

一种新的 SVM 对等增量学习算法

王晓丹,郑春颖,吴崇明,张宏达

(空军工程大学 导弹学院,陕西 三原 713800)

(afeu_w@yahoo.com.cn)

摘 要:在分析支持向量机(SVM)寻优问题的 KKT 条件和样本分布之间关系的基础上,分析了新增样本的加入对 SV 集的影响,定义了广义 KKT 条件。基于原训练样本集和新增训练样本集在增量训练中地位等同,提出了一种新的 SVM 增量学习算法。算法在及时淘汰对后继分类影响不大的样本的同时保留了含有重要分类信息的样本。对标准数据集的实验结果表明,算法获得了较好的性能。

关键词:支持向量机;增量学习

中图分类号: TP18 **文献标识码:** A

New algorithm for SVM-Based incremental learning

WANG Xiao-dan, ZHENG Chun-ying, WU Chong-ming, ZHANG Hong-da

(Missile Institute, Air Force Engineering University, Sanyuan Shaanxi 713800, China)

Abstract: Based on the analysis of the relation between the Karush-Kuhn-Tucker (KKT) conditions of Support Vector Machine(SVM) and the distribution of the training samples, the possible changes of support vector set after new samples are added to training set were analyzed, and the generalized Karush-Kuhn-Tucker conditions were defined. Based on the equivalence between the original training set and the newly added training set, a new algorithm for SVM-based incremental learning was proposed. With this algorithm, the useless samples were discarded and the useful training samples of importance were reserved. Experimental results with the standard dataset indicate the effectiveness of the proposed algorithm.

Key words: Support Vector Machine(SVM); incremental learning

0 引言

基于结构风险最小化原则的支持向量机(Support Vector Machine, SVM),由于具有很强的学习能力和较好的泛化性能,能够较好地解决小样本、高维数、非线性、局部极小等问题,可以有效地进行分类、回归、密度估计等^[1],因此得到了广泛的应用并成为了机器学习领域的研究热点^[2]。但是经典 SVM 学习算法并不直接支持增量式学习,所以对 SVM 增量学习方法的研究具有重要的理论意义和实用价值。

在增量学习的情况下,由于增量训练集的加入,打破了支持向量(Support Vector, SV)集和整个训练样本集的等价关系,从而有必要寻找新的 SV 集。如何进行有效的增量学习,利用历史训练结果让再次训练更快,并且如何在不损失分类精度的前提下抛弃无用的历史样本点等都是在 SVM 增量学习算法中需要考虑的重要问题^[3]。为此,文献[4]中提出了 α -SVM 支持增量学习,训练集主要从支持向量、误分数据并有选择的淘汰一些样本来获得,但该算法用户需选择很多参数,而确定这些参数的值也是一个非常棘手的问题。文献[5]中的增量训练由 SV 样本和新样本组成,再训练只需要进行一次即可完成,所有的非 SV 样本点都被抛弃。但是该方法同时抛弃了非 SV 样本中包含的重要信息。文献[6]分析了 SV 的性质和增量学习的过程,提出了一种新的 SVM 增量学习算法,但该方法同时淘汰了原样本集中可能成为新 SV

的非 SV 及满足 KKT 条件的新增样本,因此,一定程度上影响了增量学习精度的提高。

本文提出了一种新的 SVM 增量学习算法——对等增量 SVM 算法,该算法在增量学习中考虑了违背广义 KKT 条件的样本、以及满足广义 KKT 条件的样本中与原分类间隔距离较近的样本,从而在及时淘汰对后继分类影响不大的样本的同时保留了含有重要分类信息的样本。

1 增量学习过程分析

增量学习的主要任务是在增量样本加入时,有效利用历史训练结果,尽量避免样本的重复训练,得到比较准确的分类结果。如果新增样本带有原样本集不包含的分类信息,则学习后的 SV 集必然发生变化,以体现新信息的加入。如果考虑新增样本对 SV 集的影响,则面临下面几个问题:(1)什么样的新增样本使得 SV 集发生变化从而影响学习能力和泛化能力;(2)增量学习后原样本集中的 SV 如何变化;(3)什么样的向量可能成为最终训练结果的 SV,什么样的向量在以后的增量学习中最不可能成为 SV。

针对以上问题,首先分析 KKT 条件与样本之间的关系。

1) KKT 条件

对于分类问题,支持向量机把分类边界最大最终归结为如下凸半正定二次规划问题:

收稿日期:2006-04-10;修订日期:2006-06-27

基金项目:国家自然科学基金资助项目(50505051);陕西省自然科学基金计划项目(2004F36)

作者简介:王晓丹(1966-),女,陕西汉中,人,教授,博士,主要研究方向:模式识别、智能信息处理; 郑春颖(1979-),女,陕西西安人,硕士,主要研究方向:支持向量机、模式识别; 吴崇明(1966-),男,江苏淮安人,讲师,硕士,主要研究方向:智能信息处理; 张宏达(1981-),男,安徽临泉人,博士研究生,主要研究方向:智能信息处理。

$$\max W(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j K(x_i \cdot x_j) \quad (1)$$

$$\text{s. t. } \sum_{i=1}^l \alpha_i y_i = 0 \quad (2)$$

$$\alpha_i \in [0, C], i = 1, \dots, l \quad (3)$$

判决函数为:

$$f(x) = \sum_{i=1}^l \alpha_i y_i K(x_i \cdot x) + b \quad (4)$$

$$y = \text{sign}(f(x)) \quad (5)$$

α 为 Lagrange 乘子, 当且仅当对于每一个 x 都满足如下 KKT 条件时, $\alpha = [\alpha_1, \dots, \alpha_l]$ 才是上述规划的最优解:

$$\begin{cases} \alpha_i = 0 \Rightarrow y_i f(x_i) \geq 1 \\ 0 < \alpha_i < C \Rightarrow y_i f(x_i) = 1 \\ \alpha_i = C \Rightarrow y_i f(x_i) \leq 1 \end{cases} \quad (6)$$

$$\text{即} \begin{cases} \alpha_i = 0 \Rightarrow f(x_i) \geq 1 \text{ 或 } f(x_i) \leq -1 \\ 0 < \alpha_i < C \Rightarrow f(x_i) = 1 \text{ 或 } f(x_i) = -1 \\ \alpha_i = C \Rightarrow -1 \leq f(x_i) \leq 1 \end{cases}$$

所以, 对样本集进行训练得到 SVM 分类器, $\alpha = 0$ 对应的样本分布在分类器的分类间隔之外, $0 < \alpha < C$ 的样本位于分类间隔之上(在两条边界虚线上), $\alpha = C$ 的样本位于分类间隔之内(含边界虚线)。

综合以上三类情况可得: 满足 KKT 条件为: $y_i f(x_i) \geq 1$, 即满足 KKT 条件的样本为 SV, 及位于分类间隔之外且被分类器正确分类的样本。

设最优分类超平面为 $f(X) = 0$, $f(X) = \pm 1$ 为分类间隔的边界, 则满足 KKT 条件的几种情况分别如图 1 中 A1, A2 所示。

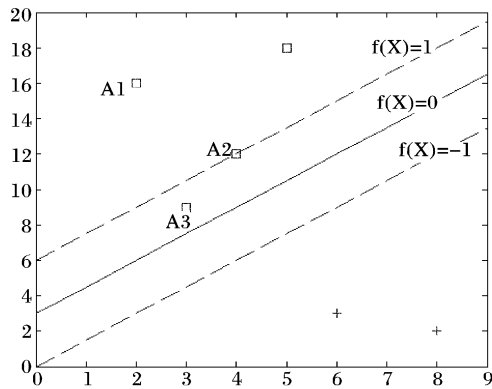


图 1 新增样本与 KKT 条件的关系

2) 新增样本对原 SV 集的影响

定理 1 对于线性约束二次规划中的 Hessian 矩阵 Q , 如果 Q 是半正定的, 则该规划是凸二次规划, 二次规划的一个可行解 α , 当且仅当每一个 x 都满足 KKT 条件(由 Lagrange 乘子 α 确定)时, α 才是上述规划的最优解^[2]。

由定理 1 可知, 新增样本中如果存在 x 不满足 KKT 条件, 则说明加入新样本后, 原始训练样本训练得到的 α 已不是问题的最优解, 根据 α 与 SV 的关系, 得定理 2。

定理 2 $f(X)$ 为 SVM 分类决策函数, $\{x_i, y_i\}$ 为新增样本, y_i 是样本点 x_i 的类别标签, 违背 KKT 条件的新增样本将使原 SV 集发生变化。

违背 KKT 条件的样本分为三类: 1) 位于分类间隔中, 与本类在分类边界同侧, 可以被原分类器正确分类的样本, 满足 $0 \leq y_i f(x_i) < 1$; 2) 位于分类间隔中, 与本类在分类边界异侧, 被原分类器错误分类的样本, 满足 $-1 \leq y_i f(x_i) \leq 0$; 3) 位于分类间隔外, 与本类在分类间隔异侧, 被原分类器错误分

类的样本, 满足 $y_i f(x_i) < -1$ 。

综合以上三类情况可得: 违背 KKT 条件等价于 $y_i f(x_i) < 1$ 。

定理表明: 对选择新增样本进行增量学习得到新的 SVM 分类器时, 采用 KKT 条件比用分类函数的分类判别作为选择标准更合理。分类错误只是样本违反 KKT 条件的特定情况。违背 KKT 条件的样本将会影响增量学习后的 SV 集。

因此新增样本可以分为两部分: 违背 KKT 条件的样本和满足 KKT 条件的样本, 在增量学习中需要考虑新增样本中违背 KKT 条件的样本。同时, 由于 SV 集反映了原 SVM 分类器所包含的信息, 因此, 在增量学习中还需要考虑原来的 SV 集。

定理 3 如果新增训练样本中存在违反 KKT 条件的样本, 则原样本集中非 SV 有可能转化为 SV^[7]。

上述定理说明, 如果在增量学习中只考虑新增样本和原来的 SV 集、而不考虑原来的非 SV 集, 虽然符合定理 2, 但定理 3 表明这样可能丢失原来样本集中有用的信息, 因而得到的结论不是最优的, 同时也舍弃了有用的样本。

由于原样本集中非 SV 是满足 KKT 条件的训练样本, 所以从定理 3 可以得出, 满足 KKT 条件的原样本有可能转化为新 SV。

以图 2 为例来说明新增样本训练后 SV 集可能发生的变化。图中 S1 ~ S5 为原 SV, A1 ~ A3 是新增样本。训练后的 SV 集由 S1, A3, N1 组成, N1 为原样本集中的非 SV。

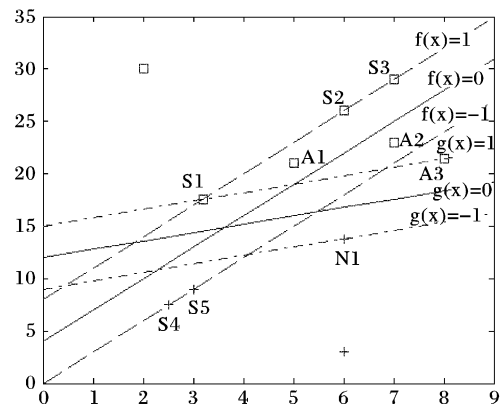


图 2 对新增样本训练后 SV 集可能发生的变化

图 2 中, 违背 KKT 条件的新增样本 A1, A2 在训练后转化为满足 KKT 条件的样本; 违背 KKT 条件的新增样本 A3 在训练后转化为 SV; 满足 KKT 条件的原样本 N1 在训练后转化为 SV; 原 SV 集中的 S2, S3, S4, S5 在训练后转化为满足 KKT 条件的样本。

3) 满足 KKT 条件的样本有可能转化成为新 SV

定理 4 如果新增训练样本中存在违反 KKT 条件的样本, 则满足 KKT 条件的样本有可能转化成为新 SV。

以图 3 为例进行说明。图 3 中 S1 ~ S5 为原 SV 集, A1 ~ A4 是新增样本。训练后的 SV 集由 S1, A3, A4, N1 组成。图 3 与图 2 的不同之处在于多增加了一个新增样本 A4, 新增样本 A4 满足 KKT 条件, 在训练后转化成为新 SV。同样, 满足 KKT 条件的原样本 N1 在训练后转化成为新 SV。由图可知, 这一部分样本分布在分类间隔附近。

所以, 满足 KKT 条件的新增样本有可能转化成为新 SV。在增量学习中, 还需要考虑一部分满足 KKT 条件的新增样本, 这一部分样本分布在分类间隔外的一定范围内。

定理 3 和 4 说明了满足 KKT 条件的新增样本、原样本都

有可能转化为新 SV。

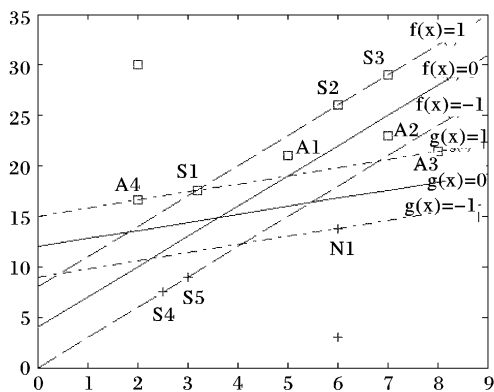


图3 对新增样本训练后 SV 集可能发生的变化

4) 广义 KKT 条件

若新增样本刚好是原 SVM 的 SV, 虽然该新增样本满足 KKT 条件, 该新增样本不影响增量学习后的最优分类面、不使 SV 集发生变化, 但该新增样本却成为了新 SV。因此, 可将上述违背 KKT 条件的定义放宽为: $y_i g(x_i) \leq 1$, 称之为违背广义 KKT 条件, 即是违背 KKT 条件的样本和 SV 集的并集。对应的满足广义 KKT 条件为: $y_i g(x_i) > 1$, 即是位于分类间隔之外且被分类器正确分类的样本。

基于上述定义, 违背广义 KKT 条件的样本将可能部分或全部成为增量学习后的新 SV。所以, 在增量学习中需要重点考虑违背广义 KKT 条件的样本, 此外, 满足广义 KKT 条件的样本中与原分类间隔距离较近的样本也需要考虑。

2 对等增量学习算法

增量学习的目标是获得原训练样本集与新增样本的并集的最优解。尽管在通常情况下, 原样本集比新增样本集大, 而且先进行学习, 但对于最终的学习结果, 所有样本的地位是等同的, 即原训练样本集和新增训练样本集地位等同。因此, 如果由新增样本集训练得到一个新的 SVM, 则原样本集可以看作这个新 SVM 的新增样本, 进而其中也可能存在违背新 SVM 的广义 KKT 条件的样本, 而这些样本也可能为增量学习后的 SV。基于以上分析, 我们提出一种新的对等增量学习算法。

基本思想: 首先用初始样本集训练 SVM₁, 检验是否存在新增样本违背 SVM₁ 的广义 KKT 条件, 如果不存在, SVM₁ 即为学习结果, 训练终止, 如果存在, 则再由新增样本训练得到 SVM₂, 检验是否存在原样本违背 SVM₂ 的广义 KKT 条件, 如果不存在, SVM₂ 即为学习结果, 训练终止, 如果仍然存在, 则由两个 SVM 的违背广义 KKT 条件的样本, 加上满足广义 KKT 条件的样本中与分类间隔距离较近的样本, 构成新的训练集。最后由新的训练集得到 SVM₃ 作为学习的最终结果。

在上述增量学习方案中, 新训练集的构成除了包含违背广义 KKT 条件的样本, 还包含了满足广义 KKT 条件的新增样本、以及原样本中与当前分类器最优分类面的垂直距离在一定范围内的样本。

设初始样本集为 X_0 , 新增样本集为 X_l , 用初始样本集训练得到的分类器为 Ω_0 , 用新增样本集训练得到的分类器为 Ω_l , X_l 中满足 Ω_0 的广义 KKT 条件的样本构成的集合为 X_l^S , X_l 中违背 Ω_0 的广义 KKT 条件的样本构成的集合为 X_l^{NS} , X_0 中满足 Ω_l 的广义 KKT 条件的样本构成的集合为 X_0^S , X_0 中违背 Ω_l 的广义 KKT 条件的样本构成的集合为 X_0^{NS} 。

算法描述如下:

1) 用 X_0 训练得到 SVM 分类器 Ω_0 ;

2) 检验 X_l 中的样本是否有违背 Ω_0 的广义 KKT 条件, 若没有, 则算法停止, Ω_0 为增量学习结果; 否则, 根据判断结果将 X_l 分为 X_l^S 和 X_l^{NS} ;

3) 用 X_l 训练得到 SVM 分类器 Ω_l ;

4) 检验 X_0 中的样本是否有违背 Ω_l 的广义 KKT 条件, 若没有, 则算法停止, Ω_l 为增量学习结果; 否则, 根据判断结果将 X_0 分为 X_0^S 和 X_0^{NS} ;

5) 将 Ω_0 的非 SV 与 X_l^S 合并得 X^S , 设定一阈值 ξ , 计算 X^S 中每个样本 X_i^S 与 Ω_0 的分类面的距离 d_i , 若 $d_i > \xi$, 将 X_i^S 从 X^S 中删除;

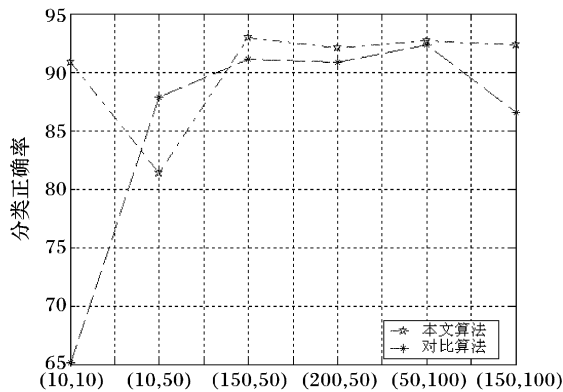
6) 将 Ω_l 的非 SV 与 X_0^S 合并得 X^S , 设定一阈值 η (η 可以与 ξ 取相同值), 计算 X^S 中每个样本 X_i^S 与 Ω_l 的分类面的距离 d_i' , 若 $d_i' > \eta$, 将 X_i^S 从 X^S 中删除;

7) 将 X_l^{NS} , X_0^{NS} , X^S , X^S 合并为 X_0 , 用 X_0 进行训练得到最终分类器 Ω 。

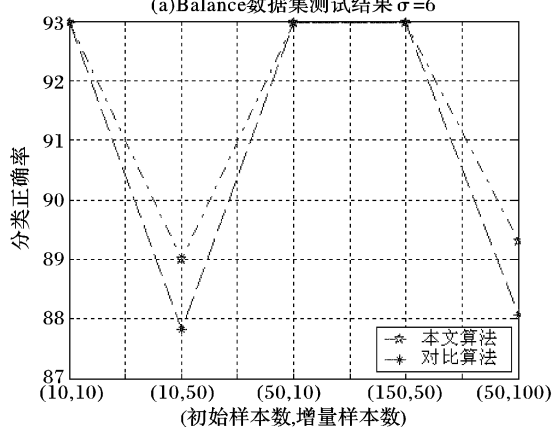
X_0 被保存下来, 作为下一次增量学习时的初始样本。

3 实验及结果

为验证对等增量学习算法的有效性, 选取两组 UCI 数据库中的数据进行了实验, 并与已有增量学习算法^[6]进行了比较, 以下将该算法称为对比算法。



(a) Balance 数据集测试结果 $\sigma=6$



(b) Balance 数据集测试结果 $\sigma=10$

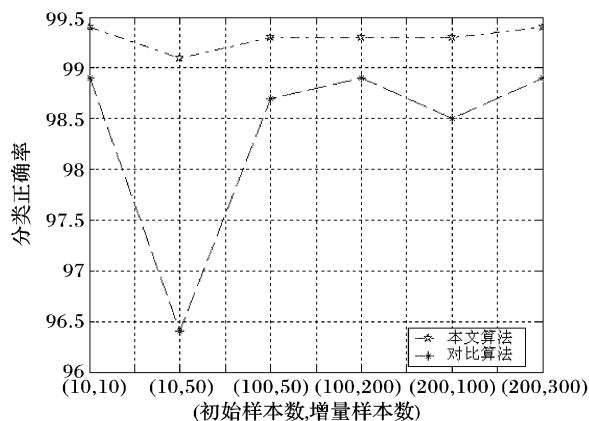
图4 对 Balance 数据集的实验结果

第一组实验数据为 Balance 数据集, 其样本特征维数为 4, 样本总数为 627, 分 3 类 (49 balanced, 290 left, 288 right), 以第 2 类作为正例, 其他类作为负类进行训练, 取前 300 个样本作为训练样本, 后 327 个样本作为测试样本; 第二组实验数据为 Letter-recognition 数据集, 共有 20 000 个样本, 样本维数为 16, 样本类别数为 26, 以第一类作为一类, 其他类作为另一类

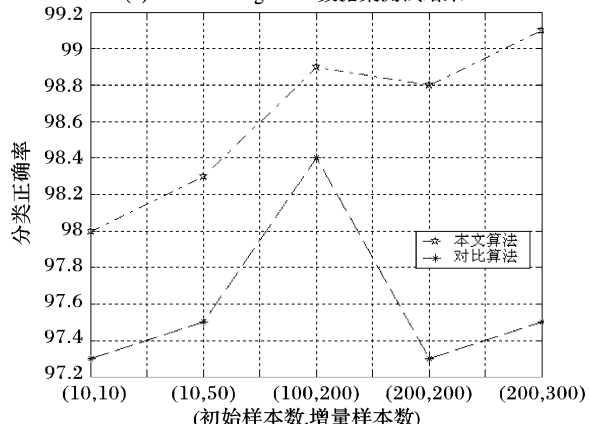
进行训练,选取其中的 600 个样本作为训练样本,1000 个样本作为测试样本。

实验表明,在其他各项条件相同的条件下,采用径向基函数比采用其他核函数效果要好。图 4、图 5 给出了以径向基函数作为 SVM 核函数的实验结果,惩罚参数 $C = 1000$, 阈值 $\xi = \eta = 1.0001$ 。图中,横坐标值标明了增量学习初始样本个数和增量样本个数,纵坐标为对应的分类正确率。

从图 4、图 5 可以看出:本文的对等增量 SVM 算法与已有算法相比精度明显提高。



(a) Letter-recognition 数据集测试结果 $\sigma=8$



(b) Letter-recognition 数据集测试结果 $\sigma=13$

图 5 对 Letter-recognition 数据集的实验结果

在对比算法中,使用 $SV \cup ASV \cup ESV$ 作为新增样本加入后的训练样本集进行训练,其中 SV 为原训练样本集中的 SV 集, ASV 为增量样本集中的错分向量, ESV 为被正确分类的,与最优分类面邻近,处在最优分类面和间隔平面之间的样本,实际上, $ASV \cup ESV$ 就是违背原 SVM 的 KKT 条件的新增样本

集, $SV \cup ASV \cup ESV$ 是违背原 SVM 的广义 KKT 条件的新增样本集。

在当样本的统计性质比较差时,历史样本和增量样本分布不相似,甚至分布差异十分显著时,若用 $SV \cup ASV \cup ESV$ 作为新增样本加入后的训练样本集进行训练,结果将会严重偏离真正的最优分类面,因此, $SV \cup ASV \cup ESV$ 还不能完全表示新增样本加入后训练样本集所含的信息。

本文提出的对等增量 SVM 算法考虑了满足广义 KKT 条件的样本中与分类间隔距离较近的样本,而这些样本也可能成为增量学习后的 SV ,从而在及时淘汰对后继分类影响不大的样本的同时保留了含有重要分类信息的样本。

4 结语

本文提出了一种新的 SVM 增量学习算法——对等增量 SVM 算法。算法在增量学习中考虑了可能成为增量学习后的新 SV 的训练样本,即违背广义 KKT 条件的样本,以及满足广义 KKT 条件的样本中与原分类间隔距离较近的样本,从而在及时淘汰对后继分类影响不大的样本的同时保留了原样本集和新增样本集中含有重要分类信息的样本。使增量学习的结果能够准确反映训练样本集的变化。对标准数据集的实验结果表明,本算法可以在新增样本加入后有效淘汰无用样本,同时保留含有重要分类信息的样本,从而获得了较好的分类性能。

参考文献:

- [1] BURGESS CJC. A tutorial on support vector machines for pattern recognition[J]. Data Mining and Knowledge Discovery, 1998, 2(2): 121-167.
- [2] VAPNIK V. 统计学习理论本质[M]. 北京: 清华大学出版社, 2000.
- [3] CAUWENBERGHS G, POGGIO T. Incremental and decremental support vector machine learning[J]. Machine Learning, 2001, 44(13): 4098-4151.
- [4] 萧嵘, 王继成, 孙正兴, 等. 一种 SVM 增量学习算法——ISVM[J]. 软件学报, 2001, 12(12): 1818-1824.
- [5] SYED N, LIU H, SUNG KK. Incremental learning with support vector machines[A]. Proc. Workshop on Support Vector Machines at the International Joint Conference on Artificial Intelligence (IJ-CAI-99)[C]. Stockholm, Sweden, 1999.
- [6] 滕月阳, 唐煥文, 张海霞. 一种新的支持向量机增量学习算法[J]. 计算机工程与应用, 2004, 36: 77-80.
- [7] 周伟达, 张莉, 焦李成. 支撑向量机推广能力分析[J]. 电子学报, 2001, 29(5): 590-594.
- [8] KAK S. On generalization by neural networks[J]. Information Sciences, 1998, 111: 293-302.
- [9] TANG KW, KAK S. Fast Classification Networks for Signal Processing[J]. Circuits Systems Signal. Processing, 2002, 21(2): 207-224.
- [10] 韩小云, 周建平, 刘瑞岩. 广义聚类神经网络 GC[J]. 数据采集与处理, 1999, 14(1): 1-4.
- [11] RAINA P. Comparison of learning and generalization capabilities of the Kak and the back propagation algorithms[J]. Information Sciences 1994, 81: 261-274.
- [12] 欧洁, 林守勋. 基于贝叶斯网络模型的信息检索[J]. 微电子学与计算机, 2005, 5(1): 83-87.
- [2] 曾春, 邢春晓, 周立柱. 个性化服务技术综述[J]. 软件学报, 2002, 13(10): 1952-1961.
- [3] SAKAGAMI H, KAMBA T, SUGIURA A. Effective personalization of push-type systems: visualizing information freshness[J]. Computer Networks and ISDN Systems, 1998, 30(1-7): 53-63.
- [4] CHEN EH, ZHEN YZ, XU FW. An extended corner classification neural Network based document classification approach [J]. Journal of Software, 2002, 13(5): 871-878.
- [5] 朱大铭, 马绍汉. 二进制神经网络分类问题的几何学习算法[J]. 软件学报, 1997, 8(8): 622-629.
- [6] 张振亚. 基于文本信息检索的知识发现技术研究[D]. 安徽: 中国科学技术大学计算机系, 2004.
- [7] TANG KW, KAK SC. A new corner classification approach to neural network training[J]. Circuits Systems Signal Processing, 1998, 17(4): 459-469.

(上接第 2439 页)