软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2024,35(4):1651−1666 [doi: 10.13328/j.cnki.jos.007010]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

局部一致性主动学习的源域无关开集域自适应 *

王 帆, 韩忠义, 苏 皖, 尹义龙

(山东大学 软件学院, 山东 济南 250101)

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

摘 要: 无监督域自适应在解决训练集(源域)和测试集(目标域)分布不一致的问题上已经取得了一定的成功. 在面向低能耗场景和开放动态任务环境时, 在资源约束和开放类别出现的情况下, 现有的无监督域自适应方法面临着严峻的挑战. 源域无关开集域自适应(SF-ODA)旨在将源域模型中的知识迁移到开放类出现的无标签目标域, 从而在无源域数据资源的限制下辨别公共类和检测开放类. 现有的源域无关开集域自适应的方法聚焦于设计准确检测开放类别的源域模型或增改模型的结构. 但是, 这些方法不仅需要额外的存储空间和训练开销, 而且在严格的隐私保护场景下难以实现. 提出了一个更加实际的场景: 主动学习的源域无关开集域自适应(ASF-ODA), 目标是基于一个普通训练的源域模型和少量专家标注的有价值的目标域样本来实现鲁棒的迁移. 为了达成此目标, 提出了局部一致性主动学习(LCAL)算法. 首先, 利用目标域中局部特征标签一致的特点, LCAL 设计了一种新的主动选择方法: 局部多样性选择, 来挑选更有价值的阈值模糊样本来促进开放类和公共类分离. 接着, LCAL 基于信息熵初步筛选出潜在的公共类集合和开放类集合, 并利用第一步得到的主动标注样本对这两个集合进行匹配纠正, 得到两个对应的可信集合. 最后, LCAL 引入开集损失和信息最大化损失来进一步促使公共类和开放类分离, 引入交叉熵损失来实现公共类的辨别. 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: 在少量有价值的目标域样本的帮助下, LCAL 不仅显著优于现有的源域无关开集域自适应方法, 还大幅度超过了现有的主动学习方法的表现, 在某些迁移任务上可以提升 20%.

关键词: 资源约束; 开集识别; 源域无关域自适应; 开集域自适应; 主动学习

中图法分类号: TP18

中文引用格式: 王帆, 韩忠义, 苏皖, 尹义龙. 局部一致性主动学习的源域无关开集域自适应. 软件学报, 2024, 35(4): 1651−1666. http://www.jos.org.cn/1000-9825/7010.htm

英文引用格式: Wang F, Han ZY, Su W, Yin YL. Local Consistent Active Learning for Source Free Open-set Domian Adaptation. Ruan Jian Xue Bao/Journal of Software, 2024, 35(4): 1651−1666 (in Chinese). http://www.jos.org.cn/1000-9825/7010.htm

Local Consistent Active Learning for Source Free Open-set Domian Adaptation

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

(School of Software, Shandong University, Jinan 250101, China)

Abstract: Unsupervised domain adaptation (UDA) has achieved success in solving the problem that the training set (source domain) and the test set (target domain) come from different distributions. In the low energy consumption and open dynamic task environment, with the emergence of resource constraints and public classes, existing UDA methods encounter severe challenges. Source free open-set domain adaptation (SF-ODA) aims to transfer the knowledge from the source model to the unlabeled target domain where public classes appear, thus realizing the identification of common classes and detection of public class without the source data. Existing SF-ODA methods focus on designing source models that accurately detect public class or modifying the model structures. However, they not only

* 基金项目: 国家自然科学基金(62176139); 山东省自然科学基金(ZR2021ZD15)
本文由“绿色低碳机器学习研究与应用”专题特约编辑封举富教授、俞扬教授、刘淇教授推荐.
收稿时间: 2023-05-13; 修改时间: 2023-07-07; 采用时间: 2023-08-24; jos 在线出版时间: 2023-09-11
CNKI 网络首发时间: 2023-11-24

require extra storage space and training overhead, but also are difficult to be implemented in the strict privacy scenarios. This study proposes a more practical scenario: Active learning source free open-set domain adaptive adaptation (ASF-ODA), based on a common training source model and a small number of valuable target samples labeled by experts to achieve a robust transfer. A local consistent active learning (LCAL) algorithm is proposed to achieve this objective. First of all, LCAL includes a new proposed active selection method, local diversity selection, to select more valuable samples of target domain and promote the separation of threshold fuzzy samples by taking advantage of the feature local labels in the consistent target domain. Then, based on information entropy, LCAL initially selects possible common class set and public class set, and corrects these two sets with labeled samples obtained in the first step to obtain two corresponding reliable sets. Finally, LCAL introduces open set loss and information maximization loss to further promote the separation of common and public classes, and introduces cross entropy loss to realize the discrimination of common classes. A large number of experiments on three publicly available benchmark datasets, Office-31, Office-Home, and VisDA-C, show that with the help of a small number of valuable target samples, LCAL significantly outperforms the existing active learning methods and SF-ODA methods, with over 20% HOS improvements in some transfer tasks.

Key words: research constraint; open-set recognition; source-free domain adaptation; open-set domain adaptation; active learning

近年来, 深度机器学习模型已经在多种类型的任务上取得了突破性进展[1−3]. 但是, 它们都隐式地假设了训练集和测试集来自同一分布. 当这种分布一致的假设不满足时, 在训练集上得到的模型难以在测试集上进行成功的泛化, 即模型在测试集上将会面临性能大幅度下降的风险[4,5]. 无监督域自适应作为解决训练集(源域)和测试集(目标域)分布不一致的有效手段, 已经在目标检测[6,7]、目标识别[8]、语义分割[9]等多种任务中取得了明显的成绩.

目前, 域自适应方法聚焦在资源完备和静态的环境下解决源域和目标域的分布不一致问题. 当这些方法受到资源限制等低能耗约束或被应用到开放动态任务环境中时, 比如源域数据不可直接被利用或目标域存在开放类(目标域中出现的新类别)等现实问题, 他们的鲁棒性将面临严峻挑战. 一方面, 当源域数据不可见时, 依赖于大量的源域数据辅助来设计的域自适应方法[10,11]难以被直接应用而失效; 另一方面, 当开放类出现时, 现有的域自适应方法会错误地将开放类识别成公共类(源域和目标域中都存在的类别), 此时不仅难以识别出开放类, 而且公共类的辨别效果也差. 举例说明: 由于成像设备和成像质量的不同, *A* 医院和 *B* 医院肺炎 CT 影像存在着明显的差异, 利用普通的域自适应来对齐两个医院的影像数据可以实现成功的迁移. 但是, 在资源约束和数据安全的限制下, *A* 医院难以提供所有的带标签的病例数据, 取而代之可以共享一个已经利用 *A* 医院病例数据训练好的病例模型. 当将此模型迁移到 *B* 医院时, 若 *B* 医院的病例数据中同时出现了新的肺炎病例类型(如 COVID-19), 基于 *A* 医院训练的模型不仅由于源域数据的缺失难以适应, 而且难以检测出新出现的肺炎病例类型.

源域无关开集域自适应(source free open-set domain adaptation, SF-ODA)的目的是利用源域模型而不是源域数据, 在目标域数据中存在开放类别时, 进行准确的公共类辨别和开放类检测[12], 如图 1 所示.

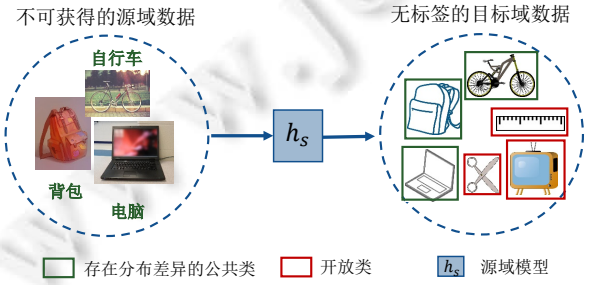


图 1 源域无关开集域自适应

SF-ODA 的核心问题是: 通过促进公共类和开放类的分离来最小化开放类数据对公共类内部辨别的影响, 从而促进鲁棒的迁移. Kundu 等人[12]对操纵源域数据生成类似开放类的样本并利用这些样本对源域模型进行训练. Liang 等人[13]和 Luo 等人[14]设计源域模型的结构并对其训练. 他们的目标都是期望获得可以识别目标域

开放类的能力的模型. 但是在严格的隐私保护限制和资源约束下, 对源域数据或源域模型进行修改设计和处理的方法在实际场景下的通用性不足. 另外, Feng 等人[15]通过增加目标域模型参数来训练开放类, 但此工作也在无形之间增加了时间和内存. 更重要的是, 现有的工作都关注于提高公共类的辨别能力, 但他们对于开放类的检测性能差. 本文提出了一个新的场景: 主动学习的源域无关开集域自适应(active learning source free open-set domain adaptation, ASF-ODA), 仅利用一个普通训练好的源域模型和少量通过专家标注的主动样本, 实现准确的公共类辨别和开放类检测.

ASF-ODA 是一个极具研究价值和挑战的问题, 其研究价值主要体现在以下 4 个方面.

(1) 满足严格的资源和隐私约束且通用性好. ASF-ODA 不需要所有的带标签的源域数据, 不需要操纵源域数据, 也不需要设计和修改源域模型的结构;

(2) 减少时间和内存. ASF-ODA 不需要对生成的额外数据训练, 也不需要增加模型结构;

(3) 现实场景的实用性. 很多工作已经将主动学习的思想引入域自适应[16−18], 通过标注少量的有价值样本来大幅度提升模型性能; 同时, 寻找领域专家对少部分样本进行标注十分合理, 比如在自动驾驶[19]、语音识别[20,21]等;

(4) 标注代价小. 对公共类样本来说, 专家需要标注其所属的具体类别, 对所有类别的开放类样本来说, 专家只需要标注其属于开放类, 大大降低了专家标注的时间成本.

其研究挑战主要体现在以下两个方面.

(1) 当普通训练的源域模型难以准确地区分公共类和开放类时, 探索哪些样本对于促进公共类和开放类分离是重要的?

(2) 如何利用挑选的主动样本实现进一步的公共类辨别和开放类检测?

为此, 我们提出了局部一致性主动学习(local consistency active learning, LCAL)算法来解决以上两个挑战.

针对问题(1), 本文发现, 挑选阈值模糊样本对于促进开放类和公共类分离是重要的. 阈值模糊样本指的是相似的开放类和公共类样本, 这些样本由于相似具备相似的熵值, 从而难以被基于阈值的方式区分. 图 2 展示了开放类和公共类样本潜在的分布情况(虚线表示分类边界), 其可以来解释阈值模糊样本产生的原因. 对公共类样本而言, 其可以被分为类似源域的公共简单类和远离源域的公共困难类[22,23]. 以信息熵为区分准则, 公共简单类(三角和方框)的样本远离分类边界且具备低信息熵值, 公共困难类(圆形)的样本靠近分类边界且具备高信息熵值. 对开放类样本而言, 理想的开放类样本(半弯月形)分布应该位于分类边界, 具备高信息熵值, 且和公共类样本不相似. 但在现实情况下, 当开放类样本(菱形)和公共简单类样本(方框)相似时, 或当开放类样本(半圆形)和公共困难类样本(圆形)相似时, 这两种情况下的相似样本难以通过设定阈值的方式被区分. 现有的主动学习方法也仅能区分部分阈值模糊样本. 基于不确定性的主动学习方法, 比如信息熵[24]、置信度[25]等, 仅能探索部分具备高不确定性的相似样本. 基于多样性的主动学习方法, 比如 *K*-means、Corset[26]等, 虽然可以关注到所有类别的信息, 但是他们挑选的样本量少, 难以有效地促进开放类和公共类分离. 用目标域局部标签一致的特点, LCAL 提出了局部多样性选择. 从探索单个样本转向探索样本局部, 从而获取到所有类别中更为全面的阈值模糊样本. LCAL 可以在有限的主动标记下扩充大量可信的标记样本, 以分离开放类和公共类.

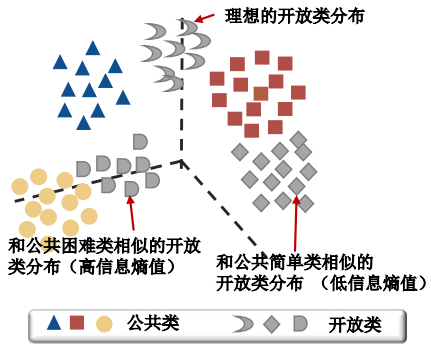


图 2 公共类和开放类样本分布分析

针对问题(2), 本文首先利用主动标注的局部区域的阈值模糊样本对潜在的公共类和开放类样本进行匹配纠正, 从而得到更为可信的公共类和开放类样本集合; 然后, 通过引入信息最大化损失和开集损失, 本文迫使模型分类器对公共类更加确定和对开放类更加不确定, 从而保证分离效果. 另外, 本文引入了交叉熵损失

来促进公共类样本的辨别性能.

本文的主要贡献包括 3 个方面.

1) 本文首次提出了一个更实际的场景: 主动学习的源域无关开集域自适应, 通过普通训练的源域模型和少量的有价值的目标域样本实现鲁棒的公共类辨别和开放类检测;
2) 本文发现挑选阈值模糊样本对促进开放类和公共类分离是重要的, 且基于目标域局部标签一致性的特点, 本文设计了局部多样性选择来挑选阈值模糊样本的区域, 从而有效地促进了开放类和公共类分离;
3) 不同数据集的实验结果表明: 我们提出的局部一致性主动学习算法可以显著提高模型的效果, 在某些迁移任务上的效果比现有的主动学习方法高 20%.

1 相关工作

1.1 无监督域自适应

无监督域自适应(unsupervised domain adaptation, UDA)的目的是将知识从大量有标注的源域中迁移到无标注的目标域中. 目前, 主流的 UDA 算法主要侧重于通过度量方法[27,28]或对抗训练方法[29,30]来对齐源域和目标域的分布. 但是在开放动态场景下, 目标域中会出现源域中没有的类别, 这些开放类的存在, 可能会造成错误的对齐, 从而大大降低域自适应的性能. 开集域自适应通过促进公共类和开放类的分离来极大地提高域自适应的效果[31−33]. 近年来, 为了显著提升目标域模型的性能, 半监督域自适应学习[34,35]和主动域自适应学习[36,37]被陆续提出, 它们都假定目标域中少量带标记的样本在训练时可以被利用. 尽管以上的场景已经获得了极大的成功, 它们在训练时需要利用所有的源域数据, 这在隐私保护的场景下不实际且难以被满足.

1.2 源域无关开集域自适应

源域无关开集域自适应(source free open-set domain adaptation, SF-ODA)的目的是利用源域模型而不是大量的源域数据, 在开放类别存在的条件下, 实现鲁棒的域自适应. 现有的 SF-ODA 方法关注于设计可以有效区分公共类和开放类的源域模型. Inheritune[12]在源域模型训练阶段, 通过对源域数据进行特征切片来构造灵活的额外样本, 同时加入了额外的分类器模块来训练这些额外样本. 在适应阶段, Inheritune 将额外分类器模型置信度较高的样本认为是开放类样本. UMAD[13]基于最大化分类器差异思想, 构造了双分类器结构的源域模型. 在源域模型训练阶段, UMAD 期望得到两个在源域上表现都很好但参数较为不同的分类器. 在适应阶段, UMAD 将两个分类器分歧较大的样本认为是开放类样本. 但是以上两个方法修改了源域数据和源域模型, 这在严格的隐私场景中和资源约束限制下通用性差. 严格来说, 我们仅可以利用一个在源域上训练好的源域模型, 而不能假定其已经具备了开放类识别的能力. 与我们设置相似的工作是 OSHT-SC[15], 但是在适应阶段, OSHT-SC 增加了目标域模型的结构, 加入了额外的内存和训练过程, 其在小设备或低能耗约束的限制下难以实现. 另外, OSHT-SC 在困难数据上对开放类别的检测性能不佳. 为此, 本文提出了一个新的场景: 主动学习的源域无关开集域自适应. 利用一个普通训练的源域模型, 在不违背隐私条件、不添加额外内存和训练时间的前提下, 仅仅在少量主动的有价值样本的代价下, 显著提升目标域模型的性能.

1.3 主动学习

主动学习(active learning, AL)的目的是, 在有限标注代价下学到一个表现性能极佳的模型. 目前, 主流的 AL 方法主要分为以下两类.

(1) 基于不确定性. 主要通过模型的输出, 比如最小置信度[25]、信息熵[24], 衡量样本的不确定性, 将不确定性较高的样本看成是模型不太确定的样本. 利用这些不确定样本对于促进明确的分类边界具有很重要的指导作用;
(2) 基于多样性. 比如 Coreset[26]期望可以获得一个可以代表整个数据集的样本集合.

但是, 对主动学习的源域无关开集域自适应来说, 有价值的样本需要落入阈值模糊样本的区域, 基于不

确定性或多样性方法挑选的样本难以实现以上目的. 基于不确定性的方式仅能促进相似且不确定性高的公共类和开放类分离, 不确定性低的且相似的公共类和开放类样本难以被探索到. 基于多样性的方法探索的样本数量较少, 难以探索到全面的相似样本来促进开放类和公共类分离. 为此, 我们设计了局部多样性选择, 利用样本局部区域标签一致的特点, 从样本选择转向区域选择, 从而选择更全面的样本, 尽量覆盖阈值模糊样本区域, 来促进相似的公共类样本和开放类样本分离. 在利用主动标注样本时, 本文还设计了不同的损失函数对这进一步促进开放类和公共类分离.

2 方 法

本节主要对本文中涉及到的问题和核心方法进行详细描述, 首先对主动学习的源域无关开集域自适应问题进行形式化定义, 然后对局部一致性主动学习算法进行详细阐述.

2.1 基本定义

本文研究的重点是: 如何通过源域模型挑选和利用有价值的主动样本, 从而有效地促进公共类和开放类分离的同时, 保证公共类样本的辨别能力. 对所要解决的问题进行形式化定义: 在主动学习的源域无关开集域自适应任务中, 我们可以获得一个已经利用源域数据 $D_s=\{x_i^s,y_i^s\}_{i=1}^{n_s}$ 经过普通训练得到的源域模型 h_s 和由 n_t 个无标签数据组成的目标域 $D_t=\{x_i^t\}_{i=1}^{n_t}$. D_s 和 D_t 来自两个相似但不同的分布, 且 D_s 在目标域适应时不可利用. $C_s \ni y_i^s$ 和 $C_t \ni y_i^t$ 分别表示源域和目标域的标签集. 在开集域自适应中, C_s 是 C_t 的子集, C_s 中包含的类别称为公共类, C_t 中存在但是 C_s 中没有的类别称为开放类, 即 $\overline{C}_t=C_t\setminus C_s$. 本文将少部分需要标注的有价值的样本定义为主动样本 $D_t^l=\{x_i^t,y_i^t\}_{i=1}^{m_l}$, 其中, $n_l=\beta n_t$ 表示主动样本的数量, β 表示主动样本的比例. 剩下的大量的无标签目标域样本被定义为 D_t^U. 源域模型包含普通的两阶段结构: 一个特征提取器和一个分类器. 在少量的主动样本 D_t^L 的帮助下, 主动学习的源域无关开集域自适应的目标是: 对公共类类别 C_s 中样本进行细分且将所有开放类别 $\overline{C}_t$ 的样本识别为‘未知’, 即同时实现公共类辨别和开放类检测.

2.2 局部一致性主动学习算法

局部一致性主动学习算法主要关注和解决以下两个问题: (1) 探索哪些样本对于促进公共类和开放类的分离是重要的? (2) 利用(1)中挑选的重要的主动标注样本, 在促进公共类和开放类的分离的同时保证公共类的辨别能力. 在探索主动样本中, 我们首先发现挑选阈值模糊样本是重要的. 然后, 设计局部多样性选择算法来探索这些阈值模糊样本. 在利用主动样本中, 本文引入开集损失和信息最大化损失来促进公共类和开放类样本进一步分离, 引入交叉熵损失来保证公共类样本内部的辨别效果.

2.2.1 探索主动样本点(见算法 1 第 2–8 行)

由于给定的源域模型仅仅是通过普通训练得到的, 其不具备区分公共类别和开放类别的能力. 参考现有的开集域自适应的工作[15,38], 基于模型输出的不确定性(比如信息熵), 本文首先可以设定阈值来评估样本属于公共类别还是开放类别. 由于源域模型是由和目标域相似的公共类样本训练得到的, 所以公共类样本的信息熵值普遍会比开放类别的信息熵值小. 具体来说, 基于每一个样本通过模型输出得到的信息熵值 H 和提前设定的阈值 w_0, 所有样本可以被分为两部分: 潜在的公共类集合 $D_{pc}=\{x_i|H(x_i)<w_0\}$ 和潜在的开放类集合 $D_{pp}=\{x_i|H(x_i)\geqslant w_0\}$(如图 3 所示). 但是由于域差异和开放类样本的存在, 模型输出的不确定性会由于未经校准和处理而变得不可信[39], 从而导致这两部分集合中往往存在较大的噪声, 即公共类集合中 D_{pc} 包含很多信息熵值小的开放类样本. 同样, 开放类集合 D_{pp} 中也会包含很多信息熵值大的公共类样本. 所以, 基于阈值划分的方式难以有效地分离熵值小的开放类样本和熵值大的公共类样本.

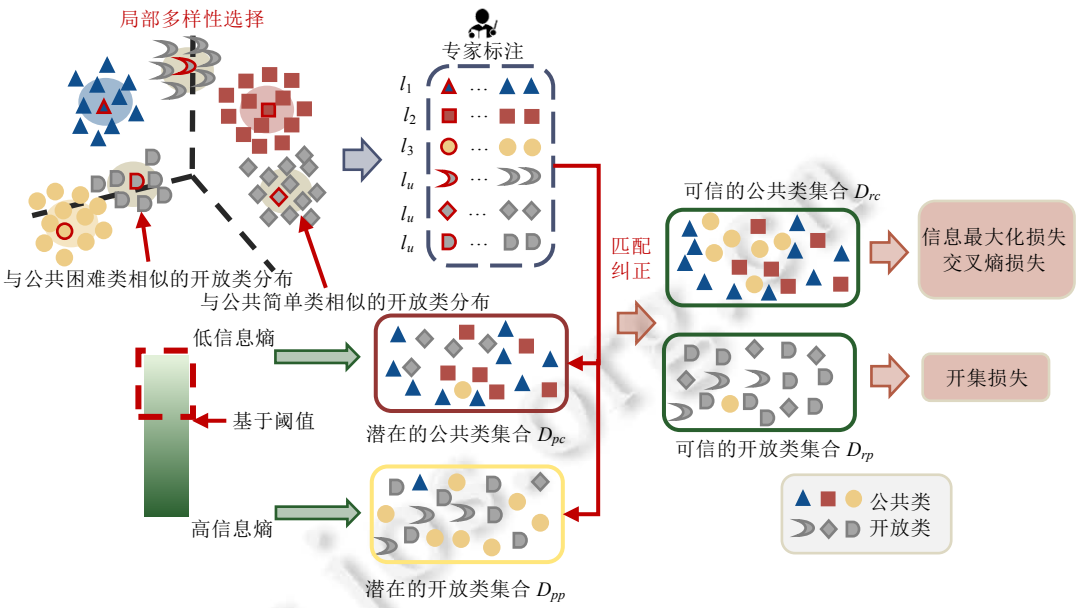


图 3　局部一致性主动学习算法的框架

- 阈值模糊样本

进一步分析，目标域中所有的公共类样本可以被分为两部分：类似源域的简单类样本 D_e 和远离于源域的困难类样本 D_h[23,40]. D_e 中包含的样本的熵值小，D_h 中包含的样本的熵值大. 如图 2 和图 3 所示：当开放类样本和 D_e 中样本相似时，这些开放类样本以较大的概率远离分类边界且熵值较小，基于阈值分离时，他们很容易落入公共类集合 D_{pc}；当 D_h 中样本和开放类别样本相似时，这些公共类样本以较大的概率靠近分类边界且熵值较大，基于阈值分离时，他们很容易落入开放类集合 D_{pp}. 如上所述，当开放类样本和公共类样本相似时，这两部分样本会具备相似的熵值，从而难以被基于阈值的方式有效分离. 本文将这些相似的样本定义为阈值模糊样本.

现有的基于不确定性和基于多样性的主动学习方法难以针对性地分离这些阈值模糊样本：基于不确定性的方法仅仅关注模型不太确定的样本，他们可以促进熵值比较大的公共类样本和开放类样本分离，而难以探索到熵值较小且相似的公共类样本和开放类样本；基于多样性的方法可以通过探索全局信息的方式挖掘到所有类别的相似的公共类和开放类样本，但是在主动样本数量较少的限制下，他们也仅能关注到相似的样本中少量的代表性样本，难以促进公共类和开放类的有效分离.

本文提出了一种新的主动选择的方式：局部多样性选择，利用局部标签一致的特点. 从探索主动样本转向探索主动样本局部区域，从而挖掘更全面，更具代表性的相似的公共类和开放类样本来促进有效的分离. 局部多样性选择基于局部标签一致的发现：即使模型难以有效地识别样本是属于开放类还是公共类，但是开放类和开放类样本在特征层面上依然会形成干净的簇[41,42]，如图 4 所示(彩色表示公共类样本，灰色表示开放类样本). 本文认为，在每一个聚类的簇中，从探索单个样本信息转为探索样本局部区域的信息，可以获得更多更全面的样本，这些样本有更大的几率落入阈值模糊样本区域. 具体来说，局部多样性选择包含以下几个步骤.

(1) 在特征层面，基于普通的 K-means 先将所有的目标域样本聚成 K 个类别；

(2) 取每个类别的中心作为锚点；

(3) 按照锚点的信息熵值进行排序，从熵值从高到低进行排序，并让专家从高到低给锚点进行标注. 在专家标注有限的情况下，标注的数量可能小于锚点的数量，排序的操作希望让专家标注信息熵值较大的锚点所在的聚类；

(4) 以锚点为中心, 以 cosine 为距离度量准则, 将距离锚点最近的 N 个邻居看成标签一致的局部区域组, 并将每一个锚点的标签赋给它当前局部组内的所有样本. 这样, 通过少量有限的标注, 可以得到更多更全面的主动样本集合 D^L.

需要注意的是, 在标注过程中, 对于公共类样本, 专家需要标注具体的类别(l_1,...,l_n); 对于开放类样本, 专家只需要将其标注为未知类别(l_u), 大大降低了专家标注的工作量(见算法 1 第 2–7 行).

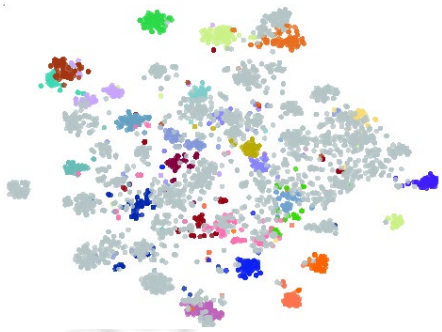


图 4 基于源域模型的目标域特征分布

得到专家标注的所有主动样本集合后, 我们对初筛得到的潜在公共类 D_{pc} 和开放类 D_{pp} 样本集合进行匹配纠正, 得到对应的可信集合. 具体来说, 对任何一个主动样本 x_i: (1) 如果其属于潜在的公共类, 同时主动标注也属于公共类, 则不做任何操作; (2) 如果其属于潜在的公共类, 但主动标注其属于开放类, 则将这个样本从公共类集合 D_{pc} 移动到 D_{pp} 中; (3) 如果其属于潜在的开放类, 同时主动标注也属于开放类, 则不做任何操作; (4) 如果其属于潜在的开放类, 但主动标注其属于公共类, 则将这个样本从开放类集合 D_{pp} 移动到公共类集合 D_{pc} 中. 经过匹配纠正之后, 原本潜在的公共类和开放类集合中的大部分阈值模糊样本会被纠正为其所属的正确集合, 从而可以得到可信的公共类集合 D_{rc} 和可信的开放类集合 D_{rp}(图 3)(算法 1, 第 8 行).

2.2.2 利用主动样本点(见算法 1 第 9–12 行)

在获得可信的公共类集合和开放类集合后, 如何利用这两部分可信样本进一步促进公共类和开放类的分离和公共类的内部辨别, 是利用主动样本点的目标. 如图 5 所示: 适应前, 阈值模糊样本难以被基于阈值的方式分开, 但随着适应过程的进行, 本文希望模型分类器对开放类别的样本拥有更大的不确定性, 公共类别的样本拥有更小的不确定性. 此时, 通过模型输出的熵值可以有效地进行对两类样本进行区分.

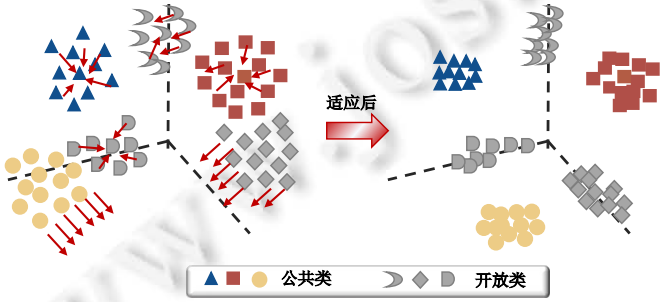


图 5 适应前后公共类样本和开放类样本的潜在分布

对于开放类别的所有样本 D_{rp}, 本文希望模型对这部分样本的输出越来越不确定, 即输出的熵值越来越大. 所以, 本文引入了现有的开集工作中[13]常用开集损失来对其进行训练:

$$L_{unk} = -E_{x_t \in D_{rp}} \sum_k \frac{1}{K} \log p_k(x_t) \tag{1}$$

其中, p_k 表示模型经过 *softmax* 输出的第 k 个元素的向量. 此损失的目的是使得模型对于开放类样本输出的概

率分布在每一个类别上趋于平均分布. 对于开放类样本而言, 若模型输出的概率分布趋于均匀分布, 那么模型对这部分样本输出的熵值很大, 进一步增强了模型分类器对于开放类样本不确定性.

对于公共类别的所有样本 D_{rc}, 本文希望模型对这部分样本的不确定性显著降低的同时, 还需要对其准确内部辨别. 为了识别公共类样本, 本文引入了交叉熵损失对这部分公共类样本进行学习. 样本的伪标签通过聚类的方式获得. 参考 SHOT[43], 本文对所有的公共类样本 D_{rc} 进行特征层面的聚类, 并得到所有样本的伪标记, 并对这些样本采用自监督的方式进行学习. 值得注意的是, 在可信的公共类样本集合 D_{rc} 中包含一部分可信的、由专家标注得到的样本 D_t^L, 这部分样本的标记相比于别的样本来说更为可靠, 故本文对这部分样本赋予更大的权重, 对剩下的公共类样本 D_t^{PL} 赋予偏小的权重:

$$L_{kn}=-E_{x_t\in D_t^L}\sum_{k=1}^{K}l_k\log p_k(x_t)-E_{x_t\in D_t^{PL}}\sum_{k=1}^{K}\alpha l_k\log p_k(x_t) \tag{2}$$

其中, l_k 表示标记向量, l_k 中真实标记处为 1, 其余位置为 0; α 表示对剩余的伪标记的公共类样本赋予的权重, α<1.0.

另外, 为了进一步促进模型对公共类样本的不确定性显著降低, 我们引入了现有的信息最大化损失[43]来促进模型对公共类样本输出的熵值不断减小:

$$L_{im}=-E_{x_t\in D_{rc}}\sum_{k=1}^{K}p_k(x_t)\log p_k(x_t)-E_{x_t\in D_{rc}}\sum_{k=1}^{K}KL(\hat{p}_k\,\|\,q_k) \tag{3}$$

其中, $\hat{p}_k=(1/m)\sum p(x)^{(k)}$ 表示所有公共类样本(数量为 m)第 k 类的样本的平均概率输出的值, $q_{\{k=1,\ldots,K\}}=1/K$ 表示均匀分布. 公式(3)的第 1 项保证了熵最小化, 在训练过程中可以促进模型对这部分公共类的不确定性显著降低, 熵值越来越小, 以此来促进模型分离公共类和开放类. 公式(3)的第 2 项 $KL(A\|B)$损失的目标是希望在学习过程中, 分布 A 不断靠近分布 B. 在本文中, 此损失希望模型输出的每一类的概率分布可以趋向均匀分布, 以防止模型在学习过程中由于过于偏向某些类而对其他类的学习性能显著下降, 即防止退化解问题的出现. $KL(\cdot)$作为一个正则化项, 已被大多数域自适应工作采用来防止这一现象.

联合开放类样本和公共类样本的损失, 整体的训练损失为

$$L_{total}=L_{kn}+L_{im}+\eta L_{unk} \tag{4}$$

其中, η表示权重.

- 测试过程

当模型达到最大训练次数后, 本文利用模型输出的信息熵来进行开放类和公共类区分. 此时, 公共类样本的不确定性越来越小, 开放类样本的不确定性越来越大, 故基于设定的平均阈值来区分开放类和公共类样本的效果对设定的阈值不敏感. 对于公共类, 直接用模型输出的类别评估公共类的辨别效果.

综上所述, LCAL 的核心在于促进开放类和公共类样本的分离. 局部多样性选择的提出, 是为了获取有价值的、且更多更全面的阈值模糊样本. 针对不同类型样本的损失也是为了进一步的促进公共类和开放类的分离, 交叉熵的引入则保证了公共类样本的辨别能力. 整体的算法流程见算法 1.

算法 1. 局部一致性主动学习算法.

输入: 源域模型 h_s, 无标签的目标域数据 D_t, 最大的训练轮次 $Epoch$, 超参数 α, β, η;

输出: 训练后的目标域模型 h_t.

1: 用源域模型参数初始化目标域模型 h_t;

第 2.2.1 节探索主动样本点

2: 对所有目标域数据的特征进行 K-means 聚类;

3: 将每个聚簇的类中心看作锚点;

4: 对锚点按照熵值顺序, 在有限标注下, 让专家为锚点赋予标记, 并将锚点放入主动样本的集合;

5: 让 $epoch$=1, $iter\_num$=0;

6: **While** $epoch\leqslant E$ **do**

7: 将 cosine 作为度量准则, 计算得到每一个锚点最近的 N 个邻居, 自动赋予邻居和锚点一样的标记,

并放入主动样本的集合;

8: 基于阈值, 得到潜在的公共类和开放类合. 利用主动样本的集合来对这两个集合进行匹配纠正, 得到可信的公共类和开放类集合;

9: 对可信公共类样本聚类, 得到可信的伪标签;

10: **While** *iter_num*<n_b **do** //n_b表示所有样本计算出来的批次总数, n_b=样本数量/*batch_size*

第 2.2.2 节利用主动样本点

11: 对开放类样本利用公式(1)计算开集损失, 对公共类样本利用公式(2)和公式(3)计算交叉熵损失和信息最大化损失;

12: 利用整体的损失(4)训练模型 h_t;

13: **end while**

14: **end while**

3 实验与结果

本节通过对比实验, 从多方面验证了本文提出方法的有效性. 实验部分将按照数据集、基准方法、实验细节和衡量指标、实验结果与分析展开介绍. 代码公布在 https://github.com/fanwang826/LCAL.

3.1 数据集

本文在 3 个无监督域自适应的公开基准数据集上评估了 LCAL 算法, 这 3 个数据集分别为 Office-31[44], Office-Home[45]和 VisDA-C[46]. Office-31 是一个标准的小型域自适应数据集, 包含了来自 Amazon(A), Dslr(D)和 Webcam(W)这 3 个 office 环境领域, 31 个类别, 共 4 110 张图片. 3 个领域可以组成 6 种迁移场景. 本文划分前 10 类作为公共类别, 后 11 类作为开放类别. Office-Home 相对是一个中型且具备挑战的数据集, 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, 共 15 588 张图片, 可以组成 12 种迁移场景. 本文划分前 25 类作为公共类别, 后 40 类作为开放类别. VisDA-C 是一个极具挑战的大型数据集, 共有 12 类. 源域包含 15.2 万张通过渲染 3D 模型生成的合成(S)图像, 而目标域包含 5.5 万张从 Microsoft COCO 采样的真实(R)物体图像. 本文划分前 6 类作为公共类别, 后 11 类作为开放类别. 以上 3 个数据集开放类和公共类的划分方式参考了现有开集域自适应工作 UMAD[13].

3.2 基准方法

本文将 LCAL 算法与目前需要额外数据和额外训练的源域无关开集域自适应方法和目前主流的主动学习的方法分别进行了比较. 对比的源域无关开集域自适应方法包括: (1) Inheritune[12], (2) OSHT-SC[15]; 和 (3) UMAD[13]. 对比的主动学习的方法包括: (1) Random: 随机选择样本作为主动样本; (2) Least Confidence (LC)[25]: 选择具备最小的模型预测输出概率样本作为主动样本; (3) Entropy[24]: 选择具备最大的信息熵的样本作为主动样本; (4) Best-Versus-Second-Best (BVSB)[47]: 选择最大的两个输出概率的差值最小的样本作为主动样本; (5) *K*-means: 对目标域样本进行 *K*-means 聚类, 并挑选每类的类中心作为主动样本; (6) Coreset[26]: Coreset 将主动样本的选择过程看成一个 set-cover 问题, 本文复现了 Coreset 提供的官方代码; (7) Batch Active Learning by Diverse Gradient Embeddings (BADGE)[48]: BADGEE 在梯度嵌入中执行 *K*-means++算法, 来挑选保证分散批次的主动样本, 本文复现了 BADGE 提供的官方代码.

3.3 实验细节和衡量指标

参考现有的源域无关开集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 ResNet 50 作为基础骨干网络, VisDA-C 利用在 ImageNet 上训练好的 ResNet101 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为 224×224. 参考 SHOT[43], 在源域模型的训练过程中, 本文引入了标签平滑技巧. 本文使用 SGD 作为优化器, 其重量衰减值为 5×10^{-4}, 动量为 0.9. 在整体模型的训练过程中, Office-31 和 Office-Home 数据集上的学习率被设置为 $1e^{-2}$, VisDA-C 上

的学习率被设置为 $1e^{-3}$. 参考文献[30], 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 m=10, q=0.75. 在总体训练损失来动态调优模型时, α=0.3, $\eta=2/(1+e^{-step/max\_step})-0.5$, $\eta\in[0.5,1.0)$. *step* 表示当前迭代次数, *max_step* 表示最大迭代次数. 随着 *step* 接近最大轮次, η从 0.5 接近于 1.0, 对开放类别赋予的训练权重也越来越大. 轮次 *Epoch* 设置为 15, *K*-means 中的 $K=\alpha\times n_t$, 邻居数量 N=15. 参考主动域自适应工作[49], 本文报告了主动样本比例β=0.05 上的结果. 另外, 本节在消融实验部分也报告了β=0.01,0.03,…,0.1 等比例的结果.

- 衡量指标

参考现有的开集域自适应方法[13,50], 本文选择 *H-score* (*HOS*)作为模型的衡量指标, *H-score* 的计算公式:

$$HOS=\frac{2\times acc_{kn}\times acc_{un}}{acc_{kn}+acc_{un}} \tag{5}$$

其中, acc_{kn} 表示公共类的每一类的识别准确率, acc_{un} 表示开放类的识别准确率. *H-score* 表示公共类识别准确率和开放类识别准确率的调和平均值, *HOS* 越大, 表示模型对于开放类和公共类的分离效果和对于公共类的辨别效果都达到最优.

3.4 实验结果

3.4.1 实验结果

表 1 报告了 LCAL, SF-ODA 方法以及最新的主动学习方法在 Office-Home 数据集上的结果. 一方面, 在不修改源域模型、不涉及多余的模型参数和训练成本外, 仅利用 5%的主动样本, LCAL 在所有的任务上均可以大幅度地提高模型对于公共类的识别能力和对开放类的辨别能力. 在平均水平上, 相比于 OSHT-SC 报告的 36.6%, LCAL 算法的效果提高了 39%; 相比于 UMAD 报告的 66.4%, LCAL 算法的效果提高了 9.5%. 以上结果表明, 将主动学习的思路融入源域无关开集域自适应可以有效提高模型的效果. 另一方面, LCAL 算法的效果远远超过了现有所有的主动学习方法的表现. 如表 1 所示, LCAL 比现有的方法高 10 个百分点左右. 值得注意的是: LCAL 和现有方法的不同仅仅体现在主动样本的挑选过程中, 其余的训练过程均一样. 在 office-home 中, 从 Art(Ar)任务到 Clipart(Cl)任务迁移是一个比较难的过程. 如表 1 中结果所示, 现有的开集域自适应工作对此任务的效果很差. 但是, 通过挑选少量有价值的主动样本后, 本文的工作可以显著提高模型的效果. 这进一步证明了本文的方法可以缓解困难任务的迁移问题. 再如, 从 Art(Ar)任务到 Product(Pr)任务迁移是一个比较容易的过程. 现有的开集域自适应工作对此任务的效果已经很好. 在这种简单任务上, 本文方法依然可以取得最优的结果. 以上说明, 本文的工作在简单和困难的迁移任务上都比现有的开集域自适应工作和主动学习方法更加有效. 表 1 中可以观察到, 使用现有的主动学习方法(如 Entropy 或 LC)挑选样本带来的效果甚至略低于随机挑选样本带来的效果, 因为这些不确定性的方法仅仅可以关注不确定性较高区域样本的分离.

表 1 5%主动样本标注后, Office-Home 数据集(ResNet50)上的 *HOS* (%)

类别	方法	Ar→Cl	Ar→Pr	Ar→Re	Cl→Ar	Cl→Pr	Cl→Re	Pr→Ar	Pr→Cl	Pr→Re	Re→Ar	Re→Cl	Re→Pr	Avg.
None	Source	53.7	65.3	72.0	55.3	61.2	65.4	56.6	47.9	66.9	65.3	50.7	64.4	60.4
SF-ODA	OSHT-SC#	40.9	32.2	40.8	30.6	23.8	24.2	49.8	31.8	40.2	31.3	46.8	46.1	36.6
	UMAD#	59.2	71.8	76.6	63.5	69.0	71.9	62.5	54.6	72.8	66.5	57.9	70.7	66.4
ASF-ODA	Random	57.6	70.2	77.0	59.0	69.0	71.6	60.7	55.6	71.7	68.3	57.8	70.8	65.9
	LC	58.1	68.7	76.3	59.5	67.6	71.2	60.2	55.7	71.1	68.0	57.2	69.2	65.2
	Entropy	57.4	69.5	76.3	58.5	67.7	71.7	59.1	55.9	71.2	68.2	56.0	69.5	65.1
	BVSB	59.3	69.2	76.3	59.0	67.8	72.0	60.7	55.7	71.4	67.6	56.6	70.2	65.5
	K-means	58.2	70.2	77.0	59.7	69.1	71.6	61.2	58.4	72.0	68.6	57.1	71.1	66.2
	Coreset	59.3	70.6	76.7	60.7	68.6	71.7	61.2	56.5	71.6	68.6	58.1	70.3	66.2
	Badge	58.8	70.0	77.4	61.4	68.8	72.2	61.2	56.6	72.0	68.1	57.2	70.5	66.2
	LCAL(Ours)	**69.7**	**84.2**	**80.6**	**66.8**	**84.5**	**79.3**	**67.6**	**69.1**	**81.2**	**72.9**	**69.1**	**82.3**	**75.6**

注: #表示结果来自 Liang 等人[13]

以上结果进一步表明: 与现有的主动学习方法相比, 使用局部多样性选择可以挑选出更全面、更具代表性阈值模糊样本, 从而在训练过程中可以有效地促进开放类和公共类的分离.

表 2 报告了 LCAL, SF-ODA 方法以及最新的主动学习方法在小型数据集 Office-31 和大型数据集 VisDA-C 上的结果. 在 Office-31 上, 在不修改源域模型、不涉及多余的模型参数和训练成本外, LCAL 的效果比 Inheritune 高 4.6%, 比 UMAD 高 1.4%; 同时, LCAL 在 D→W 和 W→D 这两个迁移任务上获得了最好的识别效果. 相比之下, LCAL 的效果低于为开放类专门设计目标域模型结构的方法 OSHT-SC, 但 OSHT-SC 方法的效果不仅依赖于增加模型参数的数量, 而且增加了额外的存储空间和训练过程. 在和 LCAL 公平比较的 ASF-ODA 的主动学习方法中, LCAL 的效果远远超过了现有主动学习方法, 最高超过了 10%. 在 VisDA-C 上, LCAL 的效果大幅度领先于现有的所有 SF-ODA 方法和目前所有的主动学习的方法. 具体来说, 和现有最好的 SF-ODA 的方法 UMAD 相比, LCAL 提升了 9%; 同时, LCAL 普遍比现有的主动学习方法高 20%. 进一步证实了: 利用本文提出的局部多样性选择的方法, 可以挑选少量却有价值的阈值模糊样本, 从而进一步有效地促进开放类区分能力的同时保证公共类样本的辨别.

表 2 5%主动样本标注后, Office-31 数据集(ResNet50)和 VisDA-C 数据集(ResNet101)上的 *HOS* (%)

类别	方法	A→D	A→W	D→A	D→W	W→A	W→D	Avg	VisDA-C
None	Source	66.5	67.9	71.5	93.0	70.2	90.7	76.6	45.2
SF-ODA	Inheritune#	78.0	81.4	83.1	92.2	91.3	**99.7**	87.6	74.8
	OSHT-SC#	**91.3**	**92.4**	**90.8**	95.2	**89.6**	96.0	**92.5**	78.6
	UMAD#	88.5	84.4	86.8	95.0	88.2	95.9	89.8	80.2
ASF-ODA	Random	74.1	80.0	77.1	97.1	76.3	90.7	82.6	69.1
	LC	72.7	77.7	75.8	95.7	76.0	88.8	81.1	68.8
	Entropy	72.7	75.6	76.4	95.5	76.1	88.8	80.9	66.0
	BVSB	72.9	78.4	75.5	95.7	77.0	88.8	81.4	68.8
	K-means	73.3	78.2	76.8	95.9	78.0	92.3	82.4	69.3
	Coreset	75.5	78.2	76.8	95.5	76.4	91.6	82.3	65.7
	Badge	71.8	77.2	77.6	96.2	78.5	88.8	81.7	69.3
	LCAL(Ours)	88.7	87.6	86.8	**97.7**	87.2	98.9	91.2	**89.2**

注: #表示结果来自 Liang 等人[13]

3.4.2 消融分析

- 模型结构

为了进一步验证 LCAL 挑选样本的有效性, 本文在不同的模型结构(VGG16)上开展了 Office-Home 的实验. 表 3 比较了 LCAL 算法和现有的主动学习方法基于 Office-Home 数据集和 VGG16 网络结构的结果. 如表 3 所示: 相比源域模型在目标域上的识别效果 55.5%, 本文的 LCAL 算法在 5%的主动标记的帮助下, 提升了 16.5%; 相比现有的主动学习方法, LCAL 算法基本上均提升了 12%. 值得注意的是: 利用现有的主动习方法, 比如 LC 和 Entropy, 带来的效果仍然低于随机挑选的方法, 说明仅促进不确定性高的开放类和公共类分离并不能有效地促进整个数据集上的公共类和开放类的分离. 以上所有结果表明: LCAL 算法可以挑选出更加有价值的主动样本, 在这些主动样本的帮助下, LCAL 不仅有效地促进开放类别和公共类别的分离, 还显著提高了公共类样本的辨别能力.

表 3 5%主动样本标注后, Office-Home 数据集(VGG16)上的 *HOS* (%)

类别	方法	Ar→Cl	Ar→Pr	Ar→Re	Cl→Ar	Cl→Pr	Cl→Re	Pr→Ar	Pr→Cl	Pr→Re	Re→Ar	Re→Cl	Re→Pr	Avg.
None	Source	46.2	61.6	67.1	52.6	56.5	60.7	53.4	40.0	64.0	60.1	42.3	61.2	55.5
ASF-ODA	Random	50.1	66.9	71.0	55.9	59.8	62.8	55.3	44.5	70.0	63.7	45.3	66.1	59.3
	LC	48.8	65.9	69.4	54.7	57.6	61.5	54.9	43.8	69.0	61.0	46.6	64.8	58.2
	Entropy	52.6	66.8	68.9	54.8	58.3	62.4	54.3	46.0	68.1	62.1	47.2	65.2	58.9
	BVSB	47.8	65.4	69.6	56.1	59.5	62.6	55.7	44.5	70.1	62.0	48.5	66.1	59.0
	K-means	51.9	67.1	70.7	58.4	60.0	64.1	56.7	46.3	70.3	63.4	47.4	67.7	60.3
	Coreset	53.4	66.5	70.2	56.2	58.8	62.7	56.5	43.9	70.1	63.2	47.3	66.9	59.6
	Badge	52.1	67.6	70.8	57.1	59.8	63.8	55.7	43.4	70.5	63.4	45.0	65.9	59.6
	LCAL(Ours)	**66.2**	**83.1**	**76.2**	**63.6**	**82.7**	**74.4**	**63.7**	**63.5**	**78.0**	**68.3**	**62.0**	**81.7**	**72.0**

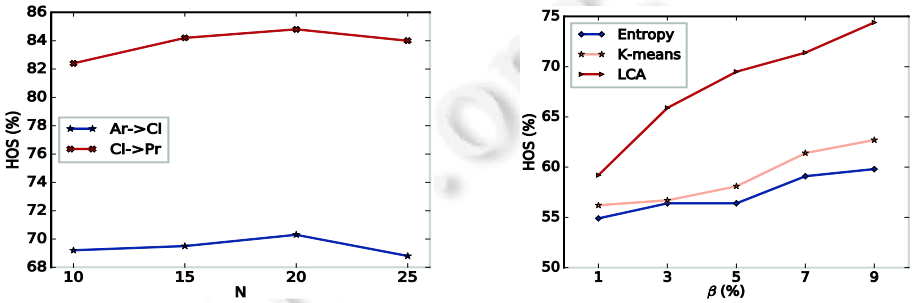
- 最近邻居的数量(*N*)

为了验证不同邻居数量对于 LCAL 算法的影响, 本文在 Ar→Cl 和 Cl→Pr 任务上基于不同的邻居数量(*N*)进行了实验. 如图 6(a)所示, 横坐标指邻居数量. 在不同的邻居数量比例下, 模型效果的波动范围始终处于

2%以内，表明本文所提的方法对于邻居数量(N)不敏感.

- 主动样本的比例(β)

为了展示主动样本的比例(β)对 LCAL 算法的影响，本文在不同的主动样本选择比例β下对 LCAL, Entropy 和 K-means 方法进行了比较. 如图 6(b)所示，横坐标指主动样本比例. 随着主动样本比例的增加，通过匹配纠正，可以纠正更多相似的公共类和开放类样本，所有主动学习的策略都带来了模型效果的显著提高. 值得注意的是：在不同的样本选择比例下，LCAL 算法始终远远高于现有的主动学习方法，进一步表明本文所提的方法可以挑选到更全面、更有价值的阈值模糊的样本，通过主动标注，可以有效地促进了开放类样本和公共类样本的分离，显著地提高了模型的效果.

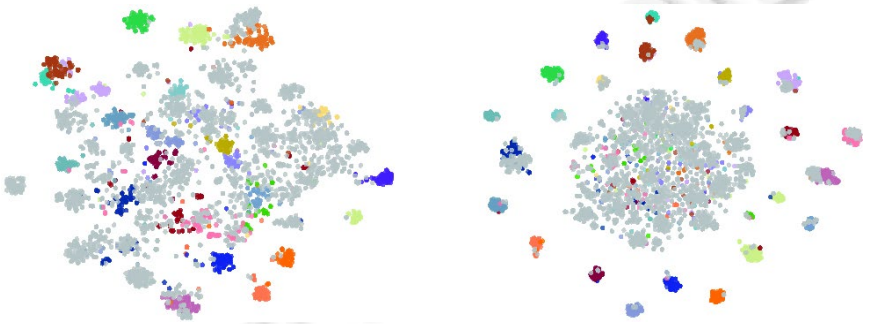


(a) 邻居数量消融实验　　(b) 主动样本比例消融实验

图 6　邻居数量与主动样本比例对 LCAL 算法的影响

- 公共类和开放类的特征分布

图 7 展示了在 Ar→Pr 任务上，训练前后目标域的 TSNE，即目标域特征分布图. 训练前，开放类和公共类样本交错分布，难以对这两类样本进行区分. 值得注意的是：在训练前，开放类的某些类别和公共类重合区域较大，表示这些开放类和公共类是比较相似的，进一步验证了本文提出的阈值模糊样本的存在. 阈值模糊样本的信息熵值相似，所以基于阈值挑选的方式难以有效地分开这部分公共类和开放类. 训练后，公共类样本内部互相远离，公共类和开放类也互相远离. 在保证了开放类和公共类分离效果的同时，保证了公共类样本内部的辨别效果.



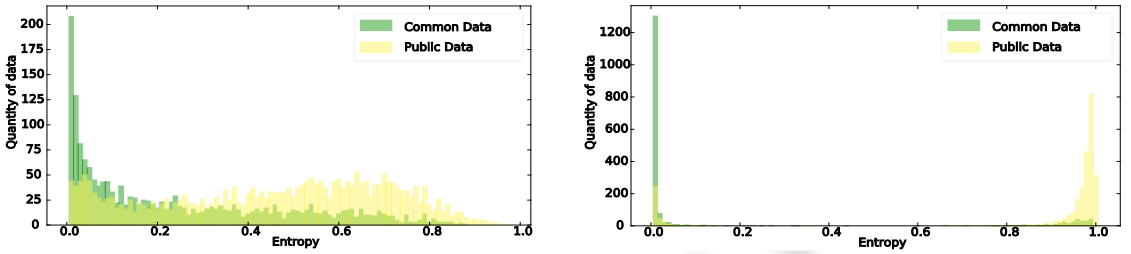
(a) 训练前　　(b) 训练后

图 7　在 Ar→Pr 任务上，训练前后目标域的 TSNE 图

- 公共类和开放类的熵值分布

图 8 展示了在 Ar→Pr 任务上，训练前后目标域样本的信息熵值分布图. 横坐标表示熵值，纵坐标表示对应熵值区间内的样本数量. 在训练前，熵值小的区域包含着部分开放类样本；同样，熵值大的区域也存在不少的公共类样本，难以通过划分阈值的方式有效区分这部分阈值模糊的相似的开放类和公共类样本. 训练后，公共类样本的熵值集中分布在较小的区域，开放类样本的熵值集中分布在较大的区域，此时，通过阈值来划分开放类和公共类时，模型的效果对设定的阈值不敏感. 本文采用设置平均熵值作为阈值来进行划分，可以

有效地区分开放类和公共类样本.



(a) 训练前 (b) 训练后

图 8 在 Ar→Pr 任务上, 训练前后目标域的信息熵值分布

- 损失函数

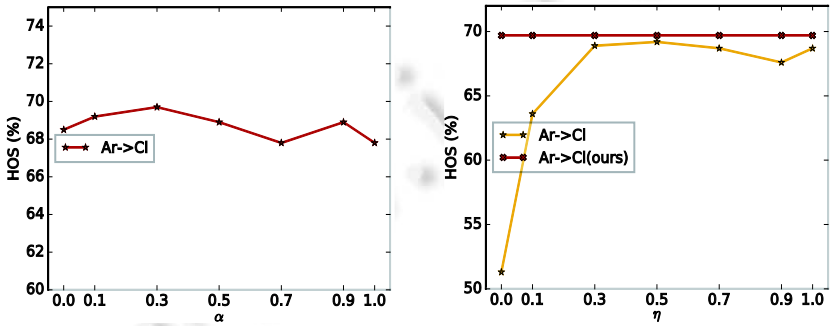
为了验证每一个损失函数的必要性, 本文在 Office-Home 数据集的所有迁移任务上进行了消融实验. 如表 4 所示: 如果去掉开集损失, 此时难以增加模型对开放类样本的不确定性, 则模型对于开放类和公共类的分离效果会大幅度下降; 同样, 若去掉信息最大化损失或交叉熵损失, 此时难以降低模型对公共类样本的不确定性, 则模型对于开放类和公共类的分离效果也会有大幅度的下降. 以上结果表明: 本文所引入的 3 个损失在模型训练中都起着至关重要的作用, 缺一不可.

表 4 损失函数的消融实验

方法	Ar→ Cl	Ar→ Pr	Ar→ Re	Cl→ Ar	Cl→ Pr	Cl→ Re	Pr→ Ar	Pr→ Cl	Pr→ Re	Re→ Ar	Re→ Cl	Re→ Pr	Avg
LCAL w.o L_{unk}	49.6	58.6	60.9	56.4	56.0	53.6	58.2	47.7	57.1	61.3	45.2	57.3	55.2
LCAL w.o L_{kn}	64.1	79.7	81.2	63.4	78.3	77.8	66.1	61.3	80.2	71.3	62.7	80.6	72.2
LCAL w.o L_{im}	63.1	81.3	78.6	61.7	81.1	73.6	60.4	65.0	76.8	66.4	62.8	80.0	70.9
LCAL(Ours)	**69.7**	**84.2**	**80.6**	**66.8**	**84.5**	**79.3**	**67.6**	**69.1**	**81.2**	**72.9**	**69.1**	**82.3**	**75.6**

- 超参数(α)和(η)

为了展示超参数α对于 LCAL 算法的影响, 本文在不同的α取值下对 Ar→Cl 任务进行了比较. 这里, α表示不太可信的公共类伪标签样本的权重. 如图 9(a)所示: 当取值为 0 时, 性能不是最佳, 说明这部分样本应该被学习; 当取值为 1.0 时, 性能也不是最佳, 说明这部分不可信样本的权重不应该设置和主动标注的样本一样, 不应该过大.对所有的迁移任务, 本文选择α=0.3.



(a) 超参数α的消融实验 (b) 超参数η的消融实验

图 9 超参数α与超参数η对 LCAL 算法的影响

为了展示超参数η对于 LCAL 算法的影响, 本文在不同的η取值下对 Ar→Cl 任务进行了比较. 这里, η表示开集损失的权重. 如图 9(b)所示, η取固定值的效果总是略低于本文设计的动态的η取值. η的动态变化使得η的取值随着训练过程逐渐变大, 对于开放类别赋予的权重也越来越大. 若训练的开始就赋予开放类别很大的权重, 会使得模型的效果被那些错误区分的开放类影响, 从而难以获得最佳结果.

4 结 论

本文最新提出并研究了主动学习的源域无关开集域自适应, 在不额外引入训练数据和模型参数的约束下, 解决了严格隐私保护场景下的域差异和新的开放类别出现的问题. 本文提出了基于局部多样性选择的局部一致性主动学习算法(LCAL), 通过挑选阈值模糊样本, 有效地促进了开放类和公共类样本的分离. LCAL 对不同类型的样本施加不同的损失, 使得模型对将主动标记样本匹配纠正后的可信开放类样本更不确定, 对可信公共类样本更确定, 从而进一步促进这两部分样本的分离. 3 个公开基准数据集上的大量实验证明了 LCAL 的有效性, 以及主动学习的源域无关开集域自适应的可行性.

References:

[1] He KM, Zhang XY, Ren SQ, *et al*. Deep residual learning for image recoginition. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE Computer Society, 2016. 770−778.

[2] Huang G, Liu Z, Laurens VDM, *et al*. Densely connected convolutional networks. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE Computer Society, 2017. 4700−4708.

[3] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Communications of the ACM, 2017, 60(6): 84−90.

[4] Torralba A, Efros AA. Unbiased look at dataset bias. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). Colorado: IEEE Computer Society, 2011. 1521−1528.

[5] Saenko K, Kulis B, Fritz M, *et al*. Adapting visual category models to new domains. In: Proc. of the 11th European Conf. on Computer Vision (Computer Vision-ECCV 2010). Grete: Springer, 2010. 213−226.

[6] 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). Salt Lake City: IEEE Computer Society, 2018. 5001−5009.

[7] Hsu HK, Yao CH, Tsai YH, *et al*. Progressive domain adaptation for object detection. In: Proc. of the Winter Conf. on Applications of Computer Vision (WACV). Snowmass Village: IEEE, 2020. 738−746.

[8] Gopalan R, Li R, Chellappa R. Domain adaptation for object recognition: An unsupervised approach. In: Proc. of the IEEE/CVF Int'l Conf. on Computer Vision (ICCV). Barcelona: IEEE Computer Society, 2011. 999−1006.

[9] Tsai YH, Hung WC, Schulter 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). Salt Lake City: IEEE Computer Society, 2018. 7472−7481.

[10] Zhang YF, Niu SC, Qiu Z, *et al*. COVID-DA: Deep domain adaptation from typical pneumonia to COVID-19. arXiv:2005.01577, 2020.

[11] Xu GX, Liu C, Liu J, *et al*. Cross-site severity assessment of COVID-19 from CT images via domain adaptation. IEEE Trans. on Medical Imaging, 2022, 41(1): 88−102.

[12] Kundu JN, Venkat N, Revanur A, *et al*. Towards inheritable models for open-set domain adaptation. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE Computer Society, 2020. 12376−12385.

[13] Liang J, Hu DP, Feng JS, *et al*. UMAD: Universal model adaptation under domain and category shift. arXiv:2112.08553, 2021.

[14] Luo YD, Wang ZJ, Chen ZX, *et al*. Source-free progressive graph learning for open-set domain adaptation. arXiv:2202.06174, 2022.

[15] Feng ZY, Xu C, Tao DC. Open-set hypothesis transfer with semantic consistency. IEEE Trans. on Image Processing, 2021, 30: 6473−6484.

[16] Su JC, Tsai YH, Sohn K, *et al*. Active adversarial domain adaptation. In: Proc. of the IEEE Winter Conf. on Applications of Computer Vision (WACV). Snowmass Village: IEEE, 2020. 739−748.

[17] Fu B, Cao ZJ, Wang JM, *et al*. Transferable query selection for active domain adaptation. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). Virtual: IEEE Computer Society, 2021. 7272−7281.

[18] Mathelin A, Deheeger F, Mougeot M, *et al*. Discrepancy-Based active learning for domain adaptation. In: Proc. of the Int'l Conf. on Learning Representations (ICLR). Virtual: Openreview.net, 2021.

[19] Dhananjaya MM, Kumar VR, Yogamani S. Weather and light level classification for autonomous driving: Dataset, baseline and active learning. In: Proc. of the IEEE Int'l Intelligent Transportation Systems Conf. (ITSC). Indianapolis: IEEE, 2021. 2816−2821.

[20] Riccardi G, Hakkani-Tur D. Active learning: Theory and applications to automatic speech recognition. IEEE Trans. on Speech and Audio Processing, 2005, 13(4), 504−511.

[21] Yuan Y, Chung SW, Kang HG. Gradient-based active learning query strategy for end-to-end speech recognition. In: Proc. of the IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP). Brighton: IEEE, 2019. 2832−2836.

[22] Yang JF, Peng XY, Wang K, *et al*. Divide to adapt: Mitigating confirmation bias for domain adaptation of black-box predictors. arXiv:2205.14467, 2022.

[23] Tian Q, Ma C, Zhang FY, *et al*. Source-free unsupervised domain adaptation with sample transport learning. Computer Science and Technology, 2021, 36(3): 606−616.

[24] Wang D, Shang Y. A new active labeling method for deep learning. In: Proc. of the Int'l joint Conf. on Neural Networks (IJCNN). Beijing: IEEE, 2014. 112−119.

[25] He T, Jin XM, Ding GG, *et al*. Towards better uncertainty sampling: Active learning with multiple views for deep convolutional neural network. In: Proc. of the IEEE Int'l Conf. on Multimedia and Expo (ICME). Shanghai: IEEE, 2019. 1360−1365.

[26] Sener O, Savarese S. Active learning for convolutional neural networks: A core-set approach. In: Proc. of the Int'l Conf. on Learning Representations (ICLR). Vancouver: Openreview.net, 2018.

[27] Long MS, Cao ZJ, Wang JM. Learning transferable features with deep adaptation networks. In: Proc. of the Int'l Conf. on Machine Learning (ICML). Lille: JMLR.org, 2015. 97−105.

[28] Long MS, Zhu H, Wang JM, *et al*. Deep transfer learning with joint adaptation networks. In: Proc. of the Int'l Conf. on Machine Learning (ICML). Sydney: PMLR, 2017. 2208−2217.

[29] Long MS, Cao ZJ, Wang JM, *et al*. Conditional adversarial domain adaptation. In: Advances in Neural Information Processing Systems, Vol.31. Montreal, 2018.

[30] Ganin Y, Lempitsky V. Unsupervised domain adaptation by backpropagation. In: Proc. of the Int'l Conf. on Machine Learning (ICML). Lille: JMLR.org, 2015. 1180−1189.

[31] Panareda Busto P, Gall J. Open set domain adaptation. In: Proc. of the IEEE Int'l Conf. on Computer Vision (ICCV). Venice: IEEE Computer Society, 2017. 754−763.

[32] Saito K, Yamamoto S, Ushiku Y, *et al*. Open set domain adaptation by backpropagation. In: Proc. of the European Conf. on Computer Vision (ECCV). Munich: Springer, 2018. 153−168.

[33] Busto PP, Iqbal A, Gall J. Open set domain adaptation for image and action recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2018, 42(2): 413−429.

[34] Jiang P, Wu AM, Han YH, *et al*. Bidirectional adversarial training for semi-supervised domain adaptation. In: Proc. of the 29th Int'l Conf. on Int'l Joint Conf. on Artificial Intelligence. ijcai.org, 2020. 934−940.

[35] Kim T, Kim C. Attract, perturb, and explore: Learning a feature alignment network for semi-supervised domain adaptation. In: Proc. of the European Conf. on Computer Vision (ECCV). Glasgow: Springer, 2020. 591−607.

[36] Fu B, Cao ZJ, Wang JM, *et al*. Transferable query selection for active domain adaptation. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). Virtual: IEEE Computer Society, 2021. 7272−7281.

[37] Xie BH, Yuan LH, Li SL, *et al*. Active learning for domain adaptation: An energy-based approach. In: Proc. of the AAAI Conf. on Artificial Intelligence. Virtual: AAAI, 2022. 8708−8716.

[38] Fu B, Cao ZJ, Long MS. Learning to detect public classes for universal domain adaptation. In: Proc. of the European Conf. on Computer Vision (ECCV). Glasgow: Springer, 2020. 567−583.

[39] Prabhu V, Chandrasekaran A, Saenko K, *et al*. Active domain adaptation via clustering uncertainty-weighted embeddings. In: Proc. of the IEEE/CVF Int'l Conf. on Computer Vision (ICCV). Monteral: IEEE, 2021. 8505−8514.

[40] Ding YH, Sheng LJ, Liang J, *et al*. ProxyMix: Proxy-based mixup training with label refinery for source-free domain adaptation. Neural Networks, 2023.

[41] Yang SQ, van de Weijer J, Herranz L, *et al*. Exploiting the intrinsic neighborhood structure for source-free domain adaptation. In: Advances in Neural Information Processing Systems. Virtual, 2021. 29393−29405.

[42] Yang SQ, Wang YX, van de Weijer J, *et al*. Generalized source-free domain adaptation. In: Proc. of the IEEE/CVF Int'l Conf. on Computer Vision (ICCV). Monteral: IEEE, 2021. 8978−8987.

[43] Liang J, Hu DP, Feng JS. Do we really need to access the source data? Source hypothesis transfer for unsupervised domain adaptation. In: Proc. of the Int'l Conf. on Machine Learning (ICML). Virtual: PMLR, 2020. 6028−6039.

[44] Saenko K, Kulis B, Fritz M, *et al*. Adapting visual category models to new domains. In: Proc. of the European Conf. on Computer Vision (ECCV). Grete: Springer, 2010. 213−226.

[45] Venkateswara H, Eusebio J, Chakraborty S, *et al*. Deep Hashing network for unsupervised domain adaptation. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE Computer Society, 2017. 5018−5027.

[46] Peng XC, Usman B, Kaushik N, *et al*.Visda: The visual domain adaptation challenge. arXiv:1710.06924, 2017.

[47] Joshi AJ, Porikli F, Papanikolopoulos N. Multi-class active learning for image classification. In: Proc. of the IEEE/CVF Int'l Conf. on Computer Vision (CVPR). Miami: IEEE Computer Society, 2009. 2372–2379.

[48] Ash JT, Zhang CC, Krishnamurthy A, *et al*. Deep batch active learning by diverse, uncertain gradient lower bounds. arXiv:1906.03671, 2019.

[49] Wang F, Han ZY, Zhang ZY, *et al*. Active Source Free Domain Adaptation. arXiv:2205.10711, 2022.

[50] Saito K, Saenko K. Ovanet: One-vs-all network for universal domain adaptation. In: Proc. of the IEEE/CVF Int'l Conf. on Computer Vision (ICCV). Monteral: IEEE, 2021. 9000–9009.

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

苏皖(1997－), 女, 博士生, 主要研究领域为机器学习, 域自适应, 开集域自适应.

韩忠义(1994－), 男, 博士, 主要研究领域为机器学习, 域自适应.

尹义龙(1972－), 男, 博士, 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.