表达来构造自表达稀疏矩阵,然后采用稀疏学习的原理将冗余属性和不相关属性进行稀疏化处理,从而本文算法能够寻找出重要的特征成分。

假设给定训练集 $X \in R^{n \times m}$,其中 n 和 m 分别表示样本数和属性数。 x_i 代表第 i 个样本,即 $X = [x_1, x_2, \dots, x_n]$,然后利用 f_j 代表第 j 个属性,即 $X = [f_1, f_2, \dots, f_m]$ 。早期的属性选择是利用一些度量来估计每个属性,比如协方差、拉普拉斯算子^[7]等,然后通过得到的权值对所有属性进行排序,取前面的一部分属性作为新的属性集。在目前的一些方法^[3,4,6]中,通常是先计算样本间的相似性或流行结构,然后建立一个响应矩阵 $Y = [y_1, y_2, \dots, y_m]$,使得属性选择问题转化为一个多输出的回归问题:

$$\min_w (Y - XW) + \lambda R(W) \quad (2)$$

其中, W 是属性权重矩阵 $l(Y - XW)$ 是损失函数 $R(W)$ 是关于

W 的正则化项 λ 是正参数,用来调节对损失项的惩罚。

尽管式(2)是一种广泛应用的属性选择的方法,但在如何选取一个适当的响应矩阵时往往比较困难。因此,本文提出一种 SR_FS(Self-Representation for Feature Selection)的属性选择算法。本文的算法使用 X 作为响应矩阵,其中 $Y = X$,即利用属性自表达的方法,使得每个属性都能被全体属性很好地表现出来。即 X 中的每个属性 f_i 均能通过用全体属性(包括自身)的线性组合来表示:

$$f_i = \sum_{j=1}^m f_j \omega_{ji} + e_i \quad (3)$$

其中 e_i 为误差项, ω_{ji} 为自表达系数。对于每个属性,可以将式(3)重写为:

$$X = XW + E \quad (4)$$

其中 $W = [\omega_{ji}] \in R^{m \times m}$ 是自表达系数矩阵 E 为误差项。式(4)是属性自表达的一种表现模型。显然,矩阵 W 有效地显示了不同属性之间的关系,而且能够使得误差项 E 尽可能小。因 L_F -范数能有效鉴别离群点,故本文用其来估计余量,即: $\min \|X - XW\|_F^2$ 。

经以上属性自表达,可以得到自表达关系矩阵 W 。此时的 W 不能够很好地选择出重要的属性,而且为使得剩余量 E 尽可能接近于 0,故本文算法添加一个有效的稀疏学习正则化来辅助属性选择。因此本文算法的模型可以重写为以下最小化问题:

$$\hat{W} = \arg \min_W \|X - XW\|_F^2 + \lambda R(W) \quad (5)$$

其中 $W = [\omega_1, \dots, \omega_i, \dots, \omega_m]$, ω_i 是 W 的第 i 行,由于 $\|\omega_i\|_2$ 能影响第 i 个属性的表示,故本文将其作为一个属性权重引入模型。通过调节参数 λ 值可以使得 W 行稀疏,例如:如果 $\|\omega_i\|_2 = 0$,就意味着第 i 个属性将不参与其他的属性表示,因此,相应的行稀疏就能够用来正则化系数矩阵 W 。故本文采用可以导致行稀疏的 $L_{2,1}$ -范数作为正则项^[8-10] 惩罚目标函数,即: $R(W) = \|W\|_{2,1} = \sum_{i=1}^m \|\omega_i\|_2$ 。

最终由式(5)就可以得到目标函数为:

$$\min_W \|X - XW\|_F^2 + \lambda \|W\|_{2,1} \quad (6)$$

其中, $X \in R^{n \times d}$ (n 和 d 分别表示样本数和属性变量数) 为训练样本, $W \in R^{d \times d}$ 是关于 X 的系数矩阵 λ 是调和参数。本文算法的伪代码如下:

算法1 SR_FS 算法

输入: 训练样本, 正则化参数 λ 。

输出: 分类准确率。

1. 通过训练样本得出类指示矩阵 $X = [x_{i,k}]$;
2. 依据所选择的模型 $\min_W \|X - XW\|_F^2 + \lambda \|W\|_{2,1}$,调用第 3 节中算法 2 求解全局最优解,并得到自表达系数矩阵 $W \in R^{d \times d}$;
3. 根据得到的 W 对原始属性集 X 进行属性选择,并将得到的属性集作为样本的新属性集;
4. 最后,对新属性集构成的样本采用 SVM 算法进行分类。

本文提出的算法具有如下优点:首先,由于算法采用了无监督属性选择模型的属性自表达方法,所以不需要标签矩阵 Y 来作为响应矩阵;其次,有效地将自表达学习和稀疏学习理论结合,该模型既能消除输入数据中的冗余和不相关的属性,又可实

现特征的自动选择; 最后, 本文的目标函数提出了一种区别于交替方向乘子法的求解方法, 该优化算法能保证目标值在每次迭代过程中减小, 并趋近于全局最优解, 最终求得全局最优解。

3 优化分析求解

由于式 (6) 的前半部分是凸的, 且正则化项是非光滑的, 本文设计了一种高效方法来求解式 (6)。具体为, 首先将式 (6) 对 $w_i (1 \leq i \leq m)$ 求微分, 且令其等于 0, 可以得到如下等式:

$$X^T x^{(i)} - XX^T w_i + \lambda D_i w_i = 0 \quad (7)$$

其中 $D_i (1 \leq i \leq m)$ 是有 k 个对角元素 $\frac{1}{2 \|w_{ki}\|}$ 的对角矩阵, 因此:

$$w_i = (\lambda D_i + XX^T)^{-1} X^T x^{(i)} \quad (8)$$

注意 D_i 是未知的且取决于 W 。由以上分析, 并根据文献 [12] 和文献 [8], 本文提出算法 2, 用一种迭代方法来优化式 (6)。

算法 2 优化求解式 (6) 的伪代码

输入: X ;

输出: $w^{(t)} \in R^{n \times n}$;

1. 初始化 $W \in R^{n \times n} \quad t = 1$;

2. while 未收敛 do

3. 计算对角矩阵 $D_i (1 \leq i \leq m)$, 其中 $D_i^{(i)}$ 的第 k 个对角元素为

$$\frac{1}{2 \|w_{ki}^{(i)}\|};$$

4. 对每一个 $i (1 \leq i \leq m)$, $w_i^{(t+1)} = (XX^T + \lambda D_i^{(i)})^{-1} X^T x^{(i)}$;

5. $t = t + 1$;

6. 结束

证明: 根据算法 2 可得到

$$W^{(t+1)} = \min_W \text{Tr}(X - X^T W)^T (X - X^T W) + \lambda \sum_{i=1}^m w_i^T D_i^{(i)} w_i \quad (9)$$

由此得到:

$$\begin{aligned} & \text{Tr}(X^T W^{(t+1)} - X)^T (X^T W^{(t+1)} - X) + \lambda \sum_{i=1}^m (w_i^{(t+1)})^T D_i^{(i)} w_i^{(t+1)} \\ & \leq \text{Tr}(X^T W^{(t)} - X)^T (X^T W^{(t)} - X) + \lambda \sum_{i=1}^m (w_i^{(t)})^T D_i^{(i)} w_i^{(t)} \\ & \Rightarrow \text{Tr}(X^T W^{(t+1)} - X)^T (X^T W^{(t+1)} - X) + \\ & \lambda \sum_{i=1}^d \sum_{j=1}^m \left(\frac{(w_{ij}^{(t+1)})^2}{2 \|w_{ij}^{(t)}\|} \right) \leq \text{Tr}(X^T W^{(t)} - X)^T (X^T W^{(t)} - X) + \\ & \lambda \sum_{i=1}^d \sum_{j=1}^m \left(\frac{(w_{ij}^{(t)})^2}{2 \|w_{ij}^{(t+1)}\|} \right) \leq \text{Tr}(X^T W^{(t)} - X)^T (X^T W^{(t)} - X) + \\ & \lambda \sum_{i=1}^d \sum_{j=1}^m \|w_{ij}^{(t)}\| \end{aligned}$$

根据文献 [19] 对任意的向量 W 和 w_0 得到 $\|w\| - \frac{\|w\|_2^2}{2 \|w_0\|_2} \leq$

$\|w_0\|_2 - \frac{\|w_0\|_2^2}{2 \|w_0\|_2}$ 。因此算法 2 的每次迭代可减少式 (6) 中的目标值。

$W^{(t)}$ 、 $D_i^{(i)} (1 \leq i \leq m)$ 在收敛时会满足式 (9)。由于式 (6) 是凸函数, 因此满足式 (9) 的矩阵 W 是式 (6) 的全局最优解。因

此, 算法 2 将收敛于目标函数式 (6) 的全局最优解。实验也表明, 优化方法能迅速收敛于最优解。

4 实验结果与分析

4.1 实验数据集和对比算法

本文用六个公开数据集测试提出的属性选择算法性能。为了测试本文算法的有效性和进行属性选择的效果, 使用有标签的数据集来测试算法, 并与其他对比算法进行比较测试。其中数据集 madelon、DBWorld、train 和 musk 来源于 UCI 数据集^[15], TOX-171、warpAP10P 来源于文献 [16], 数据集详情如表 1 所示。

表 1 数据集信息统计

数据集	样本数	属性数	类数
madelon	2000	500	2
DBWorld	64	4 702	2
musk(version2)	6598	168	2
TOX-171	171	5748	4
train	168	147	9
warpAR10P	130	2400	10

本文实验是在 Win7 系统下的 Matlab 2014a 软件上进行编程实验。然后实验选择四种公认的有代表性的算法参与比较, 包括: ① NFS 算法(Non Feature Selection): 对原始数据不做任何处理, 直接使用 LIBSVM 工具箱^[17]进行 SVM 分类, 并作为实验的基准线; ② PCA 主成分分析方法(Principal Component Analysis); ③ LPP 方法(Locality Preserving Projection); ④ RFS 方法^[9]: 通过权值大小来表示属性重要性的强弱, 损失函数和正则化项两者均采用 $L_{2,1}$ 范数来约束。其中, NFS 方法直接对原始数据集进行 SVM 分类^[10], 未对原始数据进行任何处理, 相比 SR_FS 等属性约简算法, 不仅数据处理量大, 而且结果容易受到冗余数据和噪音数据的影响; PCA 考虑数据的主成分把数据从高维空间投影到低维空间, 信息损失较为严重; LPP 只考虑了数据的局部信息, 没有考虑数据内部的整体情况, 投影时信息损失也较为严重; RFS 方法为产生稀疏系数矩阵, 由此可以赋大权重给重要的属性, 赋小权重给不重要的属性。在比较的算法中, PCA 和 LPP 算法均属于子空间学习方法, RFS 算法属于属性选择方法。

分析各算法的时间复杂度和空间复杂度。时间复杂度: 本文 SR_FS 算法是稀疏学习与属性自表达的最小二乘的有效组合, 其时间复杂度与 PCA、LPP 及 RFS 算法一致, 同为 $O(n^3)$ (其中 n 是样本量); 空间复杂度: 本文算法与所有的对比算法均需存储矩阵乘法的中间结果, 空间复杂度为线性。

4.2 实验结果和分析

本文首先用不同的算法对数据进行属性选择处理, 然后用统一的方法对数据进行分类, 并采用分类准确率作为评价指标, 分类准确率越高表示算法效果越好, 以此来检测是否提出的 SR_FS 算法能够具有很好的效果。实验通过 10 折交叉验证方法把原始数据划分为训练集和测试集, 运用训练集通过算法来训练模型, 用测试集测试模型的效果。实验采用 SVM 分类工具箱分类, 得到分类准确率。所有的算法均保证在同一实验环境下进行, 最后提取 10 次运行的实验结果的均值加方差来评估

各算法的分类效果和鲁棒性,具体各数据集实验结果见表 2 所示。为了比较不同数据量纲的各算法的鲁棒性,本文采用变异系数(变异系数 = (标准差/平均值) × 100%)作为评价指标,统计结果见表 3 所示。为更直观地比较各算法的效果,每一数据集各算法每次的实验结果对比如图 1 - 图 6 所示。

首先,分析表 2 数据可以看出,其中,SR_FS 算法在不同数据集上取得的准确率均为最高,且与 NFS 算法比较平均提高了 12.78% 效果最为明显,与 RFS 算法比较平均提高了 3.35%。其中,在 train 数据集上,SR_FS 对比其他算法准确率平均提升 3.94%,为最低;而在 musk 数据集上效果最为显著,SR_FS 与其他对比算法相比平均提高 14.08%,且与 NFS 相比提高了 19.1%。这是因为 SR_FS 算法不仅可以使属性自表达,而且同时利用属性自表达的表示方式,考虑到了全体不同属性之间的相互关系;而 PCA 算法仅做了高维到低维的投影,RFS 算法未考虑数据之间的关系,LPP 算法只考虑了数据局部的关系。而 SR_FS 算法又结合了稀疏学习,弥补了单一方法的不足,因此能有效地选择主要的判别属性和去除噪音属性,显著提高分类性能。根据图 1 - 图 6 可以直观地看出,SR_FS 算法在各数据集上 10 次运行的结果的折线大部分能在其他算法的上方,表明 SR_FS 算法表现出了更好的效果。

分析表 2 的方差统计结果可看出,SR_FS 算法在各个数据集的 10 次运行结果取得的方差与各对比算法相比,均能取得最小的方差。分析表 3 的变异系数统计结果可看出,SR_FS 算法在各数据集上得到的变异系数均为最小。通过各算法 10 次运行结果得到的方差(表 2)和变异系数(表 3)的比较可知,在多个不同数据集上,经本文 SR_FS 算法属性选择后的数据,均能达到更加准确和稳定的分类效果。因此,本文 SR_FS 算法对比算法有更好的鲁棒性。

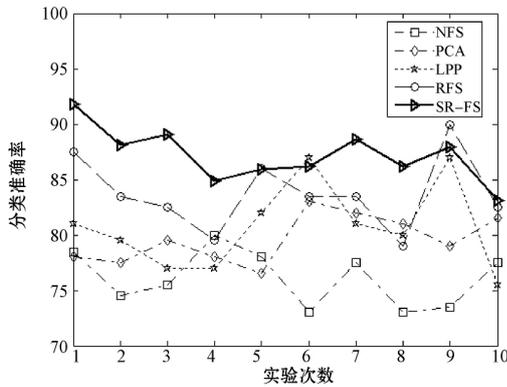


图 1 数据集 madelon 上各算法准确率对比

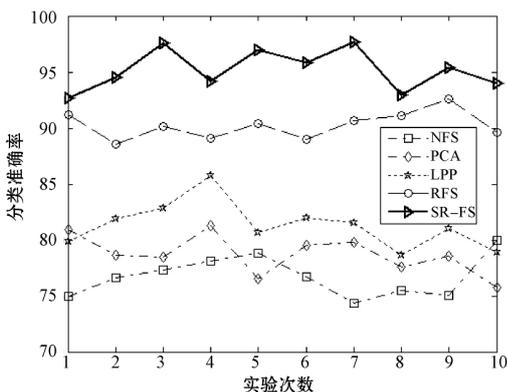


图 2 数据集 mask 上各算法准确率对比

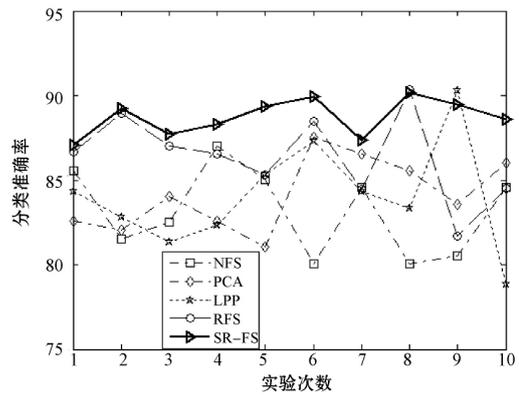


图 3 数据集 train 上各算法准确率对比

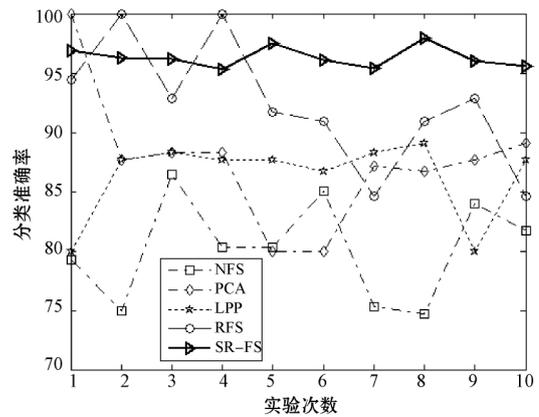


图 4 数据集 warpAR10P 上各算法准确率对比

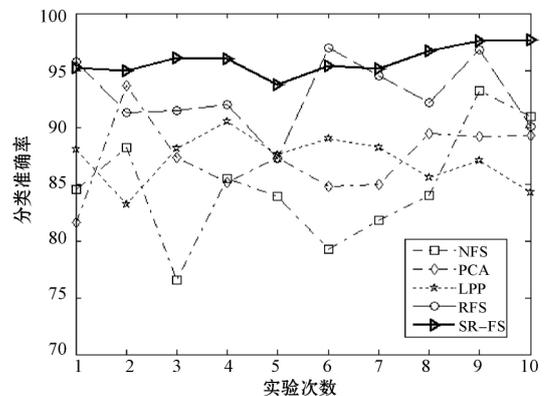


图 5 数据集 TOX-471 上各算法准确率对比

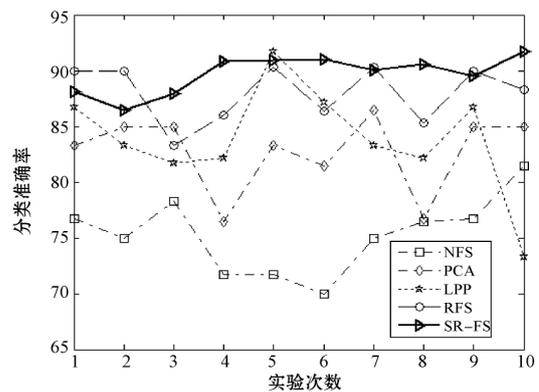


图 6 数据集 DBworld 上各算法准确率对比

(下转第 239 页)

[12] Huang J Z ,Ng M K ,Rong H ,et al. Automated variable weighting in k-means type clustering [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence 2005 27(5) : 657-668.

[13] Jing L ,Ng M K ,Huang J Z. An entropy weighting k-means algorithm for subspace clustering of high-dimensional sparse data [J]. IEEE Transactions on Knowledge and Data Engineering 2007 ,19(8) : 1026-1041.

[14] 李仁侃 ,叶东毅. 属性赋权的 K-Modes 算法优化 [J]. 计算机科学与探索 2012 6(1) : 90-96.

[15] Ji J ,Bai T ,Zhou C ,et al. An improved k-prototypes clustering algorithm for mixed numeric and categorical data [J]. Neurocomputing ,2013 ,120: 590-596.

[16] Ferreira M R ,de Carvalho F d A. Kernel-based hard clustering methods in the feature space with automatic variable weighting [J]. Pattern Recognit 2014 47(9) : 3082-3095.

[17] Zhang L ,Pedrycz W ,Lu W ,et al. An interval weighed fuzzy c-means clustering by genetically guided alternating optimization [J]. Expert Systems with Applications 2014 41(13) : 5960-5971.

(上接第 196 页)

表 2 准确率的均值统计结果(百分比%)

数据集	NFS	PCA	LPP	RFS	SR_FS
madelon	76.1±1.6	79.6±1.5	80.7±2.0	83.8±1.8	88.7±1.4
DBWorld	75.3±3.6	82.8±3.6	83.8±3.8	88.0±2.9	90.3±2.8
musk	76.6±2.1	78.1±3.9	81.7±2.5	89.9±2.3	95.7±2.0
TOX-471	83.0±2.9	85.3±2.8	86.3±2.6	93.8±2.2	95.4±2.0
train	83.1±3.6	84.2±4.4	84.0±3.8	86.4±2.9	88.4±2.8
warpAR10P	80.2±3.0	87.5±2.4	86.3±1.8	92.3±2.3	96.3±1.5
平均值	79.05	82.91	83.80	89.00	92.46

表 3 各数据集十次结果的变异系数统计结果(百分比%)

数据集	NFS	PCA	LPP	RFS	SR_FS
madelon	1.60	1.48	1.69	1.51	1.28
DBWorld	2.40	2.29	2.32	1.91	1.81
musk	1.83	2.38	1.89	1.70	1.53
TOX-471	2.15	2.02	1.92	1.67	1.53
train	2.40	2.53	2.33	1.91	1.81
warpAR10P	2.19	1.87	1.60	1.70	1.32

5 结 语

本文提出一种新的属性选择算法 SR_FS 算法,即通过使用数据本身的性质来构建自表达系数矩阵,并整合 $L_{2,1}$ -范数作为稀疏正则化项,来达到选择重要属性的目的。该算法有效地利用属性的特性来构造自表达系数矩阵并和稀疏属性选择有机融合在一起,达到了进行自动属性选择的目的。其既扩展了属性自表达的理论应用范畴,也弥补了稀疏学习在属性选择处理方面的不足。经实验表明,本文算法能够在分类准确率和稳定性上取得显著提高。在后续的工作中,将尝试在半监督属性选择方面拓展验证本文提出的算法,并尝试使用更先进的技术改进。

参 考 文 献

[1] Zhu X F ,Suk H I ,Shen D G. Matrix-Similarity Based Loss Function and Feature Selection for Alzheimer's Disease Diagnosis[C]//Computer Vision and Pattern Recognition (CVPR) ,2014 IEEE Conference on. IEEE 2014: 3089-3096.

[2] Donoho D L. Compressed sensing [J]. IEEE Transactions on Information Theory 2006 52(4) : 1289-1306.

[3] Zhao Z ,Wang L ,Liu H. Efficient spectral feature selection with minimum redundancy [C]//Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence ,Atlanta 2010: 673-678.

[4] Li Z C ,Yang Y ,Liu J ,et al. Unsupervised feature selection using non-negative spectral analysis [C]//Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence ,Toronto 2012: 1026-1032.

[5] Zhu X F ,Huang Z ,Yang Y ,et al. Self-taught dimensionality reduction on the high-dimensional small-sized data [J]. Pattern Recognition , 2013 46(1) : 215-229.

[6] Yang Y ,Shen H T ,Ma Z G ,et al. $L_{2,1}$ -norm regularized discriminative feature selection for unsupervised learning [C]//Proceedings of the Twenty-Second international joint conference on Artificial Intelligence , 2011 2: 1589-1594.

[7] Wan J W ,Yang M ,Chen Y J. Discriminative cost sensitive Laplacian score for face recognition [J]. Neurocomputing 2015 ,152: 333-344.

[8] Zhu X F ,Huang Z ,Yang Y ,et al. Self-taught dimensionality reduction on the high-dimensional small-sized data [J]. Pattern Recognition , 2013 46(1) : 215-229.

[9] Nie F P ,Huang H ,Cai X ,et al. Efficient and Robust Feature Selection via Joint $L_{2,1}$ -Norms Minimization [C]//Advances in Neural Information Processing Systems 2010: 1813-1821.

[10] Yao L ,Zhang X J ,Li D H ,et al. An interior point method for $L_{1/2}$ -SVM and application to feature selection in classification [J]. Journal of Applied Mathematics 2014(1) : 1-16.

[11] Zhu P F ,Zuo W M ,Zhang L ,et al. Unsupervised feature selection by regularized self-representation [J]. Pattern Recognition 2015 48(2) : 438-446.

[12] Zhu X F ,Huang Z ,Shen H T ,et al. Linear cross-modal hashing for efficient multimedia search [C]//MM 2013-Proceedings of the 2013 ACM International Conference on Multimedia ,2013: 143-152.

[13] Zhao Z ,Wang L ,Liu H ,et al. On Similarity Preserving Feature Selection [J]. IEEE Transactions on Knowledge and Data Engineering , 2013 25(3) : 619-632.

[14] Mitra P ,Murthy C A ,Pal S K. Unsupervised feature selection using feature similarity [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence 2002 24(3) : 301-312.

[15] Nie F P ,Xiang S M ,Jia Y Q ,et al. Trace ratio criterion for feature selection [C]//AAAI08 2008 2: 671-676.

[16] Feature selection datasets [EB/OL]. [2015-04-10]. <http://featureselection.asu.edu/datasets.php>.

[17] LIBSVM—A Library for Support Vector Machines. [EB/OL]. [2015-04-10]. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.

[18] Yang Y ,Shen H T ,Ma Z G ,et al. $L_{2,1}$ -norm regularized discriminative feature selection for unsupervised learning [C]//Proceedings of the Twenty-Second international joint conference on Artificial Intelligence. AAAI Press 2011 2: 1589-1594.

[19] Zhu X F ,Zhang L ,Huang Z. A sparse embedding and least variance encoding approach to hashing [J]. IEEE Transactions on image processing 2014 23(9) : 3737-3750.