

# 一种基于局部分类精度的多源在线迁移学习算法\*

唐诗淇<sup>2</sup>, 文益民<sup>1,2</sup>, 秦一休<sup>2</sup>



<sup>1</sup>(广西可信软件重点实验室(桂林电子科技大学), 广西 桂林 541004)

<sup>2</sup>(桂林电子科技大学 计算机与信息安全学院, 广西 桂林 541004)

通讯作者: 文益民, E-mail: ymwen2004@aliyun.com

**摘要:** 近年来,迁移学习得到越来越多的关注.现有的在线迁移学习算法一般从单个源领域迁移知识,然而,当源领域与目标领域相似度较低时,很难进行有效的迁移学习.基于此,提出了一种基于局部分类精度的多源在线迁移学习方法——LC-MSOTL. LC-MSOTL 存储多个源领域分类器,计算新到样本与目标领域已有样本之间的距离以及各源领域分类器对其最近邻样本的分类精度,从源领域分类器中挑选局部精度最高的分类器与目标领域分类器加权组合,从而实现多个源领域知识到目标领域的迁移学习.在人工数据集和实际数据集上的实验结果表明,LC-MSOTL 能够有效地从多个源领域实现选择性迁移,相对于单源在线迁移学习算法 OTL,显示出了更高的分类准确率.

**关键词:** 在线学习;迁移学习;多源在线迁移;局部分类精度

中图法分类号: TP181

中文引用格式: 唐诗淇,文益民,秦一休.一种基于局部分类精度的多源在线迁移学习算法.软件学报,2017,28(11):2940-2960.  
http://www.jos.org.cn/1000-9825/5352.htm

英文引用格式: Tang SQ, Wen YM, Qin YX. Online transfer learning from multiple sources based on local classification accuracy. Ruan Jian Xue Bao/Journal of Software, 2017, 28(11): 2940-2960 (in Chinese). http://www.jos.org.cn/1000-9825/5352.htm

## Online Transfer Learning from Multiple Sources Based on Local Classification Accuracy

TANG Shi-Qi<sup>2</sup>, WEN Yi-Min<sup>1,2</sup>, QIN Yi-Xiu<sup>2</sup>

<sup>1</sup>(Guangxi Key Laboratory of Trustworthy Software (Guilin University of Electronic Technology), Guilin 541004, China)

<sup>2</sup>(School of Computer Science and Information Safety, Guilin University of Electronic Technology, Guilin 541004, China)

**Abstract:** In recent years, transfer learning has gained more and more attention. However, most of the existing online transfer learning methods transfer knowledge from a single source, and it is hard to make effective transfer learning when the similarity between source domain and target domain is low. To solve this problem, this paper proposes a multi-source online transfer learning method, LC-MSOTL, based on local classification accuracy. LC-MSOTL stores multiple classifiers each trained on a different source, computes the distance between the new arrived sample and its  $k$ -nearest neighbor samples in the target domain as well as the local classification accuracies of each source domain classifier on the  $k$ -nearest neighbor samples, selects the classifier with the highest local classification accuracy from source domain classifiers and combines it with the target domain classifier, so as to realize the knowledge transfer from multi-source domains to a target domain. Experiments on artificial datasets and real datasets illustrates that LC-MSOTL can effectively transfer

\* 基金项目: 国家自然科学基金(61363029); 广西科学研究与技术开发项目(桂科攻 14124005-2-1); 广西自然科学基金(2014GXNSFAA118395); 广西信息科学中心项目(YB408); 桂林电子科技大学研究生教育创新计划(YJCXS201544)

Foundation item: National Natural Science Foundation of China (61363029); Guangxi Scientific Research and Technology Development Project (14124005-2-1); Guangxi Natural Science Foundation (2014GXNSFAA118395); Program of Guangxi Experiment Center of Information Science (YB408); Graduate Education Innovation Program of Guilin University of Electronic Technology (YJCXS 201544)

本文由复杂环境下的机器学习研究专刊特约编辑胡清华教授、张道强教授、张长水教授推荐.

收稿时间: 2017-05-17; 修改时间: 2017-06-26; 采用时间: 2017-08-23

knowledge selectively from multi-source domains, and can get higher classification accuracy compared with the single source online transfer learning algorithm OTL.

**Key words:** online learning; transfer learning; multi sources online transfer; local classification accuracy

随着云计算和物联网等新技术的迅猛发展,产生了多种多样的数据,各种应用全面跨入大数据时代.机器学习在很多领域都得到了密切关注,并在大数据技术中扮演着非常重要的角色.数据流作为机器学习的研究热点,应用于很多领域,比如可穿戴设备异常检测、汽车自动驾驶、信用卡欺诈检测、顾客购物兴趣跟踪、天气预报等.在这些情景中,数据都随着时间不断产生,因此需要对数据进行实时分析与处理.

传统的批量式机器学习算法需要提前获取训练样本,因此不适用于数据流分类.在线学习算法基于流式计算,按照样本到达顺序更新模型,非常适合于解决数据流问题.然而在数据流的起始阶段,由于获取到的样本较少,传统的在线学习算法只能利用当前获取到的少量样本更新模型,导致分类器的分类准确率不理想.尤其在有概念漂移的数据流<sup>[1,2]</sup>中,新概念的产生会使这种情况变得更为严重.在数据流的初始阶段或者新概念刚产生后如何有效提高分类准确率,成为在线学习无法回避的挑战.近年来已成为研究热点的迁移学习<sup>[3]</sup>,为解决该问题提供了新的思路.

迁移学习是运用源领域已有的知识对不同但相关的目标领域进行求解的一种新的机器学习方法<sup>[4]</sup>.迁移学习作为机器学习的热门研究领域,在很多领域已经得到了广泛应用,比如推荐系统、文本分类等领域<sup>[5-8]</sup>.在迁移学习中,根据辅助源领域数量的多少,可将迁移学习划分成两种:单源迁移学习和多源迁移学习.顾名思义,单源迁移学习是指用于辅助目标领域学习的源领域只有1个,Dai等人<sup>[9]</sup>提出的 TrAdaBoost 算法、Pan等人<sup>[10]</sup>提出的 TCA 算法等都属于单源迁移学习.为了减少负迁移,多源迁移学习利用多个源领域来辅助目标领域的学习,张倩等人<sup>[11]</sup>提出的多源动态 TrAdaBoost 实例迁移学习算法、Yao等人<sup>[12]</sup>提出的 TaskTrAdaBoost 和 MultiSourceTrAdaBoost 算法、Eaton等人<sup>[13]</sup>提出的 TransferBoost 算法、Huang等人<sup>[14]</sup>提出的 SharedBoost 算法、Duan等人<sup>[15]</sup>提出的 DAM 算法、Chattopadhyay等人<sup>[16]</sup>提出的 CP-MDA 算法都属于多源迁移学习算法.

以上迁移学习算法通常假设目标领域的训练样本全部已知,但在数据流场合,目标领域的训练样本常常无法一次性全部获得,所以上述迁移学习算法不适用于数据流分类.针对数据流场合的迁移学习问题,Zhao等人提出了一种在线单源迁移学习算法——OTL<sup>[17,18]</sup>.OTL 需要一个与目标领域较为相似的源领域,但在很多场景下此条件难以满足,此时,OTL 的分类准确率就不会太高.针对以上在线学习算法和迁移学习算法的不足,本文提出了一种基于局部分类精度的多源在线迁移学习算法——LC-MSOTL.LC-MSOTL 能够有效地从多个源领域实现选择性迁移.本文主要做了如下3个方面的工作.

- (1) 提出了一种局部分类精度的计算方法.
- (2) 提出了一种基于局部分类精度的多源迁移学习算法——LC-MSOTL.LC-MSOTL 在每个源领域上训练一个分类器并存入分类器池,同时使用(1)中的分类器选取方法从分类器池中挑选合适的分类器,将该分类器与目标领域上训练得到的分类器加权组合对新到样本进行分类与学习.
- (3) 从理论和实验的角度对基于局部分类精度选取源领域分类器方法的合理性和有效性进行分析.

本文第1节介绍在线学习和迁移学习的相关工作.第2节给出选取源领域分类器的方法和度量指标的合理性解释.介绍 LC-MSOTL 算法的基本思路和算法伪代码.第3节使用2组UCI数据、2组人工数据以及 Intrusion detection 数据集进行实验,与经典的在线学习算法 PA<sup>[19]</sup>和单源在线迁移学习算法 OTL<sup>[17,18]</sup>进行对比,以验证 LC-MSOTL 的合理性和有效性.第4节进行总结.

## 1 相关工作

与传统的批量大数据不同,数据流具有实时性、无限性、易失性和突发性等显著特征,很多学者针对数据流问题展开了深入研究<sup>[20]</sup>.目前,主流的在线学习算法大致包括在线线性学习算法、基于核的在线学习算法及其他经典的在线学习算法.在线线性学习算法使用一个线性分类器对数据流中的样本实施分类,当获取到样本

的真实类别后,通过求解一个特定的最优化问题实现对线性分类器权值的更新.感知器<sup>[21]</sup>作为最经典的在线线性学习算法,其更新策略为:当分类错误时,对权值向量进行修改,使其向损失函数的负梯度方向转换.很多学者在其基础上进行了改进,如二阶感知器算法<sup>[22]</sup>和置信度加权算法<sup>[23]</sup>.Crammer 等人<sup>[19]</sup>提出了 PA 算法,其基于支持向量机的最大间隔约束定义链式损失,当对新到样本分类产生损失时,使用该链式损失作为约束构造一个最优化问题并对其求解来更新权值向量.更新后的权值向量具有两个特点.

- (1) 有足够大的间隔正确分类当前样本;
- (2) 与原有的权值向量尽量靠近.

截断梯度法<sup>[24]</sup>常常被用于求解在线学习中最优化问题中的稀疏解,从而降低在线学习算法权值更新时的计算复杂度,其更新策略为:当待更新的权值向量的某维属性值小于给定的阈值时将其设为 0,否则将继续对其进行更新.当数据集为线性不可分时,通常采用基于核的在线学习算法将样本投影到高维再生希尔伯特空间进行分类.例如, Freund 等人<sup>[25]</sup>提出了核感知器算法,当样本被分类错误时,将该样本加入到支持向量集中,并使用该样本更新现有模型.基于核的 PA 算法与核感知器算法类似,当对新到的样本分类产生损失后,基于给定的核函数求解原有最优化问题更新权值向量. Fan 等人<sup>[26]</sup>提出了基于核的在线梯度下降算法 OGD. 该算法使用随机梯度下降法求解在线核学习模型,并且采用正则化项截取来进行加速. Yang 等人<sup>[27]</sup>提出了多任务环境下基于对偶平均的 DA-MTFS 算法. 该算法采用 L1 和 L2 范数的线性组合作为正则化项,具有特征选择和任务选择的作用.此外,很多学者提出了其他在线学习算法,其中较为典型的有 Domingos 等人<sup>[28]</sup>提出的 VFDT 算法、Gama 等人<sup>[29]</sup>针对 VFDT 算法进行改进提出的 VFDTc 算法、郭躬德等人<sup>[30]</sup>基于 KnnModel 模型而提出的一种增量式 KnnModel 算法等.

近年来,迁移学习取得了很多研究成果. Pan<sup>[3]</sup>、Weiss<sup>[31]</sup>、庄福振<sup>[4]</sup>等人对迁移学习进行了比较深入的综述. 在单源迁移学习领域中,依据迁移方式的不同,可将现有的方法大致分为基于样本、基于特征、基于参数以及基于相关关系这 4 种类型. 基于样本的迁移学习,顾名思义,迁移的是样本,其基本思想为:由于目标领域内带标签的样本较少,从而需要通过从源领域中寻找与目标领域样本分布较为近似的样本来帮助训练分类器. 例如, Dai 等人<sup>[9]</sup>根据 adaboosting 的思想提出了迁移学习算法 TrAdaBoost,其假设目标领域中存在很少的带标签的样本,而源领域中存在大量带标签的样本. 该算法将源领域和目标领域的带标签样本合并起来进行训练,通过多次迭代训练多个基分类器,并将它们带权组合成一个集成分类器,并使用该集成分类器对目标领域样本进行分类. Ling 等人<sup>[32]</sup>则提出了一种新的光谱分类算法. 该算法通过优化一个目标函数来寻找源领域中的监督信息与目标领域的样本结构之间最大的一致性. 基于特征的迁移学习并非迁移整个源领域,而是利用目标领域与源领域中存在的一些共有特征表示进行迁移. 例如, Jiang 等人<sup>[33]</sup>提出了一种基于特征选择的迁移学习算法. 该算法首先从源领域和目标领域中选出共有特征并以此训练一个通用分类器,然后从目标领域选择一些特定的特征对该通用分类器进行修正和精化,使其更加适应目标领域的分类. Dai 等人提出的 CoCC<sup>[34]</sup>算法利用协同聚类实现源领域的结构信息向目标领域的迁移. 基于参数的迁移学习不同于前两种迁移学习方式,这种迁移学习方式迁移的是对目标领域训练有效的参数,其中较为典型的有 Dai 等人提出的 NBTC 算法<sup>[35]</sup>. NBTC 算法假设目标领域所有样本的类别未知,首先在具有标签的源领域中训练一个分类模型,然后基于 EM 算法,使用目标领域样本对该模型进行修正,并使用修正后的模型实施分类. 另外,洪佳明等人<sup>[36]</sup>提出的 TrSVM 算法通过定义领域负相似度的概念,要求目标领域分类器对源领域的领域负相似度小. 基于相关关系的迁移学习的迁移对象是相关的关系,即通过源领域与目标领域所含知识之间的相关性进行迁移,以帮助在目标领域中进行建模. 例如, Davis 等人<sup>[37]</sup>提出了一种基于二阶马尔科夫逻辑性的算法. 该算法能够发现源领域中样本的结构,其以带有谓词变量的马尔科夫逻辑公式进行表示,然后用来自目标领域的谓词变量实例化该公式,并基于此实施知识迁移. 然而,目前大多数单源迁移学习算法要求使用一个与目标领域较为相似的源进行迁移,但是在实际中,强源不太容易获得,因此,单源迁移学习算法很容易产生“负迁移”.

为了弥补单源迁移学习的不足,很多学者提出了利用多个源领域进行迁移的多源迁移学习算法<sup>[31,38]</sup>. Eaton 等人<sup>[13]</sup>提出了 TransferBoost 算法. 该算法将多个源领域的样本与目标领域的样本合并,在合并后的数据集上训

训练分类器.该算法为不同的源领域设置不同的样本权重,该权重的大小由样本所在源领域的迁移能力<sup>[39]</sup>决定. Luo 等人<sup>[40]</sup>提出了一致性正则化框架,在每个源领域上训练一个分类器,并将这些分类器带权组合对目标领域实施分类,在训练分类器时,不仅要满足其在对应源领域上的泛化能力,也要保证其与其他分类器在目标领域上预测的一致性.王雪松等人<sup>[41]</sup>根据决策树的特点提出了 STDT 算法.该算法根据决策树的路径信息或者叶子节点中的样本数量信息计算源领域决策树和目标领域决策树之间的相似度,根据相似度,为不同的源领域决策树分配不同的权重. Yao 等人<sup>[12]</sup>将 TradataBoost 进行扩展,提出了多源迁移学习算法 MultiSourceTrAdaBoost.该算法将每个源领域中的样本与目标领域的样本进行合并,组成若干个训练集,使用 Boosting 方式在每个训练集上进行多次迭代,得到的若干个基分类器后加权集成,权值由每个基分类器对目标领域样本的分类准确率决定. Gu 等人<sup>[42]</sup>提出了一种寻找共享特征子空间的方法,在该子空间中实现各个领域样本的聚类,并将该方法用于直推式迁移学习. Chattopadhyay 等人<sup>[16]</sup>提出了 CP-MDA 算法.该算法在每个源领域上训练一个分类器,依据每个源领域与目标领域样本分布的相似程度为每个分类器赋予相应的权值,并利用该权值将所有分类器集成,以对目标领域内的样本实施分类. Tommasi 等人<sup>[43]</sup>提出了一种基于参数的多源迁移学习算法 MMKT.该算法在每个源领域上训练一个 SVM 分类器,然后将它们的 SVM 超平面信息进行迁移.为了最小化“负迁移”的影响,MMKT 算法使用文献<sup>[44]</sup>中提出的方法为每个 SVM 超平面信息赋予相应的权值,其中,与目标领域关联最紧密的源领域赋予的权值最高. Pan 等人提出了 SFA 算法<sup>[45]</sup>,其假设在源领域与目标领域中特征的表示不同,并将所有的特征分为领域特殊和领域独立两种类型. SFA 在每个源领域中建立一个领域特殊特征集合,并将其与一个公共的领域独立特征集合用二部图关联起来.当两个领域中的领域特殊特征关联到同一个领域独立特征集合后,则认为它们特征表示相同.依照此方法, SFA 将多个源领域的特征知识向目标领域实施迁移. Gao 等人<sup>[46]</sup>提出了 LWE 算法,假设目标领域中所有样本均没有获取到真实标签. LWE 算法在每个领域上训练一个基分类器,并针对每个目标领域的样本基于“流形聚类”假设为各个基分类器赋予不同的权值,然后根据赋予的权值进行带权集成.综上所述,现有的很多多源迁移学习算法大多只能对目标领域实施批量学习,不适应进行在线学习与分类.

针对数据流场景, Zhao 等人<sup>[17,18]</sup>将迁移学习和在线学习结合起来提出了 OTL 算法.该算法在源领域和目标领域上各训练一个分类器,将源领域分类器和目标领域分类器加权组合.随着目标领域数据的到来,不断调整目标领域分类器以及源领域和目标领域分类器的权重.但是 OTL 算法的本质仍然是一种单源迁移学习算法,当源领域与目标领域的相似程度较低时, OTL 同样难以取得较好的迁移效果. Ge 等人<sup>[47]</sup>提出了一种多源在线迁移学习算法 OMS-TL,其假设无法获取到目标领域中样本的真实类别. OMS-TL 算法将源领域分类器与目标领域样本之间的关系转化为二部图,每到达一个样本,更新二部图并求解以对当前样本分类.该算法本质上属于直推学习,当源领域和目标领域相似度较低时,也难以取得好的迁移学习效果.

针对以上算法的不足,受 Zhao 等人<sup>[17,18]</sup>的 OTL 算法和 Gao 等人<sup>[46]</sup>的 LWE 算法的启发,本文提出了一种基于局部分类精度的多源在线迁移学习算法——LC-MSOTL.

## 2 算法简介

### 2.1 研究动机

如图 1 所示,有两个源领域和一个目标领域. Gao 等人观察到<sup>[46]</sup>:每个源领域与目标领域都有较大的不同,但又都各自存在分布相同的局部.若在每个源领域上训练一个分类器以实施知识迁移,必然会产生“负迁移”.对于目标领域的 A1 区域,该区域的样本分布与源领域 1 中的对应区域相异,而与源领域 2 的对应区域分布相同.对于目标领域的 A1 区域中的样本,此时若使用源领域 1 的基分类器对其进行分类, A1 区域的样本就会被错误分类;而使用源领域 2 的基分类器对其进行分类, A1 区域的样本就能被正确分类.类似地,对于目标领域的 A3 区域,该区域的样本分布与源领域 2 中的对应区域相异,而与源领域 1 中的对应区域分布相同,此时若使用源领域 2 的基分类器对其进行分类, A3 区域的样本就会被错误分类;而使用源领域 1 的基分类器对其进行分类, A3 区域的样本就能被正确分类.因此,对于目标领域中的样本,若能为每个样本选择合适的源领域分类器对其进行辅助分类,就可能实现在目标领域更准确的分类,从而更好地实现知识迁移.

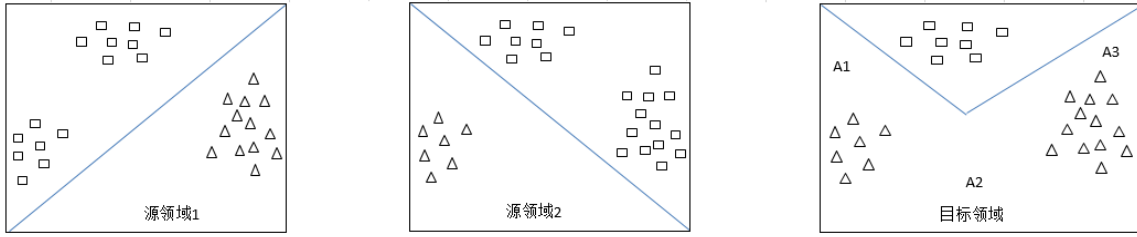


Fig.1 Sample distribution are similar in some local areas of source and target domains

图 1 在某些局部区域源领域和目标领域的样本分布相似

2.2 问题描述

设存在  $n$  个源领域  $D_{S_1}, \dots, D_{S_n}$  及一个目标领域  $D_T$ , 设  $D_T$  为按照时间顺序到达的  $L$  个样本的集合, 每次获取一个样本  $x_t$ , 分类后得到样本的真实标签  $y_t$ .

$$D_T = \{(x_t, y_t) | t=1, 2, \dots, L\},$$

其中,  $(x_t, y_t) \in X \times Y, X \subseteq \mathcal{R}^m, Y = \{c_1, \dots, c_M\}, M$  为类别数量.  $f_{S_1}, f_{S_2}, \dots, f_{S_n}$  分别为在源领域  $D_{S_1}, \dots, D_{S_n}$  上训练得到的分类器, 每个源领域  $D_{S_i} (i=1, 2, \dots, n)$  与目标领域  $D_T$  的样本分布可能有差异, 但在某些局部区域上的样本分布相似. 若使用单个源领域对目标领域进行迁移学习, 显然较难获得很好的迁移学习效果. 因此, 希望利用这种局部区域上样本分布的相似性, 从多个源领域中选择最合适的源领域分类器进行迁移学习.

对于目标领域中的样本  $x_t$ , 其存在一个真实的后验概率  $P(y|x_t)$ . 设该概率能够被描述为一个后验概率向量  $P(y|x_t) = (P(y=c_1|x_t), \dots, P(y=c_M|x_t))^T$ .  $n$  个源领域分类器  $f_{S_1}, f_{S_2}, \dots, f_{S_n}$  分别对该样本进行分类, 其给出的预测后验概率分别设为  $P(y|x_t, f_{S_i}) (i=1, 2, \dots, n)$ .

设该后验概率对应的后验概率向量为  $h_i = (P(y=c_1|x_t, f_{S_1}), \dots, P(y=c_1|x_t, f_{S_n}))^T$ , 定义  $n$  个源领域分类器构成成分类器, 给出的后验概率如公式(1)所示.

$$P_E(y|x_t) = \sum_{i=1}^n \delta(r-i) P(y|x_t, f_{S_i}) \tag{1}$$

其中,  $P_E(y|x_t) = (P_E(y=c_1|x_t), \dots, P_E(y=c_M|x_t))^T$ ;  $n$  为源领域分类器的总数;  $r$  为 1 到  $n$  之间的整数;  $\delta(u)$  为单位冲激函数, 其具体的形式如  $\delta(u) = \begin{cases} 1, & u=0 \\ 0, & \text{otherwise} \end{cases}$ . 当  $\delta(r-i)=1$  时, 有  $r=i, P_E(y|x_t) = P(y|x_t, f_{S_i})$ . 公式(1)的含义在于, 使用分类器集合中的某一个分类器的后验概率作为分类器集合的后验概率. 构造如公式(1)所示的后验概率的目的, 是要从多个源领域分类器中挑选出使  $|P(y|x_t) - P(y|x_t, f_{S_i})|$  最小的分类器  $f_{S_i}$ , 对目标领域中的样本  $x_t$  辅助分类.

2.3 局部分类精度的计算

定理 1. 设  $X$  为  $m$  维连续型随机变量,  $Y$  为离散型随机变量,  $\Omega$  为  $Y$  取值全体构成的集合, 且对随机变量  $X$  的任意取值  $x_t$ , 极限  $\lim_{\epsilon \rightarrow 0^+} \frac{P\{Y \neq y_i, U_\epsilon(\tilde{x})\}}{P\{Y = y_i, U_\epsilon(\tilde{x})\}}$  存在, 其中,  $y_i \in \Omega, U_\epsilon(\tilde{x}) = \{\tilde{x} - \epsilon I < X < \tilde{x} + \epsilon I\}, I$  为分量全为 1, 且维数与  $\tilde{x}$  相同, 则有  $\lim_{\tilde{x} \rightarrow x_t} P_{Y|X}(y_i | \tilde{x}) = P_{Y|X}(y_i | x_t)$ .

证明:

- 首先, 我们证明: 若  $P\{Y=y_i\} > 0$  时,  $P\{U_\epsilon(x_t) | Y=y_i\}$  关于  $(\epsilon, x)$  连续.

记  $\tilde{x} = (\tilde{x}^1, \tilde{x}^2, \dots, \tilde{x}^m)^T, x_t = (x_t^1, x_t^2, \dots, x_t^m)^T, X = (X^1, X^2, \dots, X^m)^T$ .

一方面,

$$|P\{U_\varepsilon(\tilde{x})|Y=y_i\}-P\{U_\varepsilon(x_t)|Y=y_i\}|\sum_{j=1}^m P\{\min(\tilde{x}^j-\varepsilon, x_t^j-\varepsilon_0) \leq X^j \leq \max(\tilde{x}^j-\varepsilon, x_t^j-\varepsilon_0)|Y=y_i\} + \sum_{j=1}^m P\{\min(\tilde{x}^j+\varepsilon, x_t^j+\varepsilon_0) \leq X^j \leq \max(\tilde{x}^j+\varepsilon, x_t^j+\varepsilon_0)|Y=y_i\};$$

另一方面,

$$\begin{aligned} &P\{\min(\tilde{x}^j-\varepsilon, x_t^j-\varepsilon_0) \leq X^j \leq \max(\tilde{x}^j-\varepsilon, x_t^j-\varepsilon_0)\} \\ &= \sum_{y_i \in \Omega} P\{Y=y_i | \min(\tilde{x}^j-\varepsilon, x_t^j-\varepsilon_0) \leq X^j \leq \max(\tilde{x}^j-\varepsilon, x_t^j-\varepsilon_0)\} \\ &= P\{Y=y_i | \min(\tilde{x}^j-\varepsilon, x_t^j-\varepsilon_0) \leq X^j \leq \max(\tilde{x}^j-\varepsilon, x_t^j-\varepsilon_0)\} \\ &= P\{Y=y_i\}P\{\min(\tilde{x}^j-\varepsilon, x_t^j-\varepsilon_0) \leq X^j \leq \max(\tilde{x}^j-\varepsilon, x_t^j-\varepsilon_0)|Y=y_i\}. \end{aligned}$$

因为  $X$  的分布函数为连续函数,则

$$\lim_{\substack{\tilde{x} \rightarrow x_t \\ \varepsilon \rightarrow \varepsilon_0}} P\{\min\{\tilde{x}^j-\varepsilon, x_t^j-\varepsilon_0\} \leq X^j \leq \max\{\tilde{x}^j-\varepsilon, x_t^j-\varepsilon_0\}\} = P\{X^j = x_t^j - \varepsilon_0\} = 0;$$

因为  $P\{Y=y_i\} > 0$ , 从而,

$$0 = P\{X^j = x_t^j - \varepsilon_0\} = P\{Y=y_i\} \lim_{\substack{\tilde{x} \rightarrow x_t \\ \varepsilon \rightarrow \varepsilon_0}} P\{\min\{\tilde{x}^j-\varepsilon, x_t^j-\varepsilon_0\} \leq X^j \leq \max\{\tilde{x}^j-\varepsilon, x_t^j-\varepsilon_0\}|Y=y_i\} = 0.$$

所以,

$$\lim_{\substack{\tilde{x} \rightarrow x_t \\ \varepsilon \rightarrow \varepsilon_0}} P\{\min\{\tilde{x}^j-\varepsilon, x_t^j-\varepsilon_0\} \leq X^j \leq \max\{\tilde{x}^j-\varepsilon, x_t^j-\varepsilon_0\}|Y=y_i\} = 0,$$

其中,  $1 \leq j \leq m$ .

同理可证:

$$\lim_{\substack{\tilde{x} \rightarrow x_t \\ \varepsilon \rightarrow \varepsilon_0}} P\{\min\{\tilde{x}^j+\varepsilon, x_t^j+\varepsilon_0\} \leq X^j \leq \max\{\tilde{x}^j+\varepsilon, x_t^j+\varepsilon_0\}|Y=y_i\} = 0.$$

于是,  $P\{U_\varepsilon(x_t)|Y=y_i\}$  关于  $(\varepsilon, x_t)$  连续,

从而有:  $P\{Y=y_i, U_\varepsilon(x_t)\} = P\{U_\varepsilon(x_t)|Y=y_i\}P\{Y=y_i\}$  关于  $(\varepsilon, x_t)$  连续.

定义:

$$\phi(\varepsilon, x_t) = \begin{cases} \frac{P\{Y=y_i, U_\varepsilon(x_t)\}}{P\{U_\varepsilon(x_t)\}}, & \varepsilon \neq 0 \\ 1 + \lim_{\substack{\tilde{x} \rightarrow x_t \\ \varepsilon \rightarrow 0^+}} \frac{P\{Y \neq y_i, U_\varepsilon(\tilde{x})\}}{P\{Y=y_i, U_\varepsilon(\tilde{x})\}}, & \varepsilon = 0 \end{cases}$$

则  $\phi(\varepsilon, x_t)$  在  $(\varepsilon, x_t) \in [0, \infty) \times \mathcal{H}^m$  上连续, 从而在  $[0, \infty) \times \mathcal{H}^m$  的任何有界闭集上一致连续.

• 其次, 我们证明: 当  $P\{Y=y_i\} > 0$  时, 对任给的  $x_t$  以及任给的  $y_i \in \Omega, P_{Y|X}(y_i|x_t)$  存在.

事实上, 由定理已知条件、概率的定义及  $\phi(\varepsilon, x_t)$  的一致连续性, 有:

$$\begin{aligned} P_{Y|X}(y_i|x_t) &= \lim_{\varepsilon \rightarrow 0^+} \frac{P\{Y=y_i, U_\varepsilon(x_t)\}}{P\{U_\varepsilon(x_t)\}} \\ &= \lim_{\varepsilon \rightarrow 0^+} \lim_{\tilde{x} \rightarrow x_t} \frac{P\{Y=y_i, U_\varepsilon(\tilde{x})\}}{P\{U_\varepsilon(\tilde{x})\}} \\ &= \lim_{\substack{\tilde{x} \rightarrow x_t \\ \varepsilon \rightarrow 0^+}} \frac{P\{Y=y_i, U_\varepsilon(\tilde{x})\}}{P\{Y=y_i, U_\varepsilon(\tilde{x})\} + P\{Y \neq y_i, U_\varepsilon(\tilde{x})\}} \\ &= \lim_{\substack{\tilde{x} \rightarrow x_t \\ \varepsilon \rightarrow 0^+}} \frac{1}{1 + \frac{P\{Y \neq y_i, U_\varepsilon(\tilde{x})\}}{P\{Y=y_i, U_\varepsilon(\tilde{x})\}}}. \end{aligned}$$

由定理已知条件可知,此极限存在,故  $P_{Y|X}(y_i|x_t)$  存在.

由  $\phi(\varepsilon, x_t)$  的一致连续性可知,下面的极限中,两极限符号可交换:

$$\lim_{\tilde{x} \rightarrow x_t} P_{Y|X}(y_i | \tilde{x}) = \lim_{\tilde{x} \rightarrow x_t} \lim_{\varepsilon \rightarrow 0^+} \frac{P\{Y = y_i, U_\varepsilon(\tilde{x})\}}{P\{U_\varepsilon(\tilde{x})\}},$$

$$\text{即 } \lim_{\tilde{x} \rightarrow x_t} P_{Y|X}(y_i | \tilde{x}) = \lim_{\tilde{x} \rightarrow x_t} \lim_{\varepsilon \rightarrow 0^+} \frac{P\{Y = y_i, U_\varepsilon(\tilde{x})\}}{P\{U_\varepsilon(\tilde{x})\}} = \lim_{\varepsilon \rightarrow 0^+} \lim_{\tilde{x} \rightarrow x_t} \frac{P\{Y = y_i, U_\varepsilon(\tilde{x})\}}{P\{U_\varepsilon(\tilde{x})\}} = \lim_{\varepsilon \rightarrow 0^+} \frac{P\{Y = y_i, U_\varepsilon(x_t)\}}{P\{U_\varepsilon(x_t)\}} = P_{Y|X}(y_i | x_t).$$

若  $P\{Y=y_i\}=0$ ,则因为

$$0 = P\{Y = y_i\} = P\{Y = y_i, U_\varepsilon(\tilde{x})\} = 0,$$

所以,

$$P\{Y = y_i, U_\varepsilon(\tilde{x})\} \equiv 0,$$

$$\phi(\tilde{x}, \varepsilon) = \frac{P\{Y = y_i, U_\varepsilon(\tilde{x})\}}{P\{U_\varepsilon(\tilde{x})\}} \equiv 0.$$

从而,  $P_{Y|X}(y_i | \tilde{x}) \equiv 0$ .

综上所述,无论  $P\{Y=y_i\}$  是否为 0,都有  $\lim_{\tilde{x} \rightarrow x_t} P_{Y|X}(y_i | \tilde{x}) = P_{Y|X}(y_i | x_t)$ .

根据定理 1 可知,在一定的条件下,在  $x_t$  的较小邻域内的样本的后验概率也较相近.若基分类器  $f_{S_t}$  对  $x_t$  的较小邻域内的样本分类准确率较高,则说明  $f_{S_t}$  对  $x_t$  的邻域内较多的样本的后验概率的估计比较准确,进而根据定理 1 可知,  $f_{S_t}$  对  $x_t$  的后验概率的估计也很可能比较准确.因此,  $f_{S_t}$  对  $x_t$  的邻域内的样本的正确分类的比例可以体现出  $f_{S_t}$  对  $x_t$  的后验概率估计的准确程度.换句话说,  $f_{S_t}$  对  $x_t$  的邻域内的样本的分类精度越高,就说明  $f_{S_t}$  对  $x_t$  的后验概率能够做出更准确的判断.为此,本文提出了一个基于局部分类精度的指标,通过该指标来选择合适的分类器.

局部分类精度在分类器选择中常常被使用<sup>[48-51]</sup>,该方法假定每个基分类器在待分类样本周围的分类精度不同,从而选择局部分类精度最高的基分类器,并以该分类器的预测结果作为整个分类器组合的输出.目前的局部分类精度可以分为类别无关的方法和类别相关的方法<sup>[48]</sup>,本文提出的计算方法是基于类别无关的方法,如下所示:

$$LC_j = \sum_{i=1}^K \frac{I(f_{S_j}(x_i) = y_i)}{e^{\zeta \text{dis}(x_i, x_t)}} \quad (2)$$

其中,  $x_t$  为待分类样本的特征向量;  $K$  为其邻域的样本数量;  $(x_i, y_i)$  为该邻域中第  $i$  个样本的特征向量和真实类别;  $\text{dis}(x_i, x_t)$  为样本  $x_t$  与待分类样本  $x_i$  的距离,本文采用欧式距离;  $\zeta$  为放大系数;  $I(x)$  为指示函数,其形式为

$$I(x) = \begin{cases} 1, & x = \text{true} \\ 0, & x = \text{false} \end{cases}.$$

若使用一个大小为  $DataSize$  的样本缓存集合  $BD$  存储部分最初获得的目标领域的样本,每获取到一个样本  $x_t$ ,在  $BD$  中选取与  $x_t$  最近的  $K$  个样本作为其邻域,就可以使用每个源领域分类器在该邻域内计算上述局部分类精度指标  $LC$ .

考虑到目标领域样本的分布通常是稳定的,因此在算法中,当缓存集合  $BD$  中充满了样本后,可以不需要更新  $BD$  中存储的样本.对于目标领域中的每一个样本,最多需要从  $DataSize$  个缓存样本中挑选  $K$  个样本构造其邻域,因此可以认为,局部分类精度计算的时间复杂度为常数.

#### 2.4 LC-MSOTL中分类器的选择与更新

结合上述分析和 Zhao 等人<sup>[17,18]</sup>提出的单源在线迁移学习算法——HomOTL-I<sup>[17,18]</sup>,本文提出一种基于局部分类精度的多源在线迁移学习算法——LC-MSOTL.HomOTL-I 算法在源领域和目标领域上各训练一个分类器.当有目标领域新样本到达时,将两个分类器加权集成,对新到样本实施分类.随着目标领域样本的不断到来,根据两个分类器的分类损失不断地调整其对应的权重以及目标领域的分类器,这种加权方式组成的集成分类

器的分类损失小于每个分类器的分类损失.LC-MSOTL 采用与 HomOTL-I 相同的集成方式,LC-MSOTL 的具体方法如下.

在算法初始阶段,LC-MSOTL 需要确定邻域样本数量  $k$ 、放大系数  $\zeta$ .在 LC-MSOTL 算法中,所有领域的分类器均为线性分类器,其形式为  $f: y = \text{sign}(w^T x)$ ,其中,  $w$  为权值向量.设在第  $i$  个源领域上训练得到的分类器表示为  $f_{s_i}$ ,对应的权值向量为  $w_{s_i}$ ,则该分类器的具体形式为  $f_{s_i}: y = \text{sign}(w_{s_i}^T x)$ . 目标领域分类器表示为  $f_t$ ,对应的权值向量为  $w_t$ ,目标领域分类器的具体形式为  $f_t: y = \text{sign}(w_t^T x)$ .使用源领域分类器池  $CF$  存储  $n$  个源领域上训练得到的分类器,  $CF = \{f_{s_1}, f_{s_2}, \dots, f_{s_n}\}$ . 每当获取到目标领域中的样本  $x_t$  时,LC-MSOTL 从  $BD$  中选取距离  $x_t$  最近的  $K$  个样本构造该样本的邻域,再使用分类器选取算法  $Select(CF, BD, x_t, K, \zeta)$  选取局部分类精度最高的源领域分类器  $f_S: y = \text{sign}(w_S^T x)$ ,其中,  $w_S$  表示选取到的局部分类精度最高的分类器对应的权值向量.

$Select(CF, BD, x_t, K, \zeta)$  的算法具体描述如图 2 所示.算法的输入为:分类器集合  $CF$ ,数据缓存器  $BD$ ,新到样本特征向量  $x_t$ ,近邻个数  $K$ ,放大系数  $\zeta$ .第 1 行~第 3 行表示当获得目标领域第 1 个样本时,样本缓存集合  $BD$  为空.此时,选取第 1 个源领域分类器,这相当于随机从源领域分类器中随机挑选.第 4 行~第 8 行表示计算缓存样本到当前样本  $x_t$  的距离.第 9 行表示对缓存样本按距离排序,第 9 行~第 15 行表示从缓存样本中选取出  $K$  个最近邻样本构造关于  $x_t$  的一个邻域.第 16 行表示从该邻域中依据局部分类精度选取分类器.第 17 行表示返回选取的最优分类器  $f_S$ .

```

算法 1.  $Select(CF, BD, x_t, K, \zeta)$  算法伪代码.
输入:分类器集合  $CF$ ,数据缓存器  $BD$ ,新到样本特征向量  $x_t$ ,近邻个数  $K$ ,放大系数  $\zeta$ .
输出:局部分类精度最高的分类器.
1. If  $BD == \emptyset$  /*如果  $BD$  中没有样本*/
2.   Return  $f_1$  /*选择第 1 个源领域分类器*/
3. End If
4.  $AllData[]$ ;
5. for  $i=1, \dots, BD.Length$  do
6.    $dis_i = dis(x_t, x_i)$ ; /*计算样本间距离,采用欧式距离*/
7.    $AllData.add(dis_i)$ ;
8. End for;
9.  $Sort(AllData)$ ; /*按照距离从小到大排序*/
10.  $Kneighbor=[]$ ; /*近邻集合*/
11. for  $i=1, \dots, AllData.Length$  do /*构造近邻*/
12.    $Kneighbor.add(AllData[i])$ ; /*将距离最近的  $K$  个样本存入近邻集合*/
13.   If  $(i - k)$  break; /*超出  $K$  时,终止循环*/
14.    $i++$ ;
15. End for; /*找到  $x_t$  的  $K$  个近邻*/
16.  $f_S = \arg \max_{f_{s_j} \in CF} \sum_{i=1}^K \frac{I(f_{s_j}(x_i) = y_i)}{e^{\zeta dis(x_i, x_t)}}$ ; /*依据  $LC$  选取局部分类精度最高的分类器*/
17. Return  $f_S$ ;
    
```

Fig.2 Pseudo codes of how to select classifier

图 2 分类器选择算法的伪代码

将使用  $Select(CF, BD, x_t, K, \zeta)$  算法挑选出来的源分类器  $f_S$  和目标领域分类器  $f_t$  的权值向量加权组合成分类器  $f$  对目标领域新到样本  $x_t$  实施分类和学习  $f$  的定义如公式(3)所示.

$$f: y = \text{sign} \left( \alpha_{1,t} \prod (w_{s_j}^T x_t) + \alpha_{2,t} \prod (w_t^T x_t) - \frac{1}{2} \right) \quad (3)$$

其中,  $\alpha_{1,t}, \alpha_{2,t}$  分别为源领域分类器和目标领域分类器的权重.当新建目标领域分类器  $f_t$  时,  $\alpha_{1,t}, \alpha_{2,t}$  的初始值都是  $\frac{1}{2}$ ,  $\mathcal{I}(z)$  是一个压缩函数<sup>[17]</sup>.

利用公式(3)对  $x_t$  实施分类之后,获取到该样本的真实标签  $y_t$ ,在  $BD$  未被存满的情况下,将  $(x_t, y_t)$  缓存到  $BD$  当中.然后使用  $(x_t, y_t)$  对目标领域基分类器  $f_t$  以及分类器的权重  $\alpha_{1,t}$  和  $\alpha_{2,t}$  进行更新,其中  $f_t$  的更新策略为:



- 1) 计算分类器损失函数  $l_t = \max(0, 1 - y_t \mathbf{w}_t^T \mathbf{x}_t)$ ;
  - 2) 若  $l_t > 0$ , 则利用  $l_t$  对  $f_t$  进行更新:  $\mathbf{w}_{t+1} = \mathbf{w}_t + \tau y_t \mathbf{x}_t$ , 其中,  $\tau = \min(C, l_t / \|\mathbf{x}_t\|^2)$ ;
- $\alpha_{1,t}$  和  $\alpha_{2,t}$  的更新策略为

$$\alpha_{1,t+1} = \frac{\alpha_{1,t} s_t(\mathbf{w}_s)}{(\alpha_{1,t} s_t(\mathbf{w}_s) + (\alpha_{2,t} s_t(\mathbf{w}_t))}, \alpha_{2,t+1} = \frac{\alpha_{2,t} s_t(\mathbf{w}_t)}{(\alpha_{1,t} s_t(\mathbf{w}_s) + (\alpha_{2,t} s_t(\mathbf{w}_t))},$$

其中,  $s_t(\mathbf{w}) = \exp\{-\eta l^*(\prod(\mathbf{w}^T \mathbf{x}_t), \prod(y_t))\}$ ,  $\eta = 0.5$ ;  $\forall \mathbf{w} \in \mathcal{H}^m, l^*(z, y) = (z - y)^2$ ;  $t$  为新到样本序号,  $C$  为用户输入的参数.

## 2.5 LC-MSOTL算法描述

LC-MSOTL 的算法描述如图 3 所示. 算法的输入为: 目标领域  $D_T$ , 近邻个数  $K$ , 放大系数  $\zeta$  以及  $n$  个源领域分类器  $f_{s_1}, f_{s_2}, \dots, f_{s_n}$ . 第 1 行~第 6 行表示算法初始化. 第 8 行表示根据  $x_t$  从源领域分类器集中选取最优分类器. 第 9 行表示使用所选局部分类精度最高分类器和目标领域分类器加权组合对样本  $x_t$  进行分类. 第 10 行表示目标领域分类器的更新. 第 11 行表示各个分类器权重的调整. 第 12 行、第 13 行缓存样本.

算法 2. LC-MSOTL 算法伪代码.

输入: 数据流  $DS$ , 近邻个数  $K$ , 放大系数  $\zeta$ , 源领域分类器  $f_{s_1}, f_{s_2}, \dots, f_{s_n}$ .

/\*初始化阶段\*/

```

1.  t=1;
2.  CF.create(); /*创建源领域分类器集合*/
3.  f_t.create(); /*创建目标领域分类器*/
4.  CF.add(f_{s_1}, f_{s_2}, \dots, f_{s_n}); /*将源领域分类器存入分类器集合*/
5.  BD=[]; /*用于缓存目标领域已获得的样本*/
6.  \alpha_{1,0} = \frac{1}{2}, \alpha_{2,0} = \frac{1}{2}; /*分类权重初始值*/
7.  While ((x_t, y_t) \in DS) do
8.    f_s = Select(CF, BD, x_t, K, \zeta);
9.    y = sign\left(\alpha_{1,t} \prod(\mathbf{w}_s^T \mathbf{x}_t) + \alpha_{2,t} \prod(\mathbf{w}_t^T \mathbf{x}_t) - \frac{1}{2}\right) /*使用集成分类器进行分类, \mathbf{w}_s 为 f_s 的权值向量, \mathbf{w}_t 为 f_t 的权值向量*/
10.   \alpha_{1,t+1} = \frac{\alpha_{1,t} s_t(\mathbf{w}_s)}{(\alpha_{1,t} s_t(\mathbf{w}_s) + (\alpha_{2,t} s_t(\mathbf{w}_t))}, \alpha_{2,t+1} = \frac{\alpha_{2,t} s_t(\mathbf{w}_t)}{(\alpha_{1,t} s_t(\mathbf{w}_s) + (\alpha_{2,t} s_t(\mathbf{w}_t))}; /*更新分类器权重*/
11.   f_t.update(x_t, y_t); /*使用第 2.4 节中的更新规则更新目标领域分类器*/
12.   If (BD.Length < DataSize); /*缓存目标领域样本*/
13.     BD.Save(x_t, y_t);
14.   End If;
15.   t++;
16. End While;

```

Fig.3 Pseudo codes of LC-MSOTL

图 3 LC-MSOTL 算法的伪代码

## 3 实验结果及分析

为了验证 LC-MSOTL 算法的合理性和有效性, 我们设计了如下 4 组实验: (1) LC-MSOTL 在不同相似度下的迁移性能实验; (2) LC-MSOTL 整合多个互补弱源的迁移性能实验; (3) LC-MSOTL 对 Intrusion detection 数据集的实验; (4) LC-MSOTL 对参数  $K$  和  $\zeta$  的敏感性实验.

### 3.1 实验数据介绍

为了精确地控制源领域和目标领域的相似度, 以探讨在各种相似度情形下 LC-MSOTL 的分类准确率, 本文利用来自 UCI 的 Letter 数据集、Waveform 数据集以及人工 Gauss 数据集构造了多个数据集.

另外, 还使用了 Intrusion detection 数据集. 在该数据集中, 完全不清楚源领域和目标领域的相似程度.

Letter 数据集被很多迁移学习的文章所使用. 该数据集是用来识别手写的英文字母, 其中一共有 26 个类别, 分别对应 26 个英文字母, 共有 2 万个样本. 每个样本包含 16 维属性. 本文分别使用其中的字母 A, B, C, D 制作第

1 个数据集,记为 ABCD 数据集.使用字母 N,O,P,Q 制作第 2 个数据集,记为 NOPQ 数据集.其中,字母 A:789 个样本,字母 B:766 个样本,字母 C:736 个样本,字母 D:805 个样本;字母 N:783 个样本,字母 O:753 个样本,字母 P:803 个样本,字母 Q:783 个样本.

Waveform 数据集是人工数据集,包含 5 000 个样本,每个样本包含 21 维属性.该数据集包含 3 个模式类,每一个模式类是由 3 个基波中的两个基波组合而成.第 1 个模式类包含 1 696 个样本,第 2 个模式类包含 1 657 个样本,第 3 个模式类包含 1 647 个样本.为了方便,3 个模式类分别标记为 0,1,2.

GaoSi 数据集也是人工数据集,包含 8 000 个样本,每个样本包含 10 维属性,每维属性都服从方差皆为 0.5 的高斯分布,且彼此相互独立.该数据集包含 4 个模式类,每个模式类各有 2 000 样本.类别 1 中,样本第 1 维的均值为 2,其他维均值为 0;类别 2 中,样本的第 2 维均值为 2,其他维均值为 0;类别 3 中,样本第 3 维的均值为 2,其他维均值为 0;类别 4 中,样本第 4 维的均值为 2,其他维均值为 0.为了方便,4 个模式类分别标记为 1,2,3,4.

Intrusion detection 数据集来自 KDD CUP99,其包括正常访问连接数据以及 DOS,R2L,U2R,Probing 这 4 种攻击连接数据.一条数据表示一次网络连接.按照 Gao 的方法<sup>[44]</sup>,随机抽取正常访问连接样本分别和从 DOS,R2L 及 Probing 中随机抽取的攻击连接样本构成 3 个领域.具体的方法是:分别从 DOS,Probing,R2L 中随机抽取 1 600 条(由于 R2L 中不满 1 600 条,所以 1 126 条全部抽取),将其分别与正常访问连接数据中随机抽取的 1 600 条样本组合形成 3 个源领域,故 3 个源领域的数据量分别为 3 200,3 200,2 726 条.这 3 个领域分别被命名为 DOS, Probing,R2L,再分别使用其中的两个领域为源领域,另外一个为目标领域.实验结果为 100 划分的平均值.

### 3.2 实验方案介绍和实验结果分析

本文将 LC-MOSTL 与 PA,OTL 算法进行对比.其中,PA 为基于核的在线学习算法<sup>[19]</sup>,所采用的核函数是高斯径向函数,其参数为  $\sigma=4$ ,惩罚系数  $C=5$ .本文中,PA 表示使用 PA 算法直接在线学习目标领域的样本.为了方便,用 OTL-S<sub>i</sub> 表示只使用第  $i$  个源对目标领域实施迁移学习的 OTL 算法,系数  $\eta=0.5$ ;在 LC-MSOTL 中,目标领域样本的缓冲上限  $DataSize=500$ ,邻域样本数量  $K=3$ ,放大系数  $\zeta=1$ .在本实验中,OTL 和 LC-MSOTL 的基分类器均采用 PA.

#### 3.2.1 LC-MSOTL 在不同相似度下的迁移性能实验

在本实验中,在 ABCD 数据集上生成 2 个源领域和 1 个目标领域.目标领域包含 1 496 个样本.每个源领域则各包含 800 个样本,其由每个字母中随机各抽取 200 个样本而成.类似地,在 NOPQ 数据集上也生成 2 个源领域和 1 个目标领域.目标领域包含 1 522 个样本.每个源领域也同样各包含 800 个样本,其由每个字母中随机各抽取 200 个样本而成.采取同样的思路,在 GaoSi 数据集上也生成两个源领域以及一个目标领域.目标领域包含 4 000 个样本.每个源领域各由随机抽取的 2 000 个样本组成,其中,每个模式类各抽取 500 个样本;在 Waveform 数据集上也生成 2 个源领域以及 1 个目标领域.目标领域包含 2 000 个样本.每个源领域则各由随机抽取的 1 500 个样本组成,其中,每个模式类各抽取 500 个样本.按照以上方式,从 ABCD,NOPQ 以及 Gaosi 数据集各生成 10 份实验数据.

将 ABCD,NOPQ 以及 Gaosi 数据集的各份实验数据,分别按照 0.25,0.50,0.75 的源领域与目标领域相似度,给各个模式类打上“+”和“-”的大类标签.“+”和“-”可以分别理解为用户对各子类别的“感兴趣”和“不感兴趣”.同样将 Waveform 数据集分别按照 0.33,0.66 的源领域与目标领域相似度,做同样的处理.

各数据集的数据分布见表 1~表 5.

本节实验的目的在于分析各算法在各种相似度情形下对目标领域样本的分类准确率及其变化情况.为了模仿在线学习场景,实验中假定目标领域样本按照数据流的形式一个一个地到来.为了减弱数据流中样本出现次序的随机性对实验结果的影响,每份实验数据再重复进行 10 次实验,在每次实验之前,随机打乱目标领域样本的顺序,因此,在每个数据集上总计要进行 100 次实验.所有的实验结果为取 100 次实验的平均值.各算法对目标领域样本的分类准确率见表 6,分类准确率的变化如图 4~图 7 所示(在图中,每隔 50 个样本计算一次累积准确率).

**Table 1** Distributions of the data sets of ABCD, NOPQ, and GaoSi when the similarity is set as 0.25**表 1** 相似度为 0.25 时,ABCD,NOPQ,GaoSi 数据集的分布

ABCD	模式类		源 1 领域	源 2 领域	目标领域
	NOPQ	GaoSi			
A	N	1	+	-	-
B	O	2	-	-	+
C	P	3	-	-	+
D	Q	4	-	+	-

**Table 2** Distributions of the data sets of ABCD, NOPQ, and GaoSi when the similarity is set as 0.5**表 2** 相似度为 0.5 时,ABCD,NOPQ,GaoSi 数据集的分布

ABCD	模式类		源 1 领域	源 2 领域	目标领域
	NOPQ	GaoSi			
A	N	1	+	-	-
B	O	2	+	-	+
C	P	3	-	+	+
D	Q	4	-	+	-

**Table 3** Distributions of the data sets of ABCD, NOPQ, and GaoSi when the similarity is set as 0.75**表 3** 相似度为 0.75 时,ABCD,NOPQ,GaoSi 数据集的分布

ABCD	模式类		源 1 领域	源 2 领域	目标领域
	NOPQ	GaoSi			
A	N	1	+	-	-
B	O	2	+	+	+
C	P	3	+	+	+
D	Q	4	-	+	-

**Table 4** Distributions of the data set of Waveform when the similarity is set as 0.33**表 4** 相似度为 0.33 时,Waveform 数据集的分布

模式类	源 1 领域	源 2 领域	目标领域
0	+	-	-
1	-	+	-
2	-	-	+

**Table 5** Distributions of the data set of Waveform when the similarity is set as 0.66**表 5** 相似度为 0.66 时,Waveform 数据集的分布

模式类	源 1 领域	源 2 领域	目标领域
0	-	+	-
1	+	-	-
2	+	+	+

**Table 6** Classification accuracy of each method for each data set under different similarity**表 6** 各算法对各个数据集在不同相似度下的分类准确率

准确率	PA	OTL-S1	OTL-S2	LC-MSOTL
ABCD0.25	0.760 48	0.759 95	0.759 87	<b>0.760 95</b>
ABCD0.50	0.763 41	0.762 54	0.765 20	<b>0.838 59</b>
ABCD0.75	0.760 68	0.768 55	0.776 14	<b>0.885 28</b>
NOPQ0.25	<b>0.699 77</b>	0.699 10	0.698 90	0.697 75
NOPQ0.50	0.699 92	0.698 79	0.702 61	<b>0.844 24</b>
NOPQ0.75	0.699 46	0.715 11	0.721 56	<b>0.865 47</b>
GaoSi0.25	<b>0.920 61</b>	0.920 47	0.920 45	0.920 55
GaoSi0.50	0.920 95	0.921 08	0.921 24	<b>0.935 31</b>
GaoSi0.75	0.921 08	0.924 45	0.924 77	<b>0.955 38</b>
Waveform0.33	0.808 37	0.808 18	0.808 34	<b>0.808 49</b>
Waveform0.66	0.808 22	0.809 49	0.809 57	<b>0.822 73</b>

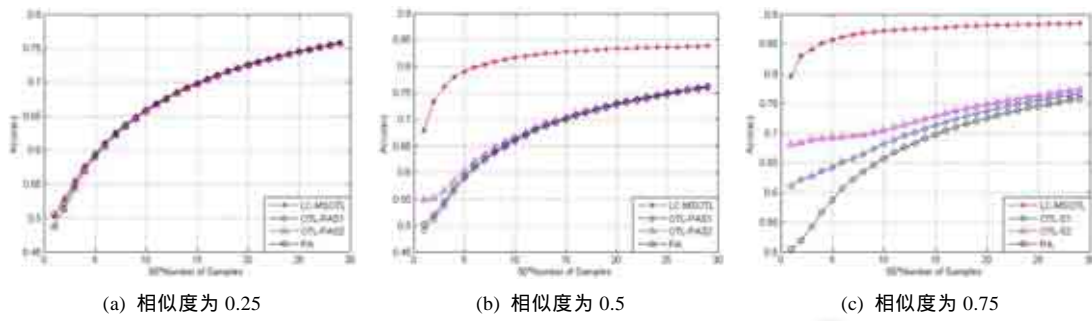


Fig.4 Classification accuracy change for ABCD data set with different similarities

图 4 ABCD 数据集在不同相似度下的分类准确率变化

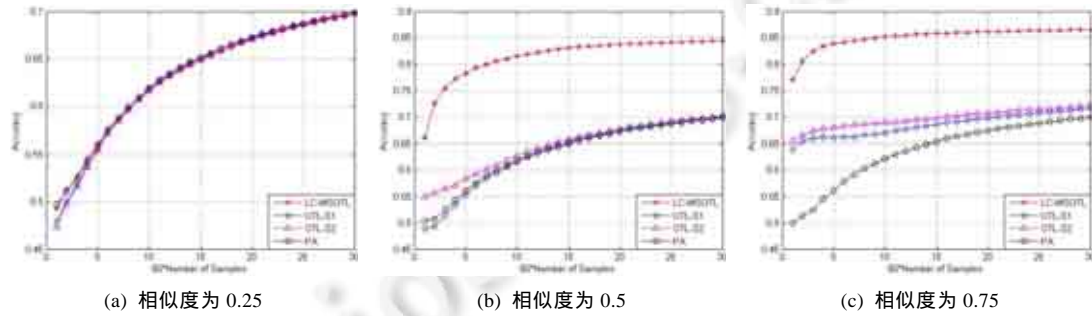


Fig.5 Classification accuracy change for NOPQ data set with different similarities

图 5 NOPQ 数据集在不同相似度下的分类准确率变化

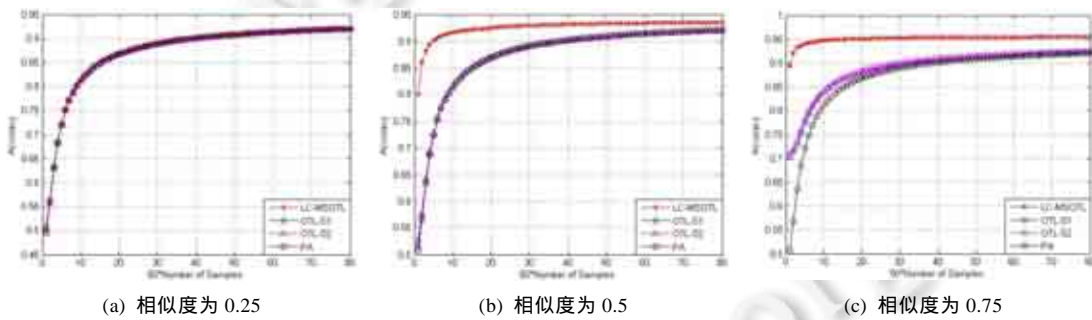


Fig.6 Classification accuracy change for GaoSi data set with different similarities

图 6 GaoSi 数据集在不同相似度下的分类准确率变化

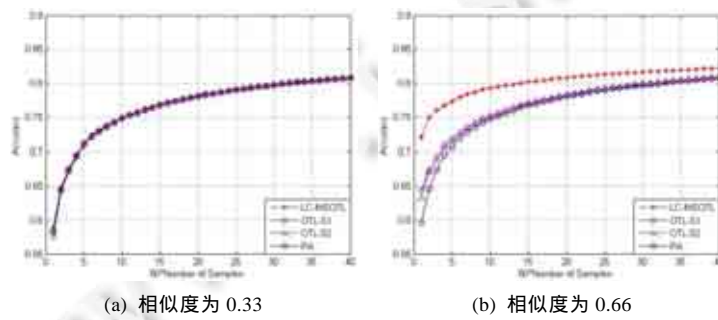


Fig.7 Classification accuracy change for Waveform data set with different similarities

图 7 Waveform 数据集在不同相似度下的分类准确率变化

从表 6 及图 4~图 7 可以看出,在源领域与目标领域的相似度较低时,如 0.25 或 0.33 的情况下,相对于 PA,OTL 的分类准确率不但没有提升反而有所下降.当相似度提高到 0.5 时,相对于 PA,OTL 的分类准确率也没有明显的提高,而显示出两者相当,同样看不出迁移学习带来的效益.这是由于源领域与目标领域的相似度较低时,OTL 算法使用弱源的知识进行迁移时产生了一定的负迁移.直到相似度提高到 0.75 或 0.66 时,OTL 的分类准确率才高于 PA.这是因为源领域和目标领域较相似时,迁移学习才能发挥作用.

非常有意思的是,从表 6 以及图 4~图 7 还可以看出,无论源领域和目标领域的相似度是低还是高,相对于 PA 和 OTL,LC-MSOTL 在大多数情况下的分类准确率都要好,而且在分类的初始阶段显示出较大的提高.这是因为相对于 PA 没有使用源领域和 OTL 只使用单个源领域,LC-MSOTL 使用了两个源领域进行迁移,其选源策略降低了负迁移的影响.从图 4~图 7 中进一步还可以看出,在相似度为 0.5,0.75 或者 0.66 情况下,相对于其他算法,LC-MSOTL 算法在数据流初期能够更快地提升分类准确率.这对于在线迁移学习意义非常大.这是因为,由于目标领域的样本是在线到来的,在初期能采集到的目标领域样本较少,这时更需要迁移学习的帮助,而 LC-MSOTL 算法能够利用多源迁移更好地提高对目标领域样本的分类准确率.

### 3.2.2 LC-MSOTL 整合多个互补弱源的迁移性能实验

为了进一步分析 LC-MSOTL 选源策略对负迁移的克服,本节进行了两组实验,以验证 LC-MSOTL 从多个互补弱源到目标领域的迁移学习.在本节实验中,各源领域与目标领域相似度不高,且各源领域与目标领域间的相似区域又互不相同,本节称这样的源领域为互补弱源.

在第 1 组实验中,分别在 ABCD,NOPQ,GaoSi 数据集上设计了 4 个源领域和 1 个目标领域.每个源领域与目标领域的相似度都仅为 0.25,且各源领域间的相似区域完全不同.数据生成的具体方法如下:在 ABCD 数据集各字母类中,分别随机抽取 100 个样本组成源 1 领域数据,进而从剩余的样本中按字母类分别随机抽取 100 个样本组成源 2 领域数据,重复上面的过程,直到生成 4 个源领域数据,剩余的 1 496 个样本组成目标领域数据.NOPQ 数据集也按照 ABCD 数据集的方式,组成 4 个源领域数据和 1 个目标领域数据.目标领域包含 1 522 样本.在 GaoSi 数据集各类中,分别随机抽取 250 个样本组成源 1 领域数据,进而从各类剩余的样本中分别随机抽取 250 个样本组成源 2 领域数据,重复上面的过程,直到生成 4 个源领域数据,剩余的 4 000 个样本组成目标领域数据.上述的实验数据按照表 7 打上“+”或“-”大类标签,由表 7 可以看出,每一个源领域与目标领域的相似度都为 0.25,且源领域间相似区域完全不相同,且各源领域与目标领域的分布相同区域的并集覆盖目标领域.

**Table 7** Distribution of the data sets of ABCD, NOPQ, and GaoSi with four source domains when the similarity is set as 0.25

表 7 相似度为 0.25 时,ABCD,NOPQ,GaoSi 数据集的 4 个互补源领域和目标领域数据分布

ABCD	模式类			源 1 领域	源 2 领域	源 3 领域	源 4 领域	目标领域
	NOPQ	GaoSi						
A	N	1	-	+	+	+	-	
B	O	2	-	-	+	-	+	
C	P	3	-	-	-	+	+	
D	Q	4	+	-	+	+	-	

按照表 7 的数据分布要求,为不同的数据集生成 10 份实验数据,为每份生成的数据重复执行 10 次实验,每次实验之前随机打乱目标领域样本,并且假定目标领域样本按照数据流的形式一个一个地到来.各算法的分类准确率及其变化情况如表 8 以及图 8 所示.所有的实验结果为 100 次实验的平均结果.

**Table 8** Classification accuracy of each method for the data sets of ABCD, NOPQ, and GaoSi when the similarity is set as 0.25

表 8 相似度为 0.25 时,各算法对 ABCD,NOPQ,GaoSi 数据集的分类准确率

准确率	PA	OTL-S1	OTL-S2	OTL-S3	OTL-S4	LC-MSOTL
ABCD	0.759 12	0.758 78	0.758 46	0.758 24	0.758 45	<b>0.899 41</b>
NOPQ	0.697 14	0.696 16	0.696 26	0.696 34	0.697 20	<b>0.894 67</b>
GaoSi	0.924 66	0.924 46	0.924 51	0.924 48	0.924 48	<b>0.946 93</b>

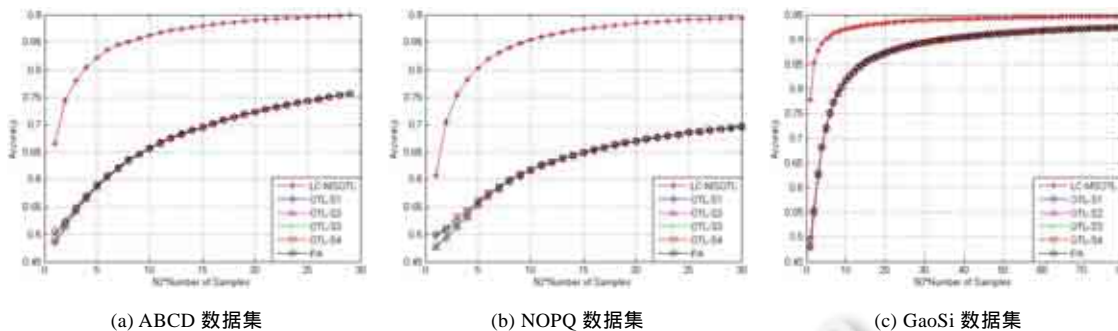


Fig.8 Classification accuracy change of each method for different data sets

图 8 各算法对各数据集的分类准确率变化

从表 8 和图 8 可以看出,绝大多数使用单个弱源进行迁移的 OTL 算法相对于 PA 的分类准确率有所下降.这再一次验证了:由于每个源领域与目标领域的相似度较低,而造成负迁移现象.但是,LC-MSOTL 的准确率相对于其他对比算法却有了显著提高,最大可以超出对比算法分类准确率的 20%.这说明在每个源领域皆与目标领域相似程度较低的情况下,LC-MSOTL 可以整合多个互补的弱源进行迁移,对目标领域的每一个样本选择合适的源对其辅助分类,从而可以有效地消除负迁移造成的影响.

在第 2 组实验中,试图探讨弱源的数量对 LC-MSOTL 分类准确率的影响.在 ABCD,NOPQ,GaoSi 数据集上,将使用两个相似度为 0.25 的源领域进行迁移的 LC-MSOTL(记为 LC-MSOTL-0.25\*2)和使用一个相似度为 0.5 的源领域进行迁移的 OTL(记为 OTL-0.5)进行对比;使用 3 个相似度 0.25 的源领域进行迁移的 LC-MSOTL(记为 LC-MSOTL-0.25\*3)和使用一个相似度为 0.75 的源领域进行迁移的 OTL(记为 OTL-0.75)进行对比.在 Waveform 数据集上,将使用两个相似度 0.33 的源领域进行迁移的 LC-MSOTL(记为 LC-MSOTL-0.33\*2)和使用一个相似度为 0.66 的源领域进行迁移的 OTL(记为 OTL-0.66)进行对比.总计 7 个对比实验.

在以上各个对比实验中,各组实验所使用的所有源领域中的原始数据相同,只是按照不同的相似度要求给样本标记上“+”或“-”大类标签.例如,2 个 0.25 的源领域的全体原始数据与 1 个 0.5 的源领域的全体原始数据相同,仅按照不同的相似度要求标记上不同的大类标签,并且目标领域完全相同.各组对比实验所使用的数据的分布见表 9~表 11.

Table 9 Experimental data sets for LC-MSOTL-0.25\*2 and OTL-0.5

表 9 LC-MSOTL-0.25\*2 和 OTL-0.5 的实验数据集

模式类			源 1 领域	源 2 领域	单源领域	目标领域
ABCD	NOPQ	Gaosi				
A	N	1	+	+	+	-
B	O	2	-	+	+	+
C	P	3	-	-	-	+
D	Q	4	-	+	-	-

Table 10 Experimental data sets for LC-MSOTL-0.25\*3 and OTL-0.75

表 10 LC-MSOTL-0.25\*3 和 OTL-0.75 的实验数据集

模式类			源 1 领域	源 2 领域	源 3 领域	单源领域	目标领域
ABCD	NOPQ	Gaosi					
A	N	1	+	+	+	+	-
B	O	2	-	+	-	+	+
C	P	3	-	-	+	+	+
D	Q	4	-	+	+	-	-

**Table 11** Experimental data sets for LC-MSOTL-0.33\*2 and OTL-0.66

表 11 LC-MSOTL-0.33\*2 和 OTL-0.66 的实验数据集

模式类	源 1 领域	源 2 领域	单源领域	目标领域
0	+	-	-	-
1	-	+	+	-
2	-	-	+	+

表 9 中,在 ABCD,NOPQ 数据集上,源 1 和源 2 领域样本量之和与单源领域的样本量相等,为 1 800 个,目标领域样本数量分别为 1 296 个和 1 322 个;GaoSi 数据集上,源 1 和源 2 领域样本量之和与单源领域的样本量相等,为 4 800 个,目标领域样本数量为 3 200 个.表 10 中,在 ABCD,NOPQ 数据集上,源 1、源 2 和源 3 领域样本量之和与单源领域的样本量相等,为 1 800 个,目标领域样本数量分别为 1 296 个和 1 322 个;在 GaoSi 数据集上,源 1、源 2 和源 3 领域样本量之和与单源领域的样本量相等,为 4 800 个,目标领域样本数量为 3 200 个.表 11 中,在 Waveform 数据集上,源 1 和源 2 领域样本量之和与单源领域的样本量相等,为 3 000 个,目标领域的样本数量为 2 000 个.

按照表 9~表 11 的数据分布要求,为不同的数据集生成 10 份实验数据,为每份生成的数据重复执行 10 次实验,每次实验之前随机打乱目标领域数据,并且假定目标领域样本按照数据流的形式一个一个地到来.各算法的分类准确率及其变化情况如表 12、表 13 及图 9、图 10 所示.所有的实验结果为 100 次实验的平均结果.需要说明的是,由于 LC-MSOTL-0.25\*2,OTL-0.5 及 PA 的分类准确率曲线几乎重合,也就是说,迁移学习没有发挥作用,所以本节没有提供其实验结果.

从表 12、表 13 可以看到,使用更多互补弱源的 LC-MSOTL 与使用单个更强源的 OTL 的分类准确率基本相等.从图 9、图 10 则可以看到,随着时间的延续,LC-MSOTL 和 OTL 的分类准确率越来越接近.这些结果说明,在对目标领域样本实施在线学习的过程中,LC-MSOTL 能够整合多个互补弱源实施迁移学习.

**Table 12** Classification accuracy contrast for ABCD, NOPQ and GaoSi data sets under different similarity

表 12 ABCD,NOPQ 和 GaoSi 数据集在不同相似度下的分类准确率对比

数据集	PA	LC-MSOTL-0.25*3	OTL-0.75
ABCD	0.745 1	0.750 7	<b>0.753 6</b>
NOPQ	0.686 0	<b>0.739 2</b>	0.719 6
GaoSi	0.917 7	0.919 7	<b>0.923 1</b>

**Table 13** Classification accuracy contrast for Waveform data set under different similarity

表 13 Waveform 数据集在不同相似度下的分类准确率对比

数据集	PA	LC-MSOTL-0.33*2	OTL-0.66
Waveform	0.803 3	<b>0.807 52</b>	0.806 96

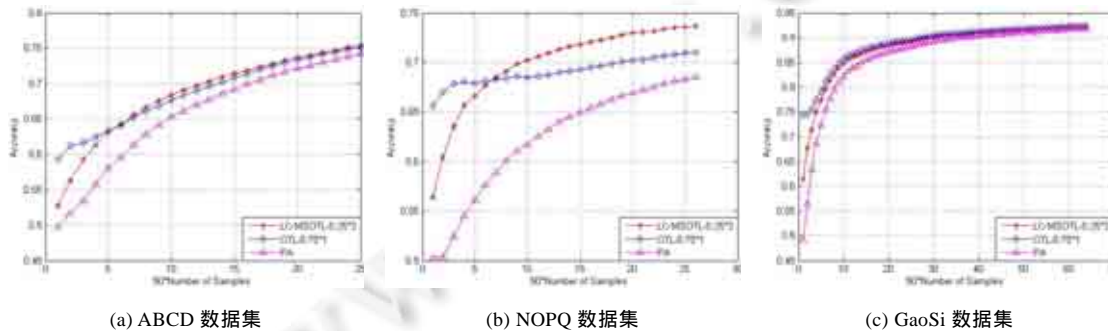


Fig.9 Classification accuracy change for different data sets with LC-MSOTL-0.25\*3 and OTL-0.75

图 9 LC-MSOTL-0.25\*3 与 OTL-0.75 对各数据集的分类准确率变化

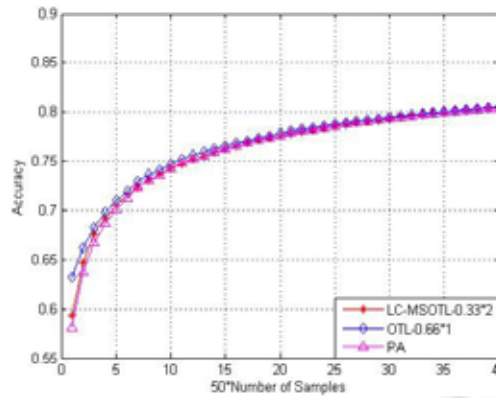
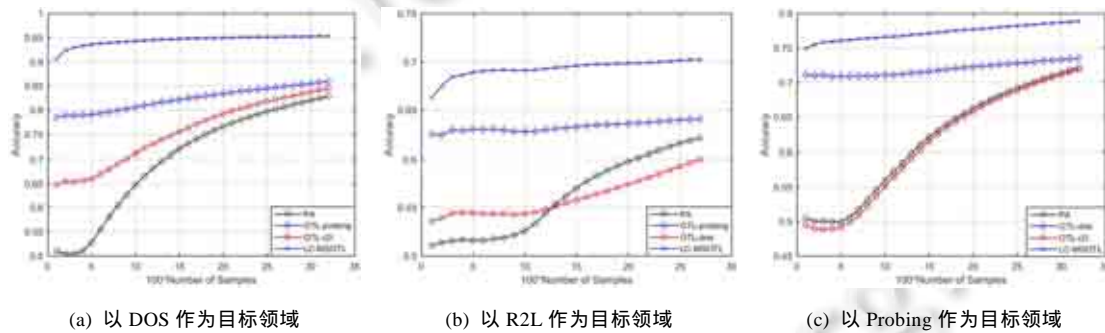


Fig.10 Classification accuracy change for Waveform data set with LC-MSOTL-0.33\*2 and OTL-0.66

图 10 LC-MSOTL-0.33\*2 与 OTL-0.66 对 Waveform 数据集的分类准确率变化

### 3.2.3 LC-MSOTL 对 Intrusion detection 数据集的实验

在本节实验中,PA 的参数为  $\sigma=16$ ,惩罚系数  $C=2$ .从图 11 可以看出,在不清楚源领域和目标领域相似程度及相似的局部区域的情形下,LC-MSOTL 的分类准确率仍然比 PA 和 OTL 要好.更进一步地,从图 11(a)还可以看到,当两个源对目标领域都有较好的迁移学习效果时,LC-MSOTL 在分类初期相对于 PA 较大幅度地提高了分类准确率;从图 11(b)、图 11(c)还可以看到,当两个源领域中只有 1 个对目标领域显示出较好的迁移学习效果时,LC-MSOTL 仍然在分类初期相对于 PA 有效地提高了分类准确率.



(a) 以 DOS 作为目标领域

(b) 以 R2L 作为目标领域

(c) 以 Probing 作为目标领域

Fig.11 Classification accuracy change of each method for different target domains

图 11 各算法在不同目标领域下的分类准确率变化

### 3.2.4 LC-MSOTL 参数敏感性实验

在不同的参数设置下,采用第 3.2.1 节中的方式进行实验,以分析各参数对 LC-MSOTL 分类准确率的影响.

#### (1) 近邻个数 $K$ 对 LC-MSOTL 算法的影响

为了验证近邻个数  $K$  对 LC-MSOTL 算法的影响,在这部分实验中,将固定放大系数  $\zeta$  设置为 1,以测试  $K$  对实验效果的影响.在该实验中, $K$  的取值依次是 3,5,10,20,50,100,200.

从图 12(a)可以看出,在 ABCD 数据集上,不管相似度是 0.75 还是 0.5,随着  $K$  值的不断增大,分类准确率越来越低.这说明在 ABCD 数据集上, $K=3$  是比较好的近邻个数取值;且随着  $K$  值的增大,分类器选择的越来越不准确.从图 12(b)可以看出,NOPQ 数据集在相似度 0.75 时,随着  $K$  的增大,分类准确率先轻微上升,之后逐步下降;在相似度为 0.5 时,NOPQ 数据集随着  $K$  值的不断增大,准确率下降很明显.从图 12(c)可以看出,在 GaoSi 数据集上,当  $K=10$  时准确率达到最大,说明此时分类器选择较为准确;随着  $K$  值的增大,准确率越来越低.从图 12(d)可以看出,在 Waveform 数据集上,当  $K=20$  时,此时的分类准确率达到最大;且当  $K<20$  时,随着  $K$  值的增大,准确率



提高的速度较快;当  $K > 20$  时,随着  $K$  值的增大,分类准确率越来越低.这些都说明,在 LC-MSOTL 算法中, $K$  的表现与  $K$  近邻分类算法中的  $K$  的表现基本一致.因此, $K$  的取值与数据集相关.

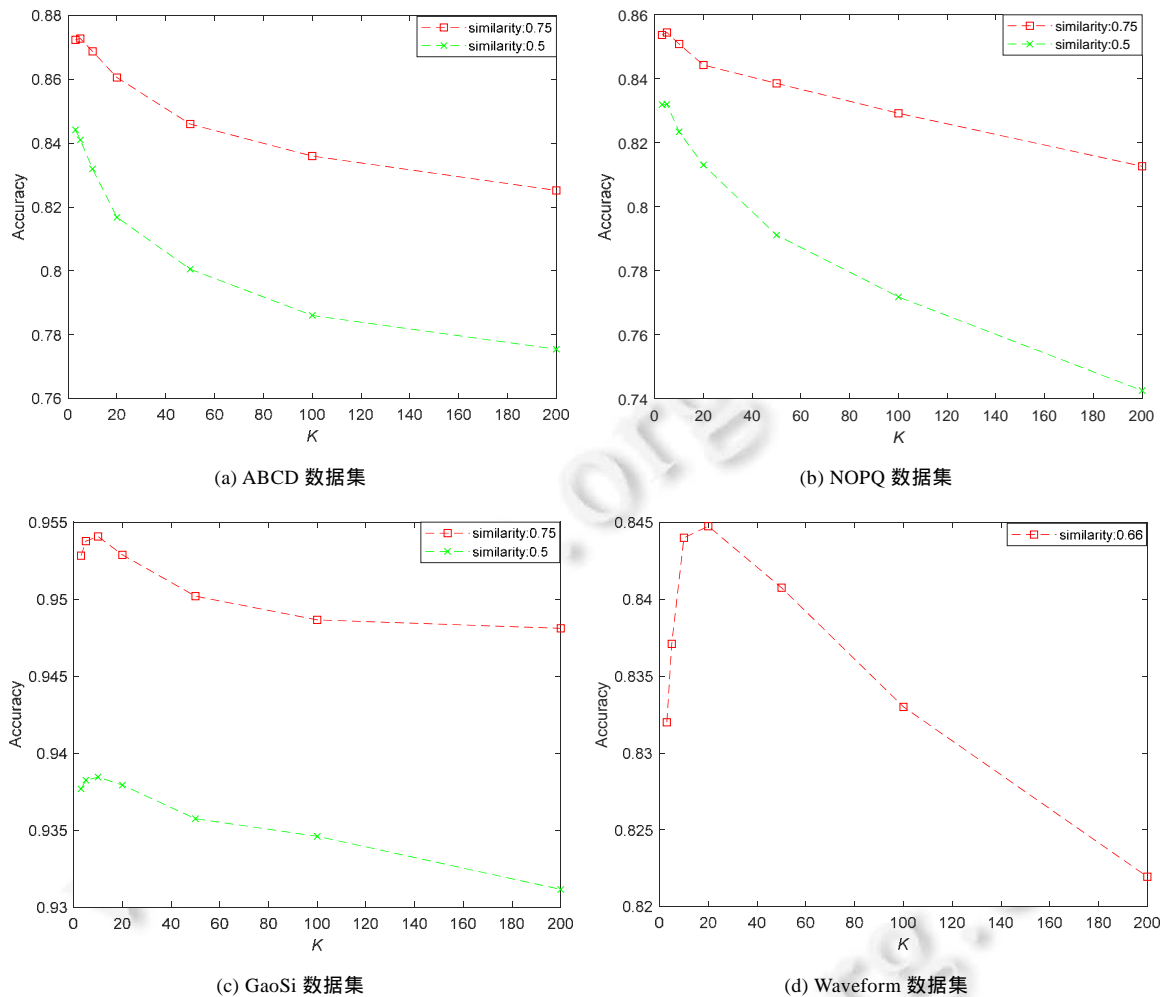


Fig.12 Classification accuracy change for different data sets with different  $K$

图 12 各数据集在不同  $K$  上的准确率变化图

## (2) 放大系数 $\zeta$ 对 LC-MSOTL 分类准确率的影响

为了分析放大系数  $\zeta$  对 LC-MSOTL 分类准确率的影响,在这部分实验中,固定近邻个数  $K$ ,以测试放大系数  $\zeta$  对分类准确率的影响.由于近邻个数  $K$  较小时  $\zeta$  对分类准确率的影响较小,故在该实验中,为了显示出放大系数  $\zeta$  对分类准确率的影响,将近邻个数  $K$  的取值设置为 200,在该实验中, $\zeta$  的取值依次是 1,3,5,10.

从图 13 中可以看出,当近邻个数  $K$  取值较大且放大系数  $\zeta$  较小时,LC-MSOTL 分类准确率相对较低;在各个数据集上,随着放大系数  $\zeta$  的不断增大,LC-MSOTL 的分类准确率逐渐提高,充分表明了放大系数  $\zeta$  可以缓解当近邻个数  $K$  较大时,LC-MSOTL 分类准确率相对较低的问题.如表 14~表 16 所示,尽管  $K$  较大,但通过增大  $\zeta$ ,可以得到与  $K$  较小时相当的分类准确率.换句话说, $\zeta$  具有放松对  $K$  的选择的功能.其中原因在于,从  $LC$  的定义可知,若  $K$  比较大,则  $x_i$  的近邻中可能存在离它较远的样本,但这些离  $x_i$  较远的样本则由于放大系数  $\zeta$  的作用而减弱对  $LC$  值的不良影响.需指出的是,表 14~表 16 中之所以没有提供 GaoSi 数据集上类似的实验结果,是因为在 GaoSi 数据集上调节近邻个数  $K$  以及放大系数  $\zeta$  时,分类准确率波动很小.

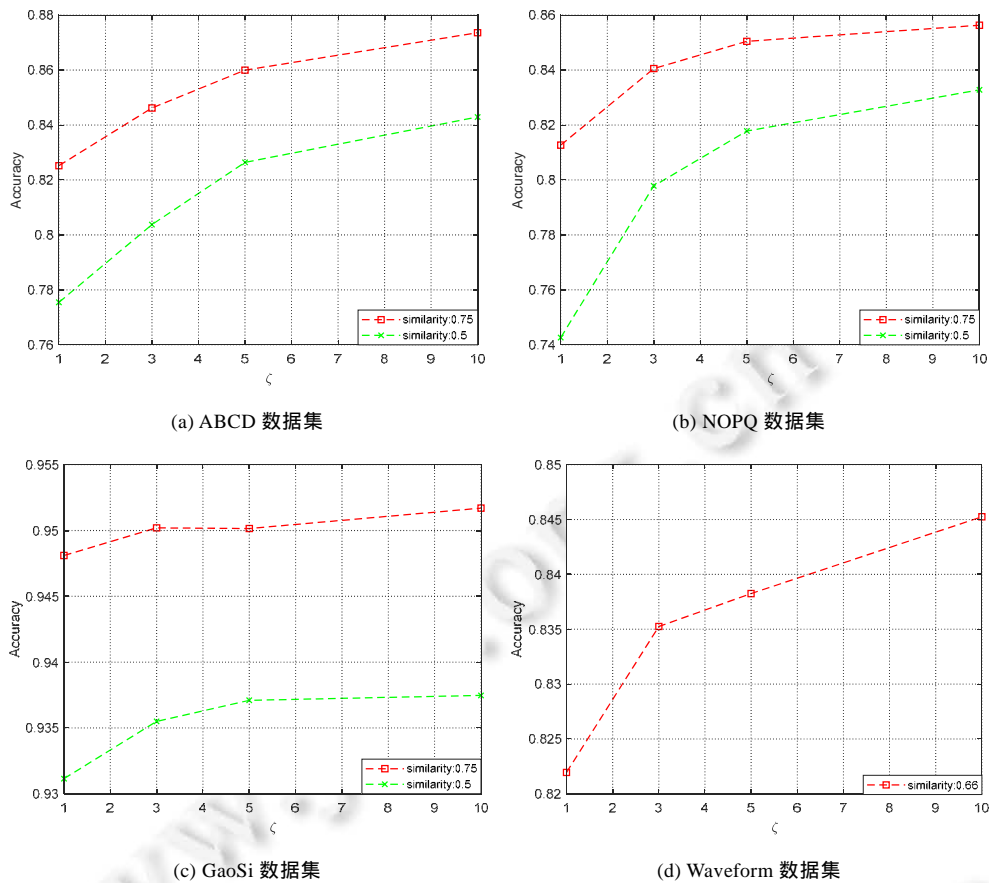


Fig.13 Classification accuracy change for different data sets with different amplification factor

图 13 各数据集在不同放大系数上的准确率变化图

Table 14 Comparison of ABCD data sets

表 14 ABCD 数据集对比

相似度 0.75			相似度 0.5		
$K$	$\zeta$	分类准确率	$K$	$\zeta$	分类准确率
5	1	0.872 794 118	5	1	0.841 042 781
200	10	0.873 596 257	200	10	0.842 914 439

Table 15 Comparison of NOPQ data sets

表 15 NOPQ 数据集对比

相似度 0.75			相似度 0.5		
$K$	$\zeta$	分类准确率	$K$	$\zeta$	分类准确率
5	1	0.854 467 806	5	1	0.832 063 075
200	10	0.856 307 49	200	10	0.832 785 808

Table 16 Comparison of Waveform data sets

表 16 Waveform 数据集对比

相似度 0.66		
$K$	$\zeta$	分类准确率
20	1	0.844 75
200	10	0.845 25

## 4 总结

本文结合多源迁移学习和在线学习,提出了一种适用于数据流环境的多源在线迁移学习算法——LC-MSOTL.本文首先提出了一种局部分类精度的计算方法,并从理论上初步分析了引入局部分类精度选取源领域分类器方法的合理性.LC-MSOTL 利用局部分类精度动态地从源领域分类器集合选取局部分类精度最高的分类器,将该分类器和目标领域分类器加权集成.进一步的实验结果表明,LC-MSOTL 可以从多个源领域中选择合适的源领域分类器,能够有效地利用多个源领域提高对目标领域样本的分类准确率.因此,当单个源领域与目标领域的相似性不高时,通过增加源领域数量,同时确保源领域之间的区别度,可以有效地提高迁移学习效果.未来的工作包括:研究其他选取源领域的方法;研究确定源领域与目标领域局部相似性的方法;研究如何将LC-MSOTL 用于概念漂移数据流分类等.

### References:

- [1] Gama J, Zliobaite I, Bifet A, Pechenizkiy M, Bouchachia A. A survey on concept drift adaptation. *ACM Computing Surveys*, 2014, 46(4):1–37. [doi: 10.1145/2523813]
- [2] Wen YM, Qiang BH, Fan ZG. A survey of the classification of data streams with concept drift. *CAAI Trans. on Intelligent Systems*, 2013,8(2):95–104 (in Chinese with English abstract).
- [3] Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans. on Knowledge and Data Engineering*, 2010,22(10):1345–1359. [doi: 10.1109/TKDE.2009.191]
- [4] Zhuang FZ, Luo P, He Q, Shi ZZ. Survey on transfer learning research. *Ruan Jian Xue Bao/Journal of Software*, 2015,26(1):26–39 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4631.htm> [doi: 10.13328/j.cnki.jos.004631]
- [5] Pan W, Zhong H, Xu CF, Ming Z. Adaptive Bayesian personalized ranking for heterogeneous implicit feedbacks. *Journal of Knowledge-Based Systems*, 2015,73(1):173–180. [doi: 10.1016/j.knosys.2014.09.013]
- [6] Pan W, Yang Q. Transfer learning in heterogeneous collaborative filtering domains. *Journal of Artificial Intelligence*, 2013,197(4): 39–55. [doi: 10.1016/j.artint.2013.01.003]
- [7] Zhuang FZ, Luo P, Yin PF, He Q, Shi ZZ. Concept learning for cross-domain text classification: A general probabilistic framework. In: *Proc. of the 23th Int'l Joint Conf. on Artificial Intelligence*. Palo Alto: AAAI, 2013. 1960–1966.
- [8] Zhuang FZ, Luo P, Du CY, He Q, Shi ZZ. Triplex transfer learning: Exploiting both shared and distinct concepts for text classification. *IEEE Trans. on Cybernetics*, 2014,44(7):1191–1203. [doi: 10.1109/TCYB.2013.2281451]
- [9] Dai W, Yang Q, Xue GR, Yu Y. Boosting for transfer learning. In: *Proc. of the 24th Int'l Conf. on Machine Learning*. New York: ACM Press, 2007. 193–200. [doi: 10.1145/1273496.1273521]
- [10] Pan SJL, Tsang IW, Kwok JT, Yang Q. Domain adaptation via transfer component analysis. *IEEE Trans. on Neural Networks*, 2011, 22(2):199–210. [doi: 10.1109/TNN.2010.2091281]
- [11] Zhang Q, Li M, Wang XS, Cheng YF, Zhu MQ. Instance-Based transfer learning for multi-source domains. *Acta Automatica Sinica*, 2014,40(6):1176–1183 (in Chinese with English abstract).
- [12] Yao Y, Doretto G. Boosting for transfer learning with multiple sources. In: *Proc. of the Computer Vision and Pattern Recognition*. New York: IEEE, 2010. 1855–1862. [doi: 10.1109/CVPR.2010.5539857]
- [13] Eaton E, Desjardins M. Selective transfer between learning tasks using task-based boosting. In: *Proc. of the AAAI Conf. on Artificial Intelligence*. Vancouver: AAAI Press, 2011. 337–342.
- [14] Huang PP, Wang G, Qin SY. Boosting for transfer learning from multiple data sources. *Pattern Recognition Letters*, 2012,33(5): 568–579. [doi: 10.1016/j.patrec.2011.11.023]
- [15] Duan LX, Tsang IW, Xu D, Chua TS. Domain adaptation from multiple sources via auxiliary classifiers. In: *Proc. of the 26th Annual Int'l Conf. on Machine Learning*. New York: ACM Press, 2009. 289–296. [doi: 10.1145/1553374.1553411]
- [16] Chattopadhyay R, Ye JP, Panchanathan S, Fan W, Davidson I. Multi-Source domain adaptation and its application to early detection of fatigue. In: *Proc. of the ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*. New York: ACM Press, 2011. 717–725. [doi: 10.1145/2020408.2020520]
- [17] Zhao PL, Hoi SCH, Wang JL, Li B. Online transfer learning. *Artificial Intelligence*, 2014,216(16):76–102. [doi: 10.1016/j.artint.2014.06.003]

- [18] Zhao PL, Hoi SCH. OTL: A framework of online transfer learning. In: Proc. of the Int'l Conf. on Machine Learning. New York: ACM Press, 2010. 1231–1238.
- [19] Crammer K, Dekel O, Keshet J, Shalev-Shwartz S, Singer Y. Online passive-aggressive algorithms. *Journal of Machine Learning Research*, 2006,7(3):551–585.
- [20] Li ZJ, Li YC, Wang F, He GL, Kuang L. Online learning algorithms for big data analytics: A survey. *Journal of Computer Research and Development*, 2015,52(8):1707–1721 (in Chinese with English abstract).
- [21] Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 1958,65(6):386–408. [doi: 10.1037/h0042519]
- [22] Cesa-Bianchi N, Conconi A, Gentile C. A second-order perceptron algorithm. *SIAM Journal on Computing*, 2002,34(3):640–668. [doi: 10.1137/S0097539703432542]
- [23] Dredze M, Grammer K, Pereira F. Confidence-Weighted linear classification. In: Proc. of the 25th Int'l Conf. on Machine Learning. New York: ACM Press, 2008. 264–271. [doi: 10.1145/1390156.1390190]
- [24] Langford J, Li L, Zhang T. Sparse online learning via truncated gradient. *Journal of Machine Learning Research*, 2008,10(2):777–801.
- [25] Freund Y, Schapire RE. Large margin classification using the perceptron algorithm. *Journal of Machine Learning*, 1999,37(3):277–296. [doi: 10.1023/A:1007662407062]
- [26] Fan HJ, Song Q, Shrestha SB. Online learning with kernel regularized least mean square algorithms. *Knowledge-Based Systems*, 2014,59(2):21–32. [doi: 10.1016/j.knosys.2014.02.005]
- [27] Yang HQ, Lyu MR, King I. Efficient online learning for multitask feature selection. *ACM Trans. on Knowledge Discovery from Data*, 2013,7(2):1–27. [doi: 10.1145/2499907.2499909]
- [28] Domingos P, Hulten G. Mining high-speed data streams. In: Proc. of the ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2000. 71–80. [doi: 10.1145/347090.347107]
- [29] Gama J, Rocha R, Medas P. Accurate decision trees for mining high-speed data streams. In: Proc. of the ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2003. 523–528. [doi: 10.1145/956750.956813]
- [30] Guo GD, Huang J, Chen LF. KNN model based incremental learning algorithm. *Pattern Recognition and Artificial Intelligence*, 2010,23(5):701–707 (in Chinese with English abstract).
- [31] Weiss K, Khoshgoftaar TM, Wang DD. A survey of transfer learning. *Journal of Big Data*, 2016,3:9. <https://doi.org/10.1186/s40537-016-0043-6>
- [32] Ling X, Dai WY, Xue GR, Yang Q, Yu Y. Spectral domain-transfer learning. In: Proc. of the ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2008. 488–496. [doi: 10.1145/1401890.1401951]
- [33] Jiang W, Zavesky E, Chang SF, Loui A. Cross-Domain learning methods for high-level visual concept classification. In: Proc. of the IEEE Int'l Conf. on Image Processing. IEEE, 2008. 161–164. [doi: 10.1109/ICIP.2008.4711716]
- [34] Dai WY, Xue GR, Yang Q, Yu Y. Co-Clustering based classification for out-of-domain documents. In: Proc. of the ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2007. 210–219. [doi: 10.1145/1281192.1281218]
- [35] Dai WY, Xue GR, Yang Q, Yu Y. Transferring naive Bayes classifiers for text classification. In: Proc. of the National Conf. on Artificial Intelligence. Palo Alto: AAAI, 2007. 540–545.
- [36] Hong JM, Yin J, Huang Y, Liu YB, Wang JH. TrSVM: A transfer learning algorithm using domain similarity. *Journal of Computer Research and Development*, 2011,48(10):1823–1830 (in Chinese with English abstract).
- [37] Davis J, Domingos P. Deep transfer via second-order Markov logic. In: Proc. of the Int'l Conf. on Machine Learning. New York: ACM Press, 2009. 217–224. [doi: 10.1145/1553374.1553402]
- [38] Sun SL, Shi HL, Wu YB. A survey of multi-source domain adaptation. *Information Fusion*, 2015,24(C):84–92. [doi: 10.1016/j.inffus.2014.12.003]
- [39] Eaton E, desJardins M, Lane T. Modeling transfer relationships between learning tasks for improved inductive transfer. In: Proc. of the European Conf. on Machine Learning and Knowledge Discovery in Databases. Berlin: Springer-Verlag, 2008. 317–332. [doi: 10.1007/978-3-540-87479-9\_39]
- [40] Luo P, Zhuang FZ, Xiong H, Xiong YH, He Q. Transfer learning from multiple source domains via consensus regularization. In: Proc. of the ACM Conf. on Information and Knowledge Management. New York: ACM Press, 2008. 103–112. [doi: 10.1145/1458082.1458099]

- [41] Wang XS, Pan J, Chen YH, Cao G. Self-Adaptive transfer for decision trees based on similarity metric. *Acta Automatica Sinica*, 2013,39(12):2186–2192 (in Chinese with English abstract).
- [42] Gu Q, Zhou J. Learning the shared subspace for multi-task clustering and transductive transfer classification. In: *Proc. of the IEEE Int'l Conf. on Data Mining*. New York: IEEE, 2009. 159–168. [doi: 10.1109/ICDM.2009.32]
- [43] Tommasi T, Orabona F, Caputo B. Safety in numbers: Learning categories from few examples with multi model knowledge transfer. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. New York: IEEE, 2010. 3081–3088. [doi: 10.1109/CVPR.2010.5540064]
- [44] Cawley GC. Leave-One-Out cross-validation based model selection criteria for weighted LS-SVMs. In: *Proc. of the Int'l Joint Conf. on Neural Networks*. New York: IEEE, 2006. 1661–1668. [doi: 10.1109/IJCNN.2006.246634]
- [45] Pan SSJ, Ni XC, Sun JT, Yang Q, Chen Z. Cross-Domain sentiment classification via spectral feature alignment. In: *Proc. of the 19th Int'l Conf. on World Wide Web*. New York: ACM Press, 2010. 751–760. [doi: 10.1145/1772690.1772767]
- [46] Gao J, Fan W, Jiang J, Han JW. Knowledge transfer via multiple model local structure mapping. In: *Proc. of the ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*. New York: ACM Press, 2008. 283–291. [doi: 10.1145/1401890.1401928]
- [47] Ge L, Gao J, Zhang AD. OMS-TL: A framework of online multiple source transfer learning. In: *Proc. of the ACM Int'l Conf. on Information and Knowledge Management*. New York: ACM Press, 2013. 2423–2428. [doi: 10.1145/2505515.2505603]
- [48] Zhou ZH, Li M. Semi-Supervised regression with co-training style algorithms. *IEEE Trans. on Knowledge and Data Engineering*, 2007,19(11):1479–1493. [doi: 10.1109/TKDE.2007.190644]
- [49] Zhou ZH, Li M. Semi-Supervised regression with co-training. In: *Proc. of the 19th Int'l Joint Conf. on Artificial Intelligence*. San Francisco: Morgan Kaufmann Publishers, 2005. 908–916.
- [50] Woods K, Bowyer K, Jr WPK. Combination of multiple classifiers using local accuracy estimates. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1997,19(4):405–410. [doi: 10.1109/34.588027]
- [51] Liu M, Yuan ZB, Miao ZJ, Tang XF, Li KL. Transformation from local accuracy to classification confidence. *Journal of Computer Research and Development*, 2008,45(9):1612–1619 (in Chinese with English abstract).

#### 附中文参考文献:

- [2] 文益民,强保华,范志刚.概念漂移数据流分类研究综述. *智能系统学报*,2013,8(2):95–104.
- [4] 庄福振,罗平,何清,史忠植.迁移学习研究进展. *软件学报*,2015,26(1):26–39. <http://www.jos.org.cn/1000-9825/4631.htm> [doi: 10.13328/j.cnki.jos.004631]
- [11] 张倩,李明,王雪松,程玉虎,朱美强.一种面向多源领域的实例迁移学习. *自动化学报*,2014,40(6):1176–1183.
- [20] 李志杰,李元香,王峰,何国良,匡立.面向大数据分析的在线学习算法综述. *计算机研究与发展*,2015,52(8):1707–1721.
- [30] 郭躬德,黄杰,陈黎飞.基于 KNN 模型的增量学习算法. *模式识别与人工智能*,2010,23(5):701–707.
- [36] 洪佳明,印鉴,黄云,刘玉葆,王甲海.TrSVM:一种基于领域相似性的迁移学习算法. *计算机研究与发展*,2011,48(10):1823–1830.
- [41] 王雪松,潘杰,程玉虎,曹戈.基于相似度衡量的决策树自适应迁移. *自动化学报*,2013,39(12):2186–2192.
- [51] 刘明,袁保宗,苗振江,唐晓芳,李昆仑.从局部分类精度到分类置信度的变换. *计算机研究与发展*,2008,45(9):1612–1619.



唐诗淇(1990 - ),男,湖南长沙人,硕士,主要研究领域为机器学习,迁移学习,数据流分类.



秦一休(1992 - ),男,学士,主要研究领域为机器学习,数据挖掘.



文益民(1969 - ),男,博士,教授,CCF 高级会员,主要研究领域为机器学习,迁移学习,数据流分类,教育数据挖掘.