



计算机科学

COMPUTER SCIENCE

融合知识的领域自适应方法综述

崔福伟, 吴璇璇, 陈钰枫, 刘健, 徐金安

引用本文

崔福伟, 吴璇璇, 陈钰枫, 刘健, 徐金安 [融合知识的领域自适应方法综述](#) [J]. 计算机科学, 2023, 50(8): 142-149.

CUI Fuwei, WU Xuanxuan, CHEN Yufeng, LIU Jian, XU Jin'an. [Survey of Domain Adaptive Methods with Knowledge Integrating](#) [J]. Computer Science, 2023, 50(8): 142-149.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于领域适应嵌入的军事命名实体识别](#)

Name Entity Recognition for Military Based on Domain Adaptive Embedding
计算机科学, 2022, 49(1): 292-297. <https://doi.org/10.11896/jsjcx.201100007>

[基于深度动态联合自适应网络的图像识别方法](#)

Image Recognition with Deep Dynamic Joint Adaptation Networks
计算机科学, 2021, 48(6): 131-137. <https://doi.org/10.11896/jsjcx.210100008>

[多跳无线网络干扰攻击建模与性能分析](#)

Modeling of Jamming Attack and Performance Analysis in Multi-hop Wireless Network
计算机科学, 2019, 46(11A): 414-416.

[一种模糊XML模型的概念设计方法](#)

Conceptual Design Methodology for Fuzzy XML Model
计算机科学, 2011, 38(12): 156-161.

[基于用户自描述标签的层次分类体系构建方法](#)

Taxonomy Construction Based on User Self-describing Tags
计算机科学, 2016, 43(7): 224-229. <https://doi.org/10.11896/j.issn.1002-137X.2016.07.040>

融合知识的领域自适应方法综述

崔福伟¹ 吴璇璇² 陈钰枫² 刘健² 徐金安²

1 北京交通大学电子信息工程学院 北京 100044

2 北京交通大学计算机与信息技术学院 北京 100044

(fuweicui@bjtu.edu.cn)

摘要 训练基于数据驱动模型时,常假设源域和目标域的数据分布相同,但是在实际场景中,这一假设通常不成立,因此容易造成模型的泛化能力较差的问题。为提高模型的泛化能力,领域自适应方法应运而生,其通过学习源域和目标域的数据特征来对齐两域数据分布,使得在源域数据上训练好的模型在有少量数据标签或者没有数据标签的目标域上也具有较好表现。为了进一步提高模型的泛化能力,现有研究探索将知识融入领域自适应方法中,该技术具有较高的实用价值。文中首先概述了融合知识的领域自适应方法的发展背景和相关综述的研究现状;其次对领域自适应的问题定义和理论基础进行了介绍;然后给出了一种融合知识的领域自适应方法的分类体系,并对其中的一些代表性方法进行了概述;最后,通过对该领域挑战性问题的分析,预测了融合知识的领域自适应方法未来的研究方向,以期为相关的研究提供一定的参考。

关键词: 泛化能力; 领域自适应; 融合知识; 分类体系

中图分类号 TP391

Survey of Domain Adaptive Methods with Knowledge Integrating

CUI Fuwei¹, WU Xuanxuan², CHEN Yufeng², LIU Jian² and XU Jin'an²

1 School of Electronic Information Engineering, Beijing Jiaotong University, Beijing 100044, China

2 School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

Abstract When training a data-driven model, it is often assumed that the data distribution of the source domain and the target domain are the same. However, in the natural scenario, this assumption is usually not tenable, and it is easy to cause poor generalization ability of the model. Domain adaptation is a method proposed to improve the generalization ability of the model. It aligns the data distribution of the source domain and the target domain by learning the data characteristics of the two domains, so that the model trained in the source domain data can also perform well in the target domain with a small number of data labels or without data labels. In order to further improve the generalization ability of the model, existing researches have explored the knowledge integrating into domain adaptive methods, which has high practical value. Firstly, we summarize the development background of domain adaptive methods with knowledge integrating and the research status of related reviews. Then, the problem definition and theoretical basis of domain adaptation are introduced. After that, a classification system of domain adaptive methods with knowledge integrating is presented, and some representative methods are summarized. Finally, through the analysis of the challenging problems in this field, the future research directions of domain adaptive methods with knowledge integrating are predicted, in the hope of providing some reference for related research.

Keywords Generalization ability, Domain adaptation, Knowledge integrating, Classification system

1 引言

随着计算机网络技术和信息采集技术的发展,基于数据驱动的机器学习算法尤其是深度学习算法在计算机视觉、自然语言处理等领域得到了广泛的应用。其中,监督学习算法是最成熟、应用最为广泛的一个分支。传统的监督学习算法假设训练数据和测试数据具有相同的概率分布。在这样的

假设条件下,利用标记好的训练数据学习到的模型在面对未标记的测试数据时也可做出较好预测。但是,在实际应用中,如人脸识别、语音识别、物体识别、对话生成等领域,上述假设条件很难成立,训练数据和测试数据往往是在不同的环境中采集且采集的数据可能会随着时间发生缓慢变化,很难保证具有相同的概率分布,此时,监督学习算法的泛化能力就较差。解决这一问题的一种方法是在测试环境中再采集和标注

到稿日期:2022-08-03 返修日期:2022-11-10

基金项目:国家重点研发计划(2019YFB1405200);国家自然科学基金(61976015,61976016,61876198,61370130)

This work was supported by the National Key R & D Program of China(2019YFB1405200) and National Natural Science Foundation of China(61976015,61976016,61876198,61370130).

通信作者:徐金安(jaxu@bjtu.edu.cn)

一些数据,对模型重新进行训练。但是,数据的采集和标注工作往往既费时又费力,特别是面对一些新的业务场景时,可能会有较少的数据或者没有数据,该方法实行起来比较困难。因此,如何有效利用标注好的训练数据,在未标注或者少量标注的测试数据上获得较好的预测性能,成为提高模型泛化能力的一个重要问题。

为了解决这一问题,国内外学者在如何将源域的知识传递到目标域以改善目标域的预测效果等方面进行了许多探索,其目标是找到源域和目标域之间的桥梁,将在源域学习到的知识迁移到目标域,增强在目标域的预测效果。解决这一问题的方法称为领域自适应(Domain Adaptation, DA)。领域自适应属于迁移学习的一种。迁移学习的定义如下:给定源域、源任务、目标域和目标任务,将源域在解决源任务过程中学习到的知识迁移到目标任务,以帮助目标任务的完成。迁移学习的方法主要包含3类^[1]:1)归纳迁移学习(Inductive Transfer Learning),无论目标域和源域是否相同,目标任务和源任务都不同,且需要目标域中有一些已标注好的数据;2)转导迁移学习(Transductive Transfer Learning),源域和目标域不同,源任务和目标任务相同;3)无监督迁移学习(Unsupervised Transfer Learning),源域和目标域不同,目标任务和源任务不同但相关。其中,领域自适应属于上述迁移学习分类的第二种,且一般认为源域和目标域的特征空间相同。

领域自适应是一种源任务和目标任务相同,但是源域和目标域的数据分布不同,并且源域中有大量的标记好的样本,而目标域中有非常少或者没有标记样本的迁移学习方法。该方法初期是机器学习与计算机视觉领域内的前沿研究方向之一,在人脸识别、物体识别等方面有极大的应用前景,随后逐渐扩展到自然语言处理领域的一些应用场景中,如机器翻译、语音识别、文本分类等。领域自适应能一定程度上缓解目标领域标注数据稀缺的问题,极大地节约重新训练模型的成本,较好地提高模型的泛化能力。

近年来,为了进一步提高模型的泛化能力,国内外研究人员将语言学知识或者非语言学知识融入领域自适应方法中,并取得了一定的提升效果。基于知识的领域自适应方法的主要思想是不同领域之间存在相同或者相关的知识关系,比如教育领域与培训领域、医疗领域和养生领域等,利用相关领域的知识辅助目标任务的完成。国内外学者从实现手段、参与域数量、是否有监督等角度对领域自适应方法进行了综述,但是,鲜有工作从融合知识的角度对该领域进行综述。鉴于此,本文从融合知识的角度对领域自适应方法进行综述。

本文的贡献如下:

- (1)首次从融合知识的角度对领域自适应方法进行综述。
- (2)提出了一种融合知识的领域自适应方法的分类体系,并对各类方法的代表性工作进行分析,总结了各类方法的优缺点。
- (3)总结了融合知识的领域自适应方法面临的挑战,并提出了未来可能的研究方向,以期为该领域的相关研究提供一定的借鉴。

本文第2节介绍了领域自适应方法的研究现状,并指出了本文与之前工作的不同之处;第3节对领域自适应问题的

定义与理论基础进行了介绍;第4节对国内外融合知识的领域自适应技术进行深入分析研究,提出了一种新的融合知识的领域自适应分类体系,并介绍和总结各类融合知识的领域自适应算法的代表性工作及其优缺点;第5节总结了融合知识的领域自适应技术存在的挑战并给出了未来可能的研究方向;最后总结全文。

2 研究现状

领域自适应技术在现实生活中有着广泛的应用,但是直到近几年才受到研究者的重视。目前,领域自适应技术的理论研究已经相对成熟,许多领域自适应方法被提出,且应用范围越来越广。针对庞杂繁多的领域自适应方法,国内外的一些学者尝试对现有的领域自适应方法进行综述,例如,文献[2]主要对统计分类领域自适应方法进行了综述,将统计分类领域自适应方法分为基于实例权重、基于半监督学习、基于表示改变、基于贝叶斯先验、基于多任务学习和基于集成六大类;文献[3]着重从概率分布适配的角度,对基于统计量适配、边际分布适配、边际分布与类条件分布适配等领域自适应方法进行了详细介绍;文献[4]创新性地将领域自适应分为基于实例、基于特征、基于参数和基于相关知识的领域间自适应学习方法;文献[5]将领域自适应分为单源域单目标域自适应学习和多源域多目标域自适应学习,其中单源域单目标域自适应学习分为基于特征和参数分解的域自适应学习、基于特征表示的域自适应学习和基于样例加权域自适应学习,多源域多目标域自适应学习包括基于流行的方法、基于正则化项组合的方法和基于源域判别函数组合的方法;文献[6]从3个角度对领域自适应方法分别进行了详细介绍,从是否有监督角度将领域自适应分为无监督、半监督和弱监督领域自适应,从参与域数量角度将领域自适应分为单源域、多源域和多目标域领域自适应,从特征空间异同角度将领域自适应分为同构域和异构域自适应;文献[7]首先根据源域的数量将领域自适应方法分为单源域领域自适应和多源域领域自适应,然后着重从源域和目标域的标签集的角度将单源领域自适应方法分为同构领域自适应和异构领域自适应两类;文献[8]主要探讨了NLP领域以模型为中心、以数据为中心和两者融合领域自适应方法,概述了从早期的传统非神经网络方法到预训练模型转移的方法;文献[9]专注于神经机器翻译(Neural Machine Translation, NMT)领域自适应问题,分别从数据选择、模型架构、参数自适应过程和推理过程等角度探讨了领域自适应的相关技术,该论文不仅为NMT领域的自适应问题提供了解决方案,而且可为其他领域提供一定的借鉴。

上述研究工作从不同的角度对领域自适应方法进行了概述,但是鲜有工作从融合知识的角度对领域自适应方法进行综述。融合的知识包括语言学知识和非语言学知识,其在领域自适应的过程中发挥着重要作用。本文在上述研究的基础上,提出了一种新的融合知识的领域自适应方法的分类体系,并对其中的代表工作进行了介绍并总结了各自的优缺点。最后,对现存挑战以及未来研究方向进行研究和分析。

3 领域自适应问题定义与理论基础

参考文献[5-6],本文首先给出了源域、目标域、源任务和

目标任务的定义,然后给出了领域自适应的定义,最后给出了融合知识的领域自适应的定义。

定义 1(源域 \mathcal{D}_S) 给定一组数据实例 $\{x_{s_1}, x_{s_2}, \dots, x_{s_n}\}$ 及其对应的标签 $\{y_{s_1}, y_{s_2}, \dots, y_{s_n}\}$, 将数据实例及其对应的标签进行组合作为源域数据, 记作 $\mathcal{D}_S = \{(x_{s_1}, y_{s_1}), (x_{s_2}, y_{s_2}), \dots, (x_{s_n}, y_{s_n})\}$ 。

定义 2(目标域 \mathcal{D}_T) 给定一组数据实例 $\{x_{t_1}, x_{t_2}, \dots, x_{t_n}\}$ 及其对应的标签 $\{y_{t_1}, y_{t_2}, \dots, y_{t_n}\}$, 将数据实例及其对应的标签进行组合作为目标域数据, 记作 $\mathcal{D}_T = \{(x_{t_1}, y_{t_1}), (x_{t_2}, y_{t_2}), \dots, (x_{t_n}, y_{t_n})\}$ 。

定义 3(源任务 T_S) 给定源领域的实例的特征向量 \mathbf{x}_S , 根据类别预测函数 $f(\mathbf{x}_S)$ 可以预测其类别标签的概率 $P(y_S | \mathbf{x}_S)$, 此过程称为源任务, 记作 $T_S = \{y_S, P(y_S | \mathbf{x}_S)\}$ 。

定义 4(目标任务 T_T) 给定目标领域的实例的特征向量 \mathbf{x}_T , 根据类别预测函数 $f(\mathbf{x}_T)$ 可以预测其类别标签的概率 $P(y_T | \mathbf{x}_T)$, 此过程称为目标任务, 记作 $T_T = \{y_T, P(y_T | \mathbf{x}_T)\}$ 。

定义 5(领域自适应) 领域自适应是一种源任务和目标任务一样, 即 $T_S = T_T$, 但是源域和目标域的数据分布不一样, 通过源域数据 $\{(x_{s_1}, y_{s_1}), (x_{s_2}, y_{s_2}), \dots, (x_{s_n}, y_{s_n})\}$ 、目标域未标识数据实例 $\{x_{t_1}, x_{t_2}, \dots, x_{t_n}\}$ 对目标域的预测函数

$f(\mathbf{x}_T)$ 进行学习的迁移学习方法。

定义 6(融合知识的领域自适应) 融合知识的领域自适应在未融合知识的领域自适应的基础上, 对预测函数 $f(\mathbf{x}_T)$ 进行学习时, 引入了外部知识, 增强了模型的自适应能力。

4 领域自适应方法分类

现实世界中的知识包含语言学知识和非语言学知识。因此, 本文将领域自适应方法分为基于语言学知识的领域自适应方法和基于非语言学知识的领域自适应方法两大类, 如图 1 所示。其中, 基于语言学知识的领域自适应方法包括基于词汇、句法和语义的领域自适应方法; 基于非语言学知识的领域自适应方法包括基于预训练模型、知识图谱和知识蒸馏的领域自适应方法。下面对各类方法的代表工作、优缺点等进行详细介绍。

4.1 基于语言学知识的领域自适应方法

语言是由句子组成的, 因此, 语言中的知识包含在句子中。而句子是由词汇按照一定的句法组成的, 其蕴含一定的语义信息。语言学中的一些知识在不同领域中是通用的, 因此, 将词汇、句法、语义等语言学知识按照一定的方式融入领域自适应方法中可能会产生积极效果。基于语言学知识的领域自适应典型方法总结如表 1 所列。

表 1 基于语言学知识的领域自适应方法

Table 1 Domain adaptive method based on linguistic knowledge

方法	模型	数据集	应用领域	评价指标	优点	缺点
基于词汇的领域自适应方法	DTM (Deep Transfer via Markov Logic) ^[10]	MIPS Comprehensive Yeast Genome Database and WebKB dataset	分子生物学、社交网络	AUC(area under the precision recall curve), CLL (conditional log-likelihood)		
	NMT based on word level weighting ^[11]	English-to-Chinese e-commerce item descriptions dataset	机器翻译	BLEU, TER	简单、实用	只考虑词汇信息而忽视了上下文信息, 不能解决一词多义的问题
	MBA (Manifold Based Alignment algorithm) ^[12]	MUSE dataset	双语词汇归纳	P@1		
基于句法的领域自适应方法	Character-level BiLSTM features with Transformer-based features ^[13]	monolingual CompLex LCP and cross-lingual the CWI Shared Dataset	复杂词识别	P (Pearson Correlation Coefficient), MAE (Mean Absolute Error)		
	TectoMT ^[14]	Batch2 answer corpus	机器翻译	BLEU	在长距离句子方面得到改进, 处理句子结构性自适应问题的能力也更强	性能依赖于句法分析的质量, 句法分析的错误容易产生错误传播, 且缺乏句子间逻辑语义关联知识的指导
基于语义的领域自适应方法	Bridge-based active domain adaptation method ^[15]	three conventional English datasets	属性抽取	F1		
	TectoMT ^[14] GAST ^[16]	Amazon-product-review dataset	情感分类	Accuracy	可以利用上下文信息, 提供更高层面的知识	严重依赖于中文的语义表示的解析设计
	DASK ^[17]	Amazon-product-review dataset	文本分类	Accuracy		

4.1.1 基于词汇的领域自适应方法

词汇是语言的基本组成部分, 通过缩小源域和目标域之间词汇的差异或者对不同领域之间的词汇进行对齐等方法, 可以实现领域自适应。

文献[10]提出了一种基于二阶马尔可夫逻辑的方法, 通过带有谓词变量的马尔可夫逻辑公式的形式发现源域中的结构规则, 并利用来自目标域的谓词实例化这些

公式。该方法可以在分子生物学、社交网络和万维网领域中传播所学知识。文献[11]提出了一种基于词级加权的神经机器翻译领域自适应算法。该方法针对领域特定语言模型和非领域特定语言模型, 根据概率的对数差计算出域外数据集中的词权值, 然后进行平滑和二值量化, 使得非域句子中的域内词具有更高的权重, 从而使神经机器翻译模型产生更多域内词。

文献[12]提出了一种新的基于流形的几何方法来学习源语言和目标语言之间的无监督对齐词嵌入,该方法将对齐学习问题描述为双随机矩阵流形上的域适应问题。文献[13]针对复杂词识别任务提出了一种领域自适应的方法,用于改进目标字符和上下文表示。该方法通过平滑探索数据集之间的差异,解决了多个域自适应的问题。

基于词汇的领域自适应方法较为简单、实用,但该方法只考虑词汇信息而忽视了上下文信息,导致领域自适应能力下降,且不能解决一词多义的问题。

术语资源是有效的。文献[15]提出了一种新的主动领域自适应方法来提取属性词,通过构建句法和语义桥梁来主动增强术语的可转移性。

基于句法的领域自适应方法在长距离句子方面得到改进,在处理句子结构性自适应问题上能力也更强。但是,该方法的性能依赖于句法分析的质量,句法分析的错误容易产生错误传播,导致领域自适应质量严重下降,且缺乏句子间逻辑语义关联知识的指导。

4.1.3 基于语义的领域自适应方法

语义是语言形式所表达的内容,其比词汇和句法更加抽象,使用不同词汇和句法表示的句子语义可能相同。

文献[16]提出了一种图自适应语义迁移(Graph Adaptive Semantic Transfer, GAST)模型,该模型能够从词序列和句法图中学习领域不变语义。该模型一方面设计了 POS-Transformer 模块,从单词序列和词性标记中提取顺序语义特征;另一方面设计了混合图注意(Hybrid Graph Attention, HGAT)模块,通过考虑可转移依赖关系来生成基于语法的语义特征,并且通过一种集成自适应策略(Integrated Adaptive Strategy, IDS)来指导两个模块的联合学习过程。为了利用词之间的隐性关系跨域,文献[17]提出了一种结构化知识领域自适应(Domain-Adaptive with Structured Knowledge, DASK)文本分类方法,利用词级语义关系来增强领域自适应。DASK 从目标领域数据构建一个知识图,以建模枢轴和非枢轴之间的关系,并通过捕获枢轴和目标领域特定的非枢轴之间的关系来学习领域不变特征。

基于语义的领域自适应方法可以利用上下文信息,相比前两种方法,其可以提供更高层面的知识。但是,该方法严重依赖于中间的语义表示的解析设计,这通常是性能改进的瓶颈。

4.2 基于非语言学知识的领域自适应方法

基于非语言学知识的领域自适应方法指不直接利用语言学方面的知识对领域自适应方法进行增强,例如使用预训练模型、知识图谱和知识蒸馏等技术获得的知识。基于非语言学知识的领域自适应典型方法总结如表 2 所列。

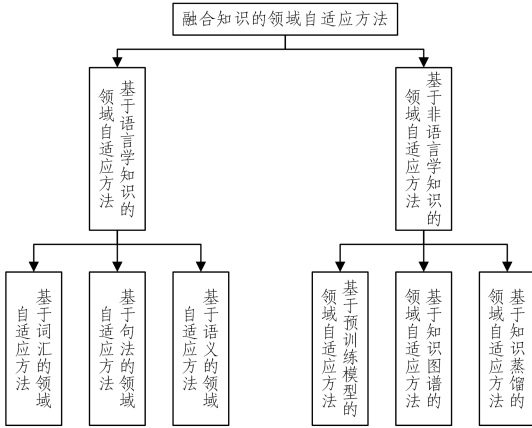


图 1 融合知识的领域自适应方法分类体系

Fig.1 Classification system of domain adaptive methods with knowledge integrating

4.1.2 基于句法的领域自适应方法

句法是句子的内部结构,以词作为基本单位,研究句子的各个组成部分和它们的排列顺序。在自然语言中,语言表达丰富多样,相比之下,句法结构比较有限。句法桥旨在识别跨域单词可传递的句法角色,在句法解析树中没有语义相关性的两个单词,也可能起到相似的作用。

文献[14]提出为双语术语添加句法结构以改进领域适应性,通过丰富源领域和目标领域的带有句法结构的多词术语,并集成到基于树的机器翻译系统 TectoMT 中,使其能够翻译出句法一致但表现形态不同的复杂术语。在 IT 领域的应用结果表明,该方法对西班牙语、巴斯克语和葡萄牙语的

表 2 基于非语言学知识的领域自适应方法

Table 2 Domain adaptive method based on non-linguistic knowledge

方法	模型名称	数据集	应用领域	评价指标	优点	缺点
	DAPT, TAPT ^[18]	S2ORC, REALNEWS, AMAZON reviews	生物医学、计算机科学、新闻、评论领域的文本分类	micro-F1		
	Pretrained language model with unsupervised domain clusters ^[19]	the multi-domain corpus	字幕、医学、法律、古兰经翻译和 IT 领域的机器翻译	SacreBLEU, Precision, Recall, F1		
基于预训练模型的领域自适应方法	K-ADAPTER ^[20]	MNIST, TFD, CIFAR-10	计算机视觉	mean log-likelihood of samples	学习了大量知识,面对下游任务能取得不错的效果	参数量较大,训练成本较高,部署困难
	PADA ^[21]	PHEME, SNLI, ABSA corpora	文本分类、序列标记,	Binary-F1, macro-F1		
	Pre-trained or Annotated models ^[22]	Wet Lab Protocols, PUBMED&M, CHEM-SYN	自然语言处理	F1, # sent et. al		
	DomBERT ^[23]	SemEval dataset, Amazon review datasets, Yelp dataset	细粒度情感分析	Accuracy, Macro-F1		

(续表)

方法	模型名称	数据集	应用领域	评价指标	优点	缺点
基于知识图谱的领域自适应方法	Domain adaptation method based on Knowledge graph embeddings ^[24]	Musiconis database, Vi-huela database	乐器识别、文化遗产保护	F1	高效利用开放的结构化领域知识,从而缓解不同领域之间知识概念上的隔阂	特定领域知识图谱难以获取、质量较差;已开放的通用知识图谱的知识覆盖度难以保证;模型参数和训练成本随图谱规模正向增大
	KEFDA ^[25]	FewRel 2.0 DA challenge	关系分类、知识抽取	Accuracy		
	KinGDOM ^[26]	Amazon-reviews benchmark database	情感分析	Accuracy		
基于知识蒸馏的领域自适应方法	Knowledge Distillation for NMT ^[27]	WMT 2014, IWSLT 2015	机器翻译	BLEU, PPL	简单、推断速度快、模型参数量小、部署容易,已经在工业界广泛应用	依赖教师模型的性能以及教师和学生模型之间的关系,且蒸馏的位置需要手动选择或者基于任务选择
	NMT model ^[28]	OpenSubtitles2018, WMT2017, WIPO COP-PA-V2 dataset	机器翻译	BLEU		
	DoKTra ^[29]	ChemProt, GAD, DDI, i2b2, HoC dataset	关系抽取、多标签分类	BLEU, F1		
	mBERT ^[30]	Conversational commerce dataset	商务会话	Precision, Recall, F1		
	Pretrained Models ^[31]	Language PubMed, EBM PICO	生物医学、计算机科学	F1, micro-F1, macro-F1		
	Meta-KD ^[32]	MNLI, Amazon Review	自然语言推理、情感分析	Accuracy		
	MDKD ^[33]	OPUS	机器翻译	BLEU		
DIFD ^[34]	SemEval 2014 task 4	情感分类	Accuracy, macro-F1			

4.2.1 基于预训练模型的领域自适应方法

近年来,随着预训练模型的发展,基于预训练模型的知识迁移成为一种有效的融合知识的领域自适应方法。预训练模型可以利用海量的无标注数据学习到通用的知识,更好地初始化其他模型,加快模型的训练速度,提升下游任务的效果。同时,因为预训练模型都是在大规模的数据上训练的,可以学到多个领域的知识,有效地实现领域间相关知识的迁移。然而,预训练模型的训练主要集中在广泛的开放域里,因此其对开放域有较好的效果。换言之,其对特定领域的适配情况可能并不好,尤其是一些专业领域,如医学、生物等领域。

为了让预训练模型适配具体的任务,可以直接用特定领域或特定任务的训练数据来对预训练模型进行微调,即领域自适应预训练(Domain-Adaptive Pretraining, DAPT)和任务自适应预训练(Task-Adaptive Pretraining, TAPT)^[18]。其中,DAPT是在所属领域的数据上继续预训练,TAPT是在特定任务的数据上继续预训练。这种简单的方式在许多任务上已经超越了之前针对特定任务精心设计的模型。文献[18]对DAPT和TAPT分别进行实验,验证了DAPT对模型有明显的提升效果。进一步的研究表明,距离开放域越远,DAPT对模型的提升幅度越明显。相比DAPT,TAPT的效果从整体来看要略差一些,但是基本上也能达到与DAPT接近的效果,将两者结合之后能进一步提升效果。值得注意的是,TAPT虽然对自己的任务提升较大,但是训练完后,预训练模型对别的任务的效果会有比较明显的下降,模型的遗忘比较明显。对于数据的选择,文献[19]在预先训练的语言模型中发现了领域特定的集群,并展示了如何利用这些集群在领域敏感训练中数据进行选择。文献[20]提出了一个能保留预训练模型原始参数并支持多种知识注入的模型K-ADAPTER,针对不同的预训练任务,定义了对应的适配器(Adapter),并在模型中注入了两种知识,分别是来自Wikipedia和Wikidata上自动对齐的文本三元组中获得的事实知识和通过依赖解析获得的语言知识。K-ADAPTER在3个知识驱动数据集上

均取得了较大增益。

此外,文献[21]提出了一种基于提示(Prompt)的领域自适应算法,该算法首先训练T5生成唯一的域相关特征DRFs(一组用于描述域信息的关键字),然后,这些DRFs可以与输入部分连接起来,形成一个模板,供下游任务进一步使用。文献[22]探讨了一些与NLP模型适应新领域的成本相关的问题,该问题可以被定义为消费者选择问题。实验结果表明,当少量的NLP模型需要适应一个新的领域时,预训练本身并不是一个经济的解决方案。文献[23]使用域内和域外数据预训练特定领域的模型,并使其适应低资源域。

基于预训练模型的领域自适应方法学习了大量知识,因此在下游任务中能取得不错的效果。但是该方法参数量较大,训练成本较高,部署困难。

4.2.2 基于知识图谱的领域自适应方法

近年来,知识图谱作为结构化语义知识库被广泛应用于自然语言处理、计算机视觉等领域,在知识驱动的人工智能研究中发挥着重要作用。其中,领域知识图谱面向各个特定领域用于复杂分析、辅助决策、知识指导与推理,具有多层次、多粒度、结构复杂、领域性强等特点。因此,利用额外的结构化领域知识或常识知识可以辅助领域自适应的任务完成。

例如,在计算机视觉领域中,文献[24]提出了一种基于知识图嵌入的领域自适应方法,将知识图谱的实体、关系投射到低维、连续的向量空间作为指导领域适应过程的关键信息。在自然语言处理领域中,知识图谱往往作为不同领域之间概念级别的“桥梁”来提升领域自适应任务的效果。例如,文献[25]提出一种知识增强的少样本关系分类模型KEFDA,将特定领域的知识图谱结合到关系分类模型中以提高领域自适应能力。该方法在概念级知识图谱的增强下使模型更好地理解文本的语义,并从少数实例中归纳出关系类型的全局语义。文献[26]引入新框架KinGDOM来实现跨领域情感分析。该模型首先基于ConceptNet知识图谱构建包含源领域和目标领域信息的子图,之后利用端到端图编码器对知识图谱进行

预训练来获取特定领域知识和通用常识背景概念,进而丰富文档语义,实验表明该方法具有明显的效果。

基于知识图谱的领域自适应方法可以高效利用开放的结构化领域知识,从而消除不同领域之间知识概念上的隔阂。但是,该方法存在一些局限:1)当前特定领域知识图谱存在难以获取、质量较差的问题;2)目前已开放的通用知识图谱与源领域和目标领域的知识覆盖度难以保证;3)模型参数与训练成本往往随着图谱规模的增大而增加。

4.2.3 基于知识蒸馏的领域自适应方法

知识蒸馏(Knowledge Distillation)是领域自适应的一种常用的方法,其通过构建一个轻量化的小模型,利用性能更好的大模型的监督信息来训练这个小模型,以达到更好的性能和精度。

知识蒸馏的概念由 Hinton 等^[35]于 2015 年首次提出并被应用在分类任务上,其中,大模型被称为教师模型,小模型被称为学生模型。之后,文献[27]又将知识蒸馏拓展到序列级文本任务,进而学到序列级的知识。教师模型输出的监督信息被称为知识,而学生模型学习迁移来自教师模型的监督信息的过程被称为蒸馏。文献[28]探究了知识蒸馏技术在机器翻译领域自适应中的应用,并设置了不同的对比实验,最终实验表明知识蒸馏在领域自适应领域有显著的提升效果,并且两步蒸馏可以得到更好的效果:第一步蒸馏得到更好的教师模型,第二步蒸馏可以将教师模型的知识迁移到目标领域的学生模型。文献[29]基于知识蒸馏的思路提出了领域知识转移框架(DoKTra),该框架不需要额外的领域预训练。DoKTra 框架的主要目标是从微调的教师模型中提取领域知识,为特定领域中的每个下游任务生成特定的学生模型。该框架包括两个主要阶段:校准的教师培训和激活边界蒸馏。在校准的教师培训中,通过校准方法生成可靠的教师模型;激活边界蒸馏则是将教师模型的特定领域知识迁移到学生模型中。文献[30]将知识蒸馏用于商业会话领域中,通过使用特定领域数据的知识蒸馏,将 mBERT 作为教师模型,利用更小的学生模型去学习教师模型的知识,解决了商业会话环境中编码器的可伸缩性问题,在缩减模型大小的同时提升了模型性能。文献[31]为在特定领域开发小型、快速和有效的预训练语言模型,提出了 AdaLM 框架,在自适应阶段针对特定领域进行了词汇量扩展,并利用语料库出现的概率来自动选择扩展词汇量的大小。然后,该论文系统地探索了不同的策略来压缩特定领域的大型预训练模型,研究了领域自适应和知识蒸馏的执行顺序对目标任务的影响。文献[32]提出了元知识蒸馏(Meta-Knowledge Distillation, Meta-KD)框架来构建元教师模型,该模型可以捕获跨领域的可转移知识,并将这些知识传递给学生。Meta-KD 由元教师学习和元蒸馏两部分组成。元教师模型与多领域数据集进行联合训练,以获得实例级和特征级的元知识。在每个领域,学生模型在元教师的指导下,学习在一个特定领域的数据集上解决任务。文献[33]提出了一个蒸馏的多领域神经机器翻译框架,该方法对训练数据进行转换,使其包含序列级知识蒸馏的输出,教师是多个领域的高性能模型,这也使得该领域自适应方法可以从特定模型中提取到更多的领域知识,进而得到更好的性能。

为解决 aspect 级别的跨领域情感分析问题,文献[34]提出了领域不变特征蒸馏来同时学习领域不变的特征和特定领域的特征,通过正交领域相关任务,aspect 级别情感分类任务可以学习到更好的领域不变特征,提升其在特定领域的性能。

基于知识蒸馏的领域自适应方法简单,推断速度快,模型参数量小,部署容易,已经在工业界得到广泛应用。但是,该方法比较依赖教师模型的性能以及教师和学生模型之间的关系,且蒸馏的位置需要手动选择或者基于任务进行选择。

5 现存挑战与未来研究方向

近年来,融合知识的领域自适应技术已经取得了一定程度的进展,但是,由于领域自适应具有复杂性,该领域仍然存在一些问题有待解决。下面总结现存的一些挑战性问题并给出未来可能的研究方向。

(1)多源领域自适应学习及负迁移问题

本文综述的多为单领域自适应学习方面的相关研究,但是在实际情况中,可能存在多个源域或者多个目标域的情况。面对多源领域时,不仅要考虑源域和目标域之间的分布差异,还需考虑多个源域或者多个目标域之间的分布差异。如果源域和目标域的相关性较小,源域知识的引入可能会导致目标任务性能下降,即出现负迁移现象。因此如何权衡多个领域之间的分布差异并解决负迁移问题面临着巨大挑战,这也是多领域自适应学习值得探索的一个研究方向。

(2)缩减模型参数

随着深度学习的发展以及数据规模的增大,模型的参数量也会不断增加,过于庞大的模型不仅会增加训练成本还会增大部署难度。如何在加强模型学习更多知识的同时,减少模型参数是实际场景中亟待解决的问题。知识蒸馏是模型压缩的一种方法,但是只有为数不多的工作将知识蒸馏与其他模型压缩方法相结合,因此,混合压缩方法以及混合压缩时不同压缩方法的应用顺序是未来值得研究的方向。

(3)高质量、大规模知识图谱的构建

面对特定领域的任务时,目前已开放的大规模通用知识图谱的知识覆盖度难以保证,而且特定领域的知识图谱获取困难。因此,如何构建高质量、大规模的知识图谱以提高领域自适应方法的效果面临着挑战。

(4)源域数据隐私保护

使用源域数据训练模型时,源域数据可能来自企业或者个人,这些数据中很有可能包含企业或者个人的隐私。此时,如何在进行领域自适应的同时保护数据隐私是一个实际场景中必须面对的重要问题。构建基于模型参数的领域自适应方法而不是基于数据特征的领域自适应方法和构建带有加密技术的领域自适应方法是值得借鉴的研究方向。

结束语 提高模型的泛化能力是基于数据驱动的模式面临的挑战性问题,融合知识的领域自适应是提高模型泛化能力的有效方法。本文首先概述了融合知识的领域自适应方法的发展背景和相关综述的研究现状;然后介绍了领域自适应的问题定义和理论基础;之后,全面地分析和总结了融合知识的领域自适应方法;最后,深入地剖析了融合知识的领域自适应方法的挑战性问题,并对该领域未来的研究方向进行了

展望。总之,融合知识的领域自适应方法的创新与进步具有十分重要的理论意义和应用前景,一方面将推动基于数据驱动的模型的发展并扩大其应用场景;另一方面将为通用人工智能技术的发展提供可行的解决方案。

参 考 文 献

- [1] PAN S J, QIANG Y. A Survey on Transfer Learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.
- [2] JING J. A Literature Survey on Domain Adaptation of Statistical Classifiers[J]. British Journal of Psychiatry, 2008, 131: 83-89.
- [3] CHEN S T. Research on domain adaptive algorithm [D]. Guangzhou: South China University of Technology, 2020.
- [4] XU M. Research on domain adaptive learning algorithm and its application [D]. Wuxi: Jiangnan University, 2014.
- [5] LIU J W, SUN Z K, LUO X L. Progress in domain adaptive learning [J]. Acta Automatica Sinica, 2014, (8): 1576-1600.
- [6] LI J J, MENG L C, ZHANG K, et al. Overview of domain adaptation research [J]. Computer Engineering, 2021, 47(6): 13.
- [7] FAN M, CAI Z, ZHANG T, et al. A survey of deep domain adaptation based on label set classification[J]. Multimedia Tools and Applications, 2022, 81: 39545-39576.
- [8] RAMPONI A, PLANK B. Neural Unsupervised Domain Adaptation in NLP--A Survey[J]. arXiv:2006.00632, 2020.
- [9] SAUNDERS D. Domain adaptation and multi-domain adaptation for neural machine translation: A survey[J]. arXiv:2104.06951, 2021.
- [10] DAVIS J, DOMINGOS P. Deep Transfer via Second-Order Markov Logic[C]// International Conference on Machine Learning. ACM, 2009.
- [11] SHEN Y, LEONARD D, PAVEL P S, et al. Word-based Domain Adaptation for Neural Machine Translation[C]// Proceedings of the 15th International Conference on Spoken Language Translation. 2018: 31-38.
- [12] PRATIK J, MAYANK M, BAMDEV M. Geometry-aware domain adaptation for unsupervised alignment of word embeddings [C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020: 3052-3058.
- [13] GEORGE-EDUARD Z, RĂZVAN-ALEXANDRU S. Domain Adaptation in Multilingual and Multi-Domain Monolingual Settings for Complex Word Identification[C]// Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2022: 70-80.
- [14] MIKEL A, GORKA L, CHAKAVEH S, et al. Adding syntactic structure to bilingual terminology for improved domain adaptation[C]// Proceedings of the 2nd Deep Machine Translation Workshop. 2016: 39-46.
- [15] ZHUANG C, TIEYUN Q. Bridge-Based Active Domain Adaptation for Aspect Term Extraction[C]// Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. 2021: 317-327.
- [16] KAI Z, QI L, ZHENYA H, et al. Graph Adaptive Semantic Transfer for Cross-domain Sentiment Classification[J]. arXiv: 2205.08772, 2022.
- [17] LI T, CHEN X, DONG Z, et al. Domain-Adaptive Text Classification with Structured Knowledge from Unlabeled Data[J]. arXiv:2206.09591, 2022.
- [18] SUCHIN G, ANA M, SWABHA S, et al. Don't Stop Pretraining: Adapt Language Models to Domains and Tasks[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020: 8342-8360.
- [19] ROEE A, YOAV G. Unsupervised domain clusters in pretrained language models[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020: 7747-7763.
- [20] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[J]. arXiv:1406.2661, 2014.
- [21] EYAL B, NADAV O, ROI R. PADA: Example-based Prompt Learning for on-the-fly Adaptation to Unseen Domains [J]. Transactions of the Association for Computational Linguistics, 2022, 10: 414-433.
- [22] FAN B, ALAN R, WEI X. Pre-train or Annotate? Domain Adaptation with a Constrained Budget [C]// Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021: 5002-5015.
- [23] HU X, LIU B, SHU L, et al. DomBERT: Domain-oriented language model for aspect-based sentiment analysis[C]// Findings of the Association for Computational Linguistics: EMNLP 2020. 2020: 1725-1731.
- [24] EYHARABIDE V, BEKKOUCH I, CONSTANTIN N D. KnowledgeGraph Embedding-Based Domain Adaptation for Musical Instrument Recognition[J]. Computers, 2021, 10(8): 94.
- [25] ZHANG J W, ZHU J Q, YANG Y, et al. Knowledge-Enhanced Domain Adaptation in Few-Shot Relation Classification [C]// Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD '21). 2021.
- [26] DEEPANWAY G, DEVAMANYU H, ABHINABA R, et al. KinGDOM: Knowledge-Guided Domain Adaptation for Sentiment Analysis[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020: 3198-3210.
- [27] KIM Y, RUSH A M. Sequence-Level Knowledge Distillation [C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016: 1317-1327.
- [28] GORDON M, KEVIN D. Distill, Adapt, Distill; Training Small, In-Domain Models for Neural Machine Translation [C]// Proceedings of the Fourth Workshop on Neural Generation and Translation. 2020.
- [29] CHOI D H, HONGSEOK C, HYUNJU L. Domain Knowledge Transferring for Pre-trained Language Model via Calibrated Activation Boundary Distillation[C]// Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2022.
- [30] KRISTEN H, JIAN W, AKSHAY H, et al. Domain-specific knowledge distillation yields smaller and better models for conversational commerce[C]// Proceedings of The Fifth Workshop

on e-Commerce and NLP(ECNLP 5). 2022.

- [31] YAO Y Z, HUANG S H, WANG W H, et al. Adapt-and-Distill: Developing Small, Fast and Effective Pretrained Language Models for Domains[C]//Findings of the Association for Computational Linguistics:ACL-IJCNLP 2021. 2021.
- [32] PAN H J, WANG C, QIU M, et al. Meta-KD: A Meta Knowledge Distillation Framework for Language Model Compression across Domains[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2021.
- [33] CURREY A, PRASHANT M, GEORGIANA D. Distilling multiple domains for neural machine translation[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020.
- [34] HU M, WU Y, ZHAO S, et al. Domain-Invariant Feature Distillation for Cross-Domain Sentiment Classification[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019: 5559-5568.

- [35] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[C]//NIPS 2014 Deep Learning Workshop. 2014.



CUI Fuwei, born in 1990, Ph. D, is a student member of China Computer Federation. His main research interests include natural language processing, text generation, dialog system, etc.



XU Jin'an, born in 1970, Ph.D, professor, Ph.D supervisor, is a senior member of China Computer Federation. His main research interests include natural language processing, machine translation, knowledge graph and its application, text emotion analysis, automatic summarization, question answering, dialogue system, human-machine interaction, etc.

(责任编辑:杨雪敏)