

知识嵌入的迁移孪生支持向量机

王洪元[†], 耿磊, 倪彤光, 王冲

(常州大学 信息学院, 江苏 常州 213016)

摘要: 孪生支持向量机(TwinSVM)相比支持向量机在解决类别不平衡数据问题上具有优势,但其在训练数据不足时受训所得分类器的泛化能力较差. 针对此问题,探讨一种知识嵌入的迁移孪生支持向量机(KE-T-TwinSVM). 该分类器不但继承了TwinSVM的优点,还可基于知识嵌入的思想利用从相关领域学到的知识来辅助学习以提高分类效果. 各种真实数据集上的实验结果表明,所提出的分类器在目标领域数据不足和不平衡情况下具有更佳的性能.

关键词: 分类; 孪生支持向量机; 迁移学习; 不平衡数据

中图分类号: TP311

文献标志码: A

Knowledge embedded transfer twin support vector machine

WANG Hong-yuan[†], GENG Lei, NI Tong-guang, WANG Chong

(School of Information, Changzhou University, Changzhou 213016, China)

Abstract: A twin support vector machines (TwinSVM) performs better than a support vector machine (SVM) in dealing with the class-imbalanced classification problem. However, the TwinSVM has weak generalization abilities when there are not enough training data. Therefore, a knowledge embedded transfer twin support vector machine (KE-T-TwinSVM) is proposed. The KE-T-TwinSVM inherits the characteristic of the TwinSVM and fully considers the data of target domain as well as the knowledge of source domain by leveraging the knowledge of source domain in the target domain. The experimental results show that the proposed KE-T-TwinSVM has better performance compared with the traditional classifiers in the situation of insufficient data and class-imbalanced data.

Keywords: classification; twin support vector machines; transfer learning; class-imbalance data

0 引言

随着社会的进步和互联网技术的快速发展,人们获取信息的手段和规模都有了极大的改善,随之而来的问题是如何利用这些数据. 机器学习作为数据挖掘和模式识别的重要手段之一,越来越引起人们的关注. 例如,遗传算法、决策树、贝叶斯网络、支持向量机等方法被广泛地应用到人脸识别检测系统、计算机视觉和网络安全检测等多个领域中^[1-2],并取得了众多成果. 虽然机器学习获取了更多数据,但是也面临着3大难题:1) 新型领域不断涌现,如何准确地标识这些数据是一件费时费力的事情;2) 出于数据安全等原因,无法获取完成分类器学习所需要的足够多的数据;3) 数据的类别不平衡性日趋严重.

传统机器学习假设训练数据与测试数据服从相

同的分布,但很多实际情况下的应用并不满足此假设,如训练数据过期、训练模型过期、标记训练样本成本过高等问题的出现,大大影响了数据分析的效能;另外,在能够采集到新的不同分布下标记清楚的数据情况下,完全丢弃这些过期但相关的数据是很浪费的,因为新采集到的数据往往在数量上是不充足的. 迁移学习就是为了更好地利用这类数据而提出的新的研究方向^[3],它放松了对训练数据和测试数据同分布的要求,充分利用原有的、相关的实例或者已训练的模型,帮助对新实例的学习,使得对新实例的学习更快速有效^[4-5]. 例如:学会了识别梨子,可以帮助识别苹果;学会了识别自行车,可以帮助识别摩托车. 在迁移学习中,新实例数据所在的领域称为目标领域,而相关历史数据所在的领域称为源领域. 目前,

收稿日期: 2017-09-10; 修回日期: 2017-12-07.

基金项目: 国家自然科学基金项目(61572085, 61502058).

责任编辑: 侯忠生.

作者简介: 王洪元(1960—),男,教授,博士,从事模式识别与智能系统的研究;耿磊(1993—),男,硕士生,从事计算机视觉的研究.

[†]通讯作者. E-mail: hywang@cczu.edu.cn.

在国内外迁移学习领域已经出现了众多成果,较有代表性的有 Quanz 等^[6]提出的一种基于特征空间的大间隔直推式迁移学习方法(LMPROJ),该方法通过寻求一个特征变换使得源领域数据和目标领域数据之间的分布距离最小化来实现跨领域学习;许敏等^[7]提出了一种基于 SVM^[8]的迁移支持向量机 TL-SVM,该算法在训练目标领域模型时参考了源领域分类面参数这一知识,从而使用目标域少量已标签数据和大量相关领域的旧数据来为目标域构建一个高质量的分类模型。

上述两种迁移方法都没有考虑数据的类别不平衡问题,而类别不平衡问题在实际生活中很常见,如质量检测、网络攻击识别和医疗诊断等问题,其中的少数类样本的识别更具意义,往往是人们关心和研究的重点。目前解决此问题较为流行的方法是通过欠采样或过采样技术来调整正负类样本的比例,如:文献[9]采用过采样技术为两类样本设置不同权重来提高少数类样本的分类精度;文献[10]先定位样本点的分布,然后抽取不同类别的代表点实现类别间数据的平衡;文献[11]提出了过采样和欠采样的结合方法来处理不平衡数据的分类问题。但是这些算法会带来新的问题,即改变样本的原始分布结构,比如采取精确复制少数类样本的策略容易造成分类器的过拟合,而采取欠抽样多类样本的策略容易丢失部分样本的分布信息。另外,由于代价敏感学习^[12]关注错分样本的代价,其相关算法也常用来解决不平衡数据的分类问题。但这些方法均有一个前提,即训练样本的数量必须是充足的。面对训练数据不足且存在类别不平衡的场景下的分类问题,上述方法均不能取得理想的效果。

针对这一问题,本文以孪生支持向量机(Twin support vector machines, TwinSVM)^[13]为基础模型,融入知识嵌入的思想,提出基于迁移学习理论的知识嵌入孪生支持向量机(Knowledge embedded transfer twin support vector machine, KE-T-TwinSVM)。选择 TwinSVM 为研究模型的原因是其对不同类别样本分别构造分类超平面,从原理上具备了良好的解决不平衡问题的能力。基于 TwinSVM 的这一特点,所提 KE-T-TwinSVM 在处理类别不平衡数据时亦对两类训练样本分别构造分类超平面,使得每一个分类超平面与其中一类样本点尽可能近,同时远离另一类样本点;此外,KE-T-TwinSVM 基于迁移学习理论的知识嵌入思想能够利用源领域数据来协同训练,缓解了目标领域数据不足的问题,从而使得分类器的分类性能更为准确。

1 SVM与TwinSVM

考虑线性可分的分类问题。设训练集为 $T = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l)\} \in (\mathbf{X} \times \mathbf{Y})^l$, 其中 $y_i \in \mathbf{Y} = \{1, -1\}$, $i = 1, 2, \dots, l$ 。传统的二分类 SVM^[8] 的优化目标函数如下所示:

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i. \\ \text{s.t.} \quad & y(\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i \geq 0; \\ & \xi_i \geq 0, i = 1, 2, \dots, l. \end{aligned} \quad (1)$$

其中: C 为正则化参数, ξ_i 为松弛变量。

由式(1)可以看出,只有采集了足够量的已标记样本作为训练集,上述传统支持向量机的分类精度才会达到满意的效果,并且训练集和测试集数据应该是相同分布的。

在处理不平衡数据的分类问题时, SVM 往往倾向于追求多数类样本的高识别率来达到整体样本分类误差的最小化。在这种情况下,分类面不可避免地会向少数类样本发生偏移,少数类样本的识别就存在较高的误判率。

由 Jayadeva 等^[13]提出的 TwinSVM 与 SVM 最大的不同是, TwinSVM 将二分类问题转换为两个支持向量机的凸规划问题。另外,如果一类样本的数量远远大于另外一类样本的数量, TwinSVM 还可以分别对这两类的错分样本设置不同的惩罚系数。已有大量的研究工作表明, TwinSVM 处理少数类精度的不平衡问题具有优势。如 Ganesh 等^[14]和 Arjunan 等^[15]将其应用于生物医学中,在对于 EMG 表面肌电图细分之后产生的不平衡数据集之上使用 TwinSVM 进行动作判断,其预测精度明显高于其他一些流行算法。另外, TwinSVM 在语音识别方面的应用也取得了不错的效果^[16]。下面简单介绍 TwinSVM 的基本思想。

假定属于两类的数据分别用矩阵 $\mathbf{A}_+$ 和 $\mathbf{B}_-$ 表示, $\mathbf{A}_+ \in R^{m_1 \times d}$, $\mathbf{B}_- \in R^{m_2 \times d}$, 其中 m_1 和 m_2 分别为正类和负类样本个数。定义矩阵 $\mathbf{A}$ 、 $\mathbf{B}$ 和 $\mathbf{v}_1$ 、 $\mathbf{v}_2$ 、 $\mathbf{e}_1$ 、 $\mathbf{e}_2$ 四个向量^[13]如下:

$$\begin{cases} \mathbf{A} = [\mathbf{A}_+, \mathbf{e}_1], \mathbf{B} = [\mathbf{B}_-, \mathbf{e}_2], \\ \mathbf{v}_1 = [\mathbf{w}_1 \quad b_1]^T, \mathbf{v}_2 = [\mathbf{w}_2 \quad b_2]^T. \end{cases} \quad (2)$$

其中: $\mathbf{e}_1 = (1, \dots, 1)^T \in R^{m_1}$, $\mathbf{e}_2 = (1, \dots, 1)^T \in R^{m_2}$ 。则相比于 SVM, TwinSVM 包含两个非平行分类超平面:

$$\mathbf{w}_1^T \mathbf{x} + b_1 = 0, \mathbf{w}_2^T \mathbf{x} + b_2 = 0. \quad (3)$$

在求解时, TwinSVM 需要分别求解如下 TSVM1 和 TSVM2 两个凸规划的优化问题:

$$\begin{aligned} \text{TSVM1: } \min_{\mathbf{v}_1, \xi_1} & \frac{1}{2} \|\mathbf{A}\mathbf{v}_1\|^2 + c_1 \mathbf{e}_2^T \xi_1; \\ \text{s.t. } & -\mathbf{B}\mathbf{v}_1 + \xi_1 \geq \mathbf{e}_2, \xi_1 \geq 0. \end{aligned} \quad (4)$$

$$\begin{aligned} \text{TSVM2: } \min_{\mathbf{v}_2, \xi_2} & \frac{1}{2} \|\mathbf{B}\mathbf{v}_2\|^2 + c_2 \mathbf{e}_1^T \xi_2; \\ \text{s.t. } & \mathbf{A}\mathbf{v}_2 + \xi_2 \geq \mathbf{e}_1, \xi_2 \geq 0. \end{aligned} \quad (5)$$

其中: c_1 和 c_2 为非负参数, ξ_1 和 ξ_2 为对应维数的松弛向量. 对于待测样本 $\mathbf{x}$, 决策函数为 $\min_{r=1,2} |\mathbf{x}^T \mathbf{w}_r + b_r|$, 即由 $\mathbf{x}$ 与两个分类面距离的远近来判断 $\mathbf{x}$ 的类别.

TwinSVM 不具备迁移学习的能力, 在训练数据不足的情况下, 该方法的分类效果并不理想.

2 知识嵌入的孪生支持向量机

为了解决训练数据不足且存在类别不平衡数据的分类问题, 在第1节传统 TwinSVM 的基础上, 本文提出了基于迁移学习理论的知识嵌入孪生支持向量机 (Knowledge embedded transfer twin support vector machine, KE-T-TwinSVM). KE-T-TwinSVM 的工作原理和参数学习示意图如图1所示. 第1步, KE-T-TwinSVM 为了在目标领域数据不足或缺失的情况下进行分类器的训练, 必须得到来自相关领域 (即源领域) 知识的辅助, 为此首先使用 TwinSVM 对源领域数据进行训练, 将得到的分类面参数作为知识传递给新的分类器. 第2步, 使用目标领域数据和源领域知识, 构造新的适用于目标领域知识嵌入的迁移孪生支持向量机 KE-T-TwinSVM.

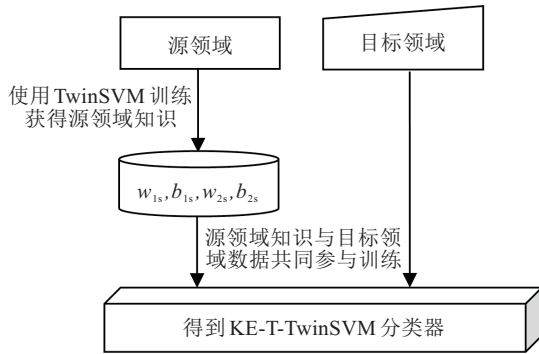


图1 KE-T-TwinSVM构造原理示意图

2.1 线性KE-T-TwinSVM

从图1所示的KE-T-TwinSVM构造原理图可以看出, 在目标领域训练数据不足以训练出可靠的分类模型的情况下, 借鉴从与目标领域相近的源领域训练得到知识是一种简单有效的方法. 这里令

$$\begin{aligned} \mathbf{A}_t &= (\mathbf{A}'_t, \mathbf{e}_1), \mathbf{B}_t = (\mathbf{B}'_t, \mathbf{e}_2), \\ \mathbf{v}_{1t} &= \begin{bmatrix} \mathbf{w}_{1t} \\ b_{1t} \end{bmatrix}, \mathbf{v}_{2t} = \begin{bmatrix} \mathbf{w}_{2t} \\ b_{2t} \end{bmatrix}, \\ \mathbf{A}_s &= (\mathbf{A}'_s, \mathbf{e}_1), \mathbf{B}_s = (\mathbf{B}'_s, \mathbf{e}_2), \end{aligned}$$

$$\mathbf{v}_{1s} = \begin{bmatrix} \mathbf{w}_{1s} \\ b_{1s} \end{bmatrix}, \mathbf{v}_{2s} = \begin{bmatrix} \mathbf{w}_{2s} \\ b_{2s} \end{bmatrix}. \quad (6)$$

其中: $\mathbf{A}'_t$ 和 $\mathbf{B}'_t$ 为目标领域的正类和负类数据, $(\mathbf{w}_{1t}, b_{1t})$ 和 $(\mathbf{w}_{2t}, b_{2t})$ 分别为目标领域正负两类数据的分类面参数, $\mathbf{A}'_s$ 和 $\mathbf{B}'_s$ 为源领域的正类和负类数据, $(\mathbf{w}_{1s}, b_{1s})$ 和 $(\mathbf{w}_{2s}, b_{2s})$ 分别为源领域正负两类数据的分类面参数. 针对正负分类面, 分别构造知识嵌入的迁移学习项 $|\mathbf{v}_{1t} - \mathbf{v}_{1s}|$ 和 $|\mathbf{v}_{1t} - \mathbf{v}_{1s}|$, 其含义是使得目标领域的正负类分类面与源领域中的正负类分类面尽可能接近. 基于知识嵌入的迁移学习项提出如下KE-T-TwinSVM目标函数表达式:

$$\begin{aligned} \min_{\mathbf{v}_1, \xi_1} & \frac{1}{2} \|\mathbf{A}_t \mathbf{v}_{1t}\|^2 + c_1 \mathbf{e}_2^T \xi_1 + D \mathbf{e}_1^T \eta; \\ \text{s.t. } & -\mathbf{B}_t \mathbf{v}_{1t} + \xi_1 \geq \mathbf{e}_2, \\ & |\mathbf{v}_{1t} - \mathbf{v}_{1s}| < \mathbf{e}_1^T \eta, \\ & \xi_1 \geq 0, \eta \geq 0. \end{aligned} \quad (7)$$

$$\begin{aligned} \min_{\mathbf{v}_{2t}, \xi_2} & \frac{1}{2} \|\mathbf{B}_t \mathbf{v}_{2t}\|^2 + c_2 \mathbf{e}_1^T \xi_2 + H \mathbf{e}_2^T \zeta; \\ \text{s.t. } & -\mathbf{A}_{1t} \mathbf{v}_{2t} + \xi_2 \geq \mathbf{e}_2, \\ & |\mathbf{v}_{2t} - \mathbf{v}_{2s}| < \mathbf{e}_2^T \delta, \\ & \xi_2 \geq 0, \zeta \geq 0. \end{aligned} \quad (8)$$

其中: D 和 H 为非负参数, 用以调节迁移项在目标函数中所占比重; η 和 ζ 为迁移项的松弛向量, $\eta \in \mathbf{R}^{m_1+1}$, $\zeta \in \mathbf{R}^{m_2+1}$. 式(7)对应的拉格朗日优化问题如下:

$$\begin{aligned} L = & \frac{1}{2} \|\mathbf{A}_t \mathbf{v}_{1t}\|^2 + c_1 \mathbf{e}_2^T \xi_1 + D \mathbf{e}_1^T \eta - \\ & \alpha^T (-\mathbf{B}_t \mathbf{v}_{1t} + \xi_1 - \mathbf{e}_2) - \beta^T (\mathbf{e}_1^T \eta - \mathbf{v}_{1t} + \mathbf{v}_{1s}) - \\ & \gamma^T (\mathbf{v}_{1t} - \mathbf{v}_{1s} + \mathbf{e}_1^T \eta) - \lambda^T \xi_1 - \sigma^T \eta. \end{aligned} \quad (9)$$

根据Karush-Kuhn-Tucker(KKT)^[17]条件

$$\begin{cases} \frac{\partial L}{\partial \mathbf{v}_{1t}} = \mathbf{A}_t^T \mathbf{A}_t \mathbf{v}_{1t} + \mathbf{B}_t^T \alpha + \beta - \gamma = 0, \\ \frac{\partial L}{\partial \xi_1} = c_1 \mathbf{e}_2 - \alpha - \lambda = 0, \\ \frac{\partial L}{\partial \eta} = D \mathbf{e}_1 - \beta - \gamma - \sigma = 0. \end{cases} \quad (10)$$

由式(10)可得

$$\mathbf{v}_{1t} = -(\mathbf{A}_t^T \mathbf{A}_t)^{-1} (\mathbf{B}_t^T \alpha + \beta - \gamma). \quad (11)$$

将式(11)代入(9), 得

$$\begin{aligned} L = & -\frac{1}{2} (\mathbf{B}_t^T \alpha + \beta - \gamma)^T (\mathbf{A}_t^T \mathbf{A}_t)^{-1} (\mathbf{B}_t^T \alpha + \\ & \beta - \gamma) + \alpha^T \mathbf{e}_2 - \beta^T \mathbf{v}_{1s} + \gamma^T \mathbf{v}_{1s}. \end{aligned} \quad (12)$$

式(12)的对偶式表示如下:

$$\begin{aligned} \min_{\Gamma} & -\frac{1}{2}\Gamma^T(\mathbf{A}_t^T\mathbf{A}_t)^{-1}\Gamma + \Gamma^T\Omega; \\ \text{s.t. } & \Gamma = \mathbf{B}_t^T\alpha + \beta - \gamma, \\ & \Omega = \begin{bmatrix} \mathbf{B}_t^{-1}\mathbf{e}_2 \\ -\mathbf{v}_{1s} \\ \mathbf{v}_{1s} \end{bmatrix}. \end{aligned} \quad (13)$$

使用同样的办法可以得到式(8)的对偶式如下:

$$\begin{aligned} \min_{\tilde{\Gamma}} & -\frac{1}{2}\tilde{\Gamma}^T(\mathbf{B}_t^T\mathbf{B}_t)^{-1}\tilde{\Gamma} + \tilde{\Gamma}^T\tilde{\Omega}; \\ \text{s.t. } & \tilde{\Gamma} = \mathbf{A}_t^T\tilde{\alpha} + \tilde{\beta} - \tilde{\gamma}, \\ & \tilde{\Omega} = \begin{bmatrix} \mathbf{A}_t^{-1}\mathbf{e}_2 \\ -\mathbf{v}_{2s} \\ \mathbf{v}_{2s} \end{bmatrix}. \end{aligned} \quad (14)$$

$$\mathbf{v}_{2t} = -(\mathbf{B}_t^T\mathbf{B}_t)^{-1}(\mathbf{A}_t^T\alpha + \beta - \gamma). \quad (15)$$

对应新样本 x , 计算下式:

$$|(\mathbf{x}, 1)^T\mathbf{v}_{1t}| \leq |(\mathbf{x}, 1)^T\mathbf{v}_{2t}|. \quad (16)$$

若式(16)为真, 则 x 属于正类, 否则属于负类.

根据上述描述, 可得线性 KE-T-TwinSVM 的算法流程如下.

算法1 算法1 线性 KE-T-TwinSVM 算法流程.

输入: 源领域数据 $\mathbf{A}'_s$ 和 $\mathbf{B}'_s$, 目标领域数据 $\mathbf{A}'_t$ 和 $\mathbf{B}'_t$;

输出: 目标领域的的数据类标签.

Step 1: 源领域知识获取阶段.

Step 1.1: 使用式(6)得到 $\mathbf{A}_s$ 和 $\mathbf{B}_s$, 设置参数 c_1 、 c_2 、 D 和 H ;

Step 1.2: 利用 TwinSVM 得到源领域数据的分类模型;

Step 1.3: 求解得到源领域分类面参数 $\mathbf{v}_{1s}$ 和 $\mathbf{v}_{2s}$.

Step 2: 目标领域迁移学习阶段.

Step 2.1: 使用式(6)得到 $\mathbf{A}_t$ 和 $\mathbf{B}_t$, 设置参数 c_1 、 c_2 、 D 和 H ;

Step 2.2: 求解式(11)、(13)~(15);

Step 2.3: 根据式(16)输出待测样本 x 的类别.

2.2 核化 KE-T-TwinSVM

本节讨论核化的 KE-T-TwinSVM 算法. 非线性情况下 KE-T-TwinSVM 的分类超平面变为如下形式:

$$\begin{cases} K\{\mathbf{x}_1^T, \mathbf{C}_1^T\}\mathbf{w}_1 + b_1 = 0, \\ K\{\mathbf{x}_2^T, \mathbf{C}_2^T\}\mathbf{w}_2 + b_2 = 0. \end{cases} \quad (17)$$

其中: K 是核函数, 且 $K\{\mathbf{x}_i, \mathbf{x}_j\} = (\Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j))$, $\Phi(\cdot)$ 是低维特征空间到高维特征空间的非线性映射; $\mathbf{C}_t$ 和 $\mathbf{C}_s$ 分别是目标领域和源领域的训练数据

集, $\mathbf{C}_t = (\mathbf{A}'_t, \mathbf{B}'_t)^T$, $\mathbf{C}_s = (\mathbf{A}'_s, \mathbf{B}'_s)^T$.

核化 KE-T-TwinSVM 的拉格朗日优化问题可以表示为

$$\begin{aligned} \min_{\mathbf{v}_{1t}, \xi_1} & \frac{1}{2}\|K\{\mathbf{A}'_t, \mathbf{C}_t^T\}\mathbf{w}_{1t} + b_{1t}\|^2 + c_1\mathbf{e}_2^T\xi_1 + D\mathbf{e}_1^T\eta; \\ \text{s.t. } & -K\{\mathbf{B}'_t, \mathbf{C}_t^T\}\mathbf{w}_{1t} - \mathbf{e}_2b_{1t} + \xi_1 \geq \mathbf{e}_2, \\ & |K\{\mathbf{A}'_t, \mathbf{C}_t^T\}\mathbf{w}_{1t} + b_{1t} - \mathbf{v}_{1s}| < \mathbf{e}_1^T\eta, \\ & \xi_1 \geq 0, \eta \geq 0. \end{aligned} \quad (18)$$

$$\begin{aligned} \min_{\mathbf{v}_{1t}, \xi_1} & \frac{1}{2}\|K\{\mathbf{B}'_t, \mathbf{C}_t^T\}\mathbf{w}_{2t} + b_{2t}\|^2 + c_2\mathbf{e}_1^T\xi_2 + H\mathbf{e}_2^T\zeta; \\ \text{s.t. } & -K\{\mathbf{A}'_t, \mathbf{C}_t^T\}\mathbf{w}_{2t} - b_{2t} + \xi_2 \geq \mathbf{e}_2, \\ & |K\{\mathbf{B}'_t, \mathbf{C}_t^T\}\mathbf{w}_{2t} + b_{2t} - \mathbf{v}_{2s}| < \mathbf{e}_2^T\delta, \\ & \xi_2 \geq 0, \zeta \geq 0. \end{aligned} \quad (19)$$

式(18)的拉格朗日优化表达式为

$$\begin{aligned} L = & \frac{1}{2}\|K\{\mathbf{A}'_t, \mathbf{C}_t^T\}\mathbf{w}_{1t} + b_{1t}\|^2 + c_1\mathbf{e}_2^T\xi_1 + D\mathbf{e}_1^T\eta - \\ & \alpha^T(-K\{\mathbf{B}'_t, \mathbf{C}_t^T\}\mathbf{w}_{1t} + b_{1t} + \xi_1 - \mathbf{e}_2) - \\ & \beta^T(\mathbf{e}_1^T\eta - \mathbf{w}_{1t} - b_{1t} + \mathbf{w}_{1s} + b_{1s}) - \\ & \gamma^T(\mathbf{w}_{1t} + b_{1t} - \mathbf{w}_{1s} - b_{1s} + \mathbf{e}_1^T\eta) - \lambda^T\xi_1 - \sigma^T\eta, \end{aligned} \quad (20)$$

其中 $\alpha, \beta, \gamma, \lambda, \sigma$ 为拉格朗日乘子. 同样, 根据如下 Karush-Kuhn-Tucker(KKT)^[17] 条件:

$$\begin{cases} \frac{\partial L}{\partial \mathbf{w}_{1t}} = K\{\mathbf{A}'_t, \mathbf{C}_t^T\}^T(K\{\mathbf{A}'_t, \mathbf{C}_t^T\}\mathbf{w}_{1t} + \\ \quad \mathbf{e}_1b_{1t}) + K\{\mathbf{B}'_t, \mathbf{C}_t^T\}^T\alpha + \beta - \gamma = 0, \\ \frac{\partial L}{\partial b_{1t}} = \mathbf{e}_1^T(K\{\mathbf{A}'_t, \mathbf{C}_t^T\}\mathbf{w}_{1t} + \mathbf{e}_1b_{1t}) + \\ \quad K\{\mathbf{B}'_t, \mathbf{C}_t^T\}^T\alpha + \beta - \gamma = 0, \\ \frac{\partial L}{\partial \xi_1} = c_1\mathbf{e}_2 - \alpha - \lambda = 0, \\ \frac{\partial L}{\partial \eta} = D\mathbf{e}_1 - \beta - \gamma - \sigma = 0. \end{cases} \quad (21)$$

令

$$\mathbf{E}_t = (K\{\mathbf{A}'_t, \mathbf{C}_t^T\}, \mathbf{e}_1), \mathbf{F}_t = (K\{\mathbf{B}'_t, \mathbf{C}_t^T\}, \mathbf{e}_2), \quad (22)$$

$$\mathbf{E}_s = (K\{\mathbf{A}'_s, \mathbf{C}_s^T\}, \mathbf{e}_1), \mathbf{F}_s = (K\{\mathbf{B}'_s, \mathbf{C}_s^T\}, \mathbf{e}_2), \quad (23)$$

$$\mathbf{v}_{1t} = (\mathbf{E}_t^T\mathbf{E}_t)^{-1}(\mathbf{F}_t^T\alpha + \beta - \gamma). \quad (24)$$

得到式(18)的对偶式如下:

$$\begin{aligned} \min_{\Lambda} & -\frac{1}{2}\Lambda^T(\mathbf{E}_t^T\mathbf{E}_t)^{-1}\Lambda + \Lambda^T\Psi; \\ \text{s.t. } & \Lambda = \mathbf{F}_t^T\alpha + \beta - \gamma, \end{aligned}$$

$$\Psi = \begin{bmatrix} F^{-1}e_2 \\ -v_{1s} \\ v_{1s} \end{bmatrix}. \quad (25)$$

使用同样的办法可以得到式(19)的对偶式如下:

$$\begin{aligned} \min_{\tilde{\Lambda}} & -\frac{1}{2} \tilde{\Lambda}^T (F_t^T F_t)^{-1} \tilde{\Lambda} + \tilde{\Lambda}^T \tilde{\Psi}; \\ \text{s.t. } & \tilde{\Lambda} = E_t^T \tilde{\alpha} + \tilde{\beta} - \tilde{\gamma}, \\ & \tilde{\Psi} = \begin{bmatrix} E_t^{-1}e_2 \\ -v_{2s} \\ v_{2s} \end{bmatrix}. \end{aligned} \quad (26)$$

$$v_{1t} = (F_t^T F_t)^{-1} (E_t^T \alpha + \beta - \gamma). \quad (27)$$

对应新样本 x , 计算下式:

$$|K\{x_1^T, C_1^T\}w_1 + b_1| \leq |K\{x_1^T, C_1^T\}w_2 + b_2|. \quad (28)$$

若式(28)为真, 则 x 属于正类, 否则属于负类.

根据上面的描述, 可得线性 KE-T-TwinSVM 的算法流程如下

算法2 核化 KE-T-TwinSVM 算法流程.

输入: 源领域样本 A'_s 和 B'_s , 目标领域样本 A'_t 和 B'_t ;

输出: 目标领域的数据类标签.

Step 1: 源领域知识获取阶段.

Step 1.1: 使用式(6)得到 A_s 和 B_s , 设置参数 c_1 、 c_2 、 D 和 H ;

Step 1.2: 利用 TwinSVM 得到源领域数据的分类模型;

Step 1.3: 求解得到源领域分类面参数 v_{1s} 和 v_{2s} .

Step 2: 目标领域迁移学习阶段.

Step 2.1: 使用式(6)得到 A_t 和 B_t , 设置参数 c_1 、 c_2 、 D 和 H ;

Step 2.2: 求解式(24)~(27);

Step 2.3: 根据式(28)输出待测样本 x 的类别.

3 实验研究

3.1 实验设置

为了验证本文方法的性能, 本节将在各种类型的迁移数据集上对 KE-T-TwinSVM 算法进行分析与验证. 有关实验数据的详细描述分别于 3.2 节、3.3 节和 3.4 节给出. 与 KE-T-TwinSVM 进行比较的算法包括 SVM 和适用于类别不平衡数据的 4 种典型的分类算法: TwinSVM、CS-SVM^[18]、BalanceCascade^[19] 和 Adac2^[20]. 为了考察所提 KE-T-TwinSVM 算法的迁移学习性能, 实验选取了 TL-SVM^[7] 和 LMPROJ^[6] 进行对比. 在本文实验部分, 各算法最优参数的设置采取

5 重交叉验证法选取, 参数的详细设置如表 1 所示. 所有算法在 Matlab2010b 环境下实现, SVM 算法由 LIBSVM 软件实现.

表 1 实验中各算法参数的定义和设置

算法	参数描述
SVM	C 在 $\{10^{-3}, 10^{-2}, \dots, 10^3\}$ 中取值, 高斯核函数核宽 σ 在 $\{10^{-2}, 10^{-1}, \dots, 10^2\}$.
CS-SVM	核参数取值为 $\{10^{-3}, 10^{-2}, \dots, 10^3\}$, 多数类的正则化参数取值为 $\{10^{-3}, 10^{-2}, \dots, 10^3\}$, 少数类的正则化参数取值为多数类正则化参数/正负类的比例
BalanceCascade	子集和弱分类器的个数分别是 4 和 10
Adac2	多数类的代价参数为 $\{0.1, 0.2, \dots, 0.9\}$, 少数类的代价参数为 1, 弱分类器的个数为 10
TL-SVM	C_t 在 $\{10^{-3}, 10^{-2}, \dots, 10^3\}$ 中取值, μ 在 $\{10^{-5}, 10^{-4}, \dots, 10^5\}$ 中取值
LMPROJ	C 在 $\{10^{-3}, 10^{-2}, \dots, 10^3\}$ 中取值
KE-T-TwinSVM	c_1 、 c_2 、 D 和 H 在 $\{10^{-3}, 10^{-2}, \dots, 10^3\}$ 中取值

为了更好地评价本文方法与其他方法在各种数据集上的性能, 特别是为了更好地呈现出不同程度的非平衡性对算法性能产生的影响, 本文采用基于目标域数据的接受者操作特性曲线下面积 (Area under curve, AUC)^[21] 评价准则来评价算法的分类性能. AUC 值在 0.5~1 之间, 在比较多个分类器时, AUC 大的算法则预示其具有比较好的分类性能.

3.2 Newsgroups 文本数据集

20 Newsgroups 数据集包含 7 个大类 20 个小类共 18 774 个文本数据, 从中抽取 4 个大类共 12 个小类来设计学习任务, 其中每 2 个大类可分别被选为正类和负类, 各子类分属不同的领域. 抽取的数据集信息见表 2, 各学习任务的领域说明及实验结果见表 3. 其中, 邻域表示为 [子类号, 占子类样本总量百分比].

表 2 抽取的 20 Newsgroups 部分大类及子类信息

大类	子类编号	子类	样本数
comp	1	comp.sys.ibm.pc.hardware	979
	2	comp.windows.x	982
	3	comp.sys.mac.hardware	958
rec	4	sport.hockey	997
	5	motorcycles	993
	6	autos	987
sci	7	crypt	989
	8	med	987
	9	Space	985
talk	10	politics.guns	909
	11	politics.mideast	940
	12	politics.misc	774

表 3 各种算法在 20Newsgroups 数据集上的 AUC 值

学习任务	源领域样本组成	目标领域样本组成	AUC							
			SVM	TwinSVM	CS-SVM	Balance	Cascade	Adac2	TL-SVM	LMPROL
1	[1, 100 %], [5, 100 %]	[3, 100 %], [6, 100 %]	0.681 2	0.685 5	0.685 5	0.672 4	0.675 8	0.821 7	0.833 4	0.849 1
	[1, 30 %], [5, 80 %]	[3, 30 %], [6, 80 %]	0.609 6	0.630 1	0.633 8	0.649 8	0.638 5	0.782 2	0.783 3	0.826 3
2	[1, 100 %], [7, 100 %]	[3, 100 %], [8, 100 %]	0.682 1	0.709 8	0.696 7	0.683 7	0.683 7	0.751 9	0.766 8	0.783 2
	[1, 30 %], [7, 80 %]	[3, 30 %], [8, 80 %]	0.601 5	0.633 5	0.636 8	0.645 5	0.617 1	0.711 1	0.700 8	0.765 3
3	[1, 100 %], [11, 100 %]	[3, 100 %], [12, 100 %]	0.740 1	0.746 3	0.759 9	0.746 8	0.772 1	0.857 9	0.861 2	0.902 6
	[1, 30 %], [11, 8 %]	[3, 30 %], [12, 80 %]	0.668 9	0.689 9	0.693 2	0.701 5	0.709 1	0.792 1	0.802 0	0.885 1
4	[5, 40 %], [7, 80 %]	[6, 100 %], [8, 100 %]	0.671 9	0.672 2	0.678 9	0.651 7	0.653 1	0.675 5	0.687 7	0.706 9
	[5, 30 %], [7, 80 %]	[6, 30 %], [8, 80 %]	0.611 2	0.626 9	0.632 2	0.624 1	0.625 8	0.638 8	0.635 2	0.678 8
5	[5, 100 %], [11, 80 %]	[6, 100 %], [12, 100 %]	0.617 8	0.620 8	0.612 6	0.620 1	0.621 3	0.719 6	0.745 2	0.739 8
	[5, 30 %], [11, 80 %]	[6, 30 %], [12, 80 %]	0.569 0	0.587 5	0.586 6	0.582 3	0.585 6	0.682 2	0.679 9	0.717 5
6	[7, 100 %], [11, 100 %]	[8, 100 %], [12, 100 %]	0.601 3	0.604 1	0.605 9	0.604 4	0.607 7	0.628 9	0.638 1	0.651 2
	[7, 30 %], [11, 80 %]	[8, 30 %], [12, 80 %]	0.530 9	0.591 0	0.589 8	0.578 8	0.588 4	0.628 1	0.616 1	0.648 1

由表 3 所示的实验结果可以看出:

1) 作为传统的机器学习方法, SVM 在迁移数据集上分类效果不佳, 当数据集类别不平衡时, 其分类效果也是最差的, 这说明传统分类方法已经不适合新领域的模式识别问题.

2) TwinSVM、CS-SVM、BalanceCascade 和 Adac2 在类别不平衡数据集上具有一定的优势, 这是因为 TwinSVM 对正负类分别构造不同的分类面, CS-SVM 为不同类别的错分样本设置不同的惩罚系数, Adc2 算法为不同类别的样本设置不同的代价参数, 而 BalanceCascade 通过欠采样技术来调整两类样本比例. 然而, 也应注意到, 这 4 个算法在目标领域和源领域数据分布不同时分类性能较差, 说明目标域数据不充足的分类场景下, 非迁移学习算法已经不再适用.

3) TL-SVM、LMPROJ 和 KE-T-TwinSVM 均是迁移学习方法, 在 20Newsgroups 上的表现普遍好于非迁移方法. 但是在数据集出现不平衡情况下, TL-SVM 和 LMPROJ 的 AUC 值下降很快, 原因在于 TL-SVM 和 LMPROJ 算法没有考虑到样本容量的差异性和样本分布的不平衡性, 这两个算法构建的分类面易向少数类样本发生偏移, 虽然在多数类样本的识别上准确率较高, 但对少数类样本的识别率较低. 本文方法在 20Newsgroups 上表现出了良好的适应性能, 均优于其他两种方法, 说明本文所提 KE-T-TwinSVM 算法能够较好地适用于训练数据不足且存在类别不平衡场景下的分类问题.

3.3 UCI 数据集

本部分选取 8 个 UCI 数据集来考察所提 KE-T-TwinSVM 的性能, 数据集的基本信息如表 4 所示. 在构造学习任务时, 分别从数据集中取出互不相同的不同比例的数据构成源领域和目标领域, 为了使两者的分布不同, 在目标领域数据中加入数据均值 3 % 大小

的随机噪声. 各个数据集上不同算法的分类性能如表 5 所示. 其中, 邻域表示为 [正/负类, 占训练集的百分比].

表 4 UCI 数据集基本信息

标号	数据集	规模	正类数	负类数	维数
1	wisconsin	683	239	444	9
2	yeast1	1 484	429	1 055	8
3	pima	768	267	501	8
4	Segment0	2 308	329	1 979	36
5	Page-blocks0	5 472	560	4 912	10
6	vehicle	846	200	646	18
7	shuttle	20 000	100 00	10 000	9
8	Vowel0	988	90	898	13

根据各算法在 UCI 数据集上的实验结果可得如下结论:

1) 传统的 SVM 在分类过程中没有考虑不同样本的类别不平衡性, 也无法将源领域的知识有效地迁移至目标领域来帮助学习, 因此在 8 种真实数据集上的分类性能均低于其余几种方法.

2) TwinSVM、CS-SVM、BalanceCascade 和 Adac2 在非不平衡情况下分类性能与 SVM 相当, 在数据集出现类别不平衡的情况下取得了比 SVM 更好的性能, 但面对迁移分类问题时没有将源领域的知识有效地迁移至目标领域来帮助学习, 因此分类效果不理想.

3) 本文方法 KE-T-TwinSVM 对两类训练样本分别构造不同的分类超平面, 使得每一个分类超平面与其中一类样本点尽可能近, 而远离另一类样本点, 在处理非不平衡数据分类问题时分类性能优于 TL-SVM 和 LMPROJ; 同时, KE-T-TwinSVM 能够利用源领域数据来协同目标领域数据训练, 缓解了目标领域数据不足的问题, 在处理不平衡数据分类问题时获得了较好的分类效果, 几乎在所有数据集上都处于领先地位.

表 5 各种算法在 8 个 UCI 数据集上的 AUC 值

数据集	源领域样本组成	目标领域样本组成	AUC							
			SVM	TwinSVM	CS-SVM	Balance	Cascade	Adac2	TL-SVM	LMPROL
1	[+, 80 %], [−, 80 %]	[+, 20 %], [−, 20 %]	0.842 7	0.868 3	0.863 0	0.842 3	0.836 4	0.929 4	0.933 6	0.944 5
	[+, 20 %], [−, 80 %]	[+, 5 %], [−, 20 %]	0.708 1	0.797 9	0.799 0	0.781 3	0.789 1	0.829 3	0.836 2	0.872 6
2	[+, 80 %], [−, 80 %]	[+, 20 %], [−, 20 %]	0.631 3	0.653 1	0.654 1	0.652 9	0.647 9	0.704 3	0.707 4	0.713 4
	[+, 20 %], [−, 80 %]	[+, 5 %], [−, 20 %]	0.587 2	0.602 1	0.606 6	0.608 2	0.607 3	0.638 5	0.645 3	0.703 2
3	[+, 80 %], [−, 80 %]	[+, 20 %], [−, 20 %]	0.612 2	0.642 6	0.635 3	0.641 0	0.646 2	0.678 8	0.686 0	0.706 3
	[+,20 %], [−, 80 %]	[+, 50 %], [−, 20 %]	0.556 6	0.611 1	0.607 6	0.592 9	0.603 7	0.622 7	0.629 9	0.663 4
4	[+, 80 %], [−, 80 %]	[+, 20 %], [−, 20 %]	0.886 6	0.910 9	0.907 1	0.906 5	0.913 7	0.937 2	0.947 3	0.952 1
	[+, 10 %], [−, 80 %]	[+, 2.5 %], [−, 20 %]	0.769 8	0.839 3	0.830 2	0.821 3	0.813 2	0.881 7	0.885 7	0.912 1
5	[+, 80 %], [−, 80 %]	[+, 20 %], [−, 20 %]	0.804 0	0.857 1	0.844 5	0.857 2	0.858 3	0.870 1	0.872 9	0.908 3
	[+, 10 %], [−, 80 %]	[+, 2.5 %], [−, 20 %]	0.700 9	0.754 1	0.753 4	0.750 2	0.762 9	0.830 1	0.837 4	0.865 4
6	[+, 80 %], [−, 80 %]	[+, 20 %], [−, 20 %]	0.803 2	0.839 8	0.856 3	0.862 8	0.850 1	0.887 3	0.907 3	0.902 1
	[+, 20 %], [−, 80 %]	[+, 5 %], [−, 20 %]	0.722 3	0.778 8	0.768 7	0.765 3	0.758 3	0.835 3	0.830 2	0.877 5
7	[+, 80 %], [−, 80 %]	[+, 20 %], [−, 20 %]	0.915 3	0.920 9	0.927 9	0.923 1	0.905 1	0.966 2	0.976 5	0.972 4
	[+, 10 %], [−, 80 %]	[+, 2.5 %], [−, 20 %]	0.703 9	0.813 3	0.815 4	0.810 1	0.806 2	0.894 0	0.894 6	0.952 1
8	[+, 80 %], [−, 80 %]	[+, 20 %], [−, 20 %]	0.872 8	0.890 1	0.886 7	0.882 2	0.886 4	0.912 0	0.913 7	0.937 7
	[+, 10 %], [−, 80 %]	[+, 2.5 %], [−, 20 %]	0.789 2	0.812 2	0.817 6	0.815 3	0.819 9	0.866 6	0.862 4	0.906 5

3.4 统计分析

为了使得对比更具有统计意义,本文采用了统计学方法Friedman检验和Host事后检验^[22]. Friedman检验是利用秩实现对多个总体分布是否存在显著差异的非参数检验方法,首先按照各个算法在不同迁移数据集上的平均精度排序;其次在各个数据集上计算每个算法精度与平均精度之间的差异;再次按降序对所有差异进行排序;最后可得每个算法的排名. 以这种方式可得Friedman图形化检验结果,其中实现最低平均值的算法,即升序排名第1的算法是最好的. 实验中将表3和表5中的分类AUC值分成2组,一组是原始迁移数据集对应的分类AUC值,另一组是构造的不平衡迁移数据集对应的分类AUC值. 这2组AUC值对应的Friedman秩如图2和图3所示. 从图2和图3可以看出,本文提出的KE-T-TwinSVM的分类性能无论在原始迁移数据集还是在不平衡迁移数据集上均优于实验中的7种对比算法.

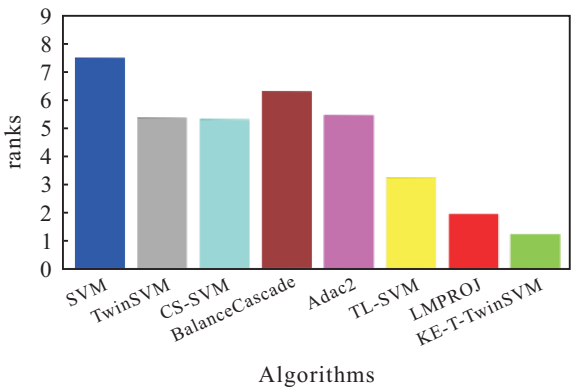


图 2 在原始迁移数据集上各个算法的Friedman秩

原始迁移数据集和不平衡迁移数据集对应的分类AUC值对应的Host检验结果如表6和表7所示. Host事后检验可以知道各算法性能之间的差异是

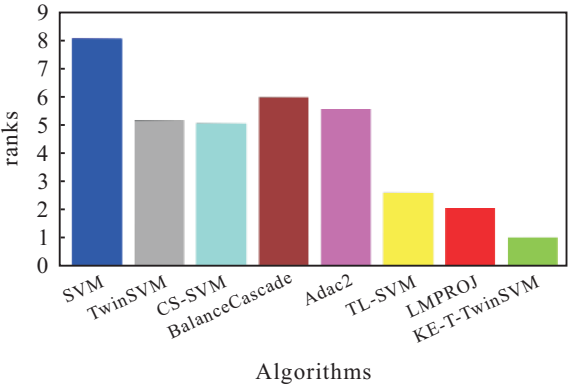


图 3 在不平衡迁移数据集上各个算法的Friedman秩

表 6 在原始迁移数据集上各个算法的Host检验结果

<i>i</i>	分类器	<i>z</i>	<i>p</i>	Holm = α/i	测试结果
7	SVM	6.71	0	0.007 14	拒绝
6	BalanceCascade	5.43	0	0.008 33	拒绝
5	Adac2	4.51	6.0e-6	0.01	拒绝
4	TwinSVM	4.43	9.0e-6	0.012 5	拒绝
3	CS-SVM	4.35	1.3e-5	0.016 7	拒绝
2	TL-SVM	2.16	0.024 9	0.025	拒绝
1	LMPROJ	0.77	0.440 4	0.05	不拒绝

表 7 在不平衡迁移数据集上各个算法的Host检验结果

<i>i</i>	分类器	<i>z</i>	<i>p</i>	Holm = α/i	测试结果
7	SVM	7.56	0	0.007 143	拒绝
6	BalanceCascade	5.32	0	0.008 333	拒绝
5	Adac2	5.32	0	0.01	拒绝
4	TwinSVM	4.47	8.0e-6	0.012 5	拒绝
3	CS-SVM	4.32	1.6e-6	0.016 7	拒绝
2	TL-SVM	1.69	0.019 6	0.025	拒绝
1	LMPROJ	1.54	0.046 7	0.05	拒绝

否是显著的. 若根据Host事后检验所得APV(Adjusted *p*-value)值小于显著性水平0.05,则说明本文的方法有着显著的优势,反之则无法说明有显著优势. 从表6可以看出,在原始迁移数据集上,即数据不平衡情况不明显的数据集上,相比于SVM、TwinSVM、

BalanceCascade、Adac2、CS-SVM和TL-SVM,本文提出的KE-T-TwinSVM的分类性能具有显著优势.从表7也可以看出,本文算法在不平衡数据集上与7种对比算法相比均具有显著优势.这一统计结果再次验证了所提方法的有效性,其不仅适用于平衡数据集的迁移场景,也适用于不平衡数据集的迁移场景.

4 结论

本文通过将迁移学习理念融入孪生支持向量机,提出了一种新的知识嵌入的迁移孪生支持向量机,即KE-T-TwinSVM分类器.KE-T-TwinSVM能够将从源领域中学到的知识应用到目标领域的分类器训练中,保证了在目标领域缺失或不足的情况下,分类器依然可以取得较好的效果.KE-T-TwinSVM也能够针对不同类别的样本分别构建不同的分类面,通过调节不同的参数,较好地解决类别不平衡数据的分类问题.多个真实数据集的仿真实验表明了本文算法的分类性能较之传统方法有相当的优势.应当指出,本文算法仍有不足,例如对KE-T-TwinSVM能否有效解决大样本等问题没有进行深入探讨.当数据容量极大时,从二次规划求解角度而言,KE-T-TwinSVM仍面临进一步提高实用性的挑战,这将作为近期的研究重点.

参考文献(References)

- [1] Zafeiriou S F, Tefas A, Pitas I. Minimum class variance support vector machines[J]. IEEE Trans on Image Processing, 2007, 16(10): 2551-2564.
- [2] 张喆, 韩德强, 杨艺. 基于证据马尔可夫随机场模型的图像分割[J]. 控制与决策, 2017, 32(9): 1607-1613. (Zhang Z, Han D Q, Yang Y. Image segmentation based on evidential Markov random field model[J]. Control and Decision, 2017, 32(9): 1607-1613.)
- [3] Pan S J, Yang Q. A survey on transfer learning[J]. IEEE Trans on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.
- [4] Wu Q Y, Wu H R, Zhou X M, et al. Online transfer learning with multiple homogeneous or heterogeneous sources[J]. IEEE Trans on Knowledge and Data Engineering, 2017, 29(7): 1494-1507.
- [5] Pan S J, Tsang I W, Kwok J T, et al. Domain adaptation via transfer component analysis[J]. IEEE Trans on Neural Networks, 2011, 22(2): 199-210.
- [6] Quanz B, Huan J. Large margin transductive transfer learning[C]. Proc of the 18th ACM Conf on Information and Knowledge Management. New York: ACM press, 2009: 1327-1336.
- [7] 许敏, 王士同, 顾鑫. TL-SVM: 一种迁移学习新算法[J]. 控制与决策, 2014, 29(1): 141-146. (Xu M, Wang S T, Gu X. TL-SVM: A transfer learning algorithm [J]. Control and Decision, 2014, 29(1): 141-146.)
- [8] Vapnik V. Statistical learning theory[M]. New York: John Wiley and Sons, 1998: 339-350.
- [9] Ramentol E, Caballero Y, Bello R, et al. SMOTE-RSB: A hybrid preprocessing approach based on oversampling and undersampling for high imbalanced data-Sets using SMOTE and rough sets theory[J]. Knowledge and Information Systems, 2012, 33(2): 245-265.
- [10] López V, Fernández A, Jesus M, et al. A hierarchical genetic fuzzy system based on genetic programming for addressing classification with highly imbalanced and borderline datasets[J]. Knowledge Based Systems, 2013(38): 85-104.
- [11] Chawla N V, Bowyer K W, Hall L O. SMOTE: Synthetic minority over-sampling technique[J]. J of Artificial Intelligence Research, 2002, 16(1): 321-357.
- [12] Tang Y, Zhang Y Q, Chawla N V. SVMs modeling for highly imbalanced classification[J]. IEEE Trans on Systems, Man, and Cybernetics, Part B: Cybernetics, 2009, 39(1): 280-288.
- [13] Jayadeva K R, Suresh C. Twin support vector machines for pattern classification [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2007, 29(5): 905-910.
- [14] Ganesh R N, Dinesh K K, Jayadeva. Twin SVM for gesture classification using the surface electromyogram[J]. IEEE Trans on Information Technology in Biomedicine, 2010, 14(2): 301-308.
- [15] Arjunan S P, Kumar D K, Naik G R. A machine learning based method for classification of fractal features of forearms EMG using Twin Support Vector Machines[C]. 2010 Annual Int Conf on Engineering in Medicine and Biology Society(EMBC). Buenos Aires: IEEE, 2010: 4821-4824.
- [16] Liu M, Xie Y, Yao Z, et al. A new hybrid GMM/SVM for speaker verification[C]. The 18th Int Conf on Pattern Recognition. Hong Kong: IEEE, 2006: 314-317.
- [17] Scholkopf B, Herbrich R, Smola A J. A generalized representer theorem[C]. Proc of Conf on Learning Theory. Amsterdam: Springer, 2001: 416-426.
- [18] Shirazi H M, Vasconcelos N, Iranmehr A. Cost-sensitive support vector machines[J]. J of Machine Learning Research, 2012, 6(5): 387-389.
- [19] Liu X Y, Wu J, Zhou Z H. Exploratory undersampling for class imbalance learning[J]. IEEE Trans on Systems, Man, and Cybernetics, Part B: Cybernetics, 2009, 39(2): 539-550.
- [20] Sun Y, Kamel M, Wong A, et al. Cost-sensitive boosting for classification of imbalanced data[J]. Pattern Recognition, 2007, 40(12): 3358-3378.
- [21] Galar M, Fernández A, Barrenechea E, et al. Ordering-based pruning for improving the performance of ensembles of classifiers in the framework of imbalanced datasets[J]. Information Sciences, 2016, 354(2): 178-196.
- [22] Jiang Y Z, Deng Z H, Chung F L, et al. Recognition of epileptic EEG signals using a novel multiview TSK fuzzy system[J]. IEEE Trans on Fuzzy Systems, 2017, 25(1): 3-20.