

## 基于共享矢量链的多任务概念漂移分类方法

史 炎 中<sup>1,2</sup>, 邓 赵 红<sup>1,3</sup>, 钱 鹏 江<sup>1</sup>, 王 士 同<sup>1†</sup>

(1. 江南大学 数字媒体学院, 江苏 无锡 214122; 2. 无锡职业技术学院 物联网学院, 江苏 无锡 214121;  
3. 江南大学 江苏省媒体设计与软件技术重点实验室, 江苏 无锡 214122)

**摘 要:** 增强型时间自适应支持向量机在针对单一概念漂移问题时展现出了良好效果,但是无法协同求解多个概念漂移问题. 然而,在很多应用场景中,有时会包含数个具有内在相关性的非静态数据集,它们各自的分类模型应充分考虑这种关联. 为了反映出各概念漂移分类模型之间的相关部分,提出共享矢量链的概念,并开发面向多任务概念漂移问题的共享矢量链支持向量机(SVC-SVM). 在模拟数据集及气体传感器阵列漂移数据集上的实验结果显示,协同求解多个具有相关性的概念漂移问题能够有效提升各自的泛化能力.

**关键词:** 多任务; 概念漂移; 支持向量机; 共享矢量链

**中图分类号:** TP181

**文献标志码:** A

## Multi-task concept drift classification method based on shared vector chain

SHI Ying-zhong<sup>1,2</sup>, DENG Zhao-hong<sup>1,3</sup>, QIAN Peng-jiang<sup>1</sup>, WANG Shi-tong<sup>1†</sup>

(1. School of Digital Media, Jiangnan University, Wuxi 214122, China; 2. School of Internet of Things, Wuxi Institute of Technology, Wuxi 214121, China; 3. Jiangsu Key Laboratory of Media Design and Software Technology, Jiangnan University, Wuxi 214122, China)

**Abstract:** The improved time adaptive supported vector machine performs well for a single concept drift problem in theoretical analysis and experimental studies, but cannot solve multiple concept drift problem. However, several parallel non-stationary datasets with inherent correlation are sometimes involved in many practical scenarios, whose respective classification models should take such cohesions into account. In this paper, the concept of shared supported vector chain is proposed to reflect the common vectors in each classification model for each nonstationary dataset and accordingly, a multi-task concept drift classification approach, called shared vector chain supported vector machines(SVC-SVM), is developed. Experimental results on synthetic datasets and gas sensor array drift dataset show the effectiveness of the proposed algorithm when solving several correlative concept drift problems together.

**Keywords:** multi task; concept drift; supported vector machine; share vector chain

## 0 引 言

随着通信和计算机信息技术的发展,每天会产生大量金融市场、交通流量、网络监控、超市零售、电信服务、电子商务等方面的数据,这些数据具有高速、持续、动态等特征. 由于数据特征随着时间不断变化,针对这些非静态数据的分类模型和聚类模型也随着时间而缓慢漂移,称为概念漂移<sup>[1-2]</sup>. 对概念漂移的研究已在理论上<sup>[3-4]</sup>及交通流量预测<sup>[5]</sup>、超市客户分析<sup>[6]</sup>、气体传感器阵列漂移<sup>[7]</sup>等应用场合取得了良好的效果. 在这些应用中,特定时刻获取的数据量通常并不充分,需要利用相近时刻的相关数据信

息来建立合适的当前时刻模型. 以往针对概念漂移分类所做的工作大多是基于滑动窗算法<sup>[8-12]</sup>的思路, Grinblat等<sup>[13]</sup>借鉴Cramer兼顾局部优化与全局优化的策略<sup>[14]</sup>,提出了时间自适应支持向量机(TA-SVM)方法来求解渐变的子分类器. 时间自适应核心向量机<sup>[15]</sup>方法在分类性能与TA-SVM方法相当的情况下,适用于对大规模数据集进行快速求解. 2015年,Shi等<sup>[16]</sup>提出了增强型时间自适应支持向量(ITA-SVM)方法,在提高分类性能的同时,从理论上保证了其对偶为凸二次规划问题.

生活中有很多概念漂移问题具有相关性,如某个

收稿日期: 2017-03-23; 修回日期: 2017-07-19.

基金项目: 国家自然科学基金项目(61572236,61300151); 江苏省产学研前瞻性研究项目(BY2016023-01).

责任编委: 程龙.

作者简介: 史炎中(1970—),男,副教授,博士,从事模式识别、图像处理等研究; 邓赵红(1981—),男,教授,博士生导师,从事人工智能、神经模糊计算等研究.

†通讯作者. E-mail: wxwangst@aliyun.com

气体传感器阵列上对多种气体的测定数据(传感器的老化会使测定数据发生漂移)、相邻城市的天气情况、相近街区的交通流量等.将它们视作孤立的概念漂移问题,并对其进行独立建模,可能并不是最佳选择.由于单个问题中包含的信息通常并不充分,对多个相关概念漂移问题同时建模,挖掘其他问题中的有效信息,能够对建模起到有益的补充.近期的研究也表明,较之于单独学习各个子任务,对多个相关子任务同时学习能够有效地提升预测性能<sup>[14,17]</sup>.因此,研究出一种能够对多个相关概念漂移问题协同建模求解的方法变得很有必要.

当前面向概念漂移问题的研究都只基于单个非静态数据集进行建模,国内外几乎看不到对多个相关概念漂移问题同时研究的文献.本文根据正则化多任务学习(rMTL)方法和ITA-SVM方法中共享矢量的思路,借鉴TA-SVM方法中同时求解随时间变化的多个子分类器的策略,提出共享矢量链支持向量机(SVC-SVM)方法.所提方法具有以下特点:1)目标函数中包含多个概念漂移问题的表达形式,即对多个概念漂移问题协同建模;2)其对偶问题等价于在核空间上的另一个SVM对偶问题,可使用传统方法直接求解所对应的二次规划问题;3)通过协同求解,能够从相关数据集中汲取有效信息,较之以往独立求解单个概念漂移问题的方法,泛化能力有较大的提高.

### 1 相关方法

围绕多任务学习及概念漂移学习,学者们已经开展了很多研究.rMTL方法利用共享矢量同时求解多个子任务,是其他多任务学习方法的基础;TA-SVM方法通过局部优化与全局优化的折衷,同时求解渐变漂移的多个子分类器,改变了以往概念漂移研究中的局部求解模式;ITA-SVM方法利用共享矢量技巧对TA-SVM方法进行改进,从理论上证明了其对偶为凸二次规划问题,并在保证分类性能的前提下达到接近线性的算法时间复杂度,适用于对大规模数据集进行快速求解.

### 1.1 多任务学习方法

多任务学习是同时求解  $K$  个任务,并利用相关任务中的信息来提高学习所得模型的泛化能力.Evgeniou等<sup>[14]</sup>在2004年提出了rMTL方法,在保持各个子学习机局部优化的同时,使多个学习机之间的全局差异最小化.这时,多个子任务的决策模型共享核空间中的矢量  $w_0$ ,第  $k$  个子任务的模型  $w_k$  由共享矢量  $w_0$  与偏差  $v_k$  构成,即  $w_k = w_0 + v_k$ .rMTL方法的目标函数如下:

$$\min \frac{1}{2} \|w_0\|^2 + \frac{\lambda}{2} \sum_{k=1}^K \|v_k\|^2 + C \sum L(f, x_i, y_i). \quad (1)$$

其中:第1项为正则化项;第2项用于保证各子任务的相似性, $\lambda$ 为对局部优化与全局优化进行权衡的因子;第3项为损失函数.rMTL方法使用了共享矢量技巧,很多后续多任务学习方法都借鉴了rMTL的思想.

### 1.2 单任务概念漂移分类方法

TA-SVM<sup>[13]</sup>方法及ITA-SVM<sup>[16]</sup>方法针对的是传统的单任务概念漂移分类.TA-SVM方法在考虑各子分类器局部优化的同时,还考虑到子分类器的变化应该比较平稳,因此对各子分类器之间的差异进行了约束,其基本思想可由下式表示:

$$\min \sum_{t=1}^T \text{Risk}(f_t) + \lambda \sum_{t=1}^{T-1} d(f_{t+1}, f_t). \quad (2)$$

其中:第1项为局部优化项;第2项为全局优化项, $\lambda$ 为对局部优化与全局优化进行权衡的因子, $f_t$ 为第  $t$  个子分类器.

### 2 SVC-SVM方法及其对偶

为了能够进一步挖掘出相关非静态数据集中蕴含的有效信息,需要协同求解多个分类模型.假定现有  $K$  个相关非静态数据集,每个非静态数据集中的数据由  $T$  个按时间顺序采集的子数据集组成,每个子数据集中的数据量为  $m$  个.将所有数据合并记为数据集  $\{(x_i, y_i) | i = 1, 2, \dots, n\}$ ,  $n = K \times T \times m$ .

本文提出面向多任务概念漂移分类的共享矢量链支持向量机方法,该方法的建模原理可用图1描述,图中任务数  $K = 3$ .

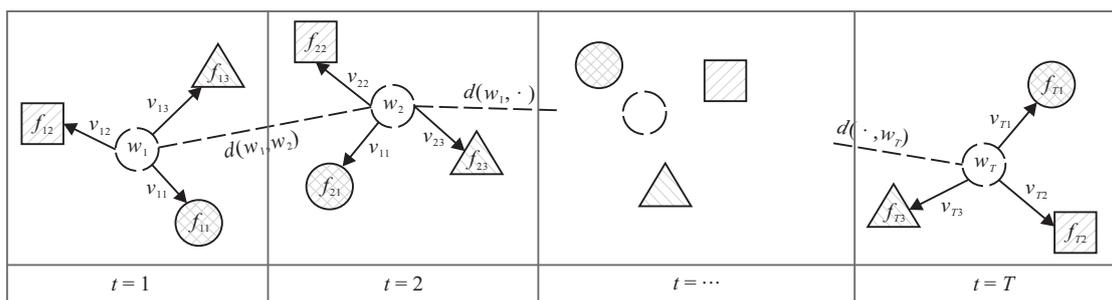


图1 多任务概念漂移SVC-SVM方法的建模原理

图1中:圆形、矩形、三角形分别表示3个概念漂移分类任务,  $f_{tk}(t = 1, 2, \dots, T; k = 1, 2, 3)$  表示任务  $k$  在  $t$  时刻的分类模型,  $\mathbf{w}_t$  表示第  $t$  个时刻3个任务的共享矢量,  $\mathbf{v}_{tk} = f_{tk} - \mathbf{w}_t (k = 1, 2, 3)$  表示第  $t$  时刻的分类模型与相应共享矢量之间的差异。

图1中包含有两类相关性: 1) 第  $t$  时刻的各子任务  $f_{tk}$  相似于共享矢量  $\mathbf{w}_t$ ; 2) 相邻时刻的共享矢量  $\mathbf{w}_{t+1}$  与  $\mathbf{w}_t$  相似。

SVC-SVM方法的原理可以通过下式表示:

$$\min \frac{1}{2T} \sum_{t=1}^T \|\mathbf{w}_t\|^2 + \frac{\lambda}{2T} \sum_{t=1}^{T-1} \|\mathbf{w}_{t+1} - \mathbf{w}_t\|^2 + \frac{\gamma}{2} \sum_{t=1}^T \sum_{k=1}^K \|\mathbf{v}_{tk}\|^2 + C \sum_{i=1}^n L(f_{tk(i)}, \mathbf{x}_i, y_i). \quad (3)$$

其中:  $\min \sum_{t=1}^T \|\mathbf{w}_t\|^2$  为正则化项;  $\min \sum_{t=1}^{T-1} \|\mathbf{w}_{t+1} - \mathbf{w}_t\|^2$  通过约束相邻时刻共享矢量的差异使共享矢量链的变化尽量平稳,  $\lambda$  为约束模型平稳漂移的参数;  $\min \sum_{t=1}^T \sum_{k=1}^K \|\mathbf{v}_{tk}\|^2$  使各子任务在同一时刻的模型尽量相似, 这是协同求解多个概念漂移问题的关键, 权衡因子  $\gamma$  表示多个任务间的相关程度;  $L(f_{tk(i)}, \mathbf{x}_i, y_i)$  为损失函数, 本文参照文献[16], 选取二次损失函数。

参照文献[13,16], 本文采用支持向量机对多任务概念漂移分类问题进行建模, 并将向量的偏置量  $b$  引入目标函数中, 记第  $t$  时刻共享矢量的偏置量为  $b_t$ , 各子任务模型中偏置量与  $b_t$  的差异记为  $d_{tk}$ , 则共享矢量记为  $(\mathbf{w}_t, b_t)$ , 第  $t$  时刻任务  $k$  的决策模型为  $f_{tk} = (\mathbf{w}_t + \mathbf{v}_{tk}, b_t + d_{tk})$ . 参照式(3)中的基本思想, SVC-SVM方法的目标函数为

$$\begin{aligned} \min_{\mathbf{w}, \mathbf{v}, b, d} & \frac{1}{2T} \sum_{t=1}^T (\|\mathbf{w}_t\|^2 + b_t^2) + \\ & \frac{\lambda}{4T} \sum_{t=1}^T \sum_{s=1}^T Q_{ts} (\|\mathbf{w}_t - \mathbf{w}_s\|^2 + (b_t - b_s)^2) + \\ & \frac{\gamma}{2} \sum_{t=1}^T \sum_{k=1}^K (\|\mathbf{v}_{tk}\|^2 + d_{tk}^2) + \frac{C}{2} \sum_{i=1}^n \xi_i^2; \\ \text{s.t. } & y_i ((\mathbf{w}_t + \mathbf{v}_{tk(i)})^T \varphi(\mathbf{x}_i) + (b_t + d_{tk(i)})) \geq 1 - \xi_i, \\ & i = 1, 2, \dots, n. \end{aligned} \quad (4)$$

其中: 当且仅当  $|t - s| = 1$  时,  $Q_{ts} = 1$ , 否则  $Q_{ts} = 0$ ;  $\mathbf{v}_{tk(i)}$  和  $d_{tk(i)}$  表示第  $i$  个样本属于任务  $k$  中的第  $t$  个子数据集. 参照文献[13,16]的推导方式, 可以得到 SVC-SVM方法的对偶形式为

$$\max_{\alpha} \alpha^T \mathbf{1} - \frac{1}{2} \alpha^T \mathbf{H} \alpha; \text{ s.t. } \alpha \geq 0. \quad (5)$$

其中  $\mathbf{H}$  是扩展核矩阵, 其详细表达形式及公式推导

过程稍后给出. 由式(5)可知, SVC-SVM方法对多个概念漂移问题同时建模, 虽然给原始问题带来了一定的复杂性, 但其对偶问题仍相当于核空间中的另一个 SVM问题, 可以用普通方法来求解此二次规划问题。

为了给出式(4)的对偶形式, 首先介绍相关符号:  $\mathbf{P}$  为  $T \times n$  矩阵, 用于标识第  $j$  个点是否为第  $t$  个时间段的数据, 当且仅当  $j \in p_t$  时,  $P_{tj} = 1$ , 否则  $P_{tj} = 0$ ;  $\mathbf{R}$  为  $TK \times n$  矩阵, 用于标识第  $j$  个点是否属于第  $tk$  个子数据集, 当且仅当  $j \in r_{tk}$  时,  $R_{tk,j} = 1$ , 否则  $R_{tk,j} = 0$ ;  $\mathbf{Q}$  为指示各共享向量之间相关性的  $T \times T$  矩阵, 实际应用中只考虑直接相邻的各共享向量. 文献[13]中提供了对于单个数据流的求解过程, 与本文方法类似, 但相对简单。

下面, 求解式(4)的对偶问题。

$$\begin{aligned} J = & \frac{1}{2T} \sum_{t=1}^T (\|\mathbf{w}_t\|^2 + b_t^2) + \frac{\lambda}{4T} \sum_{t=1}^T \sum_{s=1}^T Q_{ts} (\|\mathbf{w}_t - \\ & \mathbf{w}_s\|^2 + (b_t - b_s)^2) + \frac{\gamma}{2} \sum_{t=1}^T \sum_{k=1}^K (\|\mathbf{v}_{tk}\|^2 + d_{tk}^2) + \\ & \frac{C}{2} \sum_{i=1}^n \xi_i^2 - \sum_{i=1}^n \alpha_i (y_i ((\mathbf{w}_t + \mathbf{v}_{tk(i)})^T \varphi(\mathbf{x}_i) + \\ & (b_t + d_{tk(i)})) - 1 + \xi_i). \end{aligned} \quad (6)$$

由 KKT 条件,  $J$  取得极值时, 有  $\frac{\partial J}{\partial \xi_i} = 0, \frac{\partial J}{\partial \mathbf{w}_t} = 0, \frac{\partial J}{\partial b_t} = 0, \frac{\partial J}{\partial \mathbf{v}_{tk}} = 0, \frac{\partial J}{\partial d_{tk}} = 0$ . 因此, 有

$$\frac{\partial J}{\partial \xi_i} = 0 = C\xi_i - \alpha_i \Rightarrow \xi_i = \alpha_i / C. \quad (7)$$

将式(7)代入(6), 则有

$$J = J_w + J_v + J_b + J_d - \frac{1}{2C} \sum_{i=1}^n \alpha_i^2 + \sum_{i=1}^n \alpha_i. \quad (8)$$

其中

$$\begin{aligned} J_w = & \frac{1}{2T} \sum_{t=1}^T \|\mathbf{w}_t\|^2 + \frac{\lambda}{4T} \sum_{t=1}^T \sum_{s=1}^T Q_{ts} \|\mathbf{w}_t - \\ & \mathbf{w}_s\|^2 - \sum_{i=1}^n \alpha_i y_i \mathbf{w}_{t(i)}^T \varphi(\mathbf{x}_i), \end{aligned} \quad (9)$$

$$J_v = \frac{\gamma}{2} \sum_{t=1}^T \sum_{k=1}^K \|\mathbf{v}_{tk}\|^2 - \sum_{i=1}^n \alpha_i y_i \mathbf{v}_{tk(i)}^T \varphi(\mathbf{x}_i), \quad (10)$$

$$\begin{aligned} J_b = & \frac{1}{2T} \sum_{t=1}^T \|b_t\|^2 + \frac{\lambda}{4T} \sum_{t=1}^T \sum_{s=1}^T Q_{ts} \|b_t - \\ & b_s\|^2 - \sum_{i=1}^n \alpha_i y_i b_{t(i)}, \end{aligned} \quad (11)$$

$$J_d = \frac{\gamma}{2} \sum_{t=1}^T \sum_{k=1}^K d_{tk}^2 - \sum_{i=1}^n \alpha_i y_i d_{tk(i)}. \quad (12)$$

又

$$\frac{\partial J}{\partial \mathbf{w}_t} = 0 = \frac{1}{T} \mathbf{w}_t + \frac{\lambda}{T} \sum_{s=1}^T Q_{ts} (\mathbf{w}_t - \mathbf{w}_s) - \sum_{j \in p_{tk}} \alpha_j y_j \varphi(\mathbf{x}_j),$$

可得

$$\frac{1}{T} \left( \mathbf{w}_t + \lambda \sum_{s=1}^T Q_{ts} (\mathbf{w}_t - \mathbf{w}_s) \right) = \sum_{j \in p_{tk}} \alpha_j y_j \varphi(\mathbf{x}_j).$$

若定义矩阵  $M$  为

$$M_{st} = \begin{cases} 1 + \lambda \sum_k Q_{tk}, & s = t; \\ -\frac{\lambda Q_{ts}}{T}, & s \neq t. \end{cases}$$

因矩阵  $M$  可逆, 则

$$\mathbf{w}_t = \sum_j M_{tt(j)}^{-1} \alpha_j y_j \varphi(\mathbf{x}_j). \quad (13)$$

因此有

$$\sum_{t=1}^T \|\mathbf{w}_t\|^2 = \sum_t \sum_{ij} M_{tt(i)}^{-1} M_{tt(j)}^{-1} \alpha_i \alpha_j y_i y_j \varphi(\mathbf{x}_i) \varphi(\mathbf{x}_j).$$

又由于  $(P^T M^{-2} P)_{ij} = \sum_t M_{tt(i)}^{-1} M_{tt(j)}^{-1}$ , 则由式(13)

可得

$$\sum_{t=1}^T \|\mathbf{w}_t\|^2 = \sum_{ij} (P^T M^{-2} P)_{ij} \alpha_i \alpha_j y_i y_j \varphi(\mathbf{x}_i) \varphi(\mathbf{x}_j) = \alpha^T ((P^T M^{-2} P) \otimes K \otimes Y) \alpha, \quad (14)$$

同时有

$$\begin{aligned} \sum_{t=1}^T \sum_{s=1}^T Q_{ts} \|\mathbf{w}_t - \mathbf{w}_s\|^2 &= \sum_{ts} Q_{ts} \sum_{ij} ((M_{tt(i)}^{-1} - M_{ss(i)}^{-1}) \times (M_{tt(j)}^{-1} - M_{ss(j)}^{-1}) \times \alpha_i \alpha_j y_i y_j \varphi(\mathbf{x}_i) \varphi(\mathbf{x}_j)) = \\ &= \sum_{ij} 2 \left( \sum_t M_{tt(i)}^{-1} M_{tt(j)}^{-1} D_{tt} - \sum_{ts} M_{tt(i)}^{-1} M_{st(j)}^{-1} Q_{ts} \right) \times \alpha_i \alpha_j y_i y_j \varphi(\mathbf{x}_i) \varphi(\mathbf{x}_j) = \\ &= 2 \alpha^T ((P^T M^{-1} (D - Q) M^{-1} P) \otimes K \otimes Y) \alpha. \end{aligned} \quad (15)$$

其中:  $Q$  是对称矩阵, 且记对角矩阵  $D$  为

$$D_{ts} = \begin{cases} \sum_k Q_{tk}, & t = s; \\ 0, & t \neq s. \end{cases}$$

将式(13)和(14)代入(9), 有

$$\begin{aligned} J_w &= \frac{1}{2T} \sum_{t=1}^T \|\mathbf{w}_t\|^2 + \frac{\lambda}{4T} \sum_{t=1}^T \sum_{s=1}^T Q_{ts} \|\mathbf{w}_t - \mathbf{w}_s\|^2 - \\ &= \sum_{i=1}^n \alpha_i y_i \mathbf{w}_{t(i)} \varphi(\mathbf{x}_i) = \\ &= \frac{1}{2T} \alpha^T ((P^T M^{-2} P) \otimes K \otimes Y) \alpha + \\ &= \frac{\lambda}{2T} \alpha^T ((P^T M^{-1} (D - Q) M^{-1} P) \otimes K \otimes Y) \alpha - \\ &= \alpha^T ((P^T M^{-1} P) \otimes K \otimes Y) \alpha = \\ &= -\frac{1}{2} \alpha^T ((P^T M^{-1} P) \otimes K \otimes Y) \alpha. \end{aligned} \quad (16)$$

则由式(8)可知

$$\begin{aligned} J &= -\frac{1}{2} \alpha^T ((P^T M^{-1} P) \otimes K \otimes Y) \alpha + J_v + \\ &= J_b + J_d - \frac{1}{2} \alpha^T (I/C) \alpha + \alpha^T \mathbf{1}. \end{aligned} \quad (17)$$

下面求解  $J_v$ .

$$J_v = \frac{\gamma}{2} \sum_{t=1}^T \sum_{k=1}^K \|v_{tk}\|^2 - \sum_{i=1}^n \alpha_i y_i v_{tk(i)} \varphi(\mathbf{x}_i).$$

由  $\frac{\partial J}{\partial v_{tk}} = 0 \Rightarrow \gamma v_{tk} - \sum_{j \in p_{tk}} \alpha_j y_j \varphi(\mathbf{x}_j) = 0 \Rightarrow v_{tk} = \frac{1}{\gamma} \sum_{j \in p_{tk}} \alpha_j y_j \varphi(\mathbf{x}_j)$ , 可得

$$\begin{aligned} J_v &= \frac{\gamma}{2} \sum_{t=1}^T \sum_{k=1}^K \|v_{tk}\|^2 - \sum_{i=1}^n \alpha_i y_i v_{tk(i)} \varphi(\mathbf{x}_i) = \\ &= -\frac{1}{2\gamma} \sum_{t=1}^T \sum_{k=1}^K \sum_{i,j \in p_{tk}} \alpha_{tk(i)} \alpha_{tk(j)} \times \\ &= y_{tk(i)} y_{tk(j)} \varphi(\mathbf{x}_{tk(i)}) \varphi(\mathbf{x}_{tk(j)}) = \\ &= -\frac{1}{2} \alpha^T ((R^T R / \lambda_2) \otimes K \otimes Y) \alpha. \end{aligned} \quad (18)$$

因此有

$$\begin{aligned} J &= -\frac{1}{2} \alpha^T ((P^T M^{-1} P) \otimes K \otimes Y) \alpha - \\ &= \frac{1}{2} \alpha^T ((G/\lambda_2) \otimes K \otimes Y) \alpha + \\ &= J_b + J_d - \frac{1}{2} \alpha^T (I/C) \alpha + \mathbf{1} = \\ &= -\frac{1}{2} \alpha^T ((P^T M^{-1} P + G/\gamma) \otimes K \otimes Y) \alpha + \\ &= J_b + J_d - \frac{1}{2} \alpha^T (I/C) \alpha + \alpha^T \mathbf{1}. \end{aligned}$$

又

$$\begin{aligned} \frac{\partial J}{\partial b_t} = 0 &= \frac{1}{T} b_t + \frac{\lambda}{T} \sum_{s=1}^T Q_{st} (b_s - b_t) + \sum_{j \in p_{tk}} \alpha_j y_j \Rightarrow \end{aligned}$$

$$b_t = \sum_j M_{tt(j)}^{-1} \alpha_j y_j,$$

$$\frac{\partial J}{\partial d_{tk}} = 0 = \gamma d_{tk} - \sum_{j \in p_{tk}} \alpha_j y_j \Rightarrow$$

$$d_{tk} = \frac{1}{\gamma} \sum_{j \in p_{tk}} \alpha_j y_j,$$

代入式(17),可得

$$J = -\frac{1}{2} \alpha^T ((P^T M^{-1} P + R^T R / \gamma) \otimes K \otimes Y) \alpha - \frac{1}{2} \alpha^T ((P^T M^{-1} P + R^T R / \gamma) \otimes K) \alpha - \frac{1}{2} \alpha^T (I/C) \alpha + \alpha^T \mathbf{1},$$

$$J = -\frac{1}{2} \alpha^T ((P^T M^{-1} P + R^T R / \gamma) \otimes K \otimes (Y + E) + I/C) \alpha + \alpha^T \mathbf{1}.$$

原始问题的对偶问题如下:

$$\max_{\alpha} \alpha^T \mathbf{1} - \frac{1}{2} \alpha^T H \alpha; \text{ s.t. } \alpha \geq 0. \quad (19)$$

其中

$$H = (P^T M^{-1} P + R^T R / \gamma) \otimes K \otimes (Y + E) + I/C. \quad (20)$$

由此,SVC-SVM方法中虽然包含了多个数据流,但其对偶问题仍相当于核空间中的另一个SVM问题的对偶,可以用普通方法来求解此二次规划问题.

### 3 实验结果与分析

本节将评估SVC-SVM方法在多任务概念漂移问题上的分类能力.对比方法分别是TA-SVM方法、ITA-SVM方法和SVC-SVMs方法.其中,SVC-SVMs方法为SVC-SVM方法的特例,即不考虑多个概念漂移任务中的相关性,每次独立求解其中一个任务,以此来观察协同求解多个相关概念漂移问题的效果.为了对比的客观性,本节实验中所使用的数据集及实验的设置都参照ITA-SVM<sup>[16]</sup>.实验环境为Matlab R2010a, Windows 7操作系统,3.30 GHz奔腾处理器,8 GB内存.

#### 3.1 实验设置

研究表明<sup>[13,15-16]</sup>,当采用全局求解模式对单个概念漂移问题进行建模时,相应算法的性能主要依赖于权衡因子 $\lambda$ 的选取,与惩罚因子 $C$ 及高斯核带宽 $\sigma$ 的取值相关性较弱,本文所提方法也有相似的特性.因此,在实验中固定参数取值 $C = 1, \sigma = 1$ .本文采用

网格遍历方式寻找最优参数组合,并采用独立的训练集和测试集.

将ITA-SVM方法中所使用的高斯漂移数据集记为数据集DS1中的Task1,这是一个二类数据集,含有 $n(n = 500)$ 个样本点,每个类中数据的特征都在缓慢地变化,Task1的训练、验证、测试数据都是独立生成的.DS1中还包含另一个非静态数据集Task2,Task2中的数据由下式生成:

$$x_{t2} = \left\{ \frac{2t\pi}{n} - \pi + 0.2y_{t2} + \varepsilon_1, (1-r) \sin \left( \frac{2t\pi}{n} - \pi + 0.2y_{t2} \right) + \varepsilon_2 \right\}. \quad (21)$$

其中: $t = 1, 2, \dots, n; \varepsilon_{1,2}$ 服从于均值为0,方差为 $\sigma = 0.1$ 的正态分布; $y_{t2}$ 是 $\pm 1$ 的随机序列,并保持正、负类的个数总体均衡.为了体现出两个非静态数据集Task1与Task2的相关性及差异性,将Task1的生成模型中的第二维数据作了适度的扰动,用参数 $r$ 来表示非静态数据Task2较之Task1的偏离程度,其中 $r \in \{0.05, 0.1, 0.2\}$ .

数据集DS2中的Task1是ITA-SVM方法<sup>[16]</sup>实验所用的旋转超平面数据集,包含有500个样本点,采样于独立分布的 $d(d = 2)$ 维超立方体 $[-1, 1]^d$ .两类之间的决策边界是一个绕中心点缓慢旋转的超平面.DS2中的Task2数据由Task1模型顺时针旋转一定的角度 $r(r \in \{1, 2, 4, 6\})$ 后生成,以体现出Task2与Task1的相关性.

#### 3.2 在多任务非静态数据集上的性能

对于数据集DS1,依据ITA-SVM的策略,独立生成100组训练集、测试集和验证集用于优化每组训练集上的最优参数.参照ITA-SVM的组织方式,相应实验分为两个步骤.首先根据验证集来优化训练集上的各个参数,然后根据优化得到的参数训练出决策函数,再使用测试集来评估各方法的分类性能.当两个非静态数据Task1和Task2呈现出不同的偏离程度 $r$ 时,求得各方法在两个非静态数据Task1和Task2上的分类精度平均值.每个方法对各训练集运行100次后的平均分类精度及标准差记录在表1中.

表1 数据集DS1上各方法的平均分类准确率 %

Methods	子任务之间的偏离程度		
	$r = 0.05$	$r = 0.1$	$r = 0.2$
TA-SVM	98.33±0.13	98.28±0.13	98.19±0.13
ITA-SVM	98.50±0.12	98.45±0.11	98.36±0.10
SVC-SVMs	98.44±0.11	98.39±0.11	98.27±0.11
SVC-SVM	<b>98.72 ± 0.06</b>	<b>98.65 ± 0.06</b>	<b>98.45 ± 0.07</b>

观察表1可知:1)在数据集DS1上,对概念漂移

问题协同求解的 SVC-SVM 方法优于 SVC-SVMs 方法,表明协同求解的有效性;2) SVC-SVMs 的分类性能与 TA-SVM 方法及 ITA-SVM 方法是相当的,说明本文模型已经充分吸收了 TA-SVM 方法及 ITA-SVM 方法的优点。

对于数据集 DS2,采用相同的实验组织方法,各方法在运行 100 次后的平均分类精度及方差如图 2 所示。

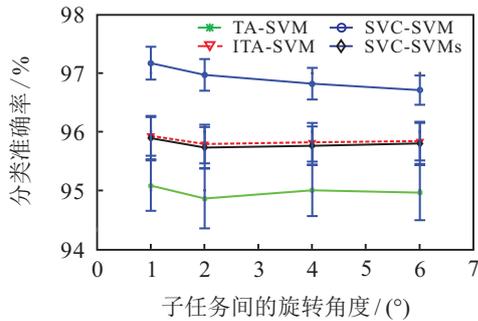


图 2 当 DS2 的各子任务间偏离程度变化时,各方法的分类性能

表 2 数据集 DS3 和 DS4 上各方法在不同噪音下的平均分类准确率

%

DataSet	Methods	在不同噪音比例下的分类准确率					
		0	2	4	6	8	10
DS3	TA-SVM	98.19±0.13	96.19±0.13	94.17±0.18	92.13±0.22	90.06±0.29	87.86±0.39
	ITA-SVM	98.36±0.10	96.18±0.15	94.02±0.20	91.89±0.23	89.62±0.32	87.32±0.39
	SVC-SVMs	98.27±0.11	96.12±0.28	93.96±0.19	91.83±0.23	89.55±0.32	87.21±0.41
	SVC-SVM	<b>98.45 ± 0.07</b>	<b>96.38 ± 0.11</b>	<b>94.34 ± 0.13</b>	<b>92.33 ± 0.16</b>	<b>90.23 ± 0.19</b>	<b>88.10 ± 0.26</b>
DS4	TA-SVM	95.00±0.44	92.68±0.61	90.28±0.64	88.14±0.68	86.07±0.65	83.91±0.65
	ITA-SVM	95.82±0.33	93.28±0.44	90.90±0.43	88.74±0.43	86.58±0.46	84.40±0.53
	SVC-SVMs	95.76±0.33	93.37±0.42	91.00±0.43	88.82±0.41	86.66±0.44	84.49±0.53
	SVC-SVM	<b>96.82 ± 0.27</b>	<b>94.49 ± 0.38</b>	<b>92.27 ± 0.38</b>	<b>90.09 ± 0.37</b>	<b>88.07 ± 0.46</b>	<b>85.95 ± 0.45</b>

观察表 2 可知: 1) 随着噪音的增加,在数据集 DS3 和 DS4 上,各方法的性能都在逐步下降,但 SVC-SVM 方法在各种情况下都优于 ITA-SVM 方法和 TA-SVM 方法;2) 虽然噪音的增加会使 SVC-SVM 方法的泛化能力下降,但随着噪音的增加, SVC-SVM 方法的相对优势进一步扩大。

本节实验表明,协同求解多个非静态数据的 SVC-SVM 方法能够充分利用相关非静态数据中的信息,且对噪音环境具有更强的适应能力。

### 3.4 气体传感器阵列漂移数据集实验

气体传感器阵列漂移数据集<sup>[7]</sup>是某传感器阵列在 2007~2011 年间对 6 种气体的测定结果。由于测量跨度长达 3 年,且有较多的闲置老化时间,传感器测定数据发生变化,若使用早期的模型对后期的数据进行预测,则预测性能急剧恶化,因此这是一个典型

的数据模型漂移问题。本文选用部分气体的测定数据,用作二类分类问题。

首先选取两个相关非静态数据集。Task1 的正类为 Ammonia 气体 (2 565 个样本),负类为 Ethylene 气体 (1 936 个样本);Task2 的正类为与 Ammonia 气体化学性质相似的 Acetaldehyde 气体 (2 565 个样本),负类为 Ethylene 气体。Acetaldehyde 气体原有 2 926 个测量数据,因为本文模型尚不能处理数据量不均衡问题,所以去除 361 个数据(在原数据集中序号为 2 100~2 460),以保持两个非静态数据集中样本量的均衡。将 Task1 及 Task2 中的数据按测定时间顺序分别划分为较均衡的 25 块子数据集,每个子数据集内数据量为 180 个(多余的 1 个样本归类到最后一块子数据集中)。

### 3.3 不同噪音情况下的分类性能

本节将继续观察所提方法在噪音数据集上的分类能力。与 ITA-SVM<sup>[16]</sup> 的实验设置相同,通过随机变换 DS1 和 DS2 上的类别标签来模拟噪音数据集 DS3 和 DS4,噪音比例从 2% 逐步增加到 10%。通过观察当数据集含有不同噪音时的分类能力来评估 SVC-SVM 方法在噪音情况下的性能。在数据集 DS3 和 DS4 上,当含有不同噪音时,各方法的 100 次实验结果记录在表 2 中。

接着进行预测工作,即用前  $N$  块数据来训练模

型, 预测第  $N+1$  块子数据集中各样本的类别. 因为较少的子数据集块数蕴含过少的信息, 而当子数据集块数较多时, 模型可能已发生了较大的漂移, 所以取子数据集块数  $N=5$ , 此时 TA-SVM 方法及 ITA-SVM 方法都能够取得较好的效果. 本次预测工作分成两步: 首先利用前 4 块数据进行训练, 并把当前第 5 块数据作为验证集来优化各方法的参数; 然后利用优化后的参数训练前 5 块数据, 用于预测第 6 块数据中的样本类别. 预测工作从第 6 块数据开始, 到第 25 块数据结束, 共进行 20 次预测. 图 3 显示了各方法每次预测的综合预测结果(两个任务的平均预测精度), 表 3 记录了各方法 20 次预测的平均准确率.

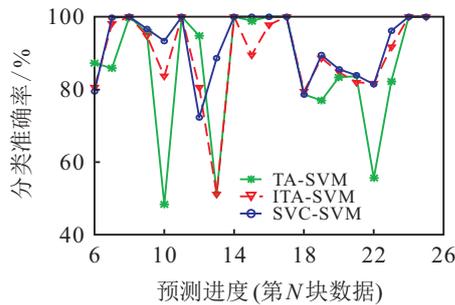


图 3 各方法的综合预测性能

表 3 各方法 20 次预测的平均预测准确率 %

Methods	Task1	Task2	平均
TA-SVM	85.72	86.5	86.11
ITA-SVM	90.36	88.03	89.19
SVC-SVM	<b>94.64</b>	<b>89.86</b>	<b>92.25</b>

观察图 3 可以得到: 1) 在 20 次预测中, TA-SVM 方法取得了 9 次最佳预测, ITA-SVM 方法取得了 8 次最佳预测, SVC-SVM 方法取得了 16 次最佳预测. SVC-SVM 方法有更多的机会取得较好效果, 表明 SVC-SVM 方法确实优于对比算法. 2) 在 20 次预测中, TA-SVM 方法预测效果较差的有 8 次, ITA-SVM 方法预测效果较差的有 5 次, SVC-SVM 方法预测效果较差的有 3 次, 表明 SVC-SVM 方法通过协求解相关模型, 提高了抗干扰能力, 而 TA-SVM 方法和 ITA-SVM 方法无法从相关数据中汲取到有效信息, 难以提高其泛化能力.

观察表 3 可知, 不管在 Task1 还是在 Task2 上, 对两个非静态数据集协同求解的 SVC-SVM 方法明显优于独立求解的 TA-SVM 方法和 ITA-SVM 方法, 显示出协同求解的优越性, 也表明 SVC-SVM 方法在实际应用中的有效性.

#### 4 结论

本文提出了面向多任务概念漂移分类的共享矢量链支持向量机方法 SVC-SVM. 由于借鉴了 rMTL

方法中共享矢量的思想及 ITA-SVM 方法中兼顾子分类器局部优化和分类器序列全局优化的优点, SVC-SVM 方法不仅能够同时求解多个概念漂移问题, 而且较以往单独求解的做法, 其分类性能有较大的提高, 且具有更强的抗噪能力. 同时, SVC-SVM 方法在应用场景的有效性得到了验证. 当然, 本文方法并没有讨论各个子任务之间数据量不均衡的现象, 也未考虑到各个时刻采样量变化问题, 如何解决此类应用场景建模与求解的复杂性, 将是更有意义的挑战.

#### 参考文献(References)

- [1] David P, Philip M Long. Tracking drifting concepts by minimizing disagreements[J]. Machine Learning, 1994, 14(1): 27-45.
- [2] Bartlett Peter L, Ben-David Shai, Kulkarni Sanjeev R. Learning changing concepts by exploiting the structure of change[J]. Machine Learning, 2000, 41(2): 153-174.
- [3] 刘志军, 张杰, 许广义. 基于自适应快速决策树的不确定数据流概念漂移分类算法[J]. 控制与决策, 2016, 31(9): 1609-1614.  
(Liu Z J, Zhang J, Xu G Y. Classifying algorithm for concept drifting of uncertain data streams based on adapting fast decision tree algorithm[J]. Control and Decision, 2016, 31(9): 1609-1614.)
- [4] 张杰, 赵峰. 流数据概念漂移的检测算法[J]. 控制与决策, 2013, 28(1): 29-35.  
(Zhang J, Zhao F. Detecting algorithm of concept drift from stream data [J]. Control and Decision, 2013, 28(1): 29-35.)
- [5] Zhou X, Wang W, Yu L. Traffic flow analysis and prediction based on GPS data of floating cars[J]. Lecture Notes in Electrical Engineering, 2013, 210: 497-508.
- [6] Kuwata S, Inaba Y, Yokogawa M, et al. Stream data analysis application for customer behavior with complex event processing[J]. Ieice Technical Report Data Engineering, 2010, 110(1): 13-18.
- [7] Vergara A, Vembu S, Ayhan T, et al. Chemical gas sensor drift compensation using classifier ensembles[J]. Sensors & Actuators B Chemical, 2012, 166/167: 320-329.
- [8] Bartlett P L. Learning with a slowly changing distribution[C]. Proc of 5th Annu Workshop Computer Learning Theory. Pittsburgh: PA, 1992: 243-252.
- [9] Klinkenberg R, Joachims T. Detecting concept drift with support vector machines[C]. Proc of 17th Int Conf on Mach Learn. San Mateo: Morgan Kaufmann Publishers Inc, 2000: 487-494.
- [10] Lanquillon C. Enhancing test classification to improve information filtering[D]. Magdeburg: Computer Science Department, Otto von Guericke University, 2001: 6.

- [11] Alippi C, Roveri M. Just-in-time adaptive classifiers-part II: Designing the classifier[J]. IEEE Trans on Neural Networks, 2008, 19(12): 2053-2064.
- [12] Street W N, Kim Y S. A streaming ensemble algorithm (SEA) for large-scale classification[C]. Proc of the 7th Int Conf on Knowledge Discovery and Data Mining. San Francisco, 2001: 377-382.
- [13] Grinblat G L, Uzal L C, Ceccatto H A, et al. Solving nonstationary classification problems with coupled support vector machines[J]. IEEE Trans on Neural Networks, 2011, 22(1): 37-51.
- [14] Evgeniou T, Pontil M. Regularized multi-task learning[C]. Proc of the 10th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2004: 109-117.
- [15] 史茨中, 王士同, 王骏, 等. 基于最小包含球的非静态大数据集的快速分类算法[J]. 控制与决策, 2013, 28(7): 1065-1072.  
(Shi Y Z, Wang S T, Wang J, et al. Fast classification for nonstationary large scale data sets using minimal enclosing ball[J]. Control and Decision, 2013, 28(7): 1065-1072.)
- [16] Shi Y, Chung F L, Wang S. An improved TA-SVM method without matrix inversion and its fast implementation for nonstationary datasets[J]. IEEE Trans on Neural Networks Learning Systems, 2015, 26(9): 2005-2018.
- [17] Duan Lixin, Tsang Ivor W, Xu Dong. Domain transfer multiple kernel learning[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2012, 34(3): 465-479.

(责任编辑: 齐 霁)

## 第30届中国控制与决策会议在沈阳召开

**本刊讯** 2018年6月9日~11日,第30届中国控制与决策会议(2018 CCDC)在沈阳玛丽蒂姆酒店隆重召开。

本次盛会由东北大学和中国自动化学会信息物理系统控制与决策专业委员会联合主办,东北大学信息科学与工程学院具体承办,IEEE控制系统协会(CSS)、中国人工智能学会(CAAI)、中国自动化学会控制理论专业委员会和东北大学流程工业综合自动化国家重点实验室协办。来自海内外高等院校和科研机构的1100余位代表出席了会议。其中:海外代表分别来自德国、西班牙、日本、澳大利亚、美国、泰国、法国、新加坡、俄罗斯、韩国、加拿大、香港等10余个国家和地区。这是一次国际学术盛会,业界精英齐聚一堂,交流学术思想,讨论学术问题,充满了浓厚的学术气氛。

6月9日上午召开了“大会开幕式暨CCDC 30周年庆典”,由大会总主席东北大学信息科学与工程学院院长杨光红教授主持,众多国内外著名专家学者出席了庆典。东北大学校长赵继教授代表会议主办单位东北大学致大会开幕辞,对会议在促进学校的一流学科建设、人才培养以及全球学术交流与合作方面起到的平台作用给予了充分的肯定,并希望会议继续发展,为经济社会发展提供人才和科技支撑。东北大学张嗣瀛院士到会致辞,作为会议的发起人,张院士回顾了创会初衷,对会议30年的蓬勃发展倍感振奋和欣慰,希望会议能够抓住机遇,迎接挑战,继续拼搏。会议顾问委员东北大学柴天佑院士对会议的学术水平和领域权威性给予了充分的认可,并表示期望本平台在控制理论、控制技术和控制系统的基础研究与应用研究中发

挥更大的作用。IEEE控制系统学会主席 Francesco Bullo 致辞,对CCDC在国际控制领域的高影响力表示认可,并愿意长期与CCDC合作,共同努力,构建优质的国际学术交流平台。会议协办单位中国人工智能学会常务副理事长北京邮电大学杨放春教授、2010年会议承办单位中国矿业大学马小平教授到会致贺辞。会议国际程序委员主席新加坡南洋理工大学温长云教授回顾了会议国际化后的近10年,程序委员会在提升会议学术水平和影响力方面所做的工作和成绩,并汇报了本届CCDC的稿件情况和程序安排。最后,大会总主席东北大学王福利教授致答谢辞,代表大会组委会对30年来支持CCDC发展的学校领导、业界同仁以及会务工作人员表示感谢,并表示CCDC将始终坚持以促进学术交流、助力科技发展为己任,本着求真务实、与时俱进、勇于创新的拼搏之心继续前行。庆典气氛庄严隆重,使在场的每位与会代表感受到作为科研工作者肩负的使命。

大会涵盖了控制与决策领域的相关课题,主题涉及理论与应用两方面。会议共设大会报告4个、杰出讲座8个、会前研讨会1个、分会报告139场次、张贴论文390余篇。

本届会议发行了《第30届中国控制与决策会议论文集》U盘,其中的1232篇论文将进入IEEE Xplore Database,并被EI检索。

下届承办单位代表——华东交通大学研究生院院长杨辉教授介绍了将于2019年在南昌举办的第31届中国控制与决策会议(2019 CCDC)的筹备情况,并欢迎各位代表积极参加会议。