

# 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 Domain 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 Domain 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







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  $\bar{C}_t = C_t \setminus C_s$ . 本文将少部分需要标注的有价值的样本定义为主动样本  $D_t^l = \{x_i^l, y_i^l\}_{i=1}^{n_l}$ , 其中,  $n_l = \beta n_t$  表示主动样本的数量,  $\beta$  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类别  $\bar{C}_t$  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<sup>[15,38]</sup>, 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  $D_{pp} = \{x_i | H(x_i) \geq w_0\}$  (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<sup>[39]</sup>, 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.























