

渐进支持向量的一种新颖选取策略

申丰山^{1,2} 马玉军³ 张军英¹

(西安电子科技大学计算机学院 西安 710071)¹ (郑州大学信息工程学院 郑州 450052)²
(南阳理工学院网络中心 南阳 473004)³

摘要 渐进支持向量决定渐进支持向量机的泛化能力,其选取至关重要。对此提出了一种新颖的基于概率计算的渐进支持向量选取策略。该方法为每个样本点构造一个通过该样本点的合适分离面,该样本点成为渐进支持向量的概率是根据该分离面对两类样本的分离率来估计的。具有较高概率值的训练样本被选为渐进支持向量,用以训练和更新渐进支持向量机。比较性的实验表明,该方法在保持渐进支持向量机泛化能力的前提下,在训练效率上具有非常突出的优势。

关键词 渐进支持向量机,渐进支持向量,支撑分离面,支撑分离率

Novel Strategy for Selecting Incremental Support Vectors

SHEN Feng-shan^{1,2} MA Yu-jun³ ZHANG Jun-ying¹

(School of Computer Science and Engineering, Xidian University, Xi'an 710071, China)¹

(School of Information Engineering, Zhengzhou University, Zhengzhou 450052, China)²

(Network Information Center, Nanyang Institute of Technology, Nanyang 473004, China)³

Abstract Incremental support vectors determine the generalization performance of incremental support vector machine, so their choosing is significant. We proposed a novel strategy to choose the incremental support vectors by computing their probabilities. The method constructs an appropriate separating hyperplane for each training sample, which goes through the corresponding training sample point. And then the probability of a sample to be incremental support vector is estimated by the separating rate of the corresponding hyperplane. Those samples with higher probabilities are chosen as the incremental support vectors to train and update the incremental support vector machine. Comparative numerical experiments of our method against the existing methods show that our method has outstanding advantage in the training efficiency without deteriorating the generalization performance of incremental support vector machine.

Keywords Incremental support vector machine, Incremental support vector, Support separating hyperplane, Support separating rate

1 引言

渐进机器学习在获取新知识的同时也不会忘记已经学过的知识^[1,2],且渐进学习不需要检查所有的历史训练样本去更新它的知识,因此在 CPU 时间和内存空间的占用上比较经济。将渐进学习方法与支持向量机(support vector machine, SVM)^[3-5]结合起来,形成 ISVM (incremental support vector machine, 渐进支持向量机),会使 SVM 的学习行为更自然,更具适应性。

传统 SVM 的泛化性能和判决函数依赖于支持向量,因此传统 SVM 的训练结果表现为支持向量,支持向量是对历史知识的概括。将支持向量结合到后继训练样本中的 SVM 渐进学习方法,有利于 ISVM 在学习新知识的同时记住已经学过的知识^[6,7]。但是在渐进学习背景下,支持向量不能完全反映渐进学习的动态特性,仅用支持向量不足以表示前瞻性的知识,进而有损 ISVM 的泛化性能。而且支持向量的获

取需要求解二次规划,耗时相当严重。因此有必要引入渐进支持向量的概念来描述既能代表历史知识又具有前瞻性的那部分训练样本,并开发高效算法来获取渐进支持向量。文献[8,9]的算法对这种思想有所体现。文献[8]提取了包含支持向量、被渐进构造的分离面误分的样本或者逾越了 ISVM 两个边界的样本在内的重要训练样本来训练更新 ISVM。文献[9]则通过计算每个样本到另一类各个样本之间的最小距离并选取相应样本来提高 ISVM 的训练效率。但是上述方法缺乏鲁棒性,当分离面发生较大旋转时,ISVM 会遗忘部分知识^[10,11],进而降低其泛化性能。为此文献[10,11]提出使用 One-Class SVM^[12]为 ISVM 选取渐进支持向量,以增强其鲁棒性,改善其泛化性能。但是当训练集增大时,该方法会产生耗时严重的问题,因为构造 One-Class SVM 相当于求解二次规划,其计算复杂度与求解传统 SVM 相当。而且,One-Class SVM 是一个超球 SVM,其性能容易受到样本分布特征的影响。

到稿日期:2009-03-06 返修日期:2009-06-18

申丰山 博士生,主要研究方向为模式识别、机器学习等,E-mail:iefssh@zhu.edu.cn;马玉军 硕士,讲师,主要研究方向为计算机应用及网络安全等;张军英 教授,博士生导师,主要研究方向为模式识别、机器学习、图像处理、生物信息技术等。

本文提出了渐进支持向量获取的一种新方法,称为支撑分离面法(support separating hyperplane, SSH), SSH 对分离面的旋转同样是鲁棒的,但是具有更低的计算代价。SSH 为每个样本点构造一个穿过该样本点的合适超平面去分两类数据。每类数据的分离率用于估计该样本点成为渐进支持向量的概率。每个训练样本成为渐进支持向量的概率都依此计算。概率较高的样本被选为渐进支持向量,用于训练和更新 ISVM。实验表明,在人工数据集和标准数据集上与文献[12]的算法相比,我们的算法在训练效率上优势突出,常常达到几百倍的加速比且没有明显牺牲 ISVM 的泛化性能。

2 ISVM (渐进支持向量机)

训练数据在 SVM 渐进学习中是渐进获得的,作为学习结果的支持向量具有动态特性,为了刻画这种动态特性,引入了渐进支持向量的概念。与传统支持向量不同的是,渐进支持向量不仅仅要包含历史知识,而且还要具有前瞻性,使 ISVM 在保持泛化能力上具有较好的鲁棒性,避免或减少知识的遗忘。ISVM 的渐进训练过程可以表示为函数:

$$[SVC, w, b] = IncSVM(X, SVC_{pre}) \quad (1)$$

其中, X 为后继训练集, SVC_{pre} 是历史渐进支持向量集, SVC 是 ISVM 在 $X \cup SVC_{pre}$ 上训练所得的新渐进支持向量集, w 和 b 分别是判决函数的权重和偏置。IncSVM 完成渐进学习的过程,其核心步骤包括渐进支持向量的获取和 ISVM 的训练,前者决定后者的效率和精度,因此前者是本文研究的重点。假定 X_1 和 X_2 是先后两个渐进训练集,那么在两个集合上的渐进学习是通过两次调用函数 IncSVM 来进行的:

$$[SVC_1, w_1, b_1] = IncSVM(X_1, \phi)$$

$$[SVC_2, w_2, b_2] = IncSVM(X_2, SVC_1)$$

其中, ϕ 表示空集,指明初始渐进支持向量集是空的。

3 渐进支持向量的选取

在几何意义上,渐进支持向量通常位于两类数据交遇区的边缘位置,如图 1 中较大的图符所示。

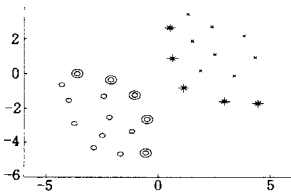


图 1 渐进支持向量(用较大图符表示)

在渐进学习过程中,当分离面发生较大旋转时,这些边缘位置的样本点极有可能成为支持向量,因此需要加以识别和保留,以作进一步的渐进学习。为此,提出了一种新的方案用于快速有效地获取渐进支持向量,加速 ISVM 的渐进训练进程。在给出该方案的原理思想之前,首先给出相关定义。下文中,在由核函数 $K(x, y) = \Phi(x) \cdot \Phi(y)$ (其中 $\Phi: R^d \rightarrow F$) 间接映射的特征空间中考虑该问题。

定义 1(分离率) 令 P 是正的样本集, M 是负的样本集, $H(w, c): f(x) = w \cdot \Phi(x) + c$ 是用于分开两类样本的分离面的函数。对于 $x_k \in S$ (其中 $S = P$ 或 $S = M, k = 1, \dots, |S|$) 和 $f(\Phi(x_k)) < 0$, 令:

$$N_S = \sum_{k=1}^{|S|} \text{sign}(f(\Phi(x_k))) \quad (2)$$

其中,

$$\text{sign}(x) = \begin{cases} 1 & (\text{if } x > 0) \\ 0 & (\text{if } x = 0) \\ -1 & (\text{if } x < 0) \end{cases} \quad (3)$$

式子

$$r_S = \frac{N_S}{|S|} \quad (4)$$

称为 $H(w, c)$ 关于数据集 S 的分离率。对于另一类数据 $x_k \in \bar{S}$ (其中 $\bar{S} = P \cup M - S, k = 1, \dots, |\bar{S}|$) 和 $f(\Phi(x_k)) > 0$, 令 $N_{\bar{S}} = \sum_{k=1}^{|\bar{S}|} \text{sign}(-f(\Phi(x_k)))$, 式子

$$r_{\bar{S}} = \frac{N_{\bar{S}}}{|\bar{S}|} \quad (5)$$

称为 $H(w, c)$ 关于数据集 \bar{S} 的分离率。

定义 2(支撑分离面和支撑分离率) 在定义 1 中,如果分离面 $H(w, c)$ 穿过特征空间的点 $\Phi(s)$ ($s \in S, S = P$ 或 $S = M$), 则该分离面称为 s 的支撑分离面, 其函数为 $(\Phi(s) - o)^T (x - \Phi(s)) = 0$, 标记为 $H(\Phi(s), o)$ 。 $H(\Phi(s), o)$ 对两类数据的分离率称为 s 的支撑分离率, 分别标记为 $r_S^{(\Phi(s), o)}$ 和 $r_{\bar{S}}^{(\Phi(s), o)}$, 其中 o 称为基准点, 与 $\Phi(s)$ 一起决定分离面的法线方向。 o 的选取原则是使得 s 具有较高的支撑分离率。式 (4) 变成:

$$r_S^{(\Phi(s), o)} = \frac{\sum_{k=1}^{|S|} \text{sign}((\Phi(s) - o)^T (\Phi(x_k) - \Phi(s)))}{|S|} \quad (6)$$

其中, $x_k \in S (k = 1, \dots, |S|)$ 且 $(\Phi(s) - o)^T (\Phi(x_k) - \Phi(s)) < 0$ 。

式(5)变成:

$$r_{\bar{S}}^{(\Phi(s), o)} = \frac{\sum_{k=1}^{|\bar{S}|} \text{sign}(-(\Phi(s) - o)^T (\Phi(x_k) - \Phi(s)))}{|\bar{S}|} \quad (7)$$

其中, $x_k \in \bar{S} (k = 1, \dots, |\bar{S}|)$ 且 $(\Phi(s) - o)^T (\Phi(x_k) - \Phi(s)) > 0$ 。

我们的方案基于这样一个事实,即样本 s 的支撑分离率越高,则该样本越可能成为渐进支持向量,因为 s 的支撑分离面可能成为或接近成为间隔边界,而间隔边界穿过支持向量, $\Phi(s)$ 是其中之一。这个事实可以用于估计一个样本成为渐进支持向量的概率。给定门限 η_1 和 η_2 , 如果 $r_S^{(\Phi(s), o)} \geq \eta_1$ 且 $r_{\bar{S}}^{(\Phi(s), o)} \geq \eta_2$, 则 s 就被选作渐进支持向量。

显然,样本 s 的支撑分离面有无限个,需要选取一个合适的分离面,使得 s 具有尽可能高的支撑分离率,实际上就是确定合适的基准点 o 。考虑到计算代价和精确性,将基准点定位于两类数据中心的连线方向上。对于正类样本集 P 和负类样本集 M , 分别确定不同的基准点 o_P 和 o_M 如下:

$$o_P = \frac{(c_P - c_M g_P)}{(1 - g_P)} \quad (g_P \in [0, 1]) \quad (8)$$

$$o_M = \frac{(c_M - c_P g_M)}{(1 - g_M)} \quad (g_M \in [0, 1]) \quad (9)$$

其中, $c_P = \frac{1}{|P|} \sum_{i=1}^{|P|} \Phi(p_i), p_i \in P$ (10)

$$c_M = \frac{1}{|M|} \sum_{j=1}^{|M|} \Phi(m_j), m_j \in M \quad (11)$$

g_P 和 g_M 是用户定义参数。

每个向量 $s \in S$ 的支撑分离率 $r_S^{(\Phi(s), o_S)}$ 和 $r_{\bar{S}}^{(\Phi(s), o_S)}$ 都可计算得到,其中 $o_S = o_P$ 如果 $S = P$ 或者 $o_S = o_M$ 如果 $S = M$ 。这样,对于 $s \in S$, 式(6)和式(7)变成:

$$r_S^{(\Phi(s), o_S)} = \frac{\sum_{k=1}^{|\bar{S}|} \text{sign}((\Phi(s) - o_S)^T (\Phi(x_k) - \Phi(s)))}{|\bar{S}|} \quad (12)$$

其中, $x_k \in S (k=1, \dots, |S|)$ 且 $(\Phi(s) - o_S)^T (\Phi(x_k) - \Phi(s)) < 0$ 。

$$r_S^{(\Phi(s), o_S)} = \frac{\sum_{k=1}^{|\bar{S}|} \text{sign}(-(\Phi(s) - o_S)^T (\Phi(x_k) - \Phi(s)))}{|\bar{S}|} \quad (13)$$

其中, $x_k \in \bar{S} (k=1, \dots, |\bar{S}|)$, $(\Phi(s) - o_S)^T (\Phi(x_k) - \Phi(s)) > 0$ 。

由此样本 s 成为渐进支持向量的条件为:

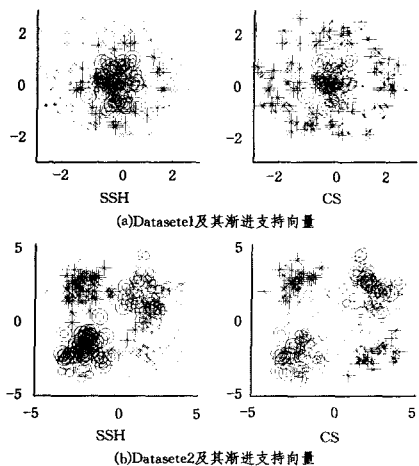
$$r_S^{(\Phi(s), o_S)} \geq \eta \text{ 且 } r_{\bar{S}}^{(\Phi(s), o_S)} \geq \eta$$

其中, η 和 η 是给定的门限。

由式(12)和式(13)易得算法的计算复杂性为 $O(N^2)$, N 为样本总数。

4 实验与分析

我们进行了两组比较实验,将所提算法 SSH 与文献[12]中的算法 CS(同心超球法,co-sphere) 进行比较,以考察 SSH 的有效性和分类性能(包括速度和泛化能力)。有效性实验在两个二维数据集上进行,观察不同算法所获得的渐进支持向量的分布位置。两个数据集分别为 Dataset1 和 Dataset2, 各由正类数据集 $P = \{(p_{i1}, p_{i2})\}_{i=1}^{200}$ 和负类数据集 $M = \{(m_{i1}, m_{i2})\}_{i=1}^{200}$ 组成。Dataset1 是一类数据环状包围另一类数据的非线性可分数据集,其中 $p_{i1} = r_1 \times \cos(\theta_1)$, $p_{i2} = r_1 \times \sin(\theta_1)$, $r_1 \sim N(2, 1)$ (服从均值为 2, 方差为 1 的高斯分布), $\theta_1 \sim N(\pi, \pi)$, $m_{i1} = r_2 \times \cos(\theta_2)$, $m_{i2} = r_2 \times \sin(\theta_2)$, $r_2 \sim N(6, 2)$, $\theta_2 \sim N(\pi, \pi)$; Dataset2 是相互交错呈菊花状的非线性可分数据集,其中 $P = P1 \cup P2, M = M1 \cup M2, P1 = \{(p_{i1}^1, p_{i2}^1)\}_{i=1}^{100}, P2 = \{(p_{i1}^2, p_{i2}^2)\}_{i=1}^{100}, M1 = \{(m_{i1}^1, m_{i2}^1 + 4)\}_{i=1}^{100}, M2 = \{(m_{i1}^2, m_{i2}^2 - 4)\}_{i=1}^{100}, p_{i1}^1, p_{i2}^1 \sim N(-2, 0.8), p_{i1}^2, p_{i2}^2 \sim N(2, 0.8), m_{i1}^1, m_{i2}^1 \sim N(-2, 0.8), m_{i1}^2, m_{i2}^2 \sim N(2, 0.8)$ 。图 2 显示了两个数据集的分布和由 SSH 和 CS 分别获取的渐进支持向量。



(左边子图表示 SSH 的结果,右图表示 CS 的结果。两类训练样本分别用圆圈和 'x' 表示,两类渐进支持向量分别用稍大的圆圈和稍大的星号来表示。)

图 2 算法获取的渐进支持向量

由图 2 可以看出,SSH 获得了绝大多数重要的样本作为渐进支持向量而没有引入太多不必要的样本,而 CS 则选取了较多不必要的样本。从图 2(b)还可以看出 CS 甚至存在明

显漏选支持向量的现象。漏选支持向量虽然有时未必会降低 ISVM 的分类精度,但是会降低其泛化能力和鲁棒性。这表明我们的方法能够更有效地获取渐进支持向量。

然后,在基准数据集上进行比较实验来考察不同算法的分类精度、训练时间以及所选渐进支持向量占样本总数的比率 ISVR (incremental support vector rate)。基准数据集的信息在表 1¹⁾ 中给出。训练数据集从基准数据集中随机选取,每个训练集由 3 个渐进训练子集和 1 个测试子集构成。训练过程就是以每个渐进训练子集和渐进支持向量集依次调用渐进学习函数(1)。

所有数据集都是二分类问题相关的。我们将分析限定于两类情况的原因在于,在 SVM 背景下,多类问题通常都可以分解为多个二分类问题。

所有算法都在配置 Intel Core2Duo 处理器(1.6GHz)、1GB 内存的 PC 机上用 Matlab7.0 实现。在所有实验中,选用高斯核 $K(x, y) = \exp(-\gamma \|x - y\|^2)$ 训练 SVM,核参数 γ 和 C 取经过 5 折交叉检验后具有较高分辨率的值。表 1 第 5 列给出了每个训练集的 γ 和 C 值。SVM 产生的对偶二次规划问题采用 SVMlight^[13] 分解方法解决。

表 1 基准数据集的信息

数据	训练集	测试集	属性	γ, C
banana	400 * 3	400	2	0.25, 32
german	700 * 3	700	20	0.25, 2
image	1300 * 3	1300	18	0.25, 256
ringnorm	400 * 3	400	20	1, 0.5
splice	1000 * 3	1000	60	0.25, 4
thyroid	140 * 3	140	5	0.25, 4
twonorm	400 * 3	400	20	0.25, 0.12
waveform	400 * 3	400	21	0.25, 2

SSH 中 g_P, g_M, η 和 η 固定为能够获得较好泛化性能的值。对每一组训练集,其具体选取办法为:任选本组训练集中的一个渐进训练子集,首先将 η 和 η 固定在 50%, g_P 和 g_M 的值依次在 $[0.99, 0.9, 0.8, 0.7, 0.6, 0.5]$ 中选取,并总是令 $g_P = g_M$,在此参数配置下从该渐进训练子集中获取渐进支持向量训练 SVM,并获得其在该子集上的分类精度,为 g_P 和 g_M 取分类精度较高的参数值。然后以 g_P (或 g_M) 为中心,在 $[g_P - 0.05, g_P + 0.05]$ 区间以步长 0.01 作微调,得到 g_P 和 g_M 的最终值。之后固定 g_P 和 g_M ,在 $[0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3]$ 中寻找最优 η 和 η ,并总是令 $\eta = \eta$ 。最后以 η (或 η) 为中心,在 $[\eta - 0.05, \eta + 0.05]$ 区间以步长 0.02 进行微调确定最终的 η (或 η)。每个数据集的参数值及其计算时间在表 2 第 2, 3, 4 列给出。我们发现 SSH 使用这种参数设置在多个数据集上能够良好地工作。我们也会调整某些数据集的参数值。

表 2 针对基准数据集的参数设置

数据	g_P, g_M	η_1, η_2	计算时间(s)	ρ, θ
banana	0.96, 0.96	0.35, 0.35	13.60	0.90, 0
german	0.99, 0.99	0.35, 0.35	36.47	0.80, 0
image	0.85, 0.85	0.35, 0.35	826.72	0.90, 0
ringnorm	0.99, 0.99	0.50, 0.50	7.90	0.90, 0
splice	0.99, 0.99	0.50, 0.50	35.81	0.90, 0
thyroid	0.99, 0.99	0.50, 0.50	5.05	0.80, 0
twonorm	0.95, 0.95	0.70, 0.70	11.70	0.90, 0
waveform	0.99, 0.99	0.50, 0.50	11.31	0.90, 0

(下转第 220 页)

¹⁾ <http://ida.fraunhofer.de/projects/bench>

Chinese with English abstract)

- [4] Li D Y, Cheung D, Ng V, et al. Uncertainty reasoning based on cloud models in controllers[J]. Computers and Mathematics with Application Elsevier Science, 1998, 35(3): 99-123
- [5] Li D Y, Liu C Y, Du Y, et al. Artificial Intelligence with Uncertainty[J]. Journal of Software, 2004, 15(11): 1583-1594 (in Chinese with English abstract)
- [6] Li D Y. Uncertainty in Knowledge Representation[J]. Engineering Science, 2000, 2(10): 73-79 (in Chinese with English abstract)
- [7] Wang P, Li J P. Quantum Interpretation of Frequency Operator [C]//ICNC'07: Third International Conference on Natural Computation, China: IEEE Computer Society Press, 2007: 613-618
- [8] Wang P, Li J P. Quantum Interpretation of Signal's Uncertainty

Principle[J]. Journal of University of Electronic Science and Technology of China, 2008, 37(1): 14-16 (in Chinese with English abstract)

- [9] Feynman R P. Simulating physics with computers[J]. International Journal of Theoretical Physics, 1982, 26(21): 467-488
- [10] Bennett C H, Divincenzo D P. Quantum information and computation[J]. Nature, 2000, 404(3): 247-255
- [11] Holland J H. Outline for a logical theory of adaptive systems [J]. Journal of the Association for Computing Machinery, 1962, 3: 297-314
- [12] Holland J H. Genetic algorithm and the optimal allocations of trials[J]. SIAM Journal of Computing, 1973, 2: 88-105
- [13] Kirkpatrick S, Gelatt C D, Vecchi M P. Optimization by Simulated Annealing[J]. Science, 1983, 220(4598): 671-680

(上接第 202 页)

CS 的参数 ρ 和 θ 设定为使该算法能得到较好的泛化性能和较低的 ISVR, 具体数值在表 2 第 5 列给出。表 3、表 4 给出了两个算法在基准数据集上的识别率(包括训练精度和测试精度)、训练时间和 ISVR。由于训练时间过长, CS 在 german, image 和 splice 数据集上的结果没有报告。

表 3 SSH 和 CS 在基准数据集上的 ISVR 和训练时间

数据	ISVR (%)		训练时间(s)	
	CS	SSH	CS	SSH
banana	73.17	59.92	1227.90	41.27
german	—	49.67	Very long	14.74
image	—	62.59	Very long	702.78
ringnorm	99.17	50.42	1622.00	2.93
splice	—	56.37	Very long	21.77
thyroid	50.71	30.71	8.28	1.26
twonorm	72.92	35.75	920.76	2.97
waveform	57.25	30.83	2435.60	5.08

表 4 SSH 和 CS 在基准数据集上的识别率

数据	训练精度(%)		测试精度(%)	
	CS	SSH	CS	SSH
banana	89.92	84.92	89.25	84.25
german	—	82.43	—	82.43
image	—	90.62	—	90.38
ringnorm	99.83	99.83	99.25	99.25
splice	—	90.80	—	91.10
thyroid	98.33	98.33	98.57	98.57
twonorm	97.58	98.00	96.75	96.50
waveform	93.42	94.17	90.25	90.25

(在表 3、表 4 中, 由于 CS 算法的一些结果没有在 4 个小时内得到, 我们不得不停止了计算。)

由表 3、表 4 可以看出, SSH 的识别率与 CS 相当, 但是 SSH 的训练时间远少于 CS。即使加上参数值的寻找时间, SSH 总的时间仍然远小于 CS。很明显, 我们的算法在计算效率上具有异常突出的优势。SSH 的 ISVR 也明显低于 CS。这表明我们的算法在内存空间和 CPU 时间的占用上比 CS 更经济。这得益于 SSH 本身的高效率节省了大量的时间, 同时其较低的 ISVR 降低了训练代价。

结束语 本文提出了一种渐进支持向量获取的新方法, 即支撑分离面法。其主要思想基于下面事实: 支持向量位于间隔边界上, 如果支撑分离面会变成间隔边界或与其比较接近, 那么位于其上的样本将极其可能成为支持向量, 进而成为 ISVM 的渐进支持向量。该方法根据样本点的支撑分离率来估计该样本成为渐进支持向量的概率, 进而选取概率值较高的样本作为渐进支持向量去训练和更新 ISVM。实验结果表

明该方法能够高效地获取渐进支持向量, 能极大地提高 ISVM 的训练速度却不牺牲其泛化能力。

参考文献

- [1] Jia Hongbin, Lu Yi, Murphey, et al. Identifying knowledge domain and incremental new class learning in SVM[C]//JCNN '05. 2005: 2742-2747
- [2] Wan Sheng, Banta L E. Parameter Incremental Learning Algorithm for Neural Networks[J]. IEEE Trans. Neural Networks, 2006, 17(6): 1424-1438
- [3] Vapnik V. The Nature of Statistical Learning Theory[M]. London, UK: Springer-Verlag, 1995
- [4] Burges C J C. A Tutorial on Support Vector Machines for Pattern Recognition[J]. Data Mining and Knowledge Discovery, 1998, 2(2): 1-47
- [5] Cristianini N, Shawe-Taylor J. An introduction to Support Vector Machines and other kernel-based learning methods[M]. Cambridge, UK: Cambridge University Press, 2000
- [6] Mitra P, Murthy, et al. Data condensation in large databases by incremental learning with support vector machines[C]//Proc. Internat. Conf. on Pattern Recognition. 2000: 2708-2711
- [7] Domeniconi C, Gunopulos D. Incremental Support Vector Machine Construction[C]//Proc. ICDM 2001. 2001: 589-592
- [8] Xiao Xianbo, Hu Guangshu. An Incremental Support Vector Machine based Speech Activity Detection Algorithm[C]//Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, Shanghai, China, 2005: 4224-4226
- [9] An Jin-long, Zhen Gou-wang, et al. An Incremental Learning Algorithm for Support Vector Machine[C]//Proceedings of the Second International Conference on Machine Learning and Cybernetics, Xi'an, 2003: 1153-1156
- [10] Katagiri S, Abe S. Selecting support vector candidates for incremental training[C]//IEEE International Conference on Systems, Man and Cybernetics. 2005(2): 1258-1263
- [11] Katagiri S, Abe S. Incremental training of support vector machines using hyperspheres[J]. Pattern Recognition Letters, 2006(27): 1495-1507
- [12] Shawe-Taylor J, Cristianini N. Kernel Methods for Pattern Analysis[M]. UK: Cambridge University Press, 2004
- [13] Joachims T. Making large-scale SVM learning practical[M]//Schölkopf B, Burges C, Smola A, eds. Advances in Kernel Methods-Support Vector Machines. Cambridge, MA: MIT Press, 1999