基于零范数特征选择的支持向量机模型

刘 峤 [1] 秦志光 [1] 陈 伟 [1] 张凤荔 [1]

摘 要 为解决高维稀疏建模问题, 本文从经验风险最小化原则出发推导出一个基于零范数约束的特征选择判据, 并利用嵌入式设计模式的特点将其与支持向量机方法相结合. 仿真实验和真实数据实验表明, 该方法不仅具备良好的特征选择性能, 而且在稀疏建模问题中表现出良好的分类准确性和泛化能力.

关键词 机器学习, 特征选择, 支持向量机, 稀疏建模, 正则化

DOI 10.3724/SP.J.1004.2011.00252

Zero-norm Penalized Feature Selection Support Vector Machine

LIU Qiao[1] QIN Zhi-Guang[1] CHEN Wei[1]
ZHANG Feng-Li[1]

Abstract To deal with the high-dimensional sparse modeling problem, this paper derives a zero-norm penalized feature selection criterion based on the the empirical risk minimization principle, and combines it with support vector machines through an embedded paradigm. Numerical results on both synthetic and real data sets show that the proposed approach does not only perform well for feature selection tasks, but also shows good performance compared to the conventional sparse modeling techniques, in the sense of classification accuracy and generalization capability.

Key words Machine learning, feature selection, support vector machine (SVM), sparse modeling, regularization

支持向量机 (Support vector machine, SVM) 是一种基于结构风险最小化原则的小样本推理统计机器学习方法. 由于其良好的理论特性, SVM 在模式识别、图像处理、生物信息学等领域获得了广泛应用[1]. 然而理论研究表明, 虽然小样本推理原则确保了 SVM 方法在高维特征空间的模式分类任务中的有效性, 但随着问题规模的扩大, 大量的冗余特征和不相干特征会对 SVM 分类器的性能造成严重干扰[2].

因此单纯依靠 SVM 方法来解决高维特征空间的模式分类问题是不够的, 需要借助特征选择来实现降维. 近年来的研究实践表明通过特征选择能够有效改善 SVM 算法的性能[3]. 针对 SVM 的特征选择方法学研究有两种主要的思路. 一种是基于过滤器的设计模式, 即采用特征排序的方法选择最优的特征子集对目标任务进行特征建模[4]. 另一种是基于封装器的设计模式, 通常采用基于范数约束的正则化方法对 SVM 的输入特征进行选择. 过滤器模式的算法效率高, 且特征选择结果不依赖于特定的机器学习算法, 其主要缺点是需要根据传统统计学方法 (如 t-检验) 来设定特征选择阈值, 而

收稿日期 2009-12-15 录用日期 2010-10-22
Manuscript received December, 15, 2009; accepted October 22, 2010
国家高技术研究发展计划 (863 计划) (2006AA01Z411) 资助
Supported by National High Technology Research and Development Program of China (863 Program) (2006AA01Z411)
1. 电子科技大学计算机科学与工程学院 成都 610054
1. School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054

基于范数约束的正则化特征选择方法则有效地解决了这一问题, 因此在近年来引起了广泛关注和深入研究[5].

由于 2-范数约束的正则化特征选择方法不具备特征选择能力, 因此 1-范数方法在近年来成为了主流研究方向[6]. 然而 1-范数约束方法也存在着一些难以克服的理论缺陷, 如目标函数的凸性不严格, 特征选择能力受样本容量限制等, 极大地制约了该方法的研究和应用[7]. 最近的研究表明, 基于 1-范数约束的特征选择方法在稀疏特征空间的决策风险相对于零范数约束方法渐近地趋于无穷大, 这一结论为突破 1-范数约束的方法框架提供了理论依据[8]. 将零范数约束方法引入支持向量机的主要困难在于目标函数非凸, 导致相应的二次规划问题难以求解. 为解决该问题, 本文通过借鉴嵌入式模式设计思想[9], 提出一个具备自动特征选择能力的支持向量机分类模型, 称为 0-Norm SVM. 实验数据表明, 0-Norm SVM 不仅具备良好的自动特征选择能力, 同时保持了传统 SVM 方法在小样本学习任务中的性能优势和泛化能力.

1　方法细节

0-Norm SVM 方法模型由两部分构成: 首先在原始特征空间通过特征评估, 在可接受的信息损失风险下筛选出一组原始特征子集; 然后采用 SVM 方法从观测数据出发针对该特征子集进行建模. 以下分别进行说明.

1.1　构造 0-Norm 特征选择算法

由于任何分类问题均可被分解为若干二分类问题. 不失一般性, 本文在如下高斯线性模型基础上对二分类问题展开讨论. 设 $\boldsymbol{y}$ 表示 n 维响应向量 ($y_i \in \{-1, 1\}$), 设计矩阵 X 代表 p 维征空间的 n 个数据样本的观测值, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^{\mathrm{T}}$ 表示 p 维特征参数向量, 以 $\boldsymbol{\epsilon} \sim \mathrm{N}(0, \sigma^2)$ 表示高斯噪音数据. 设响应向量 $\boldsymbol{y}$ 的真实生成模型为

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1}$$

特征选择的目标是在原始特征空间中寻找满足如下优化条件的 $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \cdots, \hat{\beta}_p)$:

$$\hat{\beta} = \arg\min_{\beta} \left\{ \|\boldsymbol{y} - X\boldsymbol{\beta}\|^2 \right\} \tag{2}$$

对于典型的稀疏建模问题有 $p \gg n$, 且 $\boldsymbol{\beta}$ 中的多数元素为零. 即协方差矩阵 $X^{\mathrm{T}}X$ 为奇异阵, 因此问题 (2) 不适定. 为解决该矛盾, 首先向目标函数 (2) 引入一个 2-范数约束条件:

$$\sum_{i=1}^{p} \beta_i^2 \ = a, \quad a > 0 \tag{3}$$

引入 2-范数约束不仅解决了稀疏建模的不适定问题, 而且能够显著提高模型泛化能力, 有效避免过拟合. 此外由于约束条件的严格凸性, 确保了优化解的存在性和唯一性[5]. 然而对 $\hat{\boldsymbol{\beta}}$ 施加 2-范数约束的结果只是实现了参数的平滑, 而无法将那些冗余的和不相干的特征排除在模型之外. 为实现特征选择的目标, 考虑对目标函数 (2) 进一步施加零范数约束:

$$\|\boldsymbol{\beta}\|_0 \ = b, \quad b > 0 \tag{4}$$

根据拉格朗日对偶理论将式 (4) 改写为如下的对偶问题:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\beta} \left\{ \|Y - X\boldsymbol{\beta}\|^2 + \xi\|\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|_0 \right\} \tag{5}$$

其中, 拉格朗日乘子 $\xi > 0$, $\lambda > 0$. 首先考虑在相同维度的特征模型间进行比较的情况, 此时 $\|\boldsymbol{\beta}\|_0$ 可以被视为常数. 对于给定维度的特征模型, 问题 (5) 退化为一个严格的凸优化问题, 其最优解为 $\hat{\boldsymbol{\beta}} = (X^{\mathrm{T}}X + \xi I_n)^{-1}X^{\mathrm{T}}\boldsymbol{y}$.

接下来考虑在不同维度的特征模型间进行选择的情况. 由模型 (1) 知 $\boldsymbol{y} - X\boldsymbol{\beta} \sim \mathrm{N}(0, \sigma^2)$. 以 $\log_2 f(\boldsymbol{x}^n|\boldsymbol{\beta})$ 表示从观测数据出发得到的对数似然函数, 则: $\|Y - X\hat{\boldsymbol{\beta}}\|^2 \propto -\log_2 f(\boldsymbol{x}^n|\hat{\boldsymbol{\beta}})$. 以 $\log_2 g(\boldsymbol{x}^n|\boldsymbol{\beta})$ 表示数据真实生成模型的对数联合密度函数. 以二者偏差的数学期望来刻划决策风险:

$$Risk = \mathrm{E}_{\beta}\{-\log_2 f_{\boldsymbol{\beta}}(\boldsymbol{x}^n) - [-\log_2 g_{\boldsymbol{\beta}}(\boldsymbol{x}^n)]\} \tag{6}$$

按照经验风险最小化原则, 问题 (5) 的优化目标等同于求解 $\hat{\boldsymbol{\beta}}$, 使得在任意条件下的决策风险均达到下界, 即 Minimax 最优. 根据随机复杂度理论, 当 $n \to \infty$ 时, 决策风险 (6) 在最坏情况下的下界 (极大极小值) 为[10]:

$$\min_{f}\max_{\boldsymbol{\beta}\in\Theta} Risk = \frac{\|\boldsymbol{\beta}\|_0}{2}\log_2\frac{n}{2\pi\mathrm{e}} + \log_2\int_K \sqrt{\det I(\boldsymbol{\beta})}\mathrm{d}\boldsymbol{\beta} + o(1) \tag{7}$$

其中 $I(\boldsymbol{\beta})$ 为参数分布的 Fisher 信息矩阵, K 为参数空间 Θ 的紧子集. 将 σ^2 的极大似然估计 $\hat{\sigma}^2 = \|Y - X\boldsymbol{\beta}\|^2/n$ 代入式 (7), 积分后合并与参数模型无关的常数项到 C, 得到:

$$\min_{f}\max_{\boldsymbol{\beta}\in\Theta} Risk = \left(\log_2 n - \frac{1}{2}\log_2\|Y - X\hat{\boldsymbol{\beta}}\|^2\right)\|\boldsymbol{\beta}\|_0 + C \tag{8}$$

由于式 (8) Minimax 最优, 因此 $\lambda = \log_2 n - \frac{1}{2}\log_2\|Y - X\hat{\boldsymbol{\beta}}\|^2$ 为式 (5) 的最优解, 由此确定了模型 (5) 中的待定参数 λ. 对另一个待定参数 ξ 的选择可以通过交叉验证方法经验得到. 在实际应用中, 由于数据真实生成模型 $g(\boldsymbol{x}^n|\boldsymbol{\beta})$ 不可知, 为求得最优的特征子集, 本文利用问题 (5) 对给定 $\|\boldsymbol{\beta}\|_0$ 是严格凸函数的特点, 采用前向分步回归法迭代求解.

1.2　构造 SVM 分类器

接下来在前述求得的最优特征子集上采用 SVM 方法构造分类器. 首先为确保样本可分, 采用非线性映射 $\phi(\boldsymbol{x})$ 将数据映射到高维特征空间. 由于本文的目标是解决支持向量机在高维特征空间的稀疏建模问题, 因此选择在先验信息缺乏的情况下仍然泛化能力良好的高斯径向基核函数[11]:

$$K(\boldsymbol{x}, \boldsymbol{x}^{\mathrm{T}}) = \langle\phi(\boldsymbol{x}), \phi(\boldsymbol{x}^{\mathrm{T}})\rangle = \exp\left(-\frac{\|\phi(\boldsymbol{x}) - \phi(\boldsymbol{x}^{\mathrm{T}})\|^2}{2\delta^2}\right)$$

针对二分类问题 (1), 训练学习的目标是从训练数据出发寻找超平面 $\langle\boldsymbol{w}, \phi(\boldsymbol{x})\rangle + b = 0$, 其中 $\boldsymbol{w} \in \mathbf{R}^p, b \in \mathbf{R}$, 使得两类样本集合与该超平面平行的边界之间的距离最大化. 引入拉格朗日乘数 $\boldsymbol{\alpha} \in \mathbf{R}^n$ 后, 目标函数可以改写为

$$\Psi(\boldsymbol{w}, b, \boldsymbol{\alpha}) = \frac{\|\boldsymbol{w}\|^2}{2} - \sum_{i=1}^{n}\alpha_i\left(\boldsymbol{y}_i\left[\langle\boldsymbol{w}, \phi(\boldsymbol{x}_i)\rangle + b\right] - 1\right) \tag{9}$$

对偶优化目标是:

$$\max_{\boldsymbol{\alpha}} W(\boldsymbol{\alpha}) = \max_{\boldsymbol{\alpha}}(\min_{\boldsymbol{w}, b}\Psi(\boldsymbol{w}, b, \boldsymbol{\alpha})) \tag{10}$$

求解该问题得到最优分类面的参数如下:

$$\boldsymbol{w}^* = \sum_{i=1}^{n} \alpha_i \boldsymbol{y}_i \phi(\boldsymbol{x}_i)$$
$$b^* = -\frac{1}{2}\langle \boldsymbol{w}^*, (\phi(\boldsymbol{x}_m) + \phi(\boldsymbol{x}_n))\rangle \tag{11}$$

其中, $\phi(\boldsymbol{x}_m)$ 和 $\phi(\boldsymbol{x}_n)$ 为满足如下条件的任意支持向量: $\alpha_m > 0$, $\alpha_n > 0$, $\boldsymbol{y}_m = 1$, $\boldsymbol{y}_n = -1$. 由此得到分类器模型如下:

$$\hat{\boldsymbol{y}}_i = \mathrm{sgn}\left(\langle \boldsymbol{w}^*, \phi(\boldsymbol{x}_i)\rangle + b^*\right) \tag{12}$$

以上建立了 0-Norm SVM 方法的数学模型, 应用该模型对稀疏数据进行建模的步骤为: 首先根据第 1.1 节给出的决策风险评估指标, 在特征空间搜索出一个最优特征子集. 然后应用本节描述的高斯径向基支持向量机算法对训练数据进行学习, 在最优子集上得到有关分类器的模型参数. 最后通过一组验证数据检验模型的分类性能和泛化能力.

2 实验结果分析

2.1 实验数据描述

本文采用的仿真数据通过伪随机算法生成. 其中样本容量 $n = 100$, 特征维度 $p = 100$, 真实模型中仅包含 10 个有效特征. 数据构造方法为首先按照正态分布 $\mathrm{N}(0, I_n)$ 随机生成一个 100×100 的设计矩阵 X. 然后按照正态分布 $\mathrm{N}(0, 1)$ 生成 10 个随机数作为特征的真实权重, 在 p 维零向量 $\boldsymbol{\beta}$ 中随机选择 10 个特征位置, 将上述权重值分别赋给这 10 个特征位置. 接下来按照分布 $\mathrm{N}(0, 0.1)$ 生成一个 100 维的高斯噪音向量 $\boldsymbol{\epsilon}$, 最后根据模型 (1) 计算得到观测向量 $\boldsymbol{y}$. 采用人工合成数据的好处在于数据的真实生成模型已知, 便于对特征选择实验的输出结果进行精确比对, 为此记录 $\boldsymbol{\beta}$ 中包含的 10 个真实特征的位置索引如下: {7, 11, 41, 54, 55, 57, 64, 67, 87, 95}.

为客观评价 0-Norm SVM 算法在处理真实稀疏建模问题时的性能, 本文进一步在三组公开基因芯片数据集上对该算法进行了验证, 并就实验结果与相关工作进行了比较. 这三组数据分别是白血病数据集 (Leukemia), 肺癌数据集 (Lung cancer) 和前列腺癌数据集 (Prostate cancer)[12]. 上述数据的共同特点是样本容量小, 特征维度高, 且特征高度相关, 特征冗余度高, 是典型的稀疏建模问题. 有关数据集的构成情况总结于表 1. 原始数据来源是: http://homes.esat.kuleuven.be/∼npochet/Bioinformatics/.

表 1 三组基因芯片数据集的构成情况

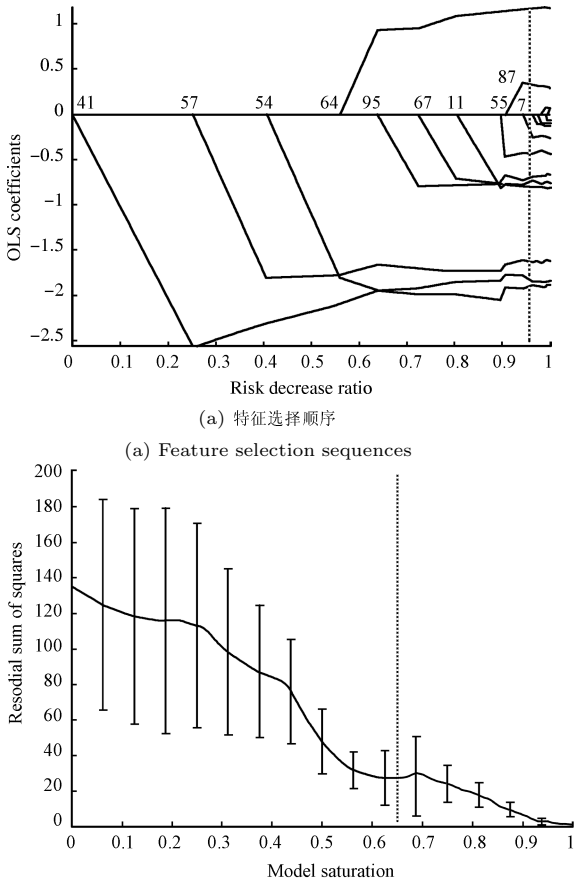
Table 1 Summaries of the three microarray datasets

数据集	样本容量	训练集	测试集	特征维度
Leukemia	72	38	34	7 129
Lung cancer	181	32	149	12 533
Prostate cancer	136	102	34	12 600

2.2 0-Norm SVM 算法的特征选择能力

为验证算法的特征选择能力, 首先通过仿真实验考察最优子集的选择过程. 为作图需要, 对特征评估标准进行了调整, 设定选出的特征数为 15. 实验结果参见图 1, 其中图 1 (a) 显示了算法运行期间特征顺序进入模型的情况. 图中的横坐标为归一化后的模型决策风险降低率, 纵坐标为相应模型的 OLS 回归系数. 图中与纵坐标轴平行的虚线表示 0-Norm SVM 的特征选择过程正常结束的位置, 可以看出 0-Norm SVM 算法恰好选出了 10 个真实特征.

图 1 (b) 一步给出了 0-Norm SVM 算法顺序选出的 15 个特征子集在仿真数据集上的 10 轮交叉验证结果. 其中横坐标为归一化后的模型饱和度, 纵坐标为残差平方和. 误差条形图的中心点纵坐标为交叉验证得到的残差均值, 误差限为残差的方差. 从中可以看出, 该算法通过特征选择实现了对模型方差与偏差之间的折衷, 有效避免了过拟合情况的出现.



(a) 特征选择顺序

(a) Feature selection sequences

(b) 十轮交叉验证结果对比

(b) 10-folds cross validation results

图 1 0-Norm SVM 算法在仿真数据集上的实验结果

Fig. 1 Experimental results on synthetic dataset

为进一步验证 0-Norm 特征选择判据的有效性, 在样本容量 $n = 100$ 的条件下设计了三组仿真试验. 这三组试验所选择的特征空间维数 p 分别为 $10^2, 10^3$ 和 10^4, 每一组仿真数据中的有效特征数均为 10. 参与比较的方法包括: 贝叶斯信息判据方法 (Bayesian information criterion, BIC), 基于 1-范数约束的 LARS 方法, 和采用 1-2 范数混合约束的弹性网络 (Elastic net, EN) 方法. 所有三组试验均采用 10 轮交叉验证, 结果取均值记入表 2. 根据 LARS 和 EN 算法中 1-范数约束条件的特点, 本文通过动态参数调整来控制选出的特征数. 由于已知三组数据中的 "真实" 特征数均为 10, 因此在表 2 中记录了这两种算法筛选出的前 10 个特征的统计数据.

观察表 2 可以看出在所有四种特征选择方法中, BIC 方法的性能表现最差, 不仅在各组数据上均存在严重的过拟合

现象, 而且随着数据特征的稀疏程度加大, 该方法的性能出现明显恶化, 表明传统信息判据方法与本文所提出的零范数方法相比, 在稀疏建模问题上存在较大的应用局限性. 接下来考察 LARS 和 EN 方法的性能表现, 从表 2 可以看出这两种方法的特征选择准确性 (漏报率和误报率指标) 仅在低维饱和数据集上与 0-Norm 方法相接近, 在其余两组稀疏数据集上的漏报率指标均明显高于 0-Norm 方法, 其中在特征稀疏度极高的第三组数据上漏报率甚至接近 45 %. 由此说明相对于 0-Norm 方法而言, LARS 和 EN 方法对噪音数据比较敏感, 在稀疏建模问题上, 0-Norm 方法的性能稳定性和特征定位准确性更好.

表 2　0-Norm SVM 与其他经典方法的仿真测试结果对比

Table 2　Comparative study of the 0-Norm SVM and other typical feature selection methods on synthetic data

P	Method	# Recall	FPR (%)	FNR (%)
10^2	0-Norm	10.62 ± 0.74	0.68 ± 0.82	0.00 ± 0.00
	BIC	16.99 ± 3.50	7.77 ± 3.89	0.00 ± 0.00
	LARS	10.00 ± 0.00	1.15 ± 0.79	10.38 ± 7.13
	EN	10.00 ± 0.00	1.12 ± 0.79	10.12 ± 7.13
10^3	0-Norm	10.65 ± 2.20	0.23 ± 0.22	15.94 ± 5.07
	BIC	59.80 ± 1.52	5.17 ± 0.16	13.52 ± 6.55
	LARS	10.00 ± 0.00	0.30 ± 0.05	30.02 ± 4.63
	EN	10.00 ± 0.00	0.30 ± 0.04	29.70 ± 4.44
10^4	0-Norm	10.08 ± 2.20	0.02 ± 0.02	16.88 ± 4.64
	BIC	89.00 ± 0.00	0.81 ± 0.01	16.00 ± 5.16
	LARS	10.00 ± 0.00	0.04 ± 0.01	44.48 ± 5.10
	EN	10.00 ± 0.00	0.04 ± 0.01	44.50 ± 5.14

可见与其他三种经典特征选择方法相比, 0-Norm 方法在各组试验中都占有性能优势, 更有利于抑制模型对数据的过拟合, 也更能够适应在高维特征空间进行稀疏建模的需求.

2.3　0-Norm SVM 算法的分类准确性

为验证 0-Norm SVM 算法的分类准确性, 本文在三组基因数据集上采用原始划分的训练集和数据集进行了测试, 测试结果记录于表 3. 实验前对全部基因数据进行了零均值规格化处理, 实验过程中 SVM 参数取相同值: $C = 10^3$, $\delta^2 = 1$.

表 3 由三部分内容构成, 首先, 为考察引入特征选择机制对 SVM 算法分类能力的影响, 本文对 0-Norm SVM 算法中包含的 SVM 分类算法进行了单独测试, 并以该实验结果作为基准 (Baseline). 然后采用 0-Norm SVM 算法在同样条件下进行测试, 并就实验结果与前述基准测试结果进行参照比对.

最后, 为客观评价 0-Norm SVM 算法的分类性能, 本文对部分近期发表的相关工作的实验结果进行了比较, 其中基于特征独立模拟退火规则的分类方法 (Features annealed independence rules, FAIR) 的实验数据引自文献 [12], 经验贝叶斯方法 (Empirical Bayes, EB) 的实验数据引自文献 [13].

通过与基准测试数据比较可以看出, 在全部三组基因芯片数据集上, 采用同样的支持向量机算法, 基于零范数约束的特征选择机制能够有效提高支持向量机的分类准确性. 这一方面验证了 0-Norm SVM 算法中特征选择算法的有效性, 另一方面表明, 通过特征选择去除原始特征空间中的不相干或冗余特征后, 有助于提高支持向量机方法的分类准确率.

表 3　0-Norm SVM, FAIR 和 EB 算法在三组基因芯片数据集上的训练与分类误差

Table 3　Classification accuracies of the 0-Norm SVM, FAIR, and EB algorithms on three microarray datasets

数据集	分类算法	训练误差	测试误差	特征维度
Leukemia	Baseline	0/38	14/34	7 129
	0-Norm	0/38	1/34	27
	FAIR	1/38	1/34	11
	EB	0/38	3/34	7 129
Lung	Baseline	0/32	15/149	12 533
	0-Norm	0/32	4/149	22
	FAIR	0/32	7/149	31
	EB	0/32	1/149	12 533
Prostate	Baseline	0/102	25/34	12 600
	0-Norm	0/102	3/34	37
	FAIR	10/102	9/34	2
	EB	38/102	4/34	12 600

通过与相关工作进行比较可以看出 0-Norm SVM 算法在全部三组数据集上的表现均优于同样引入了特征选择机制的 FAIR 算法. 对于采用全部基因特征进行学习和分类的 EB 算法, 0-Norm SVM 算法也取得了 2 : 1 的优势, 仅在肺癌数据集上的测试误差指标略高. 此外, 相对于其他两种算法结果的不稳定, 0-Norm SVM 分类算法的测试和训练误差在三组数据集上表现出稳定均衡的优势.

2.4　0-Norm SVM 算法的泛化能力

如前所述, SVM 方法在小样本学习任务中具有良好的泛化能力是其突出的理论优点之一. 本节将通过实验证明, 在引入了特征选择机制后, 0-Norm SVM 算法同样保持了良好的泛化能力. 本文采用文献 [6] 的实验方法, 对三组基因数据分别按照 1 : 1 的比例随机划分为训练集和测试集 (正反样本的相对比例与原始数据集保持一致), 然后采用 0-Norm SVM 进行学习和分类测试, 同样条件下重复实验 30 轮, 对各轮测试得到的分类准确率指标取均值和方差计入表 4. 作为参照, 表 3 还引用了文献 [6] 中 2-Norm SVM 和 1-Norm SVM 的实验结果.

表 4　基于不同范数约束的支持向量机方法的泛化能力比较

Table 4　Comparison of the generality capabilities of the support vector machines with different norm penalties

数据集	0-Norm SVM	1-Norm SVM	2-Norm SVM
Leukemia	91.4 ± 2.3	90.1 ± 1.5	82.7 ± 1.0
Lung	92.2 ± 1.3	92.6 ± 0.5	92.9 ± 0.4
Prostate	94.0 ± 1.5	88.5 ± 0.8	90.8 ± 0.6

通过观察表 4 可知 0-Norm SVM 在白血病和前列腺癌数据集上的分类精度的均值优于其他两类范数约束的支持向量机方法, 在肺癌数据集上的表现与其他两种方法相当, 但 0-Norm SVM 方法的方差稍高. 从总体上看, 这三类方法的

平均分类性能表现十分接近, 均具有良好的分类准确性和泛化能力, 其中 0-Norm SVM 在分类准确性上占有一定的优势.

3 结论

本文的主要目标是解决支持向量机方法在稀疏建模问题中的特征选择问题. 由于 2-范数约束方法不具备自动特征选择能力, 1-范数约束方法又存在难以克服的理论局限性, 因此本文利用了嵌入式设计模式中特征评估过程与参数估计过程无关的特性, 从经验风险最小化原则出发构造出一个可证优化的 0-范数特征选择判据, 并采用分步回归的方法实现了对该问题的经验求解. 仿真实验和真实基因芯片数据上的实验结果表明, 0-范数约束方法不仅具备良好的自动特征选择能力, 而且通过与基于高斯径向基核函数的支持向量机配合, 能够在稀疏建模任务中取得明显的性能优势. 此外, 0-Norm SVM 算法的性能表现十分稳定, 保持了支持向量机方法在小样本学习问题中泛化能力好的优良特性. 综上所述, 0-Norm SVM 算法的提出为稀疏建模方法学研究提供了一个有希望的解决方案, 由于稀疏建模问题是多个交叉学科领域当前的研究热点, 因此本文的成果具有一定的理论和应用价值.

References

1 Wang Hong-Wei, Qi Chao, Wei Yong-Chang, Li Bin, Zhu Song. Review on data-based decision making methodologies. *Acta Automatica Sinica*, 2009, **35**(6): 820−833
(王红卫, 祁超, 魏永长, 李彬, 朱松. 基于数据的决策方法综述. 自动化学报, 2009, **35**(6): 820−833)

2 Forman G. An extensive empirical study of feature selection metrics for text classification. *The Journal of Machine Learning Research*, 2003, **3**: 1289−1305

3 Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Machine learning*, 2002, **46**(1−3): 389−422

4 Li T, Zhang C L, Ogihara M. A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression. *Bioinformatics*, 2004, **20**(15): 2429−2437

5 Zhu J, Rosset S, Hastie T, Tibshirani R. 1-norm support vector machines. In: Proceedings of the Advances in Neural Information Processing Systems. Cambridge, USA: The MIT Press, 2004. 49−56

6 Kujala J, Aho T, Elomaa T. A walk from 2-norm SVM to 1-norm SVM. In: Proceedings of the 9th IEEE International Conference on Data Mining. Miami, USA: IEEE, 2009. 836−841

7 Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B*, 2005, **67**(2): 301−320

8 Lin D, Pitler E, Foster D P, Ungar L H. In defense of l. In: Proceedings of ICML Workshop on Sparse Optimization and Variable Selection. Helsinki, Finland: The Omnipress, 2008.

9 Guyon I, Elisseeff A. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 2003, **3**: 1157−1182

10 Myung J I, Navarro D J, Pitt M A. Model selection by normalized maximum likelihood. *Journal of Mathematical Psychology*, 2006, **50**(2): 167−179

11 Yu Zhong-Dang, Wang Long-Shan. Research on image noise suppression algorithm based on LS-SVR. *Acta Automatica Sinica*, 2009, **35**(4): 364−370
(于忠党, 王龙山. 基于 LS-SVR 的图像噪声去除算法研究. 自动化学报, 2009, **35**(4): 364−370)

12 Fan J Q, Fan Y Y. High dimensional classification using features annealed independence rules. *Annals of statistics*, 2008, **36**(6): 2605−2637

13 Greenshtein E, Park J. Application of non parametric empirical Bayes estimation to high dimensional classification. *The Journal of Machine Learning Research*, 2009, **10**: 1687−1704

刘 峤 电子科技大学计算机科学与技术学院博士研究生. 主要研究方向为智能信息处理. 本文通信作者. E-mail: cnliuqiao@gmail.com
(**LIU Qiao** Ph. D. candidate at the School of Computer Science and Engineering, University of Electronic Science and Technology of China. His main research interest is intelligent information processing. Corresponding author of this paper.)

秦志光 电子科技大学计算机科学与技术学院教授. 主要研究方向为信息安全. E-mail: qinzg@uestc.edu.cn
(**QIN Zhi-Guang** Professor at the School of Computer Science and Engineering, University of Electronic Science and Technology of China. His main research interest is information security.)

陈 伟 电子科技大学计算机科学与技术学院博士研究生. 主要研究方向为无线网络路由和网络安全. E-mail: chenwei@uestc.edu.cn
(**CHEN Wei** Ph. D. candidate at the School of Computer Science and Engineering, University of Electronic Science and Technology of China. His research interest covers wireless routing protocols and network security.)

张凤荔 电子科技大学计算机科学与技术学院教授. 主要研究方向为空间数据库与数据挖掘技术. E-mail: fzhang@uestc.edu.cn
(**ZHANG Feng-Li** Professor at the School of Computer Science and Engineering, University of Electronic Science and Technology of China. Her research interest covers spacial database and data-mining technology.)