

基于迁移权重的条件对抗领域适应

王进* 王科 闵子剑 孙开伟 邓欣

(重庆邮电大学数据工程与可视计算重点实验室 重庆 400065)

摘要: 针对条件对抗领域适应(CDAN)方法未能充分挖掘样本的可迁移性, 仍然存在部分难以迁移的源域样本扰乱目标域数据分布的问题, 该文提出一种基于迁移权重的条件对抗领域适应(TW-CDAN)方法。首先利用领域判别模型的判别结果作为衡量样本迁移性能的主要度量指标, 使不同的样本具有不同的迁移性能; 其次将样本的可迁移性作为权重应用在分类损失和最小熵损失上, 旨在消除条件对抗领域适应中难以迁移样本对模型造成的影响; 最后使用Office-31数据集的6个迁移任务和Office-Home数据集的12个迁移任务进行了实验, 该方法在14个迁移任务上取得了提升, 在平均精度上分别提升1.4%和3.1%。

关键词: 迁移学习; 领域适应; 对抗学习; 迁移权重

中图分类号: TP391.41

文献标识码: A

文章编号: 1009-5896(2019)11-2729-07

DOI: 10.11999/JEIT190115

Transfer Weight Based Conditional Adversarial Domain Adaptation

WANG Jin WANG Ke MIN Zijian SUN Kaiwei DENG Xin

(Key Laboratory of Data Engineering and Visual Computing, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: Considering the failure of the Conditional adversarial Domain Adaptation(CDAN) to fully utilize the sample transferability, which still struggle with some hard-to-transfer source samples disturbed the distribution of the target domain samples, a Transfer Weight based Conditional adversarial Domain Adaptation(TW-CDAN) is proposed. Firstly, the discriminant results in the domain discriminant model as the main factor are employed to measure the transfer performance. Then the weight is applied to class loss and minimum entropy loss. It is for eliminating the influence of hard-to-transfer samples of the model. Finally, experiments are carried out using the six domain adaptation tasks of the Office-31 dataset and the 12 domain adaptation tasks of the Office-Home dataset. The proposed method improves the 14 domain adaptation tasks and increases the average accuracy by 1.4% and 3.1% respectively.

Key words: Transfer learning; Domain adaptation; Adversarial learning; Transfer Weight(TW)

1 引言

在标准的有监督学习中, 一个鲁棒的模型往往需要大量的标记样本。然而在一些真实场景中, 存在着标记样本不足、标记困难等问题, 使得传统的监督学习方法难以训练得到一个鲁棒的模型。另外, 在真实场景中, 由于环境的不断变化, 数据的分布也随之改变, 标准的有监督学习只能拟合训练数据中出现的模式, 难以泛化存在分布偏移的新数

据^[1]。针对于上述情形, 领域适应^[2]应运而生。领域适应通常是指利用其它领域的标记样本, 对无标记的目标域进行适配。领域适应在计算机视觉和自然语言处理等领域都有广泛应用, 例如使用电商网站上已分类的图片数据对手机拍摄图片进行分类^[3], 或者使用已有情感标记的电影评论数据对快餐评论数据进行情感预测^[4], 这些问题都属于领域适应研究的范畴。领域适应消除了训练数据和测试数据必须服从独立同分布的限制, 解决了标准监督学习所面临的训练数据不足、训练数据与测试数据存在分布偏移的问题。

减少训练数据和测试数据分布偏移的关键是如何挖掘训练数据和测试数据之间的相似性。现有解决分布偏移问题的方法一般包含两类^[5]: 其一是学习训练数据和测试数据不变的表征向量; 其二是重

收稿日期: 2019-02-27; 改回日期: 2019-06-11; 网络出版: 2019-06-24

*通信作者: 王进 wangjin@cqupt.edu.cn

基金项目: 国家自然科学基金(61806033), 国家社会科学基金西部项目(18XGL013)

Foundation Items: The National Nature Science Foundation of China(61806033), The National Social Science Foundation of China(18XGL013)

新估计样本权重,使两个领域数据的分布尽量相似。由于学习不变的表征向量不限制于源域和目标域中的局部分布一致性,近年来得到更多的关注和发展。学习不变表征向量主要有两种实现途径^[5]:一种是基于偏差度量,其主要思想为通过优化偏差度量函数,求得一个能将源域样本和目标域样本映射到同一特征空间的映射函数。最具代表性的偏差度量方法为最大均值差异(Maximum Mean Discrepancy, MMD)^[6],通过减少MMD确定映射函数,达到领域适应的目的。基于偏差度量的方法主要包含:深度适应网络(Deep Adaptation Network, DAN)^[7],将多个全连接层分别使用一个多核的MMD进行适配;联合适配网络(Joint Adaptation Network, JAN)^[8],对多个连接层使用一个联合的MMD进行适配。另一种是基于对抗学习的方法,主要受生成对抗网络(Generative Adversarial Nets, GAN)^[9]的启发,通过判别模型和特征提取模型的博弈来学习共有特征。随着GAN的蓬勃发展,基于对抗学习的方法逐渐成为研究的重点。领域对抗神经网络(Domain Adversarial Neural Network, DANN)^[10]是首次将对抗学习引入领域适应的方法;对抗判别领域适应(Adversarial Discriminative Domain Adaptation, ADDA)^[11]提出了一种基于对抗学习的领域适应算法框架,使得一般的基于对抗学习的领域适应方法都可以看做是ADDA的一种特例;受条件生成对抗网络(Conditional Generative Adversarial Nets, CGAN)^[12]的启发,条件对抗领域适应(Conditional Domain Adversarial Network, CDAN)^[13]将类别预测向量作为条件信息应用到领域适应问题上,并取得了较好的结果。条件信息的使用,一方面提升了判别模型的性能,另一方面作为样本可迁移性的衡量指标,对于类别预测置信度较高的样本,赋予更高的权重进行特征对齐。

DANN和ADDA等方法将源域数据和目标域数据在特征空间上做完美对齐,并未考虑样本的迁移性。对齐难以迁移的源域样本将扰乱目标域数据分布,影响模型性能。CDAN使用源域和目标域样本预测向量的熵来设置迁移性,但对于样本迁移性的思考并不充分。具体为,如果存在多个相似度较高的类别,CDAN会降低这部分样本对模型的贡献,因此难以拟合类别相似度较高的样本。本文认为样本的可迁移性更多地体现在源域样本和目标域样本的相似度上,如果源域样本和目标域样本更相似,则这部分样本的可迁移性就越高,更应加大这部分样本对模型的贡献。

本文针对CDAN方法在样本迁移性上考虑的不

足,提出基于迁移权重的条件对抗领域适应(Transfer Weight based Conditional adversarial Domain Adaptation, TW-CDAN)方法。本文认为样本的可迁移性还体现在样本的相似度上。具体来讲,样本的可迁移性还体现在判别模型的判别结果上。如果判别模型越不能区分源域样本和目标域样本,源域样本和目标域样本相似度就更高,此时的源域样本更具有迁移性能,模型需加大对这部分样本的拟合。本文的主要贡献在于:(1)使用领域判别模型的预测结果作为条件对抗领域适应中样本迁移性能的衡量指标;(2)提出基于样本迁移权重的条件对抗领域适应,使用(1)的熵作为迁移权重应用在条件领域适应中源域的分类损失和最小熵损失上。经多个对比实验验证,使用领域判别模型的输出结果作为条件对抗领域适应迁移性能的度量指标时,具有更好的性能。

2 基于样本权重的条件对抗领域适应

在无监督领域适应中,通常包含两个领域的数据集,分别为包含 n_s 个已标记样本的源域数据集 $\mathcal{D}_s = \{(\mathbf{x}_i^s, \mathbf{y}_i^s)\}_{i=1}^{n_s}$,和包含 n_t 个未标记样本的目标域数据集 $\mathcal{D}_t = \{\mathbf{x}_j^t\}_{j=1}^{n_t}$,其中, $\mathbf{x}_i$ 为一个样本, $\mathbf{y}_i$ 为样本 $\mathbf{x}_i$ 对应的标记。 $\mathcal{D}_s$ 和 $\mathcal{D}_t$ 分别抽样于总体分布 $P(\mathbf{x}^s, \mathbf{y}^s)$ 和 $Q(\mathbf{x}^t, \mathbf{y}^t)$,且 P 和 Q 违背了样本独立同分布的假设。换言之, P 和 Q 具有一定的分布差异。本文目的就是设计一种模型来减轻 P 和 Q 之间的差异,使得使用 P 训练出的模型,仍然能有效地应用在 Q 上。

传统的领域适应方法多数基于特征对齐,主要包含了两个步骤,首先使用一个特征提取模型 $\mathbf{f} = F(\mathbf{x})$ 来提取样本的特征,然后使用一个分类模型 $\mathbf{g} = G(\mathbf{x})$ 来预测该样本的标记。受生成对抗学习的启发,DANN^[10]提出了一种基于对抗学习的特征对齐,DANN类似于两个模型的博弈:判别模型 D 用来判别样本是否来自于源域数据集;特征提取模型 G 提取样本特征,并使提取的特征最大化混淆判别模型 D 。受条件生成对抗模型的启发,CDAN^[13]将预测的类别向量 $\mathbf{f}$ 作为辅助信息来约束模型的构建:一方面利用预测的类别向量 $\mathbf{g}$ 与提取的特征向量 $\mathbf{f}$ 进行多线性映射得新的向量 $\mathbf{h} = \mathbf{f} \otimes \mathbf{g}$ 来作为判别模型 D 的输入。另一方面,利用预测的类别向量 $\mathbf{g}$ 的熵来控制样本的迁移性。

CDAN中通过样本类别预测的不确定性来降低了部分样本的迁移性能,但是关于迁移性能的探索依然不够。在基于CDAN方法使用类别预测向量和特征向量的多线性映射来作为判别模型输入的基础

上, 本文发现, 样本的迁移性可以具体地表现在判别模型 D 的判别结果 $\hat{d}_i$ 上, 如果判别结果 $\hat{d}_i$ 的置信度越高, 则该样本可迁移性就越低, 反之亦然。因此, 对于判别结果 $\hat{d}_i$ 的置信度较高的样本, 本文尝试了适当降低其对模型网络的影响。主要体现在两方面, 首先是类别预测上, 对于迁移性能较低的样本, 本文尝试减少它对类别预测的影响。其次, 模型进一步地引入了最小熵损失加大类别的分离边界, 同理, 最小熵损失同样可通过领域判别结果 $\hat{d}_i$ 进行加权。

2.1 样本迁移性

样本的迁移性可以直接体现在领域判别模型的输出结果 $\hat{d}_i$ 中, 对于领域标记不确定的样本, 迁移性能越高。因此, 可通过判别模型的输出来设置迁移权重, 使得在使用源域样本作为监督样本训练模型时, 能够根据其迁移性能的不同, 设置不同的样本迁移权重。在领域判别模型中, 可使用领域预测结果 $\hat{d}_i$ 的熵 $H(\hat{d}_i)$ 来计算样本的可迁移性 w_i 为

$$w_i = H(\hat{d}_i) \quad (1)$$

2.2 基于样本迁移权重的分类损失

当使用源域样本作为监督样本训练模型时, 模型往往能够很好地拟合源域样本的分布。在领域适应问题中, 需要使用一个判别模型将目标域数据拟合到源域的数据分布上。因此, 当存在部分迁移性能较低, 或者不能迁移的源域样本时, 判别模型依然会使目标域数据去拟合这部分迁移性能低的源域样本, 造成模型性能下降。因此, 本文需要降低这部分样本对于训练模型的贡献, 主要体现在使用源域样本做监督学习任务时, 对分类损失进行加权。对于迁移性能较高的样本, 赋予更高的权重更新分类模型。因此, 可设置源域的基于样本权重的分类损失 L_y 为

$$L_y = \frac{1}{n_s} \sum_{i=1}^{n_s} w_i \cdot L(g_i^s, y_i^s) \quad (2)$$

2.3 基于样本迁移权重的最小熵损失

模式崩溃是对抗学习的主要挑战之一, 模式崩溃主要是指在使用对抗学习训练模型时, 即使能够混淆判别模型, 也不能拟合真实样本的分布。为了减轻模式崩溃的问题, 有必要设置更加明显的分类边界, 因此引入最小熵损失^[14]加大类别的分离边界, 进一步限制模式崩溃问题。主要体现在, 对源域数据设置最小熵损失 L_h , 并通过样本的迁移性 w_i 进行加权。对于类别比较确定的样本, 本文对熵的权重更大; 反之, 对于类别不够确定的样本, 本

文尽量减少对这部分样本的拟合。因此, 可设置源域的基于样本迁移权重的最小熵损失 L_h 为

$$L_h = -\frac{1}{n_s} \sum_{i=1}^{n_s} \sum_{j=1}^c w_i \cdot g_{i,j} \cdot \lg(g_{i,j}) \quad (3)$$

2.4 基于样本迁移权重的目标函数

基于样本迁移权重的领域适应的目标函数主要包含3部分: 第1部分为源域上基于样本迁移权重的分类损失 L_y , 第2部分为源域上基于样本迁移权重的最小熵损失 L_h , 第3部分为为领域判别模型 D 的领域判别损失 L_d 。基于样本迁移权重的目标函数如式(4)所示。其中 γ 和 λ 为模型的超参数, 分别为最小熵损失的全局权重和领域判别损失的全局权重, $L(\cdot, \cdot)$ 为交叉熵损失。

$$\left. \begin{aligned} & \min_G (L_y + \gamma L_h - \lambda L_d) \\ & = \frac{1}{n_s} \sum_{i=1}^{n_s} w_i \cdot L(g_i^s, y_i^s) \\ & \quad - \frac{\gamma}{n_s} \sum_{i=1}^{n_s} \sum_{j=1}^c w_i \cdot g_{i,j} \cdot \lg(g_{i,j}) \\ & \quad - \frac{\lambda}{n_s + n_t} \sum_{i=1}^{n_s+n_t} L(\hat{d}_i, d_i) \\ & \min_D L_d = \frac{1}{n_s + n_t} \sum_{i=1}^{n_s+n_t} L(\hat{d}_i, d_i) \end{aligned} \right\} \quad (4)$$

TW-CDAN模型结构如图1所示, 在CDAN模型的基础上(黑色部分), 本文将领域判别模型判别结果的熵作为迁移权重, 并应用在源域数据的分类损失和最小熵损失上(红色部分)。对于源域样本 x^s 和目标域样本 x^t , 首先, 使用一个共用的卷积神经网络来提取样本特征, 分别得到源域样本的特征 f^s 和目标域样本的特征 f^t 。其次, 输入到一个共用的用于分类的全连接网络, 分别得到源域样本的类别预测向量 g^s 和目标域样本的类别预测向量 g^t 。然后, 将特征向量 f 和类别向量 g 的多线性映射作为领域判别模型 D 的输入, 得到关于领域的判别损失 L_d 和领域判别结果的熵 $H(\hat{d}_i)$ 。最后, 将源域样本的领域判别结果的熵 $H(\hat{d}_i)$ 作为权重应用在源域样本的分类损失和最小熵损失上, 得到关于源域样本的分类损失 L_y 和最小熵损失 L_h 。因此, 对于迁移性能较高的源域样本, 对模型的训练有更大的影响。

3 实验

本节将基于2个数据集的18个迁移任务, 与多

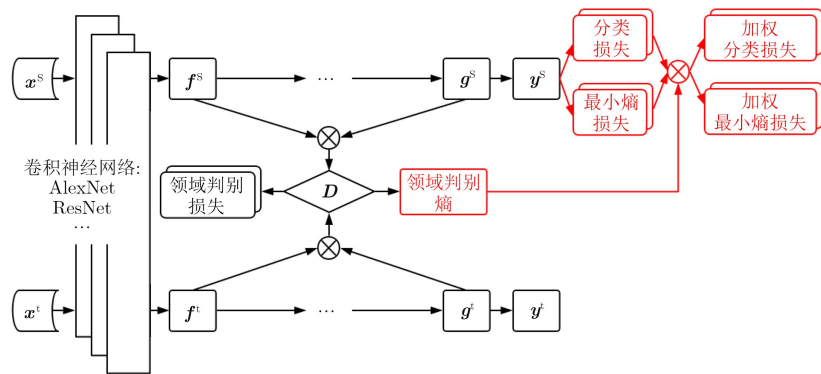


图1 TW-CDAN模型结构图

个主要方法进行实验对比, 并对实验结果进行分析。

3.1 实验设置与数据

Office-31^[15] 是一个应用最为广泛的基于图片领域适应的数据集。一共包含4652张图片, 分为31个类别。这些图片来于3个不同的领域, 分别为Amazon(**A**), Webcam(**W**)和DSLR(**D**)。其中, Amazon为电商网站Amazon.com的商品展示图片; Webcam为图像处理软件Webcam处理后的图; DSLR为数码单反相机拍摄的图片。实验中, 将这3个领域的数据集设置6种迁移任务。

Office-Home^[16] (图2)是一个更具迁移难度的数据集。一共包含了15500张图片, 分为65个类别。Office-Home数据集含有4个领域的图片, 分别为Art(**Ar**), Clipart(**Cl**), Product(**Pr**)和Real World(**Rw**)。其中, Art为艺术画图片; Clipart为插画图片; Product为商品展示图片; Real World为实际生活中所拍摄的图片。在实验中, 将这4个领域的图片设置12种迁移任务。相较于Office-31数据集, Office-Home更具迁移难度, 主要体现在: 首先, Office-Home数据集中各领域差异更大, 相似度更低; 其次, Office-Home数据集中, 包含的类别数更多, 模型需要学习更为复杂的模式。

本文与多个主流方法进行了对比实验, 包含已介绍的DAN^[7], JAN^[8], DANN^[10], ADDA^[11]和CDAN^[13]外, 还有利用最小熵损失来细化分类器性

能的残差迁移网络(Residual Transfer Network, RTN)^[17]和使用生成的像素级跨领域图片来完成领域适应的生成适应网络(Generate To Adapt, GTA)^[18]。

在实验设置中, 为了达到基于CDAN^[13]方法的实验一致性, 本文采用CDAN方法相同的实验设置。全部源域数据带有标记信息, 全部目标域数据不含标记信息。使用平均精度作为实验的度量指标, 且最终结果基于3次随机实验的平均结果。在目标函数中, 主要包含两个超参数, 分别为全局最小熵损失权重 γ 和全局领域判别损失权重 λ 。在实验中, 设置全局最小熵损失权重 $\gamma = 1$, 即与分类损失占有相同的权重。全局领域判别损失权重 λ 的设置采用了渐进式的判别器训练策略, 由模型训练完成程度 p 而定, 计算公式为 $\lambda = 1 - \exp(-\delta p) / [1 + \exp(-\delta p)]$, 其中 p 的取值范围为 $0 \sim 1$, δ 为一个常数, 在实验中设置为 $\delta = 10$ 。

在算法实现中, 使用PyTorch进行实现, 并使用ResNet-50^[17]作为基础的特征提取模型。其中ResNet的初始参数为使用ImageNet预训练的模型参数。对于优化器的设置, 采用动量为0.9的SGD来更新参数, 同时采用与CDAN相同的优化策略: 学习率 η_p 由公式 $\eta_p = \eta_0 / (1 + \alpha p)^\beta$ 算得, 其中 p 指模型训练完成程度, 范围为 $0 \sim 1$ 。并设置 $\eta_0 = 0.01$, $\alpha = 10$ 和 $\beta = 0.75$ 的优化器参数组合。

3.2 实验结果

基于Office-31数据集的实验结果如表1所示。其中, 对比方法的结果均来自于原论文报告结果。相较于其他方法, 本文的TW-CDAN在多个迁移任务上具有更好的性能。在 $D \rightarrow A$ 和 $W \rightarrow A$ 两个任务上, TW-CDAN相较于CDAN有较大提升, 在平均精度上, 较于CDAN提升了1.4%。

基于Office-Home数据集的实验结果如表2所示。在Office-Home数据集中, 有更多的迁移任



图2 Office-Home数据集

务, 且类别数更多, 各领域之间的图像相似度也更低, 因此更具挑战。

从实验结果中可以得出, TW-CDAN方法在多数任务中, 都好于其他方法的结果。相较于CDAN方法, TW-CDAN在12个迁移任务中的9个得到了较高的提升, 并且在多个任务的平均精度上, TW-CDAN方法相较于CDAN方法提高了3%。

3.3 结果分析

迁移权重损失 CDAN将样本类别预测的熵作为条件熵来限制样本的迁移性。在优化过程中, 也有最小化类别预测熵的作用。为了更好地研究样本迁移权重对模型的影响, 本文对CDAN中的目标函数进行了拆分, 将条件熵作为损失函数的一部分。实验对比了从损失函数中拆分条件熵的CDAN(S), 仅在最小熵损失上应用迁移权重的TW-CDAN(E), 仅在分类损失上应用迁移权重的TW-CDAN(C)和同时在最小熵损失和分类损失应用迁移权重的

TW-CDAN。实验结果如表3所示, 由实验结果可知, CDAN与CDAN(S)无太大差距。其次, 相较于CDAN(S), TW-CDAN(E)与TW-CDAN(C)有部分提升, 而TW-CDAN相对于其他算法, 提升更为稳定。

算法收敛性 本文对Office-31数据集上的任务 $A \rightarrow W$ 在ResNet, DANN, CDAN和TW-CDAN算法上的训练曲线进行了可视化(图3), 进一步分析算法的稳定性和收敛性。ResNet方法对源域的拟合逐渐升高但对目标域的拟合逐渐降低, 且不够稳定。DANN, CDAN和TW-CDAN方法都有着非常稳定的结果, 收敛后无负迁移现象。TW-CDAN方法相对于其他方法性能更好, 结果也较为鲁棒。

适配可视化 本文对模型提取出的特征进行了可视化分析, 并对比了ResNet, DANN, CDAN和TW-CDAN方法(图4)。在基于ResNet方法的结果中, 源域和目标域在特征空间上的重合度较低, 因

表 1 Office-31数据集结果(使用平均精度进行评价)

方法	$A \rightarrow W$	$D \rightarrow W$	$W \rightarrow D$	$A \rightarrow D$	$D \rightarrow A$	$W \rightarrow A$	平均
ResNet-50 ^[17]	68.4	96.7	99.3	68.9	62.5	60.7	76.1
DAN ^[7]	80.5	97.1	99.6	78.6	63.6	62.8	80.4
RTN ^[8]	84.5	96.8	99.4	77.5	66.2	64.8	81.6
DANN ^[10]	82.0	96.9	99.1	79.7	68.2	67.4	82.2
ADDA ^[11]	86.2	96.2	98.4	77.8	69.5	68.9	82.9
JAN ^[18]	85.4	97.4	99.8	84.7	68.6	70.0	84.3
GTA ^[19]	89.5	97.9	99.8	87.7	72.8	71.4	86.5
CDAN ^[13]	93.1	98.6	100.0	92.9	71.0	69.3	87.5
TW-CDAN	94.9	99.2	100.0	94.0	72.7	72.5	88.9

表 2 Office-Home数据集结果(使用平均精度进行评价)

方法	$Ar \rightarrow Cl$	$Ar \rightarrow Pr$	$Ar \rightarrow Rw$	$Cl \rightarrow Ar$	$Cl \rightarrow Pr$	$Cl \rightarrow Rw$	$Pr \rightarrow Ar$	$Pr \rightarrow Cl$	$Pr \rightarrow Rw$	$Rw \rightarrow Ar$	$Rw \rightarrow Cl$	$Rw \rightarrow Pr$	平均
ResNet-50 ^[17]	34.9	50.0	58.0	37.4	41.9	46.2	38.5	31.2	60.4	53.9	41.2	59.9	46.1
DAN ^[7]	43.6	57.0	67.9	45.8	56.5	60.4	44.0	43.6	67.7	63.1	51.5	74.3	56.3
DANN ^[10]	45.6	59.3	70.1	47.0	58.5	60.9	46.1	43.7	68.5	63.2	51.8	76.8	57.6
JAN ^[18]	45.9	61.2	68.9	50.4	59.7	61.0	45.8	43.4	70.3	63.9	52.4	76.8	58.3
CDAN ^[13]	50.6	65.9	73.4	55.7	62.7	64.2	51.8	49.1	74.5	68.2	56.9	80.7	62.8
TW-CDAN	48.8	71.1	76.7	61.6	68.9	70.2	60.4	46.6	77.9	71.3	55.4	81.9	65.9

表 3 不同迁移权重设置在Office-31数据集结果(使用平均精度进行评价)

方法	$A \rightarrow W$	$D \rightarrow W$	$W \rightarrow D$	$A \rightarrow D$	$D \rightarrow A$	$W \rightarrow A$	平均
CDAN ^[13]	93.1	98.6	100.0	92.9	71.0	69.3	87.5
CDAN(S)	93.0	98.7	100.0	92.7	71.0	69.1	87.4
TW-CDAN(E)	93.7	98.8	100.0	93.4	71.5	71.3	88.1
TW-CDAN(C)	94.2	98.9	100.0	93.1	72.1	71.8	88.4
TW-CDAN	94.9	99.2	100.0	94.0	72.7	72.5	88.9

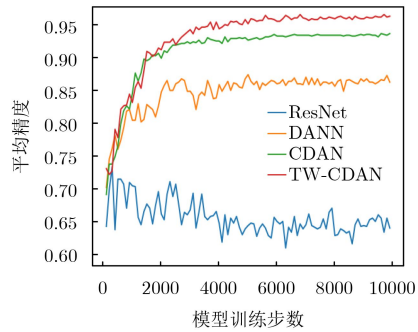


图3 算法收敛性对比实验

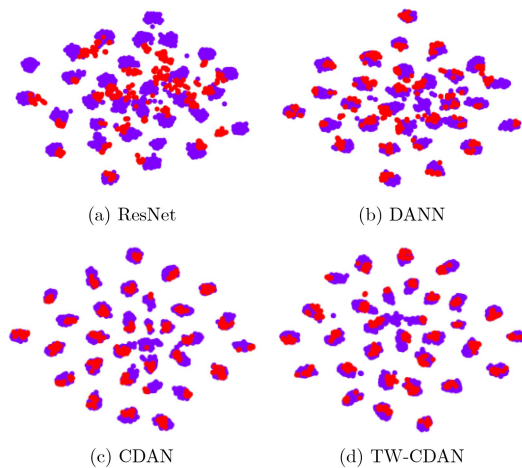


图4 T-SNE特征可视化

此源域和目标域依然存在分布偏移。在基于DANN的方法中,源域和目标域在特征空间上重合度较高,因此一定程度上解决了分布偏移的问题,但是DANN并没有把不同的类别样本分开,表明DANN虽然在整体的分布上进行了重合,但是并没有完全的拟合类别的分布。CDAN方法很好地解决了分布偏移问题,且不同类别的边界较为明显。相较于CDAN方法,TW-CDAN方法不仅延续了CDAN方法的优点,而且错分样本的数量更少,适应性能更强。

4 结束语

本文主要针对CDAN方法对于样本迁移性问题的考虑不足,提出基于迁移权重的条件对抗领域适应(TW-CDAN)方法,进一步挖掘样本的可迁移性对模型的影响。(1)将领域判别模型输出结果的熵作为样本迁移性的指标;(2)将迁移权重应用在最小熵损失和分类损失上。相比CDAN等方法,本文方法在Office-31数据集的6个迁移任务中有5个得到提升,平均精度提升了1.4%;在Office-Home数据集的12个迁移任务中有9个取得了提升,平均精度提升了3.1%。

本文方法在18个迁移任务中的14个任务得到了提升,表明本文方法仍然存在局限性。在Office-

Home数据集的其他领域向Clipart领域的迁移过程中,本文方法的性能都低于其他方法,主要原因为对于相似性较低的两个领域,模型的迁移性能被弱化。如何解决迁移能力弱化的问题,使模型更具普适性仍然是下一步工作的关键。

参考文献

- [1] YOSINSKI J, CLUNE J, BENGIO Y, *et al.* How transferable are features in deep neural networks?[C]. Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal, Canada, 2014: 3320-3328.
- [2] PAN S J and YANG Qiang. A survey on transfer learning[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2010, 22(10): 1345-1359. doi: [10.1109/TKDE.2009.191](https://doi.org/10.1109/TKDE.2009.191).
- [3] GEBRU T, HOFFMAN J, LI Feifei, *et al.* Fine-grained recognition in the wild: A multi-task domain adaptation approach[C]. Proceedings of IEEE International Conference on Computer Vision, Venice, Italy, 2017: 1358-1367. doi: [10.1109/ICCV.2017.151](https://doi.org/10.1109/ICCV.2017.151).
- [4] GLOROT X, BORDES A, and BENGIO Y. Domain adaptation for large-scale sentiment classification: A deep learning approach[C]. Proceedings of the 28th International Conference on Machine Learning, Bellevue, USA, 2011: 513-520.
- [5] WANG Mei and DENG Weihong. Deep visual domain adaptation: A survey[J]. *Neurocomputing*, 2018, 312: 135-153. doi: [10.1016/j.neucom.2018.05.083](https://doi.org/10.1016/j.neucom.2018.05.083).
- [6] GRETTON A, BORGWARDT K, RASCH M, *et al.* A kernel method for the two-sample-problem[C]. Proceedings of the 19th Conference on Neural Information Processing Systems, Vancouver, Canada, 2007: 513-520.
- [7] LONG Mingsheng, CAO Yue, WANG Jianmin, *et al.* Learning transferable features with deep adaptation networks[C]. Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 2015: 97-105.
- [8] LONG Mingsheng, ZHU Han, WANG Jianmin, *et al.* Deep transfer learning with joint adaptation networks[C]. Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 2017: 2208-2217.
- [9] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, *et al.* Generative adversarial nets[C]. Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal, Canada, 2014: 2672-2680.
- [10] GANIN Y, USTINOVA E, AJAKAN H, *et al.* Domain-adversarial training of neural networks[J]. *The Journal of Machine Learning Research*, 2016, 17(1): 2096-2030.
- [11] TZENG E, HOFFMAN J, SAENKO K, *et al.* Adversarial

- discriminative domain adaptation[C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 2962–2971. doi: [10.1109/CVPR.2017.316](https://doi.org/10.1109/CVPR.2017.316).
- [12] MIRZA M and OSINDERO S. Conditional generative adversarial nets[EB/OL]. <https://arxiv.org/abs/1411.1784>, 2014.
- [13] LONG Mingsheng, CAO Zhangjie, WANG Jianmin, *et al.* Conditional adversarial domain adaptation[C]. Proceedings of the 32nd Conference on Neural Information Processing Systems, Montréal, Canada, 2018: 1647–1657.
- [14] GRANDVALET Y and BENGIO Y. Semi-supervised learning by entropy minimization[C]. Proceedings of the 17th International Conference on Neural Information Processing Systems, Vancouver, Canada, 2004: 529–536.
- [15] SAENKO K, KULIS B, FRITZ M, *et al.* Adapting visual category models to new domains[C]. Proceedings of the 11th European Conference on Computer Vision, Heraklion, Greece, 2010: 213–226.
- [16] VENKATESWARA H, EUSEBIO J, CHAKRABORTY S, *et al.* Deep hashing network for unsupervised domain adaptation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 5385–5394. doi: [10.1109/CVPR.2017.572](https://doi.org/10.1109/CVPR.2017.572).
- [17] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al.* Deep residual learning for image recognition[C]. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770–778. doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [18] LONG Mingsheng, ZHU Han, WANG Jianmin, *et al.* Unsupervised domain adaptation with residual transfer networks[C]. Proceedings of the 30th Conference on Neural Information Processing Systems, Barcelona, Spain, 2016: 136–144.
- [19] SANKARANARAYANAN S, BALAJI Y, CASTILLO C D, *et al.* Generate to adapt: Aligning domains using generative adversarial networks[C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 8503–8512. doi: [10.1109/CVPR.2018.00887](https://doi.org/10.1109/CVPR.2018.00887).
- 王 进: 男, 1979年生, 教授, 研究方向为机器学习、数据挖掘。
- 王 科: 男, 1993年生, 硕士生, 研究方向为机器学习。
- 闵子剑: 男, 1995年生, 硕士生, 研究方向为机器学习。
- 孙开伟: 男, 1987年生, 讲师, 研究方向为机器学习、数据挖掘。
- 邓 欣: 男, 1981年生, 副教授, 研究方向为机器学习、认知计算。