DOI: 10.11992/tis.201712019
网络出版地址: http://kns.cnki.net/kcms/detail/23.1538.TP.20180417.0849.002.html

基于核心向量机的多任务概念漂移数据快速分类

史荧中[1,2], 王士同[1], 邓赵红[1,3], 侯立功[2], 钱冬杰[2]
（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. 江苏省媒体设计与软件技术重点实验室 (江南大学), 江苏 无锡 214122）

摘　要: 通过协同求解多个概念漂移问题并充分挖掘相关概念漂移问题中蕴含的有效信息, 共享矢量链支持向量机 (shared vector chain supported vector machines, SVC-SVM) 在面向多任务概念漂移分类时表现出良好性能。然而实际应用中的概念漂移问题通常有较大的数据容量, 较高的计算代价限制了 SVC-SVM 方法的推广能力。针对这个弱点, 借鉴核心向量机的近线性时间复杂度的优势, 提出了适于多任务概念漂移大规模数据的共享矢量链核心向量机 (shared vector chain core vector machines, SVC-CVM)。SVC-CVM 具有渐近线性时间复杂度的算法特点, 同时又继承了 SVC-SVM 方法协同求解多个概念漂移问题带来的良好性能, 实验验证了该方法在多任务概念漂移大规模数据集上的有效性和快速性。

关键词: 多任务; 大规模数据集; 概念漂移; 核心向量机; 线性时间复杂度
中图分类号: TP181　　**文献标志码:** A　　**文章编号:** 1673−4785(2018)06−0935−11

中文引用格式: 史荧中, 王士同, 邓赵红, 等. 基于核心向量机的多任务概念漂移数据快速分类[J]. 智能系统学报, 2018, 13(6): 935–945.
英文引用格式: SHI Yingzhong, WANG Shitong, DENG Zhaohong, et al. The core vector machine-based rapid classification of multi-task concept drift dataset[J]. CAAI transactions on intelligent systems, 2018, 13(6): 935–945.

The core vector machine-based rapid classification of multi-task concept drift dataset

SHI Yingzhong[1,2], WANG Shitong[1], DENG Zhaohong[1,3], HOU Ligong[2], QIAN Dongjie[2]
(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 shared vector chain-supported vector machine (SVC-SVM) can solve multiple concept drift problems as well as related problems, and it shows attractive performance in multi-task concept drift classification. However, in many practical scenarios, the concept drift dataset is usually large, and its high computational cost severely limits the generalization ability of the SVC-SVM. To overcome this shortcoming, a novel classifier termed shared vector chain-core vector machine (SVC-CVM) is proposed for large scale multi-task concept drift dataset, considering the asymptotic linear time complexity of the core vector machines. This classifier has the merit of asymptotic time complexity and inherits the good performance of SVC-SVM in solving multi-task concept drift problems. Furthermore, the effectiveness and rapidness of the proposed method is experimentally confirmed on large-scale multi-task concept drift datasets.
Keywords: multi-task; large-scale dataset; concept drift; core vector machines; linear time complexity

收稿日期: 2017−12−13.　**网络出版日期:** 2018−04−17.
基金项目: 国家自然科学基金项目 (61300151);江苏省杰出青年基金项目 (BK20140001); 江苏省高等教育教改研究课题 (2017JSJG282);江苏省高校自然科学研究项目 (18KJB520048).
通信作者: 史荧中. E-mail: shiyz@wxit.edu.cn.

随着计算机信息技术的发展, 每天都会产生大量电信服务、电子商务、金融市场、交通流量、网络监控、超市零售等方面的数据, 这些数据是持续增加且不断变化的。由于数据特征会随着时

间逐渐变化，针对这些非静态数据的分类、回归、聚类模型也在随着时间而缓慢漂移，称为概念漂移[1-2]。对概念漂移的研究已在理论上[1-4]及交通流量预测[5]、超市客户行为分析[6]、气体传感器阵列漂移[7]、垃圾邮件过滤[8]等应用场合取得了良好的效果。概念漂移建模过程中每个时刻的数据量都很少，因而需要借助相邻时刻的一些数据来构建合适的当前时刻模型。以往针对概念漂移分类所作的工作大多是基于滑动窗算法[9-11]的思路，即采用一定时间窗口 (区间) 内的数据进行建模。2011 年，Grinblat 等[12]借鉴 Crammer 等在多任务学习中兼顾局部优化与全局优化的策略，提出了时间自适应支持向量机[13]方法来求解渐变的子分类器。Shi 等[14]提出了增强型时间自适应支持向量机方法，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。

由于生活中的概念漂移问题并不是孤立出现的，如某个气体传感器阵列上对多种气体的测定数据可能会同时漂移；相邻城市的天气情况具有一定的关联；相近街区的交通流量会相互影响等。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，能对建模起到有益的补充。共享矢量链支持向量机[15](shared vector chain supported vector machines, SVC-SVM) 方法通过对相关概念漂移问题协同建模，有效地提升了所得模型的泛化性能。但由于具有较高的算法时间复杂度，限制了其在数据量急剧增长的社会现状下的应用能力。

现在已进入大数据时代，各种社交和电子商务等信息量都越来越大，多任务概念漂移算法的时间复杂度也变得越来越重要。SVC-SVM 方法可转化为核空间中的另一 SVM 问题，算法时间复杂度一般为 $O(n^3)$，其中 n 为样本容量。如采用 SMO(sequential minimal optimization)[16]方法来求解，其复杂度可降为 $O(n^{2.3})$，但 SVC-SVM 方法仍然无法从容面对大规模概念漂移数据集。本文旨在寻找出一种新的概念漂移学习方法，除了能保持 SVC-SVM 方法良好的分类特性外，又能在面对多任务概念漂移大规模数据集时具有较好的算法时间性能。

结合前期在概念漂移领域的研究基础[14-16]，本文提出了共享矢量链核心向量机 (shared vector chain core vector machines, SVC-CVM) 方法，并基于核心向量机[17-19](core vector machine, CVM) 理论给出了 SVC-CVM 方法的快速算法。所提 SVC-CVM 方法具有以下特点：

1) 面对多任务概念漂移问题时，SVC-CVM 方法优于独立求解单个概念漂移问题的 TA-SVM 及 ITA-SVM 方法；

2)SVC-CVM 方法采用了与 SVC-SVM 方法相同的技巧，即假设多个概念漂移问题共享渐变的矢量链序列，因而在分类性能上，SVC-CVM 方法与 SVC-SVM 方法相当；

3)SVC-CVM 方法可以借鉴 CVM 理论[17]设计出快速求解算法，以处理多任务概念漂移中数据量较大的问题，算法时间复杂度接近 $O(n)$。

1　概念漂移问题相关研究

在概念漂移研究方面，传统的研究是基本滑动窗算法，这是一类局部优化模式。TA-SVM 和 ITA-SVM 方法对局部优化和全局优化进行了权衡，取得了良好的效果。

1.1　单任务概念漂移分类方法

TA-SVM[13]方法及 ITA-SVM[14]方法针对的是传统的单任务概念漂移分类。假设有 T 个按时间顺序采集的子数据集，TA-SVM 方法在优化各子分类器的同时，还假设子分类器应该能够光滑地变化，因此约束相邻子分类器之间的差异，其基本思想可由 (1) 式来表示。

$$\min \sum_{t=1}^{T} \mathrm{Risk}(f_t) + \lambda \sum_{t=1}^{T-1} d(f_{t+1}, f_t) \tag{1}$$

式中：第 1 项为局部优化项，f_t 为第 t 个子分类器；第 2 项为全局优化项，$d(f_{t+1}, f_t)$ 为相邻两个子分类器之间的差别，以保证子分类器能平稳变化；λ 是对局部优化与全局优化进行权衡的因子。

1.2　SVC-SVM 方法及其对偶

为了能进一步挖掘出相关概念漂移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 个相关概念漂移数据集，每个概念漂移数据集中的数据由 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 $\{(\boldsymbol{x}_i, y_i) | i = 1, 2, \cdots, n\}$，$n = K \times T \times m$。记 f_{tk} 为第 $k(k = 1, 2, \cdots, K)$ 个任务在第 $t(t = 1, 2, \cdots, T)$ 时刻的分类模型，$\boldsymbol{w}_t$ 为第 t 时刻的共享矢量，$\boldsymbol{v}_{tk}$ 表示在第 t 时刻共享矢量与第 k 个任务 f_{tk} 之间的差异。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 的原理可通过式 (2) 来表示：

$$\begin{aligned} \min \frac{1}{2T} \sum_{t=1}^{T} \|\boldsymbol{w}_t\|^2 + \frac{\lambda}{2T} \sum_{t=1}^{T-1} \|\boldsymbol{w}_{t+1} - \boldsymbol{w}_t\|^2 + \\ \frac{\gamma}{2} \sum_{t=1}^{T} \sum_{k=1}^{K} \|\boldsymbol{v}_{tk}\|^2 + C \sum_{i=1}^{n} L(f_{tk}, \boldsymbol{x}, y) \end{aligned} \tag{2}$$

式中：$\min \sum_{t=1}^{T} \|\boldsymbol{w}_t\|^2$ 为正则化项，$\min \sum_{t=1}^{T-1} \|\boldsymbol{w}_{t+1} - \boldsymbol{w}_t\|^2$ 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, λ为约束各个模型平稳变化的参数; $\min\sum_{t=1}^{T}\sum_{k=1}^{K}\|\boldsymbol{v}_{\mathrm{tk}}\|^2$是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γ表示多个任务间的相关程度; $L(f_{\mathrm{tk}},\boldsymbol{x},y)$为损失函数。根据文献[15]的推导, 可得到 SVC-SVM 方法的对偶形式:

$$\max_{\alpha}\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{1}-\frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{\alpha}\ \ \text{s.t.}\quad \boldsymbol{\alpha}\geqslant 0 \tag{3}$$

式中: $\boldsymbol{H}$ 为扩展核空间上的某个核函数, 具体表达形式可以参见相关文献[15]。从式 (3) 可知, SVC-SVM 方法对多个概念漂移问题同时建模, 其对偶问题为核空间中的另一个支持向量机问题, 当采用普通方法来求解此二次规划问题时, 其算法时间复杂度为$O(n^3)$, 即便采用 SMO[16]方法来求解 SVC-SVM 的对偶问题, 使其复杂度降为$O(n^{2.3})$, 仍然是无法承受计算的代价, 难以从容面对现实生活中数据规模较大的应用场景。

2　共享矢量链核心向量机及快速算法

2.1　共享矢量链核心向量机

鉴于 SVC-SVM 方法在针对多任务概念漂移大规模数据集时算法时间复杂度偏高, 本文借鉴 CVM[17-19]的思路, 提出了与 SVC-SVM 方法在分类性能相似, 但在数据量较大的场景时又能进行快速处理的 SVC-CVM 方法。SVC-CVM 方法借鉴了 SVC-SVM 方法的思想, 为了能进一步用快速算法求解, 本文按文献[17-18]的方法对 SVC-SVM[15]的目标函数稍作变化, 采用平方损失函数, 通过推导得到可以用 CVM 方法快速求解的对偶形式。

设数据集$\{(\boldsymbol{x}_i,y_i)|i=1,2,\cdots,n\}$中含有$n$个样本点, 其中包含 K 个数据流, 每个数据流中的数据由 T 个按时间顺序采集的子数据集组成。在每个时刻引入某个共享矢量, 记第t时刻各个数据流共享某个矢量为$\boldsymbol{w}_t$,第t时刻第k个数据流的决策函数为 f_{tk}, 并记决策函数与共享矢量之间的差为$\boldsymbol{v}_{\mathrm{tk}}=f_{\mathrm{tk}}-\boldsymbol{w}_t$。$\boldsymbol{P}$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, $P_{tj}=1$当且仅当$j\in p_t$, 否则取值为 0。$\boldsymbol{R}$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$j\in r_{\mathrm{tk}}$时$R_{\mathrm{tk},j}=1$, 否则取值为 0。 $\boldsymbol{Q}$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, 即当且仅当$|s-t|=1$时有$Q_{\mathrm{st}}=1$, 否则值为 0。

SVC-CVM 方法的目标函数为

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

式 (4) 中用记号$v_{\mathrm{tk}(i)}$、$d_{\mathrm{tk}(i)}$间接表示第i个样本属于任务k中的第t个子数据集。下面求解式 (4) 的对偶问题:

$$\begin{aligned}J=&\frac{1}{2T}\sum_{t=1}^{T}(\|\boldsymbol{w}_t\|^2+b_t^2)+\\&\frac{\lambda}{4T}\sum_{t=1}^{T}\sum_{s=1}^{T}Q_{\mathrm{ts}}(\|\boldsymbol{w}_t-\boldsymbol{w}_s\|^2+(b_t-b_s)^2)+\\&\frac{\gamma}{2}\sum_{t=1}^{T}\sum_{k=1}^{K}(\|\boldsymbol{v}_{\mathrm{tk}}\|^2+d_{\mathrm{tk}}^2)-\rho+\frac{C}{2}\sum_{i=1}^{n}\xi_i^2-\\&\sum_{i=1}^{n}\alpha_i\left(y_i\left((\boldsymbol{w}_t+\boldsymbol{v}_{\mathrm{tk}(i)})^{\mathrm{T}}\varphi(\boldsymbol{x}_i)+(b_t+d_{\mathrm{tk}(i)})\right)-\rho+\xi_i\right)\end{aligned} \tag{5}$$

由 KKT 条件, J 取得极值时, 有

$$\frac{\partial J}{\partial\xi_i}=0,\frac{\partial J}{\partial\rho}=0,\frac{\partial J}{\partial\boldsymbol{w}_t}=0,$$
$$\frac{\partial J}{\partial b_t}=0,\frac{\partial J}{\partial\boldsymbol{v}_{\mathrm{tk}}}=0,\frac{\partial J}{\partial d_{\mathrm{tk}}}=0$$

因此有:

$$\begin{aligned}&\frac{\partial J}{\partial\xi_i}=0=C\xi_i-\alpha_i\Rightarrow\xi_i=\alpha_i/C\\&\frac{\partial J}{\partial\rho}=0=-1+\sum_{i=1}^{n}\alpha_i\Rightarrow\sum_{i=1}^{n}\alpha_i=1\end{aligned} \tag{6}$$

将式 (6) 代入式 (5) 则有

$$J=J_w+J_v+J_b+J_d-\frac{1}{2C}\sum_{i=1}^{n}\alpha_i^2 \tag{7}$$

式中:

$$J_w=\frac{1}{2T}\sum_{t=1}^{T}\|\boldsymbol{w}_t\|^2+\frac{\lambda}{4T}\sum_{t=1}^{T}\sum_{s=1}^{T}Q_{\mathrm{ts}}\|\boldsymbol{w}_t-\boldsymbol{w}_s\|^2-\sum_{i=1}^{n}\alpha_i y_i\boldsymbol{w}_{t(i)}\varphi(\boldsymbol{x}_i) \tag{8}$$

$$J_v=\frac{\gamma}{2}\sum_{t=1}^{T}\sum_{k=1}^{K}\|\boldsymbol{v}_{\mathrm{tk}}\|^2-\sum_{i=1}^{n}\alpha_i y_i\boldsymbol{v}_{\mathrm{tk}(i)}\varphi(\boldsymbol{x}_i) \tag{9}$$

$$J_b=\frac{1}{2T}\sum_{t=1}^{T}\|\boldsymbol{b}_t\|^2+\frac{\lambda}{4T}\sum_{t=1}^{T}\sum_{s=1}^{T}Q_{\mathrm{ts}}\|\boldsymbol{b}_t-\boldsymbol{b}_s\|^2-\sum_{i=1}^{n}\alpha_i y_i b_{t(i)} \tag{10}$$

$$J_d=\frac{\gamma}{2}\sum_{t=1}^{T}\sum_{k=1}^{K}d_{\mathrm{tk}}^2-\sum_{i=1}^{n}\alpha_i y_i d_{\mathrm{tk}(i)} \tag{11}$$

又:

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

得:

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

若定义矩阵 $\boldsymbol{M}$ 为

$$M_{\mathrm{st}}=\begin{cases}(1+\lambda\sum_k Q_{\mathrm{tk}})/T, & s=t\\ -\lambda Q_{\mathrm{ts}}/T, & s\neq t\end{cases}$$

因矩阵 $\boldsymbol{M}$ 可逆, 则

$$\boldsymbol{w}_t=\sum_j \boldsymbol{M}^{-1}_{\mathrm{tt}(j)}\alpha_j y_j\varphi(\boldsymbol{x}_j)\tag{12}$$

因此有

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

由于

$$\left(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{M}^{-2}\boldsymbol{P}\right)_{ij}=\sum_t \boldsymbol{M}^{-1}_{\mathrm{tt}(i)}\boldsymbol{M}^{-1}_{\mathrm{tt}(j)}$$

则由 (12) 可得:

$$\begin{aligned}&\sum_{t=1}^{T}\|\boldsymbol{w}_t\|^2=\sum_{ij}\left(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{M}^{-2}\boldsymbol{P}\right)_{ij}\alpha_i\alpha_j y_i y_j\varphi(\boldsymbol{x}_i)\varphi(\boldsymbol{x}_j)=\\&\qquad\boldsymbol{\alpha}^{\mathrm{T}}\left((\boldsymbol{P}^{\mathrm{T}}\boldsymbol{M}^{-2}\boldsymbol{P})\otimes\boldsymbol{K}\otimes\boldsymbol{Y}\right)\boldsymbol{\alpha}\end{aligned}\tag{13}$$

同时有

$$\begin{aligned}&\sum_{t=1}^{T}\sum_{s=1}^{T}Q_{\mathrm{ts}}\|\boldsymbol{w}_t-\boldsymbol{w}_s\|^2=\sum_{\mathrm{ts}}Q_{\mathrm{ts}}\sum_{ij}(\boldsymbol{M}^{-1}_{\mathrm{tt}(i)}-\boldsymbol{M}^{-1}_{\mathrm{ss}(i)})\times\\&(\boldsymbol{M}^{-1}_{\mathrm{tt}(j)}-\boldsymbol{M}^{-1}_{\mathrm{ss}(j)})\alpha_i\alpha_j y_i y_j\varphi(\boldsymbol{x}_i)\varphi(\boldsymbol{x}_j)=\\&\sum_{ij}2(\sum_t\boldsymbol{M}^{-1}_{\mathrm{tt}(i)}\boldsymbol{M}^{-1}_{\mathrm{tt}(j)}\boldsymbol{D}_{\mathrm{tt}}-\sum_{ts}\boldsymbol{M}^{-1}_{\mathrm{tt}(i)}\boldsymbol{M}^{-1}_{\mathrm{st}(j)}Q_{\mathrm{ts}})\times\\&\alpha_i\alpha_j y_i y_j\varphi(\boldsymbol{x}_i)\varphi(\boldsymbol{x}_j)=\\&2\boldsymbol{\alpha}^{\mathrm{T}}((\boldsymbol{P}^{\mathrm{T}}\boldsymbol{M}^{-1}(\boldsymbol{D}-\boldsymbol{Q})\boldsymbol{M}^{-1}\boldsymbol{P})\otimes\boldsymbol{K}\otimes\boldsymbol{Y})\boldsymbol{\alpha}\end{aligned}\tag{14}$$

式中: $\boldsymbol{Q}$ 是对称矩阵, 且记对角矩阵 $\boldsymbol{D}$ 为

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

将 (12)、(13) 代入 (8) 有

$$\begin{aligned}&J_{\boldsymbol{w}}=\frac{1}{2T}\sum_{t=1}^{T}\|\boldsymbol{w}_t\|^2+\frac{\lambda}{4T}\sum_{t=1}^{T}\sum_{s=1}^{T}Q_{\mathrm{ts}}\|\boldsymbol{w}_t-\boldsymbol{w}_s\|^2-\\&\sum_{i=1}^{n}\alpha_i y_i\boldsymbol{w}_{t(i)}\varphi(\boldsymbol{x}_i)=\frac{1}{2T}\boldsymbol{\alpha}^{\mathrm{T}}\left((\boldsymbol{P}^{\mathrm{T}}\boldsymbol{M}^{-2}\boldsymbol{P})\otimes\boldsymbol{K}\otimes\boldsymbol{Y}\right)\boldsymbol{\alpha}+\\&\frac{\lambda}{2T}\boldsymbol{\alpha}^{\mathrm{T}}\left((\boldsymbol{P}^{\mathrm{T}}\boldsymbol{M}^{-1}(\boldsymbol{D}-\boldsymbol{Q})\boldsymbol{M}^{-1}\boldsymbol{P})\otimes\boldsymbol{K}\otimes\boldsymbol{Y}\right)\\&\boldsymbol{\alpha}-\boldsymbol{\alpha}^{\mathrm{T}}\left((\boldsymbol{P}^{\mathrm{T}}\boldsymbol{M}^{-1}\boldsymbol{P})\otimes\boldsymbol{K}\otimes\boldsymbol{Y}\right)\boldsymbol{\alpha}=\\&-\frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\left((\boldsymbol{P}^{\mathrm{T}}\boldsymbol{M}^{-1}\boldsymbol{P})\otimes\boldsymbol{K}\otimes\boldsymbol{Y}\right)\boldsymbol{\alpha}\end{aligned}\tag{15}$$

则由 (7) 可知

$$\begin{aligned}&J=-\frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\left((\boldsymbol{P}^{\mathrm{T}}\boldsymbol{M}^{-1}\boldsymbol{P})\otimes\boldsymbol{K}\otimes\boldsymbol{Y}\right)\boldsymbol{\alpha}+\\&J_v+J_b+J_d-\frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}(\boldsymbol{I}/C)\boldsymbol{\alpha}\end{aligned}\tag{16}$$

下面求解 J_v。

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

由

$$\frac{\partial J}{\partial v_{\mathrm{tk}}}=0\Rightarrow\gamma v_{\mathrm{tk}}-\sum_{j\in p_{\mathrm{tk}}}\alpha_i y_i\varphi(\boldsymbol{x}_i)\Rightarrow v_{\mathrm{tk}}=1/\gamma\sum_{j\in p_{\mathrm{tk}}}\alpha_i y_i\varphi(\boldsymbol{x}_i)$$

得

$$\begin{aligned}&J_v=\frac{\gamma}{2}\sum_{t=1}^{T}\sum_{k=1}^{K}\|\boldsymbol{v}_{\mathrm{tk}}\|^2-\sum_{i=1}^{n}\alpha_i y_i\boldsymbol{v}_{\mathrm{tk}(i)}\varphi(\boldsymbol{x}_i)=\\&-\frac{1}{2\gamma}\sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{i,j\in p_{\mathrm{tk}}}\alpha_{\mathrm{tk}(i)}\alpha_{\mathrm{tk(j)}}y_{\mathrm{tk}(i)}y_{\mathrm{tk}(j)}\varphi(\boldsymbol{x}_{\mathrm{tk}(i)})\varphi(\boldsymbol{x}_{\mathrm{tk}(j)})=\\&-\frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\left((\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}/\lambda_2)\otimes\boldsymbol{K}\otimes\boldsymbol{Y}\right)\boldsymbol{\alpha}\end{aligned}\tag{17}$$

因此有

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

又:

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

$$\frac{\partial J}{\partial d_{\mathrm{tk}}}=0\Rightarrow\gamma d_{\mathrm{tk}}-\sum_{j\in p_{\mathrm{tk}}}\alpha_i y_i\Rightarrow d_{\mathrm{tk}}=1/\gamma\sum_{j\in p_{\mathrm{tk}}}\alpha_i y_i$$

由此可得:

$$\begin{aligned}&J=-\frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\left((\boldsymbol{P}^{\mathrm{T}}\boldsymbol{M}^{-1}\boldsymbol{P}+\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}/\gamma)\otimes\boldsymbol{K}\otimes\boldsymbol{Y}\right)\boldsymbol{\alpha}-\\&\frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\left((\boldsymbol{P}^{\mathrm{T}}\boldsymbol{M}^{-1}\boldsymbol{P}+\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}/\gamma)\otimes\boldsymbol{K}\right)\boldsymbol{\alpha}-\frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}(\boldsymbol{I}/C)\boldsymbol{\alpha}\\&J=-\frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\left((\boldsymbol{P}^{\mathrm{T}}\boldsymbol{M}^{-1}\boldsymbol{P}+\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}/\gamma)\otimes\boldsymbol{K}\otimes(\boldsymbol{Y}+\boldsymbol{E})+\boldsymbol{I}/C\right)\boldsymbol{\alpha}\end{aligned}$$

原始问题的对偶问题如下:

$$\max_{\boldsymbol{\alpha}}-\frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{\alpha}\ \text{ s.t. }\ \boldsymbol{\alpha}\geqslant 0,\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{1}=1\tag{18}$$

式中:

$$\boldsymbol{H}=(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{M}^{-1}\boldsymbol{P}+\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}/\gamma)\otimes\boldsymbol{K}\otimes(\boldsymbol{Y}+\boldsymbol{E})+\boldsymbol{I}/C\tag{19}$$

$\boldsymbol{K}$为核矩阵;

$$\boldsymbol{Y}=\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}\ \boldsymbol{y}=[y_1\ y_2\ \cdots\ y_n]$$

$$\boldsymbol{E}=\boldsymbol{1}\times\boldsymbol{1}^{\mathrm{T}};\boldsymbol{1}=[1_1\ 1_2\ \cdots\ 1_n]^{\mathrm{T}}$$

由此, SVC-CVM 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, 可以用常规方法来求解, 其算法时间复杂度为$O(n^3)$, 在算法效率上并不具有优势。因此下文将介绍 SVC-CVM 的快速求解方法。

2.2　核心向量机

求解最小包含球 (minimum enclosing ball, MEB) 是一个数学问题, 等价于求解一个二次规划问题[17-19], 如式 (20) 所示:

$$\max_{\alpha} \boldsymbol{\alpha}^{\mathrm{T}}\mathrm{diag}(\boldsymbol{K})-\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{\alpha} \ \text{ s.t. } \ \boldsymbol{\alpha}\geqslant 0, \quad \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{1}=1 \tag{20}$$

式中: $\boldsymbol{\alpha}=[\alpha_1\ \alpha_2\ \cdots\ \alpha_n]^{\mathrm{T}}$为 Lagrange 乘子, $\boldsymbol{K}_{n\times n}=[k(x_i,x_j)]=[\boldsymbol{\phi}(x_i)^{\mathrm{T}}\boldsymbol{\phi}(x_j)]$为核矩阵。$\mathrm{diag}(\boldsymbol{K})$表示由核矩阵$\boldsymbol{K}$的主对角线元素组成的一维向量。

Tsang 等在文献[17-18]中指出, 形如式(20)的二次规划问题, 如果核矩阵对角线元素为常量, 则均等价于求解 MEB 问题。他们借助求解 MEB 问题时的近似包含球方法[19], 提出了核心向量机 (core vector machines, CVM),在处理大规模数据集时有接近线性的时间复杂度。对形如式(20) 的二次规划问题, 即使核矩阵对角线元素不为常量, 也可以使用核心集方法进行求解, 这时就需要给核空间中每个样本点$\phi(x_i)$都添加一个新的维度$\delta_i\in R$,样本在新特征空间中表示为$\tilde{\boldsymbol{\phi}}(x_i)=[\ \phi(x_i)\quad \delta_i\ ]^{\mathrm{T}}$, 然后求解在新特征空间中的最小包含球。该问题的形式如下:

$$\begin{aligned}&\max_{\alpha} \boldsymbol{\alpha}^{\mathrm{T}}(\mathrm{diag}\boldsymbol{K}+\Delta)-\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{\alpha}\\ &\text{s.t.}\quad \boldsymbol{\alpha}\geqslant 0, \quad \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{1}=1\end{aligned} \tag{21}$$

式中: $\boldsymbol{\Delta}=[\delta_1^2\ \delta_2^2\ \cdots\ \delta_n^2]^{\mathrm{T}}\geqslant 0$是为了保证 (20) 式中$\alpha$的系数为常量。将式 (21) 进一步改写为如下形式:

$$\begin{aligned}&\max_{\alpha} \boldsymbol{\alpha}^{\mathrm{T}}(\mathrm{diag}\boldsymbol{K}+\boldsymbol{\Delta}-\eta\mathbf{1})-\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{\alpha}\\ &\text{s.t.}\quad \boldsymbol{\alpha}\geqslant 0, \quad \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{1}=1\end{aligned} \tag{22}$$

式中: $\eta\in R$为用户自定义常数, 目的是为了保证α的系数为非负的。

2.3　SVC-CVM 的快速算法

当使用普通方法来求解 SVC-CVM 时, 其求解时间复杂度为$O(n^3)$, 对于多任务概念漂移大规模数据集来说, 是相当大的计算开销。对比式 (18) 和式 (22), 它们具有相似的形式, 因此, SVC-CVM 方法可以利用核心向量机技巧来求解。可以将式 (18) 等价地改写为

$$\begin{aligned}&\max_{\alpha} \boldsymbol{\alpha}^{\mathrm{T}}(\mathrm{diag}(\boldsymbol{H})+\boldsymbol{\Delta}-\eta\mathbf{1})-\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{\alpha}\\ &\text{s.t.}\quad \boldsymbol{\alpha}\geqslant 0, \quad \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{1}=1\end{aligned} \tag{23}$$

这是一个标准 MEB 问题, 其中$\boldsymbol{\Delta}=-\mathrm{diag}(\boldsymbol{H})+\eta\mathbf{1}$, 通过适当调节常数$\eta$的值, 可以保证$\boldsymbol{\Delta}\geqslant 0$。

SVC-CVM 算法的输入为多任务概念漂移大规模数据集 S, 核心集逼近精度ε、η、Δ等参数; 输出为核心集S_t和权重系数α。下面给出实现步骤:

由于 SVC-CVM 算法是基于核心集理论的, 因而在描述算法的时间与空间复杂度时, 可以参考文献[17-18], 得到相关结论:

性质 1 对于给定的误差ε, 由 SVC-CVM 算法求得的核心集数量的上界、算法迭代次数的上界、得到的支持向量数目上界都为$O(1/\varepsilon)$。

输入　数据集 S, 最小包含球近似精度;

1) 初始化核心集S_0, 最小包含球半径R_0和中心c_0, 并设置初始迭代次数$t=1$;

2) 若所有点都包含在球$B(c_t,(1+\varepsilon)R_t)$中,则转 7);

3) 找到离中心c_t最远的点x,并将其加入核心集, 即$S_{t+1}=S_t\cup x$;

4) 对新的核心集进行求解, 得到新的半径R_{t+1}和中心c_{t+1}, 并更新权重系数α;

5) 计算新的中心到其他各点的距离 ;

性质 2 对于给定的近似误差ε, SVC-CVM 算法时间复杂度上界应为$O(N/\varepsilon^2+1/\varepsilon^4)$。

6) $t=t+1$, 转 2 ;

7) 终止训练, 返回求解的核心集S_t及权重系数α ;

输出　核心集S_t, 权重系数α。

3　实验研究和分析

本节将对 SVC-CVM 方法进行实验验证, 实验将从 SVC-CVM 方法的分类准确率、SVC-CVM 算法的时间性能两个方面展开。这里有必要首先验证其分类准确率。1) 需要考察引入分类间隔ρ及采用平方损失函数以后, SVC-CVM 算法是否保持了良好的分类能力; 2) 因为 SVC-CVM 方法的有效性是其快速算法有效的必要条件。本文另外选取了在单任务概念漂移建模中取得良好效果的两个方法作为对比算法,作为对比算法的共有: 1)TA-SVM 方法[13]; 2)ITA-SVM 方法[14]; 3)SVM-SVM 方法[15]。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对比算法 TA-SVM [13]。实验环境为 MATLAB R2013a, 操作系统为 Windows7, 8 GB 内存及 3.30 GHz 奔腾处理器。

3.1　实验设置

实验中涉及的各方法与相应参数在表 1 中列出。

表 1　实验所用的对比方法及相应参数

Table 1　Methods and parameters used in experiments

对比算法	求解对象	求解方法	参数
TA-SVM	单概念漂移	普通二次规划	C,σ,λ
ITA-SVM	单概念漂移	普通二次规划	C,σ,λ
SVC-SVM	多概念漂移	普通二次规划	C,σ,λ,γ
SVC-CVM	多概念漂移	核心集技术	$C,\sigma,\lambda,\gamma,\varepsilon$

本文独立生成相同分布的训练集、验证集和测试集各 10 组, 共进行 10 次重复实验, 以获得比较客观的实验结果。实验分为参数优化和建模测

试两个阶段，首先需要基于训练集，利用验证集获得各方法的最优参数；其次基于得到的优化参数对训练集建模，并利用测试集来获得各方法的性能。本文采用网格遍历法来寻找最优参数。

将旋转超平面数据集记为数据集 DS_1 中的第 1 个任务 $Task_1$，$Task_1$ 数据集的样本量为 500，采样于独立分布的 2 维立方体 $[-1,1]^d$，两类之间的边界是一个超平面，并绕原点缓慢旋转。设超平面的法向量为 $\boldsymbol{v}$，$Task_1$ 的训练、验证、测试数据由式 (24) 生成：

$$\begin{aligned} v_1(i)=\cos(2\pi i/500), v_2(i)=\sin(2\pi i/500)\\ y_i=\text{sign}(\boldsymbol{x}_i\boldsymbol{v}(i)), 1\leqslant i\leqslant 500 \end{aligned} \tag{24}$$

数据集 DS_1 中的第 2 个任务 $Task_2$ 数据则由 $Task_1$ 模型顺时针旋转一定的角度 $r(r\in\{2,4,6,8,10\})$ 后随机生成，以体现出 $Task_2$ 与 $Task_1$ 模型的相关性。

将 TA-SVM 方法中所使用的高斯漂移数据集记为数据集 DS_2 中的第 1 个任务 $Task_1$，数据集中包含两个类别，共含有 $n(n=500)$ 个数据点，每个类别中数据的特征都在缓慢变化。$Task_1$ 的训练、验证、测试数据由式 (25) 取 $r=0$ 时独立生成，DS_2 中还包含另一个概念漂移数据集 $Task_2$，其数据同样由 (25) 式生成，这时 $r\neq 0$，以体现任务之间的差异性。

$$x_t=\frac{2t\pi}{n}-\pi+0.2y_t+\varepsilon_1, (1-r)\times\sin\left(\frac{2t\pi}{n}-\pi+0.2y_t\right)+\varepsilon_2 \tag{25}$$

式中：$t=1,2,\cdots,n$，$\varepsilon_{1,2}$ 服从于均值为 0，方差为 $\sigma=0.1$ 的正态分布，y_{t2}、y_{t2} 是 ±1 的随机序列，并保持正类负类个数的均衡。为体现出两个概念漂移数据集 $Task_1$ 与 $Task_2$ 的相关性及差异性，将 $Task_2$ 的生成模型式 (11) 中的第二维数据作了适度的扰动，用参数 r 来表示概念漂移数据 $Task_2$ 较之 $Task_1$ 的偏离程度，其中 $r\in\{0.05,0.1,0.2,0.3\}$。

将 DS_1、DS_2 中的类别标签按一定比例随机替换以模拟噪音数据，得到数据集 DS_3、DS_4，用于测试 SVC-SVM 方法在噪音条件下的分类能力。

数据集 DS_5、DS_6 由 DS_1，DS_2 逐步加大采样量分别得到，它们用于测试 SVC-CVM 方法的算法时间复杂度。实验所用数据集如表 2 所示。

3.2　SVC-SVM 的分类性能

本子节基于数据集 DS_1 和 DS_2 来观察 SVC-CVM 方法的分类能力，并将在噪音数据集 DS_3、DS_4 上观察 SVC-CVM 方法在噪音条件下的性能。

针对数据集 DS_1，依据文献[13]的策略，我们独立生成 10 组训练集、测试集及用于选择最优参数的验证集。根据前述的实验设置，实验分为优化参数和建模测试两个阶段。核函数选用最常用的线性核及高斯核。当两个概念漂移数据 $Task_1$、$Task_2$ 呈现出不同的偏离程度 r 时，求得各方法在两个概念漂移数据 $Task_1$、$Task_2$ 上的分类精度及平均值 Average。每个方法对各训练集共计 10 次运行后的平均分类精度及标准差记录在图 1 中。

表 2　实验所用的数据集

Table 2　Description of artificial dataset

数据集	样本量	描述
DS_1	500	旋转超平面数据集，任务间偏移不同角度
DS_2	500	高斯漂移数据集，任务间振幅变化
DS_3	500	DS_1 中加入 2% ~ 10% 的噪音
DS_4	500	DS_2 中加入 2% ~ 10% 的噪音
DS_5	500 ~ 30 000	逐步加大 DS_1 中的采样量
DS_6	500 ~ 30 000	逐步加大 DS_2 中的采样量

由图 1 可以得到如下观察：

1) 在数据集 DS_1 上，不管采用高斯核还是线性核，当多个任务呈不同偏移程度时，协同求解多个概念漂移问题的 SVC-SVM、SVC-CVM 方法在任务 $Task_1$ 和 $Task_2$ 上总是优于独立求解单个概念漂移问题的 TA-SVM 和 ITA-SVM 方法，显示了协同求解多任务概念漂移问题是有效的。

2) 随着多个任务之间偏离程度的增加，相对于独立求解单个任务，协同求解方法的优势逐渐减弱。

3) 不管是采用高期核还是线性核，也不管任务间的偏移程度，用普通方法求解的 SVC-SVM 与核心集技术求解的 SVC-CVM的分类性能都非常接近。

对高斯漂移数据集 DS_2，按照同样的实验流程，求得当两个任务 $Task_1$、$Task_2$ 呈现出不同的偏离程度 r 时，各方法的分类性能。每个方法对各训练集共计 10 次运行后的平均分类精度及标准差记录在表 3 及表 4 中。

由表 3 及表 4 可以得到如下观察：

1) 在高斯漂移数据集 DS_2 上，不管是采用高斯核还是线性核，协同求解多个概念漂移问题的 SVC-SVM、SVC-CVM 方法总是优于独立求解单个概念漂移问题的 TA-SVM 方法及 ITA-SVM 方法，与数据集 DS_1 上的实验结果相似。

2) 采用高期核或线性核时，不管任务间的偏移程度，SVC-CVM 与 SVC-SVM 方法的分类性能是相当的。

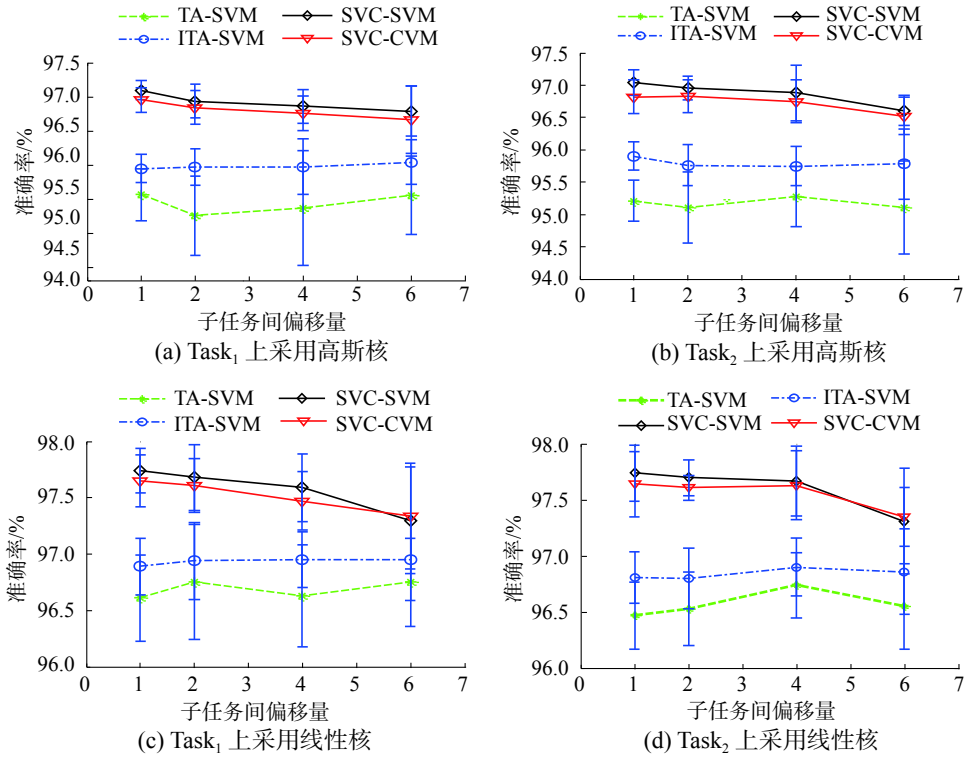


图 1　旋转超平面数据集 DS_1 上各概念漂移数据之间偏移角度变化时的分类性能

Fig. 1　Classification accuracies on DS_1 with different diversities of data stream

表 3　数据集 DS_2 上采用高斯核时各方法的平均分类精度

Table 3　Classification accuracies on dataset DS_2 with Gaussian kernel

方法	Task	不同偏移的分类精度			
		r=0.05	r=0.1	r=0.2	r=0.3
TA-SVM	$Task_1$	98.25+0.32	98.26+0.23	98.13+0.14	97.64+0.21
ITA-SVM		98.33+0.21	98.40+0.12	98.17+0.18	97.82+0.19
SVC-SVM		98.65+0.08	98.60+0.12	98.30+0.14	97.92+0.13
SVC-CVM		98.64+0.08	98.62+0.08	98.27+0.14	97.91+0.16
TA-SVM	$Task_2$	98.38+0.09	98.43+0.14	98.38+0.22	98.38+0.18
ITA-SVM		98.49+0.11	98.50+0.09	98.48+0.16	98.45+0.18
SVC-SVM		98.77+0.04	98.69+0.09	98.61+0.09	98.55+0.17
SVC-CVM		98.71+0.06	98.66+0.04	98.57+0.10	98.54+0.11

表 4　数据集 DS_2 上采用线性核时各方法的平均分类精度

Table 4　Classification accuracies on dataset DS_2 with Linear kernel

方法	Task	不同偏移的分类精度			
		r=0.05	r=0.1	r=0.2	r=0.3
TA-SVM	$Task_1$	97.92+0.39	97.76+0.50	97.70+0.26	97.30+0.49
ITA-SVM		97.80+0.23	97.71+0.19	97.46+0.39	97.05+0.31
SVC-SVM		98.36+0.06	98.17+0.13	97.89+0.20	97.55+0.29
SVC-CVM		98.31+0.10	98.15+0.13	97.82+0.21	97.56+0.35
TA-SVM	$Task_2$	98.04+0.39	98.12+0.21	97.85+0.83	97.92+0.36
ITA-SVM		97.91+0.25	97.86+0.18	97.90+0.21	97.77+0.27
SVC-SVM		98.48+0.11	98.30+0.11	98.31+0.14	97.98+0.30
SVC-CVM		98.44+0.13	98.27+0.16	98.23+0.15	98.03+0.29

下面继续评估 SVC-CVM 方法在噪音数据集 DS_3 和 DS_4 上的表现，以观察本文方法的抗噪能力。与文献[13]的实验设置相同，通过将 DS_1 和 DS_2 上的类别标签随机变换来模拟噪音数据，噪音比例分别为 2% ~ 10%。在数据集 DS_3 和 DS_4 上，当含有不同噪音时各方法的实验结果记录在表 5 到表 6 中。

由表 5 及表 6 可知：

1) 在噪音数据集 DS_3 及 DS_4 上，不管采用高斯核或是线性核时，SVC-SVM 和 SVC-CVM 方法相对于独立求解的 TA-SVM 方法及 ITA-SVM 方法，都有较大的优势。

表 5　数据集 DS_3 上各方法在不同噪音下的平均分类精度

Table 5　Classification accuracies on dataset DS_3 with Different kernel

方法	Kernel	不同噪声的分类精度				
		2	4	6	8	10
TA-SVM	Gauss	96.24+0.13	94.09+0.24	92.08+0.34	90.03+0.30	87.88+0.38
ITA-SVM		96.27+0.12	94.01+0.14	91.86+0.26	89.71+0.38	87.40+0.27
SVC-SVM		96.41+0.08	94.35+0.13	92.34+0.18	90.24+0.16	88.14+0.22
SVC-CVM		96.42+0.28	94.40+0.14	92.37+0.40	90.42+0.16	88.25+0.39
TA-SVM	Linear	95.73+0.30	93.64+0.38	91.48+0.56	89.42+0.43	86.98+0.79
ITA-SVM		95.40+0.19	93.13+0.17	90.95+0.25	88.78+0.40	86.47+0.37
SVC-SVM		96.04+0.09	93.98+0.20	91.93+0.18	89.83+0.21	87.65+0.30
SVC-CVM		95.94+0.43	93.99+0.24	91.89+0.22	89.73+0.42	87.51+0.55

表 6　数据集 DS_4 上各方法在不同噪音下的平均分类精度

Table 6　Classification accuracies on dataset DS_4 with Different kernel

方法	Kernel	不同噪声的分类精度				
		2	4	6	8	10
TA-SVM	Gauss	92.39+0.50	90.17+0.27	88.15+0.82	86.00+0.61	83.94+0.56
ITA-SVM		93.12+0.31	90.88+0.26	88.62+0.50	86.65+0.33	84.43+0.47
SVC-SVM		94.17+0.35	92.16+0.32	89.77+0.33	87.81+0.48	85.83+0.40
SVC-CVM		94.07+0.33	91.91+0.38	89.66+0.35	87.68+0.25	85.63+0.69
TA-SVM	Linear	93.96+0.39	91.76+0.32	89.57+0.41	87.65+0.47	85.73+0.51
ITA-SVM		93.99+0.42	91.66+0.33	89.48+0.56	87.58+0.41	85.37+0.56
SVC-SVM		94.89+0.43	92.78+0.33	90.73+0.37	88.72+0.58	86.70+0.45
SVC-CVM		94.95+0.34	92.88+0.28	90.62+0.39	88.63+0.38	86.54+0.30

2) SVC-CVM 与 SVC-SVM 方法在噪音情况下的分类性能是相当的。

3.3　SVC-CVM 方法的时间性能

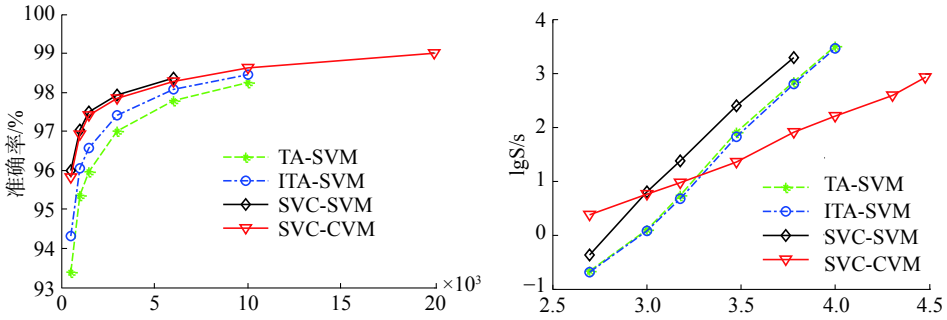
本子节将以数据集 DS_5、DS_6 为基础来评估各方法的算法时间效率。各数据集的样本量从 500 缓慢增加到 30 000 个。对于数据集 DS_5，独立生成 10 组训练集及测试集，并将各方法在取不同容量样本时的平均准确率及平均训练时间显示在图 2 中。随着数据量增加时，为了能更准确地观察各方法时间性能的量级，本文分别以 $\log_{10}n$ (n为样本量) 为横坐标，以$\log_{10}S$(S 为运行时间，单位为 s) 为纵坐标描述各方法的时间性能图，将始的指数曲线转化为线性曲线，斜率代表运行时间的指数量级，如图 2(b) 所示。

由图 2 可以得到如下观察：

图 2(a) 可知，在数据集合 DS_5 上，随着训练数据量的加大，各方法的泛化性能稳定增加。同时 SVC-SVM 和 SVC-CVM 方法优于独立求解单个概念漂移问题的 TA-SVM 和 ITA-SVM 方法。由于用普通 SVM 方法求解时需要先求解相应方法的核矩阵，因此受硬件约束较大，当数据量较大时，相应方法无法继续求解。而 SVC-CVM 方法

采用核心集技术求解, 相应的支持向量逐个添加到核心集中, 不需要预先计算核矩阵, 因而能处理更大容量的数据, 得到泛化能力更强的模型。

图 2(b) 可知, 在数据集 DS_5 上, 求解各方法所需时间与数据量之间呈现稳定的指数级关系, 其中 SVC-CVM 方法所表示的准直线的斜率明显小于其他方法, 显示了 SVC-CVM 方法在时间效率上远优于 TA-SVM、ITA-SVM 与 SVC-SVM 方法。



(a) 分类准确率　　(b) 求解时间

图 2　各方法在数据集 DS_5 上的性能

Fig. 2　Performance on DS_5

在数据集 DS_6 上, 按相同的流程进行训练及测试, 并将各方法在不同容量样本上的平均准确率和标准差、平均训练时间和标准差分别记录在表 7 及表 8 中 (其中—表示在本文实验环境中无法得到相应结果)。

由表 7 及表 8 可以得到如下观察:

表 7　在数据集 DS6 上当不同数据量情况下各方法的平均分类准确率及标准差

Table 7　Classification accuracies with different dataset size of DS_6　%

数据总量	TA-SVM	ITA-SVM	SVC-SVM	SVC-CVM
500	97.40+0.34	97.65+0.27	98.30+0.16	98.36+0.18
1 000	98.31+0.12	98.48+0.14	98.63+0.08	98.63+0.08
1 500	98.53+0.09	98.61+0.11	98.69+0.06	98.67+0.07
3 000	98.73+0.07	98.77+0.08	98.81+0.05	98.80+0.05
6 000	98.81+0.05	98.85+0.06	98.89+0.03	98.88+0.04
10 000	98.83+0.04	98.88+0.04	—	98.91+0.03
20 000	—	—	—	98.93+0.02
30 000	—	—	—	98.94+0.02

表 8　在数据集 DS_6 上当不同数据量情况下各方法的平均训练时间及标准差

Table 8　Training time with different dataset size of DS_6　s

数据总量	TA-SVM	ITA-SVM	SVC-SVM	SVC-CVM
500	0.194± 0.010	0.187± 0.012 5	0.470± 0.011	3.587± 1.512
1 000	1.092± 0.047	0.976± 0.073 7	7.177± 0.587	7.485± 3.379
1 500	4.794± 0.328	3.861± 0.178 7	27.958± 1.457	15.481± 7.193
3 000	74.483± 3.523	56.893± 3.257 8	225.199± 11.843	36.439± 14.452
6 000	612.213±27.575	524.879±39.363 2	1 874.750±106.262	135.923± 74.687
10 000	2 519.897±90.988	2 192.208±64.052 8	—	296.321±183.391
20 000	—	—	—	753.458±287.496
30 000	—	—	—	1 266.967±453.862

1) 从表 7 中可以看出, 在数据集 DS_6 上, 随着训练数据量的增加, 各方法的分类性能逐渐增高, 其中 SVC-SVM 和 SVC-CVM 的分类性能相当, 都优于独立求解单个概念漂移问题的 TA-SVM 与 ITA-SVM 方法。

2) 从表 8 可以看出, 在数据集 DS_6 上, 当数据

量较小时，SVC-CVM 方法的求解时间上并不具有优势。当数据量的逐渐增加时，SVC-CVM 方法求解时间的变化很缓慢，明显优于用普通二次规划方式进行求解。

在数据集 DS_5 和 DS_6 上的实验可知，当数据量不大时，SVC-SVM 方法和 SVC-CVM 方法都优于独立求解的方法，且两者的分类性能相当。当数据量很大时，只有 SVC-CVM 方法能处理较大规模数据集，且在算法时间性能上保持近线性时间复杂度，因而具有较强的实用性。

4 结束语

本文提出了适用于对概念漂移大规模数据集快速求解的多任务核心向量机方法 SVC-CVM。由于采用共享矢量链技术协同求解多个概念漂移问题，SVC-CVM 方法在分类精度上等价于 SVC-SVM 方法，明显优于独立求解单个概念漂移问题的 TA-SVM 及 ITA-SVM 方法，且 SVC-CVM 算法在面对多个概念漂移大数集时仍然能够进行快速建模决策。当然 SVC-CVM 方法仍需要进一步研究，特别是多任务概念漂移大规模数据集的回归问题；多任务概念漂移大规模数据集的单类问题，将是更有意义的挑战。

参考文献：

[1] HELMBOLD D P, LONG P M. Tracking drifting concepts by minimizing disagreements[J]. Machine learning, 1994, 14(1): 27–45.

[2] BARTLETT P L, BEN-DAVID S, KULKARNI S R. Learning changing concepts by exploiting the structure of change[J]. Machine learning, 2000, 41(2): 153–174.

[3] ZHOU Xiangyu, WANG Wenjun, YU Long. Traffic flow analysis and prediction based on GPS data of floating cars[C]//Proceedings of the 2012 International Conference on Information Technology and Software Engineering. [S.l.], 2013: 497–508.

[4] KUWATA S, INABA Y, YOKOGAWA M, et al. Stream data analysis application for customer behavior with complex event processing[C]//IEICE Technical Committee Submission System Conference Paper's Information. [S.l.], 2010, 110(1): 13–18.

[5] VERGARA A, VEMBU S, AYHAN T, et al. Chemical gas sensor drift compensation using classifier ensembles[J]. Sensors and actuators B: chemical, 2012, 166-167: 320–329.

[6] BARTLETT P L. Learning with a slowly changing distribution[C]//Proceedings of the Fifth Annual Workshop on Computational Learning Theory. Pittsburgh, Pennsylvania, USA, 1992: 243–252.

[7] KLINKENBERG R, JOACHIMS T. Detecting concept drift with support vector machines[C]//Proceedings of the Seventeenth International Conference on Machine Learning. San Francisco, CA, USA, 2000: 487–494.

[8] RUANO-ORDáS D, FDEZ-RIVEROLA F, MéNDEZAB J R. Concept drift in e-mail datasets: an empirical study with practical implications[J]. Information sciences, 2018, 428: 120–135.

[9] C LANQUILLON. Enhancing test classification to improve information filtering[D]. Magdeburg, Germany: Faculty Comp Sci, Univ. Magdeburg, 2001.

[10] 文益民, 强保华, 范志刚. 概念漂移数据流分类研究综述[J]. 智能系统学报, 2013, 8(2): 95–104.
WEN Yimin, QIANG Baohua, FAN Zhigang. A survey of the classification of data streams with concept drift[J]. CAAI transactions on intelligent systems, 2013, 8(2): 95–104.

[11] ALIPPI C, ROVERI M. Just-in-time adaptive classifiers-part II: designing the classifier[J]. IEEE transactions on neural networks, 2008, 19(12): 2053–2064.

[12] EVGENIOU T, PONTIL M. Regularized multi--task learning[C]//Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Seattle, WA, USA, 2004: 109–117.

[13] GRINBLAT G L, UZAL L C, CECCATTO H A, et al. Solving nonstationary classification problems with coupled support vector machines[J]. IEEE transactions on neural networks, 2011, 22(1): 37–51.

[14] SHI Yingzhong, CHUNG F L K, WANG Shitong. An improved ta-svm method without matrix inversion and its fast implementation for nonstationary datasets[J]. IEEE transactions on neural networks and learning systems, 2015, 26(9): 2005–2018.

[15] 史荧中, 邓赵红, 钱鹏江,等. 基于共享矢量链的多任务概念漂移分类方法[J]. 控制与决策, 2018, 33(7): 1215-1222.
SHI Yingzhong, DENG Zhaohong, QIAN Pengjiang, et al. Multi-task concept drift classification method based on shared vector chain[J]. Control and Decision, 2018, 33(7): 1215-1222.

[16] PLATT J. Fast training of support vector machines using sequential minimal optimization[C]//Advances in Kernel Methods-Support Vector Learning. Cambridge, MA: MIT Press, 2000: 185–208.

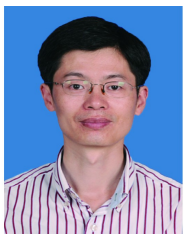
[17] TSANG I W, KWOK J T, CHEUNG P M. Core vector machines: fast SVM training on very large data sets[J].

Journal of Machine Learning Research, 2005, 6: 363–392.

[18] TSANG I W H, KWOK J T Y, ZURADA J M. Generalized core vector machines[J]. IEEE transactions on neural networks, 2006, 17(5): 1126–1140.

[19] BĂDOIU M, CLARKSON K L. Optimal core-sets for balls[J]. Computational geometry, 2008, 40(1): 14–22.

作者简介：



史荧中，男，1970 年生，副教授，博士，主要研究方向为人工智能、模式识别。参与多项省级以上科研项目，发表学术论文 10 余篇。

王士同，男，1964 年生，教授，博士生导师，主要研究方向为人工智能、模式识别。发表学术论文近百篇，其中被 SCI、EI 检索 50 余篇。

邓赵红，男，1981 年生，教授，博士生导师，CCF 高级会员，主要研究方向为人工智能与模式识别、智能计算、系统建模。

2019 IEEE 计算机视觉与模式识别会议
IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2019)

CVPR is the premier annual computer vision event comprising the main conference and several co-located workshops and short courses. With its high quality and low cost, it provides an exceptional value for students, academics and industry researchers.

Overview

The Doctoral Consortium provides a unique opportunity for students, who are close to finishing or who have recently finished their doctorate degree, to interact with experienced researchers in computer vision. A senior member of the community will be assigned as a mentor for each student based on the student's preference or similarity of research interests. All students and mentors will attend a Doctoral Consortium meeting/luncheon, giving the students an opportunity to discuss their ongoing research and career plans with their mentor. In addition, each student will present a poster, either describing their thesis research or a single recent paper, to the other participants and their mentors.

Eligibility

Students must be conducting research in computer vision and be within 6 months (before or after) of graduating with their doctoral degree. Applicants must not have attended a doctoral consortium previously (at ICCV, ECCV or CVPR).

Submission Guidelines

Students that meet the eligibility requirements should submit an application via CMT (link announced at a later date). The applicant must submit the following as a single pdf file (preferred) or a zip/rar file including multiple PDFs.

1. The applicant's CV.

2. A two-page research statement summarizing the applicant's research and progress to date.

3. The title and author list of the poster that will be presented at the consortium, which may or may not be presented at CVPR as well.

4. A signed note from their advisor confirming the date of graduation and stating his/her availability as a mentor.

5. The first and last names and email addresses for 7 potential mentors, ranked from most to least preferred. Note that there is no mentor list provided; you simply have to identify researchers from industry or academia whose work is relevant to yours, or whose feedback you think would be particularly useful to you for a different reason. Do not list your own advisor as a mentor.

Please ensure that all five pieces of information are included in the application. Incomplete applications will be rejected.