

基于图稀疏的自表达属性选择算法

钟智¹, 胡荣耀²⁺, 何威², 罗葵²

(1. 广西师范学院 计算机与信息工程学院, 广西南宁 530023;

2. 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: 为解决高维数据属性维度高, 不易直接应用的问题, 提出通过属性自表达移除不相关和冗余属性的属性选择算法。基于稀疏学习的框架, 通过属性自表达考虑属性间的相关性, 利用子空间学习的局部保留投影 (LPP) 算法, 确保属性选择时数据的局部结构保持不变。实验结果表明, 该算法在 UCI 等数据集上优于 4 种对比算法。

关键词: 属性选择; 属性自表达; 子空间学习; 属性约简; 稀疏学习

中图分类号: TP181 文献标识码: A 文章编号: 1000-7024 (2016) 06-1643-06

doi: 10.16208/j.issn1000-7024.2016.06.041

Graph self-representation for sparse feature selection

ZHONG Zhi¹, HU Rong-yao²⁺, HE Wei², LUO Yan²

(1. College of Computer and Information Engineering, Guangxi Teachers Education University, Nanning 530023, China;

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

Abstract: To solve the issues that high-dimensional data are hardly used in applications, a feature selection method using the self-representation of samples to remove the redundant and irrelevant features was proposed. The self-representation of samples was used to estimate the correlation among features in a sparse feature selection framework, and also locality preserving projection (LPP) was employed to preserve the local structures of samples. The experimental results on real datasets show that the proposed method outperforms state-of-the-art methods.

Key words: feature selection; characteristics of self-representation; subspace learning; dimensionality reduction; sparse learning

0 引言

从原始数据中提取重要的属性, 消除大量冗余的属性, 发现数据内在的关系成为了机器学习和模式识别研究的热点^[1,2]。属性约简是从高维数据中选取一个属性子集, 使得用此子集构建的分类器预测精度提高或并不显著降低。常见的属性约简包含子空间学习和属性选择两种^[3,4]方法, 其中, 子空间学习^[5]是通过投影实现高维属性空间到低维空间的映射, 常见方法如: 局部保留投影 (locality preserving projection, LPP)^[6], 主成分分析 (principal component analysis, PCA)^[7]算法等; 属性选择是从原始数据集中选取一个与原始数据的类分布最接近的子集, 常见方法如:

t-test检验法^[8]和稀疏逻辑回归方法^[9]等。早期的属性选择首先用一些度量方法, 比如: 协方差、拉普拉斯算子^[10]等, 估计每个属性的权值, 然后利用得到的权值来排列所有属性的重要度。近期的一些方法^[11-14]先得到原始数据样本的相似性, 再建立一个响应矩阵, 由此把属性选择的问题转为一个较为简单的多输出的回归问题。

然而, 选取效果好且准确率要高精准的响应矩阵对现有方法而言是一直具有挑战性。为此, 本文考虑到属性选择和子空间学习各自的优点, 提出一种简单且有效的属性选择的算法, 称作基于图稀疏的自表达属性选择算法 (graph self-representation for sparse feature selection, GSR_FS)。本文算法利用原始数据自表达的能力得到一个

收稿日期: 2015-06-22; 修订日期: 2015-08-30

基金项目: 国家自然科学基金项目 (61170131、61263035、61363009); 国家 863 高技术研究发展计划基金项目 (2012AA011005); 国家 973 重点基础研究发展计划基金项目 (2013CB329404); 广西自然科学基金项目 (2012GXNSFGA060004); 广西高校科学技术研究重点基金项目 (2013ZD041); 广西研究生教育创新计划基金项目 (YCSZ2015095、YCSZ2015096)

作者简介: 钟智 (1963-), 男, 广西梧州人, 副教授, 研究方向为机器学习、数据挖掘; 通讯作者: 胡荣耀 (1992-), 男, 江西景德镇人, 硕士研究生, 研究方向为数据挖掘、机器学习; 何威 (1989-), 男, 河南商丘人, 硕士研究生, 研究方向为数据挖掘、机器学习; 罗葵 (1989-), 男, 安徽安庆人, 硕士研究生, 研究方向为数据挖掘、机器学习。E-mail: 610273948@qq.com

自表达属性矩阵，该自表达属性矩阵可以反映原始数据的主要属性（越重要的属性其对应的系数越大）。基于 $l_{2,1}$ 范数能导致行稀疏的事实，采用其进行重要属性的选择，并减少离群样本的干扰。为了使得选出的属性更具有代表性，嵌入子空间学习算法 LPP 算法，以保持数据的局部结构不变。最后经实验验证，该算法能够达到较好的效果。

1 相关理论背景及简介

1.1 稀疏学习简介

稀疏学习 (sparse learning) 最早主要应用于图形、图像视觉等领域，由于具有强大的应用价值及内在理论，所以稀疏学习^[15]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。

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，主要实现的目标为

$$\min_w (w) = f(w) + \lambda \phi(w) \tag{1}$$

式中： $f(x)$ ——损失函数， $\phi(w)$ ——正则化项， λ 调节 w 的稀疏性。即 λ 越大，权值矩阵 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达，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，鉴于稀疏学习的正则化因子通常选用能够凸优化求解的范数，如此，更能保证本文提出的模型求得唯一的全局最优解^[16]。

1.2 LPP 算法简介

经典的流形学习方法可以很好的在高维数据中挖掘出低维的流形结构，如果没有明确的投影矩阵，就很难直接提取新样本的属性，故在模式识别领域运用比较困难。为了解决这个问题，何晓飞等提出了局部保留投影 (locality preserving projection, LPP)^[6]，对新进的样本进行属性提取有很好的效果，其理论是在拉普拉斯属性映射的基础上，使得计算更加简单，数据处理更加快速。

LPP 算法主要作用是使得数据间的近邻关系在投影到子空间后依旧得到保持，同时，以局部合并的方式来重建数据集间的整体内在规律，但是无法较好的体现不同类数据间的分布差异。假设有一个数据集 $X = \{x_1, x_2, \dots, x_n\}$ ，其中 $x_i \in R^m (i = 1, 2, \dots, n)$ ，其中 n 表示样本数， m 表示每个样本的维数，如果能找到一个矩阵 A ，使得 X 中每一个样本数据都投影到子空间 $R^d (d \ll m)$ 中，且有 $Y = \{y_1, y_2, \dots, y_n\}$ ， $y_i \in R^d (i = 1, 2, \dots, n)$ ，即 $y_i = A^T x_i$ ，算法具体步骤如下：

(1) 构建近邻图：把每个 x_i (即 $x_i \in X$) 都当作一个无向图的顶点，再用 k-NN 即 k 最近邻算法 (其中 $k \in N$) 来判断 x_i 与 x_j 是否为近邻，如果为近邻关系，则用一条线

将 x_i 与 x_j 连接起来，否则不连接，如此最终形成一个近邻关系图。

(2) 构造权重矩阵：在构建的近邻图中，没有边连接的点则说明两个样本不相关，具有独立性，而有边连接的点还需要进一步确定两者之间的相关性，即增加该条边的系数权值，可以用一种最简单的方法得到：就是有边相连的权值设置为 1，无边相连的权值设置为 0。因此，数据集 X 可以得到一个权值矩阵 $S_{ij} = e^{-\frac{\|x_i - x_j\|^2}{t}}$ ，其中 t 是一个大于零的常量。

(3) 进行属性映射：算法的最终目的是要找出最优的投影方向，而选择投影方向的准则是使得如下目标函数最小化

$$\sum_{ij} (y_i - y_j)^2 S_{ij} \tag{2}$$

利用第 (2) 步得到的权值矩阵对式 (2) 进行变换，如果存在一个向量 w ，使得 $y_i = w^T x_i$ ，则

$$\frac{1}{2} \sum_{ij} (y_i - y_j)^2 S_{ij} = \frac{1}{2} (\sum_{ij} y_i^2 S_{ij} - 2 \sum_{ij} y_i y_j S_{ij} + \sum_{ij} y_j^2 S_{ij}) = w^T X(D - S)X^T w = w^T X L X^T w \tag{3}$$

其中， $D = (D_{ij})$ 为对角矩阵，且满足 $D_{ij} =$

$$\begin{cases} \sum_{k=1}^m S_{ik}, & i = j \\ 0, & i \neq j \end{cases} . L = D - S, \text{ 叫做拉普拉斯矩阵, 而 } D_{ii} \text{ 表示第}$$

i 个样本, 如果 D_{ii} 越大, 则说明第 i 个样本的重要性越大, 因此投影后对应的 y_i 也就越重要。为了使上述目标函数能够消除在最小化时, 产生的缩放及平移的问题, 可以加上一个约束条件

$$y^T D y = 1 \Rightarrow w^T X D X^T w = 1 \tag{4}$$

从而使得式(2) 最小化问题转化为

$$w^* = \operatorname{argmin} w^T X L X^T w \tag{5}$$

上式可以通过拉格朗日乘子法来求解，最终化简为求 $X L X^T w = \lambda X D X^T w$ 的属性值和属性向量。

2 本文算法描述和优化方法

2.1 算法描述

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，而 f_1, f_2, \dots, f_m 则是对应的特征向量，其中， $f_i \in R^n$ 且 $X = [f_1, f_2, \dots, f_m]$ 。按照目前的方法^[11-14]，首先计算样本的相似性，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

$$\min_w l(Y - XW) + \lambda R(W) \tag{6}$$

式中： W ——属性权值矩阵， $l(Y - XW)$ ——损失函数， $R(W)$ ——关于 W 的正则化项， λ ——正参数用来调节对损失项的惩罚。

尽管式 (6) 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比较困难。因此, 本文提出一种 GSR_FS 属性选择的算法。本文的算法仅仅用 X 作为响应矩阵, 其中 $Y=X$, 即用属性自表达的方法, 使得每个属性都能被所有的属性很好的表现出来。对于 X 中的每个属性 f_i , 都能通过和所有属性 (包括自身) 的线性组合来表示

$$f_i = \sum_{j=1}^m f_j \omega_{ji} + e_i \quad (7)$$

式中: e_i —— 误差项, ω_{ji} —— 系数。然后对于所有的属性, 可以表示为

$$X=XW+E \quad (8)$$

其中, $W=[\omega_{ji}] \in R^{m \times m}$ 是系数表示矩阵, E 为误差项。式 (8) 只是属性自表达的一种表现模型。很明显, 矩阵 W 可以表现出不同属性的关系, 而且要使得剩余的 E 尽可能小。这里, 本文用 l_F 范数来估计余量, 即: $\min_W \|X-XW\|_F^2$ 。

接着, 令 W 作为一个 $m \times m$ 的关系矩阵, 为使得剩余量 $E=0$, 故需要一个正则化项 $R(W)$ 来辅助属性选择, 然后就有接下来的最小化问题

$$\hat{W} = \arg \min_W \|X-XW\|_F^2 + \lambda R(W) \quad (9)$$

其中, $W=[w_1, \dots, w_i, \dots, w_m]$, w_i 是 W 的第 i 行, 由于 $\|w_i\|_2$ 能影响第 i 个属性的表示, 故可以作为一个属性权重。例如: 如果 $\|w_i\|_2=0$, 就意味着第 i 个属性将不参与其它的属性表示, 因此, 相应的行稀疏^[17]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 $l_{2,1}$ 范数作为正则化项惩罚目标项^[14,18], 即: $R(W) = \|W\|_{2,1} = \sum_{i=1}^m \|w_i\|_2$ 。

同时, 本文引入 LPP (局部保留投影) 算法, 通过在提出的模型上嵌入一个正则化项 $R(W) = tr(W^T X L X^T W)$, 用来保证样本数据在空间投影变换后数据的近邻关系依旧保持。最终由式 (9) 就可以得到本文的目标函数为

$$\min_W \frac{1}{2} \|X-XW\|_F^2 + \lambda_1 tr(W^T X L X^T W) + \lambda_2 \|W\|_{2,1} \quad (10)$$

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, L 是如式 (3) 中描述的拉普拉斯矩阵, λ_1 和 λ_2 是两个调整参数。本文算法的伪代码如下:

算法 1: GSR_FS 算法

输入: 训练样本, 正则化参数 λ_1, λ_2 。

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:

$\min_W \frac{1}{2} \|X-XW\|_F^2 + \lambda_1 tr(W^T X L X^T W) + \lambda_2 \|W\|_{2,1}$ 调用

算法 2 求解全局最优解得到自表达系数矩阵 $W \in R^{d \times d}$;

3. 利用 W 对原始属性集 X 进行属性选择后得到的属性集作为样本新的属性集;

4. 对新的属性集构成的样本采用 SVM 分类。

本文提出的算法具有如下优点: 首先, 由于算法为非监督属性选择, 采用了属性自表达的方法, 所以不需要测试矩阵 Y 来作为响应矩阵; 其次, 利用稀疏学习理论来消除输入数据中的冗余和不相关属性; 最后, 由于属性选择具有很好的解释性, 而子空间学习能保证良好的效果, 结合这两种方法的各自优点, 以此保证算法在分类上的鲁棒性。

2.2 优化分析求解

式 (10) 是一个凸但非光滑的函数, 为此, 本文提出一种新的近似加速梯度法来求解目标函数。首先, 通过将式 (10) 按照近似加速梯度法拆分成两个函数

$$f(W) = \frac{1}{2} \|X-XW\|_F^2 + \lambda_1 tr(W^T X L X^T W) \quad (11)$$

$$L(W) = f(W) + \lambda_2 \|W\|_{2,1} \quad (12)$$

这里注意到, $f(W)$ 是凸且可微的, 而 $\lambda_2 \|W\|_{2,1}$ 是凸且非光滑的。为了使用近似加速梯度法来优化 W , 本文用如下的优化准则来迭代更新 W

$$W(t+1) = \underset{W}{\operatorname{argmin}} G_{\eta(t)}(W, W(t)) \quad (13)$$

其中: $G_{\eta(t)}(W, W(t)) = f(W(t)) + \langle \nabla f(W(t)), W-W(t) \rangle + \frac{\eta(t)}{2} \|W-W(t)\|_F^2 + \lambda_2 \|W\|_{2,1}$, $\nabla f(W(t)) = (X X^T + \lambda_1 X L X^T) W(t) - X X^T$, $\eta(t)$ 和 $W(t)$ 分别是调优参数和 t 次迭代所得的 W 值。

通过忽略式 (13) 中的 W 的条件独立性, 我们可以重写式 (13) 为

$$W(t+1) = \pi_{\eta(t)} W((t)) = \underset{W}{\operatorname{argmin}} \frac{1}{2} \|W-U(t)\|_2^2 + \frac{\lambda_2}{\eta(t)} \|W\|_{2,1} \quad (14)$$

其中, $U(t) = W(t) - \frac{1}{\eta(t)}$ 和 $\pi_{\eta(t)}(W(t))$ 是 $W(t)$ 在凸集 $\eta(t)$ 上的欧几里德投影。由于 $W(t+1)$ ^[19] 在每一行的可分性, 令 $w^i(t+1)$ 为 $W(t+1)$ 的第 i 行, 我们可以分别为每一行更新权重

$$w^i(t+1) = \underset{w^i}{\operatorname{argmin}} \frac{1}{2} \|w^i - u^i(t)\|_2^2 + \frac{\lambda_2}{\eta(t)} \|w^i\|_2 \quad (15)$$

其中, $u^i(t) = w^i(t) - \frac{1}{\eta(t)} \nabla f(w^i(t))$, 而 $u^i(t)$ 和 $w^i(t)$ 分别是 $U(t)$ 和 $W(t)$ 的第 i 行, 而 $w^i(t+1)$ 可以更好解决用中如下的方法

$$w^{i*} = \max \{ |w^i| - \lambda_2, 0 \} \cdot \operatorname{sgn}(w^i) \quad (16)$$

同时, 为了加快式 (13) 的近似梯度法, 我们引入一个辅助变量 $V(t+1)$

$$V(t+1) = W(t) + \frac{\alpha(t)-1}{\alpha(t+1)} (W(t+1) - W(t)) \quad (17)$$

其中, $\alpha(t+1)$ 通常取 $\frac{1 + \sqrt{1 + 4\alpha(t)^2}}{2}$ 。

算法 2: 优化求解式 (10) 的伪代码

```

输入:  $\eta(0) = 0.01, \alpha(1) = 1, \gamma = 0.002, \lambda_1, \lambda_2$ ;
输出:  $W$ ;
初始化  $t = 1$ ;
初始化  $W(1)$  作为随机对角矩阵;
重复:
while  $L(W(t)) > G_{\eta(t-1)}(\pi_{\eta(t-1)}(W(t)), W(t))$ 
do 设  $\eta(t) = \eta(t-1)$ ;
end
设  $\eta(t) = \eta(t-1)$ ;
计算  $W(t+1) = \arg \min_w G_{\eta(t)}(W, V(t))$ ;

计算  $\alpha(t+1) = \frac{1 + \sqrt{1 + 4\alpha(t)^2}}{2}$ ;
计算式 (17);
until 式 (10) 收敛;

```

定理 1^[20] 设 $\{W(t)\}$ 是由优化算法所生成的序列, 然后对于 $\forall t \geq 1$, 下式成立

$$\varrho(W(t)) - \varrho(W^*) \leq \frac{2\gamma L \|W(1) - W^*\|_F^2}{(t+1)^2} \quad (18)$$

其中, γ 是一个正的预定义的常数, L 是式 (11) 中 $f(W)$ 的梯度的李普希茨常数 (Lipschitz constant), $W^* = \operatorname{argmin}_W \varrho(W)$. 定理 1 表明算法 1 中的近端加速梯度法 (accelerated proximal gradient method) 的收敛率为 $O(\frac{1}{t^2})$, 其中 t 为迭代次数.

3 实验与结果分析

3.1 实验数据集和对比算法

本文用 4 个数据集测试提出属性选择算法的性能, 其中数据集 *madelon*, *DBWorld*, *train* 来源于 UCI^[21], *warp AP10P* 来源于文献 [22], 数据集详情见表 1.

表 1 数据集信息统计

名称	<i>madelon</i>	<i>DBWorld</i>	<i>train</i>	<i>WarpAR10P</i>
样本个数	2000	64	168	130
属性个数	500	4702	147	2400
类数	2	2	9	10

实验在 win7 系统下运行, 使用 matlab2014a 软件进行实验. 本文实验选择 4 种比较算法来与本文提出的方法进行比较: NFS 方法 (non feature selection); 对原始数据不做任何处理, 直接使用 LIBSVM 工具箱^[23]进行 SVM 分类; PCA (principal component analysis); LPP (locality preserving projection); RFS 方法^[24]; 是一种通过权值大小来保证属性的重要性的强弱, 而且在损失函数和正则化项两者均采用 $l_{2,1}$ 范数来约束, 这样既对离群样本具有鲁棒性, 还能有效计算的属性选择算法.

分析以上算法, NFS 方法直接对原始数据集进行 SVM 分类, 未对原始数据进行任何处理, 相比 GSR_FS 等属性约简算法, 不仅数据处理量大, 而且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; LPP 只考虑了数据的局部信息; RFS 方法产生稀疏系数矩阵, 由此可以赋值大权重给重要的属性, 赋值小权重给不重要的属性. 而且比较的算法中, 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.

3.2 实验结果和分析

本文采用分类准确率作为评价指标, 而且分类准确率越高表示算法效果越好. 实验通过 10-fold cross-valind (即 10 折交叉验证) 方法把原始数据划分为训练集和测试集, 再用 SVM 进行分类并得到分类准确率. 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减均方差来评估各算法的性能, 各算法在 4 个数据集上实验结果对比如图 1~图 4 所示, 具体数据结果见表 2.

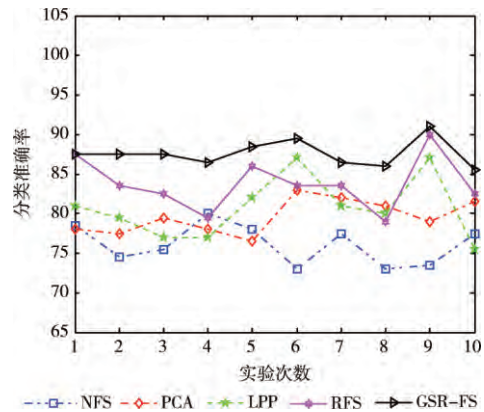


图 1 madelon 数据集

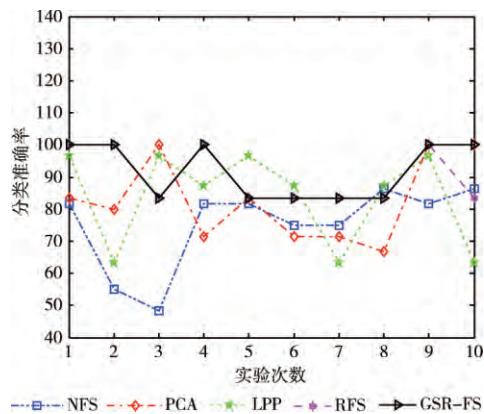


图 2 DBWorld 数据集

通过图 1~图 4 所示, 可以清楚的看到 GSR_FS 算法在 4 个数据集上的分类准确率, 由于 10-折交叉验证的随机性, 并不能保证每次的结果都是最好的, 但是每个数据集上 10 次实验结果对比算法大部分都较高, 最终的平均分类准确率也是最高的.

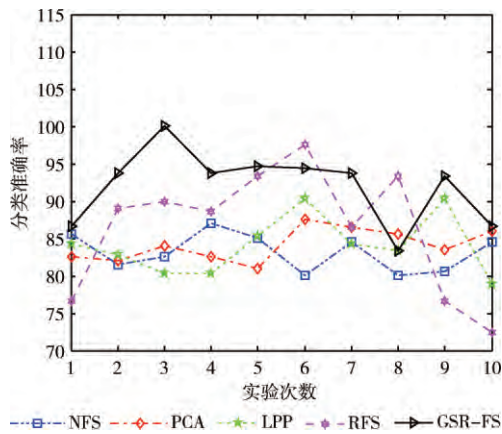


图 3 train 数据集

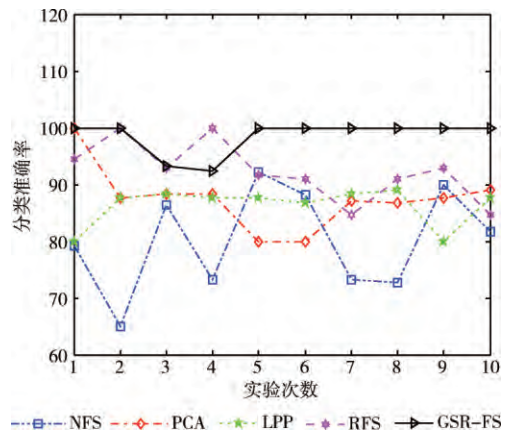


图 4 warpAR10P 数据集

表 2 准确率的均值统计结果

数据集	NFS	PCA	LPP	RFS	GSR_FS
madelon	76.10% ± 1.59	79.60% ± 1.47	80.70% ± 1.98	83.75% ± 1.84	88.70% ± 1.30
DBWorld	75.29% ± 3.62	82.76% ± 3.62	83.81% ± 3.84	90.00% ± 2.93	91.66% ± 2.96
train	83.14% ± 3.59	84.15% ± 4.37	84.02% ± 3.77	86.40% ± 2.89	89.89% ± 2.81
warpAR10P	80.21% ± 3.00	87.50% ± 2.35	86.32% ± 1.84	92.28% ± 2.29	98.56% ± 1.44
平均值	78.68	83.50	83.71	88.11	92.20

通过分析表 2 分析得出, GSR_FS 与其它 4 种对比算法比较均取得最高的分类准确率, 具体地, 与 NFS 算法比较平均提高了 13.52%, 与 RFS 算法对比平均提高了 4.09%。由于 train 数据集维度较小, 5 种算法所得准确率比较接近, GSR_FS 算法最好提高了 3.49%。在 warpAR10P 数据集上, GSR_FS 算法效果最佳, 不但与 NFS 相比提高了 18.35%, 而且比 RFS 算法提高了 6.28%。这是由于 GSR_FS 算法不仅仅利用 $l_{2,1}$ 范数来去除噪音和冗余属性, 而且结合了数据局部信息结构, 故能显著提高模型的性能。

由于不同数据集有不同的数据分布, 且含有不同的干扰因素。实验结果显示, 本文提出的 GSR_FS 算法在每个数据集上均对比算法好, 相对所有对比算法, 本文 GSR_FS 算法具有最强的鲁棒性。同时, 跟子空间学习方法比较, 本文算法不仅保持了原始属性, 还利用子空间学习把原始属性转化为新的属性, 且比单一的属性选择算法保证了数据自身的局部结构, 因此使得本文算法具有更好的解释性。

4 结束语

本文提出一种属性选择算法——GSR_FS 算法, 通过使用原始数据本身来构建自表达系数矩阵, 并整合 $l_{2,1}$ 范数正则项和局部保留投影 (LPP)。该算法有效的将子空间学习和稀疏属性选择融合在一起, 不仅能够保存数据的局部信息而且能消除冗余和不相关的属性。既扩展了子空间

学习的理论应用范畴, 也弥补了稀疏学习在局部信息处理方面的不足。实验结果表明, 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算法。

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.
- [2] Zhu Xiaofeng, Zhang Shichao, Jin Zhi, et al. Missing value estimation for mixed-attribute data sets [J]. IEEE Transactions on Knowledge & Data Engineering, 2010, 23 (1): 110-121.
- [3] Zhu Xiaofeng, Suk Heung-Il, Shen Dinggang. Matrix-similarity based loss function and feature selection for Alzheimer's disease diagnosis [C] //IEEE Conference on Computer Vision and Pattern Recognition. Columbus; IEEE, 2014: 3089-3096.
- [4] Gu Quanquan, Li Zhenhui, Han Jiawei. Joint feature selection and subspace learning [C] //Twenty-Second International Joint Conference on Artificial Intelligence, 2011: 1294-1299.
- [5] Zhu Xiaofeng, Huang Zi, Cheng Hong, et al. Sparse hashing for fast multimedia search [J]. Acm Transactions on Information Systems, 2013, 31 (2): 595-605.
- [6] Zhu Xiaofeng, Huang Zi, Yang Yang, et al. Self-taught di-

- dimensionality reduction on the high-dimensional small-sized data [J]. *Pattern Recognition*, 2013, 46 (1): 215-229.
- [7] 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.
- [8] Konietschke F, Pauly M. Bootstrapping and permuting paired t -test type statistics [J]. *Statistics & Computing*, 2013, 24 (3): 283-296.
- [9] Liimatainen K, Heikkilä R, Yli-Harja O. Sparse logistic regression and polynomial modelling for detection of artificial drainage networks [J]. *Remote Sensing Letters*, 2015, 6 (4): 311-320.
- [10] Benabdeslem K, Hindawi M. Constrained Laplacian score for semi-supervised feature selection [G]. LNCS 6911: *Machine Learning and Knowledge Discovery in Databases*, 2011: 204-218.
- [11] Zhu Xiaofeng, Li Xuelong, Zhang Shichao. Block-row sparse multiview multilabel learning for image classification [J]. *IEEE Trans Cybern*, 2015.
- [12] Zhu Xiaofeng, Huang Zi, Shen Hengtao, et al. Linear cross-modal hashing for efficient multimedia search [C] // *Proceedings of the 21st ACM International Conference on Multimedia*. New York: ACM, 2013: 143-152.
- [13] Li Zechao, Yang Yi, Liu Jing, et al. Unsupervised feature selection using nonnegative spectral analysis [C] // *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*. Toronto, 2012: 1026-1032.
- [14] Yang Yi, Shen Hengtao, Ma Zhigang, 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: 1589-1594.
- [15] Lai Hanjiang, Pan Yan, Liu Cong, et al. Sparse learning-to-rank via an efficient primal-dual algorithm [J]. *IEEE Transactions on Computers*, 2013, 62 (6): 1221-1233.
- [16] Adel A Alamri. Theory and methodology on the global optimal solution to a general reverse logistics inventory model for deteriorating items and time-varying rates [J]. *Computers & Industrial Engineering*, 2011, 60 (2): 236-247.
- [17] Zhu Xiao, Huang Zi, Cui Jiangtao, et al. Video-to-shot tag propagation by graph sparse group lasso [J]. *IEEE Transactions on Multimedia*, 2013, 15 (3): 633-646.
- [18] He Ran, Tan Tieniu, Wang Liang, et al. $L_{2,1}$ Regularized coreentropy for robust feature selection [C] // *Twenty-Fifth IEEE Conference on Computer Vision and Pattern Recognition*. Rhode Island: IEEE, 2012: 2504-2511.
- [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.
- [20] Zhu Xiaofeng, Suk Heung-I, Shen Dinggang. A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis [J]. *NeuroImage*, 2014, 100: 91-105.
- [21] UCI repository of machine learning datasets [EB/OL]. [2015-04-10]. <http://archive.ics.uci.edu/ml/>.
- [22] Feature selection datasets [EB/OL]. [2015-04-10]. <http://featureselection.asu.edu/datasets.php>.
- [23] LIBSVM—A Library for Support Vector Machines [EB/OL]. [2015-04-10]. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [24] Nie Feiping, Huang Heng, Cai Xiao, et al. Efficient and robust feature selection via joint $L_{2,1}$ -norms minimization [J]. *Advances in Neural Information Processing Systems*, 2010, 23: 1813-1821.