

# 领域适应学习算法研究与展望

孟娟 胡谷雨 潘志松 周宇欢

(解放军理工大学指挥信息系统学院 南京 210007)

**摘要** 领域适应学习旨在利用源领域中带标签的样本来解决目标领域的学习问题,其关键在于如何最大化地减小领域间的分布差异,有效解决领域间数据分布的变化。对当前领域适应学习算法进行了归纳和分类,总结了每类算法的特点,分析了 5 个相关典型算法并比较了其性能。最后指出了领域适应学习值得进一步探索的方向。

**关键词** 领域适应学习,最大均值差,实例加权,特征映射

**中图分类号** TP181 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.10.002

## Research and Perspective on Domain Adaptation Learning Algorithms

MENG Juan HU Gu-yu PAN Zhi-song ZHOU Yu-huan

(Institute of Command Information System, PLA University of Science and Technology, Nanjing 210007, China)

**Abstract** Domain adaptation learning aims to solve the learning problem of target domain by using the labeled samples of source domain. The key challenge is how to minimize the distribution distance among different domains at most and solve the change of data distribution effectively. Domain adaptation learning algorithms were summed up and classified. The characteristics of each type learning algorithm were summarized. Five typical algorithms were carefully analyzed and their performances were compared. What directions are worthy of further exploration was indicated.

**Keywords** Domain adaptation learning, Maximum mean discrepancy, Instance weighting, Feature mapping

### 1 引言

传统的机器学习方法假设训练数据和测试数据服从于相同的概率分布,即是独立同分布的(Independent and Identically Distributed, IID),但在实际应用中这个假设并不总得到满足,在许多应用中存在非 IID 数据<sup>[1]</sup>,如语际(Cross-language)文本挖掘、Wi-Fi 定位<sup>[2]</sup>等,这些应用中精确标注的任务特定数据较少,而任务相关的数据却大量存在。近年来提出的领域适应学习(Domain adaptation learning)通过学习分布不同但相关的源领域(Source domain)带标签样本来帮助目标领域(Target domain)学习。领域适应学习放松了对训练数据和测试数据同分布假设的要求,能够利用不同领域的数据来帮助目标任务的学习,有效解决了训练数据和测试数据分布不一致的学习问题。该问题吸引了越来越多研究者的关注和研究<sup>[3-13]</sup>。领域适应学习着眼于解决源领域和目标领域的分布差异,也被称为类别失衡(Class imbalance)<sup>[14]</sup>、协方差偏移(Covariate shift)<sup>[15]</sup>和样本选择偏置(Sample selection bias)<sup>[16]</sup>等。随着应用的深入,人们考虑如何从多个源领域适应到目标领域,这时不仅需要解决领域间的分布差异,还需要考虑多个源领域之间的相关性。相关研究有多任务学习(Multi-task learning)<sup>[17]</sup>和半监督学习(Semi-supervised

learning)<sup>[18]</sup>等。本文对当前领域适应学习算法进行了归纳和分类,将算法分为分布差异和源差异两大类进行分析,总结了每类算法的特点,仔细分析了 5 种典型算法,并通过实验对比其性能。最后展望了领域适应学习进一步的研究方向。

### 2 相关概念与问题描述

对于一个模式分类问题,设数据领域为  $D$ ,领域数据概率分布为  $P(x, y), x \in X, y \in Y$ ,其中  $X$  和  $Y$  分别指领域内数据实例及其对应的类标签,分类器为一个映射函数  $f(x): X \rightarrow Y$ ,其将实例  $x \in X$  映射为相应的类标签  $y \in Y$ 。在领域适应学习中,大量带标签的训练数据为源领域  $D_s$ ,没有或仅有少量标签的测试数据为目标领域  $D_t$ 。 $P_s(X, Y), P_t(X, Y)$  分别表示源领域和目标领域的联合概率分布, $P_s(X), P_s(Y), P_t(X), P_t(Y)$  分别表示  $X$  和  $Y$  在源领域和目标领域的边缘概率分布, $P_s(X|Y), P_s(Y|X), P_t(X|Y), P_t(Y|X)$  分别表示两个领域的条件概率分布。 $D_s$  中所有带标签的数据集为  $D_s = \{(x_i^s; y_i^s)\}_{i=1}^{N_s}$ ;  $D_t$  中未标记的数据集为  $D_{t,u} = \{x_i^{t,u}\}_{i=1}^{N_{t,u}}$ ,有时  $D_t$  中可能会有少量的标记数据,记为  $D_{t,l} = \{(x_i^{t,l}; y_i^{t,l})\}_{i=1}^{N_{t,l}}$ 。领域适应学习旨在通过学习分布不同但相关的源领域和目标领域上的数据来获取相应的统计模型,提高目标领域数据分类的精度。领域适应学习一个最主要的关键性问题就

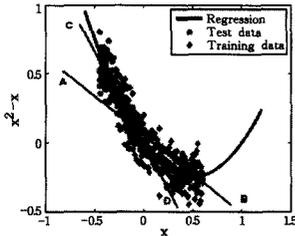
到稿日期:2014-10-11 返修日期:2015-01-16 本文受国家八六三高技术研究与发计划基金项目(2012AA01A510),国家博士后基金项目(2013M542425)资助。

孟娟(1978-),女,博士生,主要研究方向为模式识别、机器学习,E-mail:mengjuan\_cn@gmail.com;胡谷雨(1963-),男,教授,博士生导师,主要研究方向为网络管理、卫星通信;潘志松(1973-),男,教授,博士生导师,主要研究方向为模式识别、机器学习;周宇欢(1980-),男,博士后,主要研究方向为语音信号处理、模式识别。

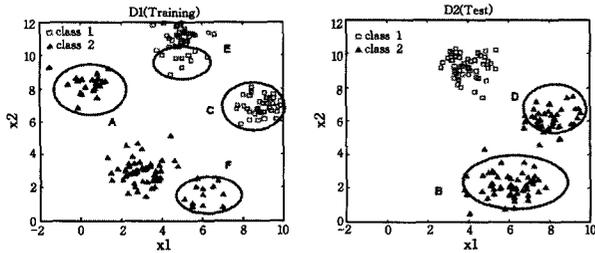
是如何选取有效的度量去反映源领域和目标领域之间的分布差异。

### 3 分布差异算法

训练数据和测试数据之间的分布差异可分为两种类型：边缘概率分布差异和条件概率分布差异。边缘概率分布差异如图 1(a)所示，不同形状的点分别代表源领域和目标领域，两个分布的数据点都是由  $Y=X^2-X$  产生，条件概率分布相同，但从图 1(a)可看出训练数据回归线 AB 明显不同于测试数据回归线 CD，源领域和目标领域的边缘概率分布不同。图 1(b)和图 1(c)表示训练数据和测试数据的边缘概率分布和条件概率分布都不同，A 数据点的边缘概率不同于 B 和 D 数据点，C 和 D 数据点的条件概率也不同。



(a) 训练数据和测试数据的边缘概率差异



(b) 二分类中训练数据的概率分布 (c) 二分类中测试数据的概率分布

图 1

按分布差异的不同可将领域适应学习算法分为基于边缘概率分布差异、基于条件概率分布差异和基于边缘与条件概率分布差异这三类算法。针对每类分布差异的特点，本文将仔细分析其相关典型算法。

#### 3.1 基于边缘概率分布差异的领域适应学习算法

基于边缘概率分布差异的领域适应学习算法是目前研究最多的一类算法。该算法主要围绕实例加权 and 均值中心对齐等方法。

在目标领域，根据统计学习理论，给定损失函数  $\ell(x, y, \theta)$  和模型的参数空间  $\Theta$ ，目标是找到参数  $\theta^*$ ，满足

$$\theta^* = \arg \min_{\theta \in \Theta} \sum_{(x,y) \in D_t} P(D_t) \ell(x, y, \theta) \quad (1)$$

由于训练集是来自源领域的联合分布，在目标领域没有被标记的数据在训练时可以被观察到，因此需要从源领域数据学习一个模型来代替。如果  $P(D_s) = P(D_t)$ ，则可以简单地通过求解式(2)的优化问题来学习模型用于目标领域。

$$\theta^* = \arg \min_{\theta \in \Theta} \sum_{(x,y) \in D_s} P(D_s) \ell(x, y, \theta) \quad (2)$$

当  $P(D_s) \neq P(D_t)$  时，优化问题变为式(3)：

$$\begin{aligned} \theta^* &= \arg \min_{\theta \in \Theta} \sum_{(x,y) \in D_s} \frac{P(D_t)}{P(D_s)} P(D_s) \ell(x, y, \theta) \\ &\approx \arg \min_{\theta \in \Theta} \sum_{i=1}^{n_s} \frac{P_t(x_i, y_i)}{P_s(x_i, y_i)} \ell(x_i, y_i, \theta) \end{aligned} \quad (3)$$

可见，利用  $\frac{P_t(x_i, y_i)}{P_s(x_i, y_i)}$  对实例  $(x_i, y_i)$  进行加权调整能够得到一个理想的学习函数。

##### 3.1.1 协方差偏移

如果源领域和目标领域  $X$  的分布不同，即  $P_s(X) \neq P_t(X)$  而  $P_s(Y|X) = P_t(Y|X)$ ，这种分布差异叫协方差偏移，在解决协方差偏移的问题中，式(3)中的  $\frac{P_t(x, y)}{P_s(x, y)}$  可改写为：

$$\frac{P_t(x, y)}{P_s(x, y)} = \frac{P_t(x)P_t(y|x)}{P_s(x)P_s(y|x)} = \frac{P_t(x)}{P_s(x)} \quad (4)$$

协方差偏移常需要对  $\frac{P_t(X)}{P_s(X)}$  进行估计。

目前，研究人员已经提出了多种估计方法。Shimodaira<sup>[16]</sup>提出了非参数的核密度估计方法。Huang<sup>[19]</sup>提出一种在核空间中学习  $P_t(X)/P_s(X)$  的方法 KMM(Kernel Means Matching)，该方法在一个再生核希尔伯特空间(Reproducing Kernel Hilbert Space, RKHS)中匹配源领域和目标领域的样本的核均值，其好处是避免了直接估计  $P_s(X)$  和  $P_t(X)$ 。Sugiyama<sup>[20]</sup>指出了 KMM 的缺陷，提出了 KLIEP(Kullback-Leibler Importance Estimation Procedure) 直接对  $P_t(X)/P_s(X)$  进行估计，这种方法基于对  $P_t(X)/P_s(X)$  的相对熵最小化的求解，KLIEP 结合交叉验证来对模型进行自动选择：首先是对源领域数据的样本权重进行估计，其次在被赋予权重的样本基础上进行模型的选择和训练。Bickel<sup>[21]</sup>避免解决求分布比率  $P_t(X)/P_s(X)$  的模型，转而对所有的源数据的权重直接进行估计，将此问题归结为一个凸优化模型并利用牛顿梯度下降来进行优化。Liao<sup>[22]</sup>提出了一个主动学习的方法，通过引入一个阈值来评估辅助数据中每一个样本对目标分类的贡献程度，在源领域数据的帮助下，选择目标领域的未标注数据进行标注。

核均值匹配算法 KMM，通过在再生核希尔伯特空间匹配源领域和目标领域的平均值来直接学习  $P_t(X)/P_s(X)$ ，KMM 可以表示为式(5)的二次规划优化问题。

$$\min_{\beta} \frac{1}{2} \beta^T K \beta - K^T \beta \quad (5)$$

$$\text{s. t. } \beta_i \in [0, B] \text{ 且 } \left| \sum_{i=1}^{n_s} \beta_i - n_s \right| \leq n_s \epsilon$$

$$K = \begin{pmatrix} K_{s,s} & K_{s,t} \\ K_{t,s} & K_{t,t} \end{pmatrix}, K_{ij} = k(x_i, x_j) \quad (6)$$

式(6)中， $K_{s,s}$  和  $K_{t,t}$  分别为源领域数据和目标领域数据的核矩阵， $K_{s,t}$  和  $K_{t,s}$  为跨域核矩阵。

$$K_i = \frac{n_s}{n_t} \sum_{j=1}^{n_t} k(x_i, x_j) \quad (7)$$

其中， $x_i \in X_s \cup X_t$ ， $x_j \in X_t$ ，可以证明  $\beta_i = P_s(X)/P_t(X)$ 。采用 KMM 的优点是它可以避免  $P_s(X)$  和  $P_t(X)$  的密度估计，当数据集很小时密度估计是很难的。

##### 3.1.2 类别不平衡

如果源领域和目标领域  $Y$  的分布不同，即  $P_s(Y) \neq P_t(Y)$  而  $P_s(X|Y) = P_t(X|Y)$ ，这种分布差异叫类别不平衡，在解决类别不平衡的问题中，式(3)中的  $\frac{P_t(x, y)}{P_s(x, y)}$  可改写为：

$$\frac{P_t(x, y)}{P_s(x, y)} = \frac{P_t(y)P_t(x|y)}{P_s(y)P_s(x|y)} = \frac{P_t(y)}{P_s(y)} \quad (8)$$

因此，需要使用  $P_t(y)/P_s(y)$  进行实例加权。Lin<sup>[23]</sup>成功

地将支持向量机应用于这种非标准的学习情形,并通过模拟实验进行验证。Chan<sup>[24]</sup>用  $P_t(y)/P_s(y)$  对每个来自源领域的训练实例进行加权,并使用 naive Bayes 分类器来解决问题。还有研究者采用训练集重采样方法,从源域重新采样训练实例,使得重新采样的数据具有和目标域大致相同的类分布。Chawla<sup>[25]</sup>提出的 SMOTE 算法首先为每个目标域类样本随机选出几个邻近样本,并且在该样本与这些邻近的样本的连线上随机取点,生成无重复的新的目标域类样本。Kubat 和 Matwin<sup>[26]</sup>采用单边采样方式,去除源域类中的噪音样本、边界样本和冗余样本。

对于分类算法而言,可通过求概率分布  $P(Y|X)$  直接建模,如 logistic regression 分类器。理论表明求  $P_t(y|x)$  问题可转化为求  $P_s(y|x)$ 。

$$P_t(y|x) = \frac{r(y)P_s(y|x)}{\sum_{y' \in \mathcal{Y}} r(y')P_s(y'|x)} \quad (9)$$

其中,  $r(y) = P_t(y)/P_s(y)$ , 现在可以先从源域估计  $P_s(y|x)$ , 然后利用  $P_s(y)$  和  $P_t(y)$  得到  $P_t(y|x)$ 。

在实践中,人们需要知道目标域中的类分布,以适用上述方法。在一些研究中,假定这个分布是已知的先验。

### 3.2 基于条件概率分布差异的领域适应学习算法

当  $P_t(Y|X) \neq P_s(Y|X)$  时, Jiang<sup>[27]</sup> 建议标注少量目标领域样本,然后构造目标领域模型,并用该目标模型过滤掉源领域中与目标领域存在争议的样本  $(x_i, y_i)$ , 通过更多地使用目标领域数据的信息来对源数据进行评估,并将源数据中被认为具有“误导”作用的样例剔除。Gao<sup>[28]</sup> 提出了局部加权集成 (Locally Weighted Ensemble, LWE), 该方法从集成的角度出发,将多个传统的分类器集成起来,并提出一种局部加权的方法对各个子分类器的权重进行动态调整。这种分类器集成了多个分类器的优点,可以同时利用多个源数据的信息来帮助目标数据学习。

LWE 利用  $k$  个模型加权集成  $E$ , 计算未标记的目标域数据  $x$  的标记  $y$ 。

$$P(y|E, x) = \sum_{i=1}^k w_{M_i, x} P(y|M_i, x) \quad (10)$$

其中,  $P(y|M_i, x)$  是  $k$  个模型中的某个模型  $M_i$  对目标数据  $x$  的预测,  $w_{M_i, x}$  为数据点  $x$  在模型  $M_i$  中获得的权重,其定义如下:

$$w_{M_i, x} = \frac{s(G_{M_i}, G_T; x)}{\sum_{i=1}^k s(G_{M_i}, G_T; x)} \quad (11)$$

其中,  $G_{M_i}$  和  $G_T$  是围绕  $x$  点的图。  $G_T$  是通过连接测试数据中与  $x$  属于同一聚类的点构建的,  $G_{M_i}$  连接由模型  $M_i$  确定的与  $x$  属于同一类的点,  $s(G_{M_i}, G_T; x)$  是测试数据在两个图中  $x$  点的相似性度量。

通过最小化风险来学习  $x$  的标记  $y^*$ :

$$y^* = \arg \min_y \int_{y \in \mathcal{Y}} \lambda(y', y) P(y|E, X) dy' \quad (12)$$

其中,  $\lambda(y', y)$  是代价函数,其真正的类标签是  $y'$ , 但预测值是  $y$ , 如以最常见的 0-1 损失函数为代价,则  $y^* = \arg \max_y P(y|E, X)$ 。

### 3.3 基于边缘概率与条件概率分布差异的领域适应学习算法

当  $P_s(X) \neq P_t(X)$  或  $P_s(Y) \neq P_t(Y)$ ,  $P_s(Y|X) \neq P_t(Y|X)$  时,不仅要解决边缘概率分布差异,还要解决条件概率分布差异。目前相关研究较少,基本方法是通过域映射来减少

边缘概率的分布差异,然后在映射域中进行样本选择,以减小条件概率的分布差异。

Zhong<sup>[29]</sup> 提出的 KMapEnsemble (KE) 使用核判别分析 (Kernel Discriminant Analysis, KDA) 进行领域适应。其目标是通过核映射空间选择出最靠近目标数据的一部分源数据来构建分类器,但是映射空间较依赖于源数据。找到一个合适的核映射空间,使得不同空间的数据在空间中具有相近边缘分布,从而能够利用 Bisecting K-means 聚类<sup>[30]</sup> 选择出那些与目标数据条件概率相近的源数据。为避免单一分类器的缺陷,通过集成的方式控制了误差边界。

算法分两步:

(1) 基于核的特征映射

KMapEnsemble 首先使用迭代过程来生成多个映射,并使用模型均值来计算最终的预测。标记的目标领域数据首先被用来执行 KDA。

$$\max \{ \lambda = (\frac{\alpha^T K W K \alpha}{\alpha^T K K \alpha}) \} \quad (13)$$

其中,核矩阵  $K_{ij} = \phi(x_i) \phi(x_j)$ ,  $W = (W_i)_{i=1, \dots, N_C}$  是块对角矩阵,  $W_i$  是  $l_i \times l_i$  的矩阵,元素值都等于  $\frac{1}{l_i}$ , 用特征值分解求解可获得映射空间。来自两个域数据的边缘分布  $P(x)$  在这个映射空间内是相似的。对于新的实例  $z$ , 可得到映射

$$v \phi(z) = \sum_{i=1}^{N_C} \sum_{j=1}^{l_i} a_{ij} K(X_{ij}, z) \quad (14)$$

其中,  $X_{ij}$  是类  $i$  的第  $j$  个实例。

(2) 基于聚类的样本选择

在迭代中,基于聚类标准选择源领域数据,这些源领域数据同目标领域数据具有类似的条件概率  $P(y|x)$ 。然后,利用 KDA 将这些选择的源领域数据和标记的目标领域数据构建为一个新的特征空间并训练新的分类器。

## 4 源差异算法

随着领域适应学习研究的进一步深入,源领域从一个扩展到多个。基于源差异,领域适应学习可分为基于单源的领域适应学习和基于多源的领域适应学习,如图 2 所示。

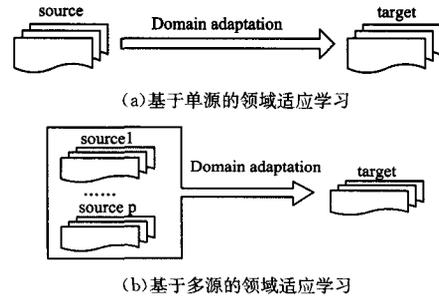


图 2

目前对单源领域适应学习的研究较多,前面提到的 KMM、KE 等算法都是研究如何减小一个源领域和目标领域的差异,而对于多源的领域适应学习,一个关键问题是如何从多个源领域中选择适应目标领域的源样本、源特征或源参数。常用方法是通过源分类器的选择纠正源偏差,进行域选择;或通过距离测度进行多源领域适应学习。

### 4.1 基于单源的领域适应学习算法

基于单源的领域适应学习算法研究如何利用一个源领域

的样本帮助目标领域学习,目标是减小这个源领域与目标领域的差异。Blitzer<sup>[3]</sup>提出的 SCL(Structure Correspondence Learning)算法,首先在领域上定义  $m$  个枢纽特征集,然后将这些特征集从数据集中去掉并将每个枢纽特征定义为一个新的标签向量,从而建立起  $m$  个分类问题。算法在目标领域当中利用无标签的数据提取出一些可以减小领域间误差的特征表示。Daume<sup>[4]</sup>提出一种核映射的方法来处理 NLP 问题,通过一个高维的特征空间将源领域和目标领域相互联系起来,然后利用传统的分类学习算法来训练分类器。Yang<sup>[31]</sup>提出的 ASVM(Adaptive SVM)以一个或多个现有的分类器适应新的数据集,它的目的是使用类似于支持向量机的目标函数学习源分类器和适应性分类器之间的“ $\delta$  函数”。Schweikert<sup>[32]</sup>提出的 MCC(Multiple Convex Combination)通过求解一个线性规划模型来优化所有凸组合系数。Gretton<sup>[33]</sup>提出了最大均值差异(Maximum Mean Discrepancy, MMD)的分布距离度量方法。基于 MMD 准则, Pan<sup>[34]</sup>首先通过 MMDE(Maximum Mean Discrepancy Embedding)学习一个低维的映射空间,并在空间中使得源数据和目标数据的均值中心对齐,以减小两个领域的差异。经过这种处理后的数据可以直接利用传统分类器对数据进行训练和泛化。由于 MMDE 的计算开销太大,在此基础上提出 TCA(Transfer Components Analysis)的领域特征适应算法,其主要学习一个领域间可适应的特征成分集,减小了领域间特征分布差距。

在再生核希尔伯特空间中源领域  $D_s$  与目标领域  $D_t$  的 MMD 可以表示为:

$$Dist(D_s, D_t) = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \Phi(x_{s_i}) - \frac{1}{n_t} \sum_{j=1}^{n_t} \Phi(x_{t_j}) \right\|_H \quad (15)$$

其中,  $\Phi$  是一个从原始输入空间到高维 Hilbert 空间的非线性映射。从式(15)可以看出, MMD 度量就是使用源领域数据集和目标领域数据集的总体均值之差来表示源领域与目标领域之间的分布差异。

TCA 基于 MMDE,该算法的主要思想是通过如下优化模型学习  $\varphi$ , 映射源领域和目标领域的的数据到一个低维嵌入子空间,减小领域间的差异并保持原始数据结构。

$$\min_{\varphi} Dist(\varphi(X_s), \varphi(X_t)) + \lambda \Omega(\varphi) \quad (16)$$

s. t. constraints on  $\varphi(X_s)$  and  $\varphi(X_t)$

$$Dist(\varphi(X_s), \varphi(X_t)) = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \Phi(\varphi(x_{s_i})) - \frac{1}{n_t} \sum_{j=1}^{n_t} \Phi(\varphi(x_{t_j})) \right\|_H \quad (17)$$

如果在 RKHS 中选取合适的核函数  $k$ , 则式(16)可以转化为下式:

$$\min_{\varphi} tr(KL) + \lambda \Omega(\varphi) \quad (18)$$

s. t. constraints on  $\varphi(X_s)$  and  $\varphi(X_t)$

其中  $K$  是定义在源域、目标域和跨域的核矩阵,  $tr(A)$  表示矩阵  $A$  的迹运算。

## 4.2 基于多源的领域适应学习算法

近年来,研究者提出了一些基于多源的领域适应学习方法<sup>[35-37]</sup>。Crammer<sup>[38]</sup>假定多个源的分布是一样的,标签的变化是由噪声变化量引起的。Luo<sup>[39]</sup>结合多源数据提出了共同最大化的学习模型。Mansour<sup>[40]</sup>估计每个源的数据分布,对不同源的权重进行调整。这些算法的理论证明基础是假设每个源域数据对目标域数据的预测能力较强。Shi<sup>[41]</sup>通过源领

域构造的特征空间来帮助目标领域的结构划分,基于谱分析方法设计了一种在源领域和目标领域数据标签空间不一致下的领域适应框架。Yang<sup>[42]</sup>提出用适应性集成方法(Adaptive ensemble approach)来解决 NLP 的多源领域适应问题,该算法通过集合生成、跨域知识学习和决策推断实现多源领域适应学习。

Duan<sup>[43]</sup>提出了基于多源的领域适应机(Domain Adaptation Machine, DAM),利用从多个源领域的标记样本独立学习到的一系列辅助分类器来进行目标领域的标签预测。受流正则<sup>[44]</sup>和基于图形的多任务学习<sup>[45]</sup>的启发,提出了基于平滑假设的数据依赖正则化,使目标域的分类与相关源域对目标域的未标记模式的辅助分类有类似的决策值,并采用了稀疏正则化使得目标分类器具有稀疏性。

DAM 采用如下优化模型:

$$\min_{f^T} \Omega(f^T) + \frac{1}{2} \sum_{i=1}^{n_t} (f_i^T - y_i^T)^2 + \Omega_D(f_i^T) \quad (19)$$

其中,  $\Omega(f^T)$  是用于控制目标分类器  $f^T$  的正则项,第二项是目标分类器  $f^T$  在目标域的标记模式  $D_t^T$  下的经验误差,最后一项是数据依赖正则化项。

DAM 不同于 MTL (Multi-Task Learning), 它侧重于利用现有的辅助分类器学习目标决策分类,并在学习阶段计算成本显著降低。此外,根据数据依赖正则化的定义,辅助分类器可以用不同的训练内核,甚至不同的学习方法。

## 5 领域适应学习算法性能比较

本节将在两个跨领域文本分类数据集 20Newsgroups (<http://people.csail.mit.edu/jrennie/20Newsgroups/>) 和情感评论 (<http://www.cs.jhu.edu/~mdredze/datasets/sentiment>) 上进行领域适应分类实验,对前面分析的 5 种典型算法做性能比较。

### 5.1 数据集描述与设置

由于数据集较大,为便于测试算法的分类性能,随机抽取数据集每个子类的 10% 样本作为测试样本。基于 20NG 数据集和情感评论数据集的领域适应学习任务的描述如表 1 所列。

表 1 领域适应学习任务

数据集	任务
1	comp. os, ms-windows, misc
2	comp. sys, ibm, pc, hardware
3	comp. sys, mac, hardware
4	comp. windows, x
5	rec. motorcycles
6	rec. sport, baseball
7	20Newsgroups rec. sport, hockey
8	sci. electronics
9	sci. med
10	sci. space
11	talk. politics, mideast
12	talk. politics, misc
13	talk. religion, misc
1	Books
2	DVD
3	情感评论 Electronics
4	Kitchen

(1) 20NG 数据集: 为了有效比较上述 5 种典型领域适应学习算法的分类性能, 本文采用与文献<sup>[46]</sup>相同的实验设置。即对于 20NG 数据集, 从 20 个类中抽取表 1 中的 13 类作为

源域和目标域, 剩余 7 类 (alt. atheism, comp. graphics, misc. forsale, rec. autos, sci. crypt, soc. religion. christian, talk. politics, guns.) 作为负类。数据基于子类进行分割, 认为不同的子类来自不同的领域。

(2) 情感评论数据集: 情感评论领域适应数据集由 4 种不同的产品类型 (书籍、DVD、电子产品和厨房设备用品) 的产

品评论数据组成。可使用来自其他产品的评论作为训练数据, 通过领域适应学习所得模型来预测某产品的评论情感类别。

## 5.2 性能比较

20Newsgroups 文本分类和情感评论数据集领域适应学习的实验结果分别记录于表 2 和表 3 中, 其中, Avg 列表示各方法在同一数据集上的平均分类精度值。

表 2 20Newsgroups 文本分类精度 (%) 比较

	1	2	3	4	5	6	7	8	9	10	11	12	13	Avg
KMM	67.93	78.62	84.41	65.51	59.82	63.44	82.32	79.48	79.92	70.46	79.55	78.07	81.76	74.72
LWE	53.98	60.88	69.32	60.75	71.21	60.63	65.15	52.76	59.77	54.12	61.18	50.47	62.39	60.20
KE	69.01	63.72	67.46	58.93	76.71	65.04	59.50	64.99	68.06	71.75	58.73	62.54	66.27	65.60
TCA	68.46	76.02	72.51	65.89	73.21	54.82	71.15	67.38	69.57	81.39	77.92	70.16	65.79	70.33
DAM	68.11	72.15	70.53	64.12	73.91	58.27	72.95	67.84	71.44	61.57	72.83	59.40	62.32	67.34

表 3 情感评论分类精度 (%) 比较

	1	2	3	4	Avg
KMM	78.66	72.20	85.21	81.06	79.28
LWE	60.34	59.51	62.44	60.87	60.79
KE	69.85	68.77	66.39	69.71	68.68
TCA	75.20	71.18	74.16	82.33	75.72
DAM	74.62	71.85	70.12	66.94	70.88

由表 2 和表 3 的实验结果可知:

由于 20Newsgroups 和情感评论数据集是基于边缘概率差异的, 因此 KMM 算法在这两个数据集上的实验结果较好。

基于特征映射的单源领域适应算法 TCA 解决的是源领域和目标领域之间的边缘概率差异, 因而在 20Newsgroups 和情感评论数据集上的实验结果也较好。

基于边缘和条件概率分布差异的 KE 算法分类性能优于基于条件概率分布差异的 LWE 算法。

多源领域适应算法 DAM 算法性能与多个源域和目标域的相关性有关。DAM 算法本质是减小多个源领域和目标领域的边缘概率差异, 在 20Newsgroups 和情感评论数据集上的性能优于 LWE。

**结束语** 通过对领域适应学习算法的分析, 可以发现不论是基于分布差异还是基于源差异, 目前的领域适应学习算法都是以传统的机器学习算法为基础, 通过领域间分布偏差的估计、公共特征学习或者是领域知识的形式化等手段来调整基础算法的优化过程, 进而使其对目标领域具有更好的泛化效果, 其主要方法是:

### (1) 实例加权

源领域的部分训练数据可以通过重加权而被用于目标领域, 通过某种方式对来自源领域的训练样本进行权重计算, 以决定每个样本在训练过程中的重要性大小。源领域训练数据的加权策略是这种技术的关键。

### (2) 特征映射

通过特征映射在源领域和目标领域之间传递有用的知识, 为目标领域学习一个“好”的特征表示。源领域的知识通过特征表示适应到目标领域, 在该特征表示下, 目标领域的学习性能能得到明显的提升。

### (3) 模型适应

源领域和目标领域的模型间共享某些参数或者某些先验分布。通过模型适应可将源领域模型的相关参数共享到目标领域, 因此这种领域适应学习的目标就是发现模型间共享的参数。

领域适应学习已成为机器学习领域一项具有挑战性的研

究课题, 涌现出一批新颖的算法。通过分析发现, 领域适应学习还存在很多问题需要并值得去研究。

### (1) 如何优化领域适应学习算法性能。

领域适应学习主要用于解决非 IID 问题, 优化领域适应学习算法性能不仅要考虑算法的正确率和计算复杂度, 还需考虑算法的泛化性、算法的简洁性、迭代次数与计算时间、对于大规模问题的可扩展性、近似解的收敛性等, 研究更高效的领域适应学习算法依然是个公共课题。

### (2) 如何在异构环境中进行领域适应学习。

目前领域适应学习研究的问题中源领域和目标领域虽然数据分布不同, 但由于它们处于同构的环境中, 因此可以找到共享的实例、特征或参数等。对于跨媒介的领域适应, 如从文本到图像、从视频到音频的领域适应学习问题中, 源领域和目标领域的环境是异构的, 因此, 如何实现异构环境下的领域适应学习是进一步的研究方向。

(3) 如何控制源领域和目标领域差异较大时出现的负适应。

源领域的判别信息对目标领域的的数据划分并不总是起正面的影响, 在差异较大的领域间进行领域适应学习可能会出现负适应的情况。目前领域适应学习的负适应问题是一个公开课题, 如何控制负适应的发生, 降低领域适应学习过程中源领域对目标领域学习的不利影响; 如何评估领域适应学习带来的风险等都是值得研究的问题。

## 参考文献

- [1] Pan S J, Yang Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22 (10): 1345-1359
- [2] 顾鑫, 王士同. 基于最小包含球的领域迁移学习新方法[J]. 计算机科学, 2013, 40(7): 187-191  
Gu Xin, Wang Shi-tong. Novel domain transfer learning approach using minimum enclosing ball [J]. Computer Science, 2013, 40(7): 187-191
- [3] Ben-David S, Blitzer J, Crammer K, et al. Analysis of representations for domain adaptation[C]// Advances in neural information processing systems. 2007: 137
- [4] Blitzer J, McDonald R, Pereira F. Domain adaptation with structural correspondence learning[C]// Proceedings of the 2006 conference on empirical methods in natural language processing. Association for Computational Linguistics, 2006: 120-128
- [5] Hal Daume III. Frustratingly easy domain adaptation[C]// Proceedings of the 45th Annual Meeting of the Association of Com-

- putational Linguistics, 2007; 256-263
- [6] Satpal S, Sarawagi S. Domain adaptation of conditional probability models via feature subsetting[M]// Knowledge Discovery in Databases, PKDD 2007. Springer Berlin Heidelberg, 2007; 224-235
- [7] Gong B, Grauman K, Sha F. Learning Kernels for Unsupervised Domain Adaptation with Applications to Visual Object Recognition[J]. International Journal of Computer Vision, 2014, 109(1-2); 3-27
- [8] Xia R, Zong C, Hu X, et al. Feature ensemble plus sample selection; domain adaptation for sentiment classification[J]. Intelligent Systems IEEE, 2013, 28(3); 10-18
- [9] Mejova Y, Srinivasan P. Crossing Media Streams with Sentiment; Domain Adaptation in Blogs, Reviews and Twitter[C]// ICWSM, 2012
- [10] Ben-David S, Urner R. On the hardness of domain adaptation and the utility of unlabeled target samples[M]// Algorithmic Learning Theory. Springer Berlin Heidelberg, 2012; 139-153
- [11] Tao J, Chung F, Wang S. On minimum distribution discrepancy support vector machine for domain adaptation[J]. Pattern Recognition, 2012, 45(11); 3962-3984
- [12] Bahirat K, Bovolo F, Bruzzone L, et al. A novel domain adaptation Bayesian classifier for updating land-cover maps with class differences in source and target domains[J]. IEEE Transactions on Geoscience and Remote Sensing, 2012, 50(7); 2810-2826
- [13] Cortes C, Mohri M. Domain adaptation and sample bias correction theory and algorithm for regression[J]. Theoretical Computer Science, 2014, 519; 103-126
- [14] Japkowicz N, Stephen S. The class imbalance problem: A systematic study[J]. Intelligent Data Analysis, 2002, 6(5); 429-449
- [15] Shimodaira H. Improving predictive inference under covariate shift by weighting the log-likelihood function[J]. Journal of Statistical Planning and Inference, 2000, 90(2); 227-244
- [16] Zadrozny B. Learning and evaluating classifiers under sample selection bias[C]// Proceedings of the twenty-first international conference on Machine learning. ACM, 2004; 114
- [17] Caruana R. Multitask learning[M]. Springer US, 1998
- [18] Zhu X. Semi-supervised learning literature survey[D]. University of Wisconsin-Madison, 2005
- [19] Huang J, Gretton A, Borgwardt K M, et al. Correcting sample selection bias by unlabeled data[C]// Advances in Neural Information Processing Systems. 2006; 601-608
- [20] Sugiyama M, Nakajima S, Kashima H, et al. Direct importance estimation with model selection and its application to covariate shift adaptation[C]// Advances in Neural Information Processing Systems. 2008; 1433-1440
- [21] Bickel S, Brückner M, Scheffer T. Discriminative learning under covariate shift[J]. The Journal of Machine Learning Research, 2009, 10; 2137-2155
- [22] Liao X, Xue Y, Carin L. Logistic regression with an auxiliary data source[C]// Proceedings of the 22nd International Conference on Machine Learning. ACM, 2005; 505-512
- [23] Lin Y, Lee Y, Wahba G. Support vector machines for classification in nonstandard situations[J]. Machine Learning, 2002, 46(1-3); 191-202
- [24] Chan Y S, Ng H T. Word Sense Disambiguation with Distribution Estimation[C]// IJCAI. 2005; 1010-1015
- [25] Chawla N V, Bowyer K W, Hall L O, et al. SMOTE: synthetic minority over-sampling technique[J]. Journal of Artificial Intelligence Research, 2002, 16; 321-357
- [26] Kubat M, Matwin S. Addressing the curse of imbalanced training sets: one-sided selection[C]// ICML. 1997, 97; 179-186
- [27] Jiang J, Zhai C X. Instance weighting for domain adaptation in NLP[C]// ACL. 2007; 264-271
- [28] Gao J, Fan W, Jiang J, et al. Knowledge transfer via multiple model local structure mapping[C]// Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2008; 283-291
- [29] Zhong E, Fan W, Peng J, et al. Cross domain distribution adaptation via kernel mapping[C]// Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2009; 1027-1036
- [30] Ren J, Shi X, Fan W, et al. Type-Independent Correction of Sample Selection Bias via Structural Discovery and Re-balancing[C]// SDM. 2008; 565-576
- [31] Yang J, Yan R, Hauptmann A G. Cross-domain video concept detection using adaptive svms[C]// Proceedings of the 15th international conference on Multimedia. ACM, 2007; 188-197
- [32] Schweikert G, Rätsch G, Widmer C, et al. An empirical analysis of domain adaptation algorithms for genomic sequence analysis[C]// Advances in Neural Information Processing Systems. 2009; 1433-1440
- [33] Borgwardt K M, Gretton A, Rasch M J, et al. Integrating structured biological data by kernel maximum mean discrepancy[J]. Bioinformatics, 2006, 22(14); e49-e57
- [34] Pan S J, Tsang I W, Kwok J T, et al. Domain adaptation via transfer component analysis[J]. IEEE Transactions on Neural Networks, 2011, 22(2); 199-210
- [35] 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
- [36] Duan L, Xu D, Chang S F. Exploiting Web images for event recognition in consumer videos: A multiple source domain adaptation approach[C]// 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2012; 1338-1345
- [37] Chattopadhyay R, Sun Q, Fan W, et al. Multisource domain adaptation and its application to early detection of fatigue[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2012, 6(4); 18
- [38] Crammer K, Kulesza A, Dredze M. Adaptive regularization of weight vectors[C]// Advances in Neural Information Processing Systems. 2009; 414-422
- [39] Luo P, Zhuang F, Xiong H, et al. Transfer learning from multiple source domains via consensus regularization[C]// Proceedings of the 17th ACM Conference on Information and Knowledge Management. ACM, 2008; 103-112
- [40] Mansour Y, Mohri M, Rostamizadeh A. Domain adaptation with multiple sources[C]// Advances in Neural Information Processing Systems. 2009; 1041-1048
- [41] Shi X, Fan W, Yang Q, et al. Relaxed transfer of different classes via spectral partition[M]// Machine Learning and Knowledge Discovery in Databases. Springer Berlin Heidelberg, 2009; 366-381

## 参 考 文 献

- [1] Murray S, Goldfarb M. Towards the use of a lower limb exoskeleton for locomotion assistance in individuals with neuromuscular locomotor deficits[C]//Proceeding of IEEE Annual International Conference on Engineering in Medicine and Biology Society, 2012; 1912-1915
- [2] Franch O, Calandre L, Álvarez-Linera L, et al. Gait disorders of unknown cause in the elderly: Clinical and MRI findings[J]. Journal of the Neurological Sciences, 2009, 280(1): 84-86
- [3] World Health Organization. World Health Statistics 2011 [R]. World Health Organization, 2011
- [4] Dietz V, Harkema S J. Locomotor activity in spinal cord-injured persons[J]. Journal of Applied Physiology, 2004, 96(5): 1954-1960
- [5] Wernig A, Müller S, Nanassy A, et al. Laufband therapy based on 'rules of spinal locomotion' is effective in spinal cord injured persons[J]. European journal of neurology, 1995, 7(4): 823-829
- [6] A Roadmap for U. S. Robotics; from Internet to Robotics[OL]. <http://robotics-vo.us/sites/default/files/2013%20Robotics%20Roadmap-rs.pdf>, 2013
- [7] Taha Z, Majeed A P P A, Tze M Y W P. Preliminary Investigation on the Development of a Lower Extremity Exoskeleton for Gait Rehabilitation; A Clinical Consideration [J]. Journal of Medical and Bioengineering, 2015, 4(1): 1-6
- [8] Krebs H, Dipietro L, Levy-Tzedek S, et al. A paradigm shift for rehabilitation robotics[J]. Engineering in Medicine and Biology Magazine, IEEE, 2008, 27(4): 61-70
- [9] Gams A, Petric T, Debevec T, et al. Effects of robotic knee-exoskeleton on human energy expenditure[J]. IEEE Transactions on Biomedical Engineering, 2013, 60(6): 1636-1644
- [10] Jimenez-Fabian R, Verlinden O. Review of control algorithms for robotic ankle systems in lower-limb orthosis, prostheses, and exoskeletons[J]. Medical Engineering & Physics, 2012, 34(4): 397-408
- [11] Zoss A B, Kazerooni H, Chu A. Biomechanical design of the berkeley lower extremity exoskeleton (BLEEX) [J]. IEEE/ASME Transactions on Mechatronics, 2006, 11(2): 128-138
- [12] Low K. Robot-assisted gait rehabilitation; From exoskeletons to gait systems[C]//Proc. Defense Science Research Conference and Expo. 2011; 1-10
- [13] Guizzo E, Goldstein H. The rise of the body bots [J]. Spectrum, IEEE, 2005, 42(10): 50-56
- [14] Walsh C J, Pasch K, Herr H. An autonomous, under actuated exoskeleton for load-carrying augmentation[C]//Proceeding of IEEE/RSJ International Conference on Intelligent Robots and Systems, 2006; 1410-1415
- [15] Liu X, Low K. Development and preliminary study of the NTU lower extremity exoskeleton[C]//Proceeding of IEEE Conference on Cybernetics and Intelligent Systems, 2004; 1243-1247
- [16] Jezernik S, Colombo G, Keller T, et al. Robotic orthotics lokomat: a rehabilitation and research tool [J]. Neuromodulation, 2003, 6: 108-115
- [17] Veneman J F, Kruidhof R, Hekman E E G, et al. Design and evaluation of the LOPES exoskeleton robot for interactive gait rehabilitation[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2007, 15(3): 379-386
- [18] van Asseldonk E H F, van der Kooij H. Robot-Aided Gait Training with LOPES[M]//Neurorehabilitation Technology, 2012; 379-396
- [19] Banala S, Agrawal S, Scholz J. Robot Assisted Gait Training with Active Leg Exoskeleton[J]. IEEE Transaction on Neural System Rehabilitation Engineering, 2009, 17(1): 2-8
- [20] Unluharsicikli O, Pietrusinski M, Weinberg B, et al. Design and control of a robotic lower extremity exoskeleton for gait rehabilitation[C]//2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2011
- [21] Bouri M, Stauffer Y, Schmitt C, et al. The WalkTrainer; A robotic system for walking rehabilitation[C]//Proc. IEEE International Conference on Robotics and Biomimetics(ROBIO'06). 2006; 1616-1621
- [22] Wang P, Low K, Tow K, et al. Initial system evaluation of an overground rehabilitation gait training robot (NaTUre-gaits) [J]. Advanced Robotics, 2011, 25(15): 1927-1948
- [23] Sankai Y. HAL; Hybrid assistive limb based on cybernetics[M]//Robotics Research. Springer, 2011; 25-34
- [24] Kawamoto H, Sankai Y. Power assist method based on phase sequence driven by interaction between human and robot suit[C]//Proceeding of IEEE International Workshop on Robot and Human Interactive Communication, 2004; 491-496
- [25] Sankai Y. Leading edge of cybernetics; Robot suit HAL[C]//SICE-ICASE International Joint Conference, 2006; 1-2
- [26] Swift T, Strausser K, Zoss A, et al. Control and experimental results for post stroke gait rehabilitation with a prototype mobile medical exoskeleton[C]//Proceeding of ASME 2010 Dynamic Systems and Control Conference, 2010
- [27] Wirz M, Rupp R. Clinical application of a robotic device for locomotion[M]//Converging Clinical and Engineering Research on Neurorehabilitation. Berlin Heidelberg; Springer, 2013; 973-975
- [28] Crowell H P, et al. Human engineering design guidelines for a powered, full body exoskeleton[R]. Aberdeen Proving Ground: U. S. Army Res Lab, ARL-TN-60, 1995
- [29] Pons J L. Wearable Robots; Biomechatronic exoskeletons[M]. Wiley & Sons, 2008

(上接第 12 页)

- [42] Yang J B, Mao Q, Xiang Q L, et al. Domain adaptation for coreference resolution: An adaptive ensemble approach[C]//Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Association for Computational Linguistics, 2012; 744-753
- [43] Duan L, Tsang I W, Xu D, et al. Domain adaptation from multiple sources via auxiliary classifiers[C]//Proceedings of the 26th Annual International Conference on Machine Learning. ACM, 2009; 289-296
- [44] Belkin M, Niyogi P, Sindhvani V. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples[J]. Journal of Machine Learning Research, 2006, 7(3): 2399-2434
- [45] Evgeniou T, Micchelli C A, Pontil M. Learning multiple tasks with kernel methods[J]. Journal of Machine Learning Research, 2005, 6: 615-637
- [46] Eaton E, desJardins M. Set-based boosting for instance-level transfer[C]//IEEE International Conference on Data Mining Workshops, 2009(ICDMW'09). IEEE, 2009; 422-428