

## 新颖的判别性特征选择方法

吴锦华<sup>1,2\*</sup>, 左开中<sup>1,2</sup>, 接标<sup>1,2,3</sup>, 丁新涛<sup>1,2</sup>

(1. 安徽师范大学 数学计算机科学学院, 安徽 芜湖 241003;

2. 安徽师范大学 网络与信息安全工程技术研究中心, 安徽 芜湖 241003;

3. 南京航空航天大学 计算机科学与技术学院, 南京 210016)

(\* 通信作者电子邮箱 ahnu\_wjh@139.com)

**摘要:**作为数据预处理的一种常用的手段,特征选择不仅能够提高分类器的分类性能,而且能增加对分类结果的解释性。针对基于稀疏学习的特征选择方法有时会忽略一些有用的判别信息而影响分类性能的问题,提出了一种新的判别性特征选择方法——D-LASSO,用于选择出更具有判别力的特征。首先 D-LASSO 模型包含一个  $L_1$ -范式正则化项,用于产生一个稀疏解;其次,为了诱导出更具有判别力的特征,模型中增加了一个新的判别性正则化项,用于保留同类样本以及不同类样本之间几何分布信息,用于诱导出更具有判别力的特征。在一系列 Benchmark 数据集上的实验结果表明,与已有方法相比较,D-LASSO 不仅能进一步提高分类器的分类精度,而且对参数也较为鲁棒。

**关键词:**特征选择;稀疏解;  $L_1$ -范式;判别正则化项;分类

**中图分类号:** TP181 **文献标志码:** A

### New discriminative feature selection method

WU Jinhua<sup>1,2\*</sup>, ZUO Kaizhong<sup>1,2</sup>, JIE Biao<sup>1,2,3</sup>, DING Xintao<sup>1,2</sup>

(1. School of Mathematics and Computer Science, Anhui Normal University, Wuhu Anhui 241003, China;

2. Network and Information Security Engineering Research Center, Anhui Normal University, Wuhu Anhui 241003, China;

3. College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing Jiangsu 210016, China)

**Abstract:** As a kind of common method for data preprocessing, feature selection can not only improve the classification performance, but also increase the interpretability of the classification results. In sparse-learning-based feature selection methods, some useful discriminative information is ignored, and it may affect the final classification performance. To address this problem, a new discriminative feature selection method called Discriminative Least Absolute Shrinkage and Selection Operator (D-LASSO) was proposed to choose the most discriminative features. In detail, firstly, the proposed D-LASSO method contained a  $L_1$ -norm regularization item, which was used to produce sparse solution. Secondly, in order to induce the most discriminative features, a new discriminative regularization term was introduced to embed the geometric distribution information of samples with the same class label and samples with different class labels. Finally, the comparison experimental results obtained from a series of Benchmark datasets show that, the proposed D-LASSO method can not only improve the classification accuracy, but also be robust against parameters.

**Key words:** feature selection; sparse solution;  $L_1$ -norm; discriminative regularization item; classification

## 0 引言

在机器学习和模式识别领域,传统学习算法经常遇到“维数灾难”问题<sup>[1]</sup>。在此情形下,降低数据维度的方法不仅能够提高计算效率和改善分类的性能,而且能够增加对分类结果的解释性。作为一种重要的降维手段,特征选择的目的是选择最能表示数据内在特性的一组最优特征子集,从而最终达到优化学习模型的目标。

目前,已有的特征选择方法被大致分为两种:特征排序方法和特征子集搜索方法。特征排序方法一般根据某一准则函数单独分析每一特征的重要性并排序,从而选择一组最重要

的特征,其代表性的方法有 LS (Laplacian Score)<sup>[2]</sup> 和 FS (Fisher Score)<sup>[3]</sup>;而特征子集搜索方法是从给定的特征空间选取与给定模型相符的最优特征子集,代表方法有:递归特征消除 (Recursive Feature Elimination, RFE) 算法<sup>[4]</sup>、LASSO (Least Absolute Shrinkage and Selection Operator)<sup>[5]</sup> 方法、顺序向前移动选择 (Sequential Floating Forward Selection, SFFS)<sup>[6]</sup> 等。相比较特征排序方法,特征子集搜索方法通常能得到更好的分类性能。因此,本文主要针对特征子集搜索方法开展研究。

在各种特征子集搜索方法中,稀疏化的方法得到了广泛研究和应用。例如:LASSO 方法通过最小化一个目标函数能

**收稿日期:** 2015-06-16; **修回日期:** 2015-06-27。 **基金项目:** 国家自然科学基金资助项目 (61472005); 安徽省自然科学基金资助项目 (1508085MF125); 模式识别国家重点实验室开放课题资助项目 (201407361)。

**作者简介:** 吴锦华 (1991-), 男,安徽安庆人,硕士研究生,主要研究方向:机器学习、信息安全; 左开中 (1974-), 男,安徽宿州人,教授,博士,CCF 会员,主要研究方向:信息安全、机器学习; 接标 (1977-), 男,安徽宿州人,副教授,博士,主要研究方向:机器学习、医学图像处理; 丁新涛 (1979-), 男,安徽芜湖人,讲师,博士,CCF 会员,主要研究方向:模式识别、图像处理。

够同时进行特征选择和回归,经验数据已经表明当有大量的不相关特征而只有少量的样本时,LASSO方法通常非常有效<sup>[7]</sup>。因此,LASSO方法已经被应用到很多问题当中,如:寻找疾病的生物标识<sup>[8]</sup>和疾病的分类<sup>[9]</sup>。然而,LASSO方法中存在的一个限制是:当特征维数 $d$ 大于样本个数 $N$ 时,LASSO最多只能选择 $N$ 个特征。针对这一不足,Zou等<sup>[10]</sup>通过增加一个 $L_2$ -范式正则化项,提出一种Elastic net特征选择方法。另外,针对一些数据特征具有平滑性的特点(即相邻的特征应该具有相同或相近的权重系数),研究人员通过在LASSO基础上增加了一个新的 $L_1$ -范式正则化项,提出了一种Fused LASSO方法<sup>[11]</sup>。Ye等<sup>[12]</sup>针对大规模数据情况,给出了对Fused LASSO进行快速优化的方法。最近,研究人员基于核的方法将LASSO从线性情况推广到非线性情况,取得了很好的效果<sup>[13]</sup>。然而,这些基于稀疏化的特征选择方法的一个主要缺点是在线性映射过程中忽略了一些有用的判别性信息,如类内样本的紧致性(即同类样本映射后应尽可能靠近)和类间样本的分异性(即不同类样本映射后应尽可能远离),这些信息的丢失可能会影响到选择特征的判别性,从而影响到分类器最终的性能。

为解决这一问题,本文提出了一种新的判别性特征选择方法——D-LASSO(Discriminative Least Absolute Shrinkage and Selection Operator)。D-LASSO方法充分保留了同类样本之间和不同类样本之间的几何信息,从而帮助诱导出更具判别能力的特征。具体而言:D-LASSO方法模型中首先包含了一个 $L_1$ -范式正则化项,用于产生一个稀疏解;其次,本文增加了一个新的正则化项,用于保留类内样本以及类间样本几何分布信息,从而帮助诱导出更具有判别力的特征;进一步,使用加速近似梯度(Accelerated Proximal Gradient, APG)<sup>[14-15]</sup>算法优化提出的模型;最后,在8个Benchmark数据集上(即6个标准UCI数据集和2个基因表达式数据集)上验证了该模型的有效性。

## 1 LASSO方法

给定训练样本集 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbf{R}^{d \times N}$ ,其中 $\mathbf{x}_i$ 表示第 $i$ 个样本的特征向量, $N$ 表示训练样本个数, $d$ 表示特征维数。 $\mathbf{Y} = [y_1, y_2, \dots, y_N] \in \mathbf{R}^N$ 表示这些样本所对应的响应向量。对于全监督分类问题, $y_i$ 表示样本的类标签,可以是离散的类标签,也可以是具体的数值。不失一般性,本文只考虑两类分类问题,即 $y_i \in \{-1, +1\}$ 。LASSO方法优化的目标函数是:

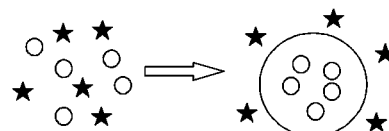
$$\min_{\mathbf{w}} (\|\mathbf{Y} - \mathbf{X}^T \mathbf{w}\|_2^2 + \lambda \|\mathbf{w}\|_1) \quad (1)$$

其中: $\mathbf{w}$ 表示特征向量的回归系数,维数为 $d$ ;正则化项 $\|\mathbf{w}\|_1$ 采用 $L_1$ -范式将在特征空间中产生一个稀疏解,即在 $\mathbf{w}$ 中不相关的和多余的特征所对应的系数将被置为零,而非零系数所对应的特征将被保留下来用于随后的分类; $\lambda$ 是正则化参数,用于平衡模型复杂度和数据拟合程度之间的相对贡献。

## 2 D-LASSO方法

在LASSO特征选择模型中,采用了线性映射模型将数据

从高维特征空间投影到一维空间,然而该模型仅考虑样本数据与类标签之间的相互关系,D-LASSO方法通过加入新的正则化项,使得同类样本投影以后尽可能靠近(即类内样本的紧致性),而不同类样本尽可能远离(即类间样本的分异性),然而在LASSO方法中忽略了判别性信息。图1为给出的图例,在图例中,“圈”和“星”分别表示两类样本,其投影后,表示“圈”和“星”的同类样本之间应该更加靠近,而表示“圈”和“星”的两类样本之间应该尽可能分离。为此,本文所提方法具体描述如下:



注:样本用不同形状区分。

图1 判别性特征选择模型图例

给定训练样本,为探索数据的局部几何结构信息,首先构造最近邻图 $G$ ,即:对于每个样本 $\mathbf{x}_i$ ,基于它的 $k$ 个最近邻 $knn(i) = \{\mathbf{x}_1^i, \mathbf{x}_2^i, \dots, \mathbf{x}_k^i\}$ 定义图 $G$ 的相应的权重矩阵 $S$ :

$$S(i, j) = \begin{cases} 1, & \mathbf{x}_i \in knn(j) \text{ 或 } \mathbf{x}_j \in knn(i) \\ 0, & \text{其他} \end{cases} \quad (2)$$

进一步,为了嵌入样本的判别信息,并保留类内样本的紧致性和类间样本的分异性,对每个样本 $\mathbf{x}_i$ ,以它为节点构建两类图,即:类内图 $G_w$ 和类间图 $G_b$ 。为此,将每个样本 $\mathbf{x}_i$ 的 $knn(i)$ 分为两个互不相交的子集:

$$knn(i)_w = \{\mathbf{x}_j^i \mid \mathbf{x}_j^i \text{ 和 } \mathbf{x}_i \text{ 属于同一类}, 1 \leq j \leq k\} \quad (3)$$

$$knn(i)_b = \{\mathbf{x}_j^i \mid \mathbf{x}_j^i \text{ 和 } \mathbf{x}_i \text{ 不属于同一类}, 1 \leq j \leq k\} \quad (4)$$

根据定义可以看出:最近相邻图 $G$ 由类内图 $G_w$ 和类间图 $G_b$ 构成。上述的 $knn_w$ 保留了在 $k$ 近邻样本点中与 $\mathbf{x}_i$ 为相同类的样本点, $knn_b$ 保留在 $k$ 近邻样本点中与 $\mathbf{x}_i$ 为不同类的样本点。根据式(3)~(4),类内图 $G_w$ 和类间图 $G_b$ 上的权重矩阵 $S_w$ 和 $S_b$ ,能被定义如下:

$$S_w(i, j) = \begin{cases} 1, & \mathbf{x}_i \in knn_w(j) \text{ 或 } \mathbf{x}_j \in knn_w(i) \\ 0, & \text{其他} \end{cases} \quad (5)$$

$$S_b(i, j) = \begin{cases} 1, & \mathbf{x}_i \in knn_b(j) \text{ 或 } \mathbf{x}_j \in knn_b(i) \\ 0, & \text{其他} \end{cases} \quad (6)$$

受最近工作<sup>[16-17]</sup>的启发,可以将两种图映射到线性空间,同时使类内图 $G_w$ 上相连接的样本点尽可能地靠近,而类间图 $G_b$ 的相连接的样本点尽可能远离,为此,引入如下正则化项:

$$Re(\mathbf{w}) = \eta P_w(\mathbf{w}) - (1 - \eta) P_b(\mathbf{w}); 0 \leq \eta \leq 1 \quad (7)$$

其中:

$$P_w(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^k \|f(\mathbf{x}_i) - f(\mathbf{x}_j)\|_2^2 S_w(i, j) \quad (8)$$

$$P_b(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^k \|f(\mathbf{x}_i) - f(\mathbf{x}_j)\|_2^2 S_b(i, j) \quad (9)$$

其中: $P_w$ 和 $P_b$ 分别测量类内样本的紧致性和类间样本的分异性; $\eta$ 是调谐参数,用于平衡两项之间相对贡献。对式(7)进行简单的代数运算,可得:

$$Re(\mathbf{w}) = \eta \mathbf{w}^T \mathbf{X} \mathbf{L}_w \mathbf{X}^T \mathbf{w} - (1 - \eta) \mathbf{w}^T \mathbf{X} \mathbf{L}_b \mathbf{X}^T \mathbf{w}; 0 \leq \eta \leq 1 \quad (10)$$

其中:  $L_w = D_w - S_w$  和  $L_b = D_b - S_b$  分别表示类内图  $G_w$  和类间图  $G_b$  的拉普拉斯矩阵,  $D_w$  和  $D_b$  分别表示相应的对角矩阵, 且  $D_w(i, j) = \sum_j S_w(i, j)$ ,  $D_b(i, j) = \sum_j S_b(i, j)$ 。

最终, 本文提出 D-LASSO 的目标函数如下:

$$\min_w \|Y - X^T w\|_2^2 + \lambda \|w\|_1 + \beta[\eta w^T X L_w X^T w - (1 - \eta) w^T X L_b X^T w] \quad (11)$$

其中: 参数  $\beta$  表示的是正则化项的参数, 用于平衡稀疏项和判别正则化项的相对贡献。

### 3 优化算法

本文采用加速近似梯度 (Accelerated Proximal Gradient, APG) 来优化所提出的 D-LASSO 模型。具体而言, 首先目标函数(11)被划分为两部分。

1) 平滑部分:

$$\min_w \|Y - X^T w\|_2^2 + \beta[\eta w^T X L_w X^T w - (1 - \eta) w^T X L_b X^T w] \quad (12)$$

2) 非平滑部分:

$$g(w) = \lambda \|w\|_1 \quad (13)$$

其次, 构造函数  $\Omega_l$  来近似  $f(w) + g(w)$ :

$$\Omega_l(w, w_k) = f(w_k) + \langle w - w_k, \nabla f(w_k) \rangle + (L/2) \|w - w_k\|^2 + g(w) \quad (14)$$

其中:  $\nabla f(w_k)$  是第  $k$  次迭代的  $w_k$  点梯度;  $L$  是步长, 其值可以通过线性搜索来确定。APG 中  $w$  的更新步骤定义如下:

$$w_{k+1} = \arg \min \left( \frac{1}{2} \|w - v_k\|_2^2 + \frac{1}{L} g(w) \right) = \arg \min \left[ \frac{1}{2} \sum_{i=1}^d \left( \|w(i) - v_k(i)\|_2^2 + \frac{\lambda}{L} |w(i)| \right) \right] \quad (15)$$

其中:  $v_k = w_k - 1/L \nabla f(w_k)$ 。

因此, 根据式(14), 优化的问题被分解成  $d$  个独立的子问题。APG 算法的关键是如何有效求解这些独立的子问题。根据文献[14-15], 这些子问题的解析解非常容易获得, 即

$$w^*(i) = \begin{cases} \left(1 - \frac{\beta}{L \|v_k(i)\|_2}\right) v_k(i), & \|v_k(i)\|_2 > \beta/L \\ 0, & \text{其他} \end{cases} \quad (16)$$

另外, 根据文献[15]使用的技巧, 本文通过计算如下搜索点来代替在  $w(i)$  上的梯度下降。

$$Q_i = w_k - \alpha_i (w(i) - w(i-1)) \quad (17)$$

其中:  $\alpha_i = \frac{(1 - \rho_{i-1}) \rho_i}{\rho_{i-1}}$ ,  $\rho_i = \frac{2}{i+3}$ 。

算法 APG 算法优化 D-LASSO 方法。

输入 令  $X = [x_1, x_2, \dots, x_N] \in \mathbb{R}^{d \times N}$  表示包含有  $N$  个训练样本的数据集, 其中  $d$  是样本的维数; 令  $Y = [y_1, y_2, \dots, y_N] \in \mathbb{R}^N$  表示每个样本所对应的标签, 其中每个样本所对应的标签  $y_i \in \{-1, +1\}$ 。

输出  $J^*$ 。

初始化  $\beta \geq 0, \lambda \geq 0, 0 \leq \eta \leq 1, l_0 > 0, w_0 = w_1 = 0, \rho_0 = 1$   
For  $i = 1$  to 最大迭代次数  $Iterator\_num$

1) 根据式(17) 计算搜索点  $Q_i$

2)  $l = l_{i-1}$ ;

3) while  $f(w_{i+1}) + g(w_{i+1}) > \Omega_l(w_{i+1}, Q_i)$ ,  $l = \sigma * l$ ,  
通过式(15) 计算  $w_{i+1}$

4)  $l_i = l$

End

计算  $J^* = \{j | w_j \neq 0\}$

## 4 实验环境

### 4.1 实验数据集

实验在 6 个标准 UCI 数据集和 2 个基因表达式数据集上评价本文方法的有效性。表 1 给出数据集的相关特性, 其中包括样本维数和样本个数。

表 1 实验中使用的数据集

数据集	维数	样本个数
heart_stalog	13	270
hepatitis	19	155
labor	26	57
ionosphere	34	351
credit	43	690
colic	60	368
colon_cancer	2000	62
embryon_tumor	7129	60

### 4.2 实验设置

在实验中, 采用 10 折交叉验证来评价分类的性能, 即所有的样本被划分为 10 等份, 对每次交叉验证, 取 1 份用于测试, 其余 9 份用于训练, 总共重复 10 次。同时为了减少在交叉验证过程中因随机划分样本而造成对分类结果的影响, 这个过程被独立地重复 10 次。最后, 平均分类精度作为最终的分类结果。在分类步骤中, 基于 RBF 核的支持向量机 (Support Vector Machine, SVM) 被用于分类, 其中参数值被设为默认值 (即  $C=1$ )。

### 4.3 实验结果

#### 4.3.1 分类精度比较

为了评价提出的方法, 首先比较 LASSO 的特征选择方法, 并比较了经典的基于特征排名特征选择方法, 包括 Laplacian Score (LS)<sup>[2]</sup> 和 Fisher Score (FS)<sup>[3]</sup> 的方法。表 2 中总结了所有方法在 8 个数据集上的分类结果。注意, 表 2 中 Baseline 表示没有进行特征选择的方法。从表 2 可看出, 本文的特征选择方法 D-LASSO 方法在所有数据集上的分类效果都要优于 LASSO 方法, 表明它诱导出了更具有判别力的特征, 以致具有更好的分类性能, 间接验证了样本数据中判别信息在分类过程中的重要性。另外, 在大部分数据上本文方法优于比较方法, 这进一步表明了本文方法的有效性。

#### 4.3.2 参数 $\lambda$ 和 $\beta$ 的影响

在提出的 D-LASSO 方法的模型中, 包含了两个正则化项参数, 即  $\lambda$  和  $\beta$ , 用于平衡模型中三个项之间的相对贡献。为了评价它们对分类结果的影响, 采用固定其中一个参数而变化另外一个参数的方式来评价 D-LASSO 方法的分类性能。

图 2 给出了固定  $\beta$  变化  $\lambda$  时本文方法在 8 个数据集上的分类精度,而且为了比较,图 2 中也给出了 LASSO 方法(即参数  $\beta = 0$ );图 3 给出了固定  $\lambda$  变化  $\beta$  时本文方法在 8 个数据集上的分类性能。

从图 2 可看出:在所有不同  $\lambda$  取值情况下,D-LASSO 方法的分类性能在 8 个数据集上都要明显好于 LASSO 方法,这再

一次表明本文方法的效能;另外,D-LASSO 在一些数据集上,特别是两个高维数据集上分类性能受  $\lambda$  的影响比较大,这是因为正则化参数  $\lambda$  控制着模型的稀疏性,从而决定选择特征的个数。从图 3 可看出,分类性能随  $\beta$  变化的曲线比较平滑,这表明 D-LASSO 方法受  $\beta$  的影响比较小,说明 D-LASSO 方法对参数  $\beta$  比较鲁棒。

表 2 不同特征选择方法的平均分类精度(±标准差)

数据集	D-LASSO	LASSO	FS	LS	Baseline
heart_statlog	84.6(±0.16)	84.5(±0.52)	83.2(±1.13)	84.2(±0.63)	<b>85.0(±0.31)</b>
hepatitis	83.9(±0.59)	81.8(±1.03)	<b>84.5(±0.25)</b>	83.4(±0.90)	83.1(±0.80)
labor	<b>94.3(±1.48)</b>	91.4(±1.70)	92.0(±2.89)	90.0(±2.82)	89.2(±1.50)
ionosphere	<b>91.2(±0.37)</b>	89.1(±0.76)	88.9(±0.41)	88.1(±0.78)	89.4(±0.51)
credit	<b>85.8(±0.07)</b>	85.2(±0.34)	85.7(±0.32)	65.8(±0.93)	85.1(±0.28)
colic	<b>84.4(±0.19)</b>	83.0(±0.71)	83.8(±0.76)	75.4(±0.57)	83.7(±0.68)
colon_cancer	<b>80.4(±2.64)</b>	79.7(±1.49)	75.0(±2.99)	69.2(±1.32)	65.0(±0.00)
embryon_tumor	<b>66.3(±1.14)</b>	65.2(±0.32)	63.8(±2.70)	62.1(±0.21)	62.0(±0.00)

注:粗体表示分类精度在当前数据集中处于最佳。

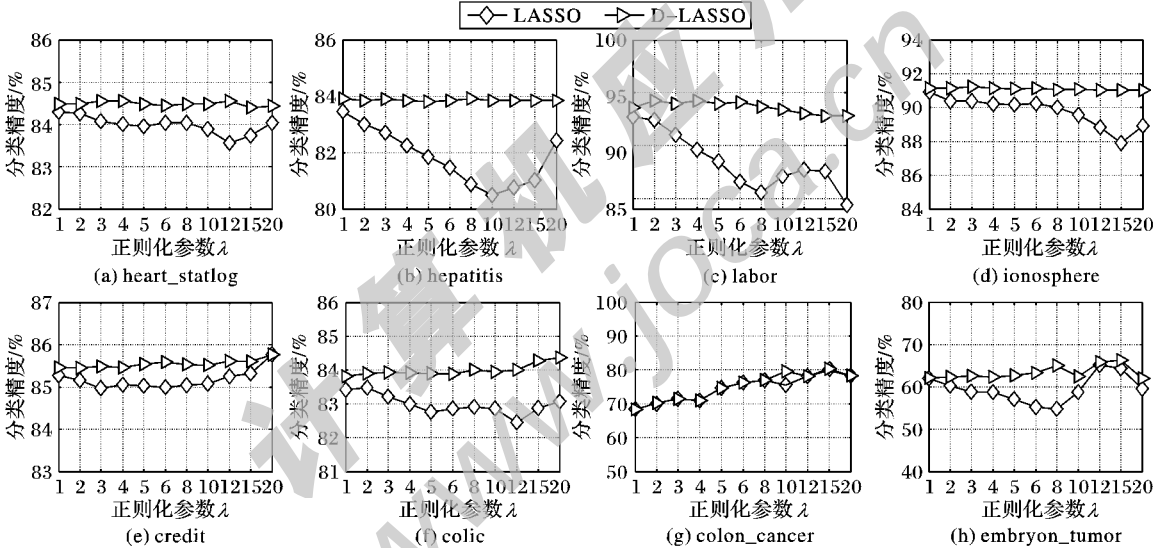


图 2 固定  $\beta$  变化  $\lambda$  时提出 D-LASSO 方法在 8 个数据集上的分类精度

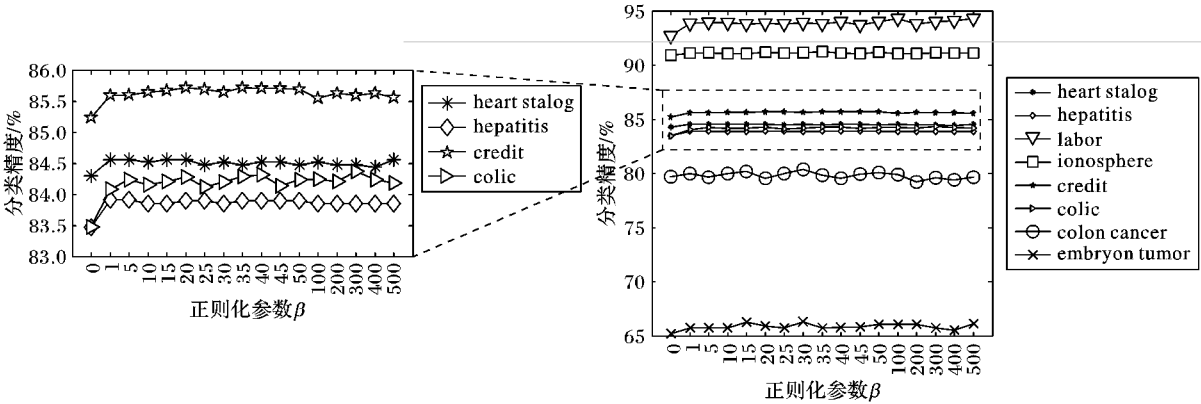


图 3 固定  $\lambda$  变化  $\beta$  时提出 D-LASSO 方法在 8 个数据集上的分类精度

5 结语

本文提出了一种新的判别性特征选择方法 D-LASSO,通过引入一个新的判别性正则化项,用于保留类内样本的紧致性和类间样本之间的分离性,从而帮助诱导出更具有判别能

力的特征,并利用 APG 算法优化提出的模型。在 8 个 Benchmark 数据集(即 6 个标准 UCI 数据集和 2 个基因表达式数据集)上,一系列的实验结果表明,与已有的特征选择方法相比,提出的模型不仅能够获得更好的分类性能,而且对参数  $\beta$  比较鲁棒。进一步工作将在更多更大规模的数据集上验

证提出的方法,并将其应用到医学图像分析中。

#### 参考文献:

- [1] GUYON I, ELISSEEFF A. An introduction to variable and feature selection [J]. *Journal of Machine Learning Research*, 2003, 3: 1157–1182.
- [2] HE X, CAI D, NIYOGI P. Laplacian score for feature selection [EB/OL]. [2014-10-10]. <http://people.cs.uchicago.edu/~niyogi/papersps/HeCaiNiyolapscore.pdf>.
- [3] YU L, LIU H. Feature selection for high-dimensional data: a fast correlation-based filter solution [EB/OL]. [2014-10-10]. <http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.68.2975>.
- [4] WESTON J, GUYON I. Support vector machine-recursive feature elimination (SVM-RFE): US, US8095483 B2[P]. 2010.
- [5] TIBSHIRANI R. Regression shrinkage and selection via the LASSO: a retrospective [J]. *Journal of the Royal Statistical Society*, 2011, 73(3): 273–282.
- [6] PUDIL P, NOVOCIOVA J, KITTLER J. Floating search methods in feature selection [J]. *Pattern Recognition Letters*, 1994, 15(11): 1119–1125.
- [7] NG A Y. Feature selection,  $L_1$  vs.  $L_2$  regularization, and rotational invariance [J]. *International Conferences on Machine Learning*, 2004, 19(5): 379–387.
- [8] ZHOU J, LU Z, SUN J, *et al.* FeaFiner: biomarker identification from medical data through feature generalization and selection [C]// *KDD 2013: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM Press, 2013: 1034–1042.
- [9] LIU F, WEE C Y, CHEN H. Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's disease and mild cognitive impairment identification [J]. *NeuroImage*, 2014, 84: 466–475.
- [10] ZOU H, HASTIE T. Regularization and variable selection via the elastic net [J]. *Journal of the Royal Statistical Society*, 2005, 67(2): 301–320.
- [11] TIBSHIRANI R, SAUNDERS M, ROSSET S. Sparsity and smoothness via the fused lasso [J]. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 2005, 67(1): 91–108.
- [12] YE G, XIE X. Split Bregman method for large scale fused lasso [J]. *Computational Statistics & Data Analysis*, 2011, 55(4): 1552–1569.
- [13] YAMADA M, JITKRITTUM W, SIGAL L. High-dimensional feature selection by feature-wise kernelized lasso [J]. *Neural Computation*, 2014, 26(1): 185–207.
- [14] CHEN X, PAN W, KWOK J T, *et al.* Accelerated gradient method for multi-task sparse learning problem [C]// *Proceedings of the Ninth IEEE International Conference on Data Mining*. Piscataway: IEEE Press, 2009: 746–751.
- [15] LIU J, YE J. Efficient  $L_1/L_q$  norm regularization [R]. Arizona: Arizona State University, 2009.
- [16] CAI D, HE X, ZHOU K. Locality sensitive discriminant analysis [C]// *Proceedings of the 2007 International Joint Conference on Artificial Intelligence*. [S. l.]: Morgan Kaufmann Press, 2007: 708–713.
- [17] XUE H, CHEN S, YANG Q. Discriminatively regularized least-squares classification [J]. *Pattern Recognition*, 2009, 42(1): 93–104.
- [11] ZHAO Y, QIN B, LIU T. Sentiment analysis [J]. *Journal of Software*, 2010, 21(8): 1834–1848. (赵妍妍, 秦兵, 刘挺. 文本情感分析 [J]. *软件学报*, 2010, 21(8): 1834–1848.)
- [12] YANG L, ZHU J, TANG S. Survey of text sentiment analysis [J]. *Journal of Computer Applications*, 2013, 33(6): 1574–1578. (杨立公, 朱俭, 汤世平. 文本情感分析综述 [J]. *计算机应用*, 2013, 33(6): 1574–1578.)
- [13] YANG A, LIN J, ZHOU Y, *et al.* Research on building a Chinese sentiment lexicon based on SO-PMI [J]. *Applied Mechanics and Materials*, 2013, 263/264/265/266: 1688–1693.
- [14] YANG A, ZHOU Y, LIN J. A method of Chinese texts sentiment classification based on Bayesian algorithm [J]. *Applied Mechanics and Materials*, 2013, 263/264/265/266: 2185–2190.
- [15] LU H, NIU Z, ZHANG N, *et al.* A model for sentiment classification of Chinese microblog based on parsing and theme extension [J]. *Transactions of Beijing Institute of Technology*, 2014, 34(8): 824–829. (陆浩, 牛振东, 张楠, 等. 基于句法与主题扩展的中文微博情感倾向性分析模型 [J]. *北京理工大学学报*, 2014, 34(8): 824–829.)
- [16] YANG J, YANG A, ZHOU Y. Sentiment classification method of Chinese micro-blog based on semantic analysis [J]. *Journal of Shandong University: Natural Science*, 2014(11): 1671–9352. (杨佳能, 阳爱民, 周咏梅. 基于语义分析的中文微博情感分类方法 [J]. *山东大学学报: 理学版*, 2014(11): 1671–9352.)
- [17] JIN X, ZHOU Y, MOBASHER B. A unified approach to personalization based on probabilistic latent semantic models of Web usage and content [C]// *Proceedings of the AAAI 2004 Workshop on Semantic Web Personalization*. Menlo Park: AAAI Press, 2004: 654–658.
- [18] CHU K, LI F. LDA model-based news topic evolution [J]. *Computer Applications and Software*, 2011, 28(4): 4–7, 26. (楚克明, 李芳. 基于 LDA 模型的新闻话题的演化 [J]. *计算机应用与软件*, 2011, 28(4): 4–7, 26.)
- [19] XU L, LIN H, PAN Y, *et al.* Constructing the affective lexicon ontology [J]. *Journal of the China Society for Scientific and Technical Information*, 2008, 27(2): 180–185. (徐琳宏, 林鸿飞, 潘宇, 等. 情感词汇本体的构造 [J]. *情报学报*, 2008, 27(2): 180–185.)
- [20] CHE W, LI Z, LIU T. LTP: a Chinese language technology platform [EB/OL]. [2014-10-10]. <http://ir.hit.edu.cn/~car/papers/coling10demo.pdf>.
- [21] WU J, TANG C, LI T, *et al.* Sentiment analysis on Web financial text based on semantic rules [J]. *Journal of Computer Applications*, 2014, 34(2): 481–485, 495. (吴江, 唐常杰, 李太勇, 等. 基于语义规则的 Web 金融文本情感分析 [J]. *计算机应用*, 2014, 34(2): 481–485, 495.)

(上接第 2751 页)