

◎热点与综述◎

提示学习研究综述

崔金满^{1,2}, 李冬梅^{1,2}, 田 萱^{1,2}, 孟湘皓^{1,2}, 杨 宇^{1,2}, 崔晓晖^{1,2}

1. 北京林业大学 信息学院, 北京 100083

2. 国家林业草原林业智能信息处理工程技术研究中心, 北京 100083

摘 要: 经过微调的预训练语言模型在各领域任务中均取得了显著的性能。但是, 预训练和微调之间在训练数据和目标函数方面存在着巨大差距, 阻碍了预训练语言模型对下游任务的有效适应。提示学习的提出缩小了预训练和微调之间的差距, 并可以很好地应用到小样本甚至零样本场景中。提示学习的核心思想是将提示模板插入到原始输入中, 将下游任务数据转化为自然语言的形式输入到预训练模型中, 输出预测结果, 然后通过语言表达器将输出映射到相应的标签。系统地梳理了当前提示学习的相关工作, 根据提示学习的实现步骤, 从提示模板和语言表达器构建两个阶段介绍该类方法的研究进展。将基于提示模板的方法细分为人工构建、自动构建、引入外部知识构建提示和思维提示方法4种; 将基于语言表达器的方法细分为人工构建的表达器、基于搜索的表达器、软表达器和引入外部知识构建表达器的方法4种。总结了提示学习在自然语言处理、计算机视觉和多模态领域的主要应用, 并对提示学习相关实验进行了分析。最后, 概述了提示学习的现状和挑战, 展望了提示学习的未来发展方向。

关键词: 提示学习; 预训练模型; 预训练和微调; 小样本学习; 零样本学习

文献标志码: A **中图分类号:** TP18 doi:10.3778/j.issn.1002-8331.2407-0436

Survey on Prompt Learning

CUI Jinman^{1,2}, LI Dongmei^{1,2}, TIAN Xuan^{1,2}, MENG Xianghao^{1,2}, YANG Yu^{1,2}, CUI Xiaohui^{1,2}

1. School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China

2. Engineering Research Center for Forestry-Oriented Intelligent Information Processing of National Forestry and Grassland Administration, Beijing 100083, China

Abstract: Fine-tuned pre-trained language models have achieved remarkable performance in various domain tasks. However, there is a significant gap in training data and objective function between pre-training and fine-tuning, which limits the effective adaptation of pre-trained language models to downstream tasks. Prompt learning has been proposed to bridge the gap in pre-training and fine-tuning, and can be well applied to few-shot or even zero-shot scenarios. The core idea of prompt learning is to wrap the original input with prompt template to convert the downstream task data into the form of natural language, and input it into the pre-trained models to output the prediction result, and then map the output to corresponding labels through the language verbalizer. This paper systematically combs the current approaches of prompt learning, and introduces its research progress from two stages of prompt template and language verbalizer construction according to the implementation steps of prompt learning. The prompt template based methods are subdivided into manually constructed, automatic constructed, introducing external knowledge to constructing prompt and thought prompting methods. The language verbalizer based methods are subdivided into manual verbalizer, search-based verbalizer, soft verbalizer and verbalizer with external knowledge introduced. In the following, the paper summarizes the main applications of prompt learning in the fields of natural language processing, computer vision and multimodal, and analyzes the related

基金项目: 国家重点研发计划(2022YFF1302700); 国家自然科学基金(82374621); 林草智能信息技术创新研究项目(QNTD202308)。

作者简介: 崔金满(1997—), 女, 博士研究生, CCF 学生会员, 研究方向为大模型、知识图谱; 李冬梅(1972—), 女, 博士, 教授, CCF 会员, 研究方向为自然语言处理、知识图谱; 田萱(1976—), 女, 博士, 副教授, CCF 会员, 研究方向为文本挖掘与多模态智能; 孟湘皓(2000—), 男, 硕士研究生, CCF 学生会员, 研究方向为大模型、知识图谱; 杨宇(2000—), 男, 硕士研究生, CCF 学生会员, 研究方向为自然语言处理、知识图谱; 崔晓晖(1984—), 通信作者, 男, 博士, 副教授, CCF 会员, 研究方向为林业大数据、智能信息处理, E-mail: cuixiaohui@bjfu.edu.cn。

收稿日期: 2024-07-23 **修回日期:** 2024-10-17 **文章编号:** 1002-8331(2024)23-0001-27

experiments of prompt learning. Finally, this paper summarizes the current situation and challenges in prompt learning, and prospects the future technological development of prompt learning.

Key words: prompt learning; pre-trained models; pre-training and fine-tuning; few-shot learning; zero-shot learning

自监督的预训练语言模型(pre-trained language models, PLMs)从大规模语料库中获取大量的知识,然后利用任务特定的标注数据对模型参数进行微调,可以使其适应各种下游任务^[1],如GPT^[2]、BERT^[3]、RoBERTa^[4]和T5^[5]等。预训练任务通常是在大规模无标注数据集上通过自监督学习方式来学习与任务无关的通用知识,而微调任务则是在小规模标注数据集上通过监督学习方式学习任务特定的知识^[6]。因此,尽管“预训练-微调”范式取得了成功,但由于预训练任务和微调任务之间在数据和目标函数方面存在着巨大差距,限制了PLMs中知识的有效转移和充分利用。

为了解决上述问题,研究人员借鉴迁移学习的思想,提出了提示学习(prompt learning)^[7-10]范式,通过设计一些有效的关键词或提示,引导语言模型(language model, LM)自动生成任务相关的答案,实现各领域下游任务^[11]。提示学习的核心思想在于利用LM在大量无标记数据上学习到的通用知识,通过特定任务的提示,帮助模型有效地表示和理解任务数据,从而将LM学习到的通用知识迁移到具体的下游任务中。这一过程减少模型对新任务数据的依赖,提高了LM在新任务上的泛化能力和适应性^[12]。

与传统“预训练-微调”范式相比,提示学习展现出多重优势。在效率方面,“预训练-微调”范式需要对每个新任务进行单独的模型训练,不仅耗时耗力,还可能导致模型过拟合,相比之下,提示学习范式更加灵活,不需要为每个新任务单独训练模型,可以根据具体的任务需求调整提示信息,使PLMs适应新任务,提高了训练和推理的效率,并在多领域应用中表现出巨大的潜力;在任务适应性方面,提示的设计可以被看作是一种主动学习的策略,通过提供任务相关的知识和约束,引导模型在学习过程中关注特定的信息,帮助其更好地理解任务要求,提高任务适应性;在泛化能力方面,“预训练-微调”范式需要大量的标注数据进行训练,而提示学习范式则能够通过提示引导,为模型提供额外的上下文信息,帮助模型在只有少量甚至零标注数据的情况下做出更准确的预测,提高数据效率,降低过拟合风险,提升泛化能力^[13]。

提示学习的概念最早兴起于自然语言处理(natural language processing, NLP)领域,弥补了预训练任务和微调任务之间的差距,在文本分类、信息抽取和逻辑推理等多种任务中取得很好的结果。受此影响,提示学习逐渐应用于计算机视觉(computer vision, CV)^[14]和多模

态领域^[13,15],在图像分类、视觉问答等任务中取得很大的进展。凭借其在效率、任务适应性和泛化能力等方面的优势,提示学习已广泛应用多种场景。如指导模型在数据匮乏的情况下学习有效知识,实现小样本学习和零样本学习;通过设计不同的提示引导模型同时学习多个任务,促进任务之间的知识共享和协同优化,实现多任务学习;帮助模型理解来自不同模态的数据,促进跨模态的信息融合和理解,实现跨模态学习;利用提示信息实现更加精准、个性化的推荐等。最近,提示学习还用于指导ChatGPT^[16]和GPT-4^[17]等大语言模型(large language models, LLMs)实现对话任务^[18]。利用提示学习可以帮助ChatGPT等LLMs更好地理解用户意图和上下文信息,提高其生成答案的准确性和可控性,同时也可以降低模型生成无意义或不相关答案的概率。

传统的基于提示学习的方法可以将下游任务重新表述为完形填空任务^[19],利用PLMs的文本生成能力来实现下游任务。其中涉及提示模板(prompt template)设计和语言表达器(language verbalizer)设计两个阶段,旨在搜索最佳提示模板和答案映射空间。如图1所示,传统的基于提示学习的方法将情感分析任务转换为完形填空任务。在提示模板设计阶段,需要为模型设计合适的包含[MASK]的提示模板,并将其与原始输入相融合(如“<S₁> It was [MASK].”“<S₁>”表示原始输入,“It was [MASK].”为提示模板);然后利用PLMs进行掩码语言建模(masked language modeling, MLM);最后,通过语言表达器将预测的[MASK]位置的标签词映射到相应的标签。但是这类方法在复杂逻辑推理任务上表现不佳,基于对人脑结构、思维方式和推理能力等方面的思考,研究人员提出思维链提示等一系列方法^[20-22],将复杂的推理问题分解为多个简单的步骤,通过构建思维推理提示来模拟人类思考的过程,从而增强模型的逻辑推理能力^[23]。

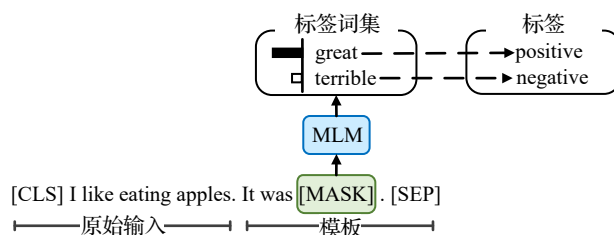


图1 提示学习示例

Fig.1 Example of prompt learning

总而言之,“预训练-微调”范式需要对PLMs的参数进行微调,使其适配NLP、CV和多模态领域的各种任务

(如图2(a)所示)。而与“预训练-微调”范式不同的是,以GPT-3为代表的提示学习范式,将任务特定的提示与下游任务的原始输入相融合,然后输入到PLMs中预测期望输出,模型可以更好地理解下游任务,灵活性更高^[24](如图2(b)所示)。提示学习作为一种新的范式,保留了PLMs同时处理多种任务的能力,即利用恰当的提示,使无监督方式训练的LM用于实现各领域下游任务^[7],同时,通过提供合适的提示,可以引导模型在小样本甚至零样本场景下进行学习和推理。

近年来,提示学习取得了很好的发展,利用提示学习范式,可以实现情感分析^[25-26]、关系抽取^[27-29]和自然语言推理^[30-31]等NLP任务,图像分类^[32]、视频动作识别^[33]、视觉定位^[34]等CV和多模态任务。目前已有部分提示学习相关的综述,如Liu等人^[24]详细介绍了NLP领域“预训练-微调”范式到提示学习范式的发展史,并系统地总结了相对前沿的工作;Chu等人^[23]引入了广义思维链的概念(X-of-thought),并从思维链、思维链结构变体和思维链增强三个角度对思维提示方法进行了分析和讨论;鲍琛龙等人^[35]对与知识相结合的提示学习方法进行了系统综述;Qiao等人^[36]从提示学习的视角对各种前沿的推

理方法进行了梳理。Liu等人^[37]系统总结了现有的包括提示学习在内的基于PLMs的实现推荐任务的方法;廖宁等人^[38]从应用任务角度对单模态视觉以及多模态视觉语言领域的提示学习方法进行了介绍。但是,上述方法各自聚焦于不同的方面和视角,缺少对提示学习系统而全面的综述,特别是从提示学习的实现步骤、应用领域等综合视角来阐述提示学习方法的研究进展。

为了使读者可以更深刻地体会提示学习的核心机制,从构建有效的提示模板到设计和训练强大的语言表达式器,逐步了解提示学习的实现过程,本文系统梳理了提示学习范式自2020年至今的技术演进,总结了代表性的算法和技术。如图3所示,本文首先根据提示学习方法的实现步骤,从基于提示模板和基于语言表达式器两阶段的方法介绍提示学习的研究进展;总结提示学习在NLP、CV和多模态领域的主要应用;然后对提示学习的研究现状和主要挑战进行了总结;最后展望了未来发展方向与趋势。

1 基于提示模板的提示学习

提示模板在提示学习方法中扮演着重要角色,通过

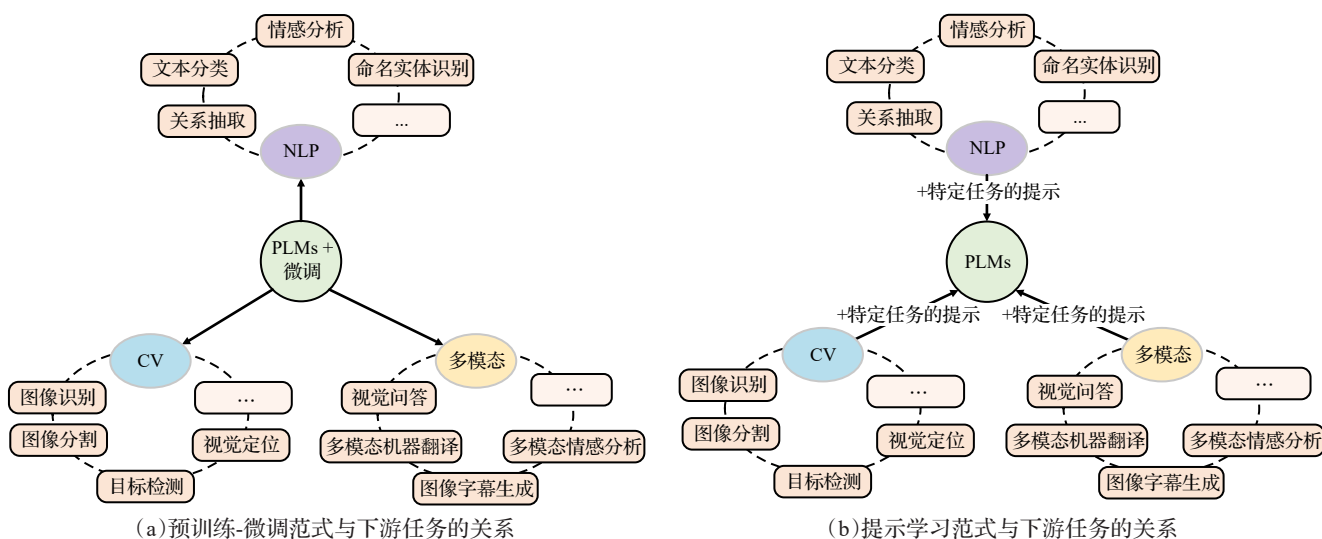


图2 不同范式与下游任务之间的关系

Fig.2 Relationship between different paradigms and downstream tasks

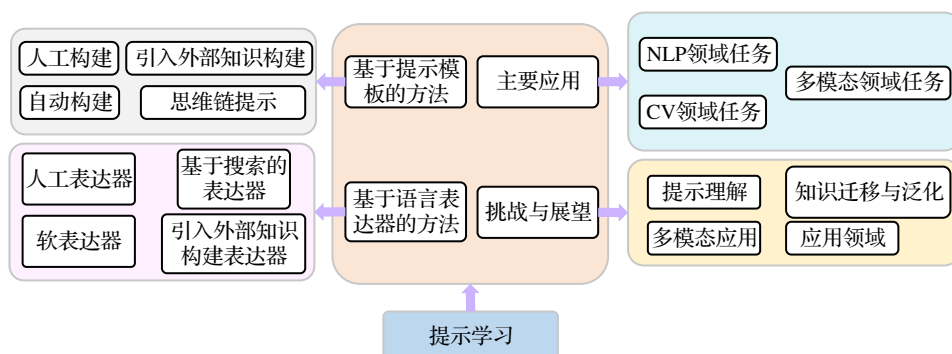


图3 提示学习整体框架

Fig.3 Overall framework of prompt learning

恰当的提示信息可以将原始任务转化为PLMs能够理解和解决的形式,而不需要对PLMs进行微调,其设计的好坏将直接影响模型的性能^[39]。例如,在情感分析任务中,可以将提示模板设计为“The sentiment of the sentence is [MASK].”将分类任务转化为完形填空任务,引导模型预测[MASK]位置的单词。根据提示模板的构建方法的不同,本文将其分为人工构建、自动构建、引入外部知识构建提示和思维提示方法4种。接下来,就各种方法中典型的算法模型及研究进展进行介绍。

1.1 人工构建提示模板的方法

构建提示模板最自然的方式是利用领域知识人工构建任务相关的提示。例如,为主题分类任务设计提示模板“This is [MASK] news.”“A [MASK] news”或“The topic is [MASK]”等。人工构建提示模板的方法在NLP、CV和多模态等领域均有应用。

(1) NLP领域的人工构建提示模板方法

Petroni等人^[19]表示在大型语料库上训练的PLMs在学习语言知识的同时,还可能存储训练数据中存在的关系知识,这些知识可以利用完形填空的方式从PLMs中提取出来。基于此,Petroni等人^[19]开创性地提出了语言模型分析(language model analysis, LAMA)方法,通过人工构建提示模板,将事实知识转换为完形填空形式的文本,探索PLMs中的知识。2020年,Brown等人^[7]提出GPT-3,使用人工构建的提示来适应不同的生成任务,它的成功表明适当的人工提示能够促进LLMs理解自然语言。自GPT-3出现以来,借助提示以充分利用模型中知识的方法受到广泛关注。但是,GPT-3由175B的参数组成,规模巨大,在实际应用场景中受到一定的限制。为此,Schick和Schütze^[40-41]引入PET(pattern-exploiting training)方法,利用人工构建的提示模板将输入实例重新表述为完形填空式短语,以帮助LM理解特定任务。

尽管上述经典模型取得了显著进展,但是提示模板的设计并没有针对具体的NLP任务。为此,一些研究人员针对命名实体识别(named entity recognition, NER)、细粒度实体分类任务和事件检测等具体任务进行了相关研究。针对NER任务,为了充分利用在源域中学习到的知识,在目标域中进行命名实体识别,Cui等人^[42]提出一种基于人工构建模板的NER方法,将NER任务视为一个序列到序列(sequence-to-sequence)问题,在推理时,根据由候选实体跨度填充的模板对应的分数,进行实体识别。针对细粒度实体分类任务,由于该任务定义明确,提示具有较明确的目的,Ding等人^[43]选择使用人工构建的陈述性提示模板,在设置提示模板时,首先复制输入文本中提及的标记实体,然后在[MASK]标记后添加一些连接动词和冠词,该方法在一定程度上避免了语法错误。针对事件检测任务,Zhang等人^[44]提出一种

基于提示预测的零样本事件检测的端到端的框架,在人工定义的提示模板中填充了事件指称,并由预训练的BERT进行编码,对应[MASK]标记的上下文向量用来预测触发词。该方法可以在不依赖于启发式规则的情况下检测新的事件类型,是一种有效的零样本事件检测方法。

上述系列方法是将下游任务转化为完形填空任务,利用MLM预测[MASK]位置的单词,并将其映射到相应的类别。这类方法都是字符级别的,要求将标签映射到固定长度的字符跨度。对于句子级模型,需要关注句子之间的关系,模型的输出通常映射到一个关系空间^[45]。基于此,Sun等人^[46]提出了句子级提示学习方法NSP-BERT,利用人工构建的提示模板将下游任务转化为自然语言推理任务,让模型判断候选答案是否正确。考虑到在关系抽取任务中,关系标签需要更详细地描述,Sainz等人^[47]将关系抽取重新表述为文本蕴含问题,首先为给定的一组关系人工构建提示模板,并用实体类型约束来补充模糊的关系表述,然后将给定包含两个实体的输入文本作为前提,根据人工模板自动生成的关系描述作为假设,利用自然语言推理模型来推断前提是否包含假设。

(2) CV和多模态领域的人工构建提示模板方法

之前先进的视觉模型通常采用监督学习方式进行训练,只能预测一组固定的目标类别,限制了模型的通用性和可用性。为了将视觉模型零样本迁移到下游任务中,2021年,Radford等人^[48]将文本数据和图像数据相结合,提出对比语言-图像预训练(contrastive language-image pre-training, CLIP)模型,其核心思想是用自然语言来训练视觉模型,即在训练时,基于图像特征和文本特征进行对比学习。预训练之后,如图4所示,与GPT-3的提示工程类似,为特定的任务生成文本提示,然后通过预训练的文本编码器得到提示编码向量,预训练图像编码器得到待分类的图像特征;最后,计算图像特征和提示文本特征的余弦相似度以实现零样本迁移学习。

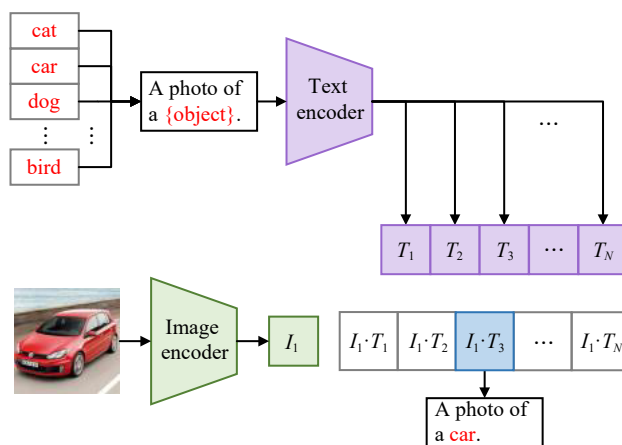


图4 CLIP模型实现零样本迁移

Fig.4 Overview of CLIP model implementing zero-shot transfer

为了使预训练视觉-语言模型适应 CV 和多模态领域的任务,Cho 等人^[49]将提示学习用于视觉-语言任务,提出一个统一的框架 VL-T5,该框架允许在 PLMs 中使用一组参数进行多任务学习,即使用不同的文本前缀作为提示,使模型适应不同下游任务。Yao 等人^[34]首次提出跨模态提示调优(cross-modal prompt tuning, CPT)方法,将视觉定位任务重新表述为完形填空题来解决跨模态差异,其中包含视觉和文本两个模态的提示,用来引导模型建立查询文本和被标记颜色的图像之间的联系。Zhang 等人^[50]首次将 CLIP 应用于 3D 点云任务,提出 PointCLIP 方法,在不进行任何 3D 训练的情况下,使用人工构建的提示对点云进行跨模态零样本分类。Jin 等人^[51]提出基于提示的低资源学习方法 FEWVLM,利用人工构建的提示模板,将预训练视觉-语言模型应用到下游任务。Li 等人^[52]提出一种新的视频-语言预训练框架 ALPRO,采用与 CLIP 类似的提示方法,提高模型捕获局部区域信息的能力,促进不同模态之间的对齐。

最近,Meta^[53]提出 SAM(segment anything model)模型,用于图像分割任务,极大地促进了 CV 领域基础模型的发展。如图 5 所示,SAM 模型主要由三部分组成,其中,图像编码器用于对图像进行编码得到描述图像的编码向量,提示编码器对点、框、mask、文本等提示信息进行编码,掩码解码器则用来将图像编码向量、提示编码向量和输出映射到掩码。SAM 模型是一种通过提示实现小样本和零样本学习的通用框架,自发布以来,SAM 模型及其衍生模型还被应用除图像分割外的其他视觉任务中,如目标检测^[54]、视频跟踪^[55]、图像标注^[56]等。SAM 发布之后,北京智源人工智能研究院紧接着推出了 SegGPT^[57],这是首个利用视觉上下文信息实现各种分割任务的通用模型。与 SAM 的交互式分割不同,SegGPT 具有上下文推理能力,可以执行任意的分割任务。

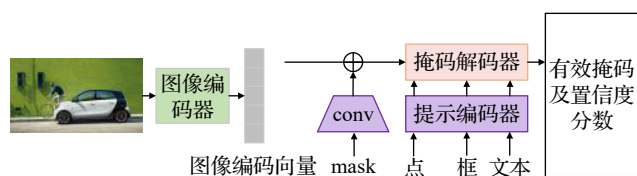


Fig.5 Framework of segment anything model(SAM)

人工构建提示模板的方法在一定程度上能够解决 NLP、CV 和多模态领域特定的下游任务,但是设计恰当的提示模板需要设计者具备极强的领域专业知识,且耗时耗力。提示模板的细微变化可能给 LM 带来巨大差异。针对该问题,有研究人员相继提出了自动构建提示模板和引入外部知识构建提示模板的方法。

1.2 自动构建提示模板的方法

自动构建提示模板的方法通常分为两种,一种是在

离散空间中自动搜索最优提示的方法,另一种是在连续嵌入空间中生成可优化的连续向量作为提示的方法。根据自动构建提示模板的方法的不同,将自动构建的提示模板分为离散提示和连续提示两种。

1.2.1 离散提示

离散提示又称为“硬提示”,通常是指在离散空间中自动搜索相应的提示模板。由于人工构建的提示模板取得的效果很可能不是最佳的,而且不恰当的提示可能导致模型无法检索到 LM 中的事实知识,Jiang 等人^[58]提出了基于挖掘和复述的方法来自动生成高质量和多样化的提示,并通过集成方法来组合来自不同提示的答案。Shin 等人^[25]提出一种基于梯度搜索的自动生成提示方法 AUTOPROMPT。该方法根据模板将原始任务输入与触发词集合结合起来,创建相应的提示,然后使用基于梯度的搜索来确定最佳提示。其中,触发词在所有的提示中共享。AUTOPROMPT 可以为一组不同的任务自动创建提示,在没有额外参数或微调的情况下,MLM 可以完成情感分析和自然语言推理任务。

针对小样本甚至零样本场景下,对话摘要任务无法扩展到新领域的问题,Zhao 等人^[59]提出一种面向领域的前缀微调模型,利用域单词初始化前缀模块,将目标域知识从源域中分离出来,并基于对话状态或查询来构建离散提示,以指导模型关注对话中的关键信息,增强对新领域的泛化能力。除此之外,还有一些研究人员利用 T5 模型自动生成离散的提示模板。例如,Gao 等人^[39]提出更好的小样本微调语言模型(better few-shot fine-tuning of language models, LM-BFF),该方法使用 T5 模型来自动创建提示模板,能够依赖少量标注样本进行通用搜索,可以更好地满足实际小样本应用场景。受 GPT-3 的启发,Jian 等人^[60]提出一个有监督的对比学习框架,该框架采用与 LM-BFF 相同的提示创建方法,通过给实例添加不同的语言提示和上下文信息来创建不同的视图,应用监督对比学习将来自同一类的输入进行聚类,进而有效地提高了模型的通用性。Song 等人^[61]提出了一个分类感知的提示学习框架 TaxonPrompt,以实现小样本事件分类任务。TaxonPrompt 利用基于 T5 模型的事件提示生成模块来自动生成特定于任务的提示,以探索事件分类答案。

在多模态领域,跨模态信息可能包含丰富的语义和上下文,单一的文本提示无法充分捕获这些信息,导致模型的理解和性能受限。针对这一问题,Yang 等人^[62]提出一种多模态概率融合提示模型 MultiPoint,对于文本模态,利用 T5 自动生成模板来获得多个有效的文本提示;对于图像模态,首先生成每个图像的文本描述,然后将其编码为文本提示,以提高图像和文本提示的兼容性。

1.2.2 连续提示

由离散字符组成的提示限制了提示的搜索空间。

为了克服离散空间的不足,进一步充分发挥提示的作用,由一系列连续向量组成的连续提示受到了更多的关注。连续提示也称为“软提示”,采用可学习向量作为提示模板,并在连续嵌入空间中学习最优提示,其在NLP、CV和多模态领域均有应用。

(1) NLP领域的连续提示方法

连续提示通常可以从如何生成最优连续提示和如何获取更好的初始化提示两个角度进行改进,本小节将详细讨论这两个改进方向的具体方法和措施。

① 最优连续提示生成

最优连续提示生成方法通常是在模型的训练阶段不断优化初始化的提示向量,以生成最优提示,提高模型的泛化能力。

微调PLMs来执行下游任务的方法会更新和保存其中所有的参数,这一过程会耗费大量的计算资源,且需要较大的存储空间。为解决这一问题,Li和Liang^[63]从提示学习中获得灵感,提出前缀调优(Prefix-tuning)方法,这是一种针对自然语言生成任务的微调替代方案。如图6所示,微调方法对于每个下游任务都需要更新所有Transformer参数,而Prefix-tuning使LM参数保持固定,只优化一个较小的、连续的、特定于任务的向量。Prefix-Tuning将提示模板优化为连续嵌入,其效果将向上传播到所有Transformer激活层,并向右传播到后续标记,该方法比需要匹配真实单词嵌入的离散提示方法性能更优。Lester等人^[6]提出提示调优(prompt tuning)方法,用于学习连续提示来调整固定参数的LM以执行特定的下游任务,该方法可以看作是Prefix-tuning方法的简化。在此基础上,Vu等人^[64]提出软提示迁移(soft prompt transfer, SPoT)方法,通过学习源任务提示来获得更好的初始化参数。对于每一个特定的任务,SPoT只需要存储一个较小的特定任务的提示,通过重用单个固定参数的预训练模型即可实现其他所有的任务。

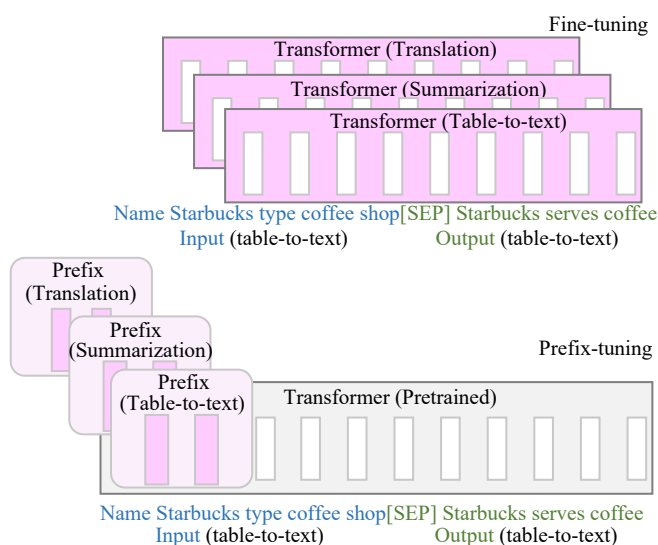


图6 微调与Prefix-tuning方法对比

Fig.6 Comparison between fine-tuning and Prefix-tuning

受CV领域的对抗性重构思想的启发,Hambardzmyan等人^[31]提出一种单词级对抗性重构(word-level adversarial reprogramming, WARP)方法,扩展了早期的自动提示生成工作。WARP采用随机梯度下降法在连续空间中学习任务特定的提示嵌入,并添加到输入文本中,实现知识从PLMs到下游任务的转移,指导LM解决特定的任务。Zhong等人^[65]提出一种连续优化提示的方法OPTIPROMPT,并应用于事实检索。OPTIPROMPT在连续的空间上优化提示,而不是将搜索空间限制在离散的字符空间,能更有效地从LM中提取事实。Liu等人^[9]提出一种通用的方法P-tuning,通过在连续空间中自动搜索更有效的提示来增强PLMs的自然语言理解能力。Wang等人^[66]在P-tuning的基础之上对其进行了扩展,提出一种可迁移的提示框架TransPrompt,采用多任务元知识获取程序来训练一个捕获跨任务可迁移知识的元学习器,并用于跨相似任务的小样本学习。

离散或连续提示方法^[5,8,63]通常独立于输入,缺乏对输入实例的语义的充分考虑,且统一的提示可能会对预测结果造成干扰,一些研究人员尝试生成与输入实例相关的提示。例如,Tang等人^[67]提出上下文调优(context-tuning)方法来微调PLMs,该方法首先根据输入文本生成提示,并从PLMs中获取有用的知识用于文本生成,然后使用连续逆提示,通过建模反向生成过程,增强生成文本与输入之间的相关性。Wu等人^[68]提出一种依赖于实例的提示生成(instance-dependent prompt generation, IDPG)模型,为每个输入实例生成提示,避免统一的提示可能会造成的干扰。对于下游任务,IDPG提取[CLS]层表示向量用于最终预测,因此,必须微调预训练MLM。为了确保下游预训练MLM的参数是固定的,Zhang等人^[69]首次提出基于提示生成模型的知识探测方法PromptGen,该方法可以根据输入句子自动生成提示。PromptGen通过微调预训练生成模型BART的参数以生成最优提示,同时保持下游预训练MLM参数的固定。

② 连续提示初始化优化

上一小节所介绍的方法主要分析如何生成最优连续提示以适应下游任务,还有一些研究人员针对连续提示初始化方面展开研究。Gu等人^[70]认为在小样本场景下提示学习性能低的原因在于初始化连续提示的方式,并提出预训练提示调优(pre-trained prompt tuning, PPT)框架,通过在预训练阶段加入连续提示来进行预训练提示,以获得更好的初始化,该方法的性能可以达到甚至优于数据充足时的微调PLMs方法。但是连续提示方法的性能在很大程度上依赖于良好的提示初始化,且提示初始化有任务限制,对于新的任务需要重新初始化提示。为了解决这一问题,Hou等人^[71]提出了一种称为MetaPrompting的通用软提示方法。该方法采用了与模

② 测试阶段的提示调优

上一小节所介绍的方法均使用下游任务中的训练数据来优化提示,缺乏对未知域的泛化能力。最近一些研究尝试利用测试时域自适应提示调优策略,使模型能更好地适应新的领域。Shu等人^[80]提出一种使用单个测试样本动态学习自适应提示的方法TPT,在测试时,给定单个样本,TPT通过生成多个随机增强视图来执行提示调优,并优化文本提示,使其具备零样本泛化能力。但是,TPT采用的是最小化边际熵来优化提示,可能会导致对错误结果产生过度信任。Ma等人^[81]提出Swap-Prompt框架,采用双重提示机制促进测试时自适应能力。以上方法动态学习文本提示,而忽略了视觉模态信息。Liu等人^[82]提出DART方法,通过学习特定于类的文本提示和实例级图像提示从每个测试样本中捕获知识,并利用双模态知识保留提示自适应地保留之前所获得的知识,增强对后续测试样本的预测能力。连续提示不再要求提示模板是由真实的单词组成的,而是可以优化的向量,因此这些方法极大地扩展了提示模板的可能性,从而获得更好的性能。

1.3 引入外部知识构建提示模板的方法

(1) NLP领域的引入外部知识构建提示模板方法

为特定的任务人工设计大量的自然语言提示模板繁琐且易出错,而自动构建的提示在标准监督场景下验证其有效性的代价较高,且特定任务的标签中往往蕴含丰富的语义信息。因此,研究人员尝试将任务标签中先验知识整合到提示调优中。在文本分类任务中,Han等人^[83]提出使用规则进行提示调优(prompt tuning with rules, PTR)的方法,将每个类别的先验知识编码为提示调优。在关系抽取任务中,Chen等人^[84]着重于将关系标签中的知识整合到提示调优进行关系提取,并提出具有协同优化的知识感知提示调优(knowledge-aware prompt-tuning, KnowPrompt)方法,旨在利用关系标签中包含的先验知识,来获取实例中给定头尾实体的可能的类型,并根据上下文信息对其进行优化,帮助模型更准确地捕获实体类型信息,提升关系抽取任务的性能。

除了可以利用类别标签中蕴含的丰富的语义知识外,还可以利用外部知识库来构建提示。Li等人^[85]提出一种知识注入提示调优(knowledge-injected prompt tuning, KiPT)模型,利用外部知识库来构建引入知识的提示,并采用提示调优策略来优化提示,有效解决了当前基于提示的事件检测方法因没有引入事件相关的语义知识而导致准确率低的问题。受到提示调优的启发,Wang等人^[86]提出一种基于知识提示的PLM(knowledge-prompt-based PLM, KP-PLM)框架,首先基于目标实体从知识库中为每个上下文分别构造知识子图,然后根据相应连续提示规则,将知识子图转换为自然语言提示。针对现有方法存在的先验知识缺失、知识噪声和知识异

质性问题,Ye等人^[87]提出一种本体增强的提示调优(ontology-enhanced prompt-tuning, OntoPrompt)方法。首先基于外部知识图谱的本体转换来弥补知识缺失问题,然后提出跨度敏感知识注入,以选择信息丰富的知识并减轻噪声注入,最后为了弥补知识和文本之间的差距,提出一种共同训练算法联合优化表示。OntoPrompt是一个通用框架,可以应用于关系抽取、事件抽取和知识图谱补全等多种下游任务。

提示学习方法在小样本场景下通常先将训练数据与提示模板融合,在训练过程中,模型侧重于学习训练数据中的知识,然后利用学习到的知识进行预测,导致模型的泛化能力弱。Chen等人^[88]提出RetroPrompt方法,从训练语料中检索相关上下文作为增强线索,减少LM对记忆的依赖,提高泛化能力。Yao等人^[89]提出一种检索增强的方法RAP,动态利用从训练数据中继承来的模式和知识作为每个实例的提示,增强了PLMs在信息抽取上的性能。

(2) CV和多模态领域的引入外部知识构建提示模板方法

在多模态领域,基于CLIP的方法受益于对比语言-图像预训练的优势,在小样本场景下表现出了良好的性能,Zhang等人^[90]尝试将更加多样化的预训练知识进行级联,并提出了CaFo模型,该模型结合了CLIP^[48]的语言对比知识、DINO^[91]的视觉对比知识、DALL-E^[92]的视觉生成知识以及GPT-3^[7]的语言生成知识,从而获得了更好的小样本学习能力。基于知识的视觉问答(visual question answering, VQA)任务通常需要利用图像之外的外部知识来回答问题,最近的研究主要将LLMs(如GPT3)作为隐式知识引擎来获取问答所需的知识^[93]。然而,LLMs的理解和推理能力并没有完全被激发,为此,Shao等人^[94]提出Prophet框架,首先针对特定的基于知识的VQA数据集训练一个VQA模型,然后从模型中提取候选答案和答案感知实例,并编码到提示中,使LLMs能够更好地理解任务,从而增强认知能力。针对视觉定位任务,Yang等人^[95]提出一种视觉提示方法SoM,首先使用交互式图像分割模型SAM将图像分割成不同粒度的图像区域,并用字母数字、掩码框等来覆盖不同图像区域,然后利用标记后的图像来提示LLMs实现视觉定位。

1.4 思维提示方法

上述通过构建简单的提示模板实现下游任务的方法在需要复杂逻辑推理能力的任务上表现很差,且通常不会随着LLMs规模的增加而取得明显的提升。以人类的思维方式为基础来构建思维提示,能够有效提升LLMs的逻辑推理能力,受到了研究人员的广泛关注^[23]。本节将从思维链提示方法和以思维链为基础的衍生方法两个方面来系统介绍基于思维提示方法如何

提升LLMs的推理能力。

1.4.1 思维链提示方法

提示学习范式经历了以离散提示为起点,连续提示为高潮的多个发展阶段。但是这些标准的提示学习方法在逻辑推理任务中的效果不佳,并且通常不会随着语言模型规模的增加而显著改善。2022年,Google^[20]首次提出了思维链提示(chain-of-thought prompting, CoT)概念,这是一种改进的提示策略,利用思维链来模拟人类思维,可以激发大模型的逻辑推理能力,帮助解决更为复杂的问题。

基于思维链的提示学习方法是一种特殊的上下文学习方法,通过生成一系列有逻辑关联的提示来指导LM生成连贯且相关的文本,提高LM在算术推理、常识推理和符号推理等复杂推理任务中的性能。如图8所示,标准提示学习范式只是给出<输入,输出>实例,而思维链提示方法的每个实例由<问题,思维链,答案>三部分组成,其中,思维链是一系列中间的自然语言推理步骤,这些步骤会使模型输出最终的推理结果,因此基于思维链的提示学习方法强调了思维推理的过程。

2023年,Google^[96]提出自一致性(self-consistency)解码策略,替代CoT中使用的贪婪解码策略,显著提高了思维链提示方法的性能。自一致性策略首先随机挑选一些样本,并使用人工标注的思维链提示来提示LM,然后通过从LM的解码器中采样以生成一组不同的推理路径,最后边缘化推理路径,在最终答案集合中选择最一致的答案作为最终答案。这种策略类似于人类经验,如果多种不同的思维方式得出相同的答案,则认为“最终生成的答案正确”的可信度更高。Li等人^[97]在Self-Consistency方法的基础上进一步优化,提出了DIVERSE方法,首先生成不同的提示来探索同一个问题的不同推理路径,然后使用验证器过滤出错误答案,最后单独验证每个推理步骤而非整个思维链。为了提高思维链提示方法在需要解决复杂问题的任务中的性

能,Zhou等人^[98]提出least-to-most prompt策略,该策略的关键思想是将复杂问题分解为一系列更简单的子问题,然后依次推理出每个子问题的答案。

CoT^[20]和self-consistency^[96]等方法需要为模型提供人工构建的思维推理路径来指示模型输出最终答案,实现小样本场景下的逻辑推理任务。但是,人工标注的<问题,思维链,答案>实例的成本高,且对于不同的任务并不一定是最有效的。Diao等人^[99]提出Activate-Prompt方法,采用基于不确定性的标注策略选择最有用、信息最丰富的问题进行标注,降低人工标注成本。Shao等人^[100]提出synthetic prompting方法,利用人工构建的种子实例来提示语言模型生成更多的实例,并选择一组有效的实例集使LM更好地进行思维推理。Shum等人^[101]提出一种自动提示增强和选择思维链的策略,Automate-CoT。第一阶段通过增强推理路径、修剪错误路径得到高质量的示例,第二阶段通过选择最佳组合来寻找更好的思维链提示。Automate-CoT不需要额外的人工成本,可以适用于任何任务和领域。

针对零样本场景下的复杂逻辑推理任务,Kojima等人^[102]提出一个两阶段的提示方式zero-shot-CoT,第一阶段先通过文字提示“Let’s think step by step”生成逻辑推理内容,第二阶段则基于第一阶段的思维链提示进行解析,使LM生成最终答案,该方法是与任务无关的,即不需要人工构建思维链提示,而是由模型自己生成问题推理过程,证明了LM出色的零样本推理能力。但是zero-shot-CoT方法仍然存在不稳定性,例如,对于语义相似的问题,该方法无法生成正确的思维链。Zhang等人^[103]提出了Auto-CoT方法,自动构建包含问题和思维链的实例。Auto-CoT方法包括两个步骤,首先通过聚类算法将数据集中的问题划分为多个簇,每个聚类簇中的问题在语义上是相似的,然后通过每个簇中选择一个代表性问题的方式,实现采样问题的多样性,并使用zero-shot-CoT生成思维链,通过自动生成思维链的方法

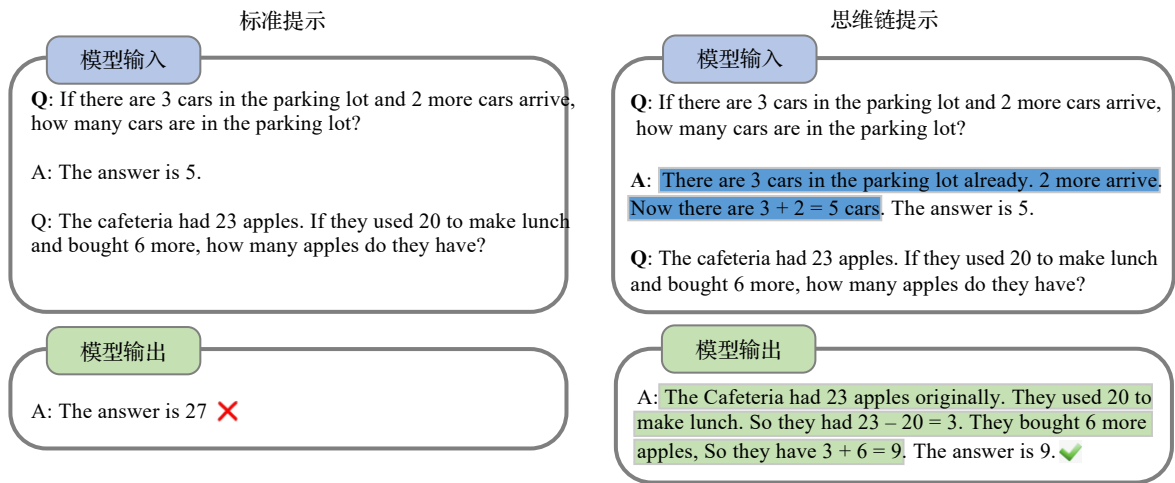


图8 思维链提示示例

Fig.8 Example of chain-of-thought

降低人工标注成本,通用性更强。Du 等人^[104]提出一种新的多智能体辩论框架,通过让多个 LM 各自独立生成推理过程,然后相互质疑、辩驳,经过多轮辩论之后生成最终答案。

为了提高 LLMs 在需要丰富领域知识和强大推理能力的任务中的信任度和性能,Zhao 等人^[105]提出了思维链提示的 verify-and-edit 框架,首先识别不确定性的实例,然后引入外部知识来验证和编辑推理链,最后根据编辑后的推理链生成最终的预测答案。Li 等人^[106]提出知识链(chain-of-knowledge, CoK)框架,通过动态合并来自异构源的基础信息来增强 LLMs。对于不确定性实例,CoK 利用查询生成器从结构化知识源中生成更可靠的事实信息,从而实现知识指导的逻辑推理。Wang 等人^[107]从外部知识库中检索结构化知识,并引入 F²-Verification 方法来评估推理链的真实性和可靠性,对不可靠的推理链进行重新思考,以提高 LLMs 的推理能力。Wang 等人^[108]提出一种知识驱动的思维链(knowledge-driven chain-of-thought, KD-CoT)框架,通过与外部知识库交互来验证和修订 CoT 中的推理链。Yao 等人^[109]提出一种通用的框架 ReAct,通过与维基百科 API 交互,克服了 CoT 推理中存在的错误传播问题,并产生类似于人类的任务解决轨迹,比没有推理过程的基线更具可解释性。Liu 等人^[110]提出一种生成知识提示的方法,将 LLMs 中的知识作为外部知识源,帮助模型进行常识推理,该方法从 LLM 中生成问题相关的知识,并将这些知识作为额外的输入,提示 LLM 生成答案。

现有的思维链提示相关研究主要集中在 NLP 领域,Zhang 等人^[111]首次提出多模态思维链提示方法 Multimodal-CoT,该方法通过两个阶段来实现多模态视觉问答任务。第一阶段利用文本和视觉多模态的输入信息生成思维链,第二阶段将原始文本输入和第一阶段生成的思维链相结合,使模型推理出答案,该方法使多模态信息得到充分利用。但是,Multimodal-CoT 采用的预训练视觉模型只能提取浅层的图像特征。He 等人^[112]提出 DPMM-CoT,通过学习一个能够捕获图像和文本之间的高级依赖关系的深层潜在空间,以深度融合图像特征和文本表示,提高多模态 CoT 的复杂推理能力。

1.4.2 基于思维链的衍生方法

思维链提示方法通过在提示中增加推理的中间步骤极大地提升了 LLMs 的逻辑推理能力,但是由于没有明确的思维分解过程,在复杂推理任务上表现不佳。近期,以链式思维提示为基础的衍生方法,如思维树、思维图和程序思维等逐渐受到关注。

(1) 结构衍生策略

尽管基于单条思维链的提示方法能够显著提高 LLMs 的逻辑推理能力,但是仍然存在信息遗失、可解释性差等问题,Yoran 等人^[113]提出多思维链推理方法,促使

LLMs 对多个思维链进行元推理,以提高推理答案的准确性。

2023 年,Yao 等人^[121]在思维链提示方法的基础上进一步提出了思维树(tree-of-thought, ToT)框架,如图 9 所示,这是一种允许语言模型在思维上探索多种推理方式的范式,用来解决需要探索或者预判的复杂任务。ToT 首先利用问题属性来设计和分解中间推理步骤,然后根据树的状态生成潜在的推理步骤,并利用状态评估器来评估中间状态以确定哪些状态要继续探索以及探索顺序,最后基于搜索算法查找解决方案。ToT 方法提高了模型决策的可解释性,但是 ToT 方法为思维过程强加了严格的树结构,导致推理能力受到限制。Besta 等人^[114]基于对人脑结构、逻辑推理等过程的思考,提出了思维图(graph of thoughts, GoT)框架,这是一种通过网络推理能力增强 LLMs 能力的方法,可以将 CoT 和 ToT 泛化到更加复杂的思维模式,其核心思想是将 LLMs 推理建模为任意的图结构,LLMs 思维是顶点,思维之间的依赖关系是边。这种方法可以将任意的 LLMs 思维组合成协同结果,提炼出整个思维网络的本质,或者使用反馈循环来增强思维。

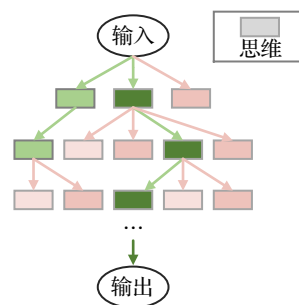


图9 ToT 框架

Fig.9 Framework of tree-of-thought (ToT)

(2) 算法衍生策略

ToT 和 GoT 方法利用生成思维和评估的迭代过程来搜索决策树,这需要大量的 LLMs 查询,增加了内存和计算资源开销。Sel 等人^[115]提出思维算法(algorithm of thoughts, AoT),这是一种通过算法化的推理路径推动 LLMs 的思考过程,提高其推理能力的策略,开创了一种新的上下文学习模式。为了降低 LLMs 的端到端生成延迟,Ning 等人^[116]受人类思维和写作过程的启发,提出了思维骨架(skeleton-of-thought, SoT)方法,首先指导 LLMs 生成答案的骨架,然后进行并行 API 调用或者处理解码来并行完成每个骨架点的内容,以加快模型的推理速度。

CoT 和 ToT 等方法利用 LLMs 实现逻辑推理任务时,不仅需要进行推理,还需在每一步中执行计算。但是,Chen 等人^[117]认为 LLMs 非常容易出现算术计算错误,且不能解决复杂的数学表达式,为此,提出程序思维(program-of-thought, PoT)提示方法,将推理步骤转化为

Python 程序,并利用外部语言解释器进行计算。Gao 等人^[22]提出程序辅助语言模型(program-aided language models, PAL)方法,使用 LLMs 理解自然语言问题,并生成程序作为中间推理步骤,而将解决方案委托给 python 解释器。

除了算术推理任务之外,研究人员还尝试利用代码提示方法来实现结构化的 NLP 任务。对于知识图谱构建任务,Bi 等人^[118]研究如何利用代码语言对结构化信息进行建模,并提出一种模式感知的代码提示和原理增强(code prompt and rationale-enhanced knowledge graph construction, CodeKGC)方法。CodeKGC 方法首先将原始的自然语言转换为代码格式,输入到任务特定的提示指导的代码语言模型中,然后利用模式感知提示来保存 KG 中的关系、属性等结构化信息,同时,受 CoT 的启发,利用原理增强模块将知识图谱构建任务分解为多个中间步骤,以提高推理能力。对于结构化的事件元素抽取任务,Wang 等人^[119]提出 Code4Struct 方法,利用文本到结构对的转化能力来处理结构化的预测任务。Code4Struct 方法首先将实体和事件类型本体转化为 python 类定义,以相关的类定义和输入句子为条件,将用于事件元素抽取的句子输入到字符串中,作为代码生成的提示,提示 LLMs 提取事件元素。对于信息抽取任务,Li 等人^[120]提出 CodeIE 方法,设计代码形式的提示,并利用 Codex^[121]等生成式的代码 LLMs 实现命名实体识别和关系抽取任务。

1.4.3 思维提示方法总结

本小节对思维提示相关的方法进行了总结,具体如表 1 所示。

思维提示方法的主要思想是向 LM 提供少量的<问题,思维推理路径,答案>实例,LM 在回答问题时也体现了其解释推理能力。随着推理过程从线性链到树再过渡到复杂的图结构和代码形式,思维提示方法显著提升了模型在算术推理、常识推理等复杂推理任务上的推理能力。但是,该类方法通常需要基于参数量在 100B 以上的 LLMs 才能诱发逻辑推理能力,对模型产生积极

影响,这也说明更大规模的 LM 蕴含更多的语义知识,利用逻辑性更强的提示可以激活 LM 中的隐式知识,以实现复杂推理任务^[36]。

1.5 基于提示模板的方法总结

根据提示模板构建方法的不同,可以进一步细分为人工构建提示模板、自动构建提示模板、引入外部知识构建提示模板和思维提示方法 4 种。本节对以上几种方法进行总结,给出各类方法的优缺点,具体如表 2 所示。

通过梳理基于提示模板方法的研究进展,可以考虑未来 3 个可能的改进方向。

(1)在嵌入空间中优化提示模板,构建通用的提示框架。由于真实世界中缺乏特定领域的标注数据,不利于 PLMs 适应各类下游任务。设计通用的、更加有效的提示模板构建方法,可以使 LM 在缺乏标注数据的场景下,在各类下游任务中取得更理想的性能。

(2)充分利用知识库、实例和标签中蕴含的事实知识,并用于构建有效的多模态提示模板。在小样本甚至零样本场景中,由于缺乏足够的数据供 LM 学习,为了提高提示学习方法的性能,可以充分利用外部知识,构建不同模态的提示信息,并融入原始输入中,从而使 PLMs 更好地适应下游任务,提升性能。

(3)思维提示方法通过模仿人类的思维方式挖掘 LM 的推理能力,在算术推理、常识推理等任务中表现出良好的性能。但是该类方法仅在参数量巨大的模型中发挥明显的积极作用,在未来的研究中需要进一步探索如何充分发挥中、小型模型的逻辑推理能力^[110,122]。

2 基于语言表达器的提示学习

语言表达式在提示学习中具有非常重要的作用,它是模型输出和最终预测结果之间的桥梁,用于将 PLMs 生成的潜在表示或输出映射到具体的单词或标签。例如,在情感分析任务中,语言表达式可以将模型预测的提示模板中[MASK]位置的输出映射到对应的标签“positive”或“negative”。根据语言表达器的构建方法的不同,本文将其分为人工构建表达式、基于搜索的表达

表 1 基于思维提示的方法总结
Table 1 Summary of methods based on thought prompting

方法	代表算法	特点	应用
思维链提示	COT ^[20]	通过人工、自动或引入外部知识的手段来构建逻辑推理的中间步骤,提高模型的推理能力和可解释性	算术、常识、符号推理任务等
	Automate-CoT ^[101]		
	Zero-shot-CoT ^[102]		
	ReAct ^[109]		
	Multimodal-CoT ^[111]		
思维链衍生方法	ToT ^[21]	对链式推理方法进行改进,通过多条思维链或者采用树、图结构进行思维推理,实现更加复杂的推理任务	迷你填字游戏、排序、创意写作等
	GoT ^[114]		
	PoT ^[117]	通过将推理步骤程序化的方式降低错误概率,提升 LLMs 的逻辑推理能力	数字推理、知识图谱构建等
	PAL ^[22]		
	CodeKGC ^[118]		

表2 基于提示模板的方法优缺点对比

Table 2 Comparison of advantages and disadvantages of methods based on prompt templates

分类	代表算法	优点	缺点	适用场景	总结
人工构建提示模板	LAMA ^[19]	可以针对具体任务需求设计具有针对性的提示模板;	确定合适的模板需要领域专业知识,且费时费力;	适用于任务目标和需求明确、规则性强的场景	[CLS] I like eating apples It was [MASK]. [SEP] 原始输入 → 人工提示 (人工设计模板)
	GPT-3 ^[7]				
	PET ^[40-41]	模板易于理解,便于引导模型学习特定的特征和知识	不易扩展到其他任务或领域,缺乏通用性		
	CLIP ^[48]				
自动构建提示模板	离散提示	AUTOPROMPT ^[25] LM-BFF ^[39]	节省人工构建成本,相比人工构建方法更加高效;任务适应性强;避免了主观因素对模型的影响	自动生成提示的方法很难达到与人工构建相当的性能;需要额外的计算成本用于生成和验证,不适合具有大规模实例和类别的标准学习设置	适用于需要对模型进行明确引导以识别特定特征或执行特定任务的场景 (利用 T5 模型自动生成离散的提示模板)
	连续提示	Prefix-Tuning ^[63] Prompt Tuning ^[8] PPT ^[70] CoOp ^[32] VPT ^[14]		适用于需要高度适应性和灵活性的复杂任务	[CLS] I like eating apples. $V_1 V_2 \dots$ [MASK] $\dots V_n$ [SEP] 原始输入 → 连续提示 (V_i 表示可学习向量)
	引入外部知识构建提示模板	PTR ^[83] KnowPrompt ^[84] KP-PLM ^[86]	节省人工成本;外部知识为模型提供更加丰富的语义信息;具有更高的灵活性和更广泛的适用性	需要有效的外部知识表示和集成方法;外部知识的准确性和完整性将直接影响模型的性能	适用于需要结合领域专业知识或大规模数据的任务 (引入外部知识构建知识提示模板)
	思维提示	CoT ^[20] Zero-shot-CoT ^[102] ToT ^[21] PoT ^[117] PAL ^[22]	可以模拟人类思维过程,通过一系列的中间推理步骤增强模型的逻辑推理能力	对参数达到 100B 的大型语言模型才能产生明显的积极影响;需要大规模的计算资源和时间成本	适用于需要深入分析和复杂逻辑推理、决策的场景 (蓝色部分为思维推理过程)

器、软表达器和引入外部知识构建表达器 4 种。如图 10 所示,人工构建的表达通过挑选一些真实标签词来描述类;基于搜索的表达器旨在通过算法从词汇表中找到合适的标签词;软表达器使用的是在调优过程中进行优化的可学习向量;引入外部知识构建的表达器则是通过引入外部知识库为每个标签生成对应标签词。接下来,就 4 种表达器中的典型算法模型及研究进展进行介绍。

2.1 人工构建表达器的方法

人工构建表达器的方法是通过人工设计一些标签词来描述一个类别。Schick 和 Schütze^[40]提出的 PET 方法引入模式-语言表达器对,对于给定的一组模式,即由人工构建的提示模板和原始输入组成的完形填空短语,人工定义一个单一表达器,预测[MASK]位置的标签词。具体来说,PET 方法将表达器定义为单射函数,将每个标签映射到词汇表中的一个具体单词,辅助模型利用 PLMs 中包含的知识实现下游任务。PET 方法在加入提示模板后,输入序列只有一个[MASK],即表达器只能

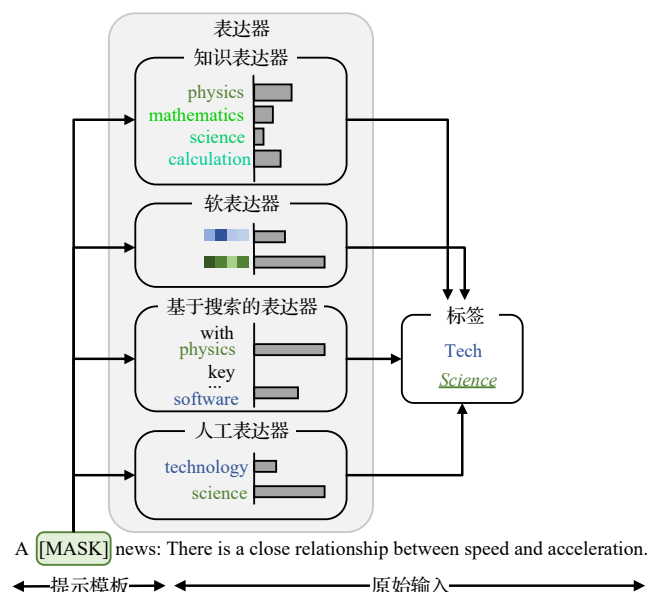


图 10 4 种表达器的构建方法图示

Fig.10 Illustration of four verbalizer construction methods

将每个输出映射到单个字符,这在很多任务中并不适用,因此 Schick 和 Schütze^[41]对 PET 方法进行了改进,支持标签对应多个字符。

尽管人工构建表达器的方法取得了一定的成功,但该类方法严重依赖于人工,耗时耗力,且要求设计者具备专业领域知识,对下游任务有精准的理解,能够用几个单词来总结每个类,如果没有特定任务的先验知识,选择恰当的标签词并非易事。基于此,研究人员相继提出了基于搜索的表达器、软表达器和引入外部知识构建表达器方法。

2.2 基于搜索的表达器方法

尽管 PET 等人工构建表达器的方法在小样本甚至零样本场景下取得了良好的效果,但是人工构建的表达器的实现 LM 的预测输出和标签之间的映射时,不仅需要特定任务的知识,还需要对 LM 内部的工作原理的理解,以识别 LM 能够充分理解的单词。Schick 等人^[123]引入带有自动标签的 PET (PET with automatic labels, PETAL),这是一种在给定少量训练数据的情况下,识别可以充当标签的单词的简单方法,其核心是将寻找最大化训练数据可能性的映射这一棘手问题分解为几个可管理的子问题。PETAL 通过自动将标签映射到单词来丰富 PET,而不需要任务或模型特定的知识。Shin 等人^[25]提出一种基于逻辑回归的两阶段启发式方法,用来自动选择标签词集合。Gao 等人^[39]基于排序搜索方法实现语言表达器的创建,通过对所有候选的标签词进行验证,选择输出概率最高的标签词。Ma 等人^[124]探索了自动选择适当的标签词的方法,并提出三种标签词的搜索方法:(1)使用数据分布进行搜索:从语料库中选择给定类中使用频率最高的单词;(2)使用 LM 输出分布进行搜索:使用 PLMs 选择标签词,获取模型在每个实体位置上的预测分布,选择每个实体位置预测 Top-*k* 的字符进行频率统计;(3)两种方法结合进行搜索:上述两种方法的频率相乘。对于冲突的可能出现在多个实体类别的高频标签词,采用阈值过滤法去除。

上述方法在探索自动化标签映射方面已经取得了一定的进展,但仍存在过程复杂、效率较低等问题。为此, Wang 等人^[125]提出一种简单而有效的自动选择标签映射与提示的小样本文本分类方法,即自动多标签提示 (automatic multi-label prompting, AMuLaP),利用一对多的标签映射和一个基于统计信息的算法来选择给定提示模板的标签映射。与之前的自动标签映射方法^[39, 123]相比, AMuLaP 更简单,且不需要访问主模型的权重,也不需要微调外部 PLMs 来搜索标签映射。Yu 等人^[126]提出一种全自动提示方法 AutoSeq,使用 T5 模型和波束搜索法自动生成大量的候选标签序列,并按对比概率重新排序,然后枚举标签序列组合,按照微调性能重新排序以获得最佳组合。其中,标签序列是指利用长度

不等的短语或句子来表示每个标签,标签序列相比单标签词更具表现力。

2.3 基于软表达器的方法

为最大限度地共享 LM 中的参数, Hambardzumyan 等人^[31]提出一种单词级对抗性重构方法 WARP,该方法不仅优化了提示嵌入,还引入答案相关的可优化参数。对抗性重构尝试学习特定任务的单词嵌入,当带有 [MASK] 标记的提示模板连接到输入文本时,将指示 LM 利用可训练的语言表达器来解决特定的任务。其中,语言表达器是使用梯度下降法优化得到的, WARP 在词嵌入部分为每个类别指定一个连续变量来表征这个类别,然后通过梯度回传来更新表征类别的词嵌入。基于搜索的表达器不便于从大规模词汇表中查找合适的单词,为此, Zhang 等人^[127]提出差异性提示 (differentiable prompt, DART) 方法,将 NLP 任务重新定义为 PLMs 任务,通过反向传播分别对提示模板和目标标签进行优化,同时学习连续的模板和标签映射。

针对低资源场景下存在训练数据缺乏和域迁移等问题,研究人员尝试构建软表达器以提升模型的性能。Chen 等人^[128]提出 LightNER 方法,为低资源 NER 进行轻量级调优范式。具体地说, LightNER 通过构建统一的可学习的实体类别表达器,以生成实体跨度序列和实体类别,不需要任何标签特定的分类器,从而解决了类迁移问题;同时, LightNER 包含一个可插入的指导模块,通过将可学习的参数加入自注意力层作为指导,进行具有跨领域知识迁移能力的轻量级调优。由于在大型词汇表或者低资源环境下的嵌入空间中进行充分搜索或优化具有一定的挑战性,因此,在某些情况下,自动构建的表达器的性能可能劣于人工构建的表达器, Cui 等人^[129]受原型网络和表达器之间存在内在相似性的启发,提出 ProtoVerb 方法,将原型向量作为语言表达器,包含了丰富的类别的语义信息,然后利用对比学习进行优化。实验表明,当人工表达器可用时, ProtoVerb 可以作为一个额外的表达器,而当样本有限,不借助人工表达器时, ProtoVerb 也能生成高质量的表达器。

2.4 引入外部知识构建表达器方法

虽然基于软表达器的方法可以与人工表达器相结合获得更好的性能,但是这种结合方式只能引入人工设计的几个单词或嵌入,这些单词或嵌入在词义或嵌入距离方面与类标签名接近,因此很难实现跨粒度的单词预测。为了提高人工表达器的覆盖范围,减少偏差,可以引入外部知识构建语言表达式以改进提示调优的性能。Hu 等人^[130]提出知识提示调优 (knowledgeable prompting, KPT) 方法,主要包含以下 3 个步骤:(1)在构建阶段,使用外部知识库为每个标签生成一组标签词,其中,扩展的标签词不仅仅是同义词,而是涵盖了不用粒

度和角度的标签词,因此更加全面、公正;(2)为了解决标签词噪声问题,更好地利用知识库,提出4种细化方法:频率细化、相关性细化、上下文校正和可学习细化;(3)最后应用损失函数将标签词得分映射到标签得分。尽管KPT方法通过细化方法降低了噪声问题,但是仍可能存在一些不合理的标签词对准确性产生消极影响。Ni等人^[131]提出KPT++,使用提示语法细化和概率分布细化来细化知识型表达式。由于KPT系列方法不是基于优化的,因此也更有利于零样本学习。

针对小样本文本分类任务,Zhang等人^[132]提出一种基于提示的元学习(prompt-based meta-learning, PBML)模型。在标签词初始化阶段,PBML引入外部知识图谱来扩展每个类的丰富标签词,然后通过平均候选词嵌入,将候选词合并到一个原型中;在优化阶段,不断合并来自支持集的上下文信息,对初始化的标签词嵌入进行优化,使其更具有鉴别性。该模型有效缓解了元学习方法需要大量元训练数据来构建元任务的不足。

对于细粒度实体分类任务,数据集通常使用分层的标签空间,如“位置/城市(LOCATION/CITY)”和“人/艺

术家(PERSON/ARTIST)”等。在这种情况下,Ding等人^[43]使用标签中各层标签所有的单词作为实体类型的标签词集,由于实体类型都是定义明确、边界明确的名词,因此,在外部知识图谱 Related Words(<https://related-words.org>)中获取与细粒度标签相关的名词来扩展标签词集,从而提升模型的性能。

2.5 基于语言表达式的方法总结

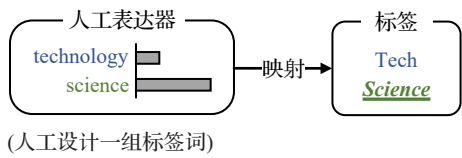
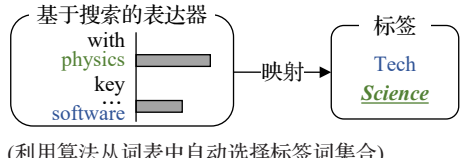
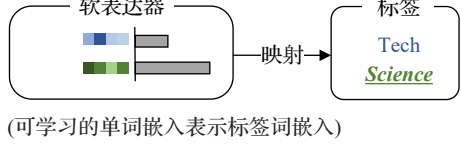
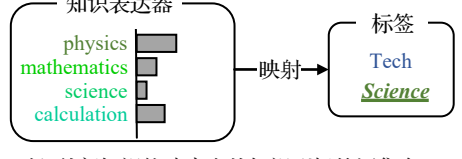
根据语言表达式构建方法的不同,可以分为人工构建表达式、基于搜索的表达式、软表达式和引入外部知识构建表达式的方法4种。本节对以上几种方法进行总结,给出各类方法的优缺点,具体如表3所示。

通过梳理基于语言表达式的方法的研究进展,可以考虑未来的两个可能的改进方向。

(1)尝试多种语言表达式构建方式相融合的方法。目前,多数提示学习方法是采用一种语言表达式构建方法来实现标签词到标签之间的映射。只有少部分的工作尝试使用多种方式构建的语言表达式相融合,例如文献[124]和文献[129]等。但是,不同的语言表达式方法具有不同的优势,未来可以进一步探索采用更加和谐的

表3 基于语言表达式的方法优缺点对比

Table 3 Comparison of advantages and disadvantages of methods based on language verbalizer

分类	代表算法	优点	缺点	适用场景	总结
人工构建表达式	PET ^[40-41]	能够实现标签词和类之间的精准映射,针对特定任务有效,并具有较强的可解释性	要求设计者对下游任务有准确的理解,并且每个类能够用几个词来概括,且耗时费力;标签词与类别之间的一对一映射限制了标签词的覆盖范围,缺乏足够的信息进行预测;可迁移性差	适用于数据规模较小、任务类别较为简单明确的应用场景,如情感分析	 (人工设计一组标签词)
基于搜索的表达式	PETAL ^[123] LM-BFF ^[39] AutoSeq ^[126]	通过算法从词表中找到合适的标签词,减轻人工构建成本;扩展性好,能够适用于较大的任务或更复杂的场景	标签词搜索过程的计算复杂度非常高;在大型词汇表中找到一个合适的目标标签词来表示特定的类别非常困难;需要足够的训练集和验证集进行优化	适用于任务规模较大、类别较多且需要一定的自动化处理的场景	 (利用算法从词表中自动选择标签词集合)
软表达式	DART ^[127] ProtoVerb ^[129]	使用可训练的向量在调优阶段进行优化,灵活性高,任务适用性强	需要大量的数据进行充分优化,在小样本场景下具有挑战性,可解释性相对较弱	适用于大规模、复杂的任务,特别是难以通过手动或搜索方式确定明确词汇的任务	 (可学习的单词嵌入表示标签词嵌入)
引入外部知识构建表达式	KPT ^[130] PBML ^[132]	可以利用外部知识,偏差小,覆盖面广,可用于零样本场景	需要引入额外的数据集或外部知识库,且外部知识的质量和覆盖范围直接影响模型的性能;知识整合难度大	适用于需要结合领域专业知识或外部数据源的任务	 (引入外部知识构建丰富的知识型标签词集合)

方式,实现多种方法的融合,提升模型的性能。

(2)探索引入外部知识的通用语言表达器构建方法。当前引入外部知识构建语言表达器的方法往往是针对特定的下游任务而设计的,即利用外部知识为特定的任务标签设计标签词集,缺乏任务通用性。针对此类问题,设计与任务无关的表达器构建方法,可以避免重新设计特定标签词集,使模型适用于各类下游任务。

3 主要应用

本章深入讨论了提示学习在NLP、CV和多模态领域的主要应用,通过对各领域中不同提示学习方法的对比和总结分析,揭示了不同提示学习方法在实际应用中的优势和局限。

3.1 NLP领域应用

本节首先对NLP领域部分下游任务及其常用数据集进行简单介绍,然后对各类提示学习方法在常用数据集上的实验结果进行了对比分析。

3.1.1 情感分析任务

情感分析旨在自动识别和提取文本中的倾向、立场、评价、观点等主观信息。在情感分析任务中,常用的数据集包括 SST-2^[133]、SST-5^[133]、MR^[134]和 CR^[135]等。其中,SST(sentence sentiment treebank)^[133]是由斯坦福大学发布的针对电影评论进行情感分析的数据集,其包括 SST-2 和 SST-5 两个单句子分类数据集;MR 是 Pang 和 Lee^[134]在 2005 年提出的二分类英文电影评论数据集;CR 是 Hu 和 Liu^[135]在 2004 年提出的二分类电商评论数据集。

本小节选取部分模型在常用数据集上的实验结果进行总结。其中,在小样本场景下均设置每个类别的训

练样本数 $K=16$ 。各模型的准确率(accuracy, Acc)对比具体如表 4 所示。

3.1.2 关系抽取任务

关系抽取(relation extraction, RE)旨在提取上下文中给定头尾实体之间的关系^[136]。在 RE 任务中,标准监督设置下常用数据集有 SemEval 2010 Task 8 (SemEval)^[137]、TACRED^[138]、TACRED-Revisit^[139]和 ReTACRED^[140]等;小样本场景下常用的数据集是 FewRel1.0^[141]和 FewRel2.0^[142]等;零样本场景下常用的数据集有 FewRel1.0^[141]和 WikiZSL^[143]等。其中,最常用的数据集为 SemEval 和 TACRED-Revisit。SemEval^[137]是一个传统的 RE 数据集,包含 9 个具有两个方向的关系和一个特殊的关系“Other”,其中,两个方向的关系是指由于两个实体的前后位置不同分成的两个关系。TACRED-Revisit^[139]是基于原始 TACRED 构建的数据集,主体提及的主题是人和组织,客体提及的主题有 16 种细粒度类型,包括日期、数字等。

本小节选取基于提示学习的 RE 方法在 SemEval 和 TACRED-Revisit 数据集上的实验结果进行总结。在 RE 任务中,所列文献在小样本场景下均设置每个类别的训练样本数 $K=8, 16, 32$ 。各模型的 F1 对比具体如表 5 所示。

3.1.3 自然语言推理任务

自然语言推理(natural language inference, NLI)旨在判断给定两个自然语言文本间的逻辑蕴含关系。在 NLI 任务中,常用的英文数据集有 SNLI^[144]、MNLI^[145]、MNLI-mm^[145]、QNLI^[146]和 RTE^[146]等,中文数据集有 OCNLI^[147]等。其中,最常用的数据集为 SNLI^[144]、MNLI^[145]和 QNLI^[146]。SNLI^[144]是 NLI 任务中最常用的数据集,包括蕴含、矛盾和中立 3 类标签。MNLI^[145]是

表 4 各模型在情感分析数据集上的准确率对比

Table 4 Comparison of accuracy of each methods on sentiment analysis datasets					单位: %
方法	方法分类	SST-2	MR	CR	
小样本	“GPT-3”in context learning ^[7]	84.8(1.3)	80.5(1.7)	87.4(0.8)	
	PET ^[40]	86.0(1.6)	80.0(1.6)	88.9(0.6)	
	NSP-BERT ^[46]	86.8(1.3)	80.5(1.5)	86.0(2.2)	
	LM-BFF ^[39]	92.3(1.0)	85.5(2.8)	89.0(1.4)	
	P-tuning ^[9]	92.7(1.4)	87.5(1.3)	90.6(1.4)	
	TransPrompt ^[66]	93.6	88.8	92.0	
	RetroPrompt ^[88]	93.9(0.4)	88.0(0.8)	91.9(0.7)	
	PET ^[40]	86.0(1.6)	80.0(1.6)	88.9(0.6)	
	AutoSeq ^[126]	89.8(1.1)	83.9(1.3)	87.2(2.5)	
	LM-BFF ^[39]	92.3(1.0)	85.5(2.8)	89.0(1.4)	
	DART ^[127]	93.5(0.5)	88.2(1.0)	91.8(0.5)	
	KPT ^[130]	90.3(1.6)	86.8(1.8)	88.8(3.7)	
零样本	PET ^[40]	83.6	80.8	79.5	
	NSP-BERT ^[46]	75.6	74.4	59.4	
	RetroPrompt ^[88]	86.8	83.5	79.7	
	KPT ^[130]	78.4	81.9	71.4	

表5 各模型在RE数据集上的F1对比

Table 5 Comparison of F1 of each model on RE datasets

单位:%

方法		方法分类	SemEval			TACRED-Revisit		
			<i>K</i> =8	<i>K</i> =16	<i>K</i> =32	<i>K</i> =8	<i>K</i> =16	<i>K</i> =32
小样本	LM-BFF ^[39]	离散提示	43.2	62.0	72.9	21.0	23.7	27.2
	PTR ^[83]	引入外部知识构建提示模板	70.5	82.9	84.8	28.7	31.4	32.4
	KnowPrompt ^[84]	引入外部知识构建提示模板	74.3	82.9	84.8	32.1	33.1	34.7
	OntoPrompt ^[87]	引入外部知识构建提示模板	52.6	65.0	83.0	28.8	33.1	34.8
	LM-BFF ^[39]	基于搜索的表达式	43.2	62.0	72.9	21.0	23.7	27.2
	DART ^[127]	软表达式	51.8	67.2	77.3	25.8	30.1	31.8
标准监督 设置	LM-BFF ^[39]	离散提示		88.0			76.4	
	PTR ^[83]	引入外部知识构建提示模板		89.9			81.4	
	KnowPrompt ^[84]	引入外部知识构建提示模板		90.2			82.4	
	OntoPrompt ^[87]	引入外部知识构建提示模板		89.1			78.2	
	LM-BFF ^[39]	基于搜索的表达式		88.0			76.4	
	DART ^[127]	软表达式		89.1			77.8	

SNLI的扩展,涵盖了更大范围的书面和口语文字类型。QNLI^[146]是从SQuAD1.0^[148]数据集转换而来的,通过将问题和上下文中的每一句话进行组合,并过滤掉词汇重叠比较低的句子对得到QNLI中的句子对。

本小节选取部分的基于提示学习的NLI模型在SNLI、MNLI和QNLI数据集上的实验结果进行总结。其中,所列文献均在小样本场景下实现,且均设置每个类别的训练样本数 $K=16$ 。各模型的准确率对比具体如表6所示,其中“—”表示模型未在该数据集上的进行实验。

3.1.4 其他任务

除了上述的任务之外,提示学习还常用于实现文本分类^[149-151]、命名实体识别^[124, 152-154]、知识图谱补全^[87, 155]和机器翻译^[156-159]等NLP领域的下游任务。其中,文本分类任务常用数据集包括AG's News^[160]、Yelp^[160]、Yahoo^[160]、Amazon^[161]和DBPedia^[162]等,常用评价指标包括Acc、精准率(precision, P)、召回率(recal, R)、F1值等;命名实体识别任务常用数据集包括CoNLL2003^[163]、OntoNotes 5.0^[164]、MIT-Movie^[165]和Few-NERD^[166]等,常用评价指标包括P、R、F1等;知识图谱补全任务常用数据集包括FB15k-237^[167]和WN18RR^[168]等,常用评价指标包括MRR、MR、Hit@k等;机器翻译任务常用数据集包括

IWSLT、WMT等,常用评价指标包括BLUE分数^[169]、COMET分数^[170]等。

由于篇幅限制,本文不再对提示学习方法在上述任务中的实验结果展开介绍和讨论。

3.1.5 实验分析

表4、5和6总结了提示学习方法在3种NLP任务中的实验结果对比,由实验结果可以综合得出以下结论。

(1)自动构建提示模板的方法和基于搜索的表达式方法不仅能够节省人工成本,而且能够达到甚至优于人工构建提示模板和表达式的方法的性能。

(2)连续提示方法通过在训练过程中不断优化,生成连续、可学习的向量作为提示模板,可以使模型生成最优提示,提升模型的性能。

(3)基于软表达式的方法使用在模型调优阶段优化的可训练向量作为软标签,不仅能够降低搜索标签词的复杂度,而且能够达到更优的性能。

(4)引入外部知识的方法明显优于其他方法,表明引入特定任务的知识来增强模型是可行的,可以更好地服务于下游任务。

(5)基于提示的方法不仅适用于资源丰富的监督学习场景,同样适用于小样本甚至零样本场景,当用于训练的各类别样本数增加时,小样本场景下的模型性能与

表6 各模型在NLI数据集上的准确率对比

Table 6 Comparison of accuracy of each methods on NLI datasets

单位:%

方法	方法分类	SNLI	MNLI	QNLI
“GPT-3”in-context learning ^[7]	人工构建提示模板	47.1(0.6)	52.0(0.7)	53.8(0.4)
LM-BFF ^[39]	离散提示	77.1(2.1)	68.3(2.5)	68.3(7.4)
P-Tuning ^[9]	连续提示	72.3(3.0)	61.5(2.1)	64.3(2.8)
TransPrompt ^[66]	连续提示	77.0	71.9	—
RetroPrompt ^[88]	引入外部知识构建提示模板	—	71.1(1.8)	71.6(1.8)
AutoSeq ^[126]	基于搜索的表达式	62.7(3.7)	51.8(1.8)	61.3(4.0)
LM-BFF ^[39]	基于搜索的表达式	77.1(2.1)	68.3(2.5)	68.3(7.4)
DART ^[127]	软表达式	75.8(1.6)	67.5(2.6)	66.7(3.7)

富资源场景下的性能相当。

3.2 CV 和多模态领域应用

受提示学习在NLP领域发展的启发,研究人员将其应用于CV和多模态领域,并取得了一定的进展。本节对提示学习在CV和多模态领域的部分应用进行总结,具体如表7所示。其中,图像分类、视觉定位和视觉问答任务常用评价指标为Acc,图像分割任务常用评价指标包括IoU(intersection over union)、mIoU(mean inter-

section over union)等,多模态情感分析任务常用评价指标包括Acc和F1等,文本-视频检索任务常用评价指标包括R@k(recall at rank k)、MnR(mean rank)和MdR(median rank)等。

表7总结了提示学习方法在CV和多模态领域的应用。由表7可以得出:

(1)提示学习在CV和多模态领域的应用以人工构建提示模板和连续提示模板为主,而在引入外部知识构

表7 应用于CV和多模态领域的提示学习方法总结
Table 7 Summary of prompt learning methods applied in CV and multimodel fields

方法分类	方法	发表年份	会议/期刊	下游任务	提示形式	实验数据集
人工构建提示模板	CLIP ^[48]	2021	ICML	图像分类、视频动作检测、地理定位等	文本	aYahoo ^[171] 、ImageNet ^[172] 、SUN397 ^[173] 、Caltech101 ^[174] 等
	CPT ^[34]	2024	AI Open	视觉定位	文本、图像	RefCOCO ^[175] 、RefCOCO+ ^[175] 、RefCOCOg ^[176]
	VL-T5 ^[49]	2021	ICML	视觉问答、参考表达式理解、多模态机器翻译等	文本	VQA v2.0 ^[177] 、Visual7W ^[178] 、RefCOCOg ^[176] 、Multi30K ^[179]
	PointCLIP ^[50]	2022	CVPR	3D点云	文本	ModelNet10 ^[180] 、ModelNet40 ^[180] 、ScanObjectNN ^[181]
	FEWVLM ^[51]	2022	ACL	视觉问答、图像字幕和图像分类	文本	VQA v2.0 ^[177] 、OKVQA ^[182] 、NoCaps ^[183] 、Flickr30K ^[184] 、miniImageNet ^[185]
	ALPRO ^[52]	2022	CVPR	细粒度视频文本对齐	文本	WebVid-2M ^[186] 、MSRVTT ^[187] 、DiDeMo ^[188] 、MSVD-QA ^[189] 、MSRVTT-QA ^[189]
	SAM ^[53]	2023	ICCV	图像分割	图像、文本	SA-1B ^[53]
自动构建提示模板	离散提示MultiPoint ^[62]	2023	ACM MM	多模态情感分析	文本	MVSA-Single ^[190] 、MVSA-Multiple ^[190] 、TumEmo ^[191]
	CoOp ^[32]	2022	IJCV	图像分类	可学习向量	ImageNet ^[172] 、Caltech101 ^[174] 、Flowers102 ^[192] 、EuroSAT ^[193] 等
	CoCoOp ^[72]	2022	CVPR	图像分类	可学习向量	ImageNet ^[172] 、Caltech101 ^[174] 、Flowers102 ^[192] 、EuroSAT ^[193] 等
	KgCoOp ^[73]	2023	CVPR	图像分类	文本、可学习向量	ImageNet ^[172] 、Caltech101 ^[174] 、Flowers102 ^[192] 、EuroSAT ^[193] 等
	SEEM ^[74]	2023	NeurIPS	全景分割、指代分割、交互式分割	可学习向量	RefCOCO ^[175] 、RefCOCOg ^[176] 、RefCOCO+ ^[175] 等
	连续提示DenseCLIP ^[75]	2022	CVPR	密集预测	图像编码	ADE20K ^[194]
	VPT ^[14]	2022	ECCV	视觉分类	可学习参数	NABirds ^[195] 、Stanford Cars ^[196] 、Flowers102 ^[192] 、Caltech101 ^[174] 等
	VoP ^[77]	2023	CVPR	文本-视频检索	可学习参数	MSRVTT ^[187] 、DiDeMo ^[188] 、ActivityNet ^[197] 等
	TPT ^[80]	2022	NeurIPS	图像分类	可学习向量	ImageNet ^[172] 、ImageNet-A ^[198] 、ImageNet-R ^[199] 等
	SwapPrompt ^[81]	2023	NeurIPS	图像分类	可学习的文本提示	ImageNet ^[172] 、ImageNet-A ^[198] 、Flowers102 ^[192] 等
	DART ^[82]	2024	AAAI	图像分类	可学习的文本和图像提示	ImageNet-A ^[198] 、ImageNet-R ^[199] 等
	引入外部知识构建提示CaFo ^[90]	2023	CVPR	图像分类	文本	ImageNet ^[172] 、Caltech101 ^[174] 、Flowers102 ^[192] 、EuroSAT ^[193] 等
思维提示	Prophet ^[94]	2023	CVPR	视觉问答	图像、文本	OKVQA ^[182] 、A-OKVQA ^[200]
	SoM ^[95]	2023	arXiv	视觉定位	图像	ADE20K ^[194] 、Flickr30K ^[184] 、RefCOCO ^[175] 等
	Multimodal-CoT ^[111]	2023	TCLR	视觉问答	图像、文本	ScienceQA ^[201]
	DPMM-CoT ^[112]	2024	AAAI	视觉问答、多模态机器翻译	图像、文本	ScienceQA ^[201] 、Multi30K ^[179]

建提示模板和思维链提示两个方面的研究相对匮乏,仍需要进一步研究,尝试利用丰富的外部知识和逻辑推理过程,提升基于提示的方法在CV和多模态领域的应用效果。

(2)CV和多模态领域的提示学习方法大多受启发于NLP领域的方法,提示模板的形式也大多采用文本形式,图像和视频等模态中的特征很少被用于生成提示以提升基于提示学习方法的模型性能。

基于以上描述可以发现,尽管提示学习方法已经在CV和多模态领域取得了一定的进展,但仍然有很大的发展和提升空间。通过有效利用多模态数据,进一步改进和优化提示策略,以充分利用预训练模型中丰富的知识,将提升图像生成^[202]、图文检索^[203]、目标检测^[204]等多种下游任务的性能。

4 提示学习挑战与展望

4.1 提示学习挑战

尽管近年来提示学习已经得到了广泛研究,并且取得了一定的进展,但仍面临着一些挑战。

(1)对提示意义的真正理解与可解释性

基于提示学习的方法在小样本和零样本场景下中取得了很大进展,但是这也自然产生了一种假设,即任务特定的提示类似于指导人类快速学习的任务指令,可以作为一种具有语义意义的指令来指导模型学习任务相关的知识。针对这一问题,Webson和Pavlick^[205]质疑基于提示的模型是否能够真正理解其提示的意义,并进行了大量的实验,结果与一般假设相悖。若模型能够理解提示的真实意义,将帮助模型更好地理解任务指令,并在下游任务中取得更好的性能。因此,进一步探索如何提升模型对提示的理解能力和可解释性,仍需要研究人员的深入研究。

(2)跨领域知识迁移与泛化能力

基于提示的方法通常是针对特定的领域任务来设计恰当的提示模板和语言表达器,实现PLMs中知识向下游任务的迁移,帮助模型快速适应不同的任务。但是,这类方法可能会过度依赖于提示的设计,忽略了提示在未知域中的泛化潜力,导致在面临未知的领域或任务时,模型可能无法准确理解问题的上下文或生成准确的输出^[206]。研究更好的迁移学习和领域适应方法以提升模型的泛化能力具有一定的挑战。

(3)多模态领域的应用

在NLP领域,提示学习方法已经取得了很大的进展,但是在图像、视频等的应用还不是很广泛,尤其是面对遥感图像^[207]、医学图像等存在的各种复杂情况,基于提示学习方法的研究更是匮乏,有很大的发挥空间。另外,跨模态知识迁移也是一个难题^[34, 208],由于不同模态间天然的语义鸿沟等问题,导致在不同模态之间进行

综合建模,构建融合多模态特征的提示仍存在一定的挑战。

4.2 提示学习展望

通过对当前提示学习研究进展的梳理,可以展望未来提示学习的发展方向。

(1)在数据层面,尝试利用外部知识库、任务相关先验知识等来构建提示模板和标签词集,或者更好地利用无标注数据。为了使模型真正理解提示的意义,未来的研究应更加专注于探索新的评估方法,测试模型对提示的理解深度,同时可以探索利用先验知识(如输入实例、任务标签中蕴含的语义信息等)使模型充分理解任务指令,提升模型在特定任务上的准确性和可解释性。

(2)提示学习方法经常会面临着特征、参数和任务迁移挑战。为了使模型设计的提示和标签词能够被迁移到其他任务领域,在未来的研究中可以着力探索通用性强的提示设计方法,增强模型在不同任务和领域的泛化能力,开发自适应提示模板和表达器生成方法,使模型在面对新领域时能够动态调整,设计更加高效、合理的迁移算法,提升模型的跨领域泛化能力。

(3)尽管提示学习在NLP领域取得了显著的进展,但其在图像、视频等多模态领域的应用仍相对不足。未来研究可以探索如何有效设计跨模态提示信息,实现文本、图像、视频等模态之间的关联,同时,可以深入探究在复杂场景下(如遥感图像、医学图像等)的提示生成与应用。

(4)尝试不同提示方法的融合。现有的提示学习方法大多是使用单一的提示模板或表达器构建方法,在以后的工作中可以尝试将多种构建方法相结合,从专家知识和模型训练两个方面进行改进,探索多种构建策略的集成,从而提升模型的性能。此外,可以设计多层次的提示模板,结合不同任务需求,优化提示与模板训练的适配性。以达到更好的效果。

(5)探索如何在中、小型LM模型中进行复杂的逻辑推理,并更好地应用于视觉和多模态等领域。目前,思维提示方法的研究主要集中于NLP领域,且小规模模型的常识知识、逻辑推理能力不佳,很难直接用于复杂逻辑推理任务。在今后的工作中,可以尝试将LLMs作为外部知识源,为小型LM提供常识知识等。

(6)在应用方面,尝试将提示学习方法应用于金融、医疗^[209-210]、法律等需要高质量标注数据的领域。这些领域通常需要专家进行高质量的标注,但是往往缺乏足够标注的标注数据供LLMs进行学习训练,提示学习可以通过少量的提示和无标注数据帮助模型快速适应新任务。在未来,研究人员可以专注于探索适合这些领域的提示设计方法,并探索如何更好地结合领域知识,提升模型的性能和效率。

5 结束语

随着BERT、GPT、T5等LLMs的出现,很多NLP的研究和应用都是以PLMs为中心进行展开的。但是,由于预训练任务和微调任务在数据和目标函数方面存在差距,尤其是在低资源场景下,微调PLMs存在巨大挑战。提示学习通过设计简洁但高效的提示指导预训练模型在小样本甚至零样本场景下进行学习,激发模型强大的语义理解和生成能力,提升模型的知识迁移能力和跨领域适应能力,逐渐受到研究人员的广泛关注。提示学习通过指导PLMs适应各类下游任务,展现出了良好的应用前景。随着提示学习在NLP领域的研究与发展,其在CV和多模态领域也逐渐受到关注,并取得了一定的研究进展。

本文对提示学习方法的相关文献进行了系统的分类整理和对比,从技术发展的角度详细叙述了提示学习范式的变革历程,对不同提示学习方法的研究进展进行了广泛而全面的阐述;介绍了提示学习在NLP、CV和多模态领域常见任务中的应用;最后,本文分析了目前提示学习面临的挑战,对其进行了前景展望。本文是首篇从提示模板和语言表达器的两个阶段,详细介绍大模型背景下的提示学习方法在NLP、CV和多模态领域研究与应用的综述。本文可以帮助读者更好地理解 and 选择合适其研究和应用的方法,此外,还对两个阶段的方法所面临的关键问题和挑战进行了深入讨论,提出了一些前瞻性的问题,以促进更深入地研究。

参考文献:

- [1] HOWARD J, RUDER S. Universal language model fine-tuning for text classification[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2018: 328-339.
- [2] RADFORD A, WU J, CHILD R, et al. Language models are unsupervised multitask learners[J]. OpenAI Blog, 2019, 1(8): 9.
- [3] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019: 4171-4186.
- [4] LIU Y, OTT M, GOYAL N, et al. RoBERTa: a robustly optimized BERT pretraining approach[J]. arXiv:1907.11692, 2019.
- [5] RAFFEL C, SHAZEER N, ROBERTS A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer[J]. The Journal of Machine Learning Research, 2020, 21(1): 5485-5551.
- [6] 舒文韬, 李睿潇, 孙天祥, 等. 大型语言模型:原理、实现与发展[J]. 计算机研究与发展, 2024, 61(2): 351-361.
- [7] SHU W T, LI R X, SUN T X, et al. Large language models: principles, implementation, and progress[J]. Journal of Computer Research and Development, 2024, 61(2): 351-361.
- [7] BROWN T, MANN B, RYDER N, et al. Language models are few-shot learners[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems, 2020: 1877-1901.
- [8] LESTER B, AL-RFOU R, CONSTANT N. The power of scale for parameter-efficient prompt tuning[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 2021: 3045-3059.
- [9] LIU X, ZHENG Y, DU Z, et al. GPT understands, too[J]. AI Open, 2024, 5: 208-215.
- [10] WEI C, XIE S M, MA T. Why do pretrained language models help in downstream tasks? an analysis of head and prompt tuning[C]//Proceedings of the 35th International Conference on Neural Information Processing Systems, 2021: 16158-16170.
- [11] 张钦彤, 王昱超, 王鹤羲, 等. 大语言模型微调技术的研究综述[J]. 计算机工程与应用, 2024, 60(17): 17-33.
- [12] ZHANG Q T, WANG Y C, WANG H X, et al. Comprehensive review of large language model fine-tuning[J]. Computer Engineering and Applications, 2024, 60(17): 17-33.
- [12] WANG Y, HUANG Z, HONG X. S-prompts learning with pre-trained transformers: an occam's razor for domain incremental learning[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems, 2022: 5682-5695.
- [13] KHATTAK M U, RASHEED H, MAAZ M, et al. Maple: multi-modal prompt learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 19113-19122.
- [14] JIA M, TANG L, CHEN B C, et al. Visual prompt tuning [C]//Proceedings of the 17th European Conference on Computer Vision(ECCV 2022), Tel Aviv, Israel, October 23-27, 2022. Cham: Springer Nature, 2022: 709-727.
- [15] SOLLAMI M, JAIN A. Multimodal conditionality for natural language generation[J]. arXiv:2109.01229, 2021.
- [16] OpenAI. ChatGPT[EB/OL]. (2023)[2024-06-15]. <https://openai.com/blog/chatgpt>.
- [17] OpenAI. GPT-4 technical report[R/OL]. (2023)[2024-06-15]. <https://cdn.openai.com/papers/gpt-4.pdf>.
- [18] 桑基韬, 于剑. 从ChatGPT看AI未来趋势和挑战[J]. 计算机研究与发展, 2023, 60(6): 1191-1201.
- [19] SANG J T, YU J. ChatGPT: a glimpse into AI's future[J]. Journal of Computer Research and Development, 2023, 60(6): 1191-1201.
- [19] PETRONI F, ROCKTÄSCHEL T, RIEDEL S, et al. Language models as knowledge bases?[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: 2463-2473.
- [20] WEI J, WANG X, SCHUURMANS D, et al. Chain-of-thought prompting elicits reasoning in large language models [C]//Proceedings of the 36th International Conference on Neural Information Processing Systems, 2022: 24824-24837.
- [21] YAO S, YU D, ZHAO J, et al. Tree of thoughts: deliberate problem solving with large language models[C]//Proceedings of the 37th International Conference on Neural Information Processing Systems, 2023: 11809-11822.
- [22] GAO L, MADAAN A, ZHOU S, et al. PAL: program-aided language models[C]//Proceedings of the International Conference on Machine Learning, 2023: 10764-10799.
- [23] CHU Z, CHEN J, CHEN Q, et al. A survey of chain of thought reasoning: advances, frontiers and future[J]. arXiv: 2309.15402, 2023.
- [24] LIU P, YUAN W, FU J, et al. Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing[J]. ACM Computing Surveys, 2023, 55(9): 1-35.
- [25] SHIN T, RAZEGHI Y, LOGAN IV R L, et al. AutoPrompt: eliciting knowledge from language models with automatically generated prompts[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020: 4222-4235.
- [26] SU Y, WANG X, QIN Y, et al. On transferability of prompt tuning for natural language processing[C]//Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2022: 3949-3969.
- [27] LIU F, LIN H, HAN X, et al. Pre-training to match for unified low-shot relation extraction[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2022: 5785-5795.
- [28] CHIA Y K, BING L, PORIA S, et al. RelationPrompt: leveraging prompts to generate synthetic data for zero-shot relation triplet extraction[C]//Findings of the Association for Computational Linguistics (ACL 2022), 2022: 45-57.
- [29] WANG J, ZHANG L, LIU J, et al. MatchPrompt: prompt-based open relation extraction with semantic consistency guided clustering[C]//Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 2022: 7875-7888.
- [30] LOGAN IV R, BALAZEVIĆ I, WALLACE E, et al. Cutting down on prompts and parameters: simple few-shot learning with language models[C]//Findings of the Association for Computational Linguistics (ACL 2022), 2022: 2824-2835.
- [31] HAMBARDZUMYAN K, KHACHATRIAN H, MAY J. WARP: word-level adversarial reprogramming[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: 4921-4933.
- [32] ZHOU K, YANG J, LOY C C, et al. Learning to prompt for vision-language models[J]. International Journal of Computer Vision, 2022, 130(9): 2337-2348.
- [33] JU C, HAN T, ZHENG K, et al. Prompting visual-language models for efficient video understanding[C]//Proceedings of the 17th European Conference on Computer Vision (ECCV 2022), 2022: 105-124.
- [34] YAO Y, ZHANG A, ZHANG Z, et al. CPT: colorful prompt tuning for pre-trained vision-language models[J]. AI Open, 2024, 5: 30-38.
- [35] 鲍琛龙, 吕明阳, 唐晋韬, 等. 与知识相结合的提示学习研究综述[J]. 中文信息学报, 2023, 37(7): 1-12.
- BAO C L, LYU M Y, TANG J T, et al. A survey of prompt learning combined with knowledge[J]. Journal of Chinese Information Processing, 2023, 37(7): 1-12.
- [36] QIAO S, OU Y, ZHANG N, et al. Reasoning with language model prompting: a survey[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023: 5368-5393.
- [37] LIU P, ZHANG L, GULLA J A. Pre-train, prompt and recommendation: a comprehensive survey of language modeling paradigm adaptations in recommender systems[J]. Transactions of the Association for Computational Linguistics, 2023, 11: 1553-1571.
- [38] 廖宁, 曹敏, 严骏驰. 视觉提示学习综述[J]. 计算机学报, 2024, 47(4): 790-820.
- LIAO N, CAO M, YAN J C. Visual prompt learning: a survey[J]. Chinese Journal of Computers, 2024, 47(4): 790-820.
- [39] GAO T, FISCH A, CHEN D. Making pre-trained language models better few-shot learners[C]//Proceedings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, 2021: 3816-3830.
- [40] SCHICK T, SCHÜTZE H. Exploiting cloze-questions for few-shot text classification and natural language inference [C]//Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, 2021: 255-269.
- [41] SCHICK T, SCHÜTZE H. It's not just size that matters: small language models are also few-shot learners[C]//Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021: 2339-2352.
- [42] CUI L, WU Y, LIU J, et al. Template-based named entity recognition using BART[C]//Findings of the Association for Computational Linguistics (ACL-IJCNLP 2021), 2021:

- 1835-1845.
- [43] DING N, CHEN Y, HAN X, et al. Prompt-learning for fine-grained entity typing[C]//Findings of the Association for Computational Linguistics (EMNLP 2022), 2022: 6888-6901.
- [44] ZHANG S, JI T, JI W, et al. Zero-shot event detection based on ordered contrastive learning and prompt-based prediction [C]//Findings of the Association for Computational Linguistics(NAACL 2022), 2022: 2572-2580.
- [45] WANG S, FANG H, KHABSA M, et al. Entailment as few-shot learner[J]. arXiv:2104.14690, 2021.
- [46] SUN Y, ZHENG Y, HAO C, et al. NSP-BERT: a prompt-based few-shot learner through an original pre-training task—next sentence prediction[C]//Proceedings of the 29th International Conference on Computational Linguistics, 2022: 3233-3250.
- [47] SAINZ O, DE LACALLE O L, LABAKA G, et al. Label verbalization and entailment for effective zero and few-shot relation extraction[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 2021: 1199-1212.
- [48] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision[C]//Proceedings of the International Conference on Machine Learning, 2021: 8748-8763.
- [49] CHO J, LEI J, TAN H, et al. Unifying vision-and-language tasks via text generation[C]//Proceedings of the International Conference on Machine Learning, 2021: 1931-1942.
- [50] ZHANG R, GUO Z, ZHANG W, et al. PointCLIP: point cloud understanding by CLIP[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022: 8542-8552.
- [51] JIN W, CHENG Y, SHEN Y, et al. A good prompt is worth millions of parameters: low-resource prompt-based learning for vision-language models[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2022: 2763-2775.
- [52] LI D, LI J, LI H, et al. Align and prompt: video-and-language pre-training with entity prompts[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022: 4943-4953.
- [53] KIRILLOV A, MINTUN E, RAVI N, et al. Segment anything[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023: 4015-4026.
- [54] YU Y, YANG X, LI Q, et al. H2rbox-v2: boosting hbox-supervised oriented object detection via symmetric learning [J]. arXiv:2304.04403, 2023.
- [55] YANG J, GAO M, LI Z, et al. Track anything: segment anything meets videos[J]. arXiv:2304.11968, 2023.
- [56] WANG T, ZHANG J, FEI J, et al. Caption anything: interactive image description with diverse multimodal controls[J]. arXiv:2305.02677, 2023.
- [57] WANG X, ZHANG X, CAO Y, et al. SegGPT: segmenting everything in context[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023: 1130-1140.
- [58] JIANG Z, XU F F, ARAKI J, et al. How can we know what language models know?[J]. Transactions of the Association for Computational Linguistics, 2020, 8: 423-438.
- [59] ZHAO L, ZHENG F, ZENG W, et al. Domain-oriented prefix-tuning: towards efficient and generalizable fine-tuning for zero-shot dialogue summarization[C]//Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2022: 4848-4862.
- [60] JIAN Y, GAO C, VOSOUGH S. Contrastive learning for prompt-based few-shot language learners[C]//Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2022: 5577-5587.
- [61] SONG C, CAI F, WANG M, et al. TaxonPrompt: taxonomy-aware curriculum prompt learning for few-shot event classification[J]. Knowledge-Based Systems, 2023, 264: 110290.
- [62] YANG X, FENG S, WANG D, et al. Few-shot multimodal sentiment analysis based on multimodal probabilistic fusion prompts[C]//Proceedings of the 31st ACM International Conference on Multimedia, 2023: 6045-6053.
- [63] LI X L, LIANG P. Prefix-tuning: optimizing continuous prompts for generation[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: 4582-4597.
- [64] VU T, LESTER B, CONSTANT N, et al. SPoT: better frozen model adaptation through soft prompt transfer[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2022: 5039-5059.
- [65] ZHONG Z, FRIEDMAN D, CHEN D. Factual probing is [MASK]: learning vs. learning to recall[C]//Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021: 5017-5033.
- [66] WANG C, WANG J, QIU M, et al. TransPrompt: towards an automatic transferable prompting framework for few-shot text classification[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 2021: 2792-2802.
- [67] TANG T, LI J, ZHAO W X, et al. Context-tuning: learning contextualized prompts for natural language generation[C]//Proceedings of the 29th International Conference on Computational Linguistics, 2022: 6340-6354.

- [68] WU Z, WANG S, GU J, et al. IDPG: an instance-dependent prompt generation method[C]//Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2022: 5507-5521.
- [69] ZHANG Y, FEI H, LI D, et al. PromptGen: automatically generate prompts using generative models[C]//Findings of the Association for Computational Linguistics (NAACL 2022), 2022: 30-37.
- [70] GU Y, HAN X, LIU Z, et al. PPT: pre-trained prompt tuning for few-shot learning[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2022: 8410-8423.
- [71] HOU Y, DONG H, WANG X, et al. MetaPrompting: learning to learn better prompts[C]//Proceedings of the 29th International Conference on Computational Linguistics, 2022: 3251-3262.
- [72] ZHOU K, YANG J, LOY C C, et al. Conditional prompt learning for vision-language models[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 16816-16825.
- [73] YAO H, ZHANG R, XU C. Visual-language prompt tuning with knowledge-guided context optimization[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 6757-6767.
- [74] ZOU X, YANG J, ZHANG H, et al. Segment everything everywhere all at once[C]//Proceedings of the 37th International Conference on Neural Information Processing Systems, 2023: 19769-19782.
- [75] RAO Y, ZHAO W, CHEN G, et al. DenseCLIP: language-guided dense prediction with context-aware prompting[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022: 18061-18070.
- [76] SOHN K, CHANG H, LEZAMA J, et al. Visual prompt tuning for generative transfer learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 19840-19851.
- [77] HUANG S, GONG B, PAN Y, et al. VoP: text-video cooperative prompt tuning for cross-modal retrieval[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 6565-6574.
- [78] ZHU J, LAI S, CHEN X, et al. Visual prompt multi-modal tracking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 9516-9526.
- [79] RUAN C, WANG H. Dynamic visual prompt tuning for parameter efficient transfer learning[C]//Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision (PRCV), 2023: 293-303.
- [80] SHU M, NIE W, HUANG D A, et al. Test-time prompt tuning for zero-shot generalization in vision-language models[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems, 2022: 14274-14289.
- [81] MA X, ZHANG J, GUO S, et al. SwapPrompt: test-time prompt adaptation for vision-language models[C]//Proceedings of the 37th International Conference on Neural Information Processing Systems, 2023: 65252-65264.
- [82] LIU Z, SUN H, PENG Y, et al. DART: dual-modal adaptive online prompting and knowledge retention for test-time adaptation[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2024: 14106-14114.
- [83] HAN X, ZHAO W, DING N, et al. PTR: prompt tuning with rules for text classification[J]. AI Open, 2022, 3: 182-192.
- [84] CHEN X, ZHANG N, XIE X, et al. KnowPrompt: knowledge-aware prompt-tuning with synergistic optimization for relation extraction[C]//Proceedings of the ACM Web Conference, 2022: 2778-2788.
- [85] LI H, MO T, FAN H, et al. KiPT: knowledge-injected prompt tuning for event detection[C]//Proceedings of the 29th International Conference on Computational Linguistics, 2022: 1943-1952.
- [86] WANG J, HUANG W, SHI Q, et al. Knowledge prompting in pre-trained language model for natural language understanding[C]//Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 2022: 3164-3177.
- [87] YE H, ZHANG N, DENG S, et al. Ontology-enhanced prompt-tuning for few-shot learning[C]//Proceedings of the ACM Web Conference, 2022: 778-787.
- [88] CHEN X, LI L, ZHANG N, et al. Decoupling knowledge from memorization: Retrieval-augmented prompt learning [C]//Proceedings of the 36th International Conference on Neural Information Processing Systems, 2022: 23908-23922.
- [89] YAO Y, MAO S, ZHANG N, et al. Schema-aware reference as prompt improves data-efficient knowledge graph construction[C]//Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2023: 911-921.
- [90] ZHANG R, HU X, LI B, et al. Prompt, generate, then cache: cascade of foundation models makes strong few-shot learners[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 15211-15222.
- [91] CARON M, TOUVRON H, MISRA I, et al. Emerging properties in self-supervised vision transformers[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 9650-9660.
- [92] RAMESH A, PAVLOV M, GOH G, et al. Zero-shot text-to-image generation[C]//Proceedings of the International Conference on Machine Learning, 2021: 8821-8831.
- [93] YANG Z, GAN Z, WANG J, et al. An empirical study of GPT-3 for few-shot knowledge-based VQA[C]//Proceedings

- of the AAAI Conference on Artificial Intelligence, 2022: 3081-3089.
- [94] SHAO Z, YU Z, WANG M, et al. Prompting large language models with answer heuristics for knowledge-based visual question answering[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 14974-14983.
- [95] YANG J, ZHANG H, LI F, et al. Set-of-mark prompting unleashes extraordinary visual grounding in GPT-4V[J]. arXiv:2310.11441, 2023.
- [96] WANG X, WEI J, SCHUURMANS D, et al. Self-consistency improves chain of thought reasoning in language models [C]//Proceedings of the Eleventh International Conference on Learning Representations, 2023.
- [97] LI Y, LIN Z, ZHANG S, et al. Making language models better reasoners with step-aware verifier[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023: 5315-5333.
- [98] ZHOU D, SCHÄRLI N, HOU L, et al. Least-to-most prompting enables complex reasoning in large language models[C]//Proceedings of the Eleventh International Conference on Learning Representations, 2023.
- [99] DIAO S, WANG P, LIN Y, et al. Active prompting with chain-of-thought for large language models[J]. arXiv:2302.12246, 2023.
- [100] SHAO Z, GONG Y, SHEN Y, et al. Synthetic prompting: generating chain-of-thought demonstrations for large language models[C]//Proceedings of the 40th International Conference on Machine Learning, 2023: 30706-30775.
- [101] SHUM K, DIAO S, ZHANG T. Automatic prompt augmentation and selection with chain-of-thought from labeled data[C]//Findings of the Association for Computational Linguistics (EMNLP 2023), 2023: 12113-12139.
- [102] KOJIMA T, GU S S, REID M, et al. Large language models are zero-shot reasoners[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems, 2022: 22199-22213.
- [103] ZHANG Z, ZHANG A, LI M, et al. Automatic chain of thought prompting in large language models[C]//Proceedings of the Eleventh International Conference on Learning Representations, 2022.
- [104] DU Y, LI S, TORRALBA A, et al. Improving factuality and reasoning in language models through multiagent debate[C]//Proceedings of the 41st International Conference on Machine Learning, 2024.
- [105] ZHAO R, LI X, JOTY S, et al. Verify-and-edit: a knowledge-enhanced chain-of-thought framework[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023: 5823-5840.
- [106] LI X, ZHAO R, CHIA Y K, et al. Chain of knowledge: a framework for grounding large language models with structured knowledge bases[J]. arXiv:2305.13269, 2023.
- [107] WANG J, SUN Q, CHEN N, et al. Boosting language models reasoning with chain-of-knowledge prompting[J]. arXiv:2306.06427, 2023.
- [108] WANG K, DUAN F, WANG S, et al. Knowledge-driven CoT: exploring faithful reasoning in llms for knowledge-intensive question answering[J]. arXiv:2308.13259, 2023.
- [109] YAO S, ZHAO J, YU D, et al. ReAct: synergizing reasoning and acting in language models[C]//Proceedings of the Eleventh International Conference on Learning Representations, 2023.
- [110] LIU J, LIU A, LU X, et al. Generated knowledge prompting for commonsense reasoning[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2022: 3154-3169.
- [111] ZHANG Z, ZHANG A, LI M, et al. Multimodal chain-of-thought reasoning in language models[J]. arXiv:2302.00923, 2023.
- [112] HE L, LI Z, CAI X, et al. Multi-modal latent space learning for chain-of-thought reasoning in language models [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2024: 18180-18187.
- [113] YORAN O, WOLFSON T, BOGIN B, et al. Answering questions by meta-reasoning over multiple chains of thought[C]//Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, 2023: 5942-5966.
- [114] BESTA M, BLACH N, KUBICEK A, et al. Graph of thoughts: solving elaborate problems with large language models[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2024: 17682-17690.
- [115] SEL B, AL-TAWAHA A, KHATTAR V, et al. Algorithm of thoughts: enhancing exploration of ideas in large language models[C]//Proceedings of the 41st International Conference on Machine Learning, 2024.
- [116] NING X, LIN Z, ZHOU Z, et al. Skeleton-of-thought: large language models can do parallel decoding[J]. arXiv:2307.15337, 2023.
- [117] CHEN W, MA X, WANG X, et al. Program of thoughts prompting: disentangling computation from reasoning for numerical reasoning tasks[J]. Transactions on Machine Learning Research, 2023: 2835-8856.
- [118] BI Z, CHEN J, JIANG Y, et al. CodeKGC: code language model for generative knowledge graph construction[J]. arXiv:2304.09048, 2023.
- [119] WANG X, LI S, JI H. Code4Struct: code generation for few-shot structured prediction from natural language[J]. arXiv:2210.12810, 2022.
- [120] LI P, SUN T, TANG Q, et al. CodeIE: large code generation

- models are better few-shot information extractors[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023: 15339-15353.
- [121] CHEN M, TWOREK J, JUN H, et al. Evaluating large language models trained on code[J]. arXiv:2107.03374, 2021.
- [122] ZELIKMAN E, WU Y, MU J, et al. STaR: self-taught reasoner bootstrapping reasoning with reasoning[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems, 2022: 15476-15488.
- [123] SCHICK T, SCHMID H, SCHÜTZE H. Automatically identifying words that can serve as labels for few-shot text classification[C]//Proceedings of the 28th International Conference on Computational Linguistics, 2020: 5569-5578.
- [124] MA R, ZHOU X, GUI T, et al. Template-free prompt tuning for few-shot NER[C]//Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2022: 5721-5732.
- [125] WANG H, XU C, MCAULEY J. Automatic multi-label prompting: simple and interpretable few-shot classification [C]//Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2022: 5483-5492.
- [126] YU Z, GAO T, ZHANG Z, et al. Automatic label sequence generation for prompting sequence-to-sequence models [C]//Proceedings of the 29th International Conference on Computational Linguistics, 2022: 4965-4975.
- [127] ZHANG N, LI L, CHEN X, et al. Differentiable prompt makes pre-trained language models better few-shot learners[C]//Proceedings of the Tenth International Conference on Learning Representations(ICLR 2022), 2022.
- [128] CHEN X, LI L, DENG S, et al. LightNER: a lightweight tuning paradigm for low- resource NER via pluggable prompting[C]//Proceedings of the 29th International Conference on Computational Linguistics, 2022: 2374-2387.
- [129] CUI G, HU S, DING N, et al. Prototypical verbalizer for prompt-based few-shot tuning[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2022: 7014-7024.
- [130] HU S, DING N, WANG H, et al. Knowledgeable prompt-tuning: incorporating knowledge into prompt verbalizer for text classification[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2022: 2225-2240.
- [131] NI S, KAO H Y. KPT++: Refined knowledgeable prompt tuning for few-shot text classification[J]. Knowledge-Based Systems, 2023, 274: 110647.
- [132] ZHANG H, ZHANG X, HUANG H, et al. Prompt-based meta-learning for few-shot text classification[C]//Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 2022: 1342-1357.
- [133] SOCHER R, PERELYGIN A, WU J, et al. Recursive deep models for semantic compositionality over a sentiment treebank[C]//Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, 2013: 1631-1642.
- [134] PANG B, LEE L. Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales[C]//Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics, 2005: 115-124.
- [135] HU M, LIU B. Mining and summarizing customer reviews [C]//Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2004: 168-177.
- [136] 李冬梅, 张扬, 李东远, 等. 实体关系抽取方法研究综述 [J]. 计算机研究与发展, 2020, 57(7): 1424-1448.
- LI D M, ZHANG Y, LI D Y, et al. Review of entity relation extraction methods[J]. Journal of Computer Research and Development, 2020, 57(7): 1424-1448.
- [137] HENDRICKX I, KIM S N, KOZAREVA Z, et al. SemEval-2010 task 8: multi-way classification of semantic relations between pairs of nominals[C]//Proceedings of the 5th International Workshop on Semantic Evaluation, 2010: 33-38.
- [138] ZHANG Y, ZHONG V, CHEN D, et al. Position-aware attention and supervised data improve slot filling[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2017: 35-45.
- [139] ALT C, GABRYSZAK A, HENNIG L. TACRED-Revisited: a thorough evaluation of the tacred relation extraction task [C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020: 1558-1569.
- [140] STOICA G, PLATANIOS E A, PÓCZOS B. Re-TACRED: addressing shortcomings of the tacred dataset[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2021: 13843-13850.
- [141] HAN X, ZHU H, YU P, et al. FewRel: a large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018: 4803-4809.
- [142] GAO T, HAN X, ZHU H, et al. FewRel 2.0: towards more challenging few-shot relation 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: 6250-6255.
- [143] CHEN C Y, LI C T. ZS-BERT: towards zero-shot relation

- extraction with attribute representation learning[C]//Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021: 3470-3479.
- [144] BOWMAN S, ANGELI G, POTTS C, et al. A large annotated corpus for learning natural language inference[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015: 632-642.
- [145] WILLIAMS A, NANGIA N, BOWMAN S. A broad-coverage challenge corpus for sentence understanding through inference[C]//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), 2018: 1112-1122.
- [146] WANG A, SINGH A, MICHAEL J, et al. GLUE: a multi-task benchmark and analysis platform for natural language understanding[C]//Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, 2018: 353-355.
- [147] HU H, RICHARDSON K, XU L, et al. OCNLI: original chinese natural language inference[C]//Findings of the Association for Computational Linguistics (EMNLP 2020), 2020: 3512-3526.
- [148] RAJPURKAR P, ZHANG J, LOPYREV K, et al. SQuAD: 100,000+ questions for machine comprehension of text [C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016: 2383-2392.
- [149] BRAGG J, COHAN A, LO K, et al. Flex: unifying evaluation for few-shot NLP[C]//Proceedings of the 35th International Conference on Neural Information Processing Systems, 2021: 15787-15800.
- [150] MIN S, LEWIS M, HAJISHIRZI H, et al. Noisy channel language model prompting for few-shot text classification [C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2022: 5316-5330.
- [151] ZHU Y, ZHOU X, QIANG J, et al. Prompt-learning for short text classification[J]. IEEE Transactions on Knowledge & Data Engineering, 2024, 36(10): 5328-5339.
- [152] LEE D H, KADAKIA A, TAN K, et al. Good examples make a faster learner: simple demonstration-based learning for low-resource ner[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2022: 2687-2700.
- [153] HUANG Y, HE K, WANG Y, et al. COPNER: contrastive learning with prompt guiding for few-shot named entity recognition [C]//Proceedings of the 29th International Conference on Computational Linguistics, 2022: 2515-2527.
- [154] WANG Y, XU C, SUN Q, et al. PromDA: prompt-based data augmentation for low-resource NLU tasks[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2022: 4242-4255.
- [155] JIANG P, AGARWAL S, JIN B, et al. Text-augmented open knowledge graph completion via pre-trained language models[C]//Findings of the Association for Computational Linguistics (ACL 2023), 2023: 11161-11180.
- [156] TAN Z, ZHANG X, WANG S, et al. MSP: multi-stage prompting for making pre-trained language models better translators[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2022: 6131-6142.
- [157] LI Y, YIN Y, LI J, et al. Prompt-driven neural machine translation[C]//Findings of the Association for Computational Linguistics (ACL 2022), 2022: 2579-2590.
- [158] GUO H, LIU J, HUANG H, et al. LVP-M3: language-aware visual prompt for multilingual multimodal machine translation[C]//Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 2022: 2862-2872.
- [159] ZHANG B, HADDOW B, BIRCH A. Prompting large language model for machine translation: a case study[C]//Proceedings of the 40th International Conference on Machine Learning, 2023: 41092-41110.
- [160] ZHANG X, ZHAO J, LECUN Y. Character-level convolutional networks for text classification[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems, 2015: 649-657.
- [161] MCAULEY J, LESKOVEC J. Hidden factors and hidden topics: understanding rating dimensions with review text [C]//Proceedings of the 7th ACM Conference on Recommender Systems, 2013: 165-172.
- [162] LEHMANN J, ISELE R, JAKOB M, et al. DBpedia—a large-scale, multilingual knowledge base extracted from wikipedia[J]. Semantic Web, 2015, 6(2): 167-195.
- [163] SANG E T K, DE MEULDER F. Introduction to the CoNLL-2003 shared task: language-independent named entity recognition[C]//Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003, 2003: 142-147.
- [164] PRADHAN S, MOSCHITTI A, XUE N, et al. Towards robust linguistic analysis using ontonotes[C]//Proceedings of the Seventeenth Conference on Computational Natural Language Learning, 2013: 143-152.
- [165] LIU J, PASUPAT P, WANG Y, et al. Query understanding enhanced by hierarchical parsing structures[C]//Proceedings of the 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, 2013: 72-77.
- [166] DING N, XU G, CHEN Y, et al. Few-NERD: a few-shot named entity recognition dataset[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: 3198-3213.
- [167] TOUTANOVA K, CHEN D, PANTEL P, et al. Representing text for joint embedding of text and knowledge bases[C]// Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015: 1499-1509.
- [168] DETTMERS T, MINERVINI P, STENETORP P, et al. Convolutional 2D knowledge graph embeddings[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2018: 1811-1818.
- [169] PAPINENI K, ROUKOS S, WARD T, et al. BLEU: a method for automatic evaluation of machine translation[C]// Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, 2002: 311-318.
- [170] REI R, STEWART C, FARINHA A C, et al. COMET: a neural framework for MT evaluation[C]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020: 2685-2702.
- [171] FARHADI A, ENDRES I, HOIEM D, et al. Describing objects by their attributes[C]// Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009: 1778-1785.
- [172] DENG J, DONG W, SOCHER R, et al. ImageNet: a large-scale hierarchical image database[C]// Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009: 248-255.
- [173] XIAO J, HAYS J, EHINGER K A, et al. SUN database: large-scale scene recognition from abbey to zoo[C]// Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010: 3485-3492.
- [174] FEI-FEI L, FERGUS R, PERONA P. Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories[C]// Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop, 2004: 178.
- [175] YU L, POIRSON P, YANG S, et al. Modeling context in referring expressions[C]// Proceedings of the 14th European Conference on Computer Vision (ECCV 2016), Amsterdam, The Netherlands, October 11-14, 2016. Cham: Springer International Publishing, 2016: 69-85.
- [176] MAO J, HUANG J, TOSHEV A, et al. Generation and comprehension of unambiguous object descriptions[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 11-20.
- [177] GOYAL Y, KHOT T, SUMMERS-STAY D, et al. Making the V in VQA matter: elevating the role of image understanding in visual question answering[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6325-6334.
- [178] ZHU Y, GROTH O, BERNSTEIN M, et al. Visual7W: grounded question answering in images[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 4995-5004.
- [179] ELLIOTT D, FRANK S, SIMA'AN K, et al. Multi30K: multilingual english-german image descriptions[C]// Proceedings of the 5th Workshop on Vision and Language, 2016: 70-74.
- [180] WU Z, SONG S, KHOSLA A, et al. 3D shapenets: a deep representation for volumetric shapes[C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 1912-1920.
- [181] UY M A C, PHAM Q H, HUA B S, et al. Revisiting point cloud classification: a new benchmark dataset and classification model on real-world data[C]// Proceedings: 2019 International Conference on Computer Vision, 2019.
- [182] MARINO K, RASTEGARI M, FARHADIA, et al. Ok-VQA: a visual question answering benchmark requiring external knowledge[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 3195-3204.
- [183] AGRAWAL H, DESAI K, WANG Y, et al. Nocaps: novel object captioning at scale[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 8948-8957.
- [184] YOUNG P, LAI A, HODOSH M, et al. From image descriptions to visual denotations: new similarity metrics for semantic inference over event descriptions[J]. Transactions of the Association for Computational Linguistics, 2014, 2: 67-78.
- [185] VINIYALS O, BLUNDELL C, LILICRAP T, et al. Matching networks for one shot learning[C]// Proceedings of the 30th International Conference on Neural Information Processing Systems, 2016: 3637-3645.
- [186] BAIN M, NAGRANI A, VAROL G, et al. Frozen in time: a joint video and image encoder for end-to-end retrieval [C]// Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 1708-1718.
- [187] XU J, MEI T, YAO T, et al. MSR-VTT: a large video description dataset for bridging video and language[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 5288-5296.
- [188] HENDRICKS L A, WANG O, SHECHTMAN E, et al. Localizing moments in video with natural language[C]// Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), 2017: 5804-5813.
- [189] XU D, ZHAO Z, XIAO J, et al. Video question answering via gradually refined attention over appearance and motion

- [C]//Proceedings of the 25th ACM International Conference on Multimedia, 2017: 1645-1653.
- [190] NIU T, ZHU S, PANG L, et al. Sentiment analysis on multi-view social data[C]//Proceedings of the 22nd International Conference on Multi-Media Modeling (MMM 2016), 2016: 15-27.
- [191] YANG X, FENG S, WANG D, et al. Image-text multimodal emotion classification via multi-view attentional network[J]. IEEE Transactions on Multimedia, 2020, 23: 4014-4026.
- [192] NILSBACK M E, ZISSERMAN A. Automated flower classification over a large number of classes[C]//Proceedings of the 2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing, 2008: 722-729.
- [193] PATRICK H, BENJAMIN B, ANDREAS D, et al. EuroSAT: a novel dataset and deep learning benchmark for land use and land cover classification[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019, 12(7): 2217-2226.
- [194] ZHOU B, ZHAO H, PUIG X, et al. Semantic understanding of scenes through the ADE20k dataset[J]. International Journal of Computer Vision, 2019, 127: 302-321.
- [195] VAN HORN G, BRANSON S, FARRELL R, et al. Building a bird recognition app and large scale dataset with citizen scientists: the fine print in fine-grained dataset collection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 595-604.
- [196] GEBRU T, KRAUSE J, WANG Y, et al. Fine-grained car detection for visual census estimation[C]//Proceedings of the 31st AAAI Conference on Artificial Intelligence, 2017: 4502-4508.
- [197] CABA HEILBRON F, ESCORCIA V, GHANEM B, et al. Activitynet: a large-scale video benchmark for human activity understanding[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 961-970.
- [198] HENDRYCKS D, ZHAO K, BASART S, et al. Natural adversarial examples[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 15262-15271.
- [199] HENDRYCKS D, BASART S, MU N, et al. The many faces of robustness: a critical analysis of out-of-distribution generalization[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 8340-8349.
- [200] SCHWENK D, KHANDELWAL A, CLARK C, et al. A-OKVQA: a benchmark for visual question answering using world knowledge[C]//Proceedings of the European Conference on Computer Vision. Cham: Springer Nature, 2022: 146-162.
- [201] LU P, MISHRA S, XIA T, et al. Learn to explain: multi-modal reasoning via thought chains for science question answering[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems, 2022: 2507-2521.
- [202] BAR A, GANDELSMAN Y, DARRELL T, et al. Visual prompting via image inpainting[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems, 2022: 25005-25017.
- [203] HE X, YANG D, FENG W, et al. CPL: counterfactual prompt learning for vision and language models[C]//Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 2022: 3407-3418.
- [204] DU Y, WEI F, ZHANG Z, et al. Learning to prompt for open-vocabulary object detection with vision-language model[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 14084-14093.
- [205] WEBSON A, PAVLICK E. Do prompt-based models really understand the meaning of their prompts?[C]//Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2022: 2300-2344.
- [206] ZHAO C, WANG Y, JIANG X, et al. Learning domain invariant prompt for vision-language models[J]. IEEE Transactions on Image Processing, 2024, 33: 1348-1360.
- [207] FANG L, KUANG Y, LIU Q, et al. Rethinking remote sensing pretrained model: instance-aware visual prompting for remote sensing scene classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-13.
- [208] TSIMPOUKELLI M, MENICK J L, CABI S, et al. Multi-modal few-shot learning with frozen language models[C]//Proceedings of the 35th International Conference on Neural Information Processing Systems, 2021: 200-212.
- [209] YANG Z, WANG S, RAWAT B P S, et al. Knowledge injected prompt based fine-tuning for multi-label few-shot ICD coding[C]//Findings of the Association for Computational Linguistics (EMNLP 2022), 2022: 1767-1781.
- [210] CODA-FORNO J, WITTE K, JAGADISH A K, et al. Inducing anxiety in large language models increases exploration and bias[J]. arXiv:2304.11111, 2023.