

文章编号: 1002-0411(2002) 01-045-06

支持向量机训练算法综述

刘江华 程君实 陈佳品

(上海交通大学信息存储研究中心 上海 200030)

摘 要: 本文介绍统计学习理论中最年轻的分支——支持向量机的训练算法,主要有三大类:以 SVM-light 为代表的分解算法、序贯分类方法和在线训练法,比较了各自的优缺点,并介绍了其它几种算法及多类分类算法.最后指出了支持向量机具体实现的方向及其在模式识别、数据挖掘、系统辨识与控制等领域中的应用.

关键词: 支持向量机; 训练算法; 统计学习理论

中图分类号: TP391.4

文献标识码: B

SUPPORT VECTOR MACHINE TRAINING ALGORITHM A REVIEW

LIU Jiang-hua CHENG Jun-shi CHEN Jia-pin

(Information Storage Research Center, Shanghai Jiaotong University, Shanghai 200030)

Abstract This article introduced the training algorithm for the newest branch of statistic learning theory, SVM (Support Vector Machine), which can be classified into three categories: the first is the Decomposition Algorithm, whose delegate is SVMlight^[2], the second is sequence algorithm, the third is online training algorithm. All the three kinds of algorithms' advantages and disadvantages were analysed. And other algorithms and multi-class algorithms are introduced too. The future direction and application of SVM in pattern recognition and data mining, and so on were introduced.

Keywords support vector machine, training algorithm, statistical learning theory

1 引言 (Introduction)

统计学习理论是目前针对小样本统计估计和预测学习的最佳理论,它从理论上系统地研究了经验风险最小化原则成立的条件、有限样本下经验风险与期望风险的关系及如何利用这些理论找到新的学习原则和方法等问题,而支持向量机作为统计学习理论的实现方法,受到广大的研究者的注意.统计学习理论因为对有限样本情况下模式识别中的一些根本性问题进行了系统的理论研究,在很大程度上解决了模型选择与过学习问题、非线性和维数灾问题、局部极小点问题等,因此成为研究的热点^[24, 22]. [24] [43] 分别对支持向量机的原理作了简要的介绍,并和神经网络作了比较,本文则对支持向量机的训练算法进行详细讨论.

Vapnik^[6] 将支持向量机问题归结为一个二次型方程求解问题. Vapnik 通过对线性不可分的两类问

题的最优分类形式提出的.即使两类无错误地分开,并使两类的分类间隙 (Margin) 最大.问题的数学形式为:

$$\begin{aligned} h(w, a) &= \frac{1}{2} (w \cdot w) + C \sum_{i=1}^l a_i \\ \text{s. t. } y_i ((w \cdot x_i) + b) &\geq 1 - a_i, i = 1, \dots, l \\ a_i &\geq 0, i = 1, \dots, l. \end{aligned}$$

通过求其对偶问题,归结为一个二次函数极值问题:

$$\begin{aligned} W(T) &= \sum_{i=1}^l T_i - \frac{1}{2} \sum_{i,j=1}^l y_i y_j T_i T_j K(x_i, x_j) \\ \text{s. t. } 0 &\leq T_i \leq C, i = 1, \dots, l \\ \sum_{i=1}^l T_i y_i &= 0 \end{aligned}$$

分类判别函数为:

$$f(x) = \text{sign} \left[\sum_{i=1}^l T_i y_i K(x, x_i) + b \right]$$

对于这个二次规划问题,经典的解法有积极方

集法、对偶方法、内点算法^[28]等,但是当训练样本增多时,这些算法便面临着维数灾,或者由于内存的限制,而导致无法训练,无法应用 SVM 进行模式分类和函数估计。所以如何训练大训练集的 SVM 便成为 SVM 实际应用的瓶颈问题。

2 各种训练算法介绍及比较 (Introduction to different training algorithms and comparison between them)

Colin Campbell^[8]对 SVM 的训练算法作了一个综述,主要介绍了以 SVM^{light}^[2]为代表的分解算法、Platt 的 SMO^[9]和 Kerrthi 的近邻算法^[14],但没有详细介绍各算法的特点,并且没有包括算法的最新进展。以下对各种算法的特点进行详细介绍,并介绍几种新的 SVM 算法,如张学工^[26]的 CSVM, Scholkopf^[18,19]的 ν -SVM 分类器, J. A. K. Suykens^[31]提出的最小二乘法支持向量机 LS-SVM, Ming-Hsuan Yang^[40]提出的训练支持向量机的几何方法, SOR(Successive Overrelaxation)^[44]以及多类时的 SVM 算法。

Osuna^[1]针对 SVM 训练速度慢及时间空间复杂度大的问题,提出了分解算法,并将之应用于人脸检测中。主要思想是将训练样本分为工作集 B 和非工作集 N , B 中的样本个数为 q 个, q 远小于总样本个数。每次只针对工作集 B 中的 q 个样本训练,而固定 N 中的训练样本。算法的要点有三: 1) 应用有约束条件下二次规划极值点存在的最优条件 KTT 条件,推出本问题的约束条件,这也是终止条件。2) 工作集中训练样本的选择算法,应能保证分解算法能快速收敛,且计算费用最少。3) 改分解算法收敛的理论证明。Osuna 给出了一个简单的证明, Chang, C. - C.^[13]证明 Osuna 的证明不严密,并详尽地分析了分解算法的收敛过程及速度。该算法的关键在于选择一种最优的工作集选择算法, Osuna 的工作集的选择算法也并不是最优的,但是 Osuna 的这一工作应该说是开创性的,并为后来的研究奠定了基础。

Platt^[9]提出了 SMO(Sequential Minimal Optimization 序贯最小优化)来解决大训练样本的问题。并和 Chunking 算法进行了比较。该算法可以说是 Osuna^[1]分解算法的一个特例,工作集 B 中只有 2 个样本。其优点是针对 2 个样本的二次规划问题可以有解析解的形式,从而避免了多样本情形下的数值解不稳定及耗时问题,同时也不需要大的矩阵存

储空间,特别适合稀疏样本。其工作集的选择也别具特色,不是传统的最陡下降法,而是启发式。通过两个嵌套的循环来寻找待优化的样本变量。在外环中寻找违背 KKT 最优条件的样本,然后在内环中再选择另一个样本,完成一次优化。再循环,进行下一次优化,直到全部样本都满足最优条件。SMO 算法主要耗时在最优条件的判断上,所以应寻找最合理即计算代价最低的最优条件判别式,同时对常用的参数进行缓存。关于 SMO 算法收敛的理论分析在 [15] 中有详尽的证明。Platt 将 SMO 同投影共轭梯度法 (PCG) 进行了比较,对于 MNIST 训练集 SMO 比 PCG 快 1.7, 对于 UCI Adult 数据库, SMO 比 PCG 快 1500 倍。

对 Platt 的 SVM 算法, S. S. Keerthi 等通过对 SVM 算法的分析在 [10] 中提出了重大改进,即在判别最优条件时用两个阈值代替一个阈值,从而使算法更加合理,更快。其收敛证明在 [15] 中,并通过实际数据库的对比,证明确实比传统 SMO 快。同时也指出 SMO 算法应用于回归时类似的问题。R. Coolober etc.^[45]将考虑了上述改进的 SMO 算法应用于分类和回归问题,实现了比 SVMlight^[2]更强的软件包。

Joachims^[2]针对 [1] [9] 提出了具体的 SVM 实现算法,并在软件包 SVM^{light} 中实现了这一算法。主要贡献在于工作集的选择和实现的细节上。工作集的选择是沿着最陡方向 d , d 仅含 q 个非零成员,对应的 q 个变量将组成工作集 B ,

$$\begin{aligned} \min V(d) &= g(T^T)^T d \\ \text{s. t. } y^T d &= 0; \quad d_i \geq 0 \text{ for } i: T = 0; \\ d_i &\leq 0 \text{ for } i: T = C; \quad -1 \leq d \leq 1; \\ |\{d_i: d_i \neq 0\}| &= q; \end{aligned}$$

解上述最优问题即得工作集。在实现细节上, Joachims 对常用的参数进行缓存 (Cache), 并对 QP 问题进行 Shrinking, 从而使算法能较好地处理大规模的训练集问题。

Chih-Wei Hsu^[11]通过改变 SVM 的提法提出了一种类似 [1] 中的简单训练算法 BSV M。主要是针对 SVM 得到一种不同的数学提法,从而使算法简单易行。

Chih-Wei Hsu 和 Chih-Jen Lin 综合 S. S. Keerthi^[10]中的修改过的 SMO 和 SVMlight 中的工作集选择算法,用 C++ 实现一个库 LIBSVM,可以说是使用最方便的 SVM 训练工具。LIBSVM 供用户选择的参数少,用训练特大训练集时,还是使用灵

活的 SVM^{light} 或 SMO.

第二类是序贯分类法,基本思想是考虑训练样本序贯加入,同时考虑其对支持向量有何影响.[46] 基于感知机中的 Adatron 算法,对 Lagrange 系数和支持向量机.特点是训练样本序贯进入, Lagrange 系数的改变采用梯度法.[47] 提出了将另一种梯度序贯算法,特点是将 SVM 原问题中的偏置 b 也看作系数,从而使该梯度算法适合软间隔 (Soft margin) 和回归情形.

第三类是在线训练,即支持向量机的训练是在训练样本单个输入的情况下训练.和序贯分类法的区别是训练样本的总的个数是未知的.最典型的应用是系统的在线辨识.[48] 最早提出了 SVM 增量训练,但只是近似的增量,即每次只选一小批常规二次规划算法能处理的训练样本,然后只保留支持向量,抛弃非支持向量,和新进的样本混合进行训练,直到训练样本用完.经实验表明误差可以接受.[49] 提出了在线训练的精确解,即增加一个训练样本或减少一个样本对 Lagrange 系数和支持向量机的影响,实验表明算法是有效的,特别是减少一个样本时,对模型选择算法 LOO (Leave one out) 的形象解释.缺点是当样本无限增多时,还是必须抛弃一些样本,使其能够实用.

此外还有许多其它算法,如:张学工^[26]提出了 CSVM 算法,将每类训练样本集进行聚类分成若干子集,用子集中心组成新的训练样本集训练 SVM,将子集中心的系数赋给子集中每个样本.考察每个子集的每个样本的系数的改变对目标函数的影响.若一个子集所有样本对目标函数的影响都一样(改良与否)不同,则进一步划分,只到没有新的拆分为止.优点是提高了算法速度,同时减少训练数据中的野值对分类结果的影响;缺点是牺牲了解的稀疏性.

S. S. Keerthi^[14]等提出了修改了的算法——NPA 最近点算法.其基本思想是将 SVM 原问题的惩罚项由线性累加改为二次累加,从而使优化问题转化为两个凸集间的最大间隔,缺点是只能用于分类问题,不适用于函数估计问题.

Scholkopf^[18,19]提出了一个新的 SVM 分类器 ν -SVM,将待优化问题变为:

$$\begin{aligned} \min & \frac{1}{2} w^T w - \nu p + \frac{1}{l} \sum_{i=1}^l a_i \\ \text{s. t. } & y_i (w^T h(x_i) + b) \geq d - a_i, \\ & a_i \geq 0, i = 1, \dots, l, d \geq 0 \end{aligned}$$

其中, x_i 是训练样本向量, $\nu = 1$ or -1 表示对应

的类, x_i 被函数映射到一个高维空间.

C.-C. Chang and C.-J. Lin^[20]分析了 ν -SVM, 将它带有上下界约束和一个简单的等式约束的二次规划问题,从而可以用已有的 SVM 算法来解决.主要优点是 1) 使用一个参数 ν 来控制支持向量的个数及误差; 2) 没有 C , 避免了数值计算的麻烦; 3) ν 表示支持向量的下界与间隙误差的上界; 4) 和常规 SVM 相比, 算法一样. SVM 中的 C 的增加导致支持向量的减少, 而 ν -SVM 中 ν 的减少导致支持向量的减少.

Ming Hsuan Yang^[40]提出了训练支持向量机的几何方法.主要是利用了训练集中的集合信息, 提出了“卫向量” (Guard vector) 的概念, “卫向量”即为通过该向量能使输入空间线性可分的向量.所有的“支持向量”都是“卫向量”, 但反之不成立.当训练集合较大时, 可以先找出“卫向量”, 再以“卫向量”构成传统的 QP 问题求出“支持向量”.“卫向量”的求解是通过判断其对偶空间中的线性规划问题的可行性而不是求其解, 从而使问题大大简化.试验表明该算法求得的最优分类面和传统 QP 问题一样, 但速度要快 30 倍, 内存要求为传统的 $1/4$. 主要原因是“卫向量”只是“支持向量”的 20 倍左右.

J. A. K. Suykens^[31]提出了最小二乘法支持向量机 LS-SVM, 主要是优化函数的不同 (只有等式约束, 而无经典 SVM 的不等式约束), 从而推出不同的一系列的等式约束, 而不是二次规划问题.其目标函数是

$$\begin{aligned} \min_{w, b, e} J &= \frac{1}{2} w^T w + \nu \frac{1}{2} \sum_{k=1}^N e_k^2, \\ \text{s. t. } & y_k [w^T h(x_k) + b] = 1 - e_k, k = 1, \dots, N \end{aligned}$$

定义 Lagrange 函数

$$L = J - \sum_{k=1}^N \mathcal{T}_k \{y_k [w^T h(x_k) + b] - 1 + e_k\}$$

其中 \mathcal{T}_k 为 Lagrange 乘子, 根据 KTT 最优条件:

$$\begin{aligned} \frac{\partial L}{\partial w} = 0 &\rightarrow w = \sum_{k=1}^N \mathcal{T}_k y_k h(x_k), \\ \frac{\partial L}{\partial b} = 0 &\rightarrow w = \sum_{k=1}^N \mathcal{T}_k y_k = 0, \\ \frac{\partial L}{\partial e} = 0 &\rightarrow \mathcal{T}_k = \nu \cdot e_k, \end{aligned}$$

$$\frac{\partial L}{\partial e} = 0 \rightarrow y_k [w^T h(x_k) + b] - 1 + e_k = 0$$

对于 $k = 1, \dots, N$ 上式消去 w 和 e , 得到如下线性系统:

$$\begin{bmatrix} 0 & Y^T \\ Y & ZZ^T + V^{-1}I \end{bmatrix} \begin{bmatrix} \bar{b} \\ T \end{bmatrix} = \begin{bmatrix} \bar{0} \\ -\bar{1} \end{bmatrix}$$

其中, $Z = [h(x_1)^T y_1; \dots; h(x_N)^T y_N]$, $Y = [y_1; \dots; y_N]$,

$$\begin{aligned} \bar{1} &= [1; \dots; 1], \quad e = [e_1; \dots; e_N], \\ T &= [T_1; \dots; T_N] \end{aligned}$$

上述线性系统用最小二乘法即可解。

[44]提出了SQR方法,通过在原目标函数中加一项 b^2 ,从而对偶问题多了一项,而约束条件少了一项等式约束,变为边界约束条件下二次规划问题,适合迭代求解。同时应用矩阵分解技术,每次只需更新Lagrange乘子的一个分量,从而不需将所有样本载入内存,提高了收敛速度。

以上的算法,大部分都是针对两类问题,那么对于多类问题,SV M的算法有以下几种。标准算法^[7]是,对于 N 类问题构造 N 个两类分类器,第 i 个SV M用第 i 类中的训练样本作为正的训练样本,而将其它的样本作为负的训练样本。这个算法称为 1- a - r (1-against-rest)。最后的输出是两类分类器输出为最大的那一类(此时,两类分类器的判决函数不用取符号函数 sgn)。其缺点是它的推广误差无界。另外一个算法是由 Knerl^[17]提出,该算法在 N 类训练样本中构造所有可能的两类分类器,每类仅仅在 N 类中的 2 类训练样本上训练,结果共构造 $K = N(N-1)/2$ 个分类器,我们称该算法为 1- a -1(1-against-1)。组合这些两类分类器很自然地用到了投票法,得票最多(Max Wins)的类为新点所属的类。U. KreUel^[18]用该方法训练多类SV M取得了很好的结果。1- a -1算法的缺点是:1)如果单个两类分类器不规范化,则整个 N 类分类器将趋向于过学习;2)推广误差无界;3)分类器的数目 $N(N-1)/2$ 随类数 N 急剧增加,导致在决策时速度很慢。

J. Weston^[12]提出了两种新的多类SV M算法。其一是 qp-mc-sv 算法,它很自然地在构造决策函数时,同时考虑所有的类,将原始优化问题推广为:

$$h(w, a) = \frac{1}{2} \sum_{m=1}^k (w_m \cdot w_m) +$$

$$c \sum_{i=1}^l \sum_{m \neq y_i} a_i^m$$

$$\text{s. t.} \quad (w_i \cdot x_i) + b_{y_i} \geq (w_m \cdot x_i) + b_m + 2 - a_i^m$$

$$a_i^m \geq 0, \quad i = 1, \dots, l, \quad m \in \{1, \dots, k\} \setminus y_i$$

相应地决策函数变为:

$$f(x) = \arg \max_k [(w_i \cdot x) + b_i], i = 1, \dots, k$$

当 k 取 2 时,和两类分类问题等价。

其二是:线性规划的方法 lp-mcsv. 原始最优问题变为:

$$\begin{aligned} & \sum_{i=1}^l T_i + c \sum_{i=1}^l \sum_{j \neq y_i} a_j, \\ \text{s. t.} \quad & \sum_{m: y_m = y_i} T_m K(x_i, x_m) + b_{y_i} \\ & \geq \sum T_n K(x_i, x_n) + b_{y_j} + 2 - a_j \\ & T_i \geq 0, \quad a_j \geq 0, \quad i = 1, \dots, l, \\ & j \in \{1, \dots, k\} \setminus y_i \end{aligned}$$

决策函数为:

$$f(x) = \arg \max_n \left(\sum_{i: y_i = n} T_i K(x, x_i) + b_n \right)$$

以上两方法的缺点是计算量都比较大,优点是得到的决策分类面的支持向量机的数量均比常规方法少。

关于多类的 LS-SVM 分类器, J. A. K. Suykens^[29]使用了类似 J. Weston^[12]中的方法一。

Platt^[16]等提出了一个新的学习架构:决策导向的循环图(Decision Directed Acyclic Graph, DDAG),将多个两类分类器组合成多类分类器。对于 N 类问题,DDAG 含有 $N(N-1)/2$ 个分类器,每个分类器对应两类。其优点是推广误差只取决于类数 N ,和节点上的类间间隙(Margin),而与输入空间的维数无关,根据 DDAG 提出算法 DAGSVM, DDAG 的每个节点和一个 1- a -1 分类器相关,其速度显著比标准算法或取最大算法(Man Wins)快。

总之,以上的算法为SV M的实用起到了推动作用,尤其是有些算法还适用于回归和函数估计问题,如:SMO。

3 支持向量机的应用及发展方向 (Application and future direction of SVM)

支持向量机在模式识别^[32, 36](字符识别^[42]、文本自动分类^[25]、人脸检测^[1, 27, 39]、头的姿态识别^[41])、函数逼近^[23, 33]、数据挖掘^[37]和非线性系统控制^[50]中均有很好的应用。在[1]中用 2 阶多项式核,上界 C 设为 200,训练样本是 50000 个 19 乘 19 大小的人脸图像和 45000 个同样大小非脸图像。训练算法是 Osuna 提出的分解算法。训练出来的 SV M 分类器同 Sung^[38]的 NN 分类器进行了比较,检测速度有所提高,但错误率也比 Sung 的方法高。其应用主要在静止图像中有人脸。[39]使用了这个 SV M 来实时检测人脸,速度是每秒 25 帧。

本文综合介绍了现有的 SVM 训练算法,说明了各种的算法的思路和优缺点.今后 SVM 的训练算法的研究方向主要是确定不同的优化目标,根据 KKT 约束优化条件寻找大规模训练样本下的实用算法,应用于模式识别时的多类问题目前还没有特别好的算法.一种训练算法不但要有理论上的收敛的证明,同时也应该有实用的可行的算法实现,这样 SVM 的应用才会更广泛.目前在线训练是研究重点,因为这是应用于系统辨识和控制及信号处理中的关键.

参 考 文 献 (References)

- Osuna E, Freund R, Girosi F. Training support vector machines: An application to face detection. In Proceedings of CVPR 97, Puerto Rico, 1997
- Joachims T. Making large-Scale SVM Learning Practical. Advances in Kernel Methods - Support Vector Learning, Schölkopf B. et al. (ed.), MIT Press, 1999
- Joachims T. Transductive Inference for Text Classification using Support Vector Machines. Inter. Conf. on Machine Learning (ICML), 1999
- Joachims T. Text Categorization with Support Vector Machines: Learning with Many Relevant Features. Proc. of the European Conf. on Machine Learning, Springer, 1998
- Joachims T. Making large-Scale SVM Learning Practical. LS8-Report, 24, University Dortmund, LS VIII-Report, 1998
- Vapnik V N. The Nature of Statistical Learning Theory. Springer, 1995
- Vapnik V N. Statistical Learning Theory. New York, Wiley, 1998
- Colin C. Algorithmic Approaches to Training Support Vector Machines: A Survey. Proceedings of ESANN2000 (D-Facto Publications, Belgium), 2000, 27- 36
- John C P. Fast Training of Support Vector Machines using Sequential Minimal Optimization. In Schölkopf B. et al. (ed.), Advances in Kernel Methods-Support Vector Learning, Cambridge, MA, MIT Press, 1999, 185- 208
- Keerthi S S, et al. Improvements to Platt's SMO Algorithm for SVM Classifier Design, TR CD-99-14, Dept. of Mecha. and Prod. Engin. National Uni. of Singapore, 1999
- Chih W H, et al. A Simple Decomposition Method for Support Vector Machines. Technical report, National Taiwan University, 1999
- Weston J, Watkins C. Multiclass Support Vector Machines. TR CSDTR9804, Department of Computer Science Egham, Surrey TW20 0EX, England, 1998
- Chang C C, Hsu C W, Lin C J. The analysis of decomposition methods for support vector machines. In Workshop on Support Vector machines, IJCAI, 1999
- Keerthi S S, et al. A Fast Iterative Nearest Point Algorithm for Support Vector Machine Classifier Design. TR- ISL- 99-03 Dept. of CS and Auto. Indian Institute of Science Bangalore, India, 1999
- Keerthi S S. Convergence of a Generalized SMO Algorithm for SVM Classifier Design. TR CD- 00- 01 Control Division Dept. of Mecha. and Prod. Engineering National University of Singapore Singapore, 2000
- Platt J, et al. Large Margin DAGs for Multiclass Classification, in Advances in Neural Information Processing Systems 12, MIT Press, 2000, 547- 553
- Knerr S, et al. Single-layer learning revisited: A stepwise procedure for building and training a neural network. In Fogelman - Soulie et al. (ed.), Neurocomputing: Algorithms, Architectures and Applications, NATO ASI. Springer, 1990
- Křel U. Pairwise classification and support vector machines. In B. Schölkopf, ed. Advances in Kernel Methods: Support Vector Learning, pages MIT Press, Cambridge, MA, 1999, 255- 268
- Schölkopf B, et al. Estimating the support of a high-dimensional distribution. TR 99- 87, Microsoft Research, 1999
- Schölkopf B, et al. New support vector algorithms. Neural Computation, 2000, 12: 1083- 1121
- Chang C C, Lin C J. Some analysis on nu-support vector classification. Technical Report, National Taiwan University, 1999
- 边肇祺,张学工. 模式识别. 北京:清华大学出版社, 2000
- 田盛丰,黄厚宽. 基于支持向量机的数据库学习算法. 计算机研究与发展, 2000, 37(1): 17- 22
- 张学工. 关于统计学习理论与支持向量机. 自动化学报, 2000, 26(1): 32- 42
- 卢增祥,李衍达. 交互支持向量机学习算法及其应用. 清华大学学报, 1999, 39(7): 93- 97
- Zhang X. Using class-center vectors to build support vector machines. In Proceedings of NNSP 99, 1999
- Lu C Y, Yan P F, Zhang C S. Face recognition using support vector machine, In proc. Of ICNNB 98, Beijing, 1998, 652- 655
- 袁亚湘,孙文瑜. 最优化理论与方法. 北京:科学出版社, 1999
- Suykens J A K, Vandewalle J. Multiclass least squares support vector machines. In IJCNN '99 International Joint Conference on Neural Networks, Washington, DC, 1999
- Massimiliano P, Alessandro V. Object Recognition with Support Vector Machines, IEEE Trans. on PAMI, 1998, 20(6): 637- 646
- Suykens J A K, J Vandewalle. Least squares support vector machine classifiers. Neural Processing Letters, 1999, 9(3): 293- 300
- Burges C J C. A Tutorial on Support Vector Machines for Pattern Recognition. Knowledge Discovery and Data Mining, 1998, 2(2)
- Smola A J, Schölkopf B. A tutorial on support vector regression. NeuroCOLT TR NC- TR- 98- 030, Royal Holloway College, University of London, UK, 1998

34 Pontil M, Verri M. Support vector machines for 3- d object recognition. IEEE Trans. PAMI, 1998, **20** 637~ 646

35 Roobaert D. Improving the generalization of linear support vec- tor machines: an application to 3d object recognition with clut- tered background. In Proc. SVM Workshop at IJCAI 99, Stockholm, Sweden, 1999

36 Roobaert D, Hulle M M Van. View-based 3d object recognition with support vector machines. In IEEE Neural Networks for Signal Processing Workshop, 1999

37 Bradley P. Mathematical Programming Approaches to Machine Learning and Data Mining. PhD thesis, University of Wiscon- sin, Computer Sciences Department, Madison, WI, USA, TR- 98-11, 1998

38 Sung K, Poggio T. Example-base Learning for View-base Hu- man Face Detection. A. I. Memo 1521, MIT A. I. Lab., De- cember 1994

39 Kumar V, Poggio T. Learning-based Approach to Real Time Tracking and Analysis of Faces. In Proceedings of the Fourth International Conference on Face and Gesture Recognition, Grenoble, France, March, 2000

40 Yang M H, Ahuja N. A Geometric Approach to Train Support V ector Machines, In Proceedings of CV PR 2000, Hilton Head Island, 2000, 430~ 437

41 Huang J, *et al.* Face Pose Discrimination Using Support Vector Machines (SVM), Proc. of 14th International Conference on Pattem Recognition (ICPR 98), 1998

42 Scholkopf B, *et al.* Comparing Support Vector Machines with Gaussian Kernels to Radial Basis Function Classifiers. IEEE Trans. On Signal Processing, 1997, **45** 2758~ 2765

43 陶卿等. 一种新的机器学习算法: Support Vetector Machines,

模式识别与人工智能, 2000, **13**(3): 285~ 290

44 Olvi L M, David R M. Successive Overrelaxiation for Support Vector Machines. IEEE Trans. On Neural Networks. 1999, **10** (5): 1032~ 1037

45 Ronan C, Samy B. Support Vector Machines for Large-Scale Regression Problems, IDIAP- RR 00- 17. <http://www.idi- ap.ch>, 2000

46 Friess T T, *et al.* The kernel-adatron algorithm: A fast and simple learning procedure for support vector machines, ICM L98, 1998 188~ 196

47 Vijayakumar S, Wu S. Sequential Support Vector Classifiers and Regression, Proc. Inter. Con. on Soft Computing (SOCO 99), Genoa, Italy, 1999 610~ 619

48 Ahmed S N, *et al.* Incremental Learning with Support Vector Machines, (IJCAI99), Workshop on Support Vector Ma- chines, Stockholm, Sweden, August 2, 1999

49 Cauwenberghs G. and Poggio T. Incremental and decremental support vector Machine. Machine Learning. 2001, 13

50 Suykens J A K, *et al.* Optimal control by least squares support vector machines, Neural Networks , 2001, **14**(1): 23~ 25

作者简介

刘江华 (1975-),男,博士研究生.研究领域为机器视 觉、人工智能、模式识别.

程君实 (1939-),男,博士生导师,教授,日本早稻田大 学访问学者.研究领域为智能机器人控制、微机电系统、工业 过程控制.

陈佳品 (1960-),男,副教授.研究领域为智能机器人和 微机电系统.