

肖铠祥, 吴晓鸽, 冯永晋, 等. 基于域不变判别特征学习的深子域自适应方法[J]. 广东工业大学学报, 2025, 42(4): 59–70. doi: 10.12052/gdutxb.240096.
Xiao Kaixiang, Wu Xiaoling, Feng Yongjin, et al. Deep subdomain adaptation method based on domain invariant discriminant feature learning[J]. Journal of Guangdong University of Technology, 2025, 42(4): 59–70. doi: 10.12052/gdutxb.240096.

基于域不变判别特征学习的深子域自适应方法

肖铠祥¹, 吴晓鸽¹, 冯永晋¹, Hoon Heo²

(1. 广东工业大学 计算机学院, 广东 广州 510006; 2. 三星电机有限公司, 京畿道 水原市 16674)

摘要: 无监督域自适应旨在将源领域已标记数据的知识迁移到无标记的目标领域。基于域对齐的无监督域自适应方法通过最小化跨域特征分布差异来学习域不变特征, 忽略了目标域判别特征的学习, 造成学习到的域不变特征在目标域中判别性差, 不同类特征易混淆, 导致模型性能下降。为了解决域对齐过程中域不变特征在目标域判别性差的问题, 本文提出了一种基于域不变判别特征学习的深子域自适应方法(deep Subdomain Adaptation method based on domain invariant Discriminant Feature learning, DFSA), 结合深子域自适应方法与最小化类混淆约束来增加域不变特征的类间差异, 并添加一致性正则化来减少域不变特征的类内差异, 通过以上组合挖掘域不变判别特征, 以更好地保留域不变特征在目标域的判别性, 提高域自适应能力。在Office-31和Office-Home基准数据集上验证了DFSA的有效性, 结果表明DFSA有效提高了域自适应性能, 比现有的深子域自适应方法在2个基准数据集上的平均准确率分别提升了0.79个百分点和3.51个百分点。

关键词: 深度学习; 域自适应; 判别特征学习; 最小类混淆; 一致性约束

中图分类号: TP183

文献标志码: A

文章编号: 1007-7162(2025)04-0059-12

Deep Subdomain Adaptation Method Based on Domain Invariant Discriminant Feature Learning

Xiao Kaixiang¹, Wu Xiaoling¹, Feng Yongjin¹, Hoon Heo²

(1. School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China;

2. Samsung Electro-mechanics Co., Ltd., Suwon City 16674, Korea)

Abstract: Unsupervised domain adaptation aims to transfer knowledge from labeled data in the source domain to the unlabeled target domain. The unsupervised domain adaptation method based on domain alignment learns domain invariant features by minimizing the difference in cross domain feature distribution, ignoring the learning of discriminative features in the target domain, resulting in poor discriminability of the learned domain invariant features in the target domain, confusion of different types of features, and performance decrease of the model. To address the problem of poor discriminative power of domain invariant features in the target domain during domain alignment, this paper proposes a deep Subdomain Adaptive method based on domain invariant Discriminative Feature learning (DFSA). The deep subdomain adaptive method is combined with the constraints of minimizing class confusion to increase the inter-class differences of domain invariant features, and consistency regularization is added to reduce the intra-class differences of domain invariant features. By doing so, domain invariant discriminative features are mined to better preserve the discriminative power of domain invariant features in the target domain and improve domain adaptability. The effectiveness of DFSA was validated on the Office-31 and Office-Home benchmark datasets, and the results showed that DFSA effectively improves domain adaptation performance, with an average accuracy improvement of approximately 0.79 percentage points and 3.51 percentage points, respectively, when compared with existing deep subdomain adaptation methods on the two benchmark datasets.

Key words: deep learning; domain adaptation; discriminant feature learning; minimum class confusion; consistency constrain

收稿日期: 2024-07-30 录用日期: 2024-10-29 网络首发日期: 2025-06-03

基金项目: 广东省重点领域研发计划项目(2019B010139002); 广东省国际科技合作领域项目(2019A050513010)

作者简介: 肖铠祥(2000-), 男, 硕士研究生, 主要研究方向为领域自适应, E-mail: 2575437457@qq.com

通信作者: 吴晓鸽(1979-), 女, 副教授, 博士, 主要研究方向为人工智能、物联网; E-mail: xl.wu@gdut.edu.cn

近年来,随着深度学习的快速发展,各种机器学习问题得到了改善。传统有监督的深度学习任务需要依赖大量标记数据,然而有些数据的可靠标记需要消耗大量的人力和财力,比如图像分割、目标识别等。因此对于没有足够标记数据的机器学习任务,需要一种从未知标签数据中学习的方法。迁移学习是一种将源域知识迁移到目标域的机器学习方法,可以解决机器学习任务中获取标记困难以及模型泛化性差的问题。

无监督领域自适应是迁移学习的一个子领域,目的是利用源域已标记数据提高模型在无标记的目标域的泛化能力。通常无监督域自适应方法通过域对齐来学习域不变特征,实现跨域知识迁移。实现域对齐常见有2种方式,一种是基于差异度量的方式,另一种是基于对抗训练的方式。

基于差异度量的域自适应模型训练过程中通过最小化源域与目标域的特征差异度量指标来提取域不变特征。常见的应用在域自适应过程中的跨域特征差异度量指标有最大均值差异(Maximum Mean Discrepancy, MMD)^[1]、Wasserstein距离^[2]、最优传输距离^[3]等。Long等^[4]提出了联合自适应网络(Joint Adaptation Networks, JAN),它通过联合最大均值差异准则对齐跨域的多个域特定层的联合分布来学习迁移网络。Xie等^[5]提出了基于能量的主动域自适应算法(Energy-based Active Domain Adaptation, EADA),选择具有代表性的有限目标域样本进行域自适应,然后通过正则化项将目标域数据压缩的自由能对准源域,隐式地减小域间隙。Zhu等^[6]提出了一个新的多源无监督域自适应两阶段对齐框架,该框架不仅可以分别对齐每对源和目标域在多个特定特征空间中的分布,还可以利用特定域的决策边界对齐分类器的输出。Wang等^[7]提出选择性伪标记技术(Selective Pseudo-Labeling, SPL)与数据增强策略(norm-AE)组合的方法,直接在高维同构特征空间中学习源域和目标域的统一分类器,而不需要显式的域对齐,实现了有竞争力的域自适应性能。

在基于对抗训练的方式中,源域与目标域使用一个共同的特征提取器,然后将提取的特征作为域鉴别器的输入,当域鉴别器无法区分特征来自源域还是目标域时,则表示特征提取器提取到了域不变特征。Ganin等^[8]提出了一种基于域对抗的域自适应的方法(Domain Adversarial Training of Neural Networks, DANN),在网络结构中添加梯度反转层和域判别器,首次将对抗训练思想引入域自适应方法。Tzeng等^[9]结合了区别建模、联合权重共享和GAN损失提出了对抗性辨别域自适应(Adversarial Discriminative Domain Adaptation, ADDA)方法,使用生成器和域辨

别器组成对抗网络,在对抗训练过程中提取域不变特征,这是一个极大极小值博弈过程,存在收敛速度慢,计算资源消耗大的问题。Gao等^[10]改进了传统的具有区分信息的对抗性域自适应方法,提出了一种新的方法称为特征梯度分布对齐(Feature Gradient Distribution Alignment, FGDA),通过约束2个域的特征梯度来减少域之间的分布差异,提高了域自适应性能。

本文定义特征的判别性为不同类特征之间可区分的特性,判别性好表示不同类特征易区分,判别性差表示不同类特征易混淆不易区分。由线性判别分析(Linear Discriminant Analysis, LDA)^[11]可知更大的类间差异与更小的类内差异代表特征更好的判别性。域不变判别特征是指在不同域都能表现良好判别性的域不变特征。文献[12]通过实验证明无监督域自适应过程中学习到的特征的域不变性和判别性是相辅相成的,而以上域对齐的方法仅学习能最小化跨域特征差异的域不变特征,没有考虑其在目标域上的判别性。源域因为使用标记数据进行训练,所以提取的域不变特征能展现较好的判别性,但未标记的目标域通过域对齐提取的域不变特征往往判别性较差,影响域自适应的性能。如图1(a)~(b)所示,其中有填充图形为源域样本,无填充图形表示目标域样本,不同形状代表不同类别。图1(a)表示源域与目标域分别进行有监督训练后的特征分布,实线表示源域决策边界,虚线表示目标域决策边界,体现了有监督训练的特征有很好的判别性。与图1(a)不同,图1(b)展示的无监督域自适应过程中,无标记的目标域依赖与源域的特征分布对齐来提取域不变特征,对比图1(a)来看,目标域提取的特征容易受源域特征分布影响(向最近的源域特征靠近),导致判别性较差(空心三角形与空心圆分布紧凑不易区分)。

本文结合深子域自适应方法和目标域正则化约束,提出了基于域不变判别特征学习的深子域自适应方法(deep Subdomain Adaptation method based on domain invariant Discriminant Feature learning, DFSA),改善了域不变特征在目标域判别性较差的问题。在子域自适应过程中,最小化目标域预测的类混淆,增大目标域数据的类间差异,并使用目标域数据的强弱数据增强伪标签一致性约束减少目标域数据的类内差异,以此挖掘域不变判别特征,提高域不变特征在目标域的判别性,实现如图1(c)所示的特征分布,提取的域不变特征在源域和目标域上都具有较好的判别性(不同形状的有填充图和无填充图都能很好地被决策边界划分)。

本文的贡献主要包括以下4个方面:

(1) 结合深子域自适应方法与类混淆最小化约

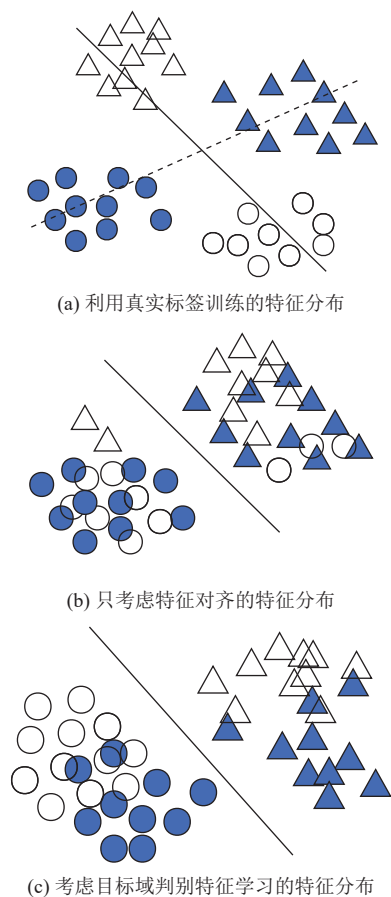


图1 不同方法的特征分布图

Fig.1 Feature distribution maps of different methods

束,增加了域不变特征在目标域类间的差异,提高了域不变特征在目标域的判别性。

(2) 使用目标域数据的强弱数据增强伪标签一致性约束减少了目标域类内差异,确保目标域决策边界位于低密度区域,提高了目标域分类准确度,增加了决策边界的抗扰动性。

(3) 从域自适应理论的角度出发,证明了域不变特征更好的判别性将有效提高域自适应性能。

(4) 本文在基于无监督域自适应的图像分类任务上验证了所提出方法的有效性,并在基准公共数据集上对比了之前的域自适应方法,结果表明本文的方法有效提高了模型的域自适应能力。

1 相关工作

域自适应旨在学习源域知识,将其运用在不同分布但相似的目标领域。根据目标领域数据标记的情况分为无监督域自适应和半监督域自适应。本文主要针对目标领域没有标记数据的无监督域自适应进行研究。常见的无监督域自适应方法通过域对齐的方式实现跨域特征提取,然而这类方法常常会因为过度拟合源域分布而忽略目标域特征的判别性,

导致域不变特征在目标域决策边界模糊,模型分类不准确。为了改善域对齐过程中目标域特征被混淆的情况,一系列基于模型概率分布预测和伪标签的域自适应方法被提出^[13-19]。

Dan等^[13]提出了一种信任感知条件对抗域自适应方法(Trust-aware Conditional Adversarial Domain Adaptation, TCADA),基于高斯均匀混合建模的后验概率,提出了一种置信度引导的对齐策略,以促进类别分布的精确对齐,并加速共享特征的学习。此外,引入了一种新的基于最优传输的策略来对齐特征规范,并促进共享特征信息变得更加丰富,还设计了一个混合信息引导的熵正则化项,以促进深层特征远离决策边界,有效地促进了条件域对齐。Zhu等^[14]提出了深子域自适应网络(Deep Subdomain Adaptation Network, DSAN),该网络通过最小化局部最大均值差异(Local Maximum Mean Discrepancy, LMMD)值来对齐子域分布,从而学习到域不变特征,很好地改善了全局域对齐导致的目标域数据混淆的情况。Zhao等^[16]提出了一种深度模糊域自适应方法(Deep Fuzzy Domain Adaptation, DFDA),该方法考虑到了源域异常类对子域自适应的影响,通过对源域和目标域中相同类别的样本分配不同的权重,增强了子域自适应能力。杜社林等^[17]针对子域自适应中未充分考虑错误伪标签在训练过程中产生的负面影响和决策边界处预测准确率较低的问题,提出基于加权分类损失和核范数的域自适应模型(Domain Adaptation Model Based on Weighted Classification Loss and Nuclear-norm, WCLN)。以上基于伪标签的域自适应方法,考虑源域与目标域的条件分布对齐来改善域对齐过程中目标域特征判别性差的问题,然而分类器受源域有标签数据的影响,会产生更加符合源域决策边界的伪标签,且基于条件分布的域对齐仍依赖源域的特征分布,没有考虑目标域判别特征的学习,不能充分发挥基于条件分布对齐的域自适应方法效能。Luo等^[20]提出基于判别流形传播的无监督域自适应方法(Discriminative Manifold Propagation, DMP),该方法关注源域类间可分性和目标域类内紧凑性,所提出的方法在目标域上改善了迁移特征的判别性,但是没有充分考虑目标域类间差异。Wang等^[15]提出具有标签分配优先权的信息最大化自适应网络(Information Maximization Adaptation Network with Label Distribution Priors, IMANLDP),通过最大化源互信息和对齐跨域特征分布来学习判别特征表示,为无监督域自适应方法提供了新的思路。

本文方法结合伪标签和目标域正则化约束,在

子域自适应过程中同时考虑域不变特征在目标域的分类区分性和类内紧凑性来充分学习目标域的判别特征,以提高域不变特征在目标域的判别性,同时增加模型决策边界的抗扰动性。

2 基于域不变判别特征学习的深子域自适应模型

2.1 无监督域自适应问题描述

在无监督域自适应问题中,设源域 $D_s = \{(\mathbf{x}_i^s, \mathbf{y}_i^s)\}_{i=1}^{n_s}$ 是有标签的数据集,其中 $n_s \in \mathbb{N}$ (\mathbb{N} 表示自然数集合)代表源域标签数据的个数, \mathbf{x}_i^s 是源域样本,它的标签 $\mathbf{y}_i^s \in \mathbb{R}^{1 \times C}$ (\mathbb{R} 表示实数集合)是一个独热向量,即 $y_{ij}^s = 1$ 表示源域样本 \mathbf{x}_i^s 属于第 j 类,目标域 $D_t = \{(\mathbf{x}_i^t)\}_{i=1}^{n_t}$ 是无标签的数据集, \mathbf{x}_i^t 是目标域样本,其中 $n_t \in \mathbb{N}$ 代表无标签数据的个数。 D_s 与 D_t 是分别来自于不同数据分布 p, q 的样本数据集合。无监督域自适应所要实现的目标是利用源域 D_s 中的有标签数据的知

识,提高模型在目标域 D_t 的泛化能力。

2.2 模型介绍

本文提出的基于域不变判别特征学习的深子域自适应模型的总体架构如图2所示。模型训练过程中首先对输入的目标域数据进行数据增强预处理,获取目标域数据的不同视图,然后将源域数据与不同视图的目标域数据输入到特征提取网络。模型使用预训练的ResNet50特征提取网络提取数据特征,提取到不同域特征后将其输入分类器得到模型预测,然后利用跨域特征和目标域模型预测计算跨域特征的子域分布差异损失。利用源域数据与模型预测计算交叉熵损失学习源域知识,而目标域弱增强视图的模型预测被用来计算类混淆损失,目标域不同视图的模型预测则被用来计算一致性损失。对以上各部分损失求和作为模型总体损失,最后利用小批量随机梯度下降算法更新模型参数寻找最小化模型总体损失的最优解。

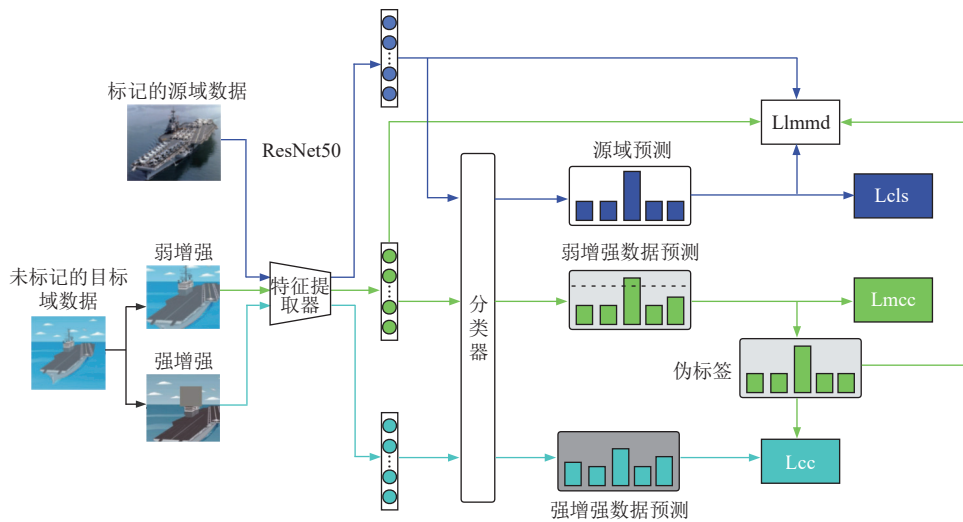


图2 DFSA模型架构

Fig.2 DFSA model architecture

2.2.1 子域分布对齐

使用已标记数据训练的模型在独立同分布的数据集上往往能表现极佳的性能,然而在无监督域自适应任务中,进行有监督训练的源域数据与无标签的目标域数据的特征分布存在差异,导致模型在目标域的泛化性不佳。MMD^[1]用于衡量2个分布之间的差异。分别来自分布 p, q 的样本 $\mathbf{X}_p, \mathbf{X}_q$ 的MMD计算表达式如式(1)所示。

$$M(p, q) = \|\mathbb{E}_p[\varphi(\mathbf{X}_p)] - \mathbb{E}_q[\varphi(\mathbf{X}_q)]\|_H^2 \quad (1)$$

其中 $\mathbb{E}[\cdot]$ 表示期望, $\varphi(\cdot)$ 函数将样本映射到再生核希尔伯特空间 H (Reproducing Kernel Hilbert Space,

RKHS),且根据MMD理论可知当 $M(p, q) = 0$ 时,表示分布 p 与 q 完全一致。基于MMD的域自适应方法^[1]通过减少源域与目标域的分布差异来学习域不变特征,从而提高模型在目标域的泛化能力,然而这种方法仅考虑全局域的分布对齐,没有考虑到不同域中更加细粒度的类信息,导致模型性能不佳。DSAN^[14]对不同类别分别进行分布对齐,并证明其所提出的基于LMMD的策略能在类级别实现特征分布对齐,一定程度上提高了域不变特征的判别性,提升了模型性能,但是并没有充分利用目标域特征来学习域不变判别特征。本文在DSAN的基础上,进一步挖掘目标域的判别特征,提高域不变特征的判别性。

本文使用LMMD表示源域与目标域分布差异,其值越小表示源域与目标域特征分布越接近,其无偏估计值 $\mathcal{L}_{\text{lmmd}}^l$ 表达式如式(2)所示。

$$\begin{aligned} \mathcal{L}_{\text{lmmd}}^l = & \frac{1}{C} \sum_{c=1}^C \left\| \sum_{\mathbf{x}_i^s \in D_s} \mathbf{Q}_i^{sc} \varphi(\mathbf{x}_i^s) - \sum_{\mathbf{x}_j^t \in D_t} \mathbf{Q}_j^{tc} \varphi(\mathbf{x}_j^t) \right\|_H^2 = \\ & \frac{1}{C} \left(\sum_{c=1}^C \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} \mathbf{Q}_i^{sc} \mathbf{Q}_j^{tc} k(\mathbf{a}_i^{sl}, \mathbf{a}_j^{tl}) + \right. \\ & \left. \sum_{c=1}^C \sum_{i=1}^{n_t} \sum_{j=1}^{n_s} \mathbf{Q}_i^{tc} \mathbf{Q}_j^{sc} k(\mathbf{a}_i^{tl}, \mathbf{a}_j^{sl}) - \right. \\ & \left. 2 \sum_{c=1}^C \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} \mathbf{Q}_i^{sc} \mathbf{Q}_j^{tc} k(\mathbf{a}_i^{sl}, \mathbf{a}_j^{tl}) \right) \end{aligned} \quad (2)$$

其中 $\varphi(\cdot)$ 函数将样本映射到再生核希尔伯特空间 H , $k(\cdot)$ 表示核函数,用于计算通过 $\varphi(\cdot)$ 函数映射的特征向量之间的点积, \mathbf{a}^{sl} 与 \mathbf{a}^{tl} 分别表示源域样本和目标域样本在第 l 层激活层的输出向量,其中 $l \in L$,且 $L = \{1, 2, 3, \dots, |L|\}$, $|L|$ 表示集合 L 的势,同时也表示网络中激活层的总层数, \mathbf{Q}_i^{sc} 与 \mathbf{Q}_j^{tc} 分别表示源域样本 \mathbf{x}_i^s 与目标域样本 \mathbf{x}_j^t 属于类别 c 的归一化权重, $c \in C$,且 $C = \{1, 2, 3, \dots, |C|\}$, $|C|$ 为集合 C 的势,同时也表示分类任务的类别总数,其计算过程如式(3)所示。

$$\mathbf{Q}_i^c = \frac{y_{ic}}{\sum_{(\mathbf{x}_i, y_i) \in D} y_{jc}} \quad (3)$$

在源域中 $y_{ic} \in \{0, 1\}$,而在目标域无标签样本中 y_{ic} 表示模型以目标域样本 \mathbf{x}_i^t 作为输入生成的类别概率分布 $\hat{\mathbf{y}}_i$ 的第 c 项的值。

2.2.2 类混淆约束

基于LMMD的域自适应方法在特征分布对齐过程中仅考虑子域分布对齐,而忽略目标域不同类特征之间的关系,导致域不变特征在目标域判别性差,不同类特征易混淆,造成域适应模型性能下降。线性判别分析^[11]表明特征之间更大的类间差异将体现特征更好的判别性,本文考虑增加目标域特征的类间差异来提高域不变特征在目标域的判别性,而直接显式地增大不同类之间的欧氏距离计算复杂度高,且需要的计算资源大。类混淆矩阵^[21]是一种基于模型预测矩阵计算不同类之间混淆程度的方法,且被证明能够有效增大类间差异,减少不同类之间的混淆。本文使用最小化类混淆矩阵约束不仅能隐式增大目标域特征类间差异,而且能充分利用 $\mathcal{L}_{\text{lmmd}}^l$ 计算过程中产生的模型预测,提高资源利用率,节省计算资源。本文用 $\mathbf{M} \in \mathbb{R}^{|C| \times |C|}$ 表示类别混淆矩阵, $\hat{\mathbf{Y}}^t \in \mathbb{R}^{B \times |C|}$ 为小批量目标域样本的模型预测矩阵,为了便于书写,后文用 $\hat{\mathbf{Y}}$ 表示 $\hat{\mathbf{Y}}^t$,其中 B 是目标域数据的批量大小, $|C|$ 表示分类任务的类别总数,类别混淆矩阵计算表达式如式(4)所示。

$$\mathbf{M}_{ij} = \hat{\mathbf{Y}}_i^T \hat{\mathbf{Y}}_j \quad (4)$$

其中 $\hat{\mathbf{Y}}_i \in \mathbb{R}^{B \times 1}$ 表示模型以批量中不同样本作为输入获得的每个样本属于类别 i 的权重值所组成的向量,称作类别 i 的类向量;同理可得 $\hat{\mathbf{Y}}_j$ 是类别 j 的类向量, $\hat{\mathbf{Y}}_i^T$ 是 $\hat{\mathbf{Y}}_i$ 的转置; \mathbf{M}_{ij} 可以用来衡量类别 i 与类别 j 的混淆程度。其次为了避免模型预测过度自信带来的负面影响,采用温度重缩放,目标域数据模型概率预测计算表达式如式(5)所示。

$$\hat{\mathbf{Y}}_{ij} = \frac{\exp(\mathbf{Z}_{ij}/T)}{\sum_{j'=1}^{|C|} \exp(\mathbf{Z}_{ij'}/T)} \quad (5)$$

其中 $\hat{\mathbf{Y}}_{ij}$ 表示批量样本中第 i 个实例属于 j 类的概率的重缩放值, \mathbf{Z}_{ij} 表示分类器输出, T 是概率重缩放的温度超参数。在计算类混淆的过程中,不同样本对类混淆的量化重要程度不同,相比于接近均匀分布的分类器输出,显然拥有几个峰值的输出更能表明分类器在几个类别之间的不确定性,更加适合体现类混淆。本文利用信息论中的熵函数 $H(p) \triangleq -\mathbb{E}_p \ln p$ 来度量这种不确定性,并为类混淆矩阵的计算引入了不确定性加权矩阵 $\mathbf{W} \in \mathbb{R}^{B \times B}$,这是一个对角矩阵,其计算表达式如式(6)所示。

$$\mathbf{W}_{ii} = \frac{B(\exp(-H(\hat{\mathbf{Y}}_i)) + 1)}{\sum_{i'=1}^B (\exp(-H(\hat{\mathbf{Y}}_{i'})) + 1)} \quad (6)$$

其中 \mathbf{W}_{ii} 表示第 i 个样本对类混淆建模的重要性权重, $\hat{\mathbf{Y}}_i \in \mathbb{R}^{1 \times |C|}$ 表示第 i 个样本的模型预测, $H(\hat{\mathbf{Y}}_i)$ 的计算表达式如式(7)所示。

$$H(\hat{\mathbf{Y}}_i) = -\sum_{c=1}^{|C|} \hat{\mathbf{Y}}_{ic} \ln \hat{\mathbf{Y}}_{ic} \quad (7)$$

其中 $H(\hat{\mathbf{Y}}_i)$ 表示以样本 \mathbf{x}_i^t 作为模型输入生成的模型预测的不确定性程度,重加权后的类混淆矩阵计算表达式如式(8)所示。

$$\mathbf{M}_{ij} = \hat{\mathbf{Y}}_i^T \mathbf{W} \hat{\mathbf{Y}}_j \quad (8)$$

由于本文方法采用小批量随机梯度下降优化算法,为解决训练过程中不同批次中的类别不平衡问题,对类混淆矩阵进行了归一化处理,如式(9)所示。

$$\tilde{\mathbf{M}}_{ij} = \frac{\mathbf{M}_{ij}}{\sum_{j'=1}^{|C|} \mathbf{M}_{ij'}} \quad (9)$$

其中 $\tilde{\mathbf{M}}_{ij}$ 表示归一化后类混淆的计算值,对规范化后的类混淆矩阵元素求平均值,作为目标域的类混淆损失 \mathcal{L}_{mcc} ,如式(10)所示。

$$\mathcal{L}_{\text{mcc}}(\hat{\mathbf{Y}}) = \frac{1}{|C|} \sum_{i=1}^{|C|} \sum_{j \neq i}^{|C|} |\tilde{\mathbf{M}}_{ij}| \quad (10)$$

2.2.3 一致性约束

由线性判别分析^[11]对特征判别性的描述可知,更大的类间差异和更小的类内差异共同决定特征好的判别性。为了进一步提高跨域特征的判别性,考虑减少域不变特征的类内差异。域自适应过程本质上是提高模型泛化能力的过程,数据增强方式通过获取数据更多的视图,增加训练数据的规模和多样性,是一种能有效增加模型泛化能力的方法^[22],且文献^[23]中证明对目标域样本进行强弱数据增强一致性约束可以获得同类样本更小的类内方差。为了降低目标域特征的类内差异,提高跨域特征的判别性和模型在目标域的泛化能力,本文使用强弱数据增强一致性约束^[22-23]。与文献^[23]基于不确定性选择的强弱数据增强一致性约束不同,本文沿用文献^[22]基于置信度选择的强弱数据增强一致性约束,计算更加简便。具体做法为最小化强弱数据增强一致性损失,确保高置信度目标域样本的强数据增强(Rand Augment^[24])与其弱数据增强(随机翻转和裁剪)保持一致的伪标签,使得同一类别的样本特征更加紧凑,实现域不变特征更好的判别性,以提高域自适应模型的分类准确率。强弱数据增强一致性约束的损失函数 \mathcal{L}_{cc} 如式(11)所示。

$$\mathcal{L}_{cc} = \frac{1}{B} \sum_{b=1}^B I(\max(\mathbf{p}(\mathbf{y} | \mathbf{a}(\mathbf{x}'_b))) \geq \tau) J(\mathbf{p}_m(\mathbf{y} | \mathbf{a}(\mathbf{x}'_b)), \mathbf{p}(\mathbf{y} | \mathbf{A}(\mathbf{x}'_b))) \quad (11)$$

式(11)中, $\mathbf{a}(\mathbf{x})$ 表示对样本 \mathbf{x} 进行弱数据增强,而 $\mathbf{A}(\mathbf{x})$ 表示对样本 \mathbf{x} 进行强数据增强, \mathbf{x}'_b 表示目标域小批量数据中第 b 个样本, $\mathbf{p}_m(\mathbf{y} | \mathbf{a}(\mathbf{x}'_b)) = \max(\mathbf{p}(\mathbf{y} | \mathbf{a}(\mathbf{x}'_b)))$ 表示样本 \mathbf{x}'_b 进行弱数据增强后的伪标签, $\mathbf{p}(\mathbf{y} | \mathbf{a}(\mathbf{x}'_b))$ 表示模型以进行弱数据增强的 \mathbf{x}'_b 为输入生成的类别概率分布, $\mathbf{p}(\mathbf{y} | \mathbf{A}(\mathbf{x}'_b))$ 则是样本进行强数据增强后的类别概率分布, τ 表示置信度阈值, $I(\cdot)$ 表示指示函数,当 $\max(\mathbf{p}(\mathbf{y} | \mathbf{a}(\mathbf{x}'_b))) \geq \tau$ 时取值为1,否则为0。 $J(\cdot, \cdot)$ 表示交叉熵损失函数。

综上所述,本文域自适应过程中总体的目标函数如式(12)所示。

$$\text{loss} = \mathcal{L}_{cls} + \mu \sum_{l \in L} \mathcal{L}_{lmm}^l + \mathcal{L}_{mcc} + \mathcal{L}_{cc} \quad (12)$$

其中 \mathcal{L}_{cls} 表示源域有标签训练的交叉熵损失,其计算表达式如式(13)所示。

$$\mathcal{L}_{cls} = \frac{1}{n_s} \sum_{i=1}^{n_s} J(\hat{\mathbf{y}}_i^s, \mathbf{y}_i^s) \quad (13)$$

其中 $\hat{\mathbf{y}}_i^s = \mathbf{p}(\mathbf{y} | \mathbf{a}(\mathbf{x}_i^s))$ 表示模型以样本 \mathbf{x}_i^s 的弱数据增

强作为输入生成的类别概率分布, \mathbf{y}_i^s 为 \mathbf{x}_i^s 的标签, μ 为子域适应因子, $J(\cdot, \cdot)$ 表示交叉熵损失函数。

为了更加清晰地理解模型处理流程,本文提出的DFSA方法的算法过程如算法1所示。

算法1 DFSA算法

- 1) 输入: $\mathbf{X}^s, \mathbf{X}^t$: 源域和目标域样本; \mathbf{Y}^s : 源域标签。
- 2) 初始化:
 - (a) 利用 $\mathbf{X}^s, \mathbf{X}^t, \mathbf{Y}^s$ 训练获取分类器输出 $\hat{\mathbf{Y}}^s, \hat{\mathbf{Y}}^t$, 强数据增强 \mathbf{X}'_A , 弱数据增强 \mathbf{X}'_a , 激活层 l 的输出 $\mathbf{a}^{sl}, \mathbf{a}^{tl}$;
 - (b) 根据式(13), 利用 $\mathbf{X}^s, \mathbf{Y}^s, \hat{\mathbf{Y}}^s$ 计算 \mathcal{L}_{cls} 损失;
 - (c) 根据式(2), 利用 $\hat{\mathbf{Y}}^s, \hat{\mathbf{Y}}^t, \mathbf{a}^{sl}, \mathbf{a}^{tl}$ 计算 \mathcal{L}_{lmm}^l 损失;
 - (d) 根据式(4)~(10), 利用 $\hat{\mathbf{Y}}^t$ 计算 \mathcal{L}_{mcc} 损失;
 - (e) 根据式(11), 利用 $\mathbf{X}'_A, \mathbf{X}'_a$ 计算 \mathcal{L}_{cc} 损失;
 - (f) 根据式(12)计算目标函数总损失;
 - (g) 进行反向传播更新模型参数。
- 3) 直至所有目标域数据都参与训练。
- 4) 输出: 使得目标函数总损失最小的模型参数。

3 理论分析

本节利用域自适应理论^[25]分析DFSA的有效性。假设 h 的误差定义为 $\epsilon(h) = \epsilon(h, f) = E_{(x,y) \sim D} |h(x) - f(x)|$, 其中 h 表示假设函数, f 表示标签函数, D 表示 x 的分布, y 表示 x 的标签。

本节借助文献^[25]中目标域假设函数误差限制理论,分析本文方法对目标域假设函数误差的影响,以反映DFSA的有效性,具体理论表述如定理1所示。

定理1 设 \mathcal{H} 为假设空间, S 和 T 分别为源分布和目标分布,则目标域的假设函数误差 $\epsilon_T(h)$ 满足式(14)

$$\epsilon_T(h) \leq \epsilon_S(h) + \frac{1}{2} d_{\mathcal{H}(\Delta \mathcal{H})}(S, T) + \lambda \quad (14)$$

其中 $\epsilon_S(h)$ 表示源域的假设函数误差, $d_{\mathcal{H}(\Delta \mathcal{H})}(S, T)$ 是用来度量 S, T 分布之间差异的域散度。上文提到的基于MMD^[1]和对抗训练^[8]的域自适应方法,以及本文的子域自适应过程就是通过最小化域散度来实现更小的目标域假设函数误差上界。 λ 是一个共享假设函数的共享误差。

定义1 设理想假设

$$h^* = \underset{h \in \mathcal{H}}{\operatorname{argmin}} \epsilon_S(h) + \epsilon_T(h)$$

则 λ 的定义如式(15)所示。

$$\lambda = \epsilon_T(h^*) + \epsilon_S(h^*) \quad (15)$$

以往的方法通常将 λ 忽略不计,本文将目标域的伪标签函数引入假设函数误差估计,分析目标域判别特征给目标域假设函数误差带来的影响。

从文献[25]中可知,任意标签函数 f_1, f_2, f_3 满足式(16)。

$$\epsilon(f_1, f_2) \leq \epsilon(f_1, f_3) + \epsilon(f_3, f_2) \quad (16)$$

所以,可得式(17)。

$$\begin{aligned} \lambda = \epsilon_S(h^*, f_S) + \epsilon_T(h^*, f_T) &\leq \epsilon_S(h^*, f_S) + \epsilon_T(h^*, f_S) + \\ \epsilon_T(f_S, f_T) &\leq \epsilon_S(h^*, f_S) + \epsilon_T(h^*, f_S) + \epsilon_T(f_S, f_T) + \\ \epsilon_T(f_T, f_T) &\quad (17) \end{aligned}$$

其中 f_T 为伪标记函数,第1项 $\epsilon_S(h^*, f_S)$ 和第2项 $\epsilon_T(h^*, f_S)$ 表达式表示理想假设函数 h^* 和源域标记函数 f_S 分别在源域和目标域上的误差。由于 h^* 使用标记的源域数据训练,所以 h^* 与 f_S 将呈现较大的一致性,实现非常小的误差。第3项 $\epsilon_T(f_S, f_T)$ 表示在目标域上源域标记函数与目标域伪标记函数的误差。由于类别属于 c 的源样本将由源标记函数 f_S 用标记 c 来预测,如果类别 c 中的目标样本的特征与类别 c 中的源特征足够相似,源标签函数与目标域伪标签函数的预测将保持一致,则 $\epsilon_T(f_S, f_T)$ 将保持较小值,本文的子域自适应过程就是通过学习域不变特征,实现更小的 $\epsilon_T(f_S, f_T)$ 。第4项 $\epsilon_T(f_T, f_T)$ 表示目标域标记函数与伪标记函数之间的误差。本文通过类混淆约束和一致性约束,挖掘目标域的判别特征,实现目标域数据的可区分结构,避免目标域模糊决策边界导致的类别混淆,以提高目标域伪标签预测准确率,可以取得更小的 $\epsilon_T(f_T, f_T)$ 损失值。

无监督域自适应的目标是估计目标域假设函数误差 $\epsilon_T(h)$,然后将其优化到一个较低的值。综上所述,本文提出的基于域不变判别特征学习的深子域自适应方法,在理论分析过程中,可以通过缩小目标域假设函数误差的上界取得更小的共享误差 λ ,理论上可以实现更小的目标域假设函数误差,提高域自适应能力。

4 实验及性能评价

本节进行了一系列实验,在域自适应基线数据集Office-31^[26]、Office-Home^[27]上将DFSA与域自适应的其他基线方法进行比较,评估了本文方法的有效性。

4.1 数据集介绍

Office-31数据集包含3个不同的数据领域,分别是amazon(A)、dslr(D)和webcam(W),每个域中都包含相同的31个类别的办公环境图片。实验中使用Office-31构建了6个迁移任务: $A \rightarrow D$ 、 $A \rightarrow W$ 、 $D \rightarrow W$ 、 $D \rightarrow A$ 、 $W \rightarrow A$ 、 $W \rightarrow D$,其中 $A \rightarrow D$ 表示以amazon(A)为源域,以dslr(D)为目标域的域自适应任务。

Office-Home是比Office-31更复杂的领域自适应基准数据集,其中包含了4个不同的领域: Artistic images (A)、Clip Art (C)、Product images (P)和Real-World images (R),每个域都包含65个不同类别的图片,给域自适应方法提出了更大的挑战。实验中使用Office-Home数据集构建了12个迁移任务: $A \rightarrow C$ 、 $A \rightarrow P$ 、 $A \rightarrow R$ 、 $C \rightarrow A$ 、 $C \rightarrow P$ 、 $C \rightarrow R$ 、 $P \rightarrow A$ 、 $P \rightarrow C$ 、 $P \rightarrow R$ 、 $R \rightarrow A$ 、 $R \rightarrow C$ 、 $R \rightarrow P$ 。图3是Office-Home数据集中的部分数据展示图。



图3 Office-Home部分数据集

Fig.3 Some samples of the Office-Home dataset

4.2 实验设置

基于以上数据集,将所提出的DFSA与深度学习方法和领域自适应方法进行了比较(包括JAN^[4]、norm-AE-SPL^[7]、DANN^[8]、ADDA^[9]、DSAN^[14]、DMP^[20]、IMANLDP^[15]、ResNet^[28]、CDAN^[29])。实验环境配置如表1所示。

本文采用batch_size为32的小批量随机梯度下降算法进行训练,且基于以往论文方法的实验设置,混淆矩阵中温度重加权重 $T=2.5$ ^[21],一致性约束中的置信度阈值 $\tau=0.95$ ^[22]。为了节省计算成本,实验过程中的学习率采用动态衰减方式: $\eta_t = \eta_0 / (1 + 10\theta)^{0.75}$,初始学习率 η_0 设置为0.001,为了在训练的早期阶段抑制噪声激活,不是固定适应因子 μ ,而是通过渐进的训练进度将其从0逐渐改变为1,表达式如式(18)所示。

$$\mu_\theta = \frac{2}{\exp(-\gamma\theta)} - 1 \quad (18)$$

其中 $\gamma=10$ 在整个实验中是固定的^[14], θ 为从0~1线性变化的训练进度。

4.3 对比实验结果与分析

由于源域进行有监督训练,提取的域不变特征在源域上的可区分性很容易实现,所以本文实验部分主要探讨域不变特征在目标域上的分布特点,如果没有明确说明,下文的域不变特征都是指在目标域提取的域不变特征。

由线性判别分析^[11]可知更大的类间差异与更小的类内差异共同体现特征更好的判别性。本文计算

表1 实验环境配置
Table 1 Experimental environment configuration

操作系统	处理器	显卡	显存	Pytorch版本
Ubuntu 9.4.0	AMD EPYC 7551P 32-Core Processor	NVIDIA RTX A4000	16G	1.11.0+cu13

特征空间中类间方差和类内方差的比值,将其作为判别性指标 I_D ,衡量特征的判别性,更大的判别性指标值意味着特征更好的判别性,计算表达式如式(19)所示。

$$I_D = \exp\left(\frac{D_{\text{inter}}}{D_{\text{intra}}}\right),$$

$$D_{\text{inter}} = \sum_{i=1}^{|C|} \sum_{j \neq i}^{|C|} (\bar{X}_{ti} - \bar{X}_{tj})^2,$$

$$D_{\text{intra}} = \sum_{c=1}^{|C|} \sum_{x^t \in X_c^t} (x^t - \bar{X}_{tc})^2 \quad (19)$$

其中 $|C|$ 表示类别总数, D_{inter} 用于衡量域不变特征类间差异, D_{intra} 用于衡量域不变特征类内差异, \bar{X}_{ti} 表示目标域中所有伪标签为 i 的样本特征向量的均值, X_c^t 表示目标域中所有伪标签为 c 的特征向量集合。

图4展示了不同方法在Office-Home数据集 $R \rightarrow P$ 任务中提取的域不变特征判别性指标计算值的变化情况。其中ResNet为基于resnet50的网络模型,没有进行域自适应过程,直接利用源域已标记数据对模型进行有监督训练,由于没有针对目标域判别特征的学习过程,所以ResNet方法学习的特征在目标域上判别性指标一直保持较低值。JAN方法试图通过挖掘使得源域与目标域分布差异最小的域不变特征将源域知识应用到目标域中,特征判别性指标有一定提升,但这是一种全局特征分布对齐方法,没有考虑域不变特征在目标域的判别性,忽略了目标域判别特征的学习,所以在图4的JAN方法中呈现的判别性指标保持较低值。DSAN考虑到目标域特征的区分性,利用伪标签对不同子域分别进行域对齐,在域自适应过程中考虑到了目标域特征的区分性,虽

然一定程度上提高了特征的判别性指标,但是目标域判别特征的学习依赖与源域特征的对齐,并没有单独针对目标域判别特征的学习过程,没有充分提取目标域的判别特征。DFSA在子域对齐的过程中考虑目标域判别特征的学习,实现了域不变特征最高的判别性指标值,表明了DFSA有效地提高了域不变特征在目标域的判别性。

图5是不同方法在Office-Home数据集上的域自适应任务 $R \rightarrow P$ 中使用t-SNE学习到的域不变特征分布的可视化图,可以更加直观地理解DFSA对域不变特征在目标域的判别性贡献。与其他方法相比,DFSA方法学习到的域不变特征呈现更好的判别性。

图6是在Office-Home数据集 $C \rightarrow R$ 任务上的最小误差值的统计图,其中共享误差为理论分析中的 λ ,而目标域理想假设函数误差为共享误差中的 $\epsilon_T(h^*)$ 。如图6所示,DFSA方法实现了最小的目标域理想假设函数误差 $\epsilon_T(h^*)$,且取得了更小的共享误差 λ ,这表明本文通过学习目标域判别特征来提高域不变特征在目标域判别性的方法能有效减少域自适应模型的共享误差,提高域自适应能力。

本文通过实验产生了针对不同数据集各个域自适应任务的分类准确率,展现了增加域不变特征的判别性对模型准确率的提升,如表2所示。

在Office-31数据集上,DFSA达到了较高的准确率,比没有充分考虑目标域判别特征学习的DMP方法准确率平均高出1.79个百分点,比DSAN方法的平均准确率高出0.79个百分点。

如表3所示,在Office-Home数据集上,DFSA方法实现的分类准确率相对于其他方法有较大的提升,尤其是在 $C \rightarrow P$ 和 $C \rightarrow R$ 两个任务上,比DSAN方法高出6.45个百分点,且在 $P \rightarrow C$ 、 $R \rightarrow A$ 、 $R \rightarrow C$ 、 $R \rightarrow P$ 四个迁移任务上实现了所有对比方法中的最高模型准确率。这表明在子域自适应过程中对域不变判别特征的挖掘,提高域不变判别特征在目标域的判别性,能有效提高域自适应能力,平均准确率比没有显式挖掘域不变判别特征的DSAN方法高出3.51个百分点,比没有充分考虑目标域特征判别性的DMP方法高出2.91个百分点,比没有考虑目标域判别特征学习的Norm-AE-SPL方法平均准确率高出2.41个百分点。模型准确率提高的同时也反映出更小的目标域假设函数误差 $\epsilon_T(h)$ 上界,以及更好的域自适应性能。但从上述实验结果发现整体准确率普遍

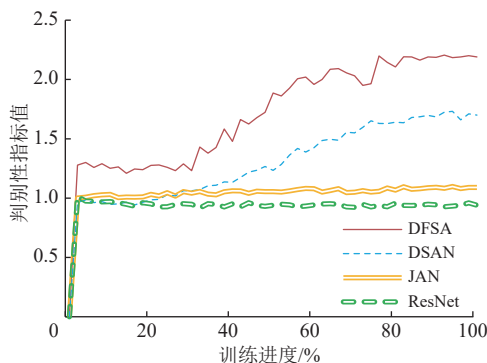


图4 不同方法特征判别性指标统计

Fig.4 Statistical analysis of discriminative indicators for the features of different methods

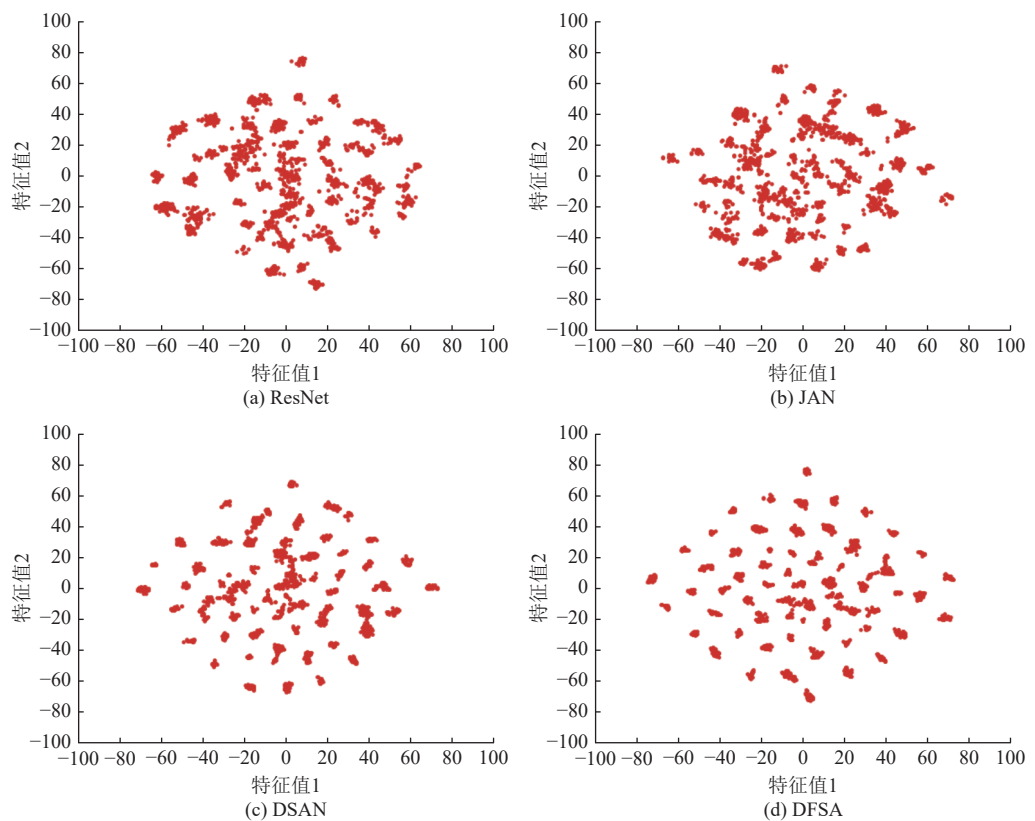


图 5 域不变特征可视化

Fig.5 Visualization of domain invariant features

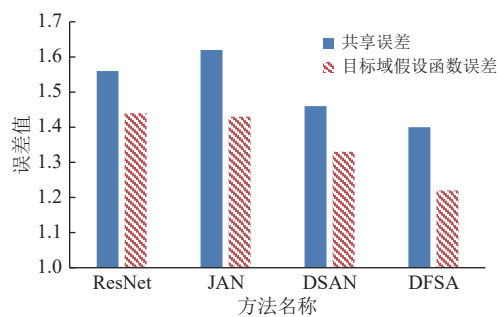


图 6 误差值统计

Fig.6 Statistics of error values

不高,这是由于源域与目标域数据分布存在差异,所以从源域学习的知识不完全适用于目标域,这导致实验结果普遍准确率较低。其次,目标域缺乏标注数据,模型无法准确学习目标域的知识,导致模型准确率不高。

不同方法基于Office-Home数据集的 $C \rightarrow P$ 任务训练域自适应模型后,针对易混淆的几个目标域图片进行了模型预测。不同模型的预测标签与真实标签的对比情况如表4所示,结果表明基于DFSA的模型提取的目标域特征拥有更大的判别性指标值(I_D),能更好地区分易混淆的图片。

表 2 在Office-31数据集上的分类准确率

Table 2 Classification accuracy of different methods on the Office-31 dataset



%

方法名称	域自适应任务						平均值
	$A \rightarrow W$	$D \rightarrow W$	$W \rightarrow D$	$A \rightarrow D$	$D \rightarrow A$	$W \rightarrow A$	
ResNet	68.4±0.5	96.7±0.5	99.3±0.1	68.9±0.2	62.5±0.3	60.7±0.3	76.1
DANN	82.0±0.4	96.9±0.2	99.1±0.1	79.7±0.4	68.2±0.4	67.4±0.5	82.2
ADDA	86.2±0.5	96.2±0.3	98.4±0.3	77.8±0.3	69.5±0.4	68.9±0.5	82.9
JAN	85.4±0.3	97.4±0.2	99.8±0.2	84.7±0.3	68.6±0.3	70.0±0.4	84.3
CDAN4	93.1±0.2	98.2±0.2	100.0±0.0	89.8±0.3	70.1±0.4	68.0±0.4	86.6
Norm-AE-SPL	88.6	98.7	97.1	93.0	73.8	74.2	87.6
DMP	93.0±0.3	99.0±0.1	100.0±0.0	91.0±0.4	71.4±0.2	70.2±0.2	87.4
DSAN	93.6±0.2	98.3±0.1	100.0±0.0	90.2±0.7	73.5±0.5	74.8±0.4	88.4
DFSA	93.65±0.4	98.42±0.1	100±0.0	92.07±0.1	75.91±0.3	75.11±0.6	89.19

表 3 在Office-Home数据集上的分类准确率
Table 3 Classification accuracy of different methods on the Office-Home dataset

方法名称	域自适应任务												平均值
	$A \rightarrow C$	$A \rightarrow P$	$A \rightarrow R$	$C \rightarrow A$	$C \rightarrow P$	$C \rightarrow R$	$P \rightarrow A$	$P \rightarrow C$	$P \rightarrow R$	$R \rightarrow A$	$R \rightarrow C$	$R \rightarrow P$	
ResNet	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
DANN	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	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	49.0	69.3	74.5	54.4	66.0	68.4	55.6	48.3	75.9	68.4	55.4	80.5	63.8
DSAN	54.4	70.8	75.4	60.4	67.8	68.0	62.6	55.9	78.5	73.8	60.6	83.1	67.6
DMP	52.3	73.0	77.3	64.3	72.0	71.8	63.6	52.7	78.5	72.0	57.7	81.6	68.1
Norm-AE-SPL	51.6	76.0	80.6	63.0	77.0	78.4	62.9	50.7	81.2	66.3	52.8	82.9	68.6
IMANLDP	59.6	77.3	79.5	67.4	75.9	74.6	66.1	56.4	81.0	74.5	61.4	84.4	71.5
DFSA	58.72	74.45	77.28	65.51	74.25	74.45	65.60	58.90	79.37	74.99	63.76	84.86	71.01

表 4 不同方法应用效果图
Table 4 Application effect diagrams of different methods

图例	不同方法的模型预测			
	ResNet ($I_D=0.85$)	JAN ($I_D=0.94$)	DSAN ($I_D=1.02$)	DFSA ($I_D=1.26$)
 椅子	沙发(×)	沙发(×)	沙发(×)	椅子(√)
 沙发	床(×)	椅子(×)	椅子(×)	沙发(√)
 床	沙发(×)	沙发(×)	沙发(×)	床(√)

综上所述,本文提出的基于域不变判别特征学习的深子域自适应方法,通过挖掘域不变判别特征,很好地保留了域不变特征在目标域的判别性,在基于图像分类的域自适应过程中能更好地区分易混淆图片,获得了目标域数据较高的分类准确度。这表明,本文针对域不变判别特征的学习能有效提高模型域自适应能力。

4.4 消融实验结果与分析

本节进行消融实验,统计了DFSA各个模块给域不变特征带来的判别性改善,以及给深子域自适应方法带来的增益。

图7与图8对Office-31数据集 $A \rightarrow C$ 任务中的域不变特征类间距与类内距进行了统计。其中resnet50表示直接使用目标域的真实标签对resnet50预训练模型进行微调训练,利用训练好的模型提取特征并统计特征的类间距与类内距,lmmd表示在resnet50基础

上添加局部均值差异损失后进行模型训练,以此类推,mcc表示类混淆损失,cc表示一致性损失。如图7所示,类混淆损失和一致性损失的添加使得模型提取的域不变特征类间距逐步增大,最后可以看到本文方法拥有仅次于resnet50模型提取特征的类间距。图8显示了一致性约束能使得目标域特征呈现更小的类内距。更大的类间距与更小的类内距,意味着更大的判别性指标值,体现了DFSA各个模块对域不变特征判别性的提高。

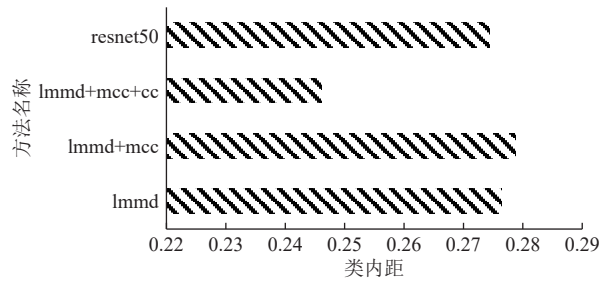


图 8 类内距统计
Fig.8 Statistics of intra-class variance

表5记录了子域自适应过程中不同损失模块在Office-31和Office-Home数据集上域自适应任务的平均准确率,其中lmmd,mcc,cc分别表示子域对齐损失,类混淆损失和一致性损失。

如表5的实验结果所示。仅考虑部分损失模块的

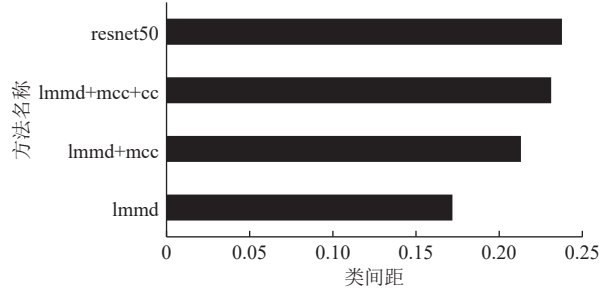


图 7 类间距统计
Fig.7 Statistics of inter-class variance

表 5 消融实验结果
Table 5 Results of ablation experiment

损失模块选择			不同数据集平均分类准确率/%	
lmmd	mcc	cc	Office-31	Office-Home
√	×	×	88.40	67.60
√	×	√	87.30	68.25
√	√	×	88.49	70.23
√	√	√	89.19	71.01

方法取得了较低的平均准确率,最后组合所有损失模块的DFSA方法取得了最高的平均准确率。表明同时考虑最小化类内差异与最大化类间差异的DFSA方法能更好地挖掘域不变判别特征,提高模型的域自适应能力。

表6统计了模型在Office-Home数据集上的 $C \rightarrow P$ 任务中每个epoch训练的平均耗时。结果表明,类混淆损失(mcc)的添加给基于LMMD的模型带来较少的训练时间代价。由于子域对齐损失的计算需要依赖模型预测生成的伪标签,而基于模型预测的类混淆损失(mcc)计算是对模型预测的再运用,提高了模型预测的利用率,而没有增加过多模型训练时间。强弱数据增强一致性损失(cc)的计算由于需要对数据进行多种方式的数据增强来生成不同的视图。利用不同视图的模型预测计算一致性损失,增加了额外的计算步骤和计算数据,给模型训练产生了较大的时间代价。

表 6 不同损失模块每个epoch平均耗时
Table 6 Average time consumption per epoch for different loss modules

损失模块	耗时/(s·epoch ⁻¹)
lmmd	56.32
lmmd+mcc	56.35
lmmd+mcc+cc	75.99

图9展现了模型在Office-Home数据集上的

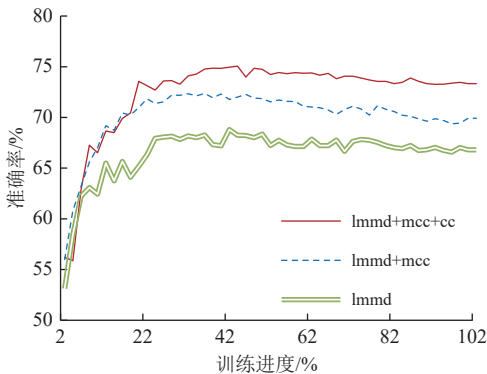


图 9 不同损失模块的收敛速度
Fig.9 Convergence speeds of different loss modules

$C \rightarrow P$ 任务中不同损失模块进行训练的收敛速度。结果显示,从模型收敛所需要的epoch数量来看,不同的损失模块对模型收敛速度影响不大,但结合每个epoch的训练时间来看,本文结合各个损失模块的方法需要更多的时间代价来达到更好的模型域自适应效果。

5 结论

常见的基于域对齐的域自适应方法忽略域不变特征的判别性,造成域不变特征在目标域的判别性差,导致域自适应模型性能不佳。针对这个问题,本文提出了一种基于域不变判别特征学习的深子域自适应方法(DFSA),结合子域对齐、混淆矩阵约束和强弱数据增强伪标签一致性约束来提高域不变特征的判别性,以改善域自适应模型性能。实验结果表明,DFSA有效提高了域不变特征在目标域的判别性,降低了目标域假设函数的误差,与其他无监督域自适应方法相比,DFSA在Office-31和Office-Home数据集上取得显著效果。这表明,通过挖掘域不变判别特征提高域不变特征的判别性可以有效降低模型误差,提高模型准确率,实现更好的域自适应效果。本文方法的实现依赖目标域更加丰富的判别信息,将更加简单有效的无监督特征学习方法引入域自适应模型将是本文未来的一个改进方向。

参考文献:

[1] ZHANG Y C, LIU T L, LONG M S, *et al.* Bridging theory and algorithm for domain adaptation[C]//36th International Conference on Machine Learning (ICML). Long Beach: IEEE, 2019: 7404-7413.

[2] SI L, DONG H, QIANG W, *et al.* Regularized hypothesis-induced wasserstein divergence for unsupervised domain adaptation[J]. *Knowledge-Based Systems*, 2024, 283: 111162.

[3] LI Y, ZHU Y, YING S. Attention-guided optimal transport for unsupervised domain adaptation with class structure prior[J]. *Neural Processing Letters*, 2023, 55(9): 12547-12567.

[4] LONG M, ZHU H, WANG J, *et al.* Deep transfer learning with joint adaptation networks[C]//34th International Conference on Machine Learning(ICML). Sydney: IEEE, 2017: 2208-2217.

[5] XIE B, YUAN L, LI S, *et al.* Active learning for domain adaptation: an energy-based approach[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, 36(8): 8708-8716.

[6] ZHU Y, ZHUANG F, WANG D. Aligning domain-specific distribution and classifier for cross-domain classification from multiple sources[C]//Proceedings of the AAAI Conference on Artificial Intelligence(AAAI). Honolulu: AAAI,

- 2019: 5989-5996.
- [7] WANG Q, MENG F, BRECKON T P. Data augmentation with norm-AE and selective pseudo-labelling for unsupervised domain adaptation[J]. *Neural Networks*, 2023, 161: 614-625.
- [8] GANIN Y, USTINOVA E, AJAKAN H, *et al.* Domain-adversarial training of neural networks[J]. *Journal of Machine Learning Research*, 2016, 17(59): 1-35.
- [9] TZENG E, HOFFMAN J, SAENKO K, *et al.* Adversarial discriminative domain adaptation[C]//30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 2962-2971.
- [10] GAO Z, ZHANG S, HUANG K, *et al.* Gradient distribution alignment certifies better adversarial domain adaptation [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021: 8917-8926.
- [11] GHIFARY M, BALDUZZI D, KLEIJN W B, *et al.* Scatter component analysis: a unified framework for domain adaptation and domain generalization[J]. *Journal of IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39: 1414-1430.
- [12] SHUANG L, SHIJI S, GAO H, *et al.* Domain invariant and class discriminative feature learning for visual domain adaptation[J]. *IEEE Trans Image Process*, 2018, 27: 4260-4273.
- [13] DAN J, JIN T, CHI H, *et al.* Trust-aware conditional adversarial domain adaptation with feature norm alignment[J]. *Neural Networks*, 2023, 168: 518-530.
- [14] ZHU Y, ZHUANG F, WANG J, *et al.* Deep subdomain adaptation network for image classification[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 32(4): 1713-1722.
- [15] WANG P, YANG Y, XIA Y, *et al.* Information maximizing adaptation network with label distribution priors for unsupervised domain adaptation[J]. *IEEE Transactions on Multimedia*, 2023, 25: 6026-6039.
- [16] ZHAO F, LIU W, WEN C. A new method of image classification based on domain adaptation[J]. *Sensors*, 2022, 22(4): 1315.
- [17] 杜社林, 黄炳赫, 李荣鹏, 等. 基于加权分类损失和核范数的领域自适应模型[J]. *计算机应用研究*, 2023, 40(6): 1734-1738.
- DU S L, HUANG B H, LI R P, *et al.* Domain adaptation based on weighted classification loss and nuclear-norm[J]. *Application Research of Computers*, 2023, 40(6): 1734-1738.
- [18] 陈辛怡, 孙涵. 聚类中心对齐的无监督域适应网络[J]. *小型微型计算机系统*, 2022, 43(4): 822-827.
- CHEN X Y, SUN H. Centroid alignment for unsupervised domain adaptation[J]. *Journal of Chinese Computer Systems*, 2022, 43(4): 822-827.
- [19] GAO J, MA X, XU C. Learning transferable conceptual prototypes for interpretable unsupervised domain adaptation[J]. *IEEE Transactions on Image Processing*, 2024, 33: 5284-5297.
- [20] LUO Y W, REN C X, DAI D Q, *et al.* Unsupervised domain adaptation via discriminative manifold propagation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(3): 1653-1669.
- [21] JIN Y, WANG X, LONG M, *et al.* Minimum class confusion for versatile domain adaptation[C]//Computer Vision-ECCV 2020: 16th European Conference(ECCV). Glasgow: ECCV, 2020: 464-480.
- [22] SOHN K, BERTHELOT D, CARLINI N, *et al.* Fixmatch: simplifying semi-supervised learning with consistency and confidence[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 596-608.
- [23] WANG X, ZHUO J, ZHANG M, *et al.* Revisiting unsupervised domain adaptation models: a smoothness perspective [C]//Proceedings of the Asian Conference on Computer Vision(ACCV). Macao: ACCV, 2022: 1504-1521.
- [24] CUBUK E D, ZOPH B, SHLENS J, *et al.* Randaugment: practical automated data augmentation with a reduced search space[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 3008-3017.
- [25] REDKO I, MORVANT E, HABRARD A. A survey on domain adaptation theory: learning bounds and theoretical guarantees[EB/OL]. arXiv: 2004.11829(2022-07-13) [2024-10-25]. <https://arxiv.org/abs/2004.11829>.
- [26] SAENKO K, KULIS B, FRITZ M, *et al.* Adapting visual category models to new domains[C]//11th European Conference on Computer Vision(ECCV). Heraklion: ECCV, 2010: 213-226.
- [27] VENKATESWARA H, EUSEBIO J, CHAKRABORTY S, *et al.* Deep hashing network for unsupervised domain adaptation[C]//30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 5385-5394.
- [28] HE K, ZHANG X, REN S, *et al.* Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer vision and Pattern Recognition(CVPR). Las Vegas: IEEE, 2016: 770-778.
- [29] LONG M S, CAO Z J, WANG J M, *et al.* Conditional adversarial domain adaptation[C]//32nd Conference on Neural Information Processing Systems (NIPS). Montreal: NIPS, 2018: 1640-1650.

(责任编辑: 王威娜 英文审核: 费伦科)