

# SVM 参数优化方法分析与决策

郭克友, 郭晓丽, 王艺伟

(北京工商大学 材料与机械工程学院, 北京 100048)

**摘要:** 针对支持向量机应用过程中的参数选择问题, 从 UCI 数据库选择样本集, 分别采用传统的网格法、智能优化算法中的粒子群法及遗传算法实现核函数参数寻优过程, 将所得最佳参数应用到样本测试中; 在深入分析优化过程中各参数关系、参数对支持向量机性能的影响以及传统与智能优化算法的优劣势, 得出了核函数优化策略; 即先使用智能优化算法初步确定最优解范围, 再结合网格法进行高精度寻优; 实验数据验证了参数优化策略的有效性, 为扩大支持向量机泛化率、提高应用性做了铺垫。

**关键词:** 支持向量机; 核函数; 传统优化算法; 智能优化算法

## Analysis and Strategy for Parameter Optimization Methods of SVM

Guo Keyou, Guo Xiaoli, Wang Yiwei

(School of Material and Mechanical Engineering, Beijing Technology and Business University, Beijing 100048, China)

**Abstract:** To select parameters of support vector machine (SVM) during application, we choose sample sets from UCI database. The parameters optimization of SVM with kernel function is achieved through traditional grid method, particle swarm optimization algorithm and genetic algorithm in intelligent optimization algorithms. And the optimum parameters are applied to testing samples. The kernel function optimization strategy is obtained after a thorough analysis of the relationship between parameters, the influence of parameters on the performance of SVM and the pros and cons of both traditional and intelligent optimization algorithms. The users should use intelligent optimization algorithm preliminarily to make sure the scope of parameters, and couple it with grid method to obtain a high degree of accuracy. Experimental data verify the effectiveness of the strategy, which expands the generalization and application of SVM.

**Keywords:** support vector machine; kernel function; traditional optimization algorithms; intelligent optimization algorithms

## 0 引言

支持向量机 (support vector machine, SVM), 可利用核函数在原空间直接计算, 既能避免“维数灾难”, 又能高效地解决非线性分类问题, 成为浅层机器学习算法中的优秀代表。人工神经网络和基于规则的分类器是以贪心学习为策略对假设空间搜索最优解, 但容易陷入局部解, 而且会出现“过拟合”<sup>[1]</sup>。SVM 可以表示为凸优化问题, 一般依靠小样本进行训练学习, 将待分类问题转化到高维线性可分空间, 找到最大分类间隔, 得到目标函数的全局最优解。Vapnik 等人<sup>[2-3]</sup>的研究成果表明, 在解决实际问题时, 核函数的参数及惩罚因子对 SVM 性能影响很大, 因此研究参数选取问题具有很大的意义。

## 1 数据准备

样本来自加州大学欧文分校 (University of California Irvine) 的 UCI 数据库<sup>[4]</sup>。该数据库中为样本提供了特征向量, 可避免使用者设计特征提取算法或因主观因素而造成的特征选取不当问题。实验中将类别标签命名为 1、2 等数字, 样本按 8:2 随机划分为训练和测试样本。为充分利用给定数据集, 降低 VC 维, 训练过程采用交叉验证法思想。选取的数据集信息如表 1 所示, 基本分布如图 1。经过多次试验测试, 得到如下实验数据集选取原则:

收稿日期:2016-03-24; 修回日期:2016-04-21。

基金项目:交通运输部信息化科技项目(2012-364-835-110);北京工商大学科研能力计划项目。

作者简介:郭克友(1975-),男,黑龙江省齐齐哈尔市人,博士,副教授,主要从事数字图像处理,模式识别,智能交通方向的研究。

表 1 UCI 标准数据库部分数据

序号	数据集名称	样本总数	属性维数	类别	类别	类别比例
1	abalone	4177	8	3	I;M;F	1342:1528:1307
2	au	25000	45	6	class1:class2: class3:class4: class5:class6	8345:1987:7981: 1813:1565:3309
3	Balance-scale	624	4	3	B;L;R	49:288:288
4	blood-transfusion	747	5	2	0:1	470:177
5	Iris	150	4	3	Iris-setosa; Iris-versicolor; Iris-Virginica	50:50:50
6	wine	178	13	3	1:2:3	59:71:48

- 1) 数据集的实例个体数目分布多样化;
- 2) 实例的特征向量维数多样化, 因为维数对机器学习训练时间和模型影响较大;
- 3) 应包含二分类和多分类情况;
- 4) 样本类别应均衡分布, 针对不均衡或欠采样的数据, SVM 需做其他处理, 或难以确定分类器性能<sup>[5-6]</sup>。

图 1 显示了 Abalone、Au 和 Wine 数据集的特征向量维数较多, Abalone、blood-transfusion 和 wine 数据集个别向量在维度上的值跨度较大。Balance-scale 数据集分布较为均衡。

## 2 数据预处理多参数全局车道线检测模型

### 2.1 数据归一化

特征向量属性不同, 在数值分布上的差别会导致奇异样本

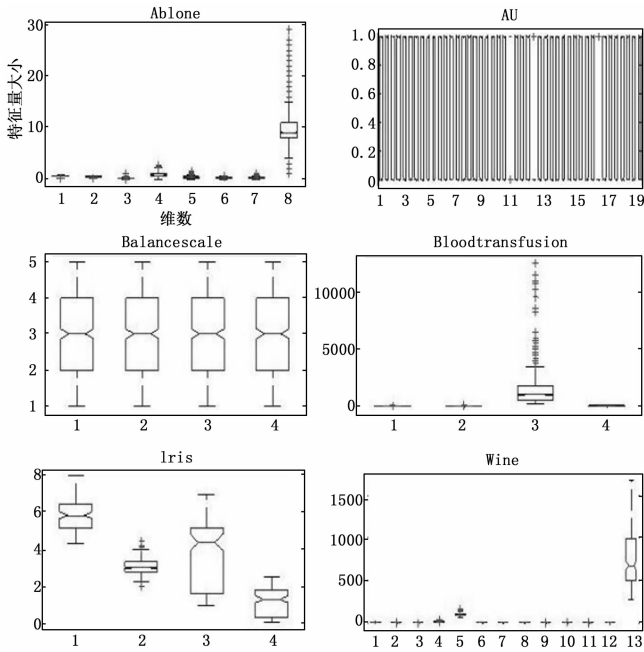


图 1 数据集样本分布盒图

数据, 引起网络不收敛, 训练时间增长等问题, 所以需要特征进行数量级统一, 即归一化。国内外学者的研究表明, 在代价函数中, 奇异样本数据中的大值特征分量对结果的影响大于小值特征分量, 特征归一化会使分类器的识别率提高, 加快训练速度<sup>[7]</sup>, 归一化公式如 (1) 所示:

$$y_i = (y_{\max} - y_{\min})(x_i - x_{\min}) / (x_{\max} - x_{\min}) + y_{\min} \quad (1)$$

其中:  $x_i$  为样本集,  $y_i$  为归一化后的值。

考虑每个维度的量纲不同, 且数量级差别悬殊, 为尽量减少样本原始信息的丢失, 应对每个维度单独归一化。经对图 1 数据分布盒形图分析, 将数据集按照公式 (1) 进行归一化到  $[-1, 1]$  区间, 归一化前后数据集样本分布如图 2 所示。

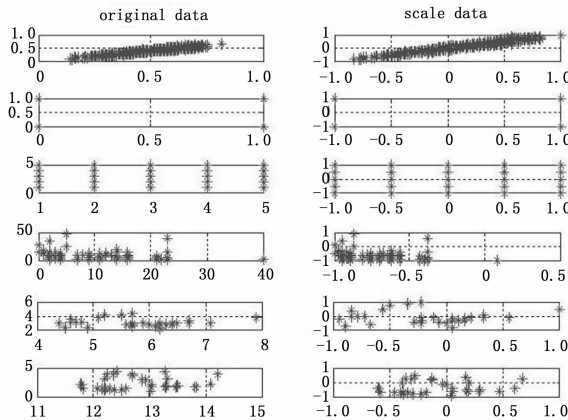


图 2 归一化后数据集样本分布

### 2.2 PCA 数据降维

统计学上, SVM 的基础是 VC 维理论和结构风险最小原理。在保证分类精度的同时, 尽量降低机器学习的 VC 维, 才能使模型在给定数据上的期望风险得到有效控制。降低 VC 维重要的手段就是对样本的特征向量进行降维处理。主成分分析法 (principal component analysis, PCA) 在最小均方意义下寻

找最能代表原始数据的投影方向, 能保存数据内在结构和原始信息, 已成功运用在人脸识别领域。本文所用实验数据维数范围是 4~45, 根据一般降维原则, 提取 90% 的主成分。6 个数据集的成分占比如图 3 所示。

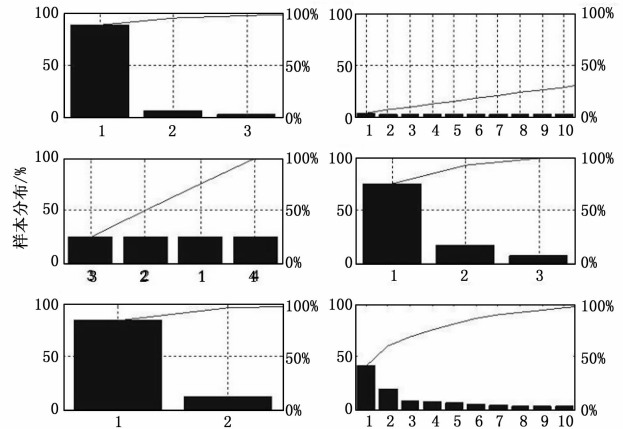


图 3 数据集主样本分布

### 3 SVM 核函数参数选取

支持向量机通过引进内积核函数巧妙的解决了高维空间中的内积运算, 而不同的核函数选择也构成了不同的支持向量机。常用的核函数有: 线性核、多项式核、高斯核、二层神经网络核, 一般高斯核最为有效<sup>[8]</sup>。设定 2 分类问题, 在 SVM 理论中, 若将 SVM 分类超平面记为  $l^{(i)}$ , 表示其相似度的高斯核函数记为  $f^{(i)}$ , 如式 (2)。则高斯支持向量机的目标函数 (cost function)  $F$  可记为式 (3):

$$F = \min\{C \sum_{i=1}^m [y^{(i)} \cos t_1(\theta^T f^{(i)}) + (1 - y^{(i)}) \cos t_0(\theta^T f^{(i)})] + \frac{1}{2} \sum_{i=1}^n \theta_j^2\} \quad (2)$$

$$f^{(i)} = \exp\left(-\frac{\|x - l^{(i)}\|^2}{2\sigma^2}\right) \quad (3)$$

在高斯核函数进行训练时, 需确定惩罚因子  $C$  和参数  $\sigma$ , 即通过适当的算法选取最优化 ( $C, \sigma^2$ ) 参数组, 提高 SVM 的推广性。设定  $g = \frac{1}{\sigma^2}$ , 找出推广识别率最高的  $C$  和  $g$  组合。优化问题的有效解决办法有多种, 本文在参数寻优过程中选取网格法、粒子群算法和遗传算法。在实验中分别进行耗时、准确率和泛化率的比较, 得出结论。实验用计算机处理器为 Intel (R) Core (TM) 2 Quad CPU Q9550 @ 2.83 GHz, 4.00 GB RAM, 软件平台为 MATLAB 7.11.0。

#### 3.1 网格算法参数寻优

网格算法属传统优化算法的一种, 其实质是穷举法, 随着排列组合情况增多, 运算量急剧增加。但在实现过程中, 能较好的找到最优参数的分布范围, 即参数组合的“好区”<sup>[9]</sup>, 摆脱局部最优解, 得到一定精度下的全局最优解。本文中  $C$  和  $g$  取值为  $[2^{-(-5)}, 2^{-(-4)}, \dots, 2^{-8}, 2^{-9}]$ , 共 255 个 ( $C, \sigma^2$ ) 组合。数据集经预处理后进行参数训练, 图 4 显示了数据集 1 以网格法进行参数寻优的过程, 表 2 汇总了 6 个数据集的训练情况。

经验风险指训练样本的错误率, 置信风险指测试样本的错误率。结构化风险为经验风险和置信风险之和。表 2 表明, 网

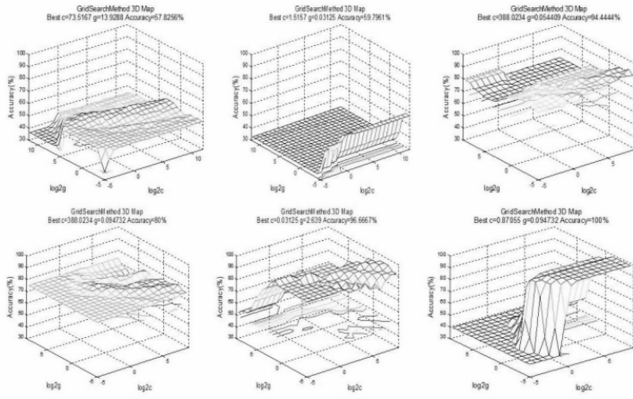


图 4 网格参数组的寻优过程

表 2 网格算法参数寻优实验数据

序号	最佳参数组 (C, g)	经验风险 / %	置信风险 / %	结构化风险 / %	训练耗时 / s
1	(73.5167, 13.9288)	42.1744	47.2156	89.39	523.734369
2	(1.5157, 0.03125)	40.204	38.9139	79.1179	23019.050751
3	(388.0234, 0.054409)	5.5556	8.4168	13.9724	7.158672
4	(388.0234, 0.094732)	20	23.2441	43.2441	10.833059
5	(0.03125, 2.639)	3.3333	7.5	10.8333	0.870338
6	(0.87055, 0.094732)	0	4.2553	4.2553	1.262056

格算法寻优的计算耗时与样本的个数、维数成正相关。数据集 2 有 25 000 个样本, 训练时间长, 且难以达到理想的效果。数据集 2 也说明置信风险不一定高于经验风险, 即在训练过程中不能过度追求低经验风险, 以结构化风险最小为原则, 合理平衡经验风险和置信风险。综合 6 组实验, 经验风险越低, 获得的置信风险越低。

### 3.2 粒子群算法参数寻优

通过对生物系统和行为动态特征的观察与模拟, 产生了较多的智能优化算法或称为启发式优化算法, 包括遗传算法 (genetic algorithm, GA)、蚁群算法 (ant colony optimization, ACO)、禁忌搜索算法 (tabu search)、模拟退火算法 (simulated annealing, SA)、粒子群算法 (particle swarm optimization, PSO) 等, 其中遗传算法和粒子群算法较适用于组合优化问题<sup>[10-11]</sup>。应用粒子群算法进行支持向量机参数寻优, 根据式 (4)、(5) 进行粒子  $i$  的第  $d$  维速度、位置更新。

$$v_{id}^k = \omega v_{id}^{k-1} + c_1 r_1 (pbest_{id} - x_{id}^{k-1}) + c_2 r_2 (gbest_d - x_{id}^{k-1}) \quad (4)$$

$$x_{id}^k = x_{id}^{k-1} + v_{id}^k \quad (5)$$

$v_{id}^k$  和  $x_{id}^k$  分别是第  $k$  次迭代, 粒子  $i$  飞行速度和位置矢量的第  $d$  维分量。 $c_1$ 、 $c_2$  是加速度常数, 用于调节学习的最大步长, 通常设置为 2。 $r_1$ 、 $r_2$  是随机数, 用于增加搜索随机性, 一般取值范围是  $[0, 1]$ 。 $\omega$  是一个非负数的惯性权重, 主要用于对解空间搜索范围的调节。 $pbest_{id}$ 、 $gbest_d$  分别是局部和全局最佳位置。Y. Shi 等已证明权重值  $\omega$  取值为  $[0.9, 1.2]$  可以得到全局最优解<sup>[12]</sup>。B. J. Rao 等人的研究<sup>[13]</sup>表明, 为避免个体搜索模式, 同时提高收敛速度, 种群最大规模一般设为 20, 终止迭代次数一般设置为 100。

实验中粒子个体位置和速度都是二维向量, 即  $(v_{i1}, v_{i2})$ 、 $(x_{i1}, x_{i2})$ ; 速度惯性权重  $\omega = 1$ ; 学习步长加速度常数

$c_1 = c_2 = 1$ ; 种群变化的极值  $c_{max} = 100$ ,  $c_{min} = 0.1$ ,  $g_{max} = 1000$ ,  $g_{min} = 0.01$ 。将预处理后的样本集进行粒子群算法参数寻优, 寻优过程中采用三折交叉验证。利用 PSO 算法优化过程如图 5 所示, 优化所得的结果如表 3 所示。

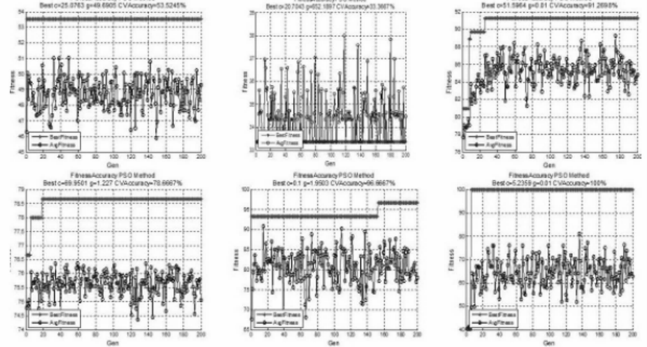


图 5 粒子群算法参数寻优过程

表 3 粒子群算法参数寻优实验数据

序号	最佳参数组 (C, g)	经验风险 / %	置信风险 / %	结构化风险 / %	训练耗时 / s
1	(25.0763, 49.6905)	46.4755	48.7126	95.1881	3115.743198
2	(20.7043, 652.1897)	66.5	66.6167	133.1167	205381.080214
3	(51.5964, 0.01)	8.7302	13.6273	22.3575	52.510017
4	(69.9501, 1.227)	21.333	21.5719	42.9049	47.044323
5	(0.1, 1.9503)	3.333	7.5	10.833	6.030470
6	(5.2359, 0.01)	0	4.9645	4.9645	8.745354

实验结果表明, 在粒子群寻优过程中, 代数较小时, 随着代数增加, 平均拟合度逐步提升, 个体趋于自我的最优状态, 整体表现出较好的适应性。代数到达一定数值后, 个体和群体的适应度更新缓慢, 种群最佳适应度不再变化, 个体最优解一致。说明粒子群算法在空间搜索时出现了停滞, 容易陷入局部最优解。6 组实验数据表明, 随着样本数量的增加, 粒子群算法收敛速度缓慢, 算法的时间复杂度为  $O(n^2)$ , 受样本影响大。

### 3.3 遗传算法参数寻优

在寻优过程中, 最优参数组具有不确定性, 当穷举范围较大时, 计算量的剧增为寻优带来更多不确定因素。对这类多项式复杂程度的非确定性问题 (non-deterministic polynomial, NP) 搜索方法中较有名的就是遗传算法 (genetic algorithm, GA)<sup>[14]</sup>。此外该算法对适应度函数无解析性质要求, 且可得到全局最优解。按照 Darwin 进化论中生物进化过程设定种群, 并对个体进行二进制或格雷码形式基因编码。根据设定的寻优目标, 得到个体适应度, 再经过基因选定父代, 通过父代的交叉或变异, 产生新个体, 最终达到寻优的目的。该启发式搜索算法采用轮盘赌法 (roulette wheel selection), 个体被选中的概率正比于适应度, 逐步淘汰适应度低者。在由  $M$  个个体组成的群体中, 设个体的适应度设为  $f_i$ , 对应的遗传给新一代的概率  $P(x_i)$  如式 (6), 累计概率  $q_i$  如式 (7)。

$$P(x_i) = \frac{f_i}{\sum_{j=1}^M f_j} \quad (6)$$

$$q_i = \sum_{j=1}^i P(x_j) \quad (7)$$

遗传算法的参数设置一般依据经验进行,不同的参数组合对其性能影响较大,其中种群个数、遗传代数、交叉率、变异率都会影响解空间搜索的深度和广度。本文根据以往经验设置种群规模  $\rho \rho = 20$ , 最大遗传代数为 200。为保证个体在有限的区域进行最优解搜索,设定参数  $c$  的变化范围为  $[0, 100]$ , 参数  $g$  的变化范围为  $[0, 1\ 000]$ , 交叉率为 0.9, 变异率为 0.1, 编码长度为 20, 将预处理后的样本集进行遗传算法参数寻优, 寻优过程中采用三折交叉验证。利用 GA 算法优化过程如图 6 所示, 训练所得结果如表 4。

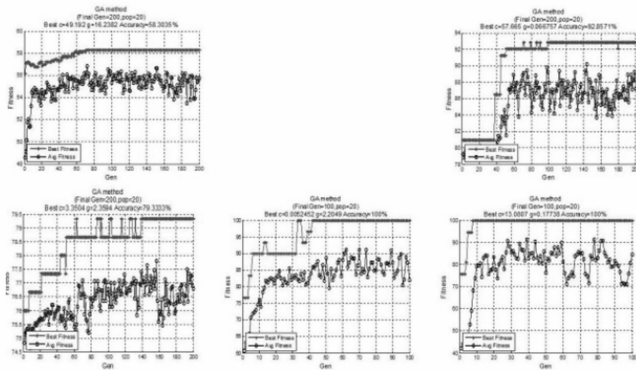


图 6 遗传算法参数寻优过程

表 4 遗传算法参数寻优实验数据

序号	最佳参数组( $C, g$ )	经验风险/%	置信风险/%	结构化风险/%	训练耗时/s
1	(11.5126, 25.4269)	41.6965	47.3353	89.0318	1964.545545
2	(*, *)	*	*	*	*
3	(0.2944, 2.1315)	7.1429	8.6172	15.7601	37.313213
4	(2.5928, 729.3432)	21.3333	24.4147	45.748	57.759948
5	(0.0052452, 2.2049)	0	7.5	7.5	2.379051
6	(6.288, 0.15831)	0	7.0922	7.0922	3.432706

由表 4 可看出在遗传算法中,随着代数的增加,遗传算法过程能实现优胜劣汰,收敛到全局最优解。实验中,数据集 2 在经历 97 小时的训练后仍未得到理想模型。数据集 3 的经验风险率低于置信风险率。大部分数据集在迭代 20 次后达到最优,但数据集 5 经过了 62 次迭代,表明 GA 的寻优能力受数据分布和规模影响。数据集 2 的寻优过程异常缓慢,其原因是算法中的随机变异和交叉率影响了个体的品质,在有限的进化代数中失去了寻优能力。以数据集 6 为例,遗传代数为 3 时达到了最优解,但因交叉和变异过程的进行,虽群体平均适应度逐步提升,但函数失去了收敛能力,只能依赖代数增加,逐渐收敛。故应适度的加入保护精英主义思想,保证当前最优解不参加变异,避免全局最优解被破坏。

### 3.4 实验分析

#### 3.4.1 优化算法分析与应用策略

对比三种算法的最优解和寻优过程可知传统的迭代算法适用于小数据计算。数据量增大时,算法从设定的初始点开始,难以实现并行、网络计算,且因针对二维参数组寻优,算法复杂度为  $O(n^2)$ , 穷举过程时间频度高,计算效率低。网格算法的优点体现在对“坏数据”——数据集 2 的处理上,样本分类正确率比智能算法高 30%。网格算法以时间和空间复杂度

为代价,可以在规定的解空间内搜索到全局最优解,但精度严重依赖所设定的步长,解空间的设置依赖初始值的选取,故需在时间、精度以及空间做出权衡。

在 SVM 参数寻优过程中,PSO 相较于 GA,收敛时间较短但置信风险高,即快速收敛是以牺牲最优解的精度为代价。PSO 实现较为简单,GA 算法中既要保证种群的多样性,维持一定的交叉率和变异率,也要对适应度高的个体进行精英保护。本文所用的 GA 算法,对并行机制和网络反馈信息利用不充分,导致数据集 5 的收敛速度较慢。终止进化条件应灵活设置,避免数据集 6 在经验风险为 0 后,继续进化。

实验中,PSO 和 GA 的算法复杂度分别为  $O(n^2)$ 、 $O(\log 2n)$ 。与传统的网格法比较,这两种智能算法表现出了极高的学习性能,有良好的全局搜索能力及并行性,可进行快速分布式计算。给定的参数组范围  $c = [0, 100]$ ,  $g = [0, 1\ 000]$ , 计算次数远多于网格算法的 225 次,但耗时与网格算法相当。智能算法的缺陷是过分依赖初始化种群的规模、迭代次数及变异率等参数,且无严格的参数设置原则,严重依赖于经验。由于训练过程中的变异和选择过程具有随机性,实验结果难以重现。故在寻优过程中,可以使用智能算法进行最优解范围初步确定,以此确定解空间的初始值,再使用网格法进行高精度运算。

#### 3.4.2 参数对 SVM 性能的影响

综合训练结果,训练样本对 SVM 的性能有较大的影响。数据集 5、6 分布均衡,个数适中,不同的优化算法都能得到较稳定的最优参数组合,得到较高的泛化机器。数据集 3、4 的数据分布不均匀,效果稍差。三种方法对数据集 2 的处理,都没有得到较理想的效果,且寻优时间长,过程缓慢,分析有以下三种原因:1) 该数据的规模不适合选择支持向量机,应尝试神经网络、弱分类器级联等方法;2) 高斯核函数对该数据不适用,可选择多项式核函数或二层神经收集核函数;3) 本文直接进行了数学机械式归一化与降维,忽略了物理意义,应尝试针对特征向量的物理意义,做不同的预处理,以期达到理想的训练模型。

目标函数式(2)中训练样本误差项的惩罚系数  $C$ ,控制着使分类间隔最大且错误率最小的折中,在确定的特征空间中调节学习机器的置信范围和经验风险的比例;而高斯函数的方差系数  $\sigma^2$ ,主要影响样本数据在高维特征空间中分布的复杂程度。通过网格算法可以看出,随着  $\sigma^2$  的增大,分类效果会变差。但为追求最高适应度,使  $\sigma^2$  不能减小超过临界点(即最优解),否则带来严重的过拟合问题,使得经验风险率与置信风险率差距悬殊。实验数据也说明 SVM 有较高的独立性,针对不同的数据集,最优参数不同,不具有泛化性。

## 4 结语

本文利用传统和智能优化算法对 SVM 应用过程中的核函数参数组  $(C, \sigma^2)$  进行优化。在对原始数据进行归一化和降维后,分别采取了网格法、粒子群法和遗传算法实现参数组寻优过程,并分析了  $(C, \sigma^2)$  对 SVM 泛化性能的影响。在应用 SVM 时,为加快收敛速度、提高最优解精度,应采取优化算法混合的策略,利用智能算法进行初步最优解定位,然后采用网格算法进行高精度最优解的确定,可以得到性能较好的支持向量机模型。

## 参考文献:

- [1] Vapnik V. The nature of statistical learning theory [M]. Springer Science & Business Media, 2013.
- [2] Vapnik V N, Vapnik V. Statistical learning theory [M]. New York: Wiley, 1998.
- [3] Chapelle O, Vapnik V, Bousquet O, et al. Choosing multiple parameters for support vector machines [J]. Machine learning, 2002, 46 (1-3): 131-159.
- [4] University of California—Irvine. UCI Machine Learning Repository. <http://archive.ics.uci.edu/ml/datasets.html>. 2015-04-11.
- [5] Eitrich T, Lang B. Parallel tuning of support vector machine learning parameters for large and unbalanced data sets [M]. Springer Berlin Heidelberg, 2005: 253-264.
- [6] He H, Garcia E A. Learning from imbalanced data [J]. Knowledge and Data Engineering, IEEE Transactions on, 2009, 21 (9): 1263-1284.
- [7] Zhang Z, Guo H. Research on Fault Diagnosis of Diesel Engine Based on PSO-SVM [A]. Proceedings of the 6th International Asia Conference on Industrial Engineering and Management Innovation [C]. Atlantis Press, 2016: 509-517.
- [8] Smits G F, Jordaan E M. Improved SVM regression using mixtures of kernels [A]. 2002. IJCNN02. Proceedings of the 2002 International Joint Conference on IEEE [C]. 2002, 3: 2785-2790.

(上接第 254 页)

索速度是 SQLite72.032 倍, 是 SQL Server 2008 的 5.495 倍, 是 MySQL58.555 倍, 是内存 B-树 1.231 倍。提出的方法和内存 B-树方法写入和读取速度远远快于其他方法, 这是由于索引存储在内存中可以获得高速的存储速度, 然而同样把索引维持在内存中, 本文提出的页面索引方法的速度要快于 B-树的速度。原因是以 B-树作为索引时, 其结点在程序中实现为数组或者指针列表, 当插入一个关键字时程序需要移动子结点在数组或者指针列表中移动, 这将会耗费时间; 当 B-树分裂一个结点时需要有效的平衡该页面所在结点的父结点和子结点。因而在此时整棵 B-树会被封锁一定的时间, 所以在这个时候不能同时进行读写操作, 而在本文所设计的索引中每个页面被页面数组结构相互隔离开, 所以当前页面进行分裂时给当前页面上锁很容易, 且只有当前的页面被锁住而其他的页面不会被锁住因此其他的页面仍然可以进行读写操作, 页面分裂的操作简单、限制少, 因此减少了时间消耗, 页面中包含 10 000 个索引项, 因此更新的频率很低, 因此对页面进行分裂耗费的时间很少。

#### 4 结束语

在 C/S 模式下, 分析客户端可视化数据缓存的存储结构和算法, 深入分析缓存的存储结构和算法对读取可视化数据所需要时间以及缓存的稳定性的影响。通过在现有的主流数据库当中进行读写速度的比较以及稳定性的比较, 选择最合适的数据库, 减少存储的数据量, 提高缓存的读写速度。详细介绍了一种新型的页面数据索引的设计, 该页面索引应用在高实时性的读取某个历史时间戳的数据, 并将当前采集的数据和读取的历史数据进行对比, 给出了查找页面的具体算法并结合具体例子对页面查找算法进行了详细分析, 介绍了存储历史数据的文件结构, 并给出了结构设计的依据, 详细介绍了数据写入和读取的功能实现, 并对各存储方法进行了比较、分析得出了结

- [9] Keerthi S S, Lin C J. Asymptotic behaviors of support vector machines with Gaussian kernel [J]. Neural computation, 2003, 15 (7): 1667-1689.
- [10] Duan S M, Mao J L, Li J L, et al. Design Implementation and Application of Swarm Intelligence Algorithm Optimization Function Simulation Platform [A]. Software Engineering and Information Technology: Proceedings of the 2015 International Conference on Software Engineering and Information Technology (SEIT2015) [C]. 2016: 196-203.
- [11] Yu S, Zhang J, Zheng S, et al. Provincial carbon intensity abatement potential estimation in China: a PSO-GA-optimized multi-factor environmental learning curve method [J]. Energy Policy, 2015, 77: 46-55.
- [12] Shi Y, Eberhart R C. Empirical study of particle swarm optimization [A]. CEC 99. Proceedings of the 1999 Congress on Evolutionary Computation [C]. IEEE, 1999, 3.
- [13] Rao B J, Babu M S P. Ongole Breed Cattle Health Expert Advisory System Using Parallel Particle Swarm Optimization Algorithm [J]. 2014.
- [14] Yang C, Xu X, Han J, et al. Energy efficiency-based device-to-device uplink resource allocation with multiple resource reusing [J]. Electronics Letters, 2015, 51 (3): 293-294.

论。可以较好地解决这一问题。通过对应用实例进行评估, 验证了该技术的有效性和易用性。

## 参考文献:

- [1] 张 棋. 信息可视化在数字图书馆中的应用研究 [D]. 湖北: 华中师范大学, 2009.
- [2] 傅雅玉, 源艳芬, 梁慎青, 等. 提高气象数据查询速度的存储分析 [J]. 电脑知识与技术, 2014, (6): 1170-1172.
- [3] 虞海燕. 基于 Cache 数据库的电子病历系统性能评价 [D]. 上海: 浙江大学, 2013.
- [4] 傅荣会. 三种关系型数据库管理系统的比较研究——Visual FoxPro, Access 和 SQL Server [J]. 重庆三峡学院学报, 2011, (3): 58-59.
- [5] 宋亚林, 任冬英. 文档型数据库与关系型数据库中数据集成的研究 [J]. 福建电脑, 2011, (2): 157-158.
- [6] 闫常友. 广域测量系统若干技术及电压稳定应用研究 [D]. 北京: 华北电力大学, 2005.
- [7] 邱胜海, 高成冲, 王云霞, 等. 大数据时代非关系型数据库教学与实验改革探索 [J]. 电脑知识与技术, 2013, (31): 7046-7048.
- [8] 李东林. 基于 Linux 的嵌入式实时数据库访问算法研究与实现 [D]. 长春: 长春理工大学, 2008.
- [9] 张 华, 顾红飞, 刘 涛. 基于 B+树的文本信息检索技术 [J]. 皖西学院学报, 2010 (2): 31-35.
- [10] 朱秋香. 流分类技术研究及其原型系统的实现 [D]. 福建: 东南大学, 2004.
- [11] 田晓辉. B-tree 检索技术及其应用 [J]. 福建电脑, 2013 (10): 22-23.
- [12] 叶晓舟, 王劲林. 混合 Chord 网络中的有效关键字检索 [J]. 计算机工程, 2008 (11): 246-248.
- [13] 汪 超, 何丕廉, 李志明. 基于 Hash 表的数据库索引结构设计及实现 [J]. 微处理机, 2007 (4): 53-56.
- [14] 雷红利. 查询执行算法的设计与优化 [D]. 长沙: 华中科技大学, 2004.