# 核典型关联性分析相关特征提取与核逻辑斯蒂回归域自适应学习

刘建伟[1],孙正康[1],刘泽宇[2],罗雄麟[1]
(1. 中国石油大学(北京)自动化系,北京 102249;
2. 中国科学院软件研究所基础软件国家工程研究中心,北京 100190)

**摘 要:** 本文提出了一种利用核典型关联性分析提取源域目标域最大相关特征,使用核逻辑斯蒂回归模型进行域自适应学习的算法,该算法称为 KCCA-DAML(Kernel Canonical Correlation Analysis for Domain Adaptation Learning).该算法基于特征集关联性分析,有效的减小源域和目标域的概率分布差异性,利用提取的最大相关特征通过核逻辑斯蒂回归模型实现源域到目标域的跨域学习.实验比较源域数据上核逻辑斯蒂学习模型、目标域上核逻辑斯蒂学习模型、源域和目标域上核逻辑斯蒂学习模型和 KCCA-DAML 模型,结果显示 KCCA-DAML 在真实数据集上成功的实现了跨域学习.

**关键词:** 域自适应;概率分布差异;相关分析;核逻辑斯蒂回归;正则化模型
**中图分类号:** TP181 **文献标识码:** A **文章编号:** 0372-2112 (2016)12-2908-08
**电子学报 URL:** http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2016.12.014

# Domain Adaptation Learning with Kernel Logistic Regression and Kernel Canonical Correlation Analysis

LIU Jian-wei[1], SUN Zheng-kang[1], LIU Ze-yu[2], LUO Xiong-lin[1]
(1. *Department of Automation, China University of Petroleum, Beijing* 102249, *China*;
2. *National Engineering Research Center for Fundamental Software, Institute of Software, Chinese Academy of Sciences, Beijing* 100190, *China*)

**Abstract:** The domain adaptive learning algorithm using kernel logistic regression model is proposed. The proposed approach use kernel canonical correlation analysis to extract the maximum relevant features of the source and target domain. We dub it as KCCA-DAML(Kernel Canonical Correlation Analysis for Domain Adaptation Learning, KCCA-DAML). Our algorithm is based on canonical correlation analysis, which simultaneously minimizes the incompatibility among source features, target features and instance labels, extract maximum relevant features from source features, target features and instance labels, and use kernel logistic regression domain adaptation learning. In experimental comparison of the kernel logistic model and KCCA-DAML model on source domain data, the target domain data, source and the target domain data, we demonstrate the power of our techniques with the following real-world data sets: Reuters 20 Newsgroups, MNIST handwritten-digits and UCI Dermatology.

**Key words:** domain adaptation; distribution discrepancy; correlation analysis; kernel logistic regression; regularization model

## 1 引言

机器学习任务中,假定训练样例-标签对组成的样本集和测试样例-标签对组成的样本集通常来自同一概率分布,这是保证良好学习性能的基本假设.但在现实应用中,这种假设过于“严苛”,具有很大的局限性.我们经常遇到训练样例-标签对组成的样本集与测试样例-标签对组成的样本集概率分布不一致的情况,例如命名实体识别(Named Entity Recognition, NER)中的文本标注问题就是一种经典的域自适应学习问题.

迁移学习中,假定源域与目标域输入样例的概率分布是一样的,存在多个标签输出预测函数,而域自适

收稿日期:2015-05-25;修回日期:2015-11-18;责任编辑:覃怀银
基金项目:国家重点基础研究发展规划(973 计划)项目(No. 2012CB720500)

应学习做相反的假设,即假定源域与目标域样例标签预测函数相同,源域与目标域输入样例的概率分布不一样. 域自适应学习通过已知源域信息对于未知目标域进行信息处理和挖掘. 目前关于域自适应学习产生了大量的理论研究成果,例如文献[1]对统计分类中的域自适应学习进行了综述;文献[2~4]对域自适应学习的各种误差界理论进行了讨论;文献[5~7]围绕域自适应核学习方法进行了研究和改进;文献[8~12]对多源域自适应学习问题进行了分析和讨论.

域自适应学习算法形式多样[13~15],如核映射函数法、结构对应学习、维数约简与协同聚类和迁移分量分析. 其中核映射函数法应用更为普遍,与域自适应学习正则化技术关联紧密. 找到合适的域自适应学习特征表示需要引入跨域数据依赖正则化项对新的特征空间进行约束. 域自适应学习研究的重点和热点是提出全新的域分布偏差度量判据和高效的域自适应学习算法. 基于特征表示的域自适应学习是当前使用最为广泛的域自适应学习方法,通过将源域和目标域数据映射到新的特征空间中,使源域与目标域的概率分布在新的特征空间下足够接近.

本文提出的核典型关联性分析域自适应学习(Kernel Canonical Correlation Analysis for Domain Adaptation Learning,KCCA-DAML)的主要观点是将源域和目标域的样本映射到再生核希尔伯特空间(Reproducing Kernel Hilbert Space,RKHS)中,保证源域和目标域在新的特征空间下线性可分,同时引入 KCCA 约束,使核空间下源域分布和目标域分布的相关性最大化,域自适应学习场景下,若两领域相关,则两域分布足够靠近,进而实现源域学习模型适应于目标域学习模型.

在 Reuters 20 Newsgroups 数据集、MNIST 手写数字识别数据集和 UCI Dermatology 数据集上进行了实验. 针对四种不同分类模型,比较分析了影响域自适应学习任务有效实现的各种因素和参数选择问题. 实验结果表明 KCCA-DAML 通过对源域学习模型进行分布偏差修正,使源域学习模型逐渐迁移为目标域学习模型,能够通过最大化源域和目标域的特征相关性,保证了源域概率分布和目标域概率分布的差异性足够小,实现跨域学习.

## 2　基本学习模型

源域样本集 $D_{\mathrm{S}}=\{(\boldsymbol{x}_{\mathrm{S},1},y_{\mathrm{S},1}),\cdots,(\boldsymbol{x}_{\mathrm{S},n},y_{\mathrm{S},n})\}$,由源域样例集 $X_{\mathrm{S}}=\{\boldsymbol{x}_{\mathrm{S},1},\cdots,\boldsymbol{x}_{\mathrm{S},n}\}\subset\mathbf{R}^{n\times d}$ 和源域类标签集 $Y_{\mathrm{S}}=\{y_{\mathrm{S},1},\cdots,y_{\mathrm{S},n}\}\subset\mathbf{R}^{n\times 1}$ 组成,其中每个样例包含 $d$ 维特征 $\boldsymbol{x}_{\mathrm{S},i}\in\mathbf{R}^{d}$,对应类标签 $y_{\mathrm{S},i}\in\{+1,-1\}$. 目标域样本集分为少量已标识样本 $D_{\mathrm{LT}}=\{(\boldsymbol{x}_{\mathrm{T},1},y_{\mathrm{T},1}),\cdots,(\boldsymbol{x}_{\mathrm{T},m},y_{\mathrm{T},m})\}$ 和大量未标识样例 $D_{\mathrm{UT}}=\{(\boldsymbol{x}_{\mathrm{T},m+1},?),\cdots,(\boldsymbol{x}_{\mathrm{T},n},?)\}$,其中每个样例包含 $d$ 维特征 $\boldsymbol{x}_{\mathrm{T},i}\in\mathbf{R}^{d}$,对应未知类标签为 $y_{\mathrm{T},i}\in\{+1,-1\}$.

域自适应分类任务的目的是利用源域已标识样本 $D_{\mathrm{S}}$,目标域少量已标识样本 $D_{\mathrm{LT}}$ 和大量未标识样例 $X_{\mathrm{UT}}$,学习一个模型能够准确地对目标域未标识样例集 $D_{\mathrm{UT}}$ 分配类标签. 即学习判别函数 $f=\mathrm{sign}(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i):X\to Y$,预测每个目标域未标识样例 $X_{\mathrm{UT}}$ 的类标签 $Y_{\mathrm{UT}}$,其中非线性映射函数 $\varphi:X\to H$ 将样例映射到特定特征空间,增广权向量 $\boldsymbol{w}=(w_1,\cdots,w_d)^{\mathrm{T}}\in\mathbf{R}^{d}$ 是确定分类平面的特征空间向量.

逻辑斯蒂模型为机器学习中常用的分类模型,逻辑斯蒂分类模型为如下无约束优化问题:

$$\arg\min_{w}L_{\mathrm{avg}}(\boldsymbol{w})=\frac{1}{n}\sum_{i=1}^{n}\sigma((\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x})y_i)\tag{1}$$

其中 $\sigma(z)=\ln(1+\exp(-z))$,对于给定的样例 $\boldsymbol{x}_i\in\mathbf{R}^{d}$,使用相应的逻辑斯蒂模型,能够得到如下的逻辑斯蒂分类器:

$$y_i=\mathrm{sign}(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i)\tag{2}$$

其中定义符号函数:

$$\mathrm{sign}(z)=\begin{cases}+1, & z>0\\ -1, & z\leqslant 0\end{cases}\tag{3}$$

逻辑斯蒂模型置信度为:

$$P(Y=y\mid X=\{\boldsymbol{x}\};\boldsymbol{w})=\frac{1}{1+\exp(-\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}y)}\tag{4}$$

域自适应学习的基本观点在于充分利用源域大量先验信息,并通过源域和目标域的偏差度量判据约束解空间,使学习得到的分类判别函数 $f(\boldsymbol{x},y;\boldsymbol{w})$ 由源域判别函数 $f_{\mathrm{S}}(\boldsymbol{x},y;\boldsymbol{w}_{\mathrm{S}})$ 逐步转变为目标域判别函数 $f_{\mathrm{T}}(\boldsymbol{x},y;\boldsymbol{w})$.

定义核矩阵:

$$K=[\boldsymbol{k}(\boldsymbol{x}_i,\boldsymbol{x}_j)]_{i,j=1}^{n}\tag{5}$$

核映射:

$$\varphi:X_S=\{\boldsymbol{x}_{S,1},\cdots,\boldsymbol{x}_{S,n}\}\to\varphi(X_S)=[\varphi(\boldsymbol{x}_{S,1}),\cdots,\varphi(\boldsymbol{x}_{S,n})]\tag{6}$$

学习判别函数:

$$f=\mathrm{sign}(\boldsymbol{w}^{\mathrm{T}}\varphi(\boldsymbol{x}_i)):X\to Y\tag{7}$$

源域核逻辑斯蒂分类模型为:

$$\arg\min_{w}L_{\mathrm{avg}}(\boldsymbol{w})=\frac{1}{n}\sum_{i=1}^{n}\sigma((\boldsymbol{w}^{\mathrm{T}}k_{\mathrm{S},i})y_{\mathrm{S},i})\tag{8}$$

其中 $k_{\mathrm{S},i}=[k(\boldsymbol{x}_{\mathrm{S},i},\boldsymbol{x}_{\mathrm{S},1}),\cdots,k(\boldsymbol{x}_{\mathrm{S},i},\boldsymbol{x}_{\mathrm{S},n})]=[k_{\mathrm{S},i,1},\cdots,k_{\mathrm{S},i,n}]$.

目标域核逻辑斯蒂分类模型为:

$$\arg\min_{w}L_{\mathrm{avg}}(\boldsymbol{w})=\frac{1}{n}\sum_{i=1}^{n}\sigma((\boldsymbol{w}^{\mathrm{T}}k_{\mathrm{T},i})y_{\mathrm{T},i})\tag{9}$$

其中 $k_{\mathrm{T},i}=[k(\boldsymbol{x}_{\mathrm{T},i},\boldsymbol{x}_{\mathrm{T},1}),\cdots,k(\boldsymbol{x}_{\mathrm{T},i},\boldsymbol{x}_{\mathrm{T},n})]=[k_{\mathrm{T},i,1},\cdots,k_{\mathrm{T},i,n}]$.

## 3 KCCA-DAML 模型

源域和目标域之间存在差异性导致源域逻辑斯蒂分类模型并不能很好的适用于目标域学习任务. 需要引入跨域数据依赖正则化项约束逻辑斯蒂分类模型的解空间,将数据嵌入到再生核希尔伯特核空间中,通过最小化源域和目标域的最大分布偏差,保证源域和目标域足够邻近,使源域和目标域在 RKHS 中具有相近的概率分布,解决跨领域学习问题.

当前域自适应学习常用的分布偏差度量为基于均值的偏差度量判据(Maximum Mean Discrepancy, MMD),是一种较为简单直观的度量判据. 但是,仅从均值特征来描述变量差异性并不能充分挖掘特征变量的差异性. 典型相关分析(Canonical Correlation Analysis, CCA)是一种分析多变量相关性的有效方法. 典型相关分析由 Hotelling 首次提出[16],并研究了两组变量之间的相关系数. 用单变量 Pearson 系数难以从整体描述两组多变量之间的关联程度,而 CCA 很好的解决了这一问题.

KCCA-DAML 的主要观点是将源域和目标域的样本映射到再生核希尔伯特空间中,保证源域和目标域在新的特征空间下线性可分,同时利用核化 CCA 约束使源域类分布和目标域类分布尽可能靠近. 域自适应学习场景下,若两域相关,则两域分类器各自的权值 $\boldsymbol{w}$ 值应相近. 通过在源域目标函数中增加 $\| \boldsymbol{w}-\boldsymbol{w}_{\mathrm{T}} \|_2^2$ 项,实现从源域到目标域的迁移学习. 其中,$\| \boldsymbol{w}-\boldsymbol{w}_{\mathrm{T}} \|_2^2$ 表示两领域分类器的差异程度,$\| \boldsymbol{w}-\boldsymbol{w}_{\mathrm{T}} \|_2^2$ 值越大则分类器间差异越大,反之越小,参数 $\lambda$ 用以控制惩罚程度. 得到如下的域自适应学习模型:

$$\arg\min_{k,f} \frac{1}{n}\sum_{i=1}^{n}\sigma((\boldsymbol{w}^{\mathrm{T}}\boldsymbol{k}_{\mathrm{S},i})y_{\mathrm{S},i}) + \lambda \| \boldsymbol{w}-\boldsymbol{w}_{\mathrm{T}} \|_2^2 \tag{10}$$

上式第二项为域自适应学习的真实分布偏差,由于不能直接获得 $\| \boldsymbol{w}-\boldsymbol{w}_{\mathrm{T}} \|_2^2$,因此如何找到合适的正则化项逼近 $\| \boldsymbol{w}-\boldsymbol{w}_{\mathrm{T}} \|_2^2$ 是域自适应学习实现的关键.

通常采用式(9)的解 $\hat{\boldsymbol{w}}_{\mathrm{T}}$ 近似 $\boldsymbol{w}_{\mathrm{T}}$,即用 $\| \boldsymbol{w}-\hat{\boldsymbol{w}}_{\mathrm{T}} \|_2^2$ 代替 $\| \boldsymbol{w}-\boldsymbol{w}_{\mathrm{T}} \|_2^2$. 由于目标域已标识样本不足,式(9)的解 $\hat{\boldsymbol{w}}_{\mathrm{T}}$ 与真实 $\boldsymbol{w}_{\mathrm{T}}$ 存在偏差,目标域样本数目越少,式(9)的解越不准确. 因此本文采用修正的权 $\boldsymbol{w}_{\mathrm{DAML}}$ 来逼近真实 $\boldsymbol{w}_{\mathrm{T}}$:

$$\boldsymbol{w}_{\mathrm{DAML}} = \hat{\boldsymbol{w}}_{\mathrm{T}} + a\Delta\boldsymbol{w} \tag{11}$$

其中 $a$ 为比例控制参数,$\boldsymbol{w}_{\mathrm{DAML}} = \boldsymbol{w}_{\mathrm{KCCA}} - \hat{\boldsymbol{w}}_{\mathrm{T}}$. $\boldsymbol{w}_{\mathrm{KCCA}}$ 为核空间下与源域权 $\boldsymbol{w}_{\mathrm{S}}$ 相关性最高的权,通过 KCCA 变量关联性分析得到.

源域和目标域在核空间下相关,则真实 $\boldsymbol{w}_{\mathrm{T}}$ 与源域权 $\boldsymbol{w}_{\mathrm{S}}$ 相关性应足够高,故与 $\boldsymbol{w}_{\mathrm{KCCA}}$ 可表示为 $\boldsymbol{w}_{\mathrm{T}} = \hat{\boldsymbol{w}}_{\mathrm{T}} + (\boldsymbol{w}_{\mathrm{T}} - \hat{\boldsymbol{w}}_{\mathrm{T}}) = \hat{\boldsymbol{w}}_{\mathrm{T}} + a(\boldsymbol{w}_{\mathrm{KCCA}} - \hat{\boldsymbol{w}}_{\mathrm{T}})$.

KCCA 数据依赖正则化项为:

$$\gamma_{\mathrm{KCCA}}(D_{\mathrm{S}}, D_{\mathrm{T}}) = \| \boldsymbol{w}-\boldsymbol{w}_{\mathrm{DAML}} \|_2^2 \tag{12}$$

得到 KCCA-DMAL 学习模型:

$$\arg\min_{w} \frac{1}{n}\sum_{i=1}^{n}\sigma((\boldsymbol{w}^{\mathrm{T}}k_{\mathrm{S},i})y_{\mathrm{S},i}) + \lambda \| \boldsymbol{w}-\boldsymbol{w}_{\mathrm{DAML}} \|_2^2 \tag{13}$$

## 4 KCCA 偏差度量判据

经过核映射后源域判别函数 $f_{\mathrm{S}} = \varphi(\boldsymbol{w}_{\mathrm{S}})^{\mathrm{T}}\varphi(X_{\mathrm{S}})$ 和目标域判别函数 $f_{\mathrm{T}} = \varphi(\boldsymbol{w}_{\mathrm{T}})^{\mathrm{T}}\varphi(X_{\mathrm{T}})$ 相关性较高,则能实现源域到目标域的迁移学习,$f_{\mathrm{S}}$ 与 $f_{\mathrm{T}}$ 相关性越高,迁移学习效果越好.

使用标准 CCA[16] 对源域和目标域进行关联性分析,对域的样本进行归一化,其中源域样本:

$$D_{\mathrm{S}} = \{(\boldsymbol{x}_{\mathrm{S},1}, y_{\mathrm{S},1}), \cdots, (\boldsymbol{x}_{\mathrm{S},n}, y_{\mathrm{S},n})\} \tag{14}$$

目标域样本:

$$D_{\mathrm{T}} = (\{\boldsymbol{x}_{\mathrm{T},1}, y_{\mathrm{T},1}), \cdots, (\boldsymbol{x}_{\mathrm{T},m}, y_{\mathrm{T},m}), (\boldsymbol{x}_{\mathrm{T},m+1}, ?), \cdots, (\boldsymbol{x}_{\mathrm{T},n}, ?)\} \tag{15}$$

定义如下向量运算:

$$f_{\mathrm{S}} = \boldsymbol{w}_{\mathrm{S}}^{\mathrm{T}}\boldsymbol{X}_{\mathrm{S}} = (\langle \boldsymbol{w}_{\mathrm{S}}, \boldsymbol{x}_{\mathrm{S},1} \rangle, \cdots, \langle \boldsymbol{w}_{\mathrm{S}}, \boldsymbol{x}_{\mathrm{S},n} \rangle) \tag{16}$$

$$f_{\mathrm{T}} = \boldsymbol{w}_{\mathrm{T}}^{\mathrm{T}}\boldsymbol{X}_{\mathrm{S}} = (\langle \boldsymbol{w}_{\mathrm{T}}, \boldsymbol{x}_{\mathrm{S},1} \rangle, \cdots, \langle \boldsymbol{w}_{\mathrm{T}}, \boldsymbol{x}_{\mathrm{S},n} \rangle) \tag{17}$$

最大化源域和目标域关联性:

$$\begin{aligned}\max\rho &= \max_{\boldsymbol{w}_{\mathrm{S}},\boldsymbol{w}_{\mathrm{T}}} \mathrm{coor}(f_S, f_T) = \max_{\boldsymbol{w}_{\mathrm{S}},\boldsymbol{w}_{\mathrm{T}}} \frac{\langle \boldsymbol{w}_{\mathrm{S}}^{\mathrm{T}}\boldsymbol{X}_{\mathrm{S}}, \boldsymbol{w}_{\mathrm{T}}^{\mathrm{T}}\boldsymbol{X}_{\mathrm{T}} \rangle}{\| \boldsymbol{w}_{\mathrm{S}}^{\mathrm{T}}\boldsymbol{X}_{\mathrm{S}} \|_2^2 \| \boldsymbol{w}_{\mathrm{T}}^{\mathrm{T}}\boldsymbol{X}_{\mathrm{T}} \|_2^2} \\ &= \max_{\boldsymbol{w}_{\mathrm{S}},\boldsymbol{w}_{\mathrm{T}}} \frac{\boldsymbol{w}_{\mathrm{S}}^{\mathrm{T}}\boldsymbol{C}_{\mathrm{ST}}\boldsymbol{w}_{\mathrm{T}}}{\sqrt{\boldsymbol{w}_{\mathrm{S}}^{\mathrm{T}}\boldsymbol{C}_{\mathrm{SS}}\boldsymbol{w}_{\mathrm{S}}\boldsymbol{w}_{\mathrm{T}}^{\mathrm{T}}\boldsymbol{C}_{\mathrm{TT}}\boldsymbol{w}_{\mathrm{T}}}}\end{aligned} \tag{18}$$

其中,$\boldsymbol{w}_1, \boldsymbol{w}_2$ 为 $d\times 1$ 维列向量,$\boldsymbol{X}_{\mathrm{S}}, \boldsymbol{X}_{\mathrm{T}}$ 分别为源域和目标域的 $d\times n$ 维样例矩阵,$\langle \boldsymbol{x}_{\mathrm{S}}, \boldsymbol{x}_{\mathrm{T}} \rangle$ 表示向量内积运算,$\boldsymbol{C}_{\mathrm{ST}}$ 表示源域样例 $\boldsymbol{x}_{\mathrm{S}}$ 与目标域样例 $\boldsymbol{x}_{\mathrm{T}}$ 的协方差矩阵,$\boldsymbol{C}_{\mathrm{SS}}$ 为源域样例 $\boldsymbol{x}_{\mathrm{S}}$ 的方差矩阵,$\boldsymbol{C}_{\mathrm{TT}}$ 为目标域样例 $\boldsymbol{x}_{\mathrm{T}}$ 的方差矩阵.

定义核函数:$K(\boldsymbol{x}_{\mathrm{S}}, \boldsymbol{x}_{\mathrm{T}}) = \langle \varphi(\boldsymbol{x}_{\mathrm{S}}), \varphi(\boldsymbol{x}_{\mathrm{T}}) \rangle$,则式(18)变为:

$$\max\rho = \max_{\boldsymbol{w}_{\mathrm{S}},\boldsymbol{w}_{\mathrm{T}}} \frac{\boldsymbol{w}_{\mathrm{S}}^{\mathrm{T}}\boldsymbol{K}_{\mathrm{S}}\boldsymbol{K}_{\mathrm{T}}\boldsymbol{w}_{\mathrm{T}}}{\sqrt{\boldsymbol{w}_{\mathrm{S}}^{\mathrm{T}}\boldsymbol{K}_{\mathrm{S}}^2\boldsymbol{w}_{\mathrm{S}}\boldsymbol{w}_{\mathrm{T}}^{\mathrm{T}}\boldsymbol{K}_{\mathrm{T}}^2\boldsymbol{w}_{\mathrm{T}}}} \tag{19}$$

其中 $\boldsymbol{K}_{\mathrm{S}}$ 为源域样本数据核矩阵,$\boldsymbol{K}_{\mathrm{T}}$ 为目标域样本数据核矩阵.

通过求解式(8)得到核空间下源域分类向量 $\boldsymbol{w}_{\mathrm{S}}$,故域自适应的 KCCA 求解与普通 KCCA 求解稍有不同,即 $\boldsymbol{w}_{\mathrm{S}}$ 已知.

由式(9)可知 $\rho$ 与 $\|\boldsymbol{w}_{\mathrm{T}}\|$ 无关,故式(9)变为:

$$\begin{aligned}&\max_{\boldsymbol{w}_{\mathrm{T}}} \frac{\boldsymbol{w}_{\mathrm{S}}^{\mathrm{T}}\boldsymbol{K}_{\mathrm{S}}\boldsymbol{K}_{\mathrm{T}}\boldsymbol{w}_{\mathrm{T}}}{\sqrt{\boldsymbol{w}_{\mathrm{S}}^{\mathrm{T}}\boldsymbol{K}_{\mathrm{S}}^2\boldsymbol{w}_{\mathrm{S}}}} \\ &\text{s.t.} \quad \boldsymbol{w}_{\mathrm{T}}^{\mathrm{T}}\boldsymbol{K}_{\mathrm{T}}^2\boldsymbol{w}_{\mathrm{T}} = 1\end{aligned} \tag{20}$$

引入 $a\geqslant 0$,式(20)表示为无约束形式:

$$L(\boldsymbol{w}_{\mathrm{T}},a)=\frac{\boldsymbol{w}_{\mathrm{S}}^{\mathrm{T}}\boldsymbol{K}_{\mathrm{S}}\boldsymbol{K}_{\mathrm{T}}\boldsymbol{w}_{\mathrm{T}}}{\sqrt{\boldsymbol{w}_{\mathrm{S}}^{\mathrm{T}}\boldsymbol{K}_{\mathrm{S}}^{2}\boldsymbol{w}_{\mathrm{S}}}}+a(\boldsymbol{w}_{\mathrm{T}}^{\mathrm{T}}\boldsymbol{K}_{\mathrm{T}}^{2}\boldsymbol{w}_{\mathrm{T}}-1)\tag{21}$$

求导:$\boldsymbol{w}_{\mathrm{T}}=[a\boldsymbol{K}_{\mathrm{T}}^{2}]^{-1}\boldsymbol{K}_{\mathrm{T}}^{\mathrm{T}}\boldsymbol{K}_{\mathrm{S}}^{\mathrm{T}}\boldsymbol{w}_{\mathrm{S}}/\sqrt{\boldsymbol{w}_{\mathrm{S}}^{\mathrm{T}}\boldsymbol{K}_{\mathrm{S}}^{2}\boldsymbol{w}_{\mathrm{S}}}$,代入式(21)并最优化$\min\limits_{a\geqslant 0}L(a)$,即可求解得到最优的 $\boldsymbol{w}_{\mathrm{T}}^{*}=\boldsymbol{w}_{\mathrm{KCCA}}$. 模型式(20)为等式约束问题,可通过 Carl Edward Rasmussen 软件包 minFun(http://learning.eng.cam.ac.uk/carl/)或其它模式识别类商业优化软件包直接求解最优的 $\boldsymbol{w}_{\mathrm{T}}^{*}=\boldsymbol{w}_{\mathrm{KCCA}}$,且避免了常规 KCCA 的奇异值分解等较为复杂的求解优化问题.

## 5　KCCA-DAML 模型求解

KCCA-DAML 模型的优化问题为:

$$\arg\min_{w}\frac{1}{n}\sum_{i=1}^{n}\sigma((\boldsymbol{w}^{\mathrm{T}}\boldsymbol{k}_{\mathrm{S},i})y_{\mathrm{S},i})+\lambda(\boldsymbol{w}-\boldsymbol{w}_{\mathrm{CCA}})^{\mathrm{T}}(\boldsymbol{w}-\boldsymbol{w}_{\mathrm{CCA}})\tag{22}$$

该问题为带正则化的 L2 范数逻辑斯蒂分类问题,优化求解如下:

令

$$L(\boldsymbol{w})=\frac{1}{m}\sum_{i=1}^{m}\sigma(y_i\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i)+(\boldsymbol{w}-\boldsymbol{w}_{\mathrm{CCA}})^{\mathrm{T}}(\boldsymbol{w}-\boldsymbol{w}_{\mathrm{CCA}})$$

更新迭代公式:

$$\boldsymbol{w}^{(t+1)}=\boldsymbol{w}^{(t)}+a^{(t)}\boldsymbol{d}^{(t)}\tag{23}$$

其中 $a^{(t)}$ 为第 $t$ 次迭代的步长,$\boldsymbol{d}^{(t)}$ 为第 $t$ 次迭代的搜索方向,$\nabla L(\boldsymbol{w})$ 为 $L(\boldsymbol{w})$ 关于 $\boldsymbol{w}$ 的导数:

$$\boldsymbol{d}^{(t)}=\begin{cases}-\nabla L(\boldsymbol{w}^{(t)})+\boldsymbol{g}^{(t-1)}\boldsymbol{w}^{(t-1)}, & t\geqslant 1\\ -\nabla L(\boldsymbol{w}^{(t)}), & t=1\end{cases}\tag{24}$$

其中:

$$\boldsymbol{g}^{(t-1)}=\{[\nabla L(\boldsymbol{w}^{(t)})]^{\mathrm{T}}\cdot\nabla L(\boldsymbol{w}^{(t)})\}/\{[\nabla L(\boldsymbol{w}^{(t-1)})]^{\mathrm{T}}\cdot\nabla L(\boldsymbol{w}^{(t-1)})\}\tag{25}$$

由于:

$$\sigma(a)=\ln(1+\exp(-a))\tag{26}$$

$$\frac{\mathrm{d}\sigma(a)}{\mathrm{d}a}=\frac{-\exp(-a)}{1+\exp(-a)}=-p(y_i|\boldsymbol{x}_i;\boldsymbol{w})\tag{27}$$

故:

$$\frac{\mathrm{d}L}{\mathrm{d}\boldsymbol{w}}=\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{x}_i y_i[-p(y_i|\boldsymbol{x}_i;\boldsymbol{w})]+\lambda(\boldsymbol{w}-\boldsymbol{w}_{\mathrm{CCA}})^{\mathrm{T}}$$
$$=-\frac{1}{m}[p(Y|X;\boldsymbol{w})]\boldsymbol{A}+\lambda(\boldsymbol{w}-\boldsymbol{w}_{\mathrm{CCA}})^{\mathrm{T}}\tag{28}$$

其中 $\boldsymbol{A}=[y_1\boldsymbol{x}_1,\cdots,y_n\boldsymbol{x}_n]^{\mathrm{T}}\in\mathbf{R}^{n\times d}$,$p(Y|X;\boldsymbol{w})=[p(y_1|\boldsymbol{x}_1;\boldsymbol{w}),\cdots,p(y_n|\boldsymbol{x}_n;\boldsymbol{w})]^{\mathrm{T}}\in\mathbf{R}^{n}$.

由等式(25)、式(28)可以确定搜索方向,式(23)中的步长可以通过如下优化问题得到:

$$\arg\min_{a^{(t)}\geqslant 0}f(\boldsymbol{w}^{(t)}+a^{(t)}\boldsymbol{d}^{(t)})\tag{29}$$

式(29)为单变量优化问题,使用 Carl Edward Rasmussen 软件包 minFun 求解. 通过逐步迭代更新,可以求解上述问题. 最后给出基于关联性分析的域自适应学习算法.

**算法 1　KCCA-DAML 域自适应学习算法**

**输入**:源域已标识样本 $D_{\mathrm{S}}$,目标域已标识样本 $D_{\mathrm{LT}}$ 和未标识样例 $X_{\mathrm{UT}}$,停止准则 $\tau=10^{-7}$.

初始化 $\boldsymbol{w}_j^{(0)}=1/\sqrt{n+1}$,$\boldsymbol{d}^{(0)}=-\nabla L(\boldsymbol{w}^{(0)})$,

$a^{(0)}:=\arg\min\limits_{a^{(0)}\geqslant 0}f(\boldsymbol{w}^{(0)}+a^{(0)}\boldsymbol{d}^{(0)})$.

令重复下面过程,直到满足 $\nabla L(\boldsymbol{w}^{(t)})=0$:

　用 minFun 求解式(23)得到 $\boldsymbol{w}^{(t)}$;

　计算 $g^{(t-1)}=\{[\nabla L(\boldsymbol{w}^{(t)})]^{\mathrm{T}}\cdot\nabla L(\boldsymbol{w}^{(t)})\}/\{[\nabla L(\boldsymbol{w}^{(t-1)})]^{\mathrm{T}}\cdot\nabla L(\boldsymbol{w}^{(t-1)})\}$;

　计算 $d^{(t)}=-\nabla L(\boldsymbol{w}^{(t)})+g^{(t-1)}\boldsymbol{w}^{(t-1)}$;

　计算 $\arg\min\limits_{a^{(t)}\geqslant 0}f(\boldsymbol{w}^{(t)}+a^{(t)}\boldsymbol{d}^{(t)})$;

**输出**:目标域权向量 $\boldsymbol{w}$.

## 6　实验结果及分析

本节通过实验对 KCCA-DAML 在分类方面的性能进行研究. 目前广泛使用的域自适应数据集有 Reuters 20 Newsgroups 数据集、Amazon reviews benchmark 数据集和 Wall Street Journal 语料库数据集等,这些数据集最先应用于自然语言处理方面的研究,随后被广泛用于跨域学习问题的研究当中,此外数据特征“飘移”导致的数据分布差异也是目前常见的域自适应学习问题. 本文选择以下三种广泛使用的真实数据集进行实验:Reuters 20 Newsgroups 数据集(http://kdd.ics.uci.edu/databases/20newsgroups);MNIST 手写数字识别数据集(http://yann.lecun.com/exdb/mnist);UCI Dermatology 数据集(http://archive.ics.uci.edu/ml/datasets.html);为讨论跨域学习的影响因素,实验按源域数据逻辑斯蒂学习模型(S-KLLM,Source-Kernel Logistic Model)、目标域逻辑斯蒂学习模型(T-KLLM,Target-Kernel Logistic Model)、源域 + 目标域逻辑斯蒂学习模型(ST-KLLM,Source and Target-Kernel Logistic Model)、KCCA-DAML 模型进行训练与测试,并给出 KCCA-DAML 在三种数据集上的实验结果和参数选择方案.

待调节参数设定为 $\lambda\in[2^{-4},\cdots,2^{-1},1,2,\cdots,2^{10}]$ 和 $p\in[0.5,0.6,0.7,\cdots,1.4,1.5]$,为简化计算复杂度,实验中使用网格搜索过程确定每组数据集参数. 对于每组参数取值,执行算法 1 中的过程.

### 6.1　Reuters 20 Newsgroups 数据集

Reuters 20 Newsgroups 报文数据集具有层次结构,包含 7 个大类:共 20 个小类,实验选择 comp 和 rec 两大类数据,使用 comp 的 4 个小类:comp.windows.x、comp.os.ms-windows、comp.sys.ibm.pc.hardware 和 comp.sys.mac.hardware. 路透社报文数据集的基本信息如表 1 所示.

**表 1 Reuters 20 Newsgroups 报文数据集**

| 大类 | 小类 | 原始数据特性 |
|---|---|---|
| Comp | comp. graphics | 970 ×61188 |
| | comp. windows. x | 982 ×61188 |
| | comp. os. ms-windows | 963 ×61188 |
| | comp. sys. ibm. pc. hardware | 979 ×61188 |
| | comp. sys. mac. hardware | 958 ×61188 |
| Rec | rec. autos | 987 ×61188 |
| | rec. motorcycles | 993 ×61188 |
| | rec. sport. baseball | 991 ×61188 |
| | rec. sport. hockey | 997 ×61188 |
| Alt | talk. politics. guns<br>talk. politics. mideast<br>sci. electronics 等 | 9954 ×61188 |
| Misc | | |
| Sci | | |
| Soc | | |
| Talk | | |

按照如下方式构造源域和目标域数据. 包含 comp 域迁移学习 rec 域的两类任务.

任务 1:comp. windows. x 作为源域中的正类、rec. autos 作为源域中的负类;comp. os. ms-windows 作为目标域中的正类、rec. motorcycles 作为目标域中的负类.

任务 2:comp. sys. ibm. pc. hardware 作为源域中的正类、rec. sport. baseball作为源域中的负类;comp. sys. mac. hardware 作为目标域中的正类、rec. sport. hockey 作为目标域中的负类. 源域和目标域数据构成如图 1 所示.

20 Newsgroups 数据集为 18774 × 61188 的词频矩阵,选用 comp 和 rec 词频数据大于 30 次的特征作为样本特征,并使用 TI-IDF 软件对数据进行处理,得到数据信息如表 2 所示:

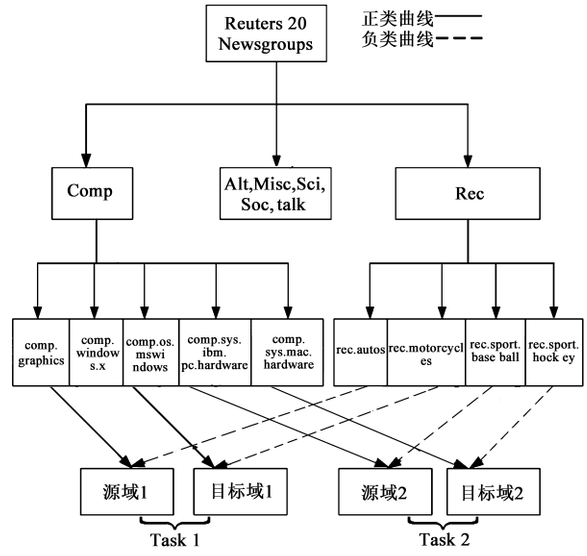


**图1 源域数据集和目标域数据集构造**

**表 2 源域数据集和目标域数据集构成**

| 学习任务 | 源域数据 | | | | 目标域数据 | | | |
|---|---|---|---|---|---|---|---|---|
| | 正样例 | 个数 | 负样例 | 个数 | 正样例 | 个数 | 负样例 | 个数 |
| Task1 | comp. windows. x | 982 ×522 | rec. autos | 987 ×522 | comp. os. ms-windows | 963 ×522 | rec. motorcycles | 993 ×522 |
| Task2 | comp. sys. ibm. pc. hardware | 979 ×522 | rec. sport. baseball | 991 ×522 | comp. sys. mac. hardware | 958 ×522 | rec. sport. hockey | 997 ×522 |

表3 是跨域学习任务 Task1 和 Task2 上的分类误差率. 从表中跨域学习任务 Task1 上的分类误差率结果可以看出,在 $a = 1.1$ 处得到了最小分类误差率 8.31,此时参数 $\lambda = 4$,$\boldsymbol{w}_{\mathrm{KCCA}}$与 $\boldsymbol{w}_{\mathrm{T}}$ 的相关性较大. 说明在两个域相关性较高的情况下,源域数据对目标域数据具有较好的迁移效果. 此外当源域数据的迁移效果较好时,即当已知源域和目标域关联性较高时,参数 $a$ 的值可在 0.8 ~ 1.2 范围内选择. 从表中 Task2 上的分类误差率结果可以看出,在 $a = 1.1$ 处得到了最小分类误差率 9.07,此时参数 $\lambda = 2$,此时两个域相关性较不高,源域数据对目标域数据具有较弱的迁移效果,如果过多考虑源域信息,会产生负迁移,使迁移学习退化为源域的学习.

**表 3 Task1 和 Task2 上的误差率**

| | a | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 |
| Task1 分类误差率(%) | 31.00 | 33.09 | 25.00 | 27.17 | 24.13 | 14.37 | **8.31** | 10.73 | 24.67 | 32.47 | 30.16 |
| Task1 $\lambda$ | 0.063 | 4 | 32 | 0.25 | 0.063 | 0.5 | 4 | 16 | 4 | 0.063 | 1 |
| Task2 分类误差率(%) | 21.10 | 31.45 | 21.75 | 30.78 | 22.39 | 26.78 | 19.22 | 15.04 | **9.62** | 36.31 | 22.57 |
| Task2 $\lambda$ | 0.063 | 0.032 | 32 | 2 | 0.063 | 4 | 0.063 | 32 | 2 | 4 | 0.5 |

表4 是模型 S-KLLM,T-KLLM,ST-KLLM 及 KCCA-DAML 在跨域学习任务 Task1 和 Task2 上的分类误差率,其中 T-KLLM 训练样本数目为 150. 从表中结果可以看出,任务 Task1 的源域与目标域的相关性高于 Task2,对应的 KCCA-DAML 的 Task1 分类误差率也小于 Task2. 此外样本的迁移学习效果越差,源域的跨域学习性能越受限,跨域学习机的学习效果也会受到影响. 当源域和目标域分布偏差足够大,甚至源域和目标域无显著关联时,实现跨域学习仍是十分困难的.

**表 4　不同模型下任务 1 和任务 2 的误差率**

| 分类误差率(%) | S-KLLM | T-KLLM | ST-KLLM | KCCA-DAML |
|---|---|---|---|---|
| Task1 | 17.71 | 27.36 | 17.82 | **8.31** |
| Task2 | 21.06 | 31.21 | 21.95 | **9.62** |

目标域训练样本不足导致 T-KLLM 学习误差较大,此外 ST-KLLM 的分类误差与 S-KLLM 的分类误差相接近,即将源域与目标域合并训练,跨域学习误差不一定减小,原因在于混合训练样本中源域样本在数量上占优,起到了主导作用. 只有在充分考虑源域信息和域关联信息的前提下,域自适应学习机才能实现良好的跨域学习.

### 6.2　MNIST 手写数字识别数据集

MNIST 手写数字识别数据集由 500 个训练样本和 300 个测试样例组成,每个样例的维数是 784,采用构造特征偏差(feature bias)数据集的方法构造源域和目标域数据集,使源域和目标域分布不同,方法为:随机选择训练样本的 375 个属性列,按数值大小选各属性值最大的 375 个训练样本作为源域训练样本,剩余样本为目标域样本集,从中随机选择 100 个训练样本构成目标域训练样本集,剩余样例作为目标域测试. 由于源域样本偏差特征值为各样本最大值,不能准确反映目标域特征的真实情况,导致源域判别函数不能准确预测目标域. 同时,目标域数已标识样本数据样本数目太少,包含目标域信息不完全,也不能准确预测目标域真实分布.

表5 是 MNIST 数据集的分类误差率. 从表中结果可以看出,参数 $a$ 在范围 0.6 ~ 1.4 范围内变化时,对分类误差率没有产生明显影响,但跨域数据依赖正则化项的引入能够保证跨域学习分类误差得以改善并不产生恶化.

**表 5　MNIST 数据集误差率**

| MNIST | $a$ | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 |
| 分类误差率(%) | 7.77 | 5.21 | 5.17 | 6.12 | 7.06 | 5.41 | 6.92 | 5.88 | **4.70** | 6.78 | 7.30 |
| $\lambda$ | 0.5 | 1 | 16 | 0.063 | 0.5 | 4 | 4 | 2 | 4 | 16 | 0.25 |

表6 是模型 S-KLLM,T-KLLM,ST-KLLM 及 KCCA-DAML 在 MNIST 数据集上的分类误差率.

**表 6　不同模型下 MNIST 数据集误差率**

| 分类误差率(%) | S-KLLM | T-KLLM | ST-KLLM | KCCA-DAML |
|---|---|---|---|---|
| MNIST | 11.53 | 12.94 | 9.65 | **4.70** |

MNIST 数据集实验中,将源域与目标域合并训练,跨域学习误差减小,这受益于数据集特性以及构造源域和目标域的方法. 和 Reuters 20 Newsgroups 数据集实验相比较,构造特征偏差数据集的方法引起的域分布差异性要小于 Reuters 20 Newsgroups 数据集子类差异.

### 6.3　UCI Dermatology 数据集

本节使用 UCI Dermatology 据集进行实验,数据集由 366 个样本数据,每个样例的维数是 33,同 MNIST 数据一样,采用构造特征偏差数据集的方法对源域和目标域数据进行构造,使源域和目标域分布不同.

选择 acanthosis, hyperkeratosis, parakeratosis, clubbing of the rete ridges, elongation of the rete ridges, exocytosis, PNL infiltrate, spongiosis, follicular horn plug 这 9 个特征作为偏差特征. 选择每个偏差特征值中最大的十个样本(样本大小为 9 × 10)作为源域训练样本,剩余样本为目标域样本集,从中随机选择 30 个样本构成目标域训练样本集,选择剩余 240 个样例作为目标域测试.

KCCA-DAML 在 Dermatology 数据集上的类误差率如表7 所示.

表 7 UCI Dermatology 数据集误差率

| UCI | a | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 |
| 分类误差率(%) | 7.92 | 10.42 | 9.17 | 11.25 | 7.06 | 5.83 | **5.42** | 6.67 | 7.08 | 8.75 | 14.17 |
| $\lambda$ | 0.5 | 1 | 16 | 0.062 | 0.5 | 4 | 4 | 2 | 4 | 16 | 0.25 |

UCI Dermatology 数据集实验在 $a=1.1$ 处得到了最小分类误差率 5.42,此时参数 $\lambda=0.0625$. 表 8 是模型 S-KLLM, T-KLLM, ST-KLLM 及 KCCA-DAML 在 UCI Dermatology 数据集上进行跨域学习的分类误差率.

表 8 不同模型下 UCI Dermatology 数据集误差率

| 分类误差率(%) | S-KLLM | T-KLLM | ST-KLLM | KCCA-DAML |
|---|---|---|---|---|
| UCI Dermatology | 8.33 | 14.58 | 7.08 | **5.42** |

## 7 结论

本文提出的域自适应学习算法 KCCA-DAML 及 KCCA 域自适应度量判据能够有效的揭示源域特征与目标域特征变的潜在关联性,从而对不同领域的差异性进行度量. 通过对源域模型进行增量修正,使源域模型逐渐迁移至目标域模型,实现跨域学习. KCCA-DAML 模型在跨域学习任务中具有可行性且学习性能良好. 此外利用跨域学习中的已知先验信息,合适的选择模型参数,可使 KCCA-DAML 获得更好的迁移效果,实现更为精确的跨域学习任务. 逻辑斯蒂模型适用于多类学习,因而 KCCA-DAML 可应用于多域自适应学习场景,这是我们下一步要做的工作.

### 参考文献


[1] 刘建伟,孙正康,罗雄麟. 域自适应学习研究进展[J]. 自动化学报,2014,40(8):1576-1600.
Liu Jianwei, SUN Zhengkang, LUO Xionglin. Review and research development on domain adaptation learning[J]. Acta Automatica Sinica, 2014, 40(8): 1576 – 1600. (in Chinese)

[2] Mansour Y, Mohri M, Rostamizadeh A. Multiple source adaptation and the Rényi divergence[A]. Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence[C]. Montreal, Canada: AUAI Press, 2009. 367 – 374.

[3] Blitzer J, Crammer K, Kulesza. A. Learning bounds for domain adaptation[A]. Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems [C]. Vancouver, British Columbia, Canada: Curran Associates, 2007. 129 – 136.

[4] Cortes C, Mansour Y, Mohri M. Learning bounds for importance weighting[A]. Proceedings of the Twenty-Four Annual Conference on Neural Information Processing Systems[C]. Vancouver, Canada: Curran Associates, 2010. 442 – 450.

[5] Tao Jianwen, Chung Fulai, Wang Shitong. A kernel learning framework for domain adaptation learning[J]. Science China Information Sciences, 2012, 55(9): 1983-2007.

[6] Malandrakis N, Potamianos A, Iosif E. Kernel models for affective lexicon creation[A]. 12th Annual Conference of the International Speech Communication Association[C]. Florence, Italy: International Speech Communication Association, 2011. 2977 – 2980.

[7] Kulis B, Saenko K, Darrell T. What you saw is not what you get: Domain adaptation using asymmetric kernel transforms[A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. Colorado, USA: Springs, 2011. 1785 – 1792.

[8] Ben-David S, Blitzer J, Crammer K. A theory of learning from different domains[J]. Machine Learning, 2010, 79(1 – 2): 151 – 175.

[9] Joshi M, Cohen W W, Dredze M. Multi-domain learning: when do domains matter? [A]. Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning[C]. Jeju, Island, Korea: Association for Computational Linguistics, 2012. 1302 – 1312.

[10] Joshi M, Dredze M, Cohen W W. What's in a domain? Multi-domain learning for multi-attribute data[A]. Proceedings of the NAACL-HLT[C]. Atlanta, Georgia, USA: Association for Computational Linguistics, 2013. 685 – 690.

[11] Mansour Y, Mohri M, Rostamizadeh A. Domain adaptation with multiple sources[A]. Proceedings of the Twenty-Second Annual Conference on Neural Information Processing Systems[C]. Vancouver, British Columbia, Canada: Curran Associates, 2008. 1041 – 1048.

[12] Chapelle O, Shivaswamy P, Vadrevu S. Boosted multi-task learning[J]. Machine Learning, 2011, 85(1 – 2): 149 – 173.

[13] Duan L, Xu D, Tsang I W. Domain adaptation from multiple sources: A domain-dependent regularization approach [J]. IEEE Transactions on Neural Networks and Learning Systems, 2012, 23(3): 504 – 518

[14] Schölkopf B, Smola A J, Williamson R C. New support vector algorithms[J]. Neural Computation, 2000, 12(5): 1207-1245.

[15] Joachims T. Transductive inference for text classification using support vector machines[A]. Proceedings of the Sixteenth International Conference on Machine Learning [C]. Bled, Slovenia: Morgan Kaufmann, 1999. 200 – 209.

[16] H Hotelling. Relations between two sets of variates[J]. Biometrika, 1936, 28(3): 312 – 377.

## 作者简介

**刘建伟(通信作者)**　男,1966 年出生. 博士,中国石油大学(北京)副研究员,主要研究方向包括智能信息处理,机器学习,非线性分析与控制,算法分析与设计等.
E-mail: liujw@cup.edu.cn

**孙正康**　男,硕士,1990 年出生. 中国石油大学(北京)地球物理与信息工程学院硕士研究生,研究方向为机器学习.
E-mail: sunzhengkang@126.com