

多分类孪生支持向量机研究进展*

丁世飞^{1,2}, 张健¹, 张谢锴¹, 安悦瑄¹



¹(中国矿业大学 计算机科学与技术学院, 江苏 徐州 221116)

²(中国科学院 计算技术研究所 智能信息处理重点实验室, 北京 100190)

通讯作者: 安悦瑄, E-mail: 1187360089@qq.com

摘要: 孪生支持向量机因其简单的模型、快速的训练速度和优秀的性能而受到广泛关注. 该算法最初是为解决二分类问题而提出的, 不能直接用于解决现实生活中普遍存在的多分类问题. 近年来, 学者们致力于将二分类孪生支持向量机扩展为多分类方法, 并提出了多种多分类孪生支持向量机. 多分类孪生支持向量机的研究已经取得了一定的进展. 主要工作是回顾多分类孪生支持向量机的发展, 对多分类孪生支持向量机进行合理归类, 分析各个类型的多分类孪生支持向量机的理论和几何意义. 以多分类孪生支持向量机的子分类器组织结构为依据, 将多分类孪生支持向量机分为: 基于“一对多”策略的多分类孪生支持向量机、基于“一对一”策略的多分类孪生支持向量机、基于“一对一对余”策略的多分类孪生支持向量机、基于二叉树结构的多分类孪生支持向量机和基于“多对一”策略的多分类孪生支持向量机. 基于有向无环图的多分类孪生支持向量机训练过程与基于“一对一”策略的多分类孪生支持向量机类似, 但其决策方式有其特殊的优缺点, 因此将其也独立为一类. 分析和总结了这 6 种类型的多分类孪生支持向量机的算法思想、理论基础. 此外, 还通过实验对比了分类性能. 为各种多分类孪生支持向量机之间建立了联系比较, 使得初学者能够快速理解不同多分类孪生支持向量机之间的本质区别, 也对实际应用中选取合适的多分类孪生支持向量机起到一定的指导作用.

关键词: 多分类; 孪生支持向量机; 多生支持向量机; 支持向量机

中图法分类号: TP18

中文引用格式: 丁世飞, 张健, 张谢锴, 安悦瑄. 多分类孪生支持向量机研究进展. 软件学报, 2018, 29(1): 89-108. <http://www.jos.org.cn/1000-9825/5319.htm>

英文引用格式: Ding SF, Zhang J, Zhang XK, An YX. Survey on multi class twin support vector machines. Ruan Jian Xue Bao/ Journal of Software, 2018, 29(1): 89-108 (in Chinese). <http://www.jos.org.cn/1000-9825/5319.htm>

Survey on Multi Class Twin Support Vector Machines

DING Shi-Fei^{1,2}, ZHANG Jian¹, ZHANG Xie-Kai¹, AN Yue-Xuan¹

¹(School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China)

²(Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100190, China)

Abstract: Twin support vector machines have drawn extensive attention for their simple model, high training speed and good performance. The initial twin support vector machine is designed for binary classification. However, multi class classification problems are also common in practice. In recent years, researchers have devoted themselves to the study of multi class twin support vector machines. Various multi class twin support vector machines have been proposed. The study of multi class twin support vector machines has made

* 基金项目: 国家自然科学基金(61672522, 61379101); 国家重点基础研究发展计划(973)(2013CB329502)

Foundation item: National Natural Science Foundation of China (61672522, 61379101); National Key Basic Research Program of China (973) (2013CB329502)

收稿时间: 2017-01-19; 修改时间: 2017-02-25, 2017-04-16; 采用时间: 2017-06-16; jos 在线出版时间: 2017-07-12

CNKI 网络优先出版: 2017-07-12 16:16:57, <http://kns.cnki.net/kcms/detail/11.2560.TP.20170712.1616.017.html>

great progress. This paper aims to review the development of multi class twin support vector machines, classify and analyze them with the respect to the basic theories and geometric meanings. According to the structures, the paper divides the machines into the following groups: “one-versus-all” strategy based multi class twin support vector machines, “one-versus-one” strategy based multi class twin support vector machines, binary tree based multi class twin support vector machines, “one-versus-one-versus-rest” strategy based multi class twin support vector machines, and “all-versus-one” strategy based multi class twin support vector machines. Although the training processes of direct acyclic graph based multi class twin support vector machines are much similar with that of “one-versus-one” based approaches, the decision processes have their own characteristics and disadvantages, and therefore they are divided into a separate group. This paper analyzes and summarizes the ideas and theories of different multi class twin support vector machines, and presents experimental results to compare the performances. This review can make it easy for novices to understand the essential differences and help to choose the suitable multi class twin support vector machine for a practical problem.

Key words: multi class classification; twin support vector machine; multiple birth support vector machine; support vector machine

孪生支持向量机(twin support vector machine,简称 TWSVM)是在传统支持向量机(support vector machine,简称 SVM)^[1,2]的基础上发展而来的一种新型机器学习算法.为了解决二分类问题,TWSVM 为每一类样本构造一个超平面,使每类样本尽可能离本类的超平面更近,而尽可能地远离另一类的超平面^[3].TWSVM 的两个超平面通过求解两个二次规划问题(quadratic programming problems,简称 QPPs)得到,每个 QPP 的约束条件都只与一类样本有关.TWSVM 不但保持了 SVM 的优点,而且训练速度比传统 SVM 要快 4 倍^[4].为了进一步提升 TWSVM 的性能,学者们已经提出了不少改进算法^[5-21],例如,Kumar 等人提出了 TWSVM 的最小二乘模型,称作最小二乘孪生支持向量机(least squares twin support vector machine,简称 LSTSVM)^[5].LSTSVM 的训练过程中仅需求解线性方程组,因此训练速度远快于 TWSVM.丁世飞等人根据样本点的位置为每个训练样本赋予不同的重要性,以降低异常点对非平行超平面的影响,并使用基于 CHKS 函数的光滑算法快速求解二次规划问题,得到了训练速度更快、鲁棒性更强的加权光滑 CHKS TWSVM^[11].此外,TWSVM 的应用研究也已经取得了不少成果^[22-25].

TWSVM 最初是为解决二分类问题而提出来的,不能直接用于多分类问题.现实问题中,更加普遍的是多分类问题,因此,多分类孪生支持向量机(multi class twin support vector machines,简称 MTWSVMs)的研究具有现实意义^[26].除了将 TWSVM 与多分类支持向量机中常用策略结合得到相应的 MTWSVMs 外,为了进一步提升分类能力和训练速度,学者们还提出了不少新颖的改进算法.目前,MTWSVMs 的研究已经取得了比较丰富的成果,但是针对 MTWSVMs 的相关综述性研究并不多,已有的综述性工作也没有给出全面的总结和分析.Tomar 和 Agarwal 仅对基于 LSTSVM 的多分类算法进行了比较^[26].Ding 等人在他们的 TWSVMs 综述中初步分析了偏二叉树 MTWSVM 和 Boosting MTWSVM 算法.Tomar 等人介绍了 Twin KVC、多生支持向量机(multiple birth support vector machine,简称 MBSVM)、决策树 TWSVM、优化的有向无环图最小二乘孪生支持向量机和最小二乘 Twin KSVC^[27],但没有对它们进行归类总结,也没有分析它们的基本思想.不同于已有工作,本文根据 MTWSVMs 的子分类器的组织结构将它们分为基于“一对多”策略的 MTWSVMs、基于“一对一”策略的 MTWSVMs、基于“一对一对余”策略的 MTWSVMs、基于二叉树的 MTWSVMs、基于“多对一”策略的 MTWSVMs.虽然基于有向无环图(directed acyclic graph,简称 DAG)结构的 MTWSVMs 具有与“一对一”MTWSVMs 类似的训练过程,但其决策过程比较特殊,因此也单独作为一类而加以分析.本文在分类基础上对比各种 MTWSVMs 模型的同时,在各个类型的 MTWSVMs 中又将改进模型与对应的原 MTWSVMs 模型进行对比.因此,本文从多方面综合性地综述了 MTWSVMs,较已有工作更全面地总结了 MTWSVMs 的研究发展状况.本文称文献[4]中最初的 TWSVM 为标准 TWSVM,TWSVMs 包括标准 TWSVM 以及所有 TWSVM 的改进算法,MTWSVMs 是指基于 TWSVM 及其改进算法的所有多分类算法.

1 二分类孪生支持向量机简介

本节介绍 TWSVM 的数学模型.假设在 R^n 的空间中有 m 个训练样本,它们都有 n 个属性,其中有 m_1 个样本属于正类, m_2 个样本属于负类,它们分别用矩阵 A 和矩阵 B 来表示.矩阵 A 的每行各代表一个属于正类的样本点.

矩阵 B 的每行各代表一个属于负类的样本点.

1.1 线性孪生支持向量机

对于线性可分的二分类问题,TWSVM 的求解过程即在 n 维实数空间 R^n 中寻找两个非平行的超平面:

$$x^T w_1 + b_1 = 0 \text{ 和 } x^T w_2 + b_2 = 0 \tag{1}$$

其中, w_1 和 w_2 是两个超平面的法向量, b_1 和 b_2 是两个超平面的偏移量. 它们一般通过求解下面两个二次规划问题而得到:

$$\left. \begin{aligned} \text{(TWSVM1)} \quad & \min \frac{1}{2} (Aw_1 + e_1 b_1)^T (Aw_1 + e_1 b_1) + c_1 e_2^T \xi \\ \text{s.t.} \quad & -(Bw_1 + e_2 b_1) + \xi \geq e_2, \xi \geq 0 \end{aligned} \right\} \tag{2}$$

$$\left. \begin{aligned} \text{(TWSVM2)} \quad & \min \frac{1}{2} (Bw_2 + e_2 b_2)^T (Bw_2 + e_2 b_2) + c_2 e_1^T \eta \\ \text{s.t.} \quad & (Aw_2 + e_1 b_2) + \eta \geq e_1, \eta \geq 0 \end{aligned} \right\} \tag{3}$$

其中, c_1 和 c_2 是两个惩罚参数, e_1 和 e_2 是元素全为 1 的列向量, ξ 和 η 是松弛变量. 在 TWSVM 的二次规划问题中, 目标函数的几何意义是每个超平面应尽可能地离本类数据点更近, 而约束条件则规定超平面要离其他类样本至少距离为 1 个单位.

直接求解 QPPs(2) 和 QPPs(3) 比较困难, 往往转化为其对偶问题进行求解, QPPs(2) 和 QPPs(3) 对应的对偶问题为

$$\left. \begin{aligned} \text{(TWSVM1)} \quad & \max e_2^T \alpha - \frac{1}{2} \alpha^T G (H^T H)^{-1} G^T \alpha \\ \text{s.t.} \quad & 0 \leq \alpha \leq c_1 \end{aligned} \right\} \tag{4}$$

$$\left. \begin{aligned} \text{(TWSVM2)} \quad & \max e_1^T \gamma - \frac{1}{2} \gamma^T P (Q^T Q)^{-1} P^T \gamma \\ \text{s.t.} \quad & 0 \leq \gamma \leq c_2 \end{aligned} \right\} \tag{5}$$

其中, $H=P=[A, e_1], G=Q=[B, e_2]$.

在得到两个分类超平面后, 线性 TWSVM 通过如下决策函数判定新样本的类别标签:

$$\text{Label}(x) = \arg \min_{k=1,2} \left(|x^T w_k + b_k| / \sqrt{w_k^T w_k} \right) \tag{6}$$

不难发现, 线性 TWSVM 的决策准则可简单描述为: 样本离哪个超平面近, 就归于该超平面对应的类.

图 1 所示为二维空间中线性 TWSVM 的示意图.

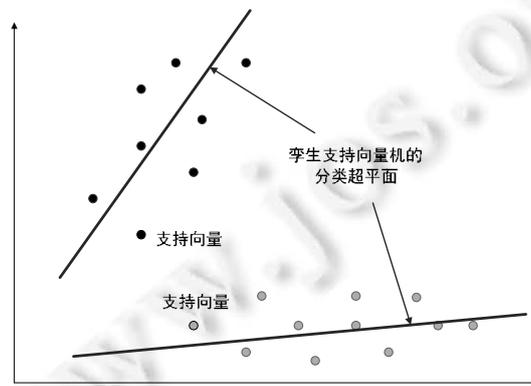


Fig.1 Illustration diagram of linear TWSVM

图 1 线性 TWSVM 示意图

本文通过图 1 说明 TWSVM 的几何解释. 图中黑色点代表正类样本, 灰色点代表负类样本. 图中两条直线表

示两个超平面.黑色点都离左上角的直线近而离另一超平面远,灰色点则反之.因此,如果待分类的样本离左上角的直线近,就应将其标为和黑色点同一类,即正类;否则,就将其标为负类.TWSVM 就是这样通过超平面来解决分类问题的.

1.2 非线性孪生支持向量机

对于非线性可分的分类问题,TWSVM 结合核技术进行处理.基本思路为:首先,将原始特征空间中的数据映射到高维再生空间中,使得映射得到的数据样本线性可分;然后,在高维空间中建立 TWSVM 的两个超平面.用 $K(x,y)$ 表示核函数,对于非线性二分类问题,TWSVM 构造如下两个基于核函数的超平面:

$$K(x^T, C^T)u_1 + b_1 = 0 \text{ 和 } K(x^T, C^T)u_2 + b_2 = 0 \quad (7)$$

类似于线性可分的情况,构造下面的二次规划问题并求解得到两个非平行超平面的法向量和偏移量:

$$\left. \begin{aligned} \text{(TWSVM1)} \quad & \min \frac{1}{2} \|K(A, C^T)u_1 + e_1 b_1\|^2 + c_1 e_1^T \zeta \\ \text{s.t.} \quad & -(K(B, C^T)u_1 + e_2 b_1) + \zeta \geq e_2, \zeta \geq 0 \end{aligned} \right\} \quad (8)$$

$$\left. \begin{aligned} \text{(TWSVM2)} \quad & \min \frac{1}{2} \|K(B, C^T)u_2 + e_2 b_2\|^2 + c_2 e_2^T \eta \\ \text{s.t.} \quad & (K(A, C^T)u_2 + e_1 b_2) + \eta \geq e_1, \eta \geq 0 \end{aligned} \right\} \quad (9)$$

其中,矩阵 C 代表所有训练样本,它的每一行即为一个训练样本.

求解公式(8)和公式(9)就得到了 TWSVM 的两个基于核的超平面.非线性情况下,TWSVM 的决策规则仍可归纳为测试样本离哪个超平面近就被归于哪个类.具体的决策函数为

$$\text{Label}(x) = \arg \min_{k=1,2,\dots,K} \left(K(x, C^T)u_k + b_k / \sqrt{u_k^T K(C, C^T)u_k} \right) \quad (10)$$

2 多分类孪生支持向量机的研究进展

算法 TWSVM 最初是为解决二分类问题而提出来的,因此需要结合额外的多分类扩展策略才能用于多分类问题.相比二分类问题,多分类问题往往包含更多的数据,不同类别的数据样本之间的分布情况也更加复杂.因此,MTWSVMs 是 TWSVMs 研究中的一个难点.本节对 MTWSVMs 作一个系统的回顾.除非特别说明,以下都考虑 n 维空间 R^n 中的一个 K 分类问题.记训练数据集为 $T = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_m, y_m)\}$, 其中, $x_i \in R^n$ 表示输入,即训练样本数据; $y_i \in \{1, 2, 3, \dots, K\}$ 表示输出,即样本标签.为表述方便,用 m_k 表示第 k 类所包含样本的个数,用矩阵 A_k 表示第 k 类训练数据,并定义 $B_k = T/A_k$, 即, B_k 表示除第 k 类样本以外的所有训练样本.

2.1 基于“一对多”策略的多分类孪生支持向量机

2.1.1 一对多孪生支持向量机的基本数学模型

“一对多(one-versus-all, 简称 OVA)”策略是最早被用于将二分类 TWSVM 扩展为 MTWSVM 的方法^[28,29]. 对于 K 分类问题,由标准 TWSVM 结合 OVA 策略得到的一对多孪生支持向量机(one-versus-all twin support vector machine, 简称 OVA TWSVM)需要为每一个类构造一个超平面,每一个超平面都对应于一个二次规划问题.在构造对应于第 k 类样本的超平面时, OVA TWSVM 把其中第 k 类作为一类,将剩余的 $K-1$ 类作为一类,构建一个 TWSVM 型的 QPP 并求解得到一个对应于第 k 类样本的超平面.该方法在每次的训练中都需要全部样本参加.对于线性可分的 K 分类问题, OVA TWSVM 分类器的第 k 个超平面的构建需要求解如下二次规划问题:

$$\left. \begin{aligned} \min \quad & \frac{1}{2} \|A_k w_k + e_k^{(1)} b_k\|^2 + c_k e_k^{(2)T} \xi_k \\ \text{s.t.} \quad & (B_k w_k + e_k^{(2)} b_k) + \xi_k \geq e_k^{(2)}, \xi_k \geq 0 \end{aligned} \right\} \quad (11)$$

其中,求得的 w_k 为第 k 个超平面的法向量, b_k 为偏移量, c_k 是取值为正数的惩罚参数, $e_k^{(1)}$ 和 $e_k^{(2)}$ 为元素全为 1 的列向量.

通过求解 K 个 QPPs 得到 K 个超平面,算法的训练阶段就完成了.在对新样本进行分类时,需要计算新样本

到 OVA TWSVM 的各个超平面的距离,未知样本的类别就是所有超平面中与该样本离得最近的超平面所对应的那一类.亦即 OVA TWSVM 的决策函数为

$$Label(x) = \arg \min_{k=1,2,\dots,K} \left(\left| x^T w_k + b_k \right| / \sqrt{w_k^T w_k} \right) \quad (12)$$

对于非线性分类问题,需要先使用核技术,即将数据点映射到高维空间,然后在高维空间中建立 OVA TWSVM.此时,OVA TWSVM 的第 k 个超平面通过求解如下最优化问题而得到:

$$\left. \begin{aligned} \min & \frac{1}{2} \| K(A_k, C^T)u_k + e_k^{(1)}b_k \|^2 + c_k e_k^{(2)T} \xi \\ \text{s.t.} & (K(B_k, C^T)u_k + e_k^{(2)}b_k) + \xi \geq e_k^{(2)}, \xi \geq 0 \end{aligned} \right\} \quad (13)$$

不难得到,非线性情况下 OVA TWSVM 的决策函数为

$$Label(x) = \arg \min_{k=1,2,\dots,K} \left(\left| K(x^T, C^T)u_k + b_k \right| / \sqrt{u_k^T K(C^T, C^T)u_k} \right) \quad (14)$$

图 2 所示为 OVA TWSVM 基本思想的一个示意图.图中展示的是用 OVA TWSVM 解决二维空间中的一个包含 3 个类的分类问题,黑色点、深灰色点、浅灰色点分别表示类 1、类 2 和类 3 的样本.对于该问题,通过求解公式(13)($k=1,2,3$)得到了 3 条直线.这 3 条直线分别表示 OVA TWSVM 对应于 3 个类的超平面.从图中不难发现,3 个超平面的特点是:类 1 的超平面离类 1 的样本点近而离类 2 和类 3 的样本点则远得多,其他两个超平面也是离其对应类的样本点近而离其他类的样本点要远.因此,假设待分类点离类 1 的超平面很近而离类 2 和类 3 的超平面很远,则将该新样本标为类 1.

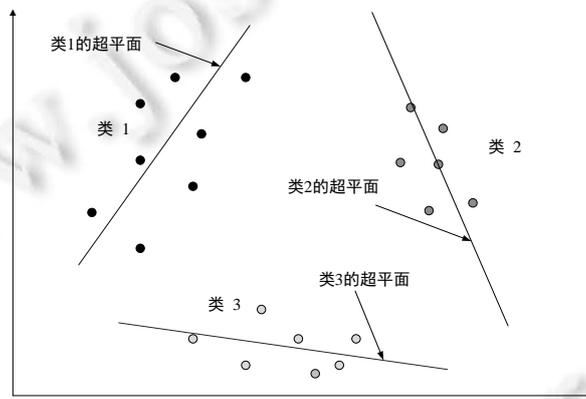


Fig.2 Illustration diagram of OVA TWSVM
图 2 OVA TWSVM 示意图

OVA TWSVM 思想简单易懂,而且容易实现,因此受到应用领域专家的青睐^[30].

2.1.2 OVA TWSVM 的改进算法

OVA TWSVM 需要求解 K 个二次规划问题,而当训练数据较大时,常用的求解二次规划问题的方法不但耗时长而且占用内存多.为了加快训练速度,可以使用一对多多分类最小二乘孪生支持向量机(one-versus-all multi class least squares twin support vector machine,简称 OVA MLSTSVM)^[31,32].与 OVA TWSVM 一样,OVA MLSTSVM 也是将 K 分类问题转化为寻找 K 个超平面的问题;但不同于 OVA TWSVM,它最终将二次规划问题转化为线性方程组进行求解,求解线性方程的算法相对要快速便捷,因此,OVA MLSTSVM 比 OVA TWSVM 要快得多.同时,OVA MLSTSVM 的分类准确率总是和 OVA TWSVM 相近.

在 TWSVM 基础上得到的二分类方法投影孪生支持向量机(projection twin support vector machine,简称 PTSVM)在复杂交叉数据的处理方面具有优秀的性能.Li 等人将 PTSVM 与 OVA 策略相结合,得到了多分类递归投影孪生支持向量机(multiple projection twin support vector machine,简称 Multi-PTSVM)^[33].不同于 OVA

TWSVM, Multi PTSVM 为每一个类寻找最佳投影轴,使得该类样本投影后尽可能集中,而其他类样本投影后离本类投影中心尽可能地远. Multi PTSVM 继承了 PTSVM 在处理复杂交叉数据集时高效的优点,在复杂交叉数据的应用中分类性能优于 OVA TWSVM. 由于 Multi PTSVM 需要求解二次规划问题,使得训练速度有提升空间. 为了提升 Multi PTSVM 的训练速度, Yang 等人参考 LSTSVM 设计得到了多分类最小二乘投影孪生支持向量机 (multiple least squares recursive projection twin support vector machine, 简称 MLSPTSVM)^[34]. MLSPTSVM 将 Multi PTSVM 模型中的不等式约束改为等式约束,然后将二次规划问题转化为线性方程组. 该算法只需求解线性方程组,在保持分类性能的同时,具有更快的训练速度.

数据不平衡现象不但在二分类问题中存在,在多分类问题中更为普遍. 在分类问题中,如果某一类样本数目要明显多于其他类,则在分类器训练过程中样本数目少的类很难发挥作用,这会严重影响得到的分类器的性能. 为了使 OVA LSTSVM 更好地处理数据不平衡的多分类问题, Tomar 和 Agarwal 在损失函数中为每一个样本添加了一个与其所属类别的样本数目相关的权重^[35]. 他们的权重分配基本原则是:为样本数据少的类别赋予较大的权重,而为样本数目大的类别赋予较小的权重. 该算法在一定程度上克服了数据不平衡可能造成的性能损失,相对于原 OVA LSTSVM,在处理数据不平衡问题时性能有一定的提升.

损失函数对基于 TWSVM 的分类算法具有决定性影响. Shao 等人用加权线性损失替换 TWSVM 中的 Hinge 损失,得到了加权线性损失孪生支持向量机 (weighted linear loss twin support vector machine, 简称 WLTSVM),并结合“一对多”思想,得到了可用于多分类问题的 Multiple WLTSVM^[36]. Multiple WLTSVM 和 OVA MLSTSVM 都是将分类问题转化为求解线性方程组,因此, Multiple WLTSVM 的训练速度接近 OVA MLSTSVM. 它们之间的本质区别是,选用了不同的损失函数来估计错分风险. 这也造成两者在几何解释上稍有不同, Multiple WLTSVM 与 OVA TWSVM 一样,在为某一类构造超平面时都要求其他类的样本尽可能地远离该超平面. 而 OVA LSTSVM 则不同,它要求其他类样本尽可能均匀地处于离该超平面距离为 1 的平行超平面的两侧. 因此, Multiple WLTSVM 的泛化性能更接近 OVA TWSVM.

2.1.3 基于“一对多”策略的多分类孪生支持向量机的优缺点

OVA 策略原理简单而且容易实现. 基于 OVA 策略的 MTWSVMs 只需为每个类构造一个超平面,每类的超平面与该类样本接近. 不难发现,这类多分类算法继承了 TWSVM 的思想,是 TWSVM 的一种自然扩展. 这类算法的缺点有:

- 其一,模型中的每个 QPP 或者等价的线性方程组的约束条件个数是 $K-1$ 个类的样本数目,约束条件较多. 而约束的数目越多,训练速度就越慢. 当训练数据为大规模数据时,这类算法的效率往往较低,所以这类算法对于样本规模较大的多分类问题不适用;
- 其二,在训练某一类的超平面时,该类算法将这个类视为正类,而将其他 $K-1$ 个类全部视为负类,这就造成了数据不平衡现象. 如果不对数据不平衡现象进行处理,则会导致训练得到的分类器性能不佳,从而影响分类时的准确率,因此,需要额外的数据不平衡处理方法;
- 其三,该类方法性能存在“木桶现象”,即:如果有一个分类器的性能比较低,就会严重影响整个分类器的泛化性能;
- 其四,这类方法存在不可分区域,即:存在与两个超平面距离相近的区域,在该区域的样本点难以判别其类别^[26].

2.2 基于“一对一”策略的多分类孪生支持向量机

2.2.1 一对一孪生支持向量机的数学模型

“一对一 (one-versus-one, 简称 OVO)”策略分类方法最早是由 Knerr 为多分类 SVM 而提出来的,将该策略与标准 TWSVM 相结合,得到的一对一孪生支持向量机 (one-versus-one twin support vector machine, 简称 OVO TWSVM) 具有比 OVO SVM 更好的分类性能^[36]. 对于 K 类分类问题,该算法是在任意两类样本间构建一个二分类 TWSVM 子分类器,总共需要构造 $K(K-1)/2$ 个二分类 TWSVM,也即构造 $K(K-1)$ 个超平面. OVO TWSVM 的每个二分类器的生成只需要用到训练样本中的两类数据,实现 K 类分类问题中两类的区分. 具体地,对于线性可

分的分类问题,在训练 i, j 两类样本间的二分类器时,需要获取两个超平面:

$$x^T w_{ij} + b_{ij} = 0 \text{ 和 } x^T w_{ji} + b_{ji} = 0 \quad (15)$$

为获取以上两个超平面的法向量和偏移,需要解决如下二次规划问题:

$$\left. \begin{aligned} \min & \frac{1}{2} \|A_i w_{ij} + e_{ij}^{(1)} b_{ij}\|^2 + \frac{C_{ij}}{2} e_{ij}^{(2)T} \xi_{ij} \\ \text{s.t.} & (A_j w_{ij} + e_{ij}^{(2)} b_{ij}) + \xi_{ij} \geq e_{ij}^{(2)}, \xi_{ij} \geq 0 \end{aligned} \right\} \quad (16)$$

$$\left. \begin{aligned} \min & \frac{1}{2} \|A_j w_{ji} + e_{ij}^{(2)} b_{ji}\|^2 + \frac{C_{ji}}{2} e_{ij}^{(1)T} \xi_{ji} \\ \text{s.t.} & (A_i w_{ji} + e_{ij}^{(1)} b_{ji}) + \xi_{ji} \geq e_{ij}^{(1)}, \xi_{ji} \geq 0 \end{aligned} \right\} \quad (17)$$

构造完所有二分类子分类器后,就可以联合它们进行新样本的类别决策了.OVO TWSVM 一般采用“投票法”进行类别的判断:“投票法”将所有类别作为“候选者”,而每个二分类子分类器都是“选民”.每个“选民”有且仅有一张“选票”,只能投给一名“候选者”.在决定待分类样本 x 的归属时,依次用 OVO TWSVM 的 $K(K-1)/2$ 个二分类 TWSVM 进行判别,如果 i, j 两类间的分类函数将新样本 x 分到 i 类,则该子分类器将“票”投给了 i 类,第 i 类的“得票”加 1;否则, j 类得到该“选票”.当遍历完 $K(K-1)/2$ 个二分类 TWSVM 后,根据“票数”来确定待分类样本 x 所属的类别.样本 x 将被分到“票数”最高的那个类.

对于训练数据非线性可分的情况,OVO TWSVM 在训练 i, j 两类样本间的分类器时,需要解决如下最优化问题:

$$\left. \begin{aligned} \min & \frac{1}{2} \|K(A_i, C^T) w_{ij} + e_{ij}^{(1)} b_{ij}\|^2 + \frac{C_{ij}}{2} \xi_{ij}^T \xi_{ij} \\ \text{s.t.} & (K(A_j, C^T) w_{ij} + e_{ij}^{(2)} b_{ij}) + \xi_{ij} = e_{ij}^{(2)}, \xi_{ij} \geq 0 \end{aligned} \right\} \quad (18)$$

$$\left. \begin{aligned} \min & \frac{1}{2} \|K(A_j, C^T) w_{ji} + e_{ij}^{(2)} b_{ji}\|^2 + \frac{C_{ji}}{2} \xi_{ji}^T \xi_{ji} \\ \text{s.t.} & (K(A_i, C^T) w_{ji} + e_{ij}^{(1)} b_{ji}) + \xi_{ji} = e_{ij}^{(1)}, \xi_{ji} \geq 0 \end{aligned} \right\} \quad (19)$$

构建好 OVO TWSVM 后,对新样本进行类别标签预测时,与线性情况下一样,一般采用“投票”法.下面结合图 3 说明 OVO TWSVM.

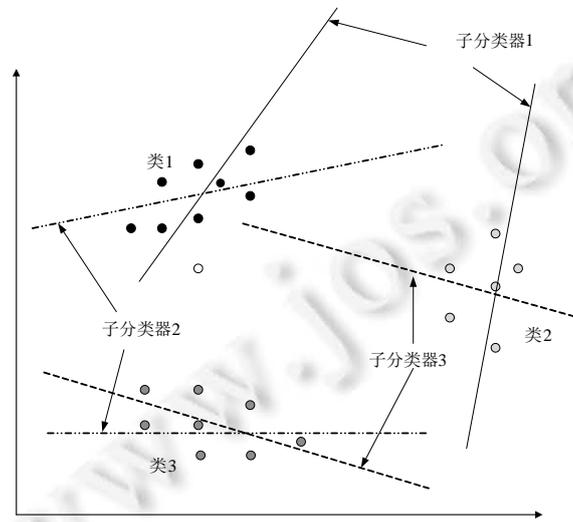


Fig.3 Illustration diagram of OVO TWSVM

图 3 OVO TWSVM 示意图

图3所示为使用OVO TWSVM解决二维空间中的三分类问题示意图.取公式(18)和公式(19)的 i 和 j 为1和2并求解,可以得到一个能够对类1和类2进行分类的二分类TWSVM,即图3所示的子分类器1.通过类似方法得到子分类器2和子分类器3,它们共同构成了解决该三分类问题的OVO TWSVM.假设图3所示空心点是待分类的点,下面以确定图中空心点的类别为例说明OVO TWSVM的决策过程,也即“投票法”的使用过程.

- 首先,使用子分类器1,计算空心点到子分类器1的两个超平面的距离.因为空心点离子分类器1对应于类1的超平面更近,因此类1得1票;
- 类似地,计算空心点到子分类器2的两个超平面的距离.因为空心点离子分类器2对应于类1的超平面更近,因此类1再得1票;
- 计算空心点到子分类器3的两个超平面的距离.因为空心点离子分类器3对应于类2的超平面更近,因此类2得1票.

最终,类1得到2票,类2得到1票,类3得到0票.OVO TWSVM将待分类样本归为类1.

2.2.2 OVO TWSVM的改进算法

将最小二乘思想引入OVO TWSVM,就可以使得求解 $K(K-1)$ 个二次规划问题转化为求解 $K(K-1)$ 个线性方程组的问题.Tomar等人在详细分析了几种基于LSTSVM的多分类方法后认为:一对一多分类最小二乘孪生支持向量机(one versus one multi class least squares twin support vector machine,简称OVO MLSTSVM)具有更好的分类性能,训练速度快而且分类精度高^[26].类似于LSTSVM,OVO MLSTSVM将OVO TWSVM中的不等式约束改为等式约束,然后将等式约束代入目标函数,进而将二次规划问题化为无约束求最小值问题.对无约束求最小值问题的目标函数求导,就得到了等价的线性方程组,通过求解方程组,就能得到原二次规划的解.OVO MLSTSVM的预测阶段与OVO MTWSVM完全一样,也采用“投票法”.OVO MLSTSVM具有与OVO TWSVM接近的分类精度,但其训练速度要远快于OVO TWSVM.

2.2.3 基于“一对一”策略的多分类孪生支持向量机的优缺点

基于OVO策略的MTWSVMs及其改进算法的优点是每个子分类器在训练时只需要两个类的训练样本,因此每个子分类器处理的样本较少,训练很快就能完成^[37].虽然基于OVO策略的多分类孪生支持向量机及其改进算法需要构造的超平面数目较多,但当样本规模很大而所含类别数不是特别大时,相比于OVA MTWSVMs,整体所需的训练时间依然更少.对于每个子分类器,训练样本只涉及两类,两类别之间的样本数目是相对平衡的,因此往往不需要额外的手段来处理数据不平衡问题就能训练得到性能较好的子分类器,进而使得整个分类模型有更高的分类准确率.这类多分类算法的缺点是:当训练数据包含的类别数很大时,需要训练的子分类器个数太多.随着样本类别数 K 的增加,训练的子分类器个数按平方倍增加,这使得整个分类系统变得复杂.此外,这类方法在决策阶段需要遍历所有子分类器,因此样本类别数很大时,分类判决的速度也很慢.另一个问题是,这类方法存在不可分现象,即:采用“投票法”进行决策时,可能存在多个获得票数一样的类别,随机将样本判为其中一类或者直接判为序号最靠前的一类都会影响了分类的精度.

2.3 基于有向无环图的多分类孪生支持向量机

2.3.1 有向无环图多分类孪生支持向量机的基本思想

DAG TWSVM采用图论中的有向无环图将多个二分类TWSVM组合成多分类器.如果图中每条边都有单一的方向,并且图中无环,这样的图就称为有向无环图^[38].对于 K 类问题,DAG TWSVM采用与OVO TWSVM一样的方式构造 $K(K-1)/2$ 个二分类TWSVM.不同的是,DAG TWSVM将这些子分类器分布于 K 层DAG结构中^[39].DAG TWSVM的DAG中顶层只有一个节点,称为根节点,第 i 层含 i 个节点,如果 $1 < i < K$ 这些节点称为内部节点,则第 K 层即最底层含 K 个节点,称为叶子节点.这些节点由有向边连接,第 j 层的第 i 个节点指向第 $j+1$ 层的第 i 个和第 $i+1$ 个节点.每个内部节点都是一个子分类器,即一个二分类TWSVM,叶子节点为最终的分类类别.一个3层的DAG拓扑结构如图4所示.对测试样本进行分类时,DAG TWSVM不采用“投票法”,该类方法先将测试样本输入根节点分类器,根据根节点分类器的输出值,判定走向左侧或右侧子节点,再根据下一层内部节点分类器的输出值,继续选择左侧或者右侧前进,直至走到某一叶子节点,就得到了该测试样本的类别^[40].

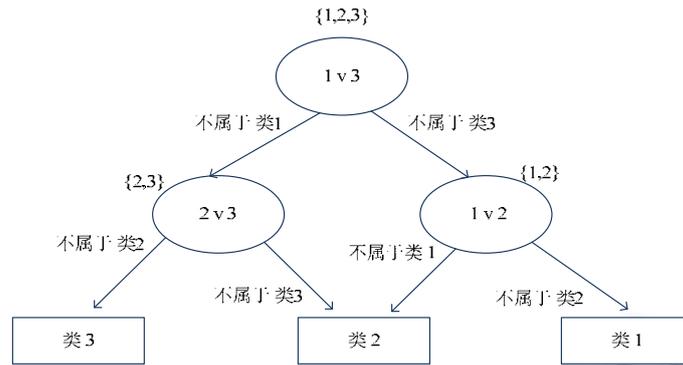


Fig.4 Illustration diagram of the topology of DAG MTWSVM for a 3 class classification problem

图 4 DAG MTWSVM 对于 3 分类的拓扑结构示意图

2.3.2 DAG TWSVM 的改进算法

Tomar 等人将 LSTSVM 与 DAG 相结合,得到了基于 DAG 的最小二乘多分类孪生支持向量机(directed acyclic graph multiclass least squares twin support vector machine,简称 DAG MLSTSVM)^[26].该方法的训练和分类过程与 DAG MTWSVM 类似,不同的是,该算法继承了 LSTSVM 的优点,即训练过程只需处理线性方程组.最小二乘思想的引入,在保证分类精度的同时提升了算法的整体运行速度.

Chen 等人发现,DAG 依然具有树形结构所固有的缺点,即误差会在向下传递的过程中累积并放大^[41].因此,应将分类性能好的二分类子分类器尽可能地安排在 DAG 的上层,而将误差大的二分类子分类器安排在接近 DAG 的叶子节点处.考虑到二分类器的性能很大程度上取决于训练数据集不同类别之间可分性的大小,Chen 等人参考参数选择中距离判断的思想设计了基于类内和类间样本间平均距离的数据类可分性评判准则,并结合该准则设计得到了最优有向无环图孪生支持向量机(optimal directed acyclic graph multiclass twin support vector machine,简称 ODAG TWSVM).ODAG TWSVM 提升了分类精度,但也增加了训练时间开销;并且随着数据类别的增大,计算类可分性值的额外开销也会增加,带来的分类器性能提升也越明显.

在 DAG MLSTSVM 的基础上,张谢锴等人引入马氏距离得到了基于马氏距离的多分类孪生支持向量机^[42].该方法利用了马氏距离的全局性提升了算法的泛化能力,对于数据方差大的数据集,其分类结果要好于 DAG TWSVM.

李侃等人为了对多传感器感应到的目标进行分类,设计了一种带置信度的最小二乘投影孪生支持向量机,并结合 DAG 策略提出了多路有向无环图支持向量机^[43].该算法的每个节点除了输出分类结果以外,还会输出一个该分类结果的置信度.为了合理安排 DAG 结构,该算法使用最小超球体矩对类进行排序,优先选择平均最小超球体矩大的类别进行训练,得到上层的子分类器,以减少累计误差.该算法在多传感器探测目标分类问题的应用中取得了预期的效果.

2.3.3 基于有向无环图的多分类孪生支持向量机的优缺点

基于 DAG 的 MTWSVMs 的优点是,在进行判别时,不需要对每个分类函数进行计算.对于 K 类分类问题,通过对 $K-1$ 个分类器的判决就可得出一个未知样本所属的类别.与“投票法”相比,分类需要调用的分类器个数明显减少,从而提高了分类的速度.此外,DAG TWSVMs 不存在拒分区域,泛化性能更好.该类方法的缺点包括:其一,该类方法在训练时存在与一对一分类方法同样的缺点,即,训练的子分类器个数与样本的类别数有关,随着类别数 K 的增加,分类器个数呈平方倍增加,使得分类判决的速度太慢;其二,DAG 存在从上层节点到下层节点的“误差累积”现象,假如在某一节点处发生了错误的判别,这种错误将会一直累积并在向下传递的过程中被放大,直到整个判决过程的结束.同时,当样本类别数目很多时,得到的累积误差将会更大,判别出错情况发生得越早,这种累积效应越明显,对分类性能造成的影响越大.因此,在分类精度方面,基于 DAG 的 MTWSVMs 分类方法往往不如基于 OVO 的 MTWSVMs 分类方法.

2.4 基于“一对一对余”策略的多分类孪生支持向量机

2.4.1 Twin KSVC 的数学模型

OVO TWSVM 由多个二分类 TWSVM 组合而成,决策阶段也是依赖这些二分类 TWSVM.OVO TWSVM 的子分类器总是将样本划分到两个类的其中一类,然而当测试样本不属于当前子分类器对应的两个类的任何一类时,该测试样本可能与该两类差异都很大,这就可能造成当前子分类器对该测试样本的判决是无意义的,进而产生随机性“投票”.为了避免这种随机性,Xu 等人在 KSVC 的基础上提出了一种新的多分类孪生支持向量机,称作 Twin KSVC^[44].该方法将训练样本组织为“一对一对余”形式,即:每个子分类器将样本分为正类、负类,还有一部分既不属于正类又不属于负类,称为其余类,对应的输出为{1,0,+1}.Twin KSVC 的子分类器是在二分类 TWSVM 基础上修正得到的一种变型.为表述方便,此处作特殊符号说明,本节使用 A, B 分别表示一个类, C 表示除 A, B 外的其余类.Twin KSVC 的子分类器的数学模型可表述为如下二次规划:

$$\left. \begin{aligned} \min & \frac{1}{2} \|Aw_1 + e_1b_1\|^2 + c_1e_2^T\xi + c_2e_3^T\eta \\ \text{s.t.} & -(Bw_1 + e_2b_1) + \xi \geq e_2 \\ & -(Cw_1 + e_3b_1) + \eta \geq e_3(1 - \varepsilon) \\ & \xi \geq 0e_2, \eta \geq 0e_3 \end{aligned} \right\} \quad (20)$$

$$\left. \begin{aligned} \min & \frac{1}{2} \|Bw_2 + e_2b_2\|^2 + c_3e_1^T\xi^* + c_4e_3^T\eta^* \\ \text{s.t.} & (Aw_2 + e_1b_2) + \xi^* \geq e_1 \\ & (Cw_2 + e_3b_2) + \eta^* \geq e_3(1 - \varepsilon) \\ & \xi^* \geq 0e_1, \eta^* \geq 0e_3 \end{aligned} \right\} \quad (21)$$

Twin KSVC 通过求解上面的二次规划问题构造两个超平面,要求正类和负类离各自的超平面尽可能地近,而其余的样本离这两个超平面的距离都要至少为 1.上面的两个二次规划问题一般通过其对偶问题进行求解.

$$\left. \begin{aligned} \max & -\frac{1}{2}\gamma^T N(H^T H)^{-1} N^T \gamma + e_4^T \gamma \\ \text{s.t.} & 0 \leq \gamma \leq F \end{aligned} \right\} \quad (22)$$

$$\left. \begin{aligned} \max & -\frac{1}{2}\rho^T P(G^T G)^{-1} P^T \rho + e_5^T \rho \\ \text{s.t.} & 0 \leq \rho \leq F^* \end{aligned} \right\} \quad (23)$$

其中, $F=[c_1e_1; c_2e_3]$, $F^*=[c_3e_1; c_4e_3]$, $e_4=[e_2; e_3(1-\varepsilon)]$, $e_5=[e_1; e_3(1-\varepsilon)]$, $H=[A, e_1]$, $G=[B, e_2]$, $M=[C, e_3]$, $N=[G, M]$, $P=[H; M]$.

Twin KSVC 也采用“投票法”来确定新样本的最终类别.即:如果子分类器输出为+1,则对应于正类的类别增加 1 票;如果输出为 1,对应于负类的那个类增加 1 票;否则,任何类都不得票.最终,票数最多的类别即为新样本的所属类别.Xu 等人将该算法应用到说话人识别(speaker recognition),仿真实验结果表明了算法的有效性^[44].

2.4.2 Twin KSVC 的改进算法

Twin KSVC 子分类器的训练需要全部数据的参与,因此速度相对慢.Nasiri 等人^[45]将最小二乘思想引入到 Twin KSVC,得到了最小二乘 Twin KSVC,缩写为 LST KSVC.LST KSVC 和 Twin KSVC 一样,每个子分类器输出结果可能为{1,0,+1}.最小二乘思想的引入,提升了算法的训练速度.LST KSVC 将 Twin KSVC 中不等式约束用等式约束加以替换,采用二次损失函数估计分类误差,最终,通过求解线性方程组就可以解决分类问题.LST KSVC 不但具备更快的训练速度,而且保持了 Twin KSVC 所具有的高分类准确率.

现实生活中,获取数据标签不是一个简单的任务,尤其是现在数据量庞大,很多情况下都不能获得所有数据点的分类标签信息,因此,半监督问题越来越受到关注.Khemchandani 和 Pal 针对半监督多分类问题提出了拉普拉斯多分类孪生支持向量机(multi-category Laplacian least squares twin support vector machine,简称 Lap LST KSVC)^[46].Lap LST KSVC 能够在充分利用有标签训练数据的同时,将无标签数据所包含的流形信息也利用起来,以提升算法的泛化性能.Lap LST KSVC 与 Twin KSVC 采用相同的“一对一对余”策略,为了使用无标签数据

信息,Lap LST KSVC 使用流形正则化项技术发现无标签数据中的流形结构,将数据量更大的无标签数据应用到分类中,不仅适用于更多实际情况,而且提升了算法的性能.由于采用了最小二乘思想,不论是线性还是非线性情况,Lap LST KSVC 的模型都能很容易地转化为一组线性方程组,因此,Lap LST KSVC 的训练速度也较快.

2.4.3 基于“一对一对余”策略的多分类孪生支持向量机的优缺点

与基于 OVO 的 MTWSVMs 相比,基于“一对一对余”策略的多分类孪生支持向量机的子分类器构造的两个超平面不但满足各个超平面离本类尽可能近而离另一个类尽可能远的要求,而且充分利用了其他 $K-2$ 个类的信息,要求其他 $K-2$ 类的样本离正负类的超平面都距离至少为 1 个单位.这类方法采用 $\{1,0,+1\}$ 形式的输出,避免了 OVO 策略中可能出现的随机“投票”,因此,这类方法的泛化性能更好一些.但是这类方法的子分类器需要求解一个具有 $K-1$ 类样本数目的约束的二次规划问题,子分类器比较复杂,时间复杂度高于 OVO TWSVMs.为了提升训练速度,应该结合最小二乘等加速策略.这类方法同样采用“投票”法进行新样本的分类,也可以采用类似 DAG 的决策规则来加快决策速度.

2.5 基于二叉树的多分类孪生支持向量机

2.5.1 二叉树多分类孪生支持向量机的基本思想

基于二叉树的多分类孪生支持向量机(binary tree based twin support vector machine,简称 BT TWSVM)针对基于一对一结构和一对多结构的 MTWSVMs 中存在的拒分问题而提出,这类方法先将训练数据中的所有类别划分为两个组,一组看作正类,另一组看作负类,训练得到根节点子分类器;然后,每个组又划分为两个更小的组,训练得到下一层的子分类器.以此继续,直到划分出最终类别.每次划分后,两类分类问题的规模是逐级下降的.BT TWSVM 的思路如图 5 所示,图中展示的是一个五分类问题的分类二叉树,每个节点对应一个二分类 TWSVM,根节点的二分类 TWSVM 是将第 1 类~第 3 类的样本一起看作正类样本,将第 4 类~第 6 类的样本看作负类样本进行训练得到的.因此,如果一个新样本被根节点分为正类,则该新样本可能是属于第 1 类~第 3 类中的一个;之后,由根节点的孩子节点继续缩小新样本可能所属类的范围,直至确定其所属类别.BT TWSVM 不存在不可分区域,分类精度高,算法复杂度低,已被应用于钢带表面缺陷识别、彩色图像分类等领域^[47-51].

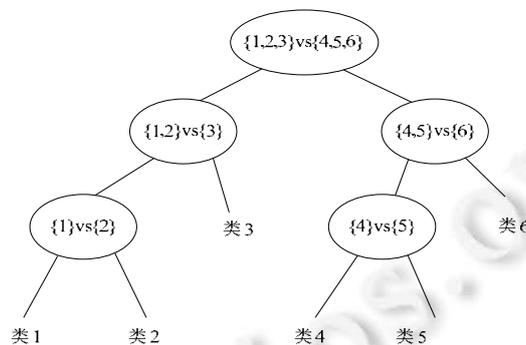


Fig.5 Illustration diagram of the topology of BT TWSVM for a 6 class classification problem

图 5 BT TWSVM 对于 6 分类的拓扑结构示意图

2.5.2 二叉树多分类孪生支持向量机的改进算法

聂盼盼等人^[50]为了减少 BT TWSVM 中的误差向下传递和累计,使用 k 均值算法先对训练数据集进行聚类,再使用聚类中心间距离代表类间距离,类间距离越大的分组方式,就对应二叉树中越上层的节点,以此得到合理的二叉树结构.他们将这种方法应用于入侵检测,得到了很好的效果.

Khemchandani 等人将 Laplacian TWSVM 与二叉树相结合,得到可以用于半监督多分类的 TB Lap TWSVM^[51].其训练过程与 BT TWSVM 一样,只是子分类算法选用的是 Lap TWSVM 而不是 TWSVM.TB Lap

TWSVM 能够将无标签数据的信息利用起来,以有效处理半监督多分类问题,具有较低的时间复杂度和较高的分类性能.

谢娟英等人^[52]将孪生支持向量机与偏二叉树相结合,得到了偏二叉树多分类孪生支持向量机.该方法首先选取类 1 作为正类,第 2 类~第 K 类作为负类训练,得到根节点子分类器;然后,将类 2 作为正类,类 3~类 K 作为负类,训练得到根节点的子节点.依次继续,直至训练得到第 $K-1$ 类和第 K 类的二分类器.该方法总是选取一个类作为正类,剩余多个类作为负类,容易出现数据不平衡现象.谢娟英等人对正负类实施不同的惩罚系数来克服这一缺陷.该算法的缺点是利用偏二叉树来解决多分类问题,树的层多,所需训练时间也长,而利用偏二叉树的优点是实现简单.

李秋林等人提出了一种基于二叉树和超球孪生支持向量机的多分类算法^[53].为了减少累计误差,该方法使用一种以类样本分布情况、超球体半径或超长长方体为依据生成二叉树的方法.为了加快训练速度,该方法使用坐标轮换法和收缩技术求解超球孪生支持向量机子分类器的规划问题.该算法训练速度较快,分类精度较高,适合数据稀疏性较强的多分类问题.

为了得到良好的二叉树结构,Shao 等人提出了“最优分离原则”^[54].该原则包含两条:在将众多的类分为两组的过程中,称分组是满足最优分离原则的,如果该分组得到两个组的数据的距离在所有可能的分组中是最大的;称分组满足最优分离原则,如果该分组得到的两组数据的分类错误率是所有分组中最低的.根据以上原则,Shao 等人提出了最优分离树孪生支持向量机(best separating decision tree twin support vector machine,简称 DTTSVM). DTTSVM 通过合理的二叉树结构,可以较好地消除累积误差,从而得到更好的分类结果.

2.5.3 基于二叉树的多分类孪生支持向量机的优缺点

基于决策二叉树的 MTWSVMs 采用广为人知的二叉树,使得算法具有简单的结构,容易实现.

这类 MTWSVMs 的一个重要特点和优点是解决了一对多和一对一方法中存在的不可分问题,从而算法本身更加稳健.对于 K 类分类问题,该方法只需要训练 $K-1$ 个二分类器,并且随着二叉树层次的增加,训练的规模不断缩小,训练的时间也随之减少,训练阶段时间复杂度较低;分类判决时不一定需要对每个分类器进行计算,从而减少了分类判决所需时间.缺点在于:决策二叉树的结构是影响分类器性能的关键因素,结构不同的决策树,其训练速度和分类能力都有较大差异.因此,该方法的关键是构建合理的树结构.当树结构趋于满二叉树时,这种分类方法的训练和分类速度达到最佳.树形结构必然存在累计误差,因此越是上层的节点越应该合理设计,从而减少累计误差.

2.6 基于“多对一”策略的多分类孪生支持向量机

2.6.1 多生支持向量机

Yang 人^[55]将 TWSVM 与“多对一(all-versus-one,简称 AVO)”策略相结合,提出了多生支持向量机(multiple birth support vector machine,简称 MBSVM).MBSVM 因其较低的算法复杂度而受到学者关注.MBSVM 每次仅选取一个类作为负类,剩下的类都当作正类构建一个 TWSVM 型的 QPP.MBSVM 的每个 QPP 的约束条件只与一类样本有关,因此,MBSVM 的二次规划模型中,约束条件较“一对多”策略实现的 MTWSVMs 要少得多.MBSVM 共需求解 K 个二次规划问题以获得 K 个分类超平面.对于线性情况,MBSVM 的求解过程即在 R^n 空间中找到 K 个非平行的超平面,这 K 个超平面通过求解下面的一组二次规划得到:

$$\left. \begin{aligned} \min_{w_k, b_k, \xi_k} & \frac{1}{2} \|B_k w_k + e_{k1} b_k\|^2 + c_k e_{k2}^T \xi_k \\ \text{s.t.} & (A_k w_k + e_{k2} b_k) + \xi_k \geq e_{k2} \\ & \xi_k \geq 0 \\ & k = 1, 2, \dots, K \end{aligned} \right\} \quad (24)$$

当需要对新样本进行分类时,只需将该样本代入到已经确定好的 K 个超平面表达式,新样本将被分入距离最远的超平面对应的类,即决策函数为

$$Label(x) = \arg \max_{j=1,2,\dots,K} \left((x^T w_j + b_j) / \sqrt{w_j^T w_j} \right) \quad (25)$$

注意,MBSVM 采用的是“max”的决策方式,这与 TWSVM 是完全不同的.

2.6.2 其他基于“多对一”策略的多分类孪生支持向量机

MBSVM 的二次规划问题通过松弛迭代法求解,得到的解往往精度不高.Ju 等人提出的非平行超平面多分类支持向量机(nonparallel hyperplanes support vector machine for multi class classification,简称 NHCMC)继承了 MBSVM 的思想,即,采用“多对一”结构和“max”的决策方式^[56].不同于 MBSVM,NHCMC 的模型可以通过序列最小优化算法求解,并且不必区分线性和非线性情况,核方法可以直接应用.NHCMC 算法与 MBSVM 一样,在样本类别多时,算法复杂度方面具有优势.值得注意的是:NHCMC 的二次规划问题与 SVM 的模型一样可以使用最小序列优化算法,因此,NHCMC 训练速度快,最小序列优化算法求解的结果精度较高,从而,NHCMC 算法较 MBSVM 稳定并且往往具有更好的分类准确率.该算法的不足在于,其理论依据有待进一步充实.

Xu 等人将孪生超球支持向量机与“多对一”多分类策略相结合,得到了孪生超球多分类支持向量机(twin hyper sphere support vector machine,简称 THKSVM)^[57].THKSVM 在训练过程中不需要进行矩阵的求逆操作,因此训练速度快,可以应用于大规模分类问题.此外,THKSVM 为每个类构造一个超球而不是超平面,能够更准确地获取和利用训练数据的信息,获取更好的分类面.

文献[58]提出了一种基于 1 范数正则项的孪生支持向量机,并使用多对一策略将该算法推广至多分类.该多分类方法通过求解线性方程组以解决分类问题,并且不需要进行大规模的矩阵求逆运算,因此训练速度快.缺点是分类精度一般,泛化能力有待提高.

针对 MBSVM 在处理数据存在交叉的多分类问题时效果不佳的情况,文献[59,60]在多生支持向量机中添加一个修正项,使得生成的超平面与对应类的样本之间距离尽量大而离其他类样本尽量远的同时,每个该类样本与超平面之间的距离都尽量相等.在对新样本进行预测阶段,不同于 MBSVM 将待分类点简单地分到离该点最远的超平面所对应的类,该算法首先判断带分类点离各分类超平面的距离是否落入对应的一个区间内,在满足此条件的分类超平面之间进行比较,最后确定分类结果.与 MBSM 相比,改进算法在处理数据存在交叉的多分类问题时效果更佳.

2.6.3 基于“多对一”策略的多分类孪生支持向量机的优缺点

基于 AVO 的 MTWSVMs 每次仅选取一个类作为负类,剩下的类都当作正类来构建超平面,得到的超平面离一个类远,而尽可能靠近其余类的样本.这样做的优点是时间复杂度低,约为 $O(m^3/K^2)$.基于 OVA 策略的 MTWSVMs 的时间复杂度为 $O(m^3K)$,因此类别数目 K 较大时,基于 AVO 方法的 MTWSVMs 在训练速度方面具有明显优势.此外,该类方法和 OVA MTWSVMs 一样,具有思想简单、容易实现的优点.

这类方法的缺点是分类精度不够高,其原因有两点.

- 其一是这种方法的约束条件少,因此容易陷入局部最优,应采取全局性更好的二次规划求解算法;
- 其二是被作为正类的那些类之间可能相距较远,很难使得超平面离它们同时都近.

综上所述,基于 AVO 方法的 MTWSVMs 适合分类类别 K 较大、但精度要求不太高的应用情景.

2.7 方法对比

本节的前 6 个小节已经对 6 类 MTWSVMs 的原理、数学模型、研究发展作了详细介绍,也简要分析了各类算法的优缺点.为了更加清晰地展现各类方法之间的优势和劣势,本小节对 MTWSVMs 进行更深入的对比分析.

首先对比分析基于 OVO 的 MTWSVMs 和基于 OVA 的 MTWSVMs,这两种是最早被提出的 MTWSVMs.这两种方法相比,基于 OVO 的 MTWSVMs 的优势是不会导致数据不平衡现象,分类准确率更高;而基于 OVA 的 MTWSVMs 由于每一步训练将多个类看作一个类,因此常伴随数据不平衡现象.与此同时,基于 OVA 的 MTWSVMs 还忽略了被看作一个类的那些类别之间的差别,因此分类准确率往往不高.但是 OVA MTWSVMs 算法的思想简单明了,容易实现,因此在所要处理的问题不是非常复杂的时候,领域专家往往还是会选择应用

OVA MTWSVMs.在训练速度上,OVO MTWSVMs 在训练集较大时训练速度要明显快于 OVA MTWSVMs.分类问题规模小时,建议选择使用 OVA MTWSVMs.

DAG MTWSVMs 是为了提升 OVO MTWSVMs 的决策速度而提出来的,并且解决了 OVO MTWSVMs 存在的拒分区域问题.但在决定整个算法执行速度的训练速度上,DAG MTWSVMs 与 OVO MTWSVMs 是相同的.此外,由于 DAG MTWSVMs 的 DAG 是一个层次结构,因此,DAG MTWSVMs 会存在累计误差,对算法性能造成一定的影响.总体上说,DAG MTWSVMs 和 OVO MTWSVMs 不论是在算法时间复杂度上还是在分类准确率方面都表现相当.因为 DAG MTWSVMs 的决策时间相对较短,因此,当待分类样本数量远多于训练样本时,应该选用 DAG MTWSVMs 代替 OVO MTWSVMs.

基于二叉树的 MTWSVMs 解决了 OVA MTWSVMs 和 OVO MTWSVMs 都存在的拒分问题.与 OVA MTWSVMs 相比,这类方法并非每步训练都需要使用所有训练数据,基于二叉树的 MTWSVMs 随着训练步骤的增加和二叉树层次的增加,训练的规模不断下降,总体的训练速度快于 OVA MTWSVMs.但是基于决策二叉树的 MTWSVMs 的实现较为复杂,因此,当所处理的问题并不复杂时,直接选用 OVA MTWSVMs 更好.与 OVO MTWSVMs 相比,基于二叉树的 MTWSVMs 仅需 $K-1$ 个子分类器,加之使用二叉树结构组织其子分类器,因此整个分类器结构更加简洁、稳定.与 DAG MTWSVMs 相比,基于决策二叉树的 MTWSVMs 的累计误差现象更加明显,因此,在构造二叉树结构的过程中要更加注重结构的优化,而 DAG MTWSVMs 的累计误差相对要少得多.在训练数据集给定之后,DAG MTWSVMs 的层次是确定的,而基于决策二叉树的 MTWSVMs 的层数不够确定,如果构造成偏二叉树,则会增加算法的训练时间.

基于“一对一对余”策略的多分类孪生支持向量机的每个子分类器以 $\{+1,0,-1\}$ 的形式标注样本,这使得每个子分类器在处理两类样本的同时识别其他类的样本,避免了 OVO 策略中对其他类样本的随机“投票”.因此,基于“一对一对余”策略的多分类孪生支持向量机在保持了 OVO MTWSVMs 的优点的同时获得了更加优秀的鲁棒性.但是,基于“一对一对余”策略的多分类孪生支持向量机为识别其他类的样本需要处理两类约束条件,模型求解难度大于 OVO MTWSVMs.与基于二叉树的 MTWSVMs 相比,基于“一对一对余”策略的多分类孪生支持向量机的优势是不存在累计误差,不需要结构优化.但是,基于“一对一对余”策略的多分类孪生支持向量机依然存在拒分区域,而且训练速度往往比基于二叉树的 MTWSVMs 要慢.

与 OVO MTWSVMs,DAG MTWSVMs 以及二叉树 MTWSVMs 相比,基于 AVO 的 MTWSVMs 和 OVA MTWSVMs 一样具有原理易懂、实现简单的优点.同时,和 OVA MTWSVMs 一样,基于 AVO 的 MTWSVMs 训练过程中也将多类看作一类,从而出现数据不平衡问题.与 OVA MTWSVMs 相比,基于 AVO 的 MTWSVMs 模型中约束条件更少,求解更加简单.训练速度上,基于 AVO 的 MTWSVMs 总是快于基于 OVA 的 MTWSVMs.基于 AVO 的 MTWSVMs 与基于 OVO 的 MTWSVMs 相比具有的优势是所需训练的子分类器个数少.因此,当分类问题包含许多类别时,基于 AVO 的 MTWSVMs 总体结构依然简单,而 OVO MTWSVMs 将会是一个包含许多子分类器的复杂分类系统.当所需处理的问题包含的类别很大时,基于 AVO 的 MTWSVMs 的时间复杂度要低于其他 5 类方法.因此,AVO MTWSVMs 适合包含类别多的分类问题.但是,基于 AVO 的 MTWSVMs 的分类准确率总体上要比其他 5 类方法要差.

3 算法实验比较

本节对各种算法进行数值实验对比.实验选取各类多分类孪生支持向量机中具有代表性的算法进行比较.为了保证实验的可靠性,本文均采用十交叉验证,即:将数据随机均分为 10 份,依次选取 1 份作为测试数据,其余作为训练数据,对每个数据集,各种算法都要进行 10 次训练和测试.运行环境为 2G 内存,CORE 2 处理器,2.19GHz 主频,Windows7 操作系统.全部算法都使用 Matlab2012a 实现和运行.本文选用 UCI 机器学习数据库中几个常用的数据集对算法进行测试,数据集的具体信息见表 1.采用 RBF 核作为核函数,因为大量研究表明,RBF 核函数具有比其他函数更佳的性能.表中分类精度为十交叉验证所得的平均结果.参数选择都使用网格搜索法,选取范围为 $-6\sim 6$,步长为 0.5.训练数据在用于训练分类器前已进行归一化.表 2 和表 3 中,ACC 表示对应

算法的平均分类准确率,STD 表示 10 次实验所得分类准确率的标准差.

Table 1 Description of UCI data sets for test
表 1 数据集详细信息

编号	数据集	样本个数	样本维数	类别数
1	Wine	178	13	3
2	Glass	214	9	6
3	Balance	625	4	3
4	Iris	150	4	3
5	Vowel	528	10	11
6	Landsat	2 000	36	6
7	Segment	2 310	18	7
8	Seeds	210	7	3
9	DNA	3 186	180	3
10	Optdigits	5 620	64	10

Table 2 Comparison of experimental results in linear case
表 2 线性情况下分类结果比较

编号	数据集	OVA TWSVM	OVO TWSVM	ODAG TWSVM	BT TWSVM	MBSVM	Twin KSVC
		ACC±STD	ACC±STD	ACC±STD	ACC±STD	ACC±STD	ACC±STD
1	Wine	95.16±1.20	98.41±2.23	98.41±3.47	98.39±5.24	95.77±1.41	96.67±3.42
2	Glass	45.91±4.81	49.66±5.87	46.29±7.80	46.96±5.36	<u>43.56±6.32</u>	45.32±3.46
3	Balance	83.52±3.63	85.54±0.79	85.32±2.12	85.76±1.51	85.66±0.86	<u>84.86±6.83</u>
4	Iris	88.33±2.32	<u>86.32±1.33</u>	87.32±1.65	86.37±3.68	88.67±3.26	86.76±4.39
5	Vowel	82.65±1.63	83.64±2.87	<u>80.37±3.53</u>	83.66±3.65	84.73±5.67	85.83±2.28
6	Landat	<u>76.16±1.37</u>	77.68±4.39	79.98±3.74	78.36±3.23	76.76±0.98	76.19±3.27
7	Segment	93.25±1.37	94.85±0.87	92.35±0.84	90.10±0.81	<u>89.89±0.67</u>	90.15±2.30
8	Seeds	95.19±3.65	96.65±2.62	96.65±1.36	<u>93.47±4.74</u>	95.72±2.69	94.65±2.65
9	DNA	85.82±3.86	86.99±1.66	84.18±2.92	83.21±3.28	<u>78.26±2.63</u>	80.23±7.28
10	Optdigits	97.75±1.02	96.13±1.37	97.24±1.32	<u>95.14±2.71</u>	96.28±2.64	96.25±1.97

Table 3 Comparison of experimental results in nonlinear case
表 3 非线性情况下分类结果比较

编号	数据集	OVA TWSVM	OVO TWSVM	ODAG TWSVM	BT TWSVM	MBSVM	Twin KSVC
		ACC±STD	ACC±STD	ACC±STD	ACC±STD	ACC±STD	ACC±STD
1	Wine	97.02±1.69	98.62±2.22	96.63±3.24	96.32±6.32	<u>95.93±0.71</u>	97.64±3.37
2	Glass	50.91±5.24	50.65±6.32	51.29±7.80	52.23±5.36	<u>43.33±6.32</u>	52.32±3.46
3	Balance	85.98±2.68	85.46±0.79	85.70±0.84	85.36±1.96	86.35±0.86	<u>84.20±5.06</u>
4	Iris	96.99±1.15	97.32±2.15	97.32±1.70	<u>96.87±1.19</u>	97.33±4.23	97.63±2.35
5	Vowel	86.45±4.78	82.64±2.18	<u>82.37±2.75</u>	83.89±3.75	84.57±5.69	83.68±1.37
6	Landat	<u>80.63±1.37</u>	82.68±3.55	82.62±3.43	85.36±2.78	80.76±1.11	84.25±3.98
7	Segment	<u>91.51±3.65</u>	93.67±1.41	93.54±1.34	94.15±1.61	<u>89.68±2.58</u>	91.68±1.65
8	Seeds	93.19±1.20	95.65±2.53	93.13±3.16	<u>89.35±5.98</u>	95.17±7.98	90.32±2.66
9	DNA	85.75±3.84	89.42±1.66	87.42±2.03	83.23±5.26	<u>78.82±3.46</u>	87.23±6.32
10	Optdigits	98.15±1.33	97.38±0.80	97.27±1.29	96.55±2.61	97.42±1.15	<u>95.47±3.94</u>

从表 2 可以看出:

- 线性情况下,在 UCI 数据集上,OVO TWSVM 总体分类精度是最高的,在 5 个数据集上获得了最佳结果.此外,OVO TWSVM 仅在一个数据集 Iris 上的分类准确率是最低的.但是 OVO TWSVM 在 Vowel 和 Optdigits 这两个样本类数超过 10 的数据集上表现不佳;
- 其次,ODAG TWSVM 的表现也不错,在数据集 Landat 上的分类精度高于其他算法,在 3 个数据集上取得了最佳分类结果,其分类结果接近 OVO TWSVM,在 Glass 和 Balance 上分类结果略好于 OVO TWSVM;
- 相对地,MBSVM 分类结果比较不够理想,在 3 个数据上的分类精度是所有方法中最差的;
- BT TWSVM 与 Twin KSVC 的分类精度相当,都在一个数据集上取得了最佳结果.

从表 3 可以看出:

- 在使用 RBF 核的情况下,在 UCI 数据集上,OVO TWSVM 总体分类精度依然是最好的,在 4 个数据集上

获得了最佳结果,特别是在数据集 DNA 上,OVO TWSVM 的分类精度明显高于其他算法.此外,OVO TWSVM 在所有数据集上的分类精度都不是最低的;

- MBSVM 相对分类结果比较差,在 4 个数据上的分类精度是所有方法中最差的,但也在一个数据上取得了最佳精度;
- BT TWSVM 和 Twin KSVC 都在两个数据集上取得了最佳结果,在一个数据集上取得了最差结果;
- ODAG TWSVM 虽然在任何数据集上都没有取得最佳结果,但其分类结果都接近 OVO TWSVM,在 Glass 和 Balance 上的分类结果好于 OVO TWSVM.

结合以上分析可得:总体而言,OVO TWSVM 和 ODAG TWSVM 分类性能是最佳的,而 MBSVM 的分类精度有待提升.

为了验证各种算法的复杂度,为实际应用中选择合适的多分类孪生支持向量机提供更为全面的参考,选取 Vowel 和 Pendigits 两个数据集作进一步的实验分析.多分类孪生支持向量机的训练时间主要与分类问题的训练样本数和样本类别数有关,为了测试各类多分类孪生支持向量机的时间复杂度与分类问题的训练样本数的关系,分别从 Pendigits 的各个类别中随机抽取 10%、20%、30%、40%、50%、60%、70%、80%、90% 和全部样本组成类别数相同但样本数目均匀递增的 10 个数据集.图 6 是各种算法在这 10 个数据集上的训练时间的折线图.实验所采用的是线性核,因为线性核的计算占用时间少,可以使实验中采集到的训练时间数据更加真实地反映各类多分类孪生支持向量机模型本身的时间复杂度.从图 6 中可以清晰地看出,随着样本数的增加,各类多分类孪生支持向量机训练时间的变化趋势.OVA TWSVM 和 Twin KSVC 的训练时间随样本数的增加而增加的速度最快,而 MBSVM 则最慢.结合表 2 和表 3 中分类准确率的测试结果可以看出:实际应用中,如果对分类准确率的要求并不苛刻而训练样本数量又较大,宜采用 MBSVM.如果希望尽可能地提升分类结果,则可选用 OVO TWSVM,因为 OVO TWSVM 的分类准确率总体是最高的,并且时间复杂度也不是很高.

接着,用包含 11 个类的 Vowel 测试各种算法训练时间随样本数目增加的变化情况以分析各算法时间复杂度随样本类别数的变化情况.取 Vowel 中的第 1 类~第 4 类作为第 1 类,第 5 类~第 8 类作为第 2 类,第 9 类~第 11 类作为第 3 类构成一个三分类数据集,取 Vowel 中的第 1 类~第 3 类作为第 1 类,第 4 类~第 6 类作为第 2 类,第 7 类~第 9 类作为第 3 类,第 10 类~第 11 类作为第 4 类构成一个四分类数据集,按照类似做法得到五分类数据集、六分类数据集、七分类数据集、八分类数据集、九分类数据集、十分类数据集和十一分类数据集.这样得到的 9 个数据集包含的数据样本个数是一样的而样本类别数不同.图 7 是各种算法在这 10 个数据集上的训练时间的折线图.

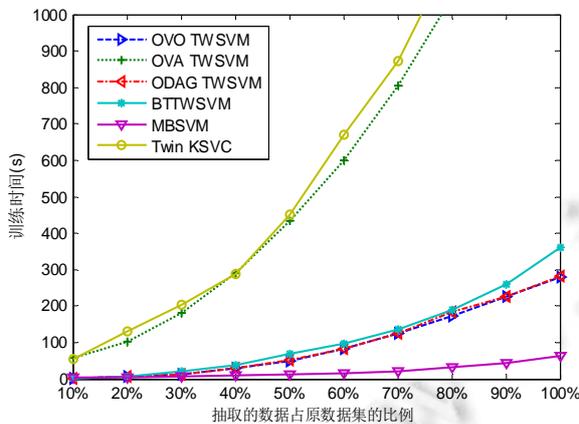


Fig.6 Relationship between training time and the number of samples

图 6 训练时间与训练样本数的关系

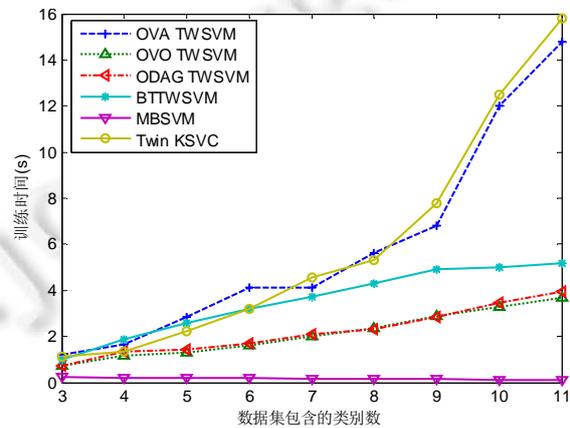


Fig.7 Relationship between training time and the class number

图 7 训练时间与训练样本类别数的关系

可以看出:当类别数为3和4时,6种算法的训练速度没有很大的差别.随着训练样本类别数的增加,MBSVM的训练时间略有减少.OVA TWSVM和Twin KSVC的训练时间近似于线性增长.OVO TWSVM和Twin KSVC训练时间的增加也并不明显.结合表2和表3中分类准确率的测试结果可以看出:实际应用中,如果所面临的问题包含类别数较多而分类准确率要求不是很高,则采用MBSVM是最为快速的.若要兼顾速度和分类结果,宜选用OVO TWSVM、ODAG TWSVM或BT TWSVM.仅在问题规模小、包含的类别数少时,建议使用OVA TWSVM和Twin KSVC.

4 总 结

将TWSVM推广到多分类是TWSVM研究中的一个重要内容,本文对近年来MTWSVMs的研究进展进行了一个总结.已有的研究主要可分为基于“一对一”策略、“一对多”策略、DAG结构、决策二叉树结构、“一对一对余”策略和“多对一”这6种策略的MTWSVMs.各种策略有着各自的优缺点.本文以多分类TWSVM采取的策略,也即组织子分类器的方式为依据,分门别类地回顾了已有的MTWSVMs算法,分析了各类算法的优缺点,通过数值实验,对比了各类算法中代表性的算法.MTWSVMs的研究虽然已经取得了不少成果,但仍然存在许多问题需要继续完善.本文认为,还需在以下几个研究方向上加以完善和改进.

(1) 理论研究.MTWSVMs的研究大多数是实验驱动的,理论基础不够深厚.完善MTWSVMs的基础理论系统是MTWSVMs发展的基石,因此特别值得关注;

(2) 研究MTWSVMs的并行性.现在是大数据时代,对于比较大的数据,几乎所有MTWSVMs的处理效率都明显不足.并行化是快速解决大型且复杂计算问题的重要方法.对MTWSVMs进行并行化,也许是未来提升MTWSVMs处理大数据能力和效率的发展方向.MTWSVMs是否适合并行化计算、如何合理分割多计算过程以便实现并行化,都是未解决的问题;

(3) 拓宽MTWSVMs的应用领域.目前,MTWSVMs的应用仍有限,大部分优秀的MTWSVMs算法都仍处于理论研究阶段.为了推动多分类TWSVMs的进一步发展,需要领域专家结合专业知识拓宽这些算法的应用领域.只有实际的应用价值,才能让MTWSVMs得到进一步的发展.

References:

- [1] Cortes C, Vapnik VN. Support vector networks. *Machines Learning*, 1995,20(2):273–297. [doi: 10.1023/A:1022627411411]
- [2] Vapnik VN. *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1998. [doi: 10.1007/978-1-4757-3264-1]
- [3] Khemchandni R, Chandra S. Twin support vector machines for pattern classification. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2007,29(5):905–910. [doi: 10.1109/TPAMI.2007.1068]
- [4] Ding SF, Yu JZ, Qi BJ, Huang HJ. An overview on twin support vector machines. *Artificial Intelligence Review*, 2014,42(2): 245–252. [doi: 10.1007/s10462-012-9336-0]
- [5] Kumar MA, Gopal M. Least squares twin support vector machines for pattern classification. *Expert Systems with Applications*, 2009,36(4):7535–7543. [doi: 10.1016/j.eswa.2008.09.066]
- [6] Peng XJ, Xu D. Twin Mahalanobis distance based support vector machines for pattern recognition. *Information Sciences*, 2012,200: 22–37. [doi: 10.1016/j.ins.2012.02.047]
- [7] Shao YH, Chen WJ, Zhang JJ, Wang Z, Deng NY. An efficient weighted Lagrangian twin support vector machine for imbalanced data classification. *Pattern Recognition*, 2014,47(9):3158–3167. [doi: 10.1016/j.patcog.2014.03.008]
- [8] Ding SF, Zhang XK, Yu JZ. Twin support vector machines based on fruit fly optimization algorithm. *Int'l Journal of Machine Learning and Cybernetics*, 2016,7(2):193–203. [doi: 10.1007/s13042-015-0424-8]
- [9] Kumar MA, Gopal M. Application of smoothing technique on twin support vector machines. *Pattern Recognition Letters*, 2008, 29(13):1842–1848. [doi: 10.1016/j.patrec.2008.05.016]
- [10] Huang HJ, Ding SF, Shi ZZ. Smooth CHKS twin support vector regression. *Journal of Computer Research & Development*, 2015, 52(3):561–568 (in Chinese with English abstract). [doi: 10.7544/issn1000-1239.2015.20131444]

- [11] Ding SF, Huang HJ, Shi ZZ. Weighted smooth CHKS twin support vector machines. *Ruan Jian Xue Bao/Journal of Software*, 2014, 24(11):2548–2557 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4475.htm> [doi: 10.3724/SP.J.1001.2013.04475]
- [12] Ding SF, Huang HJ, Xu XZ, Wang J. Polynomial smooth twin support vector machines. *Applied Mathematics and Information Sciences*, 2014,8(4):2063–2071. [doi: 10.12785/amis/080465]
- [13] Peng XJ, Kong LY, Chen DJ. Improvements on twin parametric margin support vector machine. *Neurocomputing*, 2015,151(2): 857–863. [doi: 10.1016/j.neucom.2014.10.010]
- [14] Xie XJ, Sun SL. Multitask centroid twin support vector machines. *Neurocomputing*, 2015,149(2):1085–1091. [doi: 10.1016/j.neucom.2014.07.025]
- [15] Gu B, Sheng VS, Wang ZJ, Ho D, Osman S, Li S. Incremental learning for ν support vector regression. *Neural Networks*, 2015,67: 140–150. [doi: 10.1016/j.neunet.2015.03.013]
- [16] Peng X, Xu D. A twin hypersphere support vector machine classifier and the fast learning algorithm. *Information Sciences*, 2013, 221:12–27. [doi: 10.1016/j.ins.2012.09.009]
- [17] Mehrkanoon S, Huang XL, Johan S. Non parallel support vector classifiers with different loss functions. *Neurocomputing*, 2014, 143:294–301. [doi: 10.1016/j.neucom.2014.05.063]
- [18] Hua XP, Ding SF. Weighted least squares projection twin support vector machines with local information. *Neurocomputing*, 2015, 160:228–237. [doi: 10.1016/j.neucom.2015.02.021]
- [19] Gu B, Sheng VS. A robust regularization path algorithm for ν support vector classification. *IEEE Trans. on Neural Networks and Learning Systems*, 2016. [doi: 10.1109/TNNLS.2016.2527796]
- [20] Tanveer M, Shubham K, Aldhaifallah M, Ho SS. An efficient regularized K nearest neighbor based weighted twin support vector regression. *Knowledge Based Systems*, 2016,94:70–87. [doi: 10.1016/j.knosys.2015.11.011]
- [21] Shao YH, Wang Z, Chen WJ, Deng NY. A regularization for the projection twin support vector machine. *Knowledge Based System*, 2013,37:203–210. [doi: 10.1016/j.knosys.2012.08.001]
- [22] Nasiri JA, Charkari NM, Mozafari K. Energy based model of least squares twin support vector machines for human action recognition. *Signal Processing*, 2014,104(6):248–257. [doi: 10.1016/j.sigpro.2014.04.010]
- [23] Yuan CS, Sun XM, Lü R. Fingerprint liveness detection based on multi scale LPQ and PCA. *China Communications*, 2016,13(7): 60–65. [doi: 10.1109/CC.2016.7559076]
- [24] Khemchandani R, Sharma S. Robust least squares twin support vector machine for human activity recognition. *Applied Soft Computing*, 2016,47:33–46. [doi: 10.1016/j.asoc.2016.05.025]
- [25] Al Dhaifallah M. Twin support vector machine method for identification of wiener models. *Mathematical Problems in Engineering*, 2015,2015(2):1–7. [doi: 10.1155/2015/125868]
- [26] Tomar D, Agarwal S. A comparison on multi class classification methods based on least squares twin support vector machine. *Knowledge Based Systems*, 2015,81:131–147. [doi: 10.1016/j.knosys.2015.02.009]
- [27] Tomar D, Agarwal S. Twin support vector machine: A review from 2007 to 2014. *Egyptian Informatics Journal*, 2015,20(1):55–69. [doi: 10.1016/j.eij.2014.12.003]
- [28] Xie JY, Hone K, Xie WX, Gao XB, Shi Y, Liu XH. Extending twin support vector machine classifier for multi category classification problems. *Intelligent Data Analysis*, 2013,17(4):649–664. [doi: 10.3233/IDA-130598]
- [29] Wang Z, Chen J, Qin M. Non parallel planes support vector machine for multi class classification. In: *Proc. of the 2010 Int'l Conf. on Logistics Systems and Intelligent Management*. 2010. 581–585. [doi: 10.1109/ICLSIM.2010.5461354]
- [30] Cong HH, Yang CF, Pu XR. Efficient speaker recognition based on multi class twin support vector machines and GMMs. In: *Proc. of the IEEE Conf. on Robotics, Automation and Mechatronics (RAM)*. Chengdu, 2008. 348–352. [doi: 10.1109/RAMECH.2008.4681433]
- [31] Chen SG, Xu J. Least squares twin support vector machine for multi class classification. *Int'l Journal of Database Theory & Application*, 2015,8(5):65–76. [doi: 10.14257/ijdt.2015.8.5.06]
- [32] Tomar D, Agarwal S. Multiclass least squares twin support vector machine for pattern classification. *Int'l Journal of Database Theory and Application*, 2015,8(6):285–302. [doi: 10.14257/ijdt.2015.8.6.26]

- [33] Li CN, Huang YF, Wu HJ, Shao YH, Yang ZM. Multiple recursive projection twin support vector machine for multi class classification. *Int'l Journal of Machine Learning and Cybernetics*, 2016,7(5):729–740. [doi: 10.1007/s13042-014-0289-2]
- [34] Yang ZM, Wu HJ, Li CN, Shao YH. Least squares recursive projection twin support vector machine for multi class classification. *Int'l Journal of Machine Learning & Cybernetics*, 2016,7(3):411–426. [doi: 10.1007/s13042-015-0394-x]
- [35] Tomar D, Agarwal S. An effective weighted multi class least squares twin support vector machine for imbalanced data classification. *Int'l Journal of Computational Intelligence Systems*, 2015,8(4):761–778. [doi: 10.1080/18756891.2015.1061395]
- [36] Shao YH, Chen WJ, Wang Z, Li CN, Deng NY. Weighted linear loss twin support vector machine for large scale classification. *Knowledge Based Systems*, 2015,73:276–288. [doi: 10.1016/j.knsys.2014.10.011]
- [37] Tomar D, Agarwal S. Multi class twin support vector machine for pattern classification. In: *Proc. of the 3rd Int'l Conf. on Advanced Computing, Networking and Informatics*, Vol.43. Springer-Verlag, 2016. 97–110. [doi: 10.1007/978-81-322-2538-6_11]
- [38] Platt JC, Cristianini N, Shawe Taylor J. Large margin dags for multiclass classification. *Advances in Neural Information Processing Systems*, 2000,12(3):547–553.
- [39] Gu HB, Niu B, Gao ZX. A directed acyclic graph algorithm for multi class classification based on twin support vector machine. *Journal of Information & Computational Science*, 2014,11(18):6529–6536. [doi: 10.12733/jics20105038]
- [40] Tomar D, Agarwal S. Direct acyclic graph based multi class twin support vector machine for pattern classification. In: *Proc. of the ACM Ikdd Conf. 2015*. 80–85. [doi: 10.1145/2732587.2732598]
- [41] Chen J, Ji GR. Multi class LSTSVM classifier based on optimal directed acyclic graph. In: *Proc. of the Int'l Conf. on Computer and Automation Engineering*. IEEE, 2010. 100–104. [doi: 10.1109/ICCAE.2010.5452037]
- [42] Zhang XK, Ding SF, Sun TF. Multi class LSTMSVM based on optimal directed acyclic graph and shuffled frog leaping algorithm. *Int'l Journal of Machine Learning and Cybernetics*, 2016,7(2):241–251. [doi: 10.1007/s13042-015-0435-5]
- [43] Li K, Huang WX, Huang ZH. Multi sensor detected object classification method based on support vector machine. *Journal of Zhejiang University (Engineering Science Edition)*, 2013,47(1):15–22 (in Chinese with English abstract). [doi: 10.3785/j.issn.1008-973X.2013.01.003]
- [44] Xu YT, Guo R, Wang LS. A twin multi class classification support vector machine. *Cognitive Computation*, 2013,5(4):580–588. [doi: 10.1007/s12559-012-9179-7]
- [45] Nasiri JA, Charkari NM, Jalili S. Least squares twin multi class classification support vector machine. *Pattern Recognition*, 2015, 48(3):984–992. [doi: 10.1016/j.patcog.2014.09.020]
- [46] Khemchandani R, Pal A. Multi category Laplacian least squares twin support vector machine. *Applied Intelligence*, 2016. [doi: 10.1007/s10489-016-0770-6]
- [47] Chu MX, Gong RF, Wang AN. Strip steel surface defect classification method based on enhanced twin support vector machine. *Trans. of the Iron & Steel Institute of Japan*, 2014,54(1):119–124. [doi: 10.2355/isijinternational.54.119]
- [48] Chu M, Wang A, Gong R, Sha M. Multi class classification methods of enhanced LS TWSVM for strip steel surface defects. *Int'l Journal of Iron and Steel Research*, 2014,21(2):174–180. [doi: 10.1016/S1006-706X(14)60027-3]
- [49] Khemchandani R, Saigal P. Color image classification and retrieval through ternary decision structure based multi category TWSVM. *Neurocomputing*, 2015,165:444–455. [doi: 10.1016/j.neucom.2015.03.074]
- [50] Nie PP, Li Z, Liu LL. Application of multi class classification algorithm based on twin support vector machine in intrusion detection. *Journal of Computer Applications*, 2013,33(2):426–429 (in Chinese with English abstract). [doi: 10.3724/SP.J.1087.2013.00426]
- [51] Khemchandani R, Pal A. Tree based multi category Laplacian TWSVM for content based image retrieval. *Int'l Journal of Machine Learning & Cybernetics*, 2016. [doi: 10.1007/s13042-016-0493-3]
- [52] Xie JY, Zhang BQ, Wang WZ. A partial binary tree algorithm for multiclass classification based on twin support vector machines. *Journal of Nanjing University (Natural Sciences)*, 2011,47(4):354–363 (in Chinese with English abstract).
- [53] Li QL, Wang JJ, Gao BB. Multi classification algorithm for twin binary tree support vector machine. *Journal of Southwest University (National Science Edition)*, 2014,36(7):162–168 (in Chinese with English abstract). [doi: 10.13718/j.cnki.xdzk.2014.07.026]

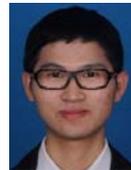
- [54] Shao YH, Chen WJ, Huang WB, Yang ZM, Deng NY. The best separating decision tree twin support vector machine for multi class classification. *Procedia Computer Science*, 2013,17:1032–1038. [doi: 10.1016/j.procs.2013.05.131]
- [55] Yang ZX, Shao YH, Zhang XS. Multiple birth support vector machine for multi class classification. *Neural Computing and Applications*, 2013,22(Suppl 1):S153–S161. [doi: 10.1007/s00521-012-1108-x]
- [56] Ju XC, Tian YJ, Liu DL, Qi ZQ. Nonparallel hyperplanes support vector machine for multi class classification. *Procedia Computer Science*, 2015,51(1):1574–1582. [doi: 10.1016/j.procs.2015.05.287]
- [57] Xu YT, Guo R. A twin hyper sphere multi class classification support vector machine. *Journal of Intelligent and Fuzzy Systems*, 2014,27(4):1783–1790. [doi: 10.3233/IFS-141145]
- [58] Li DW, Tian YJ. Twin support vector machine in linear programs. *Procedia Computer Science*, 2014,29:1770–1778. [doi: 10.1016/j.procs.2014.05.162]
- [59] Zhang XK, Ding SF, Xue Y. An improved multiple birth support vector machine for pattern classification. *Neurocomputing*, 2017, 225:119–128. [doi: 10.1016/j.neucom.2016.11.006]
- [60] Ding SF. *Twin Support Vector Machine: Algorithm, Theory and Extension*. Beijing: Science Press, 2017 (in Chinese).

附中文参考文献:

- [10] 黄华娟,丁世飞,史忠植.光滑 CHKS 孪生支持向量回归机. *计算机研究与发展*,2015,52(3):561–568. [doi: 10.7544/issn1000-1239.2015.201314444]
- [11] 丁世飞,黄华娟,史忠植.加权光滑 CHKS 孪生支持向量机. *软件学报*,2013,24(11):2548–2557. <http://www.jos.org.cn/1000-9825/4475.htm> [doi: 10.3724/SP.J.1001.2013.04475]
- [43] 李侃,黄文雄,黄忠华.基于支持向量机的多传感器探测目标分类方法. *浙江大学学报(工学版)*,2013,47(1):15–22. [doi: 10.3785/j.issn.1008-973X.2013.01.003]
- [50] 聂盼盼,臧洌,刘雷雷.基于对支持向量机的多类分类算法在入侵检测中的应用. *计算机应用*,2013,33(2):426–429. [doi: 10.3724/SP.J.1087.2013.00426]
- [52] 谢娟英,张兵权,汪万紫.基于双支持向量机的偏二叉树多类分类算法. *南京大学学报(自然科学)*,2011,47(4):354–363.
- [53] 李秋林,王建军,高斌斌.孪生二叉树支持向量多分类机算法. *西南大学学报(自然科学版)*,2014,36(7):162–168. [doi: 10.13718/j.cnki.xdzk.2014.07.026]
- [60] 丁世飞. *孪生支持向量机:理论、算法及其拓展*.北京:科学出版社,2017.



丁世飞(1963—),男,山东青岛人,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为人工智能与模式识别,机器学习,数据挖掘,大数据智能分析.



张谢喆(1992—),男,硕士,主要研究领域为机器学习,支持向量机,孪生支持向量机.



张健(1990—),男,博士生,主要研究领域为机器学习与数据挖掘,支持向量机,深度神经网络,深度学习.



安悦瑄(1993—),女,硕士生,主要研究领域为机器学习与强化学习,支持向量机,孪生支持向量机.