

对抗式域适配迁移学习研究

张雪梅¹, 周艳聪^{2*}

¹天津商业大学理学院, 天津

²天津商业大学信息工程学院, 天津

收稿日期: 2021年11月6日; 录用日期: 2021年12月6日; 发布日期: 2021年12月13日

摘要

对抗式域自适应以对抗学习的方式最小化源领域中的任务损失和最大化域混淆损失, 学习到领域间的共享特征空间, 从而进行特征适配(分布式配), 辅助目标领域学习任务, 是当前域自适应研究, 被广泛应用于在行人重识别、图像分类和情感分析等领域。本文梳理了当前对抗式域适配研究工作, 按照源领域和目标领域标签类别空间异同将其分为同构对抗式域自适应和异构对抗式域自适应两种子类型。依次详细介绍边缘分布同构对抗式域自适应、条件分布对抗式域自适应、联合分布对抗式域自适应和动态分布对抗式域自适应、开放集对抗式域自适应、局部对抗式域自适应和通用对抗式域适配六类子域自适应的研究问题、研究思路及主要研究工作。

关键词

迁移学习, 同构对抗式域自适应, 异构对抗式域自适应

Research on Adversarial Domain Adaptation

Xuemei Zhang¹, Yancong Zhou^{2*}

¹School of Science, Tianjin University of Commerce, Tianjin

²School of Information Engineering, Tianjin University of Commerce, Tianjin

Received: Nov. 6th, 2021; accepted: Dec. 6th, 2021; published: Dec. 13th, 2021

Abstract

Adversarial domain adaptation can be used to learn a shared feature space across domains by minimizing the task loss in the source domain and maximizing the domain confusion loss. Then the

*通讯作者。

task in the target domain can get assistance from feature adaptation (distributed adaptation). Now, adversarial domain adaptation is widely used in the fields of pedestrian re-recognition, image classification, sentiment analysis, and so on. This paper summarizes the current research work of adversarial domain adaptation and divides it into two subtypes: homogeneous adversarial domain adaptation, heterogeneous adversarial domain adaptation according to whether the label space of the source domain and the target domain is the same. This paper also introduces in detail the research problems, research ideas, and main research work of six kinds of subdomain adaptation: marginal distribution adversarial domain adaptation, conditional distribution adversarial domain adaptation, joint distribution adversarial domain adaptation, dynamic distributed adversarial domain adaptation, open set adversarial domain adaptation, partial adversarial domain adaptation, and universal adversarial domain adaptation.

Keywords

Transfer Learning, Homogeneous Adversarial Domain Adaptation, Heterogeneous Adversarial Domain Adaptation

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

深度学习被广泛运用到图像视觉、自然语言处理和语音识别等各个领域, 研究成果丰富。训练出一个泛化能力足够强的深度学习模型需要大量带标注的数据。在实际应用中, 标注数据却是十分稀缺且获取难度较大的。特别是在大数据时代, 海量数据不仅结构各异, 分布也不同, 难以确保收集的训练数据和测试数据满足机器学习模型训练中独立同分布的假设。因此, 深度学习的发展面临着标注数据和训练数据稀缺的问题。

迁移学习有助于解决深度学习中标注数据稀缺的问题。它不但放松了数据独立同分布这一限制假设的要求, 而且能够借助辅助领域知识来研究新领域, 节约模型建立成本。同时, 深度学习特有的分层结构带来的多层次特征学习, 能够显示出不同领域数据的共享特征, 即不能用来区分领域的特征, 为迁移学习的实现提供了多种实现结构。一方面深度学习提供多视角, 研究人员可以根据不同层的输入进行适当的处理, 找到不同领域间的共享特征; 另一方面深度学习的对抗结构能够自动找到不同领域间的共享特征, 用特征提取器来提取领域特征, 让判别器根据特征区分不同的领域, 双方在博弈的过程中达到平衡时, 领域间重叠部分即共享特征即可被筛选出。简而言之, 深度学习天然适合迁移学习, 其对抗网络结构可能成为迁移学习的范式。因此, 基于对抗网络的深度迁移算法是迁移学习的研究热点之一。

对抗式域适配是深度对抗迁移学习算法的子领域, 本文对当前对抗式域适配研究的梳理, 主要贡献如下:

- 1) 本文对对抗式域适配进行定义, 按照源领域与目标领域的标签类别空间异同分为同构域自适应和异构对抗域自适应两类;
- 2) 详细介绍了同构对抗式域适配和异构对抗式域适配研究的主要思路及重要研究。

2. 对抗域自适应

首先给出文中所用的相关符号, 具体说明如表 1 所示。

Table 1. Symbol description
表 1. 符号说明表

符号	含义
上(下)标 s/t	表示源领域或目标领域
\mathcal{D}	领域
\mathcal{T}	任务
$x/X/\mathcal{X}$	特征向量/矩阵/空间
y/\mathcal{Y}	标签类别向量/空间
n	样本数
$P(X)$	边缘概率分布
$P(y X)$	条件分布
$f(X)$	目标领域分类模型

对抗域自适应属于迁移学习领域, 常用到迁移学习定义中“领域”和“任务”两个重要概念, 目前广泛使用的定义是 Pan 等人[1]在 2010 年关于迁移学习的综述中所提出的, 具体描述如下:

领域 \mathcal{D} 定义为由 d 维特征空间 \mathcal{X} 和边缘概率分布 $P(X)$ 组成, 即 $\mathcal{D} = \{\mathcal{X}, P(X)\}$, 其中样本 $X = \{x_1, \dots, x_n\} \in \mathcal{X}$ 。给定领域 \mathcal{D} , 任务 \mathcal{T} 定义为由标签类别空间 \mathcal{Y} 和预测模型 $f(X)$, 即 $\mathcal{T} = \{\mathcal{Y}, f(X)\}$, 按统计学的观点, 预测模型 $f(X)$ 可以解释为条件概率分布 $P(y|X)$, 即 $f(X) = P(y|X)$, 其中 $y \in \mathcal{Y}$ 。故迁移学习的定义如下:

定义 2.1 (迁移学习) [1]: 给定标注的源领域 $\mathcal{D}_s = \{(x_1^s, y_1^s), \dots, (x_{n_s}^s, y_{n_s}^s)\}$ 和学习任务 \mathcal{T}_s , 无标注目标领域 $\mathcal{D}_t = \{x_1^t, \dots, x_{n_t}^t\}$ 和学习任务 \mathcal{T}_t , 迁移学习的目标是当 $\mathcal{D}_s \neq \mathcal{D}_t$ 或者 $\mathcal{T}_s \neq \mathcal{T}_t$ 时, 降低目标领域预测模型 $f_t(X)$ 的泛化误差。

迁移学习研究思路主要是确定源领域和目标领域的相似性, 减小领域间的差异, 对抗网络结构进行功能实现。对抗网络的生成器可以作为特征提取器, 提取领域间的共享特征即领域间的“相似性”; 判别器可以作为领域判别器, 区分源领域和目标领域特征, 与特征提取器进行博弈, 双方在博弈的过程中达到平衡时, 领域间重叠部分即共享特征即可被筛选出, 其差异也被减小, 因此研究者将对抗网络与迁移学习结合在一起, 形成一个新的研究领域——对抗式域自适应。对抗域自适应网络一般由特征提取器 F , 标签分类器 G 和领域判别器 D 三部分组成, 其定义如下:

定义 2.2 (对抗式域自适应): 给定一个迁移学习任务 $\langle \mathcal{D}_s, \mathcal{T}_s, \mathcal{D}_t, f_t(\cdot) \rangle$, $f_t(\cdot)$ 是由对抗网络组成的非线性的函数。

3. 对抗域自适应的类型

根据源领域和目标领域标签类别空间异同, 对抗域自适应包括同构对抗式域自适应和异构对抗式域自适应两种子类型。同构对抗式域自适应根据对边缘分布和条件分布的侧重可以分为边缘分布同构对抗域自适应、条件分布对抗域自适应、联合分布对抗域自适应和动态分布对抗域自适应四种子类型; 异构对抗式域自适应根据源领域和目标领域标签类空间交集情况可以分为开放集对抗式域自适应、局部对抗式域自适应和通用对抗式域自适应三种。

本节将逐一介绍上述对抗式域自适应算法。

3.1. 同构对抗式域适配迁移学习

在同构域自适应中, 常给定源领域 $D_s = \{(x_i^s, y_i^s)\}_{i=1}^{n_s}$, 有 n_s 个带标签样本, 目标领域 $D_t = \{(x_i^t)\}_{i=1}^{n_t}$, 有 n_t 个无标签样本, 假设源领域和目标领域的标签类别空间是相同的, 即 $\mathcal{Y}_s = \mathcal{Y}_t$ 。源领域数据的概率分布为 p , 目标领域的概率分布为 q , 且 $p \neq q$ 。现需要函数 $f(\cdot)$ 使得 p 和 q 的差异减小, 且目标领域的分类误差有界, 即 $R_t(f_t) = \mathbb{E}_{(x,y)} \sim [f_t(x) \neq y]$ 有界。

这里的数据分布主要是指领域的边缘分布 $P(X)$ 、条件概率分布 $P(y|X)$ 和联合分布 $P(X, y)$, 根据源领域数据分布和目标领域不同的数据分布情况, 在研究中需要对领域间的概率分布适配有所侧重, 如图 1 所示(改编自文献[2]), 源领域和不同分布的目标领域, 其中不同的形状代表不同标签类的数据, 线表示标签类之间的边界。这三种概率分布由对抗式域自适应算法中不同的结构负责输出。

对抗式域自适应算法结构一般由特征提取器 F , 标签分类器 G 和领域判别器 D 三部分组成。特征提取器 F 的输入是源领域数据 X_s 和目标领域数据 X_t , 输出是数据特征 f ; 标签分类器 G 的输入是 f , 输出是样本估计标签 $\hat{y} = P(y|X)$; 领域判别器的输入是 f , 输出是样本估计领域标签 $\hat{d} \approx P(X)$ 。

后文将介绍边缘分布同构对抗域自适应、条件分布对抗域自适应、联合分布对抗域自适应和动态分布对抗域自适应四种同构对抗式域自适应。

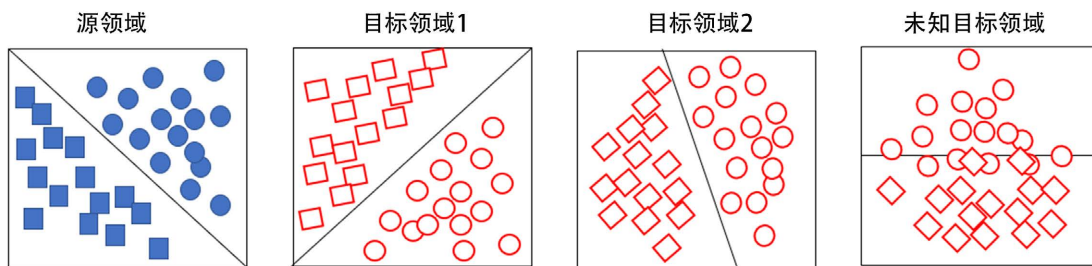


Figure 1. Source domain and target domain with different distribution
图 1. 源领域和不同分布情况的目标领域图示

3.1.1. 边缘分布同构对抗式域自适应

当源领域是对目标领域 1 进行适配时, 两个领域的整体分布相似度低, 即 $P_s(X) \neq P_t(X)$, 具体到每个标签类中数据分布差异较小即 $P_s(y|X) \approx P_t(y|X)$ 时, 对抗域自适应的主要任务是减小源领域和目标领域间边缘分布的差异, 其目标函数可以简写为:

$$\min_{D_s, D_t} \text{Dist}(P_s(X), P_t(X)) \quad (1)$$

例如, 经典的对抗域自适应算法 Domain-Adversarial Neural Network (DANN) [3] 由特征提取器 F , 标签分类器 G 和领域判别器 D 三部分组成, 其中 D 是对整个源领域和目标领域进行特征适配即边缘分布适配; Adversarial Discriminative Domain Adaptation (ADDA) [4] 网络结构在 DANN 基础上将特征提取器分为源领域特征提取器和目标领域特征提取器两部分, 且源领域的特征提取器提前由源领域数据预训练, 在对抗域训练过程中, 源领域特征提取器参数固定不变, 目标领域特征提取器采用源领域特征提取器的参数作为初始化参数, 随着目标领域数据特征进行调整, 确保目标领域特征提取器既可以学习目标领域独有特征, 也又可以学到领域间的共享特征来迷惑领域判别器, 使得领域判别器难以区分源领域和目标领域特征分布, 实现领域间的边缘分布适配; 文献[5]则引入胶囊网络和残差机制作为 ADDA 中特征提取器, 更好的拟合源域和目标域内有限的样本特征边缘分布[5]。Adversarial Domain Adaptation with Domain

Mixup (DM-ADA) [6]网络结构与 ADDA 相同, 增加源领域和目标领域输入数据混肴和训练过程中领域特征混肴这两部分, 以此迷惑和加强训练判别器, 使得领域整体的边缘分布更好的适配; Domain Symmetric Networks (SymNets) [7]则通过源领域标签分类器和目标领域标签分类器共享神经元组成一个领域判别器进行领域整体特征混肴, 拟合边缘分布。

3.1.2. 条件分布同构对抗式域自适应

当源领域是对目标领域 2 进行适配时, 两个领域的整体分布相似度高即 $P_s(X) \approx P_t(X)$, 具体到每个标签类中数据分布差异较大时, 对抗域自适应的主要任务是减小源领域和目标领域间条件分布的差异, 其目标函数可以简写为

$$\min_{D_s, D_t} \text{Dist}(P_s(y|X), P_t(y|X)) \quad (2)$$

例如, TANG 等人[8]在对运动脑电波图进行无监督学习时, 将标签分类器输出 $\hat{y} = P(y|X)$ 与样本特征 f 进行交叉乘, 即借助条件分布进行特征适配; Adversarial-Learned Loss for Domain Adaptation (ALDA) [9]根据领域判别器 D 的输出 \hat{d} 生成混淆矩阵, 据此对目标领域无标签数据的分类估计 $\hat{y} = P(y|X)$ 进行修正, 从而拟合源领域和目标领域的条件分布; Adversarial Branch Architecture Search (ABAS) [10]旨在为 DANN 和 ALDA 这类对抗式域自适应寻找最优领域判别器, 改进它们的网络结构, 进行分布适配。

3.1.3. 联合分布同构对抗式域自适应

联合分布同构对抗式域自适应其目标是减小源领域和目标领域的联合概率分布 $P(X, y)$ 距离, 完成迁移学习。由于目标领域的标注类别空间是未知, 其联合概率分布则无法获知, 源领域和目标领域的联合概率分布差异无法直接进行度量。因此, 源领域和目标领域之间联合概率分布距离可以用边缘分布距离和条件分布距离之和近似估计, 或者用边缘分布和条件分布的乘积估计源领域和目标领域的联合概率, 再进行距离度量, 即

$$\min_{D_s, D_t} \text{Dist}(P_s(X, y), P_t(X, y)) \approx \text{Dist}(P_s(y|X), P_t(y|X)) + \text{Dist}(P_s(X), P_t(X)) \quad (3)$$

或

$$\min_{D_s, D_t} \text{Dist}(P_s(X, y), P_t(X, y)) \approx \text{Dist}(P_s(y|X)P_s(X), P_t(y|X)P_t(X)) \quad (4)$$

例如, Multi-Adversarial Domain Adaptation (MADA) [11]中领域判别器分为 $|\mathcal{Y}_s|$ 个分类网络 G_d^K , $K = 1, 2, \dots, |\mathcal{Y}_s|$, G_d^K 的输入对应每个标签类的样本特征 f 和标签分类器 G 输出的预测值 \hat{y}^K 的乘积, 根据条件概率对特征进行赋权, 即对每个样本点加权特征进行边缘分布自适应后实现领域整体边缘分布自适应, 避免每个标签类下的样本特征适配错分, 造成负迁移, 学习到标签类级特征, 同时减小领域间的联合概率分布差异; Conditional Adversarial Domain Adaptation (CDANs) [12]则对样本特征 f 和标签分类器 G 输出标签预测 \hat{y} , 采取线性组合的方式进行条件分布适配; Joint Adversarial Domain Adaptation (JADA) [13]的分类器由源领域标签分类器 G_y^s 和目标领域标签分类器 G_y^t 两部分组成, 它们对目标领域的条件概率估计分别是 \hat{y}_s^t 、 \hat{y}_t^t , 按照 L_1 正则项计算方式计算 \hat{y}_s^t 和 \hat{y}_t^t 之间的差异, 作为 JADA 算法目标函数的正则项, 实现条件概率适配。

3.1.4. 动态分布对抗式域自适应

前面三种情况均对源领域和目标领域的数据分布差异作出假设, 且默认边缘分布和条件分布自适应是同等重要的。若目标领域的数据分布是未知的, 源领域和目标领域边缘分布和条件分布的差异是无从

获知的, 则对抗域自适应中边缘分布适配和条件分布适配的重要性是难以评估的, 如图 1 中源领域对未知目标领域进行对抗域自适应。因此, WANG 等人[2] [14]提出了动态分布自适应(Dynamic Distribution Adaptation, DDA)的方法, 一方面能够根据特定的数据领域动态衡量分布适配过程中的边缘分布和条件分布的重要性, 另一方面自动调整这两种分布的重要性, 取得良好的分布适配效果。调整方式是通过计算动态平衡因子 μ 实现, $\mu \in [0,1]$, 则对抗域自适应的目标函数可以简写为

$$\min_{D_s, D_t} (1-\mu)Dist(P_s(X), P_t(X)) + \mu Dist(P_s(y|X), P_t(y|X)) \tag{5}$$

Dynamic Adversarial Adaptation Network (DAAN) [15]是率先将动态分布适配和对抗域自适应结合在一起的方法。DANN 由一个特征提取器 G_f , 分类器 G_g , 全局领域判别器 G_d , 局部子领域判别器 G_d^k , $k=1,2,\dots,|\mathcal{Y}_s|$ 。这里 G_d 用来度量源领域和目标领域的领域整体数据分布差异 L_g , 即领域的边缘分布差异, G_d^k 用来度量源领域目标领域每个标签类 k 中数据分布差异 L_t^k , 即条件分布差异, 动态平衡因子 μ 由 G_d 和 G_d^k 的交叉熵估计得出, 即

$$\mu = \frac{2(1-L_g)}{2(1-L_g) + \frac{1}{|\mathcal{Y}_s|} \sum_{k=1}^{|\mathcal{Y}_s|} 2(1-2L_t^k)} \tag{6}$$

3.2. 异构对抗式域适配迁移学习

如图 2 所示, 实线圈代表源领域标签类别空间, 虚线圈代表目标领域标签类别空间, 图 2(a)表示闭集域自适应场景即源领域和目标领域的标签类别空间一致, $\mathcal{Y}_s = \mathcal{Y}_t$, 3.1 节中同构对抗式域自适应算法均适用于此应用场景; 图 2(b)、图 2(c)和图 2(d)依次为开放集、局部和通用域自适应场景, 源领域和目标领域标签类别空间是有差异的, 即 $\mathcal{Y}_s \neq \mathcal{Y}_t$, 下文将依次给出这三种场景的详细介绍。

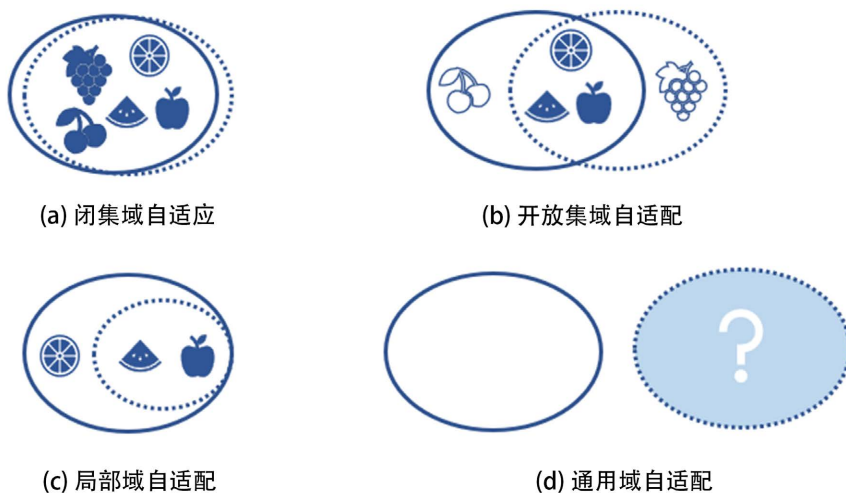


Figure 2. From totally closed set to universal domain adaptation setting
图 2. 从全封闭集到通用域自适应标签类别空间示意图

3.2.1. 开放集对抗式域自适应

开放集域自适应(Open Set Domain Adaptation, OSDA) [16], 给定源领域 $D_s = \{x_i^s, y_i^s\}_{i=1}^{n_s}$ 有 n_s 个带标签的样本, 分为 $|C_s|$ 个类别, 目标领域 $D_t = \{x_i^t\}_{i=1}^{n_t}$ 有 n_t 个无标签样本, 分为 $|\mathcal{Y}_t|$ 个类别, 且 $\mathcal{Y}_s \neq \mathcal{Y}_t$,

$\mathcal{Y} = \mathcal{Y}_s \cap \mathcal{Y}_t \neq \emptyset$ 。两个领域的概率分布不一致。现在需要寻找一个可转移的特征映射函数 $f = G_f(X)$ 和分类器 $y = G_y(f)$, 使得模型分类误差期望 $E_{(x,y)} [G_y(G_f(X)) \neq y]$ 达到最小。

Saito 等人[17]在 OSDA 上提出了一种更具挑战性的开放集情形, 即 \mathcal{Y}_s 已知, \mathcal{Y}_t 未知, $\mathcal{Y}_s \cap \mathcal{Y}_t = \mathcal{Y}$, 源领域标注类别空间相当于目标领域标注类别空间的子集, 这意味着源领域中没有数据帮助模型将目标领域中独有的标注类别数据标注出来, 并且他们率先采用对抗网络应用到此场景中解决问题; **Separate to Adapt (STA)** [18]分两步解决 OSDA 中的问题, 第一步训练 $|\mathcal{Y}_s|$ 个二分类器, 计算目标领域样本与每个标签类源领域数据的相似性 p_y , 根据 p_y 对源领域的数据进行聚类, 分为高、中和低相似三类, 取相似度高标签类的 p_y 均值作为选出 \mathcal{Y} 阈值, 取相似度低的标签类的 p_y 均值作为选出不属于 \mathcal{Y} 标签类的阈值, 第二步运用对抗网络结构对筛选出来的属于 \mathcal{Y} 的样本进行分类。

3.2.2. 局部对抗式域自适应

局部域自适应(Partial Domain adaptation, PDA) [19], 给定源领域 $D_s = \{x_i^s, y_i^s\}_{i=1}^{n_s}$ 有 n_s 个带标签的样本, 分为 $|C_s|$ 个类别, 目标领域 $D_t = \{x_i^t\}_{i=1}^{n_t}$ 有 n_t 个无标签样本, 分为 $|\mathcal{Y}_t|$ 个类别, 且 $|\mathcal{Y}_s| > |\mathcal{Y}_t|$, 即 $\mathcal{Y}_s \supset \mathcal{Y}_t$, $\mathcal{Y}_s \cap \mathcal{Y}_t = \mathcal{Y}$ 。两个领域的概率分布不一致。现在需要寻找一个可转移的特征映射函数 $f = G_f(X)$ 和分类器 $y = G_y(f)$, 使得模型分类误差期望 $E_{(x,y)} [G_y(G_f(X)) \neq y]$ 达到最小。

此场景由清华大学龙明盛教授团队提出, 可以看作开集的一种特殊情况。他们在深度对抗网络的基础上提出 **Selective Adversarial Network (SAN)** [20]和 **Partial Adversarial Domain Adaptation (PADA)** [19]两种适用于局部迁移背景的深度迁移算法。SAN 和 PADA 是利用对抗网络中的领域判别器来对样本赋权, 减少非目标领域类别数据带来的负迁移, 但并没有避免这些数据对源领域分类器的影响。因此, 龙明盛团队又对这两种算法作出进一步的改进, 提出 **Example Transfer Network (ETN)** [21], 能够在判别器和源领域分类器上同时减弱源领域中非目标领域标注类别数据造成的负迁移。

3.2.3. 通用对抗式域自适应

通用领域适配(Universal Domain Adaptation, UniDA) [22]中, 给定源领域 $D_s = \{x_i^s, y_i^s\}_{i=1}^{n_s}$ 有 n_s 个带标签的样本, 分为 $|\mathcal{Y}_s|$ 个类别, 目标领域 $D_t = \{x_i^t\}_{i=1}^{n_t}$ 有 n_t 个无标签样本, 分为 $|\mathcal{Y}_t|$ 个类别。源领域数据的概率分布为 p , 目标领域的概率分布为 q 。 $\mathcal{Y} = \mathcal{Y}_s \cap \mathcal{Y}_t$, 是源领域和目标领域共享的标签类别空间。 $\bar{\mathcal{Y}}_s = \mathcal{Y}_s \setminus \mathcal{Y}$ 和 $\bar{\mathcal{Y}}_t = \mathcal{Y}_t \setminus \mathcal{Y}$ 分别是独属于源领域和目标领域的标签类别集。 $p_{\bar{\mathcal{Y}}_s}$ 和 $p_{\mathcal{Y}}$ 分别是源领域中属于 $\bar{\mathcal{Y}}_s$ 和 \mathcal{Y} 数据的概率分布, $q_{\bar{\mathcal{Y}}_t}$ 和 $q_{\mathcal{Y}}$ 分别是源领域中属于 $\bar{\mathcal{Y}}_t$ 和 \mathcal{Y} 数据的概率分布。利用杰卡德距离定义领域间的相似性, 即

$$\xi = \frac{|\mathcal{Y}_s \cap \mathcal{Y}_t|}{|\mathcal{Y}_s \cup \mathcal{Y}_t|} \quad (7)$$

闭集领域自适应问题可以看作是 UniDA 的一种特殊情况, $\xi = 1$ 。 ξ 越小, 说明领域间可迁移的知识越少, 领域自适应的难于越大。在 UniDA 情况中, ξ 是未知的, 领域自适应模型需要在不同 ξ 值下取得较好的结果。同时, 模型必须要能够对目标领域中来自 \bar{C}_t 和 C 的数据进行分类, 即得到一个模型 f , 使得模型分类误差期望 $E_{(x,y)-q_C} [f(X) \neq y]$ 最小。

Universal Adaptation Network (UAN) [22]适用于 UniDA, 它在常用的对抗式域自适应网络结构中引入一个非对抗领域判别器 D' , 即 UAN 由特征提取器 F , 标签分类器 G 和领域判别器 D 和非对抗领域判别器 D' 四部分组成。 D' 用来计算领域相似性和权重 w_s 、 w_t , 辅助源领域和目标领域在 \mathcal{Y} 内进行域自适

配; Calibrated Multiple Uncertainties (CMU) [23]在常用对抗式域自适应网络结构的基础上, 将标签分类器分为多个分类器并根据它们的输出计算神经网络预测结果的熵、一致性和置信度, 避免单一分类器带来的不确定性问题, 造成分类错误。此外, 研究者也根据算法在域自适应方法对 \mathcal{Y} 、 $\bar{\mathcal{Y}}_s$ 和 $\bar{\mathcal{Y}}_t$ 的分类准确率, 提出迁移能力评价指标 H-score。

4. 总结

本文对对抗式域自适应研究进行了总结和分析。对抗式域自适应按照源领域和目标领域标签类别空间的异同可分为同构对抗式域自适应和异构对抗式域自适应两种子类。同构对抗式域自适应属于闭集域自适应, 运用对抗方式实现源领域和目标领域的概率分布适配, 包含边缘概率分布适配、条件概率分布适配、联合分布适配和动态分布适配四种; 异构对抗式域自适应根据源领域和目标领域标签类别空间的关系, 可以分为开放集对抗式域自适应、局部集对抗式域自适应和通用对抗式域自适应三种, 当前这方面的研究具有巨大潜力。

参考文献

- [1] Pan, S.J. and Yang, Q. (2010) A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, **22**, 1345-1359. <https://doi.org/10.1109/TKDE.2009.191>
- [2] Wang, J., Chen, Y., Feng, W., et al. (2020) Transfer Learning with Dynamic Distribution Adaptation. *ACM Transactions on Intelligent Systems and Technology*, **11**, 1-25. <https://doi.org/10.1145/3360309>
- [3] Ganin, Y., Ustinova, E., Ajakan, H., et al. (2017) Domain-Adversarial Training of Neural Networks. In: Csurka, G., Ed., *Domain Adaptation in Computer Vision Applications*, Springer International Publishing, Cham, 189-209. http://link.springer.com/10.1007/978-3-319-58347-1_10
https://doi.org/10.1007/978-3-319-58347-1_10
- [4] Tzeng, E., Hoffman, J., Saenko, K., et al. (2021) Adversarial Discriminative Domain Adaptation. <http://arxiv.org/abs/1702.05464>
- [5] 戴宏, 盛立杰, 苗启广. 基于胶囊网络的对抗判别域适应算法[J]. 计算机研究与发展, 2021, 58(9): 1997-2012.
- [6] Xu, M., Zhang, J., Ni, B., et al. (2020) Adversarial Domain Adaptation with Domain Mixup: 04. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 6502-6509. <https://doi.org/10.1609/aaai.v34i04.6123>
- [7] Zhang, Y., Tang, H., Jia, K., et al. (2019) Domain-Symmetric Networks for Adversarial Domain Adaptation. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 6-20 June 2019, 5026-5035. <https://ieeexplore.ieee.org/document/8953920>
<https://doi.org/10.1109/CVPR.2019.00517>
- [8] Tang, X. and Zhang, X. (2020) Conditional Adversarial Domain Adaptation Neural Network for Motor Imagery EEG Decoding. *Entropy*, **22**, 96. <https://doi.org/10.3390/e22010096>
- [9] Chen, M., Zhao, S., Liu, H., et al. (2020) Adversarial-Learned Loss for Domain Adaptation: 04. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 3521-3528. <https://doi.org/10.1609/aaai.v34i04.5757>
- [10] Robbiano, L., Rahman, M.R.U., Galasso, F., et al. (2021) Adversarial Branch Architecture Search for Unsupervised Domain Adaptation. <http://arxiv.org/abs/2102.06679>
- [11] Pei, Z., Cao, Z., Long, M., et al. (2021) Multi-Adversarial Domain Adaptation. *Thirty-Second AAAI Conference on Artificial Intelligence*, New Orleans, 2-7 February 2018, 3934-3941. <https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/17067>
- [12] Long, M., Cao, Z., Wang, J., et al. (2018) Conditional Adversarial Domain Adaptation. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, Curran Associates Inc., Red Hook, 1647-1657.
- [13] Li, S., Liu, C.H., Xie, B., et al. (2019) Joint Adversarial Domain Adaptation. *Proceedings of the 27th ACM International Conference on Multimedia*, Nice, 21-25 October 2019, 729-737. <https://doi.org/10.1145/3343031.3351070>
- [14] Wang, J., Feng, W., Chen, Y., et al. (2018) Visual Domain Adaptation with Manifold Embedded Distribution Alignment. *Proceedings of the 26th ACM International Conference on Multimedia*, Seoul, 22-26 October 2018, 402-410. <https://doi.org/10.1145/3240508.3240512>
- [15] Yu, C., Wang, J., Chen, Y., et al. (2019) Transfer Learning with Dynamic Adversarial Adaptation Network. 2019 *IEEE International Conference on Data Mining (ICDM)*, Beijing, 8-11 November 2019, 778-786.

-
- <https://doi.org/10.1109/ICDM.2019.00088>
- [16] Busto, P.P. and Gall, J. (2017) Open Set Domain Adaptation. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 754-763. <http://ieeexplore.ieee.org/document/8237350>
<https://doi.org/10.1109/ICCV.2017.88>
- [17] Saito, K., Yamamoto, S., Ushiku, Y., *et al.* (2018) Open Set Domain Adaptation by Backpropagation. In: Ferrari, V., Hebert, M., Sminchisescu, C., *et al.*, Eds., *Computer Vision—ECCV 2018*, Springer International Publishing, Cham, 156-171. http://link.springer.com/10.1007/978-3-030-01228-1_10
https://doi.org/10.1007/978-3-030-01228-1_10
- [18] Liu, H., Cao, Z., Long, M., *et al.* (2019) Separate to Adapt: Open Set Domain Adaptation via Progressive Separation. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 16-20 June 2019, 2922-2931. <https://ieeexplore.ieee.org/document/8953906>
<https://doi.org/10.1109/CVPR.2019.00304>
- [19] Cao, Z., Ma, L., Long, M., *et al.* (2018) Partial Adversarial Domain Adaptation. *ECCV 2018: 15th European Conference*, Munich, 8-14 September 2018, 139-155.
- [20] Cao, Z., Long, M., Wang, J., *et al.* (2018) Partial Transfer Learning with Selective Adversarial Networks. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-22 June 2018, 2724-2732. <https://ieeexplore.ieee.org/document/8578386>
<https://doi.org/10.1109/CVPR.2018.00288>
- [21] Cao, Z., You, K., Long, M., *et al.* (2019) Learning to Transfer Examples for Partial Domain Adaptation. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 16-20 June 2019, 2980-2989. <https://ieeexplore.ieee.org/document/8954395>
<https://doi.org/10.1109/CVPR.2019.00310>
- [22] You, K., Long, M., Cao, Z., *et al.* (2019) Universal Domain Adaptation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 2720-2729. https://openaccess.thecvf.com/content_CVPR_2019/html/You_Universal_Domain_Adaptation_CVPR_2019_paper.html
- [23] Fu, B., Cao, Z., Long, M., *et al.* (2020) Learning to Detect Open Classes for Universal Domain Adaptation. In: Vedaldi, A., Bischof, H., Brox, T., *et al.*, Eds., *Computer Vision—ECCV 2020*, Springer International Publishing, Cham, 567-583. https://link.springer.com/10.1007/978-3-030-58555-6_34
https://doi.org/10.1007/978-3-030-58555-6_34