

面向大规模数据的隐私保护学习机

刘忠宝^{1,2}, 王士同¹

(1. 中北大学电子与计算机科学技术学院 太原 030031; 2. 江南大学数字媒体学院 江苏 无锡 214122)

【摘要】随着海量数据不断涌入, SVM隐私泄露问题日益严重。在分析已有隐私保护支持向量机基础上, 提出一种面向大规模数据的隐私保护学习机(PPLM)。该方法首先通过核心向量机对大规模样本进行采样, 然后在核心集上选取两个样本点并将两点连线的法平面作为最优分类面。通过对标准数据集和人工数据集的实验表明, PPLM可有效地解决大规模样本分类问题, 且分类效果良好。

关键词 大规模数据集; 模式分类; 隐私保护; 支持向量机

中图分类号 TP391

文献标志码 A

doi:10.3969/j.issn.1001-0548.2013.02.018

Privacy Preserving Learning Machine for Large Scale Datasets

LIU Zhong-bao^{1,2} and WANG Shi-tong¹

(1. School of Electronics and Computer Science Technology, North University of China Taiyuan 030051;

2. School of Digital Media, Jiangnan University Wuxi Jiangsu 214122)

Abstract Support vector machine (SVM) is widely used in pattern classification. In order to solve the privacy preserving problem in SVM, a privacy preserving learning machine for large scale datasets (PPLM) is proposed in this paper. First, core vector machine (CVM) is introduced for sampling the large scale datasets; then two points from different classes are chosen in the core set and the hyperplane orthogonal to the line connecting these two points is treated as the optimal separating hyperplane. Experimental results obtained from synthetic and standard datasets verify that the PPLM is effective and competitive.

Key words large scale datasets; pattern classification; privacy preserving; support vector machine

信息时代带来了数据爆炸式增长, 也催生了从海量数据中提取有用信息的研究领域——数据挖掘。信息社会高度的数据收集和共享机制为各种组织机构间的合作研究提供了便利条件, 同时也增加了隐私信息泄露的风险。随着个人和企业信息泄露问题日益严重, 隐私保护问题引起人们的极大关注。

分类是数据挖掘的主要目标和任务之一, 其目的是构造一个分类器用于预测未来数据的类属。支持向量机(SVM)^[1-3]是一种运用广泛的分类方法, 它建立在统计学习VC维理论和结构风险最小原理基础上, 成功地将最大分类间隔思想和基于核的方法结合在一起。随着支持向量机的发展及隐私保护问题的扩展, 隐私保护支持向量机研究得到学术界的重视, 越来越多的人致力于隐私保护支持向量机的研究并取得了一些成果^[4-7]。

隐私保护支持向量机主要从数据水平分布和垂直分布两方面进行研究。在水平分布数据隐私保护

方面, 文献[4]通过计算集合交的势求得布尔向量的内积和支持向量机的核函数, 最终得到所有数据建立的隐私保护支持向量机; 文献[5]将简约支持向量机(reduced support vector machine, RSVM)^[8-9]引入隐私保护支持向量机, 通过构造所有数据的简约核矩阵, 进而求得隐私保护支持向量机。在垂直分布数据隐私保护方面, 文献[6]将求解垂直分布数据整体核函数问题分解为求解各部分数据核函数问题, 在不泄露信息前提下得到所有数据建立的支持向量机; 文献[7]将简约支持向量机引入隐私保护支持向量机, 通过对所有站点的核矩阵相加求和得到隐私保护支持向量机。

上述方法在中小规模数据集均可达到较好分类效果, 但面对大规模数据便无能为力。基于上述分析, 本文提出一种面向大规模数据的隐私保护学习机(PPLM)。该方法具有以下优势: 1) 有效解决大规模数据分类问题; 2) 保证分类过程隐私安全。

收稿日期: 2011-06-16; 修回日期: 2012-11-26

基金项目: 国家863项目(2007AA1Z158, 2006AA10Z313); 国家自然科学基金(60773206, 60704047)。

作者简介: 刘忠宝(1981-), 男, 博士生, 主要从事模式识别与机器学习方面的研究。

1 支持向量机及隐私泄露问题

1.1 支持向量机

对于包含 N 个模式两类划分问题, 设给定训练集合 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, 其 $x_i \in R^d$ ($1 \leq i \leq N_1 + N_2 = N$) 为输入数据, $y_i \in \{1, -1\}$ 为类别标签。 $1 \leq i \leq N_1$ 时, $y_i = 1$; $N_1 + 1 \leq i \leq N$ 时, $y_i = -1$ 。 第一类含有 N_1 个模式 $\{x_i, y_i\}_{i=1}^{N_1}$; 第二类含有 N_2 个模式 $\{x_j, y_j\}_{i=N_1+1}^N$ 。 SVM的目的是寻找一个最优超平面将两类模式正确分开。 设超平面方程为 $\mathbf{w}^T \mathbf{x} + b = 0$, 分类间隔为 $2/\|\mathbf{w}\|$, 则寻找最优分类面的过程可转化为如下优化问题:

$$\begin{aligned} \min_{\mathbf{w}, b, \xi_i} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \\ \text{s.t.} \quad & y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i \\ & \xi_i \geq 0; i = 1, 2, \dots, N \end{aligned} \quad (1)$$

式中, C 为惩罚因子, 它控制对错分样本的惩罚程度: $C=0$ 时表示线性可分, $C>0$ 时表示线性不可分。 对于线性不可分或事先未知是否线性可分的情况, 通过引入松弛因子 ξ_i 允许错分样本的存在。

利用对偶理论将上述优化问题转化为二次规划问题, 即:

$$\begin{aligned} \max \quad & \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j \alpha_i \alpha_j \mathbf{x}_i^T \mathbf{x}_j \\ \text{s.t.} \quad & \sum_{i=1}^N y_i \alpha_i = 0 \\ & 0 \leq \alpha_i \leq C; i = 1, 2, \dots, N \end{aligned} \quad (2)$$

式中, α_i 为每个样本对应的Lagrange乘子。 上述优化问题的最优解为 $\boldsymbol{\alpha}^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*)^T$, 其中不为零的 α_i^* 对应的样本为支持向量。 可计算最优分类面的权系数向量:

$$\mathbf{w}^* = \sum_{i=1}^N y_i \alpha_i^* \mathbf{x}_i \quad (3)$$

选择最优解 $\boldsymbol{\alpha}^*$ 的一个正分量 $0 < \alpha_j^* < C$, 计算分类阈值:

$$b^* = y_j - \sum_{i=1}^N y_i \alpha_i^* \mathbf{x}_i^T \mathbf{x}_j \quad (4)$$

可求得决策函数为:

$$f(\mathbf{x}) = \text{sgn}(\boldsymbol{\alpha}^* \mathbf{y}_i \mathbf{x}_i^T \mathbf{x} + b^*) \quad (5)$$

对于非线性问题, SVM通过一个非线性映射 ϕ 将输入空间变换到高维空间求解最优分类面, 则优化问题可转化为:

$$\min_{\mathbf{w}, b, \xi_i} \quad \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i$$

$$\begin{aligned} \text{s.t.} \quad & y_i (\mathbf{w}^T \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \\ & \xi_i \geq 0; i = 1, 2, \dots, N \end{aligned} \quad (6)$$

其中, $\phi(\mathbf{x})$ 表示从原始样本空间到高维特征空间的映射。

利用对偶理论将上述优化问题转化为二次规划问题, 即:

$$\begin{aligned} \max \quad & \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) \\ \text{s.t.} \quad & \sum_{i=1}^N y_i \alpha_i = 0 \\ & 0 \leq \alpha_i \leq C; i = 1, 2, \dots, N \end{aligned} \quad (7)$$

其中, 核函数 $k(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{x})^T \phi(\mathbf{y})$ 。

可求得决策函数为:

$$f(\mathbf{x}) = \text{sgn}(\sum_{i=1}^N \alpha_i^* y_i k(\mathbf{x}_i, \mathbf{x}) + b^*) \quad (8)$$

选择最优解 $\boldsymbol{\alpha}^*$ 的一个正分量 $0 < \alpha_j^* < C$, 计算分类阈值:

$$b^* = y_j - \sum_{i=1}^N y_i \alpha_i^* k(\mathbf{x}_i, \mathbf{x}_j) \quad (9)$$

1.2 SVM隐私泄露问题

线性SVM本身实现了隐私保护, 原因是线性SVM决策函数 $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b$ 中权重 \mathbf{w} 是所有支持向量的线性组合, 每个支持向量的敏感内容由于权重 \mathbf{w} 的存在而不被暴露。 但线性SVM使用范围有限, 核方法的引入导致线性SVM隐私保护作用消失: 式(11)表明判断样本 \mathbf{x} 的类属必须事先得知所有训练样本信息, 这对隐私保护带来极大挑战。

2 面向大规模数据的隐私保护学习机

2.1 分类超平面

为了减小线性SVM时空复杂度, 文献[10]提出如下分类超平面(separating hyperplane, SH):

$$d(\mathbf{x}) = \text{sgn}(y_i \mathbf{x}_i^T \mathbf{x} - y_j \mathbf{x}_j^T \mathbf{x} - b) \quad (10)$$

该超平面工作原理如下: 对于两类问题, 两类中各选一个样本点, 将两点连线的法平面作为候选分类面。 由于存在无穷多个这样的分类面, 因此规定该分类面至少经过一个样本点。 选取误分率最低的分类面作为最优分类面, 设分类面方程为 $\mathbf{a}^T \mathbf{x} = b$, 其中 $\mathbf{a} = y_i \mathbf{x}_i - y_j \mathbf{x}_j$ ($y_i y_j = -1$), $b = \mathbf{a}^T \mathbf{x}_k$ 。

上述思想如图1所示, 其中 o 和 $+$ 分别代表两类样本。

文献[10]主要讨论了线性情况下上述分类器的性能, 对非线性情况未做详尽的理论和实验分析。 更重要的是, 文献[10]没有意识到上述分类器的核化形式

能有效解决核SVM隐私泄露问题。因此,本文在文献[10]基础上对上述分类器的核化形式在隐私保护方面的作用做进一步研究。式(13)的核化形式如下:

$$d(\mathbf{x}) = \text{sgn}(y_i k(\mathbf{x}_i, \mathbf{x}) - y_j k(\mathbf{x}_j, \mathbf{x}) - b) \quad (11)$$

式中,分类阈值 $b = y_i k(\mathbf{x}_i, \mathbf{x}_k) - y_j k(\mathbf{x}_j, \mathbf{x}_k)$ ($1 \leq i \leq N_1, N_1 + 1 \leq j \leq N, 1 \leq k \leq N$) 且 $y_i y_j = -1$ 。

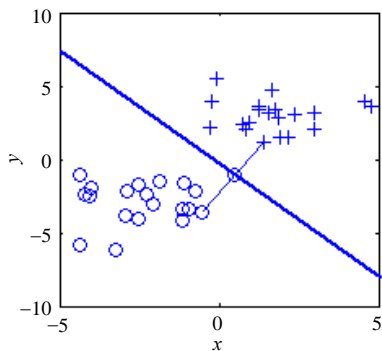


图1 分类超平面工作原理示意图

2.2 CVM

核心向量机(core vector machine, CVM)^[11]把QP问题的求解转化为最小包含球问题,并使用一个逼近率为 $(1+\varepsilon)$ 的近似算法得到核心集(core set)。该核心集规模远小于原始样本规模,通过对该核心集训练可得到理想的分类效果。核心集规模仅与参数 ε 有关,与样本数及样本维数无关,该结论从理论上保证CVM适用于大规模样本分类问题。

2.3 PPLM

实际应用往往面临大规模数据,SH由于时空复杂度过大而无法求解。PPLM对SH进行了改进,不仅有效解决大规模数据分类问题,而且保证分类过程隐私安全。PPLM算法如下:

参数说明: \mathbf{z} 为训练样本点, $B(\mathbf{c}, R)$ 是球心为 \mathbf{c} ,半径为 R 的最小包含球, \mathbf{S}_t 为核心集, t 为迭代次数, ε 为终止参数。

- 1) 初始化 $\mathbf{c}_t, R_t, \mathbf{S}_t, \varepsilon, t=0$;
- 2) 当所有样本点 \mathbf{z} 均在最小包含球 $B(\mathbf{c}_t, (1+\varepsilon)R)$ 内,则转到步骤6),否则转到步骤3);
- 3) 如果 \mathbf{z} 距离球心 \mathbf{c}_t 最远,则 $\mathbf{S}_{t+1} = \mathbf{S}_t \cup \{\varphi(\mathbf{z})\}$;
- 4) 寻找最新最小包含球 $B(\mathbf{S}_{t+1})$,并设置 $\mathbf{c}_t = \mathbf{c}_{B(\mathbf{S}_{t+1})}, R_t = R_{B(\mathbf{S}_{t+1})}$;
- 5) $t=t+1$,转到步骤2);
- 6) SH对核心集 \mathbf{S}_t 进行训练并由式(11)得到分类结果。

2.3 理论分析

由文献[12]可知,SVM满足如下VC维复杂度:

$$h \leq \min([R^2 / \varepsilon_{\min}^2], N) + 1 \quad (12)$$

式中, N 为样本数; R 为包含所有样本的最小球半径; ε_{\min} 为类间最小间隔。

用CVM对大规模样本采样后,式(12)中的 R 变为 $(1+\varepsilon)R$; 采样后原样本规模减小,则 ε_{\min} 变大。当 ε 取值较小(经验性结论: ε 的取值范围为 $[10^{-5}, 10^{-6}]$) 时, $(1+\varepsilon)R^2 / \varepsilon_{\min}^2$ 变化不大,则VC维 $h \leq \min([(1+\varepsilon)R^2 / \varepsilon_{\min}^2], N) + 1$ 近似成立。因此近似地先用CVM采样,再用SH分类(即PPLM)是基本可行的。

2.4 重要性质

性质1: PPLM的VC维小于等于9。

借鉴文献[10]有关思想,可得PPLM的VC维小于等于9。具体证明过程如下:

1) 由VC维定义可知: 对于一个函数集,如果存在 h 个样本能被函数集中的函数按所有可能的 2^h 种形式分成两类,则称函数集能把 h 个样本打散。函数集的VC维就是它能打散的最大样本数 h 。

2) 由PPLM算法可知: 两类样本共产生有序样本对的数量为 $h(h-1)$; 由于分类面至少经过一个样本点,则共有 $h^2(h-1)$ 种形式将样本分开。

当 $h > 9$ 时,有 $2^h > h^2(h-1)$, 即当可打散的最大样本数大于9时, PPLM不能打散它们,因此PPLM的VC维小于等于9。

3 实验分析

验证PPLM的有效性。实验环境为3 GHz Pentium4 CPU、2G RAM、Windows XP及Matlab7.0。实验选取的核函数为高斯核函数:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\delta^2) \quad (13)$$

3.1 实验参数设置

目前参数选择的主流方法有单一验证估计、留一法、交叉验证法以及基于样本相似度的方法等,同时随着遗传算法、蚁群算法等智能优化方法的成功应用,陆续有学者采用这些方法来优化参数选择^[13-16]。目前交叉验证法应用广泛、易于实现且性能优良,因篇幅有限,本文仅考虑交叉验证法,且采用5倍交叉验证法。

参数通过网格搜索策略选择。高斯核函数中 δ^2 在网格 $\{m/8, m/4, m/2, m, 2m, 4m, 8m\}$ 中搜索选取,其中 m 为训练样本平均范数; C-SVC中的惩罚因子 C 在网格 $\{0.01, 0.05, 0.1, 0.5, 1, 5, 10\}$ 中搜索选取。

3.2 实验数据集

实验数据集如表1所示。数据集 Chess、Contraceptive、Magic 下载于 www.ics.uci.edu/~mlern/MLRepository.html; 数据集 Forest 下载于 www.cse.ust.hk/~ivor/cvm.html; Checkboard 为人工数据集, 如图2所示。

表1 中大规模数据集

Dataset	#total	#class1	#class2	dim
Chess	3196	2294	904	37
Contraceptive	1140	629	511	10
Magic	190 20	123 32	6688	11
Checkboard	450 000	250 000	200 000	2
Forest	581 02	283 29	297 73	54

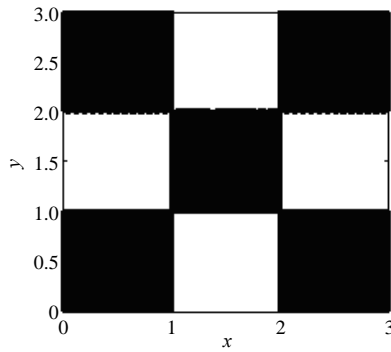
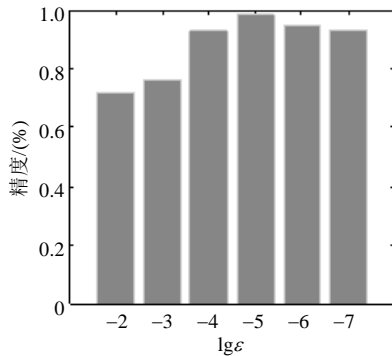
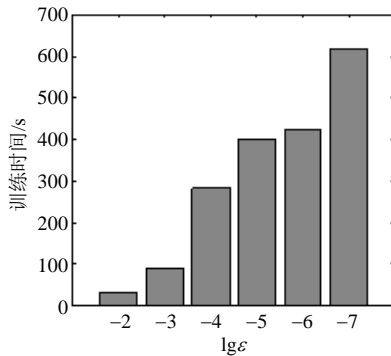


图2 Checkboard数据集



a. ϵ 与分类精度的关系



b. ϵ 与样本训练时间的关系

图3 参数 ϵ 对PPLM的影响

3.3 参数 ϵ 对PPLM的影响

由2.3节PPLM算法可知: 终止参数 ϵ 越小, 则算法迭代次数越多, 样本训练时间越长。因此选择恰当的终止参数 ϵ 至关重要。

实验选取60%的Chess数据集作为训练样本, 剩下的作为测试样本。将训练样本等分为5份, 其中4份作为训练集, 1份作为验证集。经5倍交叉验证得到高斯核函数的 $\delta^2 = m/4$ 。

终止参数 ϵ 在网格 $\{10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}\}$ 中搜索选取。实验结果如图3所示。

由图3可以看出: 终止参数 ϵ 不仅影响到算法的分类精度, 而且影响到样本的训练时间。不失一般性, 选取 $\epsilon = 10^{-6}$ 。

3.4 PPLM分类性能

分别取Contraceptive、Magic、Forest、Checkboard数据集的20%、40%、60%、80%作为训练样本, 从剩下样本中任取500个作为测试样本。实验参数通过对CVM得到的核心集作5倍交叉验证获得。实验参数和实验结果分别记录于表2和表3。表3中“Time”表示训练时间(Train Time)。

由表3可以看出: 随着训练样本规模扩大, PPLM分类精度呈上升趋势, 但训练时间并未迅速增加。由此可见, PPLM能在较短时间内解决大规模样本分类问题。

表2 PPLM实验参数

Datasets	Optimal Parameter (δ^2)
Contraceptive	$8m$
Magic	$8m$
Checkboard	$8m$
Forest	$4m$

表3 PPLM分类结果

Train Size	Contraceptive		Magic		Checkboard		Forest	
	accuracy / (%)	Time/s	accuracy / (%)	Time/s	accuracy / (%)	Time/s	accuracy / (%)	Time/s
20	72.5	15.4	74.2	62.4	79.6	132.2	92.3	332.8
40	82.7	25.1	77.6	134.7	92.6	316.9	91.9	286.4
60	84.1	39.2	98.8	209.6	93.9	564.2	96.8	323.4
80	100	42.0	96.4	357.8	94.2	745.8	92.9	479.2

3.5 对比实验

通过与C-SVC、SH比较, 验证PPLM在中大规模数据集上的有效性。实验选取的数据集为Forest, 训练样本规模依次取 $1 \times 10^2, 3 \times 10^2, 5 \times 10^2, 1 \times 10^3, 3 \times 10^3, 5 \times 10^3, 1 \times 10^4, 3 \times 10^4, 5 \times 10^4$, 剩余样本中任取500个作为测试样本。实验结果记录于表4, 其中“-”表示在有限时空范围内无法求解。

由表4可看出, 当样本规模在500以内时, 3种算法均可在有限时间内获得分类结果; 样本规模为100时, 3种算法分类精度基本相当且训练时间较短; 样本规模为300和500时, SH分类精度优于C-SVC和PPLM, 但SH的训练时间最长。当样本规模达到1000时, SH由于空间代价过高而无法求解, C-SVC和

PPLM均可在有限时间内获得分类结果,但PPLM的训练时间远小于C-SVC。当样本规模上升至3000及以上时,C-SVC和SH无法在有限时空范围内获得分类结果,但PPLM仍可较快地获得分类结果,且分类精度良好。由此可见,PPLM在解决大规模样本分类问题上的有效性是传统方法所不具备的。

表4 C-SVC、SH、PPLM分类结果

Train Size	C-SVC		SH		PPLM	
	accuracy / (%)	Time/s	accuracy / (%)	Time/s	accuracy / (%)	Time/s
1×10^2	66.00	0.75	65.62	2.48	66.84	2.44
3×10^2	70.67	12.54	79.30	26.81	68.56	12.40
5×10^2	72.67	77.72	82.84	627.42	80.69	25.52
1×10^3	77.21	1 222.10	—	—	86.57	55.72
3×10^3	—	—	—	—	85.65	68.34
5×10^3	—	—	—	—	91.31	97.15
1×10^4	—	—	—	—	92.80	118.14
3×10^4	—	—	—	—	93.36	144.72
5×10^4	—	—	—	—	96.00	219.16

4 结论

针对核SVM在实际应用中存在的信息泄露问题,提出一种面向大规模数据的隐私保护学习机PPLM。该方法首先通过核心向量机对大规模样本进行采样,然后在核心集上选取两个样本点并将两点连线的法平面作为最优分类面。该方法具有以下优势:1)有效解决大规模数据分类问题;2)保证分类过程隐私安全。标准数据集和人工数据集上的实验验证了PPLM的有效性。

参考文献

- [1] VAPNIK V. The nature of statistical learning theory[M]. New York: Springer Verlag, 1995.
- [2] 邓乃扬, 田英杰. 支持向量机——理论、算法与拓展[M]. 北京: 科学出版社, 2009.
DENG Nai-yang, TIAN Ying-jie. Support vector machine—theory, algorithm and extension[M]. Beijing: Chinese Science Press, 2009.
- [3] PAL M, FOODY G M. Feature selection for classification of hyper spectral data by SVM[J]. IEEE Transactions on Geoscience and Remote Sensing, 2010, 48(5): 2297-2307.
- [4] YU H, JIANG Xiao-qian, VAIDYA J. Privacy-preserving SVM using nonlinear kernels on horizontally partitioned data[C]//Proceedings of the 2006 ACM Symposium on Applied Computing. New York, USA: ACM, 2006: 603-610.
- [5] MANGASARIAN O L, WILD E W. Privacy-preserving classification of horizontally partitioned data via random kernels[C]//Proceedings of the 2008 International Conference on Data Mining. Las Vegas: [s.n.], 2008.
- [6] YU H, VAIDYA J, JIANG Xiao-qian. Privacy-preserving SVM classification on vertically partitioned data[C]//Proceedings of PAKDD'06. Berlin Heidelberg: Springer Verlag, 2006: 647-656.
- [7] MANGASARIAN O L, WILD E W, FUNG G M. Privacy-preserving classification of vertically partitioned data via random kernels[J]. ACM Transactions on Knowledge Discovery from Data, 2008, 3(2):1-16.
- [8] LEE Y J, MANGASARIAN O L. Reduced support vector machines[C]//Proceedings of the First SIAM International Conference on Data Mining. Chicago, Philadelphia, UAS: ACM, 2001: 57-64.
- [9] LIN Kuan-ming, LIN C J. A study on reduced support vector machines[J]. IEEE Transactions on Neural Network, 2003, 45(2): 199-204.
- [10] MARCIN O. New separating hyperplane method with application to the optimization of direct marketing campaigns[J]. Pattern Recognition Letters, 2011(32): 540-545.
- [11] TSANG I W, KWOK J T, CHEUNG P M. Core vector machines: Fast SVM training on very large data sets[J]. Journal of Machine Learning Research, 2005(6): 363-392.
- [12] SEEP H, KLAUS O. Support vector machines for dyadic[J]. Neural Computation, 2006, 18(6), 1472-1510.
- [13] 奉国和. SVM分类核函数及参数选择比较[J]. 计算机工程与应用, 2011, 47(3): 123-128.
FENG Guo-he. Optimizing for support vector machines classification[J]. Computer Engineering and Applications, 2011, 47(3): 123-128.
- [14] 吴景龙, 杨淑霞, 刘承水. 基于遗传算法优化参数的支持向量机短期负荷预测方法[J]. 中南大学学报(自然科学版), 2009, 40(1): 180-184.
WU Jing-long, YANG Shu-xia, LIU Cheng-shui. Parameter selection for support vector machines based on genetic algorithms to short-term power load forecasting[J]. Journal of Central South University(Science and Technology), 2009, 40(1): 180-184.
- [15] 杨旭, 纪玉波, 田雪. 基于遗传算法的SVM参数选取[J]. 辽宁石油化工大学学报, 2004, 24(1): 54-58.
YANG Xu, JI Yu-bo, TIAN Xue. Parameters selection of SVM based on genetic algorithm[J]. Journal of Liaoning University of Petroleum & Chemical Technology, 2004, 24(1): 54-58.
- [16] 庄严, 白振林, 许云峰. 基于蚁群算法的支持向量机参数选择方法研究[J]. 计算机仿真, 2011, 28(5): 216-219.
ZHUANG Yan, BAI Zhen-lin, XU Yun-feng. Research on parameters of support vector machine based on ant colony algorithm[J]. Computer Simulation, 2011, 28(5): 216-219.

编辑 张俊