

## 加权解耦语义表达的多源领域自适应方法<sup>\*</sup>

蔡瑞初, 郑丽娟, 李梓健

(广东工业大学 计算机学院, 广东 广州 510006)

通信作者: 蔡瑞初, E-mail: cairuichu@gmail.com



**摘要:** 近年来, 深度学习受到越来越多研究者的重视并成功应用于许多领域. 虽然深度学习在这些领域获得了巨大的成功, 但是数据采集和标注成本高, 严重限制了深度学习的推广应用. 迁移学习不仅可以打破训练集数据和测试集数据独立同分布的假设, 而且可以利用有标签的迁移源数据和没有标签的迁移目标数据训练得到具有良好泛化能力的模型, 是扩展深度学习应用场景的重要研究方向. 在众多的迁移学习方法中, 多源领域自适应方法可以充分利用多个迁移源的信息, 具有重要的实际价值. 从数据的因果生成机制出发, 假设观测数据由语义隐变量和领域隐变量这两组独立的隐变量同时生成. 基于上述假设, 提出了一种基于多种距离度量框架和加权解耦语义表达的多源领域自适应方法. 该方法利用了双重对抗网络来提取解耦的语义信息和领域信息; 另一方面, 采用了3种不同的语义信息聚合策略获得领域不变的语义表达; 最后使用领域不变的语义表达进行图片分类. 在多个多源领域自适应数据上的对比及鲁棒性分析实验中, 充分地验证了所提出方法的有效性.

**关键词:** 迁移学习; 多源领域自适应; 解耦表达; 变分推理

**中图法分类号:** TP181

中文引用格式: 蔡瑞初, 郑丽娟, 李梓健. 加权解耦语义表达的多源领域自适应方法. 软件学报, 2022, 33(12): 4517-4533. <http://www.jos.org.cn/1000-9825/6366.htm>

英文引用格式: Cai RC, Zheng LJ, Li ZJ. Multi-source Domain Adaptation of Weighted Disentangled Semantic Representation. Ruan Jian Xue Bao/Journal of Software, 2022, 33(12): 4517-4533 (in Chinese). <http://www.jos.org.cn/1000-9825/6366.htm>

### Multi-source Domain Adaptation of Weighted Disentangled Semantic Representation

CAI Rui-Chu, ZHENG Li-Juan, LI Zi-Jian

(School of Computer, Guangdong University of Technology, Guangzhou 510006, China)

**Abstract:** Recent years have witnessed the widespread use of domain adaptation. Though having achieved significant performance in different fields, these methods are hungry for a large amount of labeled data, which requires unaffordable cost to meet the data quality and quantity and hinders the further application of deep learning model. Fortunately, domain adaptation, which not only relaxes the I.I.D assumption between the source and the target domain but also uses the labeled source domain data and the unlabeled target domain data simultaneously, is beneficial to achieve a well-generalized model. Among all the domain adaptation setting, multi-source domain adaptation, which takes full advantage of the information of multiple source domains, are more suitable to the real-world application. This study proposes a multi-source domain adaptation method via multi-measure framework and weighted disentangled semantic representation. Motivated from the data generation process in causal view, it is first assumed that the observed samples are controlled by the semantic latent variables and the domain latent variables, and it is further assumed that these variables are independent. As for the extraction of these variables, the duel adversarial training schema is used to extract and disentangle the semantic latent variables and the domain latent variables. As for the multi-domain aggregation, three different domain aggregation strategies are employed to obtain the weighted domain-invariant semantic representation. Finally, the weighted domain-invariant semantic representation is used for classification. Experiment studies not only testify that the proposed method yields state-of-the-art performance on many multi-source domain adaptation benchmark datasets but also validate the robust of the proposed method.

\* 基金项目: 国家自然科学基金(61876043, 61976052); 广州市科技计划(201902010058)

收稿时间: 2020-11-24; 修改时间: 2021-02-04, 2021-03-16; 采用时间: 2021-04-30

**Key words:** transfer learning; multi-source domain adaptation; disentangle representation; variational inference

深度学习算法取得巨大成功的关键因素之一是大规模标签数据的可用性<sup>[1-7]</sup>。但是,在许多实际应用中,仅提供了有限带标签的训练数据,甚至没有提供有标签的数据,导致获得大量标记数据的成本昂贵且代价高。另一方面,视觉数据本质上具有差异性,例如:在图像识别问题中,使用真实的人物数据训练得到的模型不能准确识别动漫人物。这种差异性限制了有监督学习模型在处理带有少量标记示例的新场景时的可扩展性和适用性,降低了传统深度学习方法的性能,而产生这种现象的根本原因是域偏移(domain shift)<sup>[8,9]</sup>,因此无监督的领域适应(domain adaptation, DA)<sup>[10-54]</sup>算法应运而生。无监督领域适应(unsupervised domain adaptation, UDA)<sup>[55]</sup>通过建立从标记源域到未标记目标域的知识转移,以及探索域不变结构和表示来拟合差距以解决此类问题。

目前已经有不少工作证明了无监督领域自适应的目标是减小迁移源和迁移目标之间的距离<sup>[15-26]</sup>。基于以上的理论,目前主要有两类无监督领域自适应方法:基于最大平均差异(maximum mean discrepancy, MMD)<sup>[56]</sup>的无监督领域自适应方法和基于对抗学习(adversarial learning)<sup>[57]</sup>的无监督领域自适应方法。基于最大平均差异是通过最小化深度前馈架构内部的特征分布之间的差异来提取域不变表示的特征。基于对抗学习的方法则是借鉴了生成对抗网络的思想,该方法通过“欺骗”域分类器来提取领域不变的表达。由于实际应用场景中通常拥有多个迁移源的数据,如图1所示,多源领域自适应学习是指学习来自多个不同的迁移源的知识,并将知识成功“迁移”到未知的领域,从而解决针对未知领域任务的方法,因此多源领域自适应的问题开始受到越来越多学者的关注<sup>[30,37,39,41,44-46,50,52-54,58,59]</sup>。



图1 多源领域自适应问题图示

将单迁移源的领域自适应方法扩展到多迁移源场景是多源领域自适应问题的研究方向之一。但是简单地将单源领域自适应方法套用到多源领域自适应问题上,部分迁移源和目标领域之间的相关性比较小的情况下,强制地迁移不仅不能提升学习效果,甚至降低模型性能,即产生负迁移<sup>[60-63]</sup>,负迁移即在源域上学习到的知识,对于目标域上的学习产生负面作用。目前有两种主流的方法来缓解负迁移:一是增加迁移源数据数量,但是由于有标签的数据难以获取,所以这种方法难以采用。二是减少不同领域间数据分布的差异,即提取领域不变的特征。因此,虽然多源领域自适应的目标依然是从多个迁移源中提取领域不变的特征,但是与单源领域自适应不同的是,多源领域自适应是有选择地从多个源中提取并且组合成领域不变信息。目前的多源领域自适应方法目标在于如何学习每个迁移源的权重,以选择合适的迁移源特征来提取可以迁移的信息。但是由于数据分布的基础是复杂的流形结构,领域不变的语义信息和领域特有的信息通常是以高度耦合的形式存在,因此所有的这些领域自适应方法主要会遇到错误的对齐问题,导致迁移源的类别和迁移目标的类别发生错配现象,于是源域上学习到的类别信息在迁移目标上是错误的,最后得到负迁移的结果。如图2(a)所示,假设数据生成过程由特征流形空间中独立的域隐变量和语义隐变量控制。但是,如图2(b)所示,如若没有完全排除域信息,“残余”领域信息会加剧负迁移的产生,无标签的数据样本就会在特征流形空间上失真,从而导

致错误的对齐问题, 即导致了迁移源的类别匹配到迁移目标错误的类别上. 例如, 黑色镜框在某些领域上“看起来”像是黑色的自行车, 因此黑色眼镜框的特征可能更接近变形特征流形空间中的自行车的特征, 从而导致错误的类别匹配, 由此可知产生错误对齐问题的原因是语义和领域信息的高度耦合, 更具体地表示为, 样本由两组独立的隐变量控制, 但是, 这两组存在于高维特征流形空间上的变量相互纠缠和扭曲. 本文从解耦的角度出发, 在领域信息和语义信息高度耦合的空间中进行信息解耦, 尽可能排除领域信息对任务造成的不利影响, 从而缓解负迁移的问题, 但是, 在复杂的高维耦合特征流形空间上保留语义信息的同时, 完全排除领域信息干扰是非常具有挑战性的.

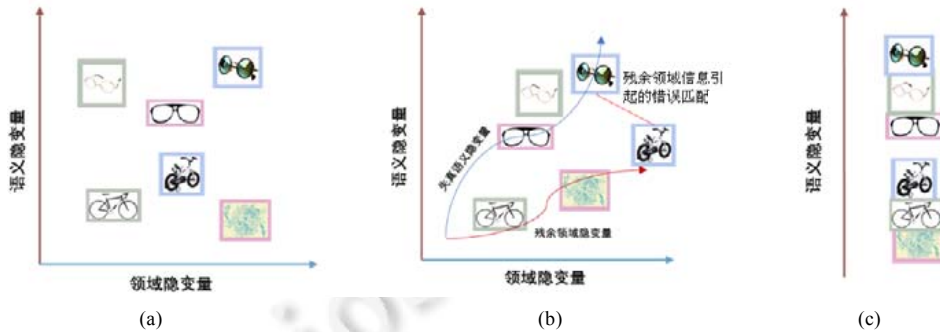


图2 在“艺术”和“真实世界”领域中的“自行车”和“眼镜”样本域自适应示例

受数据生成机制的因果图的启发(如后文的图3所示), 本文假设语义变量和领域变量之间是相互独立的, 并且观测数据是由语义变量和领域变量共同生成的. 因此提出了基于多距离度量框架和加权解耦语义表达的多源领域自适应方法. 本文的模型重构了高度耦合的隐变量空间, 同时使用语义隐变量来预测目标标签. 如图2(c)所示, 可直观地理解为通过使用独立于领域隐变量的语义隐变量来预测目标标签, 即: 仅根据语义轴就可以实现对标签进行分类.

受到上述的启发, 本文提出了基于多种距离度量框架的加权解耦语义表达的多源领域自适应方法. 通过采用变分自动编码器以及借助双重对抗网络, 重构了高度耦合的语义隐变量和领域隐变量. 并进一步采用三种不同的特征距离度量方法(平均距离度量, 欧氏距离度量, Wasserstein distance 度量<sup>[64]</sup>), 计算不同迁移源的语义特征和迁移目标的特征计算距离, 从而计算出最合适的迁移源选择权重, 最终提取更好的领域不变特征. 且本文在标准数据集中验证了该方法在多源域自适应基准上优于最新的无监督多源域自适应方法.

对比现有的方法, 本文提出的方法主要有如下3个优势: (1) 针对在深度领域适应现存方法中出现的负迁移及欠适配等问题, 提出一种新的网络模型, 主要解决无监督领域适应问题, 该模型基于因果生成机制, 采用解耦学习的思路建模. (2) 基于现存相关性对齐方法存在错配的缺点, 提出了新的改进方案, 所提方法利用 variational autoencoders (VAEs)<sup>[65,66]</sup>技术与基于对抗学习解耦方法来增强深度网络的领域适应能力; (3) 不同于当前单一的源领域和目标领域的分布对齐方法, 本文通过不同的度量方式来度量源域对目标域影响程度计算最佳多源领域迁移的权重分配策略.

## 1 相关工作

### 1.1 单源无监督域适应

无监督域自适应(unsupervised domain adaptation, UDA)的目的是将所学习的知识从标记的源领域转移到未标记的目标领域. 深度特征学习方法是对解决监督领域自适应问题有效的方法. 深度特征学习的关键思想是通过对齐不同的领域来提取域不变表示. Ganin 等人<sup>[31]</sup>提出了领域对抗神经网络 DANN, 以学习结合分类器的判别能力和投影空间的域不变性的特征, 同时学习自适应和判别式投影. Ghifary 等人<sup>[32]</sup>提出的 DRCN 将特征重构为图像, 并使转换后的图像与原始图像相似. Long 等人<sup>[55]</sup>通过基于联合最大平均差异准则对齐跨域的

多个特定于域的层的联合分布以学习领域不变表示, 还有一些利用领域对齐的域对抗层的方法, 如 Ganin 和 Lempitsky<sup>[33]</sup>引入了一个梯度反转层来欺骗领域分类器并提取领域不变表示, Tzeng 等人<sup>[34]</sup>通过在对抗性训练的跨源域和目标域上适用一个基于 CNN 的深度特征提取器/分类器来解决无监督域自适应。

有研究显示语义对齐有利于对无监督领域自适应问题的解决。假设具有相同标签但来自不同领域的样本之间的距离应尽可能小, Xie 等人<sup>[35]</sup>通过对齐标记的源质心和伪标记的目标质心来学习语义表示; 但是, 大多数语义对齐方法都需要借助目标伪标签, 以最大程度地减少跨域的品种差异, 但是伪标签的不准确性会导致对齐错误的累积。

## 1.2 多源无监督领域适应

大部分的领域自适应方法都针对单个“源领域-目标领域”这种形式展开, 然而实际数据集中通常是拥有多个源的数据, 从而先前表现很好的“单源领域-目标领域”模式就不能很好的适用, 因此多源领域自适应(multi-source unsupervised domain adaptation, MUDA)获得越来越多研究者的关注。多源领域自适应算法是单源领域自适应算法的拓展, 一方面, 多源领域迁移学习解决了因目标领域标签数据过少而不能训练有效分类器的窘况, 另一方面, 在训练过程中依据分类器误差, 区别的对待源领域和目标领域, 降低了源数据与目标数据之间不对等的负影响。同时也面临了一些现实的挑战: (1) 源数据具有多个领域, 限制了主流单源领域自适应方法的有效性; (2) 领域之间存在域偏移; (3) 大规模有效多源域数据的匮乏, 限制了多源领域自适应研究的发展。对于 MUDA 的研究, 早期的多源领域自适应方法主要集中在浅层模型上<sup>[32,36-42]</sup>, 最早的理论支持有: Blitzer 和 Crammer 等人<sup>[36]</sup>通过最小化近源的经验损失, 建立模型预期损失的一般界限。Mansour 等人<sup>[37]</sup>假设目标假设可以用源假设的加权组合来表示。Ben-David 等人<sup>[38]</sup>在源领域和目标领域的加权组合之间引入了。源分类器的加权组合成为流行的方法, Hoffman 等人<sup>[39,40]</sup>为 MUDA 问题提出了一种基于交叉熵损失的理论保证的正则化方法, 并计算了更加精确的加权组合; 在分类和回归的设置下, Zhao 等人<sup>[41]</sup>将半理论模型的泛化边界拓展到了多源领域; Ghifary 等人<sup>[32]</sup>提出了 DRCN 来重建图像特征, 并使转换后的图像近似于原始图像。通常也有一些任务模型是基于带有相应任务损失的源领域数据来学习的, 类如交叉熵损失。大部分的 MUDA 方法都是基于共享特征提取来对称的将多个源领域和目标领域映射到同一特征空间, 虽然有考虑到成对的训练以及每个源领域和目标领域之间的关系, 但是以上方法都并没有从决定数据特征分布的隐变量出发, 因此, 与上述工作不同, 本文结合变分自动编码器<sup>[63,64]</sup>和对抗学习<sup>[40]</sup>的概念来提取领域不变语义表达, 以获取每个源的特定的语义特征表示和领域特征表示。

## 1.3 解耦表达

解耦表达最早来源于 Bengio<sup>[42]</sup>的工作, 其定义是: 在生成模型中, 如果改变表达的一个维度能使得生成模型的输出只改变一个特性, 其他特性不会发生改变, 那么具有这种特性的表达就叫做解耦表达。解耦表达的优良特性获得越来越多的研究人员关注。Bengio 等人<sup>[42]</sup>首先使用了波尔曼兹机来解耦数据的两个不同的因子; 在迁移学习方面, 蔡瑞初等人<sup>[43]</sup>提出了基于提取解耦语义表达的领域自适应, 并且在单源领域自适应中取得了很好的效果。本文基于蔡瑞初等人<sup>[43]</sup>的工作拓展到多源领域自适应中, 而且进一步探究基于解耦的语义隐变量提出了多源隐变量信息聚合策略。

## 2 基于多距离度量框架和加权解耦语义表达的多源领域自适应方法

### 2.1 问题定义

假设有  $M$  个迁移源  $S_1, S_2, \dots, S_M$ , 一个迁移目标  $T$ 。在无监督领域自适应(unsupervised domain adaptation)场景中  $S_1, S_2, \dots, S_M$  的有标签的数据, 目标领域  $T$  是没有标签的数据。对于第  $i$  个源领域  $S_i$ , 根据源分布  $p_i(x, y)$  观察到的图像和得出的相应的标签分别为  $X_i = \{x_i^j\}_{j=1}^{N_i}$  和  $Y_i = \{y_i^j\}_{j=1}^{N_i}$ ,  $N_i$  是迁移源样本的数量。从迁移目标的分布  $p_T(x, y)$  采样得出的样本为  $X_T = \{x_T^j\}_{j=1}^{N_T}$ , 且没有标签, 其中,  $N_T$  表示迁移目标样本的数量。本文假设: (1) 同质

性, 即  $x_i^d \in \mathbb{R}^d$ ,  $x_i^t \in \mathbb{R}^d$ , 这表明来自不同领域的数据是在同一特征空间中观察到的, 但呈现不同的分布; (2) 闭合集, 即  $y_i^d \in y$ ,  $y_i^t \in y$ , 其中,  $y$  是类别标签空间, 指示所有域共享其类别. 本文的目标是同时利用  $M$  个有标签的迁移源数据和一个没有标签的迁移目标数据训练模型, 使得模型可以准确估计迁移目标的样本类别.

### 2.2 模型概览

本节提出基于多距离度量框架和加权解耦语义表达的多源领域自适应方法, 旨在使用带标签的多个源领域数据  $D_S = \{D_{S_1}, D_{S_2}, \dots, D_{S_1}, \dots, D_{S_M}\}$  (其中  $D_{S_i} = \{x_i^{S_i}, y_i^{S_i}\}_{i=1}^n$ ) 和一个无标签的目标领域数据  $D_T = \{x_j^T\}_{j=1}^n$  进行无监督的多源领域自适应学习. 因此需要解决以下 3 个问题: (1) 跨域的领域不变表示是什么? (2) 如何设计可以提取此类领域不变表示的框架? (3) 如何聚合多个源领域的可用信息?

关于第(1)点, 根据数据生成过程的因果机制, 如图 3 所示. 给定  $x$  它是由两个独立的隐变量生成的, 即  $z_d$  对领域信息进行编码,  $z_y$  对语义信息进行编码.  $z_d \in \mathbb{R}^{K_d}$  和  $z_y \in \mathbb{R}^{K_y}$  分别表示领域隐变量和语义隐变量. 考虑到领域信息可能在不同领域之间存在很大差异, 本文假设语义隐变量在提取领域不变表示中起着重要作用. 令  $z = \{z_y, z_d\}$ . 通过进一步发展隐空间中的独立性, 进而假设  $z_y \perp z_d$ .

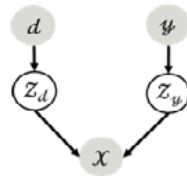


图 3 数据生成过程的因果模型, 由语义隐变量和领域隐变量控制

关于第(2)点, 通过如图 3 所示的数据生成机制, 提出一种解耦的语义表示领域自适应框架, 该框架首先通过变分自动编码器重构两个独立的隐变量, 然后通过对偶的对抗学习对两个独立的隐变量进行解耦. 图 4 中给出了所提出框架的关键结构, 在“重构”架构中, 首先通过特征提取器  $G(x)$ . 例如, ResNet 50 提取图片特征, 其次使用类似于变分自动编码器 VAE (variational auto-encoder)<sup>[65,66]</sup>的方案先将其编码为整体隐变量来重构特征, 继而通过隐变量解耦的两个独立的隐变量  $z_y$  和  $z_d$  重构特征  $G(x)$ . 但是, 与 VAE 不同的是, 本文在模型上进一步设计了解耦模块. 在这部分架构中, 两个对抗模块分别置于语义隐变量和领域隐变量之下. 对于标签对抗学习模块, 其目的是将所有语义信息提取到  $z_y$  中, 并从  $z_d$  中推送所有领域信息. 对于领域对抗学习模块, 其目的是将所有领域信息提取到  $z_d$  中, 并从  $z_y$  推送所有语义信息. 凭此可以获得领域不变的语义信息和不被污染领域信息. 模型结构如图 4 所示.

### 2.3 语义隐变量重构

由于语义隐变量很大程度上控制着数据标签的决策, 为能够最大限度地提取到所有有用的语义信息, 需要对模型的进行语义重构, 确保有效信息在不丢失的情况下进行有效训练. 框架中的重构体系结构遵循 VAE 中的配置, 针对每一对源领域和目标领域进行学习, 将提取的特征重新参数化. 令  $q_\phi(z|x)$  表示为相对于  $\phi$  的编码器, 以近似难以处理的真实后验  $p(z|x)$ . 边际可能性的变化下界如下:

其中,  $P_{\theta'}(x|z)$  表示相对于参数  $\theta'$  的解码器,  $P(z)$  是先验分布. 对于领域  $i$  我们进一步将潜在变量  $z^i$  分解为  $z_y^i$  和  $z_d^i$ , 进而可以得出如下:

$$\begin{aligned} \mathcal{L}_{ELBO}^i(\Phi_y^i, \Phi_d^i, \theta_r^i) = & -D_{KL}(q_{\Phi_y^i}(z_y^i | G(x)) \| P(z_y^i)) - D_{KL}(q_{\Phi_d^i}(z_d^i | G(x)) \| P(z_d^i)) + \\ & E_{q_{\Phi_y^i, \Phi_d^i}(z_y^i, z_d^i | x)}[\log P_{\theta_r^i}(G(x) | z_y^i, z_d^i)] \end{aligned} \quad (1)$$

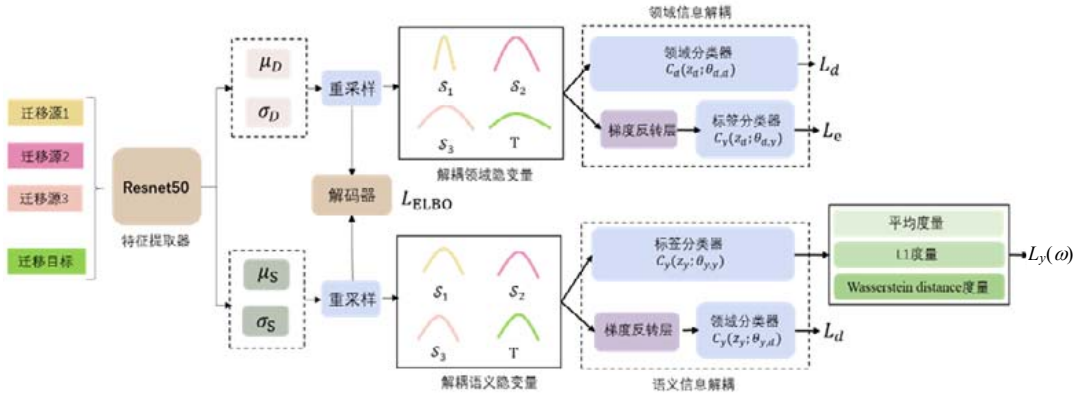


图 4 基于解耦语义表达对齐的多源领域自适应模型

假设  $P(z_y^i)$ ,  $P(z_d^i)$  服从正太分布, 以及  $\Phi_y^i$  和  $\Phi_d^i$  是编码器的参数. 与 VAE 相似, 通过应用重新参数化技巧, 使用 MLP  $H_y(G(x); \Phi_y^i)$  和  $H_d(G(x); \Phi_d^i)$  将数据编码为  $z_y^i$  和  $z_d^i$  并且用于拟合  $P(z_y^i)$ ,  $P(z_d^i)$ . 最后, 将  $M$  对源领域和目标领域的目标函数进行聚合. 本文将会在第 4.6 节详细介绍如何使用多种不同的距离度量方法聚合不同领域的语义隐变量  $z_y^i$ .

2.4 语义隐变量解耦

大部分情况源和目标领域之间数据分布的不同, 因此存在各种程度的差异, 若强制性地迁移不仅不能提升学习效果, 甚至降低模型性能, 即: 产生“负迁移”, 负迁移<sup>[61-64]</sup>是指源数据或者知识对目标任务造成负面影响. 由于数据分布的基础是复杂的流形结构, 领域不变的语义信息和领域特有的信息通常呈现高度耦合, 因此所有的领域自适应方法主要会遇到错误的对齐问题. 对于本文框架中的分离架构, 它由两个对抗模块组成, 这些模块按照 Ganin 和 Lempitsky<sup>[33]</sup>中的典型配置一起工作. 语义隐变量  $z_y^i$  下的是样本标签对抗学习模块, 它融合了样本语义信息并排除所有领域信息, 这是通过使用标签分类器  $C_y$  和领域分类器  $C_d$  完成的. 为了排除领域信息, 领域鉴别器  $C_d$  使用了梯度反转层 GRL<sup>[55]</sup>. 通过最大化  $C_d$  的损失  $L_d$  的同时最小化  $C_y$  的损失  $L_y$  来学习  $H_y$  中的参数  $\Phi_y^i$ .  $C_d$  的参数  $\theta_{y,d}^i$  由  $L_d$  获知. 标签对抗学习模块的总体目标函数如下所示:

$$\begin{aligned} \mathcal{L}_{sem}^i(\Phi_y^i, \theta_{y,y}^i, \theta_{y,d}^i) &= \frac{\delta}{n_S^i M} \sum_{D_{S^i} \in D_S} \alpha^i \sum_{x_j^i \in D_{S^i}} L_y(C_y(H_y(G(x); \Phi_y^i); \theta_{y,y}^i); y^i) - \\ &\quad \frac{\lambda}{n} \sum_{x_j \in (D_{d^i}, D_T)} L_d(C_y(H_y(G(x); \Phi_y^i); \theta_{y,d}^i); d^i) \end{aligned} \tag{2}$$

其中,  $n=n_S+n_T$ ,  $\lambda$  是权衡两个目标的权衡参数, 而  $\delta$  是控制  $L_y$  权重的参数. 较大的  $\delta$  可使标签分类器学习更多的语义信息. 本文的  $\delta$  的默认值为 1. 其中,  $\alpha^i$  表示领域  $i$  的重要性权重, 之后的章节通过多源隐藏变量信息聚合策略将介绍如何获取这个相对重要性权重.

类似地, 在领域对抗学习模块, 其将领域  $i$  的信息融合到  $z_d^i$  并从  $z_y^i$  中排除语义信息. GRL 梯度反转层置于标签分类器上, 以便从  $z_y^i$  吸收所有领域信息, 排除所有的语义信息. 但是, 与语义模块不同, 由于目标域中的无监督学习, 因此这里不使用交叉熵作为标签损失. 为了利用目标域中的数据, 要求标签分类器  $C_y$  采用最大熵损失  $L_e$ . 通过使标签分类器  $C_y$  的损失  $L_e$  最大而使领域  $i$  标签分类器  $C_d$  的损失  $L_d$  最小来学习  $H_d$  中的参数  $\Phi_d^i$ . 另外, 通过使自身的损失  $L_e$  最小化来学习标签分类器  $C_y$  的参数  $\theta_{d,y}^i$ . 关于领域对抗学习模块的目标函数如下:

$$\begin{aligned} \mathcal{L}_{dom}^i(\Phi_d^i, \theta_{d,d}^i, \theta_{d,y}^i) &= \frac{1}{n} \sum_{x_j \in (D_{S^1}, D_T)} L_d(C_d(H_d(G(x); \Phi_d^i); \theta_{d,d}^i), d_i) - \\ &\quad \frac{\omega}{n} \sum_{x_j \in (D_s, D_t)} L_E(C_y(H_d(G(x); \Phi_d^i); \theta_{d,y}^i)) \end{aligned} \quad (3)$$

其中,  $\omega$ 是在建模过程中塑造特征的两个目标之间的权衡参数.

### 2.5 目标函数

提出的针对领域适应模型的解耦语义表示学习的总损失公式为

$$\mathcal{L}(\Phi_y, \theta_{y,y}, \theta_{y,d}, \Phi_d, \theta_{d,d}, \theta_{d,y}, \theta_r) = \sum_{x_j \in (D_{S^1}, D_{S^2}, D_{S^3}, \dots, D_T)} \mathcal{L}_{ELoB}^i + \beta \mathcal{L}_{dom}^i + \sum_{x_j \in (D_{S^1}, D_{S^2}, D_{S^3}, \dots, D_T)} \gamma \mathcal{L}_{sem}^i \quad (4)$$

其中,  $\beta$ 和 $\gamma$ 为不非常敏感的超参数, 设置 $\beta=1$ 和 $\gamma=1$ . 在上述目标函数下, 本文的模型在源领域上使用以下过程进行训练:

$$\left. \begin{aligned} (\hat{\Phi}_y, \hat{\theta}_{y,y}, \hat{\Phi}_d, \hat{\theta}_{d,y}, \hat{\theta}_r) &= \arg \min_{\Phi_y, \theta_{y,y}, \Phi_d, \theta_{d,y}, \theta_r} \mathcal{L}(\Phi_y, \theta_{y,y}, \theta_{y,d}, \Phi_d, \theta_{d,d}, \theta_{d,y}, \theta_r) \\ (\hat{\theta}_{y,d}, \hat{\theta}_{d,d}) &= \arg \min_{\theta_{y,d}, \theta_{d,d}} \mathcal{L}(\Phi_y, \theta_{y,y}, \theta_{y,d}, \Phi_d, \theta_{d,d}, \theta_{d,y}, \theta_r) \end{aligned} \right\} \quad (5)$$

以下分类器具有经过训练的最佳参数.

$$y = C_y(H_y(G(x); \hat{\Phi}_y); \hat{\theta}_{y,y}) \quad (6)$$

### 2.6 多源隐藏变量信息聚合策略

通常情况下迁移源样本和迁移目标样本的是不同分布的, 而且大多情况下, 迁移目标样本和不同迁移源的距离也是不相同的. 一个简单的方法是直接将单源领域自适应方法套用到多源领域自适应问题上, 但是这样的效果往往不会很好. 这是因为在部分迁移源和目标领域之间的相关性比较小的情况下, 强制地迁移不仅不能提升学习效果, 甚至降低模型性能, 即产生“负迁移”. 因此多源领域自适应问题中其中一个重要的挑战就是如何对多个不同源的信息进行聚合. 之前的方法仅仅使用特征提取器提取出来的领域不变性特征来计算不同领域之间的相对重要性权重, 由于领域不变性特征不等于领域语义特征, 而实际上领域不变性特征往往包含领域特有的特征. 仅使用领域特有特征来计算领域相对重要性权重会导致相对重要性权重估计不准确. 为了能最大化发挥领域之间的相似性作用, 以及有效信息的影响力, 本文基于语义解耦框架, 通过计算迁移源和迁移目标的解耦语义隐变量之间的距离来估计不同领域的相对重要性权重. 本文提出了两种信息聚合策略: 基于特征空间距离度量的欧氏距离聚合方法和基于分布距离度量的 Wasserstein distance<sup>[64]</sup>聚合方法.

#### 2.6.1 基于特征空间距离度量的欧式距离聚合方法

本文的第 1 种聚合方法是欧式距离聚合, 对抗学习解耦方法来增强深度网络的领域适应能力, 公式(3)不同于当前单一的源领域和目标领域的分布对齐方法, 本文通过不同的度量方式来度量源域对目标域影响程度计算最佳多源领域迁移的权重分配策略. 具体解释为通过计算不同迁移源和迁移目标之间的欧氏距离来估计不同领域的相对重要性权重, 计算过程如下:

$$\alpha_i = \|z_S^i - z_T\|_2 \quad (7)$$

#### 2.6.2 基于分布距离度量 Wasserstein distance 的聚合方法

Wasserstein distance<sup>[64]</sup>是用于衡量两个分布之间的距离的度量, 当假设语义隐变量和领域隐变量均服从高斯分布时, 采用基于 Wasserstein distance 的聚合方法, 相比于基于欧式距离度量的方法, 更能描述不同分布之间形态特征变换. 考虑使用 Wasserstein distance 度量领域之间的距离的原因在于本模型中不同领域的语义隐变量服从多维高斯分布, 本质上符合 Wasserstein distance 作为度量不同分布的工具的要求. Wasserstein distance 是在一个度量空间上定义, 其中表示集合中两个实例和的距离函数, 给定概率分布  $P$  和概率分布  $Q$ , 他们之间的 Wasserstein distance 定义为

$$W_p(P, Q) = \left( \inf_{\mu \in \mathcal{F}(P, Q)} \int \rho(x, y)^p d\mu(x, y)^{\frac{1}{p}} \right) \quad (8)$$

其中,  $\Pi(P, Q)$ 是在集合  $M \times M$  内以  $P$  和  $Q$  为边缘分布的联合分布. 由 Kantorovich-Rubinstein<sup>[54]</sup>定理可知, 当  $M$  是分离的时候, Wasserstein distance 能够用以下积分概率度量表示:

$$W_1(P, Q) = \sup_{\|H(G(x))\|_{L \leq 1}} E_{x \in Q}[H(G(x))] - E_{y \in P}[H(G(x))] \quad (9)$$

其中,  $\|f\|_{L \leq 1}$  是 Lipschitz 约束. 由公式(24)可知, 给定迁移源  $S^i$  和迁移目标  $T$ ,  $z_y^{S^i} \sim N(\cdot, \cdot)$ ,  $z_y^T \sim N(\cdot, \cdot)$ , 于是

$$W_1(\mathbb{P}_{h^S}, \mathbb{P}_{h^T}) = \sup_{\|H(G(x))\|_{L \leq 1}} E_{\mathbb{P}_{h^S}}[z_y] - E_{\mathbb{P}_{h^T}}[z_d] \quad (10)$$

由于 Lipschitz 约束可通过构造谱归一化层的方法来满足, 根据公式(2)–公式(5), 迁移源头  $S^i$  和迁移目标  $T$  之间的 Wasserstein distance 可以表示为

$$\mathcal{L}_{wd}(x^S, x^T) = \frac{1}{n^S} \sum_{x^S \in X^S} H(G(x^S)) - \frac{1}{n^T} \sum_{x^T \in X^T} H(G(x^T)) \quad (11)$$

### 3 实验设计及结果分析

本文在 3 个公开基准 Office-Home<sup>[58]</sup>数据集和 Office-31<sup>[69]</sup>数据集以及 Image-CLEF 数据集进行了实验验证, 所有的实验使用准确率作为评价指标. 考虑数据集的图片分辨率较高, 所以算法采用 Resnet 50 作为网络骨架提取 2 048 维的特征. 所有数据集实验的学习率均设置为 0.015, 迭代次数为 100 000 次, 每次反向传播的批量大小为 128, 对于所有的迁移方向本文使用了相同的超参数. 对于每组实验本文分别使用 5 个不同的随机种子并且选取最好的实验结果. 其中, 对比算法的超参数和正则项均使用原始论文或源代码中所使用的设计. 模型训练运行环境包含 GTX1080 的 GPU, Intel 7700K CPU 和 32 GB 内存.

#### 3.1 Office-Home 实验

Office-Home<sup>[68]</sup>数据集由 4 个域的日常物体图像组成. 艺术: 绘画、素描和/或艺术描绘, 剪贴画: 剪贴画, 产品: 无背景的图像和真实世界: 用相机拍摄的常规图像; 且各自的样本数目分别为 2 427、4 365、4 439、4 357. 数据集中的图像是使用 Python 网络爬虫收集的, 该网络爬虫通过多个搜索引擎和在线图像目录进行爬网. 最初的运行搜索了大约 120 个不同的对象, 并生成了超过 100 000 种跨不同类别和域的图像. 然后将这些图像过滤以确保所需对象在图片中. 还对类别进行了过滤, 以确保每个类别至少具有一定数量的图像. 该数据集的最新版本包含来自 65 个不同类别的约 15 500 张图像.

在本组实验中, 对比了多种多源领域自适应方法, 其中包括 DANN<sup>[33]</sup>、CDAN<sup>[70]</sup>、M3SDA<sup>[44]</sup>和 MDAN<sup>[45]</sup>, 这些对比方法着重考虑了学习不同的迁移源之间的权重, 但是他们都是基于耦合的特征提取来计算不同迁移源之间的权重, 并不是基于解耦的语义特征. 同时为了表现本文多源领域自适应方法的优越性, 也对比的目前比较流行的单源领域自适应的方法, 包括 DANN<sup>[33]</sup>、DAN<sup>[47]</sup>、CDAN+E<sup>[70]</sup>、CDAN+BSP<sup>[48]</sup>、SAFN<sup>[49]</sup>以及 FDAN-UDA<sup>[24]</sup>、Meta-DANN<sup>[50]</sup>、Meta-MCD<sup>[50]</sup>和 MOSDANET<sup>[51]</sup>这些最新的方法. 根据该数据集的组成形式, 所有算法在四种领域组合下评估其在目标领域的图片分类准确性的性能, 这些领域组合分别为 Clipart、Product、RealWolrd→Art, Art、Product、RealWolrd→Clipart, Art、Clipart、RealWold→Product 和 Art、Clipart、Product→RealWold, 其中“→”两边的字母分别代表源领域和目标领域.

#### 3.2 Office-31 实验

Office-31<sup>[69]</sup>是用于实际领域适应任务的通用数据集. 它由属于 31 个类别的 4 110 张图像组成. 此数据集包含 3 个不同的域, 即分别从亚马逊网站(Amazon 域)、网络摄像头(Webcam 域)和数码 SLR 相机(DSLR 域)收集的图像. 数据集跨域不平衡, 其中, A 域中有 2 817 张图像, W 域中有 795 张图像, D 域中有 498 张图像.

在本组实验中, 除了对比经典的单源领域自适应学习方法例如迁移成分分析 TCA<sup>[66]</sup>和 GFK<sup>[46]</sup>之外, 在基



于分布对齐的方面, 对比了 DAN<sup>[47]</sup>、DDC<sup>[40]</sup>、RevGrad<sup>[33]</sup>、RTN<sup>[55]</sup>和 DRCN<sup>[32]</sup>这些主流算法以及最近的 ResAN<sup>[25]</sup>. 多源领域自适应方面对比了 DCTN<sup>[53]</sup>、sFRAME<sup>[42]</sup>、MDAN<sup>[45]</sup>、MDDA<sup>[54]</sup>以及最新的 LtC-MSDA<sup>[58]</sup>方法. 根据该数其中“ $\rightarrow$ ”两边的字母分别代表源领域和目标领域.

### 3.3 Image-CLEF实验

Image-CLEF 数据集来源于 2014 年的 Domain Adaptation Challenge 比赛, 本文参考最新工作的实验设置, 使用其中的 3 个领域, 他们分别是 Caltech-256 (C)、ImageNet ILSVRC (I)和 Pascal VOC (P). 这个数据集中有 12 个类, 每个类别包含 50 张图片.

在本组实验中, 除了对比一些主流的和最近的单源领域自适应学习方法之外, 基于源领域组合的多源域自适应对比实验有 DAN<sup>[47]</sup>、D-CORAL<sup>[59]</sup>和 RevGrad<sup>[33]</sup>, 并对比了最新的 DCTN<sup>[53]</sup>方法. 根据该数据集的组成形式, 所有算法在 3 种领域组合下评估性能. 这些领域组合分别为 C、P $\rightarrow$ I, I、P $\rightarrow$ C 和 I、C $\rightarrow$ P, 其中“ $\rightarrow$ ”两边的字母分别代表源领域和目标领域.

### 3.4 分离实验

为了验证模型的优越性, 本文设计了如下两组分析实验.

#### (1) 解耦性分析实验

为了验证双重对抗学习模块解耦的有效性, 首先训练基于标准的原始模型直到收敛, 即: 平均加权策略, 逐步将语义信息的权重从 1 提升至 3、5 和 7. 分别在 3 个不同图片分类任务的多源领域自适应基准数据集完成实验验证. 由图 5–图 7 可见, 随着领域信息的相对权重不断减小, 实验效果反而出现下降趋势, 性能下降是因为语义隐变量和领域隐变量的解耦关系被破坏导致语义隐变量中包含领域信息. 这个实验证明了解耦的语义隐变量有利于提升模型泛化性能.

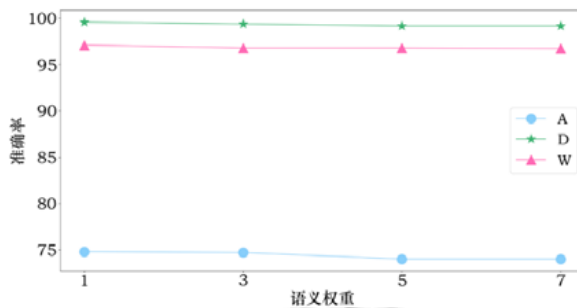


图 5 Office31 数据集在图像分类上的准确性

#### (2) 语义隐变量信息聚合分析实验

本文设计了 3 种聚合实验.

##### (a) 基于平均权重的解耦语义表达的多源领域自适应方法(Ours mean)

最初, 本文假设所有的源领域对目标领域的影响力都是相等的, 那么在此基础上, 设置所有源领域的对目标领域的影响因子即权重都是相等的.

##### (b) 基于 L2 距离度量加权的解耦语义表达的多源领域自适应方法(Ours L2)

因为不同的迁移源对迁移目标的距离不相同, 为了更加突出不同的源领域对目标领域的影响力, 相比于前一种方法, 本文进一步对特征作了 L2 度量, 以此作为一种领域影响力加权策略, 从特征的角度来区分不同的源领域对目标领域的影响程度.

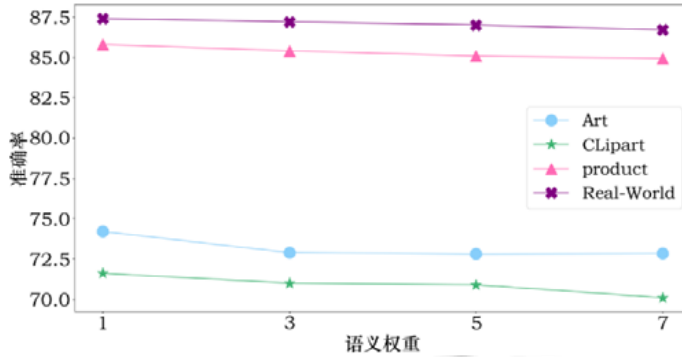


图 6 Office-Home 数据集在图像分类上的准确性

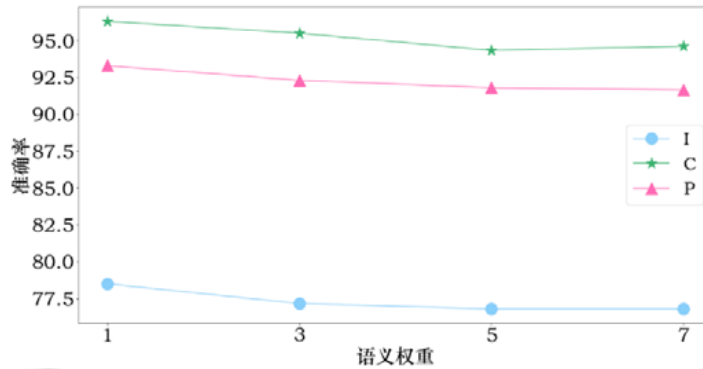


图 7 Image-CLEF 数据集在图像分类上的准确性

(c) 基于 Wasserstein distance 距离度量加权的解耦语义表达的多源领域自适应方法(Ours W-distance)

Wasserstein distance 度量是从分布的层面上来衡量源领域与目标领域的相似度，分布越相似的“源-目标”领域所对应的源领域相对目标领域来说更重要，权重也愈大。语义隐变量信息聚合实验分布在表 1-表 3 中展示，由实验结果我们可以得出以下结论。

表 1 Office-Home 实验结果及其对比实验

Standards	Models	Art	Clipart	Product	Real-World	Avg.
Single best	DANN	63.2	51.8	76.8	70.1	65.5
	DAN	63.1	51.5	74.3	67.9	64.2
	CDAN+E	70.9	56.7	81.6	77.3	71.7
	CDAN+BSP	72.2	59.3	81.9	77.6	72.8
	FDAN-UDA	60.6	53.9	71.5	73.1	64.
	SAFN	70.9	57.1	81.5	77.1	71.7
Source combine	SRC	58.5	57.6	74.5	78.2	67.2
	DANN	58.0	57.7	74.1	78.3	67.0
	Meta-DANN	70.6	59.1	80.2	82.8	73.2
	MOSDANET	62.0	66.0	76.3	78.0	70.5
	CDAN	69.21	67.81	78.36	79.34	73.68
Multi-source	M3SDA	64.7	63.2	76.5	78.9	70.8
	MDAN	68.7	67.3	81.3	82.9	75.1
	Meta-MCD	70.2	60.5	81.2	83.4	73.8
	MDMN	69.2	68.0	83.0	84.0	76.1
	Ours (Mean)	<b>73.0</b>	<b>68.9</b>	<b>84.8</b>	<b>86.6</b>	<b>78.3</b>
	Ours (L2)	<b>74.2</b>	<b>71.6</b>	<b>85.8</b>	<b>87.4</b>	<b>79.5</b>
	Ours (W-distance)	<b>74.3</b>	<b>71.6</b>	<b>86.2</b>	<b>87.5</b>	<b>79.9</b>

(1) 在源领域权重平均的情况下，本文的方法仍然比其他方法优越；这证明了解耦的语义隐变量具有更好的泛化性。

(2) 经过 L2 加权策略的引入, 本文的方法的性能相比平均加权略有提升, 进一步验证通过欧式距离对不同的迁移源加权选择比较重要的迁移源, 从而减轻“负迁移”的影响。

(3) Wasserstein distance 领域加权策略不仅在性能上超过了之前的方法, 甚至也优于前两种加权方法, 这是因为我们的解耦语义隐变量服从高斯分布, 采用 Wasserstein distance 可以更好地计算不同迁移源和迁移目标之间的距离, 从而获得更好的聚合效果。

表 2 Office-31 实验结果及其对比实验

Standards	Models	A	D	W	Avg.
Single best	TCA	51.6	95.2	93.2	68.8
	GFK	52.4	95.0	95.6	68.7
	DDC	52.2	98.5	95.0	70.7
	DRCN	56.0	99.0	96.4	73.6
	RevGrad	53.4	99.2	96.4	74.3
	DAN	54.0	99.0	96.0	72.9
	FDAN-UDA	67.4	92.0	90.9	83.4
	ResAN	69.1	91.5	91.4	83.9
	RTN	51.0	99.6	96.8	73.7
Source combine	DAN	67.6	99.6	97.8	88.3
	D-CORAL	67.1	99.3	<b>98.0</b>	88.1
	RevGrad	67.6	99.7	<b>98.1</b>	88.5
Multi-source	Source only	51.6	98.2	92.7	80.8
	sFRAME	32.1	54.5	52.2	46.3
	MDDA	56.2	99.2	97.1	84.2
	LtC-MSDA	56.9	<b>99.6</b>	<b>97.2</b>	84.6
	DCTN	54.9	<b>99.6</b>	96.9	83.8
	Ours (Mean)	74.5	99.6	96.4	90.1
	Ours (L2)	<b>74.8</b>	<b>99.6</b>	<b>97.1</b>	<b>90.5</b>
	Ours (W-distance)	<b>74.9</b>	<b>99.6</b>	<b>97.1</b>	<b>90.5</b>

表 3 Image-CLEF 实验结果及其对比实验

Standards	Models	I	C	P	Avg.
Single Best	Resnet	74.8	91.5	83.9	83.4
	DDC	74.6	91.1	85.7	83.8
	DAN	75.0	93.3	86.2	84.8
	FDAN-UDA	92.4	94.0	79.0	88.5
	D-CORAL	76.9	93.6	88.5	86.3
	RTN	75.6	95.3	86.9	85.9
Source Combine	DAN	77.6	93.3	92.2	87.7
	D-CORAL	77.1	93.6	91.7	87.5
	RevGrad	77.9	93.7	91.8	87.8
Multi-source	DCTN	75	95.7	90.3	87.0
	Ours(mean)	<b>78.0</b>	<b>95.9</b>	<b>92.8</b>	<b>88.9</b>
	Ours(L2)	<b>78.5</b>	<b>96.3</b>	<b>93.3</b>	<b>89.4</b>
	Ours(W-distance)	<b>79.2</b>	<b>96.4</b>	<b>93.3</b>	<b>89.6</b>

### 3.5 可视化实验

#### 3.5.1 实验结果对比分析

如图 8 所示, 蓝色线代表的是 MDAN, 绿色线代表的是本文的模型在 Office-Home 图片数据集上的预测准确性曲线, 其中源域为 office-Home 数据集中的 Art、Clipart 和 Product; 目标域为 Realworld; 在遵循无监督域自适应的实验设置的情况下, 使用所有标记的源示例和所有未标记的目标示例, 采用 ResNet-50 作为特征提取网络以及其对比算法 MDAN 的超参数和正则项均使用原始论文或源代码中所使用的设计, 在与对比实验相同训练步数的情况下通过图 8 可得出以下结论。

(1) 从图可直观看出, 两种方法都能最终都能比较好的收敛, 且收敛速度差别不明显, 从最终的趋于稳定的结果可以对比得出, 本文模型在一定程度上表现出了更优的性能。

(2) 即使本文的方法和 MDAN 最终都能收敛于不错的结果, 但仍能比较明显的看出本文的模型比 MDAN 的性能更加不稳定, 验证了本文模型的鲁棒性。

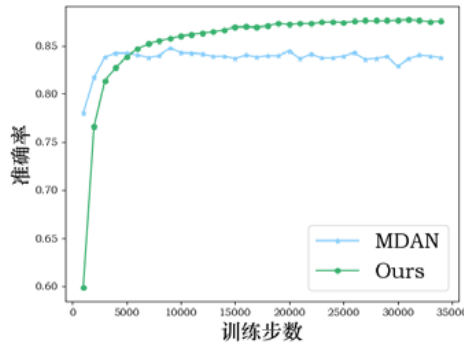


图 8 MDAN 和我们的方法从(Art, Clipart, Product)迁移到 RealWorld 的收敛性结果

### 3.5.2 特征可视化分析

为了验证本文的模型提取的基于多距离度量框架和加权解耦语义表达特征是领域不变的, 通过使用 t-SNE 对提取的特征进行可视化分析. 为了能够更加清晰地表现实验结果的对比差异, 通过在包含了 65 个类别的 Office-Home 数据集中, 选取其中的 10 个类别, 如图 9 所示, 4 种颜色代表了 4 个领域, 其中包括了来自 3 个源域和 1 个目标域 10 个类别的样本, 分别为 Art、Clipart、Product 和 Realworld, 其中, MDAN 和本文的模型都能将每个种类的数据点划分成簇, 表现了非常好的性能, 相对而言, 本文方法划分的类别簇的界线更加分明, 表明了本文的实验性能更优, 进一步验证了其有效性.

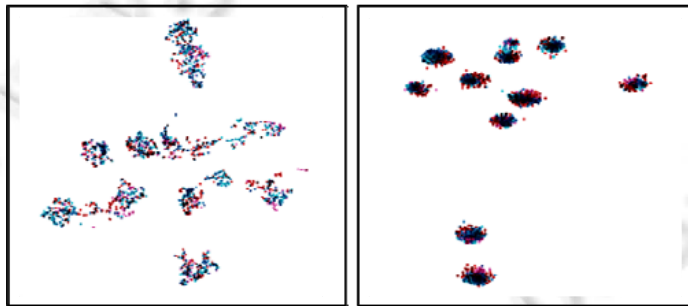


图 9 由 Ours (右), MDAN (左)提取的深层特征的 t-SNE 可视化

### 3.5.3 混淆矩阵可视化分析

为了进一步验证该方法可以缓解负迁移的问题, 本文添加了可视化混淆矩阵分析, 实验结果如图 10 所示. 图 10(a)为 ResNet 50 分类以及图 10(b)为 DANN 方法的混淆矩阵, 根据图 10(a)、图 10(b)所示, ResNet 50 与 DANN 方法分类错误的情况相对较多, 其原因在于迁移源和迁移目标的信息发生了错误的匹配, 最终导致了较严重的负迁移结果. 而本文方法通过解耦语义信息和领域信息, 一定程度上缓解了特征信息错配, 从而缓解了负迁移现象. 通过比较图 10(a)和图 10(b)中绿色方框部分, 可以说明 DANN 具有正向迁移的效果. 与此同时, DANN 方法也发生了错误匹配所引起的负迁移现象, 体现为图 10(a)和图 10(b)中红色方框内的颜色更深, 相比而言, 本文的方法有效地缓解了负迁移问题, 如图 10(c)所示, 且本文方法整体分类性能相对前两种方法更佳.

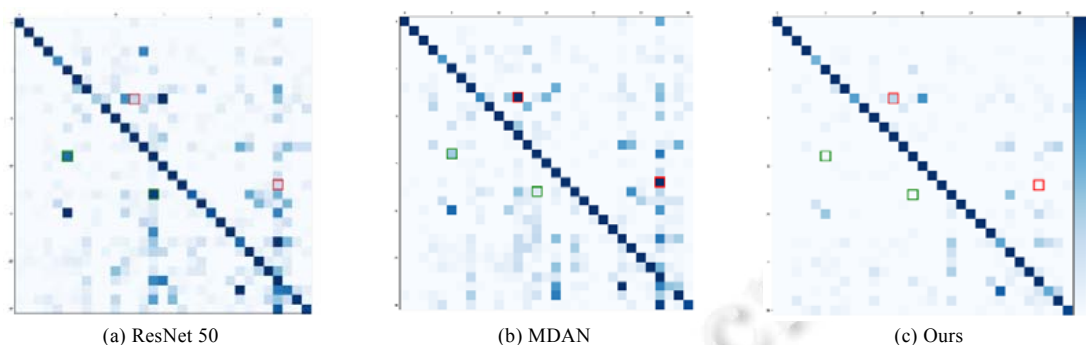


图 10 Office 31 数据集中 DSRL 领域的混淆矩阵示意图

### 3.5.4 定量分析

为了证明本文提出的方法比现有方法更加有效, 本文进一步在不同隐变量维度下进行实验, 实验结果如图 11 所示, 由此可以得出如下结论: 在不同维度下, 本文提出的方法和 MDAN 方法性能相比有明显差异且效果更好, 这说明了本文提出的方法具有良好的稳定性.

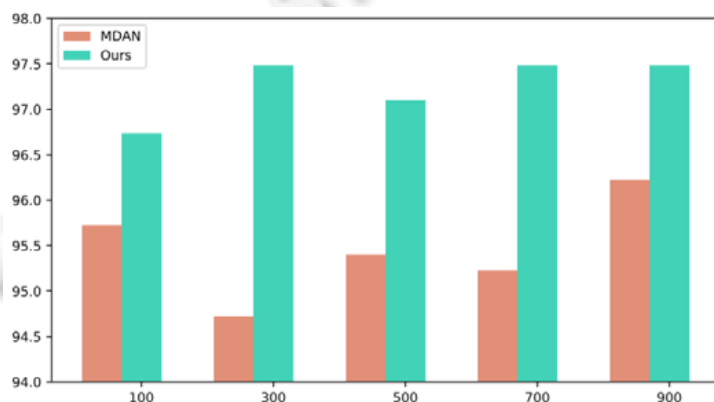


图 11 不同隐变量维度下 MDAN 和本文提出方法的实验效果柱形图

## 4 结束语

本文提出了一种有效的多源领域自适应方法, 每个迁移源分别进行预训练的特征提取器和分类器, 可以充分探索标记源数据的可分辨性. 在这个基础上, 本文设计了一种新颖的解耦表示方法, 充分保护了语义信息和领域信息的独立性, 其次通过基于语义隐变量加权策略结合来自不同来源分类器的预测, 最后采用了多种语义隐变量聚合策略学习不同迁移源关于迁移目标的权重, 并提取加权的解耦语义隐变量用于分类. 在 Office-home、Office-31 以及 Image-CLEF 基准数据集上进行的广泛实验表明了架构的有效性. 在未来的研究中, 我们计划将该模型扩展和优化到更具挑战性的多源领域适应.

### References:

- [1] Xu M, Wang ST, Gu X, *et al.* Support vector regression for large domain adaptation. *Ruan Jian Xue Bao/Journal of Software*, 2013, 24(10): 2312–2326 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4375.htm> [doi: 10.3724/SP.J.1001.2013.04375]
- [2] Ge H. Research on image classification method based on deep learning [Ph.D. Thesis]. Chengdu: University of Electronic Science and Technology of China, 2020 (in Chinese with English abstract).

- [3] Wang JX, Wang ZY, Tian X. Review of natural scene text detection and recognition based on deep learning. *Ruan Jian Xue Bao/Journal of Software*, 2020, 31(5): 1465–1496 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5988.htm> [doi: 10.13328/j.cnki.jos.005988]
- [4] Jiang F, Gu Q, Hao HZ, *et al.* Survey on content-based image segmentation methods. *Ruan Jian Xue Bao/Journal of Software*, 2017, 28(1): 160–183 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5136.htm> [doi: 10.13328/j.cnki.jos.005136]
- [5] Gao YS. Research on the application and development of deep learning in the field of computer vision. *Computer Programming Skills and Maintenance*, 2020(9): 125–127 (in Chinese with English abstract).
- [6] Wang YW. Research on deep learning classification network and its application in computer vision. *Communication World*, 2019, 26(3): 195–196 (in Chinese with English abstract).
- [7] Zhang X, Xiao XL, Xu GY. Fuzzy support vector machine based on affinity among samples. *Ruan Jian Xue Bao/Journal of Software*, 2006, 17(5): 951–958 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/17/951.htm>
- [8] Sankaranarayanan S, Balaji Y, Jain A, *et al.* Learning from synthetic data: Addressing domain shift for semantic segmentation. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2018.
- [9] Luo YW, Zheng L, Guan T, *et al.* Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. In: *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*. 2019.
- [10] Zhuang FZ, Luo P, He Q, *et al.* Research progress of transfer learning. *Ruan Jian Xue Bao/Journal of Software*, 2015, 26(1): 26–39 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4631.htm> [doi: 10.13328/j.cnki.jos.004631]
- [11] Cai RC, Li JH, Hao ZF. Unsupervised domain adaptive algorithm based on intra-class maximum mean difference. *Computer Application Research*, 2020, 37(8): 2371–2375 (in Chinese with English abstract).
- [12] Tang S, Ye M, Li XD. Overview of domain adaptive target recognition. *ZTE Technology*, 2017, 23(4): 25–31 (in Chinese with English abstract).
- [13] Tao JW, Wang ST. Multiple kernel local leaning-based domain adaptation. *Ruan Jian Xue Bao/Journal of Software*, 2012, 23(9): 2297–2310 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4240.htm> [doi: 10.3724/SP.J.1001.2012.04240]
- [14] Xu M. Research on domain adaptive learning algorithm and its application [Ph.D. Thesis]. Wuxi: Jiangnan University, 2014 (in Chinese with English abstract).
- [15] Gu X. Cross-domain classification learning method and application research [Ph.D. Thesis]. Wuxi: Jiangnan University, 2014 (in Chinese with English abstract).
- [16] Han X. Open-set image classification and semantic segmentation based on domain adaptation [Ph.D. Thesis]. Nanjing: Nanjing University, 2019 (in Chinese with English abstract).
- [17] Fu JH. Research on deep transfer learning algorithm and its application [Ph.D. Thesis]. Nanjing: Nanjing University of Posts and Telecommunications, 2020 (in Chinese with English abstract).
- [18] Zang WH. Research on transfer learning algorithm based on generative adversarial networks [Ph.D. Thesis]. Chengdu: University of Electronic Science and Technology of China, 2018 (in Chinese with English abstract).
- [19] Li X. Research on cross-domain image classification method based on transfer learning [Ph.D. Thesis]. Xi'an: Xidian University, 2017 (in Chinese with English abstract).
- [20] Sun Q, Ling WX. Source domain selection in transfer learning based on domain similarity rank. *Science Technology and Engineering*, 2020, 20(20): 8245–8251 (in Chinese with English abstract).
- [21] Huang L. Research on a class of natural image classification algorithms based on transfer learning [Ph.D. Thesis]. Zhengzhou: Zhengzhou University, 2020 (in Chinese with English abstract).
- [22] Hu H. Research on cross-domain image classification algorithm based on domain adaptation [Ph.D. Thesis]. Chongqing: Chongqing University of Posts and Telecommunications, 2020 (in Chinese with English abstract).
- [23] He Z, Guo H. Comparative study on transfer learning algorithms based on data distribution adaptive. *Software Guide*, 2020, 19(8): 80–83 (in Chinese with English abstract).
- [24] Huang XH. Research on image classification based on deep domain adaptation [Ph.D. Thesis]. Xuzhou: China University of Mining and Technology, 2020 (in Chinese with English abstract).

- [25] He ZH. Research on unsupervised domain adaptive algorithm based on adversarial methods [Ph.D. Thesis]. Chengdu: University of Electronic Science and Technology of China, 2020 (in Chinese with English abstract).
- [26] Bai YW. Research on multi-source transfer learning method based on graph structure [Ph.D. Thesis]. Harbin: Harbin Institute of Technology, 2020 (in Chinese with English abstract).
- [27] Li T. Research on multi-source transfer learning method based on graph convolutional neural network [Ph.D. Thesis]. Harbin: Harbin Institute of Technology, 2020 (in Chinese with English abstract).
- [28] Wu HR. Research on transfer learning for multiple source domains [Ph.D. Thesis]. Guangzhou: South China University of Technology, 2020 (in Chinese with English abstract).
- [29] Fang M, Xu X, Du H, *et al.* Image Classification Method Based on Adversarial Fusion Multi-source Transfer Learning. Shaanxi Province: CN111738315A, 2020-10-02 (in Chinese with English abstract).
- [30] Liu Z, Yang JA, Liu H, *et al.* Multi-source domain transfer learning classification algorithm based on domain correlation and manifold constraints. *Computer Application Research*, 2017, 34(2): 351–356 (in Chinese with English abstract).
- [31] Ganin Y, Ustinova E, Ajakan H, *et al.* Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 2016, 17(1): 2096–2030.
- [32] Ghifary M, Kleijn WB, Zhang MJ, *et al.* Deep reconstruction-classification networks for unsupervised domain adaptation. In: *Proc. of the European Conf. on Computer Vision*. Cham: Springer, 2016.
- [33] Ganin Y, Lempitsky V. Unsupervised domain adaptation by backpropagation. In: *Proc. of the ICML*. 2015.
- [34] Tzeng E, Hoffman J, Zhang N, *et al.* Deep domain confusion: Maximizing for domain invariance. arXiv:1412.3474, 2014.
- [35] Xie J, Hu W, Zhu SC, *et al.* Learning sparse FRAME models for natural image patterns. *Int'l Journal of Computer Vision*, 2015, 114(2–3): 91–112.
- [36] Gong B, Shi Y, Sha F, *et al.* Geodesic flow kernel for unsupervised domain adaptation. In: *Proc. of the CVPR*. 2012.
- [37] MehryarMohri MY, Rostamizadeh A. Domain adaptation with multiple sources. *Advances in Neural Information Processing Systems*, 2008, 21: 1041–1048.
- [38] Ben David S, Blitzer J, Crammer K, *et al.* Analysis of representations for domain adaptation. *Advances in Neural Information Processing Systems*, 2007, 19: 137.
- [39] Tzeng E, Hoffman J, Darrell T, *et al.* Simultaneous deep transfer across domains and tasks. In: *Proc. of the IEEE Int'l Conf. on Computer Vision*. 2015. 4068–4076.
- [40] Tzeng E, Hoffman J, Zhang N, *et al.* Deep domain confusion: Maximizing for domain invariance. 2014.
- [41] Lin C, Zhao S, Meng L, *et al.* Multi-source domain adaptation for visual sentiment classification. In: *Proc. of the AAAI Conf. on Artificial Intelligence*. 2020, 34(3): 2661–2668.
- [42] Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2013, 35(8): 1798–1828.
- [43] Cai R, Li Z, Wei P, *et al.* Learning disentangled semantic representation for domain adaptation. In: *Proc. of the IJCAI Conf. NIH Public Access*, 2019. 2060.
- [44] Peng X, Bai Q, Xia X, *et al.* Moment matching for multi-source domain adaptation. In: *Proc. of the IEEE Int'l Conf. on Computer Vision*. 2019.
- [45] Zhao H, Zhang S, Wu G, *et al.* Adversarial multiple source domain adaptation. *Advances in Neural Information Processing Systems*, 2018, 31: 8559–8570.
- [46] Peng X, Huang Z, Zhu Y, *et al.* Federated adversarial domain adaptation. arXiv:1911.02054, 2019.
- [47] Long M, Cao Y, Wang J, *et al.* Learning transferable features with deep adaptation networks. In: *Proc. of the ICML*. 2015.
- [48] Chen XY, Wang S, Long MS, *et al.* Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation. In: *Proc. of the Int'l Conf. on Machine Learning*. PMLR, 2019.
- [49] Xu RJ, Li GB, Yang JH, *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*. 2019.
- [50] Li D, Hospedales T. Online meta-learning for multi-source and semi-supervised domain adaptation. In: *Proc. of the European Conf. on Computer Vision*. Cham: Springer, 2020.

- [51] Sayan R, Tamboli D, Shuddhodhan P, *et al.* Multi-source open-set deep adversarial domain adaptation. In: Proc. of the European Conf. on Computer Vision. Cham: Springer, 2020.
- [52] Pan SJ, Tsang IW, Kwok JT, *et al.* Domain adaptation via transfer component analysis. TNNLS, 2011, 22(2): 199–210.
- [53] Xu R, Chen Z, Zuo W, *et al.* Deep cocktail network: Multi-source unsupervised domain adaptation with category shift. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2018. 3964–3973.
- [54] Zhao SC, Wang GZ, Zhang SH, *et al.* Multi-source distilling domain adaptation. In: Proc. of the AAAI Conf. on Artificial Intelligence, Vol. 34. 2020.
- [55] Long MS, Zhu H, Wang JM, *et al.* Unsupervised domain adaptation with residual transfer networks. arXiv:1602.04433, 2016.
- [56] Dziugaite GK, Roy DM, Ghahramani Z. Training generative neural networks via maximum mean discrepancy optimization. arXiv: 1505.03906, 2015.
- [57] Daniel L, Meek C. Adversarial learning. In: Proc. of the 11th ACM SIGKDD Int'l Conf. on Knowledge Discovery in Data Mining. 2005.
- [58] Wang H, Xu MH, Ni BB, *et al.* Learning to combine: Knowledge aggregation for multi-source domain adaptation. In: Proc. of the European Conf. on Computer Vision. Cham: Springer, 2020.
- [59] Sun BC, Saenko K. Deep Coral: Correlation alignment for deep domain adaptation. In: Proc. of the European Conf. on Computer Vision. Cham: Springer, 2016.
- [60] Lisa T, Shavlik J. Transfer learning. In: Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques. IGI Global, 2010. 242–264.
- [61] Rosenstein MT, Marx Z, Kaelbling LP, *et al.* To transfer or not to transfer. In: Proc. of the NIPS 2005 Workshop on Transfer Learning, Vol.898. 2005.
- [62] Ge L, Gao J, Ngo H, *et al.* On handling negative transfer and imbalanced distributions in multiple source transfer learning. Statistical Analysis and Data Mining: The ASA Data Science Journal, 2014, 7(4): 254–271.
- [63] Li YS, He JZ, Zhao GQ, *et al.* Research on the direction of negative transfer in transfer learning. Command Control and Simulation, 2020, 42(4) :28–33.
- [64] Shen J, Qu Y, Zhang W, *et al.* Wasserstein distance guided representation learning for domain adaptation. arXiv:1707.01217, 2017.
- [65] Carl D. Tutorial on variational autoencoders. arXiv:1606.05908, 2016.
- [66] Zhai ZL, Liang ZM, Zhou W, *et al.* Review of variational autoencoder models. Computer Engineering and Applications, 2019, 55(3): 1–9.
- [67] Lellmann J, Lorenz DA, Schonlieb C, *et al.* Imaging with Kantorovich-Rubinstein discrepancy. SIAM Journal on Imaging Sciences, 2014, 7(4): 2833–2859.
- [68] Venkateswara H, Eusebio J, Chakraborty S, *et al.* Deep hashing network for unsupervised domain adaptation. In: Proc. of the CVPR. 2017.
- [69] Saenko K, Kulis B, Fritz M, *et al.* Adapting visual category models to new domains. In: Proc. of the ECCV. 2010.
- [70] Long MS, Cao ZJ, Wang JM, *et al.* Conditional-adversarial-domain-adaptation. arXiv:1705.10667, 2017.

#### 附中文参考文献:

- [1] 许敏, 王士同, 顾鑫, 等. 大样本领域自适应支撑向量回归机. 软件学报, 2013, 24(10): 2312–2326. <http://www.jos.org.cn/1000-9825/4375.htm> [doi: 10.3724/SP.J.1001.2013.04375]
- [2] 葛昊. 基于深度学习的图像分类方法研究 [博士学位论文]. 成都: 电子科技大学, 2020.
- [3] 王建新, 王子亚, 田萱. 基于深度学习的自然场景文本检测与识别综述. 软件学报, 2020, 31(5): 1465–1496. <http://www.jos.org.cn/1000-9825/5988.htm> [doi: 10.13328/j.cnki.jos.005988]
- [4] 姜枫, 顾庆, 郝慧珍, 等. 基于内容的图像分割方法综述. 软件学报, 2017, 28(1): 160–183. <http://www.jos.org.cn/1000-9825/5136.htm> [doi: 10.13328/j.cnki.jos.005136]
- [5] 高玉双. 深度学习在计算机视觉领域的应用发展探究. 电脑编程技巧与维护, 2020(9): 125–127.
- [6] 王一玮. 深度学习分类网络研究及其在计算机视觉中的应用. 通信世界, 2019, 26(3): 195–196.



- [7] 张翔, 肖小玲, 徐光祐. 基于样本之间紧密度的模糊支持向量机方法. 软件学报, 2006, 17(5): 951–958. <http://www.jos.org.cn/1000-9825/17/951.htm>
- [10] 庄福振, 罗平, 何清, 等. 迁移学习研究进展. 软件学报, 2015, 26(1): 26–39. <http://www.jos.org.cn/1000-9825/4631.htm> [doi: 10.13328/j.cnki.jos.004631]
- [11] 蔡瑞初, 李嘉豪, 郝志峰. 基于类内最大均值差异的无监督领域自适应算法. 计算机应用研究, 2020, 37(8): 2371–2375.
- [12] 唐宋, 叶茂, 李旭冬. 领域自适应目标识别综述. 中兴通信技术, 2017, 23(4): 25–31.
- [13] 陶剑文, 王士同. 多核局部领域适应学习. 软件学报, 2012, 23(9): 2297–2310. <http://www.jos.org.cn/1000-9825/4240.htm> [doi: 10.3724/SP.J.1001.2012.04240]
- [14] 许敏. 领域自适应学习算法及其应用研究 [博士学位论文]. 无锡: 江南大学, 2014.
- [15] 顾鑫. 跨领域分类学习方法及应用研究 [博士学位论文]. 无锡: 江南大学, 2014.
- [16] 韩旭. 基于领域自适应的开放集图像分类与语义分割 [博士学位论文]. 南京: 南京大学, 2019.
- [17] 付家慧. 深度迁移学习算法及其应用研究 [博士学位论文]. 南京: 南京邮电大学, 2020.
- [18] 臧文华. 基于生成对抗网络的迁移学习算法研究 [博士学位论文]. 成都: 电子科技大学, 2018.
- [19] 李晓. 基于迁移学习的跨域图像分类方法研究 [博士学位论文]. 西安: 西安电子科技大学, 2017.
- [20] 孙俏, 凌卫新. 基于域间相似度序数的迁移学习源领域的选择. 科学技术与工程, 2020, 20(20): 8245–8251.
- [21] 黄露. 基于迁移学习的一类自然图像分类算法研究 [博士学位论文]. 郑州: 郑州大学, 2020.
- [22] 胡昊. 基于领域自适应的跨领域图像分类算法研究 [博士学位论文]. 重庆: 重庆邮电大学, 2020.
- [23] 和泽, 郭辉. 基于数据分布自适应的迁移学习算法比较研究. 软件导刊, 2020, 19(8): 80–83.
- [24] 黄霄汉. 基于深度领域适应的图像分类研究 [博士学位论文]. 徐州: 中国矿业大学, 2020.
- [25] 何志海. 基于对抗方法的非监督域自适应算法研究 [博士学位论文]. 成都: 电子科技大学, 2020.
- [26] 白雅雯. 基于图结构的多源迁移学习方法研究 [博士学位论文]. 哈尔滨: 哈尔滨工业大学, 2020.
- [27] 李韬. 基于图卷积神经网络的多源迁移学习方法研究 [博士学位论文]. 哈尔滨: 哈尔滨工业大学, 2020.
- [28] 吴汉瑞. 面向多个源领域的迁移学习研究 [博士学位论文]. 广州: 华南理工大学, 2020.
- [29] 方敏, 徐筱, 杜辉, 等. 基于对抗融合多源迁移学习的图像分类方法. CN111738315A, 2020-10-02.
- [30] 刘振, 杨俊安, 刘辉, 等. 基于域相关性与流形约束的多源域迁移学习分类算法. 计算机应用研究, 2017, 34(2): 351–356.
- [63] 李永盛, 何佳洲, 赵国清, 等. 关于迁移学习中的负迁移方向研究. 指挥控制与仿真, 2020, 42(4): 28–33.
- [66] 翟正利, 梁振明, 周炜, 等. 变分自编码器模型综述. 计算机工程与应用, 2019, 55(3): 1–9.



蔡瑞初(1983—), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为因果发现, 图神经网络, 领域自适应学习, 自然语言处理.



李梓健(1994—), 男, 博士生, 主要研究领域为迁移学习.



郑丽娟(1996—), 女, 硕士, 主要研究领域为领域自适应学习, 多源领域迁移学习.