

模式识别中的支持向量机方法

杜树新, 吴铁军

(工业控制技术国家重点实验室, 浙江大学 智能系统与决策研究所, 浙江 杭州, 310027)

摘 要: 针对模式识别问题, 描述了支持向量机的基本思想, 着重讨论了 ν -SVM、最小二乘 SVM、加权 SVM 和直接 SVM 等新的支持向量机方法, 用于降低训练时间和减少计算复杂性的海量样本数据训练算法分块法、分解法, 提高泛化能力的模型选择方法, 以及逐一鉴别法、一一区分法、M-ary 分类法、一次性求解等多类别分类方法。最后给出了污水生化处理过程运行状态监控的多类别分类实例。作为结构风险最小化准则的具体实现, 支持向量机具有全局最优性和较好的泛化能力。

关键词: 支持向量机; 模式识别; 泛化能力

中图分类号: TP181

文献标识码: A

文章编号: 1008-973X(2003)05-0521-07

Support vector machines for pattern recognition

DU Shu-xin, WU Tie-jun

(National Key Laboratory of Industrial Control Technology, Institute of Intelligent Systems and Decision Making, Zhejiang University, Hangzhou 310027, China)

Abstract An overview of the basic ideas underlying Support Vector Machine (SVM) for pattern recognition was given. New methods such as ν -SVM, LS-SVM, weighted SVM and direct SVM, training algorithms including chunking method and decomposition method for the sake of fast computational speed and ease of implementation, model selection approaches minimizing the generalization error, and multiclass classification methods such as one-against-the-rest method, one-against-one method, M-ary classification were concentrated. Finally, an example of multiclass classification for monitoring operation status of wastewater treatment processes was given. As a direct implementation of the structure risk minimization (SRM) inductive principle, SVM provides good performances such as global optimization and good generalization ability.

Key words support vector machines; pattern recognition; generalization ability.

从观测数据中学习归纳出系统运动规律, 并利用这些规律对未来数据或无法观测到的数据进行预测一直是智能系统研究的重点。传统学习方法中采用的经验风险最小化准则 (empirical risk minimization, ERM) 虽然可以使训练误差最小化, 但并不能最小化学习过程的泛化误差。ERM 不成功的例子就是神经网络的过学习问题。为此, Vapnik^[1,2]提出

了结构风险最小化准则 (structural risk minimization, SRM), 通俗地说就是通过对推广误差 (风险) 上界的最小化达到最大的泛化能力。Vapnik 提出的支持向量机 (support vector machine, SVM) 就是这种思想的具体实现。支持向量机的基本思想是在样本空间或特征空间, 构造出最优超平面, 使得超平面与不同类样本集之间的距离最大, 从而达到最大

收稿日期: 2002-09-30

作者简介: 杜树新 (1967-), 男, 浙江东阳人, 副研究员, 从事模式识别与智能系统理论及应用、工业过程计算机控制等方面研究。

E-mail: shxd@iipc.zju.edu.cn

©1994-2018 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

的泛化能力.

支持向量机结构简单,并且具有全局最优性和较好的泛化能力,自 20 世纪 90 年代中期提出以来得到了广泛的研究. 支持向量机方法是求解模式识别和函数估计问题的有效工具. 本文针对模式识别的问题求解,描述了支持向量机方法的基本思想,并着重讨论了支持向量机新方法. 降低训练时间和减少计算复杂性的训练算法. 提高泛化能力的模型选择方法以及多类别分类方法等方面的研究内容.

1 支持向量机基本方法

假定大小为 l 的训练样本集 $\{(x_i, y_i), i = 1, 2, \dots, l\}$, 由二类别组成, 如果 $x \in \mathbf{R}^N$ 属于第 1 类, 则标记为正 ($y = 1$), 如果属于第 2 类, 则标记为负 ($y = -1$). 学习的目标是构造一判别函数, 将测试数据尽可能正确地分类. 针对训练样本集为线性、非线性两种情况分别讨论.

1.1 线性情况

如果存在分类超平面

$$w^{\circ} \cdot x + b = 0, \tag{1}$$

使得

$$\left. \begin{aligned} w^{\circ} \cdot x_i + b \geq 1, y_i = 1, \\ w^{\circ} \cdot x_i + b \leq -1, y_i = -1, i = 1, 2, \dots, l, \end{aligned} \right\} \tag{2}$$

则称训练集是线性可分的, 其中 $w^{\circ} \cdot x$ 表示向量 $w \in \mathbf{R}^N$ 与 $x \in \mathbf{R}^N$ 的内积. 式 (1) 和式 (2) 中的 $w \in \mathbf{R}^N, b \in R$ 都进行了规范化, 使每类样本集中与分类超平面距离最近的数据点满足式 (2) 的等式要求. 对于式 (2), 可写成如下形式:

$$y_i(w^{\circ} \cdot x_i + b) \geq 1, i = 1, 2, \dots, l. \tag{3}$$

由统计学习理论知, 如果训练样本集没有被超平面错误分开, 并且距超平面最近的数据点与超平面之间的距离最大, 则该超平面为最优超平面 (如图 1 所示), 由此得到的判别函数

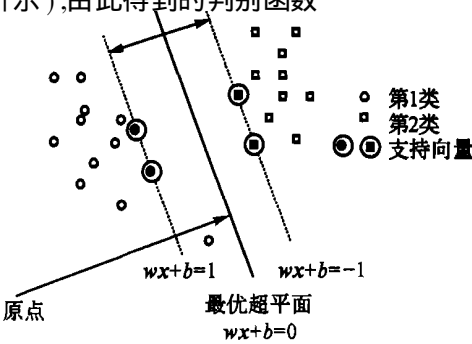


图 1 最优分类超平面

Fig. 1 Optimal separating hyperplanes

$$y(x) = \text{sign}(w^{\circ} \cdot x + b), \tag{4}$$

其泛化能力最优, 其中 $\text{sign}(\cdot)$ 为符号函数. 最优超平面的求解需要最大化 $2/\|w\|$, 即最小化

$\frac{1}{2}\|w\|^2$, 这样可转换成如下的二次规划问题:

$$\left. \begin{aligned} \min_{w,b} \frac{1}{2}\|w\|^2, \\ \text{s.t. } y_i(w^{\circ} \cdot x_i + b) \geq 1, i = 1, 2, \dots, l, \end{aligned} \right\} \tag{5}$$

训练样本集为线性不可分时, 需引入非负松弛变量 $a_i, i = 1, 2, \dots, l$, 分类超平面的最优化问题为

$$\left. \begin{aligned} \min_{w,b,a} \frac{1}{2}\|w\|^2 + C \sum_{i=1}^l a_i, \\ \text{s.t. } y_i(w^{\circ} \cdot x_i + b) \geq 1 - a_i, \\ a_i \geq 0, i = 1, 2, \dots, l. \end{aligned} \right\} \tag{6}$$

其中 C 为惩罚参数, C 越大表示对错误分类的惩罚越大. 采用拉格朗日乘子法求解这个具有线性约束的二次规划问题, 即

$$\left. \begin{aligned} \max_{a,b} \min_{w,b} \{L_P = \frac{1}{2}\|w\|^2 + C \sum_{i=1}^l a_i - \sum_{i=1}^l T_i [y_i(w^{\circ} \cdot x_i + b) - 1 + a_i] - \sum_{i=1}^l U_i a_i\}, \\ \text{s.t. } 0 \leq T_i, \\ 0 \leq U_i. \end{aligned} \right\} \tag{7}$$

其中 T_i, U_i 为拉格朗日乘子, 由此得到

$$\frac{\partial L_P}{\partial w} = 0 \rightarrow w = \sum_{i=1}^l T_i y_i x_i, \tag{8}$$

$$\frac{\partial L_P}{\partial b} = 0 \rightarrow \sum_{i=1}^l T_i y_i = 0, \tag{9}$$

$$\frac{\partial L_P}{\partial a} = 0 \rightarrow C - T_i - U_i = 0. \tag{10}$$

将式 (8)~(10) 代入式 (7), 得到对偶最优化问题

$$\left. \begin{aligned} \max \{L_D = \sum_{i=1}^l T_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l T_i T_j y_i y_j x_i^{\circ} \cdot x_j\}, \\ \text{s.t. } 0 \leq T_i \leq C, \\ \sum_{i=1}^l T_i y_i = 0. \end{aligned} \right\} \tag{11}$$

最优化求解得到的 T_i 可能是: ① $T_i = 0$; ② $0 < T_i < C$; ③ $T_i = C$. 后两者所对应的 x_i 为支持向量 (support vector, SV). 由式 (8) 知只有支持向量对 w 有贡献, 也就对最优超平面、判别函数有贡献, 所对应的学习方法称之为支持向量机. 在支持向量中, ③ 所对应的 x_i 称为边界支持向量 (boundary support vector, BSV), 实际上是错分的训练样本点; ② 所对应的 x_i 称为标准支持向量 (normal support vector, NSV). 根据 Karush-Kuhn-Tucker 条件^[1] (简称 KKT 条件) 知, 在最优点, 拉格朗日乘子与约束的积

为 0,即

$$\left. \begin{aligned} T_i [y_i (\mathbf{w}^\circ \cdot \mathbf{x}_i + b) - 1 + a_i] &= 0, \\ U a_i &= 0. \end{aligned} \right\} \quad (12)$$

对于标准支持向量 ($0 < T_i < C$),由式 (10)得到 $U_i > 0$,则由式 (12)得到 $a_i = 0$,因此,对于任一标准支持向量 \mathbf{x}_i ,满足

$$y_i (\mathbf{w}^\circ \cdot \mathbf{x}_i + b) = 1, \quad (13)$$

从而计算参数 b 为

$$b = y_i - \mathbf{w}^\circ \cdot \mathbf{x}_i = y_i - \sum_{\mathbf{x}_j \in \text{NSV}} T_j y_j \mathbf{x}_j \cdot \mathbf{x}_i, \quad \mathbf{x}_i \in \text{NSV}. \quad (14)$$

为了使计算可靠,对所有标准支持向量分别计算 b 的值,然后求平均,即

$$b = \frac{1}{N_{\text{NSV}}} \sum_{\mathbf{x}_i \in \text{NSV}} (y_i - \sum_{\mathbf{x}_j \in \text{NSV}} T_j y_j \mathbf{x}_j \cdot \mathbf{x}_i). \quad (15)$$

式中: N_{NSV} 为标准支持向量数.

由式 (13)知,支持向量机就是满足式 (3)等式要求的样本数据,支持向量如图 1 所示.

1.2 非线性情况

训练集为非线性时,通过一非线性函数 $\phi(\cdot)$ 将训练集数据 \mathbf{x} 映射到一个高维线性特征空间,在这个维数可能为无穷大的线性空间中构造最优分类超平面,并得到分类器的判别函数.因此,在非线性情况下,分类超平面为

$$\mathbf{w}^\circ \cdot \phi(\mathbf{x}) + b = 0, \quad (16)$$

判别函数为

$$y(\mathbf{x}) = \text{sign} [\mathbf{w}^\circ \cdot \phi(\mathbf{x}) + b]; \quad (17)$$

最优分类超平面问题描述为

$$\left. \begin{aligned} \min_{\mathbf{w}, b} & \frac{1}{2} \|\mathbf{w}\|^2 + \sum_{i=1}^l a_i, \\ \text{s.t.} & y_i (\mathbf{w}^\circ \cdot \phi(\mathbf{x}_i) + b) \geq 1 - a_i, \\ & a_i \geq 0, i = 1, 2, \dots, l. \end{aligned} \right\} \quad (18)$$

类似于 1.1,得到对偶最优优化问题

$$\left. \begin{aligned} \max_{\alpha} & \left\{ L_D = \sum_{i=1}^l T_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l T_i T_j y_i y_j \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) = \right. \\ & \left. \sum_{i=1}^l T_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l T_i T_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \right\}, \\ \text{s.t.} & 0 \leq T_i \leq C, \\ & \sum_{i=1}^l T_i y_i = 0. \end{aligned} \right\} \quad (19)$$

式中: $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)$ 称为核函数. 判别函数为

$$y(\mathbf{x}) = \text{sign} \left[\sum_{\mathbf{x}_j \in \text{SV}} T_j y_j K(\mathbf{x}_j, \mathbf{x}) + b \right], \quad (20)$$

其中阈值 b 为

$$b = \frac{1}{N_{\text{NSV}}} \sum_{\mathbf{x}_i \in \text{NSV}} \left[y_i - \sum_{\mathbf{x}_j \in \text{SV}} T_j y_j K(\mathbf{x}_j \cdot \mathbf{x}_i) \right]. \quad (21)$$

由式 (19)~ (21)知,尽管通过非线性函数将样本数据映射到具有高维甚至为无穷维的特征空间,并在特征空间中构造最优分类超平面,但在求解最优优化问题和计算判别函数时并不需要显式计算该非线性函数,而只需计算核函数,从而避免特征空间维数灾难问题.核函数的选择必须满足 Mercer 条件^[1].常见的核函数有线性函数 $K(\mathbf{x}_i, \mathbf{x}) = \mathbf{x}_i \cdot \mathbf{x}$ 、多项式函数 $K(\mathbf{x}_i, \mathbf{x}) = (\mathbf{x}_i \cdot \mathbf{x} + 1)^d$ 、径向基函数 $K(\mathbf{x}_i, \mathbf{x}) = \exp(-\|\mathbf{x} - \mathbf{x}_i\|^2 / \sigma^2)$ 、多层感知器函数 $K(\mathbf{x}_i, \mathbf{x}) = \tanh(k \mathbf{x}_i \cdot \mathbf{x} + \theta)$.

1.3 关于支持向量机方法的说明

1)由统计学习理论^[1,2]知,对于分类器 f ,实际风险 $R_{\text{str}}(f)$ 和经验风险 $R_{\text{emp}}(f)$ 之间以至少有 $1-Z$ 的概率满足 ($Z > 0$)

$$R_{\text{str}}(f) \leq R_{\text{emp}}(f) + \sqrt{\frac{h \ln(2/lh + 1) - \ln(l/4)}{l}}, \quad (22)$$

其中 h 为 VC 维,对于线性分类器,满足^[2]

$$h \leq \|\mathbf{w}\|^2 r^2 + 1. \quad (23)$$

其中 r 为包络训练数据的最小球半径.机器学习过程不仅要使经验风险最小,还要使 VC 维尽量小,对未来样本才会有较好的泛化能力,这是结构风险最小化准则的基本思想.式 (5)中的约束条件约束了 \mathbf{w} 、 b ,使得经验误差为 0,同时最小化 $\|\mathbf{w}\|^2$ 使 VC 维最小,因此式 (5)的最优化体现了结构风险最小化准则,具有较好的泛化能力.

2)支持向量机方法本质上是一个非负的二次型优化问题,在理论上可以得到全局最优的解析解,因此 SVM 不存在局部最优化问题.

3)SVM 的重要特征之一是解的稀疏性,即多数最优值 T_i 为 0,只有少量的 T_i 不为 0,也就是说只需少量样本(支持向量)就可构成最优分类器,这样有用的样本数据大大压缩.

4)对于式 (12)的 KKT 条件,也可以写为(非线性情况)

$$\left. \begin{aligned} y_i (\mathbf{w}^\circ \cdot \phi(\mathbf{x}_i) + b) &\geq 1, T_i = 0, \\ y_i (\mathbf{w}^\circ \cdot \phi(\mathbf{x}_i) + b) &= 1, 0 < T_i < C, \\ y_i (\mathbf{w}^\circ \cdot \phi(\mathbf{x}_i) + b) &\leq 1, T_i = C. \end{aligned} \right\} \quad (24)$$

由于 KKT 条件是充要条件,利用上式可判别 T 是否为最优.

2 支持向量机方法的主要研究内容

2.1 支持向量机新方法研究

2.1.1 ν -SVM 在 SVM 中,通过参数 C 实行对错误分类的惩罚,但如何选择 C 并没有直观的解释.在 Scholkoph 等人^[3]提出的 ν -SVM 方法中,优化问题为

$$\left. \begin{aligned} \min_{w,b,\xi} & \frac{1}{2} \|w\|^2 + \nu d + \frac{1}{l} \sum_{i=1}^l a_i, \\ \text{s.t.} & y_i(w \cdot \phi(x_i) + b) \leq d - a_i, \\ & a_i \geq 0, \\ & d > 0. \end{aligned} \right\} \quad (25)$$

类似于 1.2,得到对偶形式为

$$\left. \begin{aligned} \max_{\alpha} & \left\{ L_D = -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l T_i T_j y_i y_j K(x_i, x_j) \right\}, \\ \text{s.t.} & 0 \leq T_i \leq 1/l, \\ & \sum_{i=1}^l T_i y_i = 0, \\ & \sum_{i=1}^l T_i \geq \nu. \end{aligned} \right\} \quad (26)$$

由 KKT 条件知,在最优点满足

$$\sum_{i=1}^l T_i = \nu. \quad (27)$$

对于边界支持向量, $T_i = 1/l$, 因此对于 N_{BSV} 个边界

支持向量有 $(N_{BSV}/l) \leq \sum_{i=1}^l T_i = \nu$; 而对于支持向量, $T_i \leq 1/l$, 因此对于 N_{SV} 个支持向量有 $\sum_{i=1}^l T_i \leq N_{SV}/l$, 即 $N_{SV}/l \geq \nu$, 因此有

$$N_{BSV}/l \leq \nu \leq N_{SV}/l. \quad (28)$$

由式 (28) 知, ν -SVM 中参数 ν 的物理意义非常明确, $l \nu$ 表示 BSV 数量的上限和 SV 数量的下限.

2.1.2 LS-SVM 在 Suykens 等人^[4]提出的最小二乘支持向量机 (LS-SVM) 中, 优化指标采用了平方项, 从而将不等式约束转变成等式约束. 最优化为问题为

$$\left. \begin{aligned} \min_{w,b,\xi} & \frac{1}{2} \|w\|^2 + \frac{1}{2} \sum_{i=1}^l \xi_i^2, \\ \text{s.t.} & y_i(w \cdot \phi(x_i) + b) = 1 - \xi_i. \end{aligned} \right\} \quad (29)$$

可得到线性方程组

$$\begin{bmatrix} 0 & y^T \\ y & Q + r^{-1}I \end{bmatrix}_{(l+1) \times (l+1)} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ e \end{bmatrix}. \quad (30)$$

式中: $e \in \mathbb{R}$ 是元素为 1 的向量; $I \in \mathbb{R}^{l \times l}$ 为单位阵; $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_l]^T \in \mathbb{R}^l$; $y = [y_1, y_2, \dots, y_l]^T \in \mathbb{R}^l$; Q

$= [q_{ij}]_{l \times l}$; $q_{ij} = y_i y_j K(x_i, x_j)$. 在 LS-SVM, 将二次规划问题转变成线性方程组的求解, 简化了计算复杂性. 但由于每个样本数据对分类器都有贡献, LS-SVM 失去了稀疏性这一 SVM 的优点.

2.1.3 W-SVM 在第 1 节描述的支持向量机中, 引入惩罚参数 C 实行对错误分类的惩罚. 在实际应用中, 某些重要样本正确分类的要求高, 而某些样本正确分类的要求低, 因此, 在优化问题描述中, 对每个采样点数据采用不同的惩罚系数, 以得到更准确的分类, 这种支持向量机称之为加权支持向量机 (Weighted SVM, W-SVM). 另外, 不同类别的样本数量差异比较大时, 存在着分类结果偏向于多样本数类别的问题^[5], 这类问题的解决方法本质上也是 W-SVM. W-SVM 的最优化问题描述为

$$\left. \begin{aligned} \min_{w,b,\xi} & \frac{1}{2} \|w\|^2 + \sum_{i=1}^l s_i a_i, \\ \text{s.t.} & y_i(w \cdot \phi(x_i) + b) \geq 1 - a_i, \\ & a_i \geq 0, i = 1, 2, \dots, l. \end{aligned} \right\} \quad (31)$$

其中 s_i 为加权系数. 对偶最优化问题为

$$\left. \begin{aligned} \max_{\alpha} & \left\{ L_D = \sum_{i=1}^l T_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l T_i T_j y_i y_j K(x_i, x_j) \right\}, \\ \text{s.t.} & 0 \leq T_i \leq C s_i, \\ & \sum_{i=1}^l T_i y_i = 0. \end{aligned} \right\} \quad (32)$$

W-SVM 需要合理选择 s_i , 文献 [6] 认为在样本集中第 1 个样本的重要性最低, s_i 设定为小于 1 的值, 最后一个样本的重要性最高, s_l 设定为 1, 采用线性插值方法得到其他样本点的加权系数. 文献 [5] 针对 ν -SVM 提出了 W-SVM 方法, s_i 只有 2 个值 s_+ 和 s_- , s_+ 表示正类对应的加权系数, s_- 表示负类对应的加权系数, 根据二类别中每类样本的数量来确定 s_+ 和 s_- 的值, 即 $s_+ : s_- = l_- : l_+$, 其中 l_+ 、 l_- 分别表示正类、负类的样本数. 由此得到的 SVM 避免了常规 ν -SVM 在各类别数量不均衡时引起的分类结果会偏向于数量较多一类的问题. Suykens 等人^[7]针对 LS-SVM 提出了 W-SVM.

2.1.4 DirectSVM 前面所讨论的支持向量机, 都需要采用二次规划或线性方程组求解最优化问题, 而 Roobaert^[8]提出的直接支持向量机 (DirectSVM) 采用了启发式搜索方法在所有训练样本集中搜索支持向量, 从而避免了二次规划最优求解. 启发规则主要包括: ① 距离最近的 2 个不同类别的数据点有可能是支持向量, ② 最大偏离最优超平面的数据点, 也就是错误分类最严重的数据点有可能是支持向

量. 根据这两条启发式规则搜索支持向量及最优超平面. 该方法具有直观、求解速度快等优点.

2.2 海量样本数据的优化算法研究

采用支持向量机方法求解最优分类问题, 其本质上是一个二次规划问题. 对于 l 个样本的数据集, 这个二次规划包括了 l 个优化变量、 $2l$ 个线性不等式约束和 1 个等式约束, 同时涉及 K/l 维核函数矩阵的计算及矩阵与向量的相乘运算. 样本数据较小时, 可以通过解析法、数值优化算法求解, 但对于海量样本数据 (样本数在 $10^5 \sim 10^6$ 以上), 常规的数值优化算法及软件已无法实现二次规划问题的求解. 运行时间和计算内存是海量样本数据求解 SVM 的主要瓶颈. 针对海量样本数据如何减少二次规划求解过程的计算时间和内存一直是 SVM 的研究热点, 目前主要有以下两种方法:

2.2.1 分块法 由支持向量机方法得到的判别函数只与支持向量有关, 与其他样本数据无关, 也就是说, 如果只取支持向量作为训练样本, 得到的判别函数与所有样本作为训练样本得到的判别函数是一致的. 分块法 (chunking method) 就是基于这种思想而提出的用于求解海量样本数据的优化算法, 即将海量样本数据集分成若干个小规模的样本集, 按顺序逐个对各样本子集进行训练学习. 在对每个样本子集学习时, 只需要根据上个样本子集得到的支持向量以及当前的样本子集进行新的最优化计算. 后来提出的增量学习方法 (incremental learning)^[9] 本质上是分块法.

分块法求解规模随着支持向量数量的增加而增加, 因此在支持向量数目非常大时, 优化计算仍难以实现.

2.2.2 分解法 分解法 (decomposition method) 也是将大规模的二次规划问题转化成一系列小规模的二次规划求解. 分解法是选择 q 个 \mathbb{T} 作为优化变量 (q 为固定), 而其他 \mathbb{T} 的值固定不变, 因此分解法的子问题求解不像分块法那样求解规模会随着支持向量数量的增加而增加, 而是固定不变的. 分解的基本思想是将样本数据的序号集 $\{1, 2, \dots, l\}$ 分为工作集 B 和非工作集 N , 工作集 B 的大小为 q , 这样将大规模的二次规划问题转化成只有 q 个优化变量、 $2q$ 个线性不等式约束、1 个等式约束的小规模二次规划问题.

分解法的关键问题在每次迭代过程中如何选择工作集 B 以及算法的收敛性. Osuna 等人^[10] 根据 KKT 条件背离程度选择工作集, 而 Joachims^[11]、Laskey^[12]、Hsu 等人^[13] 采用可行方向法选择工作

集, 这些分解方法的收敛性已得到了证明^[14].

Platt^[15] 提出的顺序优化算法 (sequential minimal optimization, SMO) 将工作集 B 大小 q 限定为 2, 这样在进行小规模二次规划求解时用解析法直接计算, 而不需要采用数值优化算法计算. Keerthi 等人^[16] 证明了 SMO 算法的收敛性.

2.3 支持向量机中模型选择研究

采用支持向量机求解模式识别问题需要选择一核函数. 尽管只要满足 Merce 条件的函数在理论上都可选为核函数, 但不同的核函数, 其分类器的性能完全不同. 因此, 针对某一特定问题, 选择什么样的核函数是至关重要的. 另一方面, 即使选择了某一类核函数, 其相应的参数 (如多项式的阶次、径向基函数的宽度) 也有如何选择的问题. 另外, 支持向量机中二次规划参数如 C 参数、 ν 参数 (ν -SVM)、 r 参数 (LS-SVM) 对分类器的泛化能力也有影响. 核函数类别选择、核函数参数选择、二次规划参数选择统称为模型选择. 尽管模型选择方面的研究成果不多, 但作为支持向量机的重要研究内容已日益受到研究者的重视.

Ayat 等人^[17] 认为核函数在 0 点附近应该有较快的下降, 在无穷远点仍有适当的下降而不应为 0, 由此构造了一个新的核函数 KMOD, 实验表明 KMOD 的推广能力高于径向基核函数. Amari 等人^[18] 通过增加不同模式间的边界距离来增加分类器的可分性, 针对径向基函数, 给出了动态修正径向基函数的算法. Chapelle 等人^[19] 针对不同模型参数得到的验证误差, 拟合出一条关于模型参数与验证误差的曲线. 根据拟合曲线的梯度修正模型参数. Baesens 等人^[20] 针对 LS-SVM 方法, 采用粗细网格搜索算法优化模型参数.

2.4 支持向量机的多类别分类方法研究

支持向量机方法是针对二类别的分类而提出的, 如何将二类别分类方法扩展到多类别分类是支持向量机研究的重要内容之一. 假定多类别分类问题有 k 个类别 $S = \{1, 2, \dots, k\}$, 训练样本为 $\{(x_i, y_i), i = 1, 2, \dots, l\}$, 其中 $y_i \in S$. 主要有以下 4 种方法实现 SVM 的多类别分类.

2.4.1 逐一鉴别方法 逐一鉴别方法 (one-against-the-rest method)^[21] 是构造 k 个 SVM 子分类器. 在构造第 j 个 SVM 子分类器时, 将属于第 j 类别的样本数据标记为正类, 不属于 j 类别的样本数据标记为负类. 测试时, 对测试数据分别计算各个子分类器的判别函数值, 并选取判别函数值最大所对应的类别为测试数据的类别.

2.4.2 一一区分法 一一区分法 (one-against-one method)^[21]分别选取 2 个不同类别构成一个 SVM 子分类器,这样共有 $k(k-1)/2$ 个 SVM 子分类器.在构造类别 i 和类别 j 的 SVM 子分类器时,样本数据集选取属于类别 i 类别 j 的样本数据作为训练样本数据,并将属于类别 i 的数据标记为正,将属于类别 j 的数据标记为负.测试时,将测试数据对 $k(k-1)/2$ 个 SVM 子分类器分别进行测试,并累计各类别的得分,选择得分最高者所对应的类别为测试数据的类别.一一区分法存在不可区分区域.

2.4.3 M-ary分类方法 在 Sebal d 等人^[22]提出的 M-ary 分类方法中,充分运用了 SVM 的二类别分类特点,将多类别分类的各个类别重新组合,构成 $\lceil \log_2^k \rceil$ 个 SVM 子分类器.以 4 个类别 {1, 2, 3, 4} 为例,构造 $\lceil \log_2^4 \rceil = 2$ 个 SVM 子分类器.对于第 1 个 SVM 子分类器,类别 2 4 所对应的样本数据全标记为正,类别 1 3 所对应的样本数据全标记为负.对于第 2 个 SVM 子分类器,类别 2 3 所对应的样本数据标记为正,类别 1 4 所对应的样本数据标记为负.由第 1 节的方法分别求得这 2 个 SVM 子分类器.测试时,根据这 2 个子分类器的结果可得到测试样本的类别.例如针对某一测试数据 x ,第 1 2 SVM 子分类器结果分别为 + 1 - 1,则由第 1 个子分类器可知, x 属于类别 2 或类别 4,由第 2 个子分类器可知, x 属于类别 1 或类别 4,因此 x 属于类别 4.

M-ary 分类方法很巧妙地将多类别分类问题转变为数量较少的二类别 SVM 分类器,使训练计算量、测试计算量大大减少,是一种很好的多类别分类方法.

2.4.4 一次性求解方法 上面描述的多类别分类方法都是通过一系列二类别子分类器的求解得到,一些 SVM 研究者提出一次性求解多类别分类问题的方法^[2],其基本思想类似于逐一鉴别法需要构造 k 个二类别 SVM 分类器,但不同的是一次性求解方法是由 1 个优化问题同时求解 k 个 SVM 分类器.

3 实例: 污水处理过程状态监控

污水处理过程是一个包含许多复杂处理工艺的生化过程,涉及到众多的机械设备、电气设备、自动化仪表等,同时由于进水流量、成分及浓度剧烈波动,污水处理过程处于非稳定状态,因此如何监控污水处理过程,即如何通过各处理阶段的可测信息自动监控处理过程的运行状态,是污水处理厂运行与管理的关键.对污水处理过程运行状态的监控本质

上是多类别分类问题.这一节,以污水处理过程运行状态的监控作为支持向量机应用的一个例子简单地加以描述.实验数据来自 UCI 机器学习数据库 (<http://www.ics.uci.edu/~mllearn/MLRrepository.htm>),这些数据都是在某一污水处理厂实测得到的.

表 1 6 个类别的样本组成

| Tab. 1 Sample of each class for 6-class classification | | | | | | |
|--|-----|-----|----|---|---|---|
| 类别 | 1 | 2 | 3 | 4 | 5 | 6 |
| 样本数 | 332 | 116 | 65 | 7 | 3 | 4 |

整个数据集有 527 个样本,每个样本维数为 38 (即 38 个测量变量),共有 6 个类别,表 1 给出了每个类别的样本数,表中类别 1 为正常状态,类别 2 为性能超过平均值的正常状态,类别 3 为低进水流量的正常状态,类别 4 为二沉池故障,类别 5 为暴雨引起的非正常状态,类别 6 为固体浓度过载引起的非正常状态.全部样本分成两部分,60% 用于训练,40% 用于测试.采用支持向量机分类时,核函数选为径向基函数, $K(x_i, x) = \exp(-\|x - x_i\|^2 / \epsilon^2)$,采用一一区分法实行多类别分类,表 2 给出了采用不同支持向量机方法得到的分类精度,为加以区别,标准支持向量机方法记为 C-SVM.

表 2 不同 SVM 方法的精度比较

| Tab. 2 Comparison of accuracy with different SVM | | | |
|--|---------|--------------|---------|
| 样本 | C-SVM % | ν -SVM % | W-SVM % |
| 测试数据 | 75 | 77.63 | 76.32 |
| 训练数据 | 99.12 | 100 | 94.73 |
| 总体数据 | 89.47 | 91.05 | 87.37 |

4 结 语

本文针对模式识别问题描述了支持向量机的基本方法及当前支持向量机研究中的主要内容,侧重讨论了支持向量机的新方法.降低训练时间和减少计算复杂性的训练算法,提高泛化能力的模型选择方法及用于多类别的 SVM 分类方法.作为新理论和新技术,支持向量机还有很多尚未解决或尚未充分解决的问题,尤其在应用方面的研究.

参考文献 (References):

[1] VAPNIK V N. The nature of statistical learning [M]. Berlin: Springer, 1995.
[2] VAPNIK V N. Statistical learning theory [M]. New York: John Wiley & Sons, 1998.
[3] SCHOLKOPH B, SMOLA A J, BARTLETT P L. New support vector algorithms [J]. Neural Computation, 1994-2018 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

- 2000, 12(5): 1207– 1245.
- [4] SUYKENS J A K, VANDEWALE J. Least squares support vector machine classifiers [J]. **Neural Processing Letters**, 1999, 9(3): 293– 300.
- [5] CHEW H-G, BOGNER R E, LIM C-C. Dual v -support vector machine with error rate and training size beasing [A]. **Proceedings of 2001 IEEE Int Conf on Acoustics, Speech, and Signal Processing** [C]. Salt Lake City, USA: IEEE, 2001. 1269– 1272.
- [6] LIN C-F, WANG S-D. Fuzzy support vector machines [J]. **IEEE Trans on Neural Networks**, 2002, 13(2): 464– 471.
- [7] SUYKENS J A K, BRANBANTER J D, LUKAS L, *et al.* Weighted least squares support vector machines robustness and spare approximation [J]. **Neurocomputing**, 2002, 48(1): 85– 105.
- [8] ROOBAERT D. Direct SVM: A fast and simple support vector machine perception [A]. **Proceedings of IEEE Signal Processing Society Workshop** [C]. Sydney, Australia: IEEE, 2000. 356– 365.
- [9] DOMENICONI C, GUNOPULOS D. Incremental support vector machine construction [A]. **Proceedings of IEEE Int Conf on Data Mining** [C]. San Jose, USA: IEEE, 2001. 589– 592.
- [10] OSUNA E, FREUND R, GIROSI F. An improved training algorithm for support vector machine [A]. **Proceedings of 1997 IEEE Workshop on Neural Networks for Signal Processing** [C]. Anelea Island, FL: IEEE, 1997. 276– 285.
- [11] JOACHIMS T. Making large-scale SVM learning practical [A]. SCHOLKOPH B, **Advances in Kernel Method-Support Vector Learning** [C]. Cambridge, MA: MIT Press, 1999. 169– 184.
- [12] LASKOV P. Feasible direction decomposition algorithms for training support vector machines [J]. **Machine Learning**, 2002, 46(1): 315– 349.
- [13] HSU C-W, LIN C-J. A simple decomposition method for support vector machines [J]. **Machine Learning**, 2002, 46(1): 291– 314.
- [14] LIN C-J. On the convergence of the decomposition method for support vector machines [J]. **IEEE Trans on Neural Networks**, 2001, 12(6): 1288– 1298.
- [15] PLATT J C. Fast training of support vector machines using sequential minimal optimization [A]. SCHOLKOPH B, *et al.* eds. **Advances in Kernel Method-Support Vector Learning** [C]. Cambridge, MA: MIT Press, 1999. 185– 208.
- [16] KEERTHI S S, GILBERT E G. Convergence of a generalized SMO algorithm for SVM classifier design [J]. **Machine Learning**, 2002, 46(1): 351– 360.
- [17] AYAT N E, CHERIET M, REMAKI L, *et al.* KMOD—a new support vector machine kernel with moderate decreasing for pattern recognition, application to digit image recognition [A]. **Proceedings of 6th Int Conf on Document Analysis and Recognition** [C]. Seattle, USA: IEEE, 2001. 1215– 1211.
- [18] AMARI S-I, WU S. An information-geometrical method for improving the performance of support vector machine classifier [A]. **Proceedings of 9th Int conf on Artificial neural networks** [C]. Edinburgh, UK: IEEE, 1999. 85– 90.
- [19] CHAPELLE O, VAPNIK V, BACSQUEST O, *et al.* Choosing multiple parameters for support vector machines [J]. **Machine Learning**, 2002, 46(1): 131– 159.
- [20] BAESENS B, VIAENE S, GESTEL T V, *et al.* An empirical assessment for kernel type performance for least squares support vector machine classifiers [A]. **Proceedings of 4th Int Conf on Knowledge-based Intelligent Engineering Systems and Allied Technologies** [C]. Brighton, UK: IEEE, 2000. 313– 316.
- [21] HSU C-W, LIN C-J. A comparison of methods for multiclass support vector machines [J]. **IEEE Trans on Neural Networks**, 2002, 13(2): 415– 425.
- [22] SEBALD D J, BUCHLEW J A. Support vector machines and the multiple hypothesis test problem [J]. **IEEE Trans on Signal Processing**, 2001, 49(11): 2865– 2872.