

# 伪标签不确定性估计的源域无关鲁棒域自适应\*

王帆, 韩忠义, 尹义龙



(山东大学 软件学院, 山东 济南 250100)

通信作者: 韩忠义, E-mail: hanzhongyicn@gmail.com; 尹义龙, E-mail: ylyin@sdu.edu.cn

**摘要:** 无监督域自适应是解决训练集(源域)和测试集(目标域)分布不一致的有效途径之一. 现有的无监督域自适应的理论和方法在相对封闭、静态的环境下取得了一定成功, 但面向开放动态任务环境时, 在隐私保护、数据孤岛等限制条件下, 源域数据往往不可直接获取, 现有无监督域自适应方法的鲁棒性将面临严峻的挑战. 鉴于此, 研究了一个更具挑战性却又未被充分探索的问题: 源域无关的无监督域自适应, 目标是仅依据预训练的源域模型和无标签目标域数据, 实现源域向目标域的正向迁移. 提出一种基于伪标签不确定性估计的源域无关鲁棒域自适应的方法 PLUE-SFRDA (pseudo label uncertainty estimation for source free robust domain adaptation). PLUE-SFRDA 的核心思想是: 根据源域模型的预测结果, 联合信息熵和能量函数充分挖掘目标域数据的隐含信息, 探索类原型和类锚点, 以准确估计目标域数据的伪标签, 进而调优域自适应模型, 实现源域数据无关的鲁棒域自适应. PLUE-SFRDA 包含提出的二元软约束信息熵, 解决了标准信息熵不能有效估计处于决策边界样本的不确定性的问题, 增强了所挖掘的类原型和类锚点的可信度, 进而提高了目标域伪标签估计的准确率. PLUE-SFRDA 包含了提出的加权对比过滤方法, 通过比较每个样本距离该类的类锚点和其他类的类锚点的加权距离, 过滤掉处于决策边界的类别信息模糊样本, 进一步提高了伪标签不确定性估计的安全性. PLUE-SFRDA 还包含一个信息最大化损失, 实现源域分类器和伪标签估计器迭代优化, 逐渐将源域模型中蕴含的源域知识迁移至目标域, 进一步提高了伪标签不确定性估计的鲁棒性. 在 Office-31, Office-Home 和 VisDA-C 这 3 个公开的基准数据集上的大量实验表明: PLUE-SFRDA 不仅超过了最新的源域无关的域自适应方法的表现, 还显著优于现有的依赖源域数据的域自适应方法.

**关键词:** 无监督域自适应; 源域无关的域自适应; 伪标签学习; 信息熵; 能量函数; 不确定性估计  
中图法分类号: TP18

中文引用格式: 王帆, 韩忠义, 尹义龙. 伪标签不确定性估计的源域无关鲁棒域自适应. 软件学报, 2022, 33(4): 1183–1199. <http://www.jos.org.cn/1000-9825/6467.htm>

英文引用格式: Wang F, Han ZY, Yin YL. Source Free Robust Domain Adaptation Based on Pseudo Label Uncertainty Estimation. Ruan Jian Xue Bao/Journal of Software, 2022, 33(4): 1183–1199 (in Chinese). <http://www.jos.org.cn/1000-9825/6467.htm>

## Source Free Robust Domain Adaptation Based on Pseudo Label Uncertainty Estimation

WANG Fan, HAN Zhong-Yi, YIN Yi-Long

(School of Software, Shandong University, Jinan 250100, China)

**Abstract:** Unsupervised domain adaptation is one of the effective ways to solve the inconsistent distribution of training set (source domain) and test set (target domain). Existing unsupervised domain adaptation theories and methods have achieved some success in relatively closed and static environments. However, for open dynamic task environments, the robustness of existing unsupervised domain adaptation methods will face serious challenges under the constraints of privacy protection and data silos, where source domain data are

\* 基金项目: 国家自然科学基金(62176139)

本文由“面向开放场景的鲁棒机器学习”专刊特约编辑陈恩红教授、李宇峰副教授、邹权教授推荐.

收稿时间: 2021-03-10; 修改时间: 2021-07-16; 采用时间: 2021-08-27; jos 在线出版时间: 2021-10-26

often not directly accessible. In view of this, this paper investigates a more challenging yet under-explored problem: source free unsupervised domain adaptation, with the goal of achieving positive transfer from the source domain to the target domain based only on the pre-trained source domain model and unlabeled target domain data. In this paper, we propose a method called PLUE-SFRDA (pseudo label uncertainty estimation for source free robust domain adaptation). The core idea of PLUE-SFRDA is to combine information entropy and energy function to fully explore the implicit information of the target domain data based on the prediction results of the source domain model, explore the class prototypes and class anchors to accurately estimate the pseudo label of the target domain data, and then tune the domain adaptation model to achieve the source free robust domain adaptation. PLUE-SFRDA contains a proposed binary soft constraint information entropy, which solves the problem that the standard information entropy cannot effectively estimate the pseudo label uncertainty of samples at the decision boundary, enhances the confidence of the mined class prototypes, and thus improves the accuracy of pseudo label estimation in the target domain. PLUE-SFRDA contains a weighted comparison filtering method proposed by this paper. By comparing the weighted distances of each sample to the class anchors of other classes, the fuzzy samples of class information at the decision boundary are filtered out, which further improves the security of the new pseudo label uncertainty estimation. PLUE-SFRDA also contains an information maximization loss to achieve iterative optimization of the source domain classifier and the pseudo label estimator, which gradually migrates the source domain knowledge embedded in the source domain model to the target domain, further improving the robustness of the pseudo label uncertainty estimation. Extensive experiments on three publicly available datasets, Office-31, Office-Home and VisDA-C, show that PLUE-SFRDA not only outperforms the state-of-the-art source-free domain adaptation methods but also significantly outperforms standard domain adaptation methods which depend on the source-domain data.

**Key words:** unsupervised domain adaptation; source-free domain adaptation; pseudo label learning; information entropy; energy function; uncertainty estimation

无监督域自适应的研究作为当前机器学习研究的热点和难点, 备受关注. 机器学习技术, 作为人工智能的核心分支之一, 在图像分类和语音识别等应用上取得了突破性进展, 并被广泛应用于众多数据分析领域. 大部分的机器学习方法都是基于训练集和测试集的概率分布一致的假设, 但当数据分布一致的条件不满足时, 现有机器学习模型或方法的性能将受到不同程度的影响. 针对非同分布条件的任务, 无监督域自适应 (unsupervised domain adaptation) 是有效提升模型或方法的鲁棒性和泛化性的可行途径之一<sup>[1,2]</sup>. 域自适应的目标是在训练集(源域)和测试集(目标域)概率分布不一致的情况下实现有效学习. 现有的无监督域自适应的方法已经在目标检测<sup>[3,4]</sup>、目标识别<sup>[5]</sup>、语义分割<sup>[6]</sup>等具体任务中取得了突出的成绩.

近年来, 国内外同行已在域自适应的理论和方法上开展了较为深入的研究, 取得了较大进展, 初步奠定了域自适应方向的研究基础. 在域自适应理论研究层面, 从分布变化假设的视角, 现有的域自适应理论研究工作主要分为基于协变量变化假设的理论分析和基于目标变化假设的理论分析两类: (1) 基于协变量变化假设的理论分析工作假设源域和目标域之间仅特征边缘分布不同, 其他条件均一致. 其主要思想是: 提出一种分布差异损失, 衡量边缘分布变化的程度<sup>[7-9]</sup>; (2) 基于目标变化假设的理论分析工作假设源域和目标域之间除了特征边缘分布不同之外, 标签集(分布)也可能不同<sup>[10,11]</sup>; 而在域自适应方法研究层面, 已有的域自适应方法大概可以分为 4 类: 基于样本重要性估计的方法<sup>[12]</sup>、基于高阶距匹配的方法<sup>[13,14]</sup>、基于伪标签的方法<sup>[15]</sup>以及基于对抗训练的方法<sup>[16]</sup>.

现有的域自适应理论和方法主要集中在如何解决数据分布不一致问题, 且基本都隐式地假设了除数据分布不一致之外, 其他条件都是相对较为理想的或者是稳态的. 但是很多实际应用面临的则是开放动态任务环境, 并不满足上述假设, 比如存在源域数据不可见、源域数据受损(含噪)、源域或目标域存在私有(未知)类等现实问题. 如果将应用拓展到开放动态任务环境中, 域自适应方法的鲁棒性将面临严峻的挑战. 鉴于此, 本文研究了一个更具挑战性的问题: 源域无关的无监督域自适应, 目标是仅依据预训练的源域模型和无标签目标域数据, 实现源域向目标域的正向迁移. 具体来说, 现有的域自适应研究工作在相对封闭、静态的环境中取得成功基于一个非常重要的假设和前提: 大量且带标签的源域数据可以自由获取. 但是随着人们对人工智能安全研究的进一步深入和对数据隐私、数据安全的关注程度的不断提高, 这个前提在现实情况下极有可能是无法满足的. 由于数据隐私和数据孤岛等约束<sup>[17]</sup>, 源域数据往往不可以被直接获取, 我们只能获得基于源域数据的预训练模型和目标域无标签的数据. 比如, 某医院建立了基于机器学习的某疾病筛查模型, 并计划将此

模型迁移到其他地区的定点医院. 由于地区之间的某疾病数据存在分布差异(由性别、年龄、成像设备差异等因素引起), 需要采取无监督域自适应策略, 但是根据当地的政策规定, 病人的数据受到法律保护, 不能将某疾病病人的源影像数据提供给对方医院, 只能提供预训练的模型. 而如何依据预训练模型和目标域无标签数据实现鲁棒的无监督域自适应, 这一挑战性问题尚未得到充分的探索.

源域无关的鲁棒域自适应是极具重要研究价值和挑战性的问题, 主要体现在以下 3 个方面.

- (1) 用户隐私保护的必要性. 大数据时代催生了数据泄露等问题的潜在隐患, 在某些情况下, 用户不愿共享自己的个人信息, 例如病人可选择自己的医学影像数据是否用于临床实验;
- (2) 数据安全的必要性. 商业竞争环境下, 国家和企业对于数据安全的关注程度不断提高, 受法律或政策等的限制, 某些关键性数据是无法获取的;
- (3) 减轻数据依赖, 转换研究角度的必要性. 过度依赖源域数据进行的域自适应虽然在理论和应用中取得了一定的成功, 但是其结果显示, 其迁移性能仍然有限. 研究角度的转换, 可能会推动域自适应领域的进一步发展.

源域数据无关的域自适应虽然具有较大研究价值, 但也具有较强的挑战性, 主要体现在以下 3 个方面.

(1) 分布对齐的困难性. 如图 1 所示(上方: 传统的无监督域自适应设置中, 可同时获得带标签源域数据和无标签目标域数据; 下方: 源域无关的域自适应设置中, 仅可以获取预训练的源域模型和无标签的目标域数据. 由于源域数据无法直接获得, 源域分布和目标域分布对齐困难). 由于源域数据无法直接获取, 使得源域分布和目标域分布无法进行样本级别或特征级别的对齐.

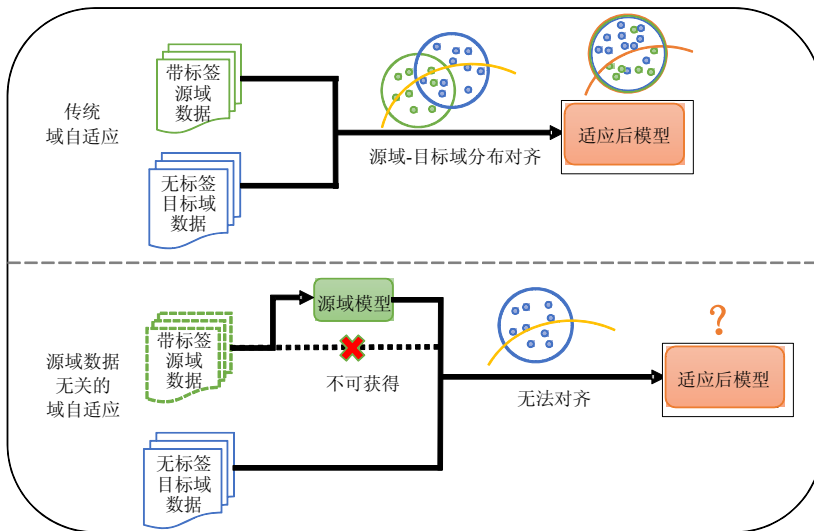


图 1 传统的无监督域自适应和源域无关的域自适应设置的对比图

(2) 负迁移问题. 由于基于源域数据预训练的模型中仅蕴含少量源域数据信息, 不足以支撑源域信息的正向域自适应, 容易造成负迁移问题.

(3) 目标域伪标签准确估计的困难性. 传统的域自适应方法可以利用源域数据和目标域数据优化分布间的高维统计量差异, 同时准确估计目标域样本标签. 而以源域数据不可获取为前提, 实现源域无关的鲁棒域自适应的主要手段之一只能是利用预训练模型充分挖掘目标域数据的伪标签信息, 并无偏差地估计目标域数据的伪标签, 进而实现域自适应模型的调优. 但是, 受分布差异的条件限制, 使用源域模型预测目标域样本的标签是极其不准确的. 在此情况下, 如何提升目标域样本伪标签的准确率, 是源域数据无关的鲁棒域自适应的关键挑战性问题.

为了准确估计目标域样本的伪标签进而调优模型实现正向迁移, 本文提出了一种基于伪标签不确定性估

计的源域无关鲁棒域自适应的方法(pseudo label uncertainty estimation for source free robust domain adaptation, PLUE-SFRDA). PLUE-SFRDA 的核心思想是: 根据源域模型的预测能力, 联合信息熵和能量函数挑选类原型, 并为每个目标域实例赋予伪标签, 进一步挖掘可信实例实现目标域模型的监督训练. 本文发现: 标准信息熵在挑选类原型时, 不能准确估计处于分类边界类别比较模糊样本的不确定性, 从而导致类原型挑选准确率过低的现象. 为解决此问题, 本文提出了一种新颖的二元软约束信息熵, 解决了标准信息熵不能有效估计处于决策边界样本伪标签不确定性的问题, 增强了所挑选的类原型和所挖掘的类锚点的可信度, 进而提高了目标域伪标签估计的准确率. 另外, 本文发现: 在基于类原型为目标域实例赋予伪标签的过程中, 极有可能引入大量的标签噪声, 这些噪声容易被具有较强记忆功能的神经网络完全充分的学习, 并随着训练过程的进行, 神经网络逐渐过拟合至噪声数据的错误标签, 从而导致模型的性能显著下降<sup>[18]</sup>. 为此, 本文引入了加权对比过滤方法, 通过比较每个样本距离该类的类锚点和其他类的类锚点的加权距离, 过滤掉处于决策边界的类别信息模糊样本, 进一步提高了伪标签不确定性估计的安全性. 最后, 为了充分利用源域模型中存储的有关源域数据的信息, PLUE-SFRDA 使用源域模型的分器模块, 引入信息最大化损失, 实现源域分类器和伪标签估计器迭代优化, 逐渐将源域模型中蕴含的源域知识迁移至目标域, 显著提高了伪标签不确定性估计的鲁棒性. 本文工作的主要贡献可以总结如下.

- (1) 研究了一个极具挑战性和重要研究价值问题: 源域无关域自适应, 并提出了基于伪标签不确定性估计的源域无关鲁棒域自适应方法, 实现了安全、可靠、鲁棒的源域无关域自适应;
- (2) 发现了标准信息熵在挑选类原型时的不适用现象, 提出了一种二元软约束信息熵, 解决了标准信息熵不能有效估计处于决策边界样本伪标签不确定性的问题, 有效提高了类原型挑选的正确率;
- (3) 提出了加权对比过滤方法, 可过滤掉处于决策边界的类别信息模糊样本, 进一步降低了伪标签的不确定性. 公开基准数据集上的大量实验, 验证了本文所提方法的鲁棒性.

## 1 相关工作

### 1.1 无监督域自适应方法研究进展

目前, 主流的无监督域自适应算法主要分为基于样本重要性估计、高阶矩匹配、伪标签以及对抗训练的学习. 这些方法有各自不同的优缺点, 其在实现思路方面也各不相同.

(1) 基于样本重要性估计的域自适应方法, 其基本思想是假设源域和目标域分布之间存在着交集, 通过计算源域样本点与重叠分布的距离估计源域样本中的重要性, 优化加权后的损失函数<sup>[12]</sup>;

(2) 基于高阶矩匹配的方法, 其核心思路是试图最小化源域和目标域分布之间的高维统计量差异. 部分代表性工作为: Huang 等人从样本选择偏差的角度分析分布变化, 发现最小化分布间希尔伯特再生核空间的特征距离等效于基于样本重要性估计的无监督域自适应思路, 并提出最大化平均差异(maximum mean discrepancy)的思想<sup>[9]</sup>; 龙明盛等人提出深度域自适应网络, 使用多核选择方法进行均值嵌入匹配<sup>[13,14]</sup>;

(3) 基于伪标签的无监督域自适应算法试图为目标域的每个实例赋予伪标签, 并筛选可信实例实现模型的监督训练. 部分代表性工作为: 非对称式三重训练(asymmetric tri-training)<sup>[15]</sup>、协作与对抗网络<sup>[20]</sup>、基于图半监督的伪标签学习<sup>[21]</sup>等;

(4) 基于对抗训练的方法, 通过引入域判别器来区分样本是来自源域还是目标域, 并在此过程中学习有效的特征表示. Ganin 等人提出了第 1 个深度对抗域自适应网络<sup>[16]</sup>, 是此类方法的开创性工作. 其他部分代表性的工作为: Saito 等人发现, 当两个训练好的分类器的输出差异较大时, 输入的实例往往是目标域中远离支撑集的实例, 并基于此提出了一种新的对抗训练策略来最小化分类器之间的差异<sup>[22]</sup>; 龙明盛等人提出了基于条件对抗的域自适应网络, 将数据的类别信息融入到域判别器中<sup>[23]</sup>.

### 1.2 源域无关的域自适应方法研究进展

基于源域信息的透明程度角度, 现有的源域无关域自适应工作主要可分为需要利用额外信息和完全不需

要额外信息两类. (1) 需要利用额外信息工作的主要思想是: 假设除源域模型和无标签目标域样本外, 还可利用其他外部信息, 代表性工作有 AOT<sup>[24]</sup>和 MCS<sup>[25]</sup>. 其中, AOT 额外通过少量带标签的目标域样本来估计分布差异和标签差异, 从而实现源域分布和目标域分布的部分对齐. 与 AOT 不同, MCS 需要获取源域数据中每类的均值和方差, 通过对齐各类子空间的均值和方差来对齐源域目标域分布. 而这些额外信息在现实场景下极难获取, 完全无法获取源域额外信息的情况下进行源域数据无关的域自适应研究显得更加现实, 也是本文所研究的场景. (2) 完全不需要额外信息工作的主要思想是从实际出发, 严格遵守源域数据不可得假设, 试图恢复源域数据信息或深度挖掘目标域数据隐含的信息. Chidlovskii 等人在 2016 年首次提出源域数据不可见问题, 并对该问题进行了初步讨论, 但由于技术和方法的限制, 导致其效果不佳<sup>[26]</sup>. 随着深度模型的发展, 该问题被重新提起, 但是对源域无关域自适应的研究仍然有限. 其方法在实现思路方面各不相同, 主要可分为以下两类——基于伪标签技术的模型微调方法和基于生成模型的方法: ① 基于模型微调的方法的思路是: 充分利用模型在目标域上的预测结果, 挖掘隐含的有价值信息. 代表性的工作主要是 2020 年提出的 3 种方法: SHOT<sup>[27]</sup>, PrDA<sup>[28]</sup>和 PPDA<sup>[29]</sup>. SHOT 将模型分为固定的编码器模块和分类器模块, 基于聚类方法获得的目标域伪标签, 利用信息最大化损失和监督损失优化目标域模型. 但是通过聚类获得的样本伪标签含有极大比例的噪声, 导致目标域模型有可能会过拟合于噪声样本. PrDA 利用训练阶段的目标模型来预测伪标签, 并进行样本级别的过滤后结合固定的源域模型分类器模块输出的伪标签共同来动态优化模型. 但是其结果显示, 利用该方法训练模型的性能有限. PPDA 是对 PrDA 方法的改进, 在过滤阶段加上了缓冲系数. 但是其结果显示, 该方法效果依然不高. 本文所提方法也基于伪标签思想, 但本方法有效地挑选了可信的类原型, 进而显著降低了伪标签的不确定性, 并且利用源域的分类器模块进一步迁移了源域知识, 最终使得目标模型的性能明显提高; ② 基于生成模型的方法仅有 2020 年提出的 MA<sup>[30]</sup>一项工作. 该方法提出了 3C-GAN 框架, 作者表示, 预测模型可以通过生成对抗模型生成的带标签的伪目标域数据不断微调模型. 这反过来促进生成模型生成更加准确的伪目标域数据, 生成模型和预测模型在不使用源域数据的条件下相互合作, 共同优化. 这个方法极大地提高了模型在目标数据上的性能, 但是该方法计算代价过大, 并且对于一些具有挑战性的数据集来说, 生成准确的目标域数据非常困难. 另外, 在源域无关自适应研究中, 除了上述针对分类任务的工作之外, 还有针对语义分割任务的方法: UBNA<sup>[31]</sup>, Privacy Preserving<sup>[32]</sup>和 Source-relaxed<sup>[33]</sup>; 针对目标检测任务的方法: SF-UDA<sup>[34]</sup>和 SFOD- Mosaic<sup>[35]</sup>. 更进一步, Tent<sup>[36]</sup>探讨了更具挑战的情况, 即在线学习(online learning)中的源域无关域自适应问题.

### 1.3 基于伪标签学习的域自适应研究进展

伪标签技术属于半监督学习的有效解决方案, 在半监督学习问题上取得了一定的成功. 其主要思想是: 基于在标记数据上训练的模型, 为未标记数据赋予初始化伪标签, 然后使用部分可信伪标签迭代优化分类器. 而近几年, 该思想也被诸多无监督域自适应工作借鉴. 这些工作充分挖掘了无标签目标域数据的隐含信息, 为其赋予伪标签, 使得域自适应模型效果有了进一步的明显提升<sup>[37-39]</sup>. 而在本文研究的源域无关的无监督的域自适应场景中, 也有相关的工作试图借鉴伪标签思想. Tent<sup>[36]</sup>利用为样本的信息熵预先设定阈值来为目标数据赋予伪标签, PrDA<sup>[28]</sup>和 PPDA<sup>[29]</sup>使用最小最大化思想为样本信息熵动态设定阈值来挑选类原型, 作为进一步估计目标域样本伪标签的基础. 但如何根据预训练模型的预测结果, 充分挖掘目标域数据的伪标签信息, 从而降低伪标签的不确定性, 是解决无源域无监督域自适应问题的一个关键性难点. 现阶段, 该问题并未得到充分探索. 对此, 本文提出的二元软约束信息熵可以有效提高挑选的类原型的准确度, 从而可以进行正确的伪标签预测, 加权对比过滤的阶段更是进一步降低了所赋予伪标签样本的不确定性.

## 2 方 法

本节首先概述了源域无关域自适应的问题设置, 然后详细介绍了本文所提出的基于伪标签不确定性估计的源域无关鲁棒域自适应方法.

### 2.1 问题设置

无监督域自适应问题和源域无关的无监督域自适应问题的最大的区别在于后者无法直接获取到源域数据,但是可以获取预先训练好的模型(如图 1 所示).为了阐明两者设置上的区别,本节分别描述无监督域自适应和源域无关域自适应的问题设置.

无监督域自适应设置:  $n_s$  个有标签样本  $D_s = \{x_i, y_i\}_{i=1}^{n_s}$  可从定义在  $\mathcal{X} \times \mathcal{Y}$  上的源域分布  $Q$  中抽样得到,其中,  $\mathcal{X}$  表示特征集合,  $\mathcal{Y}$  表示标签集合. 在多分类问题中,  $\mathcal{Y} \in \{1, \dots, K\}$ ,  $K$  表示类别数量. 同时, 存在  $n_t$  个无标签样本  $D_t = \{x_i\}_{i=1}^{n_t}$  可从目标域分布  $P$  中抽样得到. 假设分布  $Q$  和  $P$  相似但不同, 即假设理想联合误差很小, 基本不影响目标域期望风险的上界. 无监督域自适应的目标是, 从标签函数(分类器)假设集  $\mathcal{H}$  中挑选一个在目标域中期望风险最小的标签函数  $h_t \in \mathcal{H}$ ;

源域无关的无监督域自适应设置: 由于数据隐私或数据孤岛等原因, 不可以直接从分布为  $Q$  源域数据中抽样获取源域样本  $D_s$ . 具体来说, 源域无关的域自适应可以获得预训练好的源域模型  $h_s$  和从目标域分布  $P$  中抽样得到  $n_t$  个无标签样本  $D_t = \{x_i\}_{i=1}^{n_t}$ , 同样假设分布  $Q$  和  $P$  相似但不同, 源域无关的域自适应的目标是: 在不使用有标签样本  $D_s$  的前提下, 从标签函数(分类器)假设集  $\mathcal{H}$  中挑选一个在目标域中期望风险最小的标签函数  $h_t \in \mathcal{H}$ , 实现源域向目标域的正向迁移. 本文将源域无关的无监督域自适应统一简称为源域无关的域自适应.

### 2.2 方法框架

PLUE-SFRDA 的核心在于有效降低伪标签的不确定性, 这是正向迁移的关键, 也是进一步训练深度目标域模型的重要前提. 为实现目标域伪标签不确定性的无偏估计, PLUE-SFRDA 融合信息熵和能量函数的物理特性, 准确估计可信类原型, 进一步降低了目标域伪标签的不确定程度. PLUE-SFRDA 采取 3 步走学习策略(如图 2 所示).

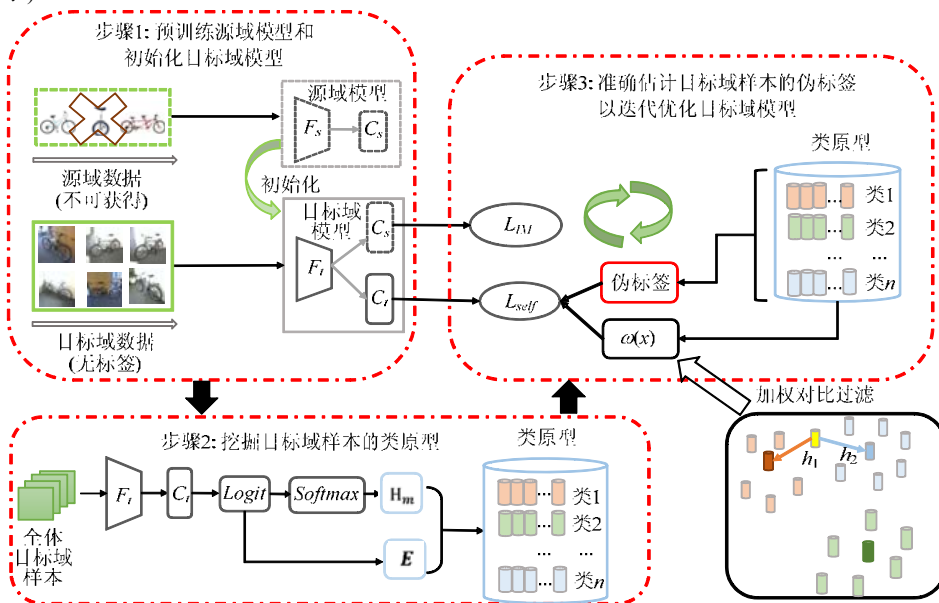


图 2 PLUE-SFRDA 总体的框架图

第 1 步预训练源域模型和初始化目标域模型, 模拟第三方预训练源域模型, 用于学习源域领域的知识和初始化目标域模型. 但由于源域和目标域分布的差异性, 仅利用源域模型预测的输出值来赋予全体目标域样本伪标签是不合理的; 因此, 在第 2 步中, 利用新提出的二元软约束熵和能量函数衡量全体目标域样本伪标签的不确定程度, 挖掘每类的可信样本作为类原型; 第 3 步是估计目标域样本的伪标签以迭代优化目标域模型, 利用第 2 步得到的类原型为全体目标域数据赋予伪标签, 并基于加权对比过滤进一步筛选处于分类边界

类别模糊的样本. 然后, 利用传统的交叉熵损失和信息最大化损失迭代优化目标模型  $h_t$ .

### 2.2.1 预训练源域模型和初始化目标域模型

#### • 预训练源域模型

源域无关的域自适应问题设置通常假设真实任务中的源域模型由第三方提供, 但在现实环境下无第三方, 故首先模拟第三方, 利用带标签的源域样本训练具有较好学习能力的源域模型  $h_s$ , 并将此源域模型作为目标域模型的初始化模型. 源域模型的总体结构如图 2 所示, 其中:  $F_s$  表示特征提取器, 用以提取源域的特征;  $C_s$  表示源域分类器模块, 以此来保留源域高层知识, 实现进一步迁移. 为了防止训练的源域模型在源域数据中过拟合, PLUE-SFRDA 在交叉熵损失的基础上引入了标签平滑化(label smooth)技巧<sup>[40]</sup>. 基于标签平滑化的交叉熵损失函数为

$$L_{\text{label\_smooth}}(D_s) = -E_{x_s \in D_s} \sum_{k=1}^K 1_{[k=y_s]}^{ls} \log(\sigma(C_s(F_s(x_s)))) \quad (1)$$

其中,

$$1_{[k=y_s]}^{ls} = (1 - \varepsilon) 1_{[k=y_s]} + \frac{\varepsilon}{K} \quad (2)$$

$\sigma$  表示 *softmax* 操作,  $1_{[k=y_s]}$  表示指示函数,  $\varepsilon$  表示平滑参数,  $y_s$  表示源域数据的标签.

#### • 初始化目标域模型

目标域模型的总体结构如图 2 所示, 其中,  $F_t$  表示特征提取器;  $C_s$  表示源域分类器;  $C_t$  表示目标域独享的分类器, 用于识别目标域样本. 首先, 将源域特征提取器  $F_s$  的参数赋予目标域特征提取器  $F_t$ ; 接着, 利用  $C_s$  初始化目标域分类器  $C_t$ , 保留源域的分类器模块用以进一步迁移源域知识. 在后续步骤的域自适应学习过程中, 仅调优初始化的目标域模型, 不会再涉及到源域数据.

### 2.2.2 挖掘目标域样本的类原型

类原型表示目标域每类样本中每类的可信样本. 首先, 利用源域模型预测全体目标域样本的标签, 但由于源域数据分布和目标域数据分布的差异性, 会导致预测结果不够准确; 然后将源域模型预测的标签视为初始伪标签, 使用信息熵和能量函数为每类样本挑选可信的伪标签作为每类样本的类原型, 作为进一步估计目标域样本伪标签的基础. 从某种程度来说, 类原型挑选的准确率和目标域样本伪标签的准确率成正相关关系, 所以挑选准确率高的类原型是算法的关键.

经过广泛的调研发现: 信息熵能够估计模型预测的不确定性程度, 信息熵越小, 表示预测结果的正确率越大. 标准的信息熵衡量模型预测结果的计算方式为

$$H(x'_i) = -\sum p(x'_i) \log p(x'_i) \quad (3)$$

其中,  $x'_i$  表示目标域样本,  $p(x'_i)$  表示 *Softmax* 函数输出的该样本属于每一类的概率集合. 但是分析发现, 标准信息熵的计算方式不能准确评估某些处于决策边界的分类模糊样本的伪标签的不确定性. 例如, 对于三分类任务来说, 假设  $[x'_1, x'_2, x'_3, x'_4, x'_5, x'_6]$  表示 6 个目标域实例, 具有相同的标签  $y=1$ , 模型的预测结果皆为  $\hat{y}=1$ , 假设模型预测 6 个实例属于第一类的概率分别为  $[0.9, 0.8, 0.7, 0.6, 0.5, 0.4]$ . 从不确定性角度, 以上 6 个实例的不确定性程度依次增加, 相应地, 信息熵应该是依次增加的. 但如图 3 所示: 假定一个三分类问题,  $[x'_1, x'_2, x'_3, x'_4, x'_5, x'_6]$  是 6 个预测为相同标签的样本,  $[0.9, 0.1, 0.0]$  表示 *softmax* 函数输出的概率向量, 向量中每个元素表示该实例属于某类的概率. 蓝色线表示标准信息熵的计算结果, 绿色线表示二元软约束熵的计算结果. 由图中折线走势可以看出, 二元软约束熵的结果最符合预期, 结果却不符合预期:  $H(x'_4) \ll H(x'_3)$ ,  $H(x'_4) \cong H(x'_5)$ . 由于  $x'_4$  和  $x'_5$  处于决策边界, 故说明标准信息熵的计算方式不能准确评估某些处于决策边界样本伪标签的不确定性.

为了解决此问题, 本文改进了标准熵的计算方式, 提出了一种二元软约束信息熵  $H_m(x'_i)$ . 二元软约束信息熵只关注概率向量中最大的元素, 将概率向量中最大的元素分为大于等于  $\lambda(1 > \lambda > 0)$  和小于  $\lambda$  两种情况:

$$H_m(x'_i) = \begin{cases} -(m(x'_i) \log m(x'_i) + (1 - m(x'_i)) \log(1 - m(x'_i))), & \text{if } m(x'_i) \geq \lambda \\ -((1 - m(x'_i)) \log m(x'_i) + m(x'_i) \log(1 - m(x'_i))), & \text{else} \end{cases} \quad (4)$$

其中,  $m(x'_i) = \max(p(x'_i))$  表示概率向量  $p(x'_i)$  中最大的元素. 如图 3 所示: 当  $\lambda=0.5$ ,  $H(x'_3) < H(x'_4)$ , 说明二元软约束信息熵可以有效估计某些处于决策边界样本伪标签的不确定性.

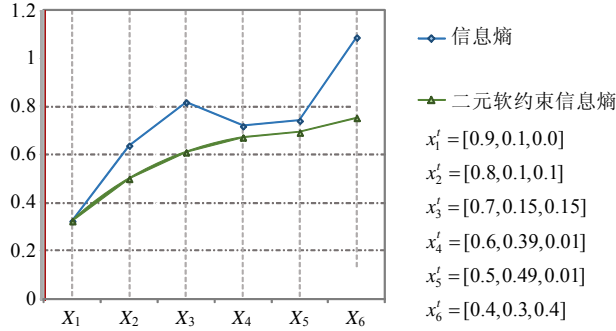


图 3

为准确估计类原型, 进一步计算每类样本的二元软约束信息熵, 得到二元软约束平均信息熵  $\bar{H}_m^k$  作为该类的类原型挑选标准.  $\bar{H}_m^k$  的计算方式为

$$\bar{H}_m^k = \frac{H_m(x'_1) + H_m(x'_2) + \dots + H_m(x'_{n_k})}{n_k} \Big|_{k \in K} \quad (5)$$

其中,  $\bar{H}_m^k$  表示第  $k$  类的二元软约束平均信息熵. 当  $H(x'_i) < \bar{H}_m^k$ , 且  $x'_i \in k$  时, 认为  $x'_i$  是该类的类原型. 但由于 *Softmax* 函数会生成一个标签过拟合输出空间<sup>[41]</sup>, 如果只使用 *Softmax* 输出的概率向量衡量样本预测结果的不确定性, 可能会存在过度自信问题. 因此, 本文利用基于 Gibbs 分布的能量函数进一步衡量样本预测结果的不确定性. 因为基于 Gibbs 分布的能量函数能够挑选分布内样本和分布外样本<sup>[42]</sup>, 故能利用能量函数计算目标域样本  $x'_i$  的负能量值  $E(x'_i)$ , 用于衡量  $x'_i$  与源域分布的相似概率. 负能量越高, 表示样本属于源域分布内的概率越大. 如果  $x'_i$  与源域相似, 可以作为类原型的候选, 实现正向迁移. 负能量值  $E(x'_i)$  的计算方式为

$$E(x'_i) = T \times \log \sum_{k=1}^K e^{\frac{f_k(x'_i)}{T}} \quad (6)$$

其中,  $f_k$  表示模型的 *Logit* 向量,  $T$  表示温度参数. 计算每类样本的负平均能量  $\bar{E}_k$  作为该类的类原型挑选标准:

$$\bar{E}_k = \frac{E(x'_1) + E(x'_2) + \dots + E(x'_{n_k})}{n_k} \Big|_{k \in K} \quad (7)$$

最后, 联合二元软约束平均信息熵和负平均能量函数挑选可靠的类原型. 熵越小, 表示属于可信样本的概率越大; 负能量越高, 表示属于源域分布内样本的概率越大. 当某样本的熵小于等于该类的二元软约束平均信息熵, 并且该样本的负能量大于等于该类的负平均能量时, 将该样本视为该类的类原型. 当  $x'_i$  满足可信样本的标准时,  $c_b=1$ ; 反之,  $c_b=0$ , 说明  $x'_i$  不是类原型. 具体挑选方式为

$$c_b = \begin{cases} 1, & \text{if } H_m(x'_i) \leq \alpha \bar{H}_m^k \text{ and } E(x'_i) \geq \beta \bar{E}_k \\ 0, & \text{if else} \end{cases} \quad (8)$$

其中,  $\alpha$ 和 $\beta$ 是超参数.

最后, 利用公式(9)存储类原型中的特征值以备计算伪标签, 利用公式(10)存储类原型中的熵均值以备过滤伪标签:

$$M_k = \{F_i(x'_i) | c_b(x'_i) = 1 \text{ and } x'_i \in k\} \quad (9)$$



$$H_{mc}^k = \frac{H_m(x_1^t) + H_m(x_2^t) + \dots + H_m(x_{n_k}^t)}{n_k} \Big|_{C_b(x_i^t) = 1 \text{ and } x_i^t \in k} \quad (10)$$

挑选类原型的具体流程见算法 1.

**算法 1.** 初始化/更新类原型.

Input: 无标签的目标数据  $D_t$ , 目标特征提取器模块  $F_t$ , 分类器模块  $C_t$ , 超参数  $\alpha, \beta$ ,

Output: 类原型  $M$ .

```

1  for  $c \leftarrow 1$  to  $num\_classes$  do
2  |  $H_m = []$ ; //Eq.4
3  |  $E_m = []$ ; //Eq.6
4  end
5  for  $t \leftarrow 1$  to  $N_t$  do
6  |  $(y_i^t, H, E) = (F_t, C_t, x_i^t)$ ;
7  |  $H_{y_i^t}.append(H)$ ;
8  |  $E_{y_i^t}.append(E)$ ;
9  end
10 for  $c \leftarrow 1$  to  $num\_classes$  do
11 |  $\bar{H}_m^k = []$ ; //Eq.5
12 |  $\bar{E}_k = []$ ; //Eq.7
13 end
14 for  $c \leftarrow 1$  to  $num\_classes$  do
15 | if  $H(x_i^t) \leq \alpha * \bar{H}_m^k$  and  $E(x_i^t) \geq \beta * \bar{E}_k$ ; //Eq.8
16 | then
17 | |  $M_k \leftarrow F_t$ ; //Eq.9
18 | | update  $\bar{H}_{mc}^k$ ; //Eq.10
19 | end
20 end
```

### 2.2.3 准确估计目标域样本的伪标签以迭代优化目标域模型

首先进行目标域样本类锚点的估计, 计算目标域中每类样本类原型特征值的均值, 作为该类的类锚点; 然后, 进一步估计目标域样本的伪标签, 从特征层面出发, 比较每个目标域样本的特征值与每类的类锚点的距离, 距离最近的类别作为该样本的伪标签, 利用类锚点对伪标签进行加权对比过滤, 进一步剔除不可信的伪标签样本; 最后, 采用迭代优化策略优化目标域模型. 具体步骤如下.

#### (1) 准确估计目标域样本伪标签

##### • 伪标签初始化

给定一个目标域样本  $x_i^t$ , 将其输入特征提取器  $F_t$ , 比较该样本的特征值和每个类的类锚点(类原型特征值的均值)的距离. 首先计算该样本与各类类锚点的相似性:

$$S_k(x_i^t) = \frac{1}{|N_k|} \sum_{p_k \in M_k} \frac{p_k^T f_t(x_i^t)}{\|p_k\|_2 \|f_t\|_2} \quad (11)$$

其中,  $S_k(x_i^t)$  表示  $x_i^t$  与第  $k$  类类原型的相似度,  $|N_k|$  表示第  $k$  类类原型的数量,  $M_k$  表示第  $k$  类类原型的特征集合,  $p_k$  表示第  $k$  类的某一个类原型,  $f_t(x_i^t)$  表示样本  $x_i^t$  特征值. 然后, 将最相似的类别作为  $x_i^t$  的伪标签  $\hat{y}_i^t$ :

$$\hat{y}_i^t = \arg \max_{k \in K} S_k(x_i^t) \quad (12)$$

##### • 加权对比过滤

一旦通过这些类原型获取了全部目标样本的伪标签, 这时本可以通过传统的交叉熵损失来训练  $F_t$  和  $C_t$ , 但基于相似度得分赋予的伪标签中含有少许噪声, 若直接使用全部的伪标签样本对网络进行训练, 依然可能会误导网络学习错误的内容. 故在给目标数据赋予伪标签时, 利用类原型对伪标签进行加权对比过滤, 进一步过滤不可信的伪标签样本. 加权对比过滤的具体过程为: 先计算目标域样本距离其所属类别类锚点的 cosine 距离  $h_1$ , 再计算其与其他所有类别类锚点的 cosine 距离, 选择最小值  $h_2$ . 在  $\gamma h_1$  小于  $h_2$  时, 选择相信此样本  $\gamma$  为超参数, 权重  $\omega$  赋予 1; 否则过滤掉此样本, 权重  $\omega$  赋予 0. 此时, 可准确得到每个目标域样本的最终伪标签, 用于优化目标域模型. 另外, 为了防止网络过滤掉太多样本, 当该样本的二元软约束信息熵小于其当前所属类的二元软约束平均信息熵时, 保留该样本. 权重  $\omega$  的具体计算方式为

$$\omega(x_i^t) = \begin{cases} 1, & \gamma h_1 \leq h_2 \text{ or } H_m(x_i^t) \leq \bar{H}_{mc}^k | x_i^t \in k \\ 0, & \text{else} \end{cases} \quad (13)$$

这里,  $\gamma$  表示加权的超参数 ( $\gamma \in (1, 2]$ ).

## (2) 迭代优化目标域模型

得到过滤后的伪标签后, 联合交叉熵损失和信息最大化损失优化目标域模型.

- 基于加权过滤的交叉熵损失

$$L_{self}(D_t) = -E_{x^t \in D_t} \omega(x^t) \sum_{k=1}^K \mathbf{1}_{[k=\hat{y}_t]} \log(\sigma(C_t(F_t(x^t)))) \quad (14)$$

其中,  $\mathbf{1}_{[k=\hat{y}_t]}$  表示  $x^t$  伪标签的 one-hot 形式.

- 信息最大化损失

为了充分利用保留的源域的分类器模块, 进一步迁移源域的知识, 受 SHOT<sup>[27]</sup> 的启发, 在基于加权过滤的交叉熵损失的基础上引入了信息最大化损失. 信息最大化损失通过促进分类结果多样性, 从而规避所有样本被分为同一类. 信息最大化损失  $L_{IM}$  包括熵最小化损失  $L_{ent}$  和类平均熵最大化损失  $L_{div}$  两部分:

$$L_{IM} = L_{ent} + L_{div} \quad (15)$$

其中,

- 熵最小化损失  $L_{ent}$  为

$$L_{ent}(f_t; D_t) = -E_{x^t \in D_t} \sum_{k=1}^K \sigma(C_s(F_t(x^t))) \log \sigma(C_s(F_t(x^t))) \quad (16)$$

- 类平均熵最大化损失  $L_{div}$  为

$$L_{div}(f_t; D_t) = \sum_{k=1}^K \hat{p}_k \log \hat{p}_k \quad (17)$$

其中,  $\hat{p}_k = E_{x^t \in D_t} [\sigma(C_s(F_t(x^t)))]$  表示所有类别为  $k$  的样本经 *Softmax* 之后的平均概率向量.

联合交叉熵损失和信息最大化损失, 总体损失函数为

$$L_{total} = \partial L_{IM} + (1 - \partial) L_{self} \quad (18)$$

其中,  $\partial$  是一个动态权衡  $L_{IM}$  和  $L_{self}$  的参数, 在训练的前期, 利用不确定性降低的伪标签样本训练模型,  $\partial$  根据迭代次数的增加而不断增加. 后期训练中, 为了不丢掉底层的源域信息, 本文利用信息最大化损失, 随着迭代次数不断增加,  $\partial$  从 0 不断增加到 1.

最后, 采用迭代优化策略优化目标域模型.

- 1) 固定模型, 联合二元软约束信息熵和能量函数挑选可靠的类原型挑选可信的样本, 更新类原型, 并计算更新后的各类的类原型平均熵;
- 2) 固定类原型, 利用更新后且更加可信的类原型给目标数据赋予伪标签, 加权对比过滤后, 利用  $L_{total}$  损失微调模型.

通过上述两个步骤不断迭代优化  $F_t$  和  $C_t$ , 调优目标模型. 具体流程如算法 2 所示.

**算法 2.** 优化过程.

**Input:** 源域分类器模块  $C_s$ , 目标网络  $(F_t, C_t)$ , 无标签的目标数据  $D_t$ ;

**Output:** 更新后的目标网络  $(F_t, C_t)$ .

```

1   $(\Theta_{F_i}, \Theta_{C_i}, \Theta_{C_s}) \leftarrow \Theta$ ; //初始化网络
2   $M \leftarrow M_k, \bar{H}_{mc}^k$ ; //算法1, 初始化类原型
3  for  $epoch\_id \leftarrow 1$  to  $max\_epoch$  do
4  |    $(\hat{y}_i^t, H) \leftarrow (F_i, C_i, x_i^t)$ ; //Eq.11; Eq.12
5  |    $\omega \leftarrow (M_k, \bar{H}_m^k)$ ; //Eq.13
6  |    $L\_total \leftarrow (x_i^t, \hat{y}_i^t, \omega, F_i, C_i)$ ; //Eq.18
7  |    $\Theta^{(t+1)} \leftarrow \Theta - \xi \nabla L\_total(MiniBatch, \Theta)$ ;
8  |    $M \leftarrow M_k, H_{mc}^k$ ; //算法1, 更新类原型
9  end

```

综上所述, PLUE-SFRDA 的核心在于获取不确定性更低的伪标签, 所提出的二元软约束信息熵、能量函数和加权对比过滤方法都是为了增强伪标签的可靠程度, 进而增强模型对于目标域数据的学习能力. 另外, 信息最大化损失的引入也有效提升了模型性能, 进一步迁移了源域知识.

### 3 实验研究

本文所提出的方法 PLUE-SFRDA 在 3 个公开数据集上与现有的标准域自适应方法和源域无关域自适应方法进行了比较. 本节介绍了实验的具体设置, 报告了 PLUE-SFRDA 方法取得的效果, 最后通过消融实验深入分析了 PLUE-SFRDA 关键模块的作用. 代码公开在 <https://github.com/sailww/SFRDA/>.

#### 3.1 实验设置

本文在 3 个无监督域自适应的公开基准数据集上评估了 PLUE-SFRDA 方法.

这 3 个数据集分别为 Office-31<sup>[43]</sup>, Office-Home<sup>[44]</sup>和 VisDA-C<sup>[45]</sup>.

- Office-31 是一个小型数据集, 包含了来自于 Amazon (A), Dslr (D)和 Webcam (W)这 3 个领域、31 个类别的图片. 3 个领域可组成 6 种迁移场景;
- Office-Home 相对来说是一个中型数据集, 其包含 Artistic (Ar), Clipart (Cl), Product (Pr)和 Real-world (Re)这 4 个领域, 每个领域含有 65 个类别图片, 可组成 12 种迁移场景;
- VisDA-C<sup>[45]</sup>相对来说是一个大型数据集, 涉及从合成图像到真实场景图像的迁移.

本文的方法与标准域无监督自适应和源域无关域自适应两类方法分别进行了比较. 对比的无监督域自适应方法(使用源域数据)包括 DAN<sup>[14]</sup>, DANN<sup>[16]</sup>, CDAN+E<sup>[23]</sup>, rRevGard+CAT<sup>[46]</sup>, SAFN+ENT<sup>[47]</sup>, CDAN+BSP<sup>[48]</sup>, CDAN+TransNorm<sup>[49]</sup>, CDAN<sup>[23]</sup>, ADR<sup>[50]</sup>和 SWD<sup>[51]</sup>; 对比的源域无关域自适应方法包括 PrDA<sup>[28]</sup>, PPDA<sup>[29]</sup>和 SHOT<sup>[27]</sup>. 除此之外, 本文还比较了只在源域数据上训练的模型(source model). 本文中比较的方法中, Source model 是单独实现的, 其余的方法均取其文献中报告的最优结果. 本文报告的结果为运行 5 次的平均结果.

参考现有的源域无关域自适应方法的标准实验设置, Office-31 和 Office-Home 利用在 ImageNet 上训练好的 ResNet50<sup>[52]</sup>作为基础骨干网络, 而 VisDA-C 利用在 ImageNet 上训练好的 ResNet101<sup>[52]</sup>作为骨干网络. 实验中, 训练图片的尺寸被重设置为 256×256, 并且使用随机水平翻转的手段将其随机裁剪成 224×224. 本文使用 SGD 作为优化器, 其重量衰减为  $5 \times 10^{-4}$ , 动量为 0.9. 预训练源域模型时, 基本的学习率设置  $5^{-2}$ . 目标模型基本的学习率设置为  $10^{-3}$ , 预先训练的特征提取器中的所有微调层的学习率设置为  $10^{-4}$ . 根据文献[16], 本文使用动态的学习率  $lr_p = lr_0(1+mz)^{-q}$ , 其中,  $lr_0$  表示基础学习率;  $z$  表示相对步骤, 在训练期间从 0 到 1 变化. 设置  $m=10$ ,  $q=0.75$ . 在预训练源域模块的标签平滑技巧中,  $\epsilon$  统一设置为 0.1; 在挖掘类原型模块, 设置  $\alpha=1.1$ ,  $\beta=0.9$ ,  $\lambda=0.5$ ; 在加权对比过滤模块, 设置  $\gamma=1.2$ . 在动态调优模型时, 设置  $\delta=2(1+\exp(-step/(max\_step/3)))^{-1}-0.5$ . 在 Office-31 数据集上, 最大迭代次数( $max\_step$ )为 5 000; 在 Office-Home 数据集上, 最大迭代次数( $max\_step$ )为 10 000; 在 VisDA-C 数据集上, 最大迭代次数( $max\_step$ )为 20 000. 在所有的数据集中, 类原型的更新周期是一个 Epoch. 本文使用 PyTorch 框架实现 PLUE-SFRDA 方法.

### 3.2 实验结果

表 1 报告了 PLUE-SFRDA、无监督域自适应方法以及最新的源域无关域自适应方法在 Office-31 数据集上的结果. PLUE-SFRDA 的平均准确率优于所有源域无关域自适应的方法: 相比 PrDA 报告的 87.2%, PLUE-SFRDA 整体提高了 1.8%; 相比 SHOT 报告的 88.6%, PLUE-SFRDA 整体提高了 0.4%. PLUE-SFRDA 在 A→D 任务上的预测结果不仅超过源域无关域自适应的方法, 而且优于无监督域自适应方法. 同时, PLUE-SFRDA 在 W→D 的迁移任务下取得了 100% 的分类准确率. 以上结果证明: 使用二元软约束信息熵和能量函数能够挑选出可信的类原型, 加权对比过滤方法可以赋予目标域样本准确的伪标签, 使得最终的域自适应的结果有所提高. 以上结果也证明: PLUE-SFRDA 能够准确地衡量伪标签不确定性、无偏差地估计目标域伪标签.

表 1 Office 31 数据集上的分类准确率(%) (ResNet50)

Method	A→D	A→W	D→A	D→W	W→A	W→D	Avg.
ResNet50 <sup>[52]</sup>	68.9	68.4	62.5	96.7	60.7	99.3	76.1
DAN <sup>[14]</sup>	78.6	80.5	63.6	97.1	62.8	99.6	80.4
DANN <sup>[16]</sup>	79.7	82.0	68.2	96.9	67.4	99.1	82.2
SAFN+ENT <sup>[47]</sup>	90.7	90.1	73.0	98.6	70.2	99.8	87.1
rRevGrad+CAT <sup>[46]</sup>	90.8	94.4	72.2	98.0	70.2	99.8	87.6
CDAN+E <sup>[23]</sup>	92.0	94.1	71.0	98.6	69.3	<b>100.0</b>	87.7
CDAN+BSP <sup>[48]</sup>	93.0	93.3	73.6	98.2	72.6	<b>100.0</b>	88.5
CDAN+TransNorm <sup>[49]</sup>	94.0	<b>95.7</b>	73.4	<b>98.7</b>	74.2	<b>100.0</b>	<b>89.3</b>
source model only	80.9	76.2	66.2	97.1	64.5	99.3	80.7
PrDA <sup>[28]</sup>	92.2	91.1	71.0	98.2	71.2	99.5	87.2
SHOT <sup>[27]</sup>	94.0	90.1	<b>74.7</b>	98.4	<b>74.3</b>	99.9	88.6
Ours w.o 对比过滤	94.1	91.5	71.6	98.4	70.2	<b>100.0</b>	87.6
PLUE-SFRDA	<b>96.4</b>	92.5	74.5	98.3	72.2	<b>100.0</b>	89.0

表 2 报告了 PLUE-SFRDA、无监督域自适应方法以及最新的源域无关域自适应方法在 Office-Home 数据集上的结果. 从表 2 可以看出: 即使在无法直接获取源域数据的限制条件下, PLUE-SFRDA 的平均准确率也显著优于源域可见的无监督域自适应方法. 源域可见域自适应达到的最优分类结果为 67.6%, 而 PLUE-SFRDA 的分类结果为 68.6%, 相较最优结果提升了 1.0%, 证明了 PLUE-SFRDA 在源域数据无法获取的情况下也能实现正向迁移. 与源域无关的域自适应方法相比, PLUE-SFRDA 显著高于 PrDA 和 PPDA 方法(分别高了 3.6%和 2.9%). 和 SHOT 相比, PLUE-SFRDA 在 Cl→Ar 和 Re→Ar 两个非常困难的任務上取得了比 SHOT 更好的结果. 这也表明 PLUE-SFRDA 从困难领域向简单数领域进行迁移时, 具有较强的鲁棒性和泛化性.

表 3 报告了 PLUE-SFRDA、无监督域自适应方法以及最新的源域无关域自适应方法在 VisDA-C 数据集上的结果. 从表 3 可以看出: 即使在无法直接获取源域数据的限制条件下, PLUE-SFRDA 的平均准确率也显著优于源域可见的无监督域自适应方法. 源域可见域自适应达到的最优分类结果为 76.4%, 而 PLUE-SFRDA 的分类结果为 78.9%, 相较最优结果提升了 2.5%, 同样证明了 PLUE-SFRDA 在源域数据无法获取的情况下也能实现正向迁移.

表 2 Office-Home 数据集上的分类准确率(%) (ResNet50)

Method	Ar→Cl	Ar→Pr	Ar→Re	Cl→Ar	Cl→Pr	Cl→Re
ResNet50 <sup>[52]</sup>	34.9	50.0	58.0	37.4	41.9	46.2
DAN <sup>[14]</sup>	43.6	57.0	67.9	45.8	56.5	60.4
DANN <sup>[16]</sup>	45.6	59.3	70.1	47.0	58.5	60.9
CDAN+E <sup>[23]</sup>	50.7	70.6	76.0	57.6	70.0	70.0
CDAN+BSP <sup>[48]</sup>	52.0	68.6	76.1	58.0	70.3	70.1
SAFN <sup>[47]</sup>	52.0	71.7	76.3	64.2	69.9	71.9
CDAN+TransNorm <sup>[49]</sup>	50.2	71.4	77.4	59.3	72.7	73.1
source model only	42.0	65.9	73.8	52.0	63.6	65.7
PPDA <sup>[29]</sup>	48.5	71.3	75.6	63.9	69.0	72.1
PrDA <sup>[28]</sup>	48.4	73.4	76.9	64.3	69.8	71.7
SHOT <sup>[27]</sup>	<b>57.1</b>	<b>78.1</b>	<b>81.5</b>	68.0	<b>78.2</b>	<b>78.1</b>
Ours w.o 对比过滤	47.6	75.1	76.4	67.2	70.4	74.6
PLUE-SFRDA	51.1	75.3	77.8	<b>69.2</b>	72.8	75.5

表 2 Office-Home 数据集上的分类准确率(%) (ResNet50) (续)

Method	Pr→Ar	Pr→Cl	Pr→Re	Re→Ar	Re→Cl	Re→Pr	Avg.
ResNet50 <sup>[52]</sup>	38.5	31.2	60.4	53.9	41.2	59.9	46.1
DAN <sup>[14]</sup>	44.0	43.6	67.7	63.1	51.5	74.3	56.3
DANN <sup>[16]</sup>	46.1	43.7	68.5	63.2	51.8	76.8	57.6
CDAN+E <sup>[23]</sup>	57.4	50.9	77.3	70.9	56.7	81.6	65.8
CDAN+BSP <sup>[48]</sup>	58.6	50.2	77.6	72.2	59.3	81.9	66.3
SAFN <sup>[47]</sup>	63.7	51.4	77.1	70.9	57.1	81.5	67.3
CDAN+TransNorm <sup>[49]</sup>	61.0	53.1	79.5	71.9	59.0	82.9	67.6
source model only	54.8	38.3	73.0	66.7	43.7	76.8	60.0
PPDA <sup>[29]</sup>	62.4	43.5	76.0	70.4	50.1	76.1	64.9
PrDA <sup>[28]</sup>	62.7	45.3	76.6	69.8	50.5	79.0	65.7
SHOT <sup>[27]</sup>	<b>67.4</b>	<b>54.9</b>	<b>82.2</b>	73.3	58.8	<b>84.3</b>	<b>71.8</b>
Ours w.o 对比过滤	65.4	47.4	76.7	72.0	51.8	81.0	67.1
PLUE-SFRDA	66.3	49.2	77.2	<b>73.4</b>	54.0	81.2	68.6

表 3 VisDA-C 数据集上的分类准确率(%) (ResNet101)

Method	plane	bycle	bus	car	horse	knife	mcycl	person	plant	sktbrd	train	truck	Per
ResNet101 <sup>[52]</sup>	55.1	53.3	61.9	59.1	80.6	17.9	79.7	31.2	81.0	26.5	73.5	8.5	52.4
DANN <sup>[16]</sup>	81.9	77.7	82.8	44.3	81.2	29.5	65.1	28.6	51.9	54.6	82.8	7.8	57.4
DAN <sup>[14]</sup>	87.1	63.0	76.5	42.0	90.3	42.9	85.9	53.1	49.7	36.3	85.8	20.7	61.1
ADR <sup>[50]</sup>	94.2	48.5	84.0	<b>72.9</b>	90.1	74.2	<b>92.6</b>	72.5	80.8	61.8	82.2	28.8	73.5
CDAN <sup>[23]</sup>	85.2	66.9	83.0	50.8	84.2	74.9	88.1	74.5	83.4	76.0	81.9	38.0	73.9
CDAN+BSP <sup>[48]</sup>	92.4	61.0	81.0	57.5	89.0	80.6	90.1	77.0	84.2	77.9	82.1	38.4	75.9
SAFN <sup>[47]</sup>	93.6	61.3	<b>84.1</b>	70.6	<b>94.1</b>	79.0	91.8	79.6	89.9	55.6	<b>89.0</b>	24.4	76.1
SWD <sup>[51]</sup>	90.8	82.5	81.7	70.5	91.7	69.5	86.3	77.5	87.4	63.6	85.6	29.2	76.4
source model only	74.0	28.5	66.4	77.5	78.6	13.1	87.4	30.8	80.6	19.0	75.4	8.8	53.3
PPDA <sup>[29]</sup>	81.5	79.4	80.3	61.8	92.3	91.9	84.5	82.7	86.5	58.4	74.2	43.5	76.4
PrDA <sup>[28]</sup>	86.9	81.7	84.6	63.9	93.1	91.4	86.6	71.9	84.5	58.2	74.5	42.7	76.7
SHOT <sup>[27]</sup>	<b>94.3</b>	<b>88.5</b>	80.1	57.3	93.1	<b>94.9</b>	80.7	<b>80.3</b>	<b>91.5</b>	<b>89.1</b>	86.3	<b>58.2</b>	<b>82.9</b>
Ours w.o 对比过滤	90.2	81.8	85.3	66.6	93.0	87.7	86.3	80.5	85.1	58.2	77.6	44.0	78.0
PLUE-SFRDA	89.4	82.7	83.5	61.2	92.8	92.1	86.0	79.8	69.0	69.0	79.6	44.1	78.9

从 Office-Home 和 VisDA-C 的实验中都可以表明: PLUE-SFRDA 从简单领域向困难的领域迁移时, 迁移能力相对来说没有 SHOT 的迁移能力强. 原因如下: 从类原型挑选的角度来讲, 从简单数据集训练得到的源域模型迁移能力相对较弱, 一开始使用源域模型为目标域样本赋予伪标签时, 伪标签含有较大比例的噪声, 导致后面挑选类原型的准确度下降, 从而导致伪标签的不确定性估计更加困难, 尽管使用二元软约束信息熵、能量函数、加权对比过滤方法能够剔除部分噪声标签, 但由于初始化伪标签的噪声率过大, 仍会有一部分错误伪标签被引入了训练集中. 但是 PLUE-SFRDA 在从简单领域向困难领域进行迁移时的表现远超过了 Source model (Office-Home 平均高了 8.3%, VisDA-C 平均高了 25.6%), 证明了 PLUE-SFRDA 估计目标域伪标签不确定性的有效性.

图 4 报告了 3 种源域无关域自适应方法 PLUE-SFRDA, SHOTs 和 PrDA 在 Office-31 数据集上 A→D 任务中的精度随着 Epoch 增加的变化情况, 表现了提出的方法和已有方法在该任务上的时间开销情况(紫色: PrDA 方法; 橙色: SHOT 方法; 蓝色: PLUE-SFRDA 方法. 横坐标表示 Epoch 数, 纵坐标表示准确率). 从算法层面来看, 本文所提方法包括两个核心算法: 挑选可信类原型和为目标域数据赋予伪标签并过滤从而优化网络. 算法 1 中需要为每一个样本计算熵和能量并进行比较, 时间复杂度为  $O(n_i)$ , 算法 2 中需要计算每一个样本和类锚点之间的距离从而确定伪标签, 并进一步过滤, 时间复杂度为  $O(n_i)$ , 总的时间复杂度为  $O(n_i)$ . 从实验效果层面来看(横坐标表示迭代的 Epoch 的数目, 纵坐标表示 A→D 的域自适应分类准确率), 可以看到: SHOT 最快收敛但效果不明显; PrDA 的实验收敛慢, 并且精度最低; 本文提出的方法效果最佳, 且在 200 个 Epoch 之内取得最优的结果.

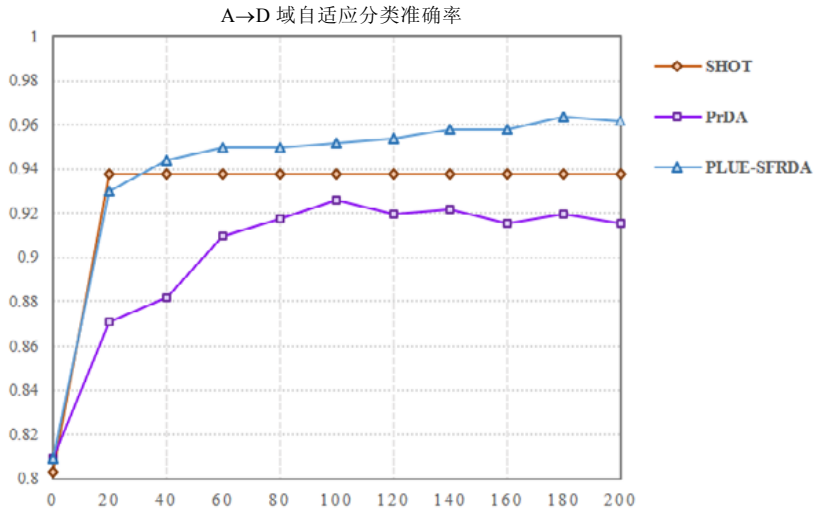


图 4 以 A→D 为例, 比较随着 Epoch 的进行, 时域自适应方法对应的准确率

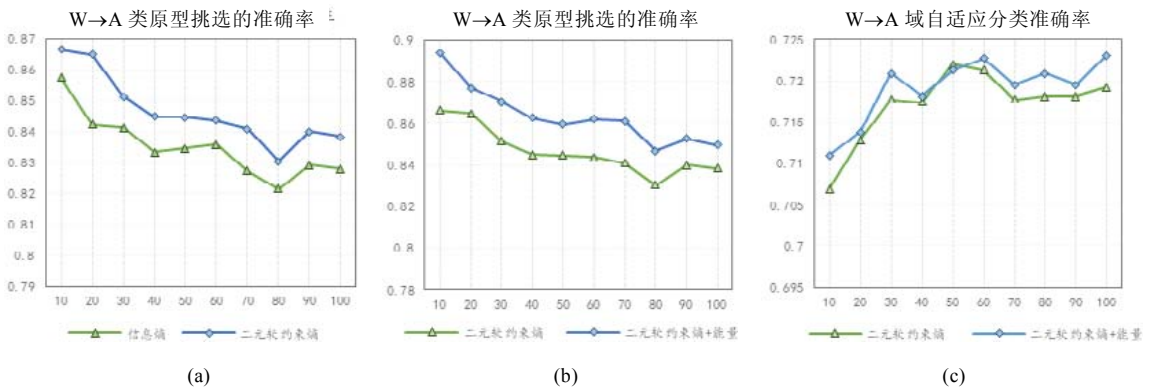
### 3.3 消融实验

本节通过消融实验, 分别分析了二元软约束信息熵、能量函数和加权对比过滤这 3 个关键模块的重要性.

二元软约束信息熵: 以 W→A 迁移任务为例, 图 5(a)报告了利用标准信息熵和二元软约束信息熵挑选类原型的对比结果. 图 5(a)显示: 随着网络训练的进行, 二元软约束信息熵挑选的类原型准确率始终比标准信息熵高 1%左右. 说明与标准信息熵相比, 二元软约束信息熵更能准确估计目标域伪标签的不确定性.

能量函数: 以 W→A 的迁移任务为例, 图 5(b)报告了使用二元软约束信息熵和能量函数两者共同作为挑选类原型的标准, 和仅仅使用二元软约束信息熵作为挑选类原型的标准时的类原型挑选准确度结果对比图. 图中显示: 当采用两者共同作为标准来挑选的可信样本时, 类原型的准确率始终比仅用二元软约束平均信息熵高 2%~3%, 说明引入能量函数能提升目标域伪标签不确定性的估计. 图 5(c)报告了使用二元软约束信息熵和能量一起约束的域自适应的分类准确率不仅稍稍有提高, 而且效果较为稳定.

加权对比过滤: 表 1~表 3 的最后两行报告了使用加权对比过滤和不使用加权对比过滤的对比结果. 使用加权对比过滤的方法在每一个迁移任务的结果上都明显优于不使用加权对比过滤的方法. 说明加权对比过滤模块可以有效过滤掉处于决策边界的类别模糊样本, 提高了目标域伪标签估计的准确率, 从而进一步提升了目标域模型的性能.



(a) 和 (b) 比较了利用标准信息熵, 二元软约束信息熵和二元软约束信息熵加能量为标准时挑选出的类原型准确率; (c) 比较了利用二元软约束信息熵和二元软约束信息熵加能量为标准时的域自适应分类准确率

图 5

## 4 结 论

本文提出了一种基于伪标签不确定性估计的源域无关鲁棒域自适应的方法 PLUE-SFRDA, 解决了无监督域自适应中一个更加现实且极具挑战的问题: 仅依据预训练的源域模型和无标签目标域数据, 实现源域向目标域的正向迁移. PLUE-SFRDA 联合二元软约束熵和能量函数挑选了可信的样本作为类原型, 其正向促进了挖掘类锚点的准确度, 从而大幅度提高了为目标数据赋予伪标签的正确率; 而进一步的加权对比过滤, 再次降低了目标域伪标签样本的不确定性. PLUE-SFRDA 联合监督损失和信息最大化损失迭代优化目标域模型, 成功实现了源域数据无关的鲁棒域自适应. 3 个公开的基准数据集上的大量实验, 证明了 PLUE-SFRDA 的有效性以及二元软约束熵和加权对比过滤的可行性.

### References:

- [1] Liu JW, Sun ZK, Luo XL. Review and research development on domain adaptation learning. *Acta Automatica Sinica*, 2014, 40(8): 1576–1600 (in Chinese with English abstract). [doi: 10.3724/SP.J.1004.2014.01576]
- [2] Fan CN, Li P, Xiao T, *et al.* A review of deep domain adaptation: General situation and complex situation. *Acta Automatica Sinica*, 2021, 47(3): 515–548 (in Chinese with English abstract). [doi: 10.16383/j.aas.c200238]
- [3] Inoue N, Furuta R, Yamasaki T, *et al.* Cross-Domain weakly-supervised object detection through progressive domain adaptation. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). 2018. 5001–5009.
- [4] Hsu H, Yao C, Tsai Y, *et al.* Progressive domain adaptation for object detection. In: Proc. of the IEEE Winter Conf. on Applications of Computer Vision (WACV). 2020. 738–746.
- [5] Gopalan R, Li R, Chellappa R. Domain adaptation for object recognition: An unsupervised approach. In: Proc. of the IEEE Int'l Conf. on Computer Vision (ICCV 2011). 2011. 999–1006.
- [6] Tsai Y, Hung W, Schuler S, *et al.* Learning to adapt structured output space for semantic segmentation. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). 2018. 7472–7481.
- [7] Ben-David S, Blitzer J, Crammer K, *et al.* Analysis of representations for domain adaptation. In: Proc. of the 19th Int'l Conf. on Neural Information Processing Systems. 2006. 137–144.
- [8] Ben-David S, Blitzer J, Crammer K, *et al.* A theory of learning from different domains. 2010, 79: 151–175.
- [9] Ben-David S, Lu T, Luu T, *et al.* Impossibility theorems for domain adaptation. In: Proc. of the 13th Int'l Conf. on Artificial Intelligence and Statistics, Vol.9. Chia Laguna Resort, 2010. 129–136.
- [10] Zhang K, Schölkopf B, Muandet K, *et al.* Domain adaptation under target and conditional shift. In: Proc. of the 30th Int'l Conf. on Machine Learning, Vol.28. 2013. III-819–III-827.
- [11] Germain P, Habrard A, Laviolette F, *et al.* A PAC-Bayesian approach for domain adaptation with specialization to linear classifiers. In: Proc. of the 30th Int'l Conf. on Machine Learning, Vol.28. 2013. 738–746.
- [12] Sugiyama M, Suzuki T, Nakajima S, *et al.* Direct importance estimation for covariate shift adaptation. *Annals of The Institute of Statistical Mathematics*, 2008, 60(4): 699–746.
- [13] Long M, Zhu H, Wang J, *et al.* Deep transfer learning with joint adaptation networks. In: Proc. of the 34th Int'l Conf. on Machine Learning (ICML 2017). 2017. 2208–2217.
- [14] Long M, Cao Y, Wang J, *et al.* Learning transferable features with deep adaptation networks. In: Proc. of the 32nd Int'l Conf. on Machine Learning (ICML 2015). 2015. 97–105.
- [15] Saito K, Ushiku Y, Harada T. Asymmetric tri-training for unsupervised domain adaptation. In: Proc. of the 34th Int'l Conf. on Machine Learning (ICML), Vol.70. 2017. 2988–2997.
- [16] Ganin Y, Lempitsky V. Unsupervised domain adaptation by backpropagation. In: Proc. of the 32nd Int'l Conf. on Machine Learning, Vol.37. 2015. 1180–1189.
- [17] Kairouz P, McMahan HB, Avent B, *et al.* Advances and open problems in federated learning. arXiv: 1912.04977, 2019.
- [18] Arpit D, Jastrzëbski S, Ballas N, *et al.* A closer look at memorization in deep networks. In: Proc. of the 34th Int'l Conf. on Machine Learning (ICML 2017). 2017. 233–242.
- [19] Huang J, Smola AJ, Gretton A, *et al.* Correcting sample selection bias by unlabeled data. In: Proc. of the 19th Int'l Conf. on Neural Information Processing Systems. 2006. 601–608.
- [20] Zhang W, Ouyang W, Li W, *et al.* Collaborative and adversarial network for unsupervised domain adaptation. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). 2018. 3801–3809.
- [21] Han L, Zou Y, Gao R, *et al.* Unsupervised domain adaptation via calibrating uncertainties. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPR Workshops 2019). 2019. 99–102.

- [22] Saito K, Watanabe K, Ushiku Y, *et al.* Maximum classifier discrepancy for unsupervised domain adaptation. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). 2018. 3723–3732.
- [23] Long M, Cao Z, Wang J, *et al.* Conditional adversarial domain adaptation. In: Proc. of the Advances in Neural Information Processing Systems 31: Annual Conf. on Neural Information Processing Systems (NeurIPS). 2018. 1647–1657.
- [24] Nelakurthi AR, Maciejewski R, He J. Source free domain adaptation using an off-the-shelf classifier. In: Proc. of the IEEE Int'l Conf. on Big Data (Big Data). 2018. 140–145.
- [25] Liang J, He R, Sun Z, *et al.* Distant supervised centroid shift: A simple and efficient approach to visual domain adaptation. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). 2019. 2975–2984.
- [26] Chidlovskii B, Clinchant S, Csurka G. Domain adaptation in the absence of source domain data. In: Proc. of the 22nd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2016. 451–460.
- [27] Liang J, Hu D, Feng J. Do we really need to access the source data? Source hypothesis transfer for unsupervised domain adaptation. In: Proc. of the 37th Int'l Conf. on Machine Learning (ICML). 2020. 6028–6039.
- [28] Kim Y, Cho D, Panda P, *et al.* Domain adaptation without source data. arXiv: 2007.01524, 2020.
- [29] Kim Y, Cho D, Hong S. Towards privacy-preserving domain adaptation. IEEE Signal Processing Letters, 2020, 27: 1675–1679.
- [30] Li R, Jiao Q, Cao W, *et al.* Model adaptation: Unsupervised domain adaptation without source data. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). 2020. 9638–9647.
- [31] Klingner M, Termöhlen JA, Ritterbach J, *et al.* Unsupervised BatchNorm adaptation (UBNA): A domain adaptation method for semantic segmentation without using source domain representations. arXiv: 2011.08502, 2020.
- [32] Stan S, Rostami M. Privacy preserving domain adaptation for semantic segmentation of medical images. arXiv: 2101.00522, 2020.
- [33] Bateson M, Kervadec H, Dolz J, *et al.* Source-Relaxed domain adaptation for image segmentation. In: Proc. of the Int'l Conf. on Medical Image Computing and Computer Assisted Intervention (MICCAI). Part I. 2020. 490–499.
- [34] Altori C, Lathuilière S, Sebe N, *et al.* SF-UDA<sup>3D</sup>: Source-free unsupervised domain adaptation for LiDAR-based 3D object detection. In: Proc. of the 8th Int'l Conf. on 3D Vision (3DV 2020). 2020. 771–780.
- [35] Li X, Chen W, Xie D, *et al.* A free lunch for unsupervised domain adaptive object detection without source data. arXiv: 2012.05400, 2020.
- [36] Wang D, Shelhamer E, Liu S, *et al.* Fully test-time adaptation by entropy minimization. arXiv: 2006.10726, 2020.
- [37] Busto PP, Iqbal A, Gall J. Open set domain adaptation for image and action recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2020, 42: 413–429.
- [38] Liang J, He R, Sun Z, *et al.* Exploring uncertainty in pseudo-label guided unsupervised domain adaptation. Pattern Recognition, 2019, 96.
- [39] Lee DD, George CS. Graph matching and pseudo-label guided deep unsupervised domain adaptation. In: Proc. of the Artificial Neural Networks and Machine Learning 2018—27th Int'l Conf. on Artificial Neural Networks (ICANN). Part III. Vol.11141. Rhodes, 2018. 342–352.
- [40] Muller R, Kornblith S, Hinton GE. When does label smoothing help? In: Proc. of the Advances in Neural Information Processing Systems 32: Annual Conf. on Neural Information Processing Systems (NeurIPS). 2019. 4696–4705.
- [41] Nguyen A, Yosinski J, Clune J. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2015. 427–436.
- [42] Liu W, Wang X, Owens JD, *et al.* Energy-Based out-of-distribution detection. In: Proc. of the Advances in Neural Information Processing Systems 33: Annual Conf. on Neural Information Processing Systems (NeurIPS). 2020.
- [43] Saenko K, Kulis B, Fritz M, *et al.* Adapting visual category models to new domains. In: Proc. of the Computer Vision 11th European Conf. on Computer Vision (ECCV), Part IV. Vol.6314. 2010. 213–226.
- [44] Venkateswara H, Eusebio J, Chakraborty S, *et al.* Deep Hashing network for unsupervised domain adaptation. In: Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2017. 5018–5027.
- [45] Peng X, Usman B, Kaushik N, *et al.* VisDA: The visual domain adaptation challenge. arXiv: 1710.06924, 2018.
- [46] Deng Z, Luo Y, Zhu J. Cluster alignment with a teacher for unsupervised domain adaptation. In: Proc. of the IEEE/CVF Int'l Conf. on Computer Vision (ICCV). 2019. 9943–9952.
- [47] Xu R, Li G, Yang J, *et al.* Larger norm more transferable: An adaptive feature norm approach for unsupervised domain adaptation. In: Proc. of the IEEE/CVF Int'l Conf. on Computer Vision (ICCV). 2019. 1426–1435.
- [48] Chen X, Wang S, Long M, *et al.* Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation. In: Proc. of the 36th Int'l Conf. on Machine Learning (ICML). 2019. 1081–1090.
- [49] Wang X, Jin Y, Long M, *et al.* Transferable normalization: Towards improving transferability of deep neural networks. In: Proc. of the Advances in Neural Information Processing Systems 32: Annual Conf. on Neural Information Processing Systems (NeurIPS). 2019. 1951–1961.
- [50] Saito K, Ushiku Y, Harada T, *et al.* Adversarial dropout regularization. In: Proc. of the 6th Int'l Conf. on Learning Representations (ICLR 2018). 2018.



- [51] Lee CY, Batra T, Baig M, *et al.* Sliced wasserstein discrepancy for unsupervised domain adaptation. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2019. 10285–10295.
- [52] He K, Zhang X, Ren S, *et al.* Deep residual learning for image recognition. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). 2016. 770–778.

**附中文参考文献:**

- [1] 刘建伟, 孙正康, 罗雄麟. 无监督域自适应学习研究进展. 自动化学报, 2014, 40(8): 1576–1600, [doi: 10.3724/SP.J.1004.2014.01576]
- [2] 范苍宁, 刘鹏, 肖婷, 等. 深度域适应综述: 一般情况与复杂情况. 自动化学报, 2021, 47(3): 515–548 [doi: 10.16383/j.aas. c200238]



王帆(1999—), 女, 硕士生, 主要研究领域为机器学习, 域自适应, 源域无关域自适应.



尹义龙(1972—), 男, 教授, 博士生导师, CCF 杰出会员, 主要研究领域为机器学习, 数据挖掘.



韩忠义(1994—), 男, 博士生, 主要研究领域为机器学习, 数据挖掘.