

文章编号: 1003-0077(2023)07-0001-12

与知识相结合的提示学习研究综述

鲍琛龙, 吕明阳, 唐晋韬, 李莎莎, 王 挺

(国防科技大学 计算机学院, 湖南 长沙 410073)

摘 要: 近年来, 提示学习方法由于可以充分激发预训练语言模型的潜能而得到了研究者越来越多的关注, 特别是在知识抽取任务中取得了较好进展。为了提升提示学习性能, 研究者也开展了基于知识的提示学习模板工程、答案工程优化等多项研究。该文对提示学习与知识相结合的相关研究进行了系统综述, 包括知识抽取中的提示学习方法以及基于知识约束的提示学习相关进展。在此基础上, 该文还探讨了目前方法存在的局限性, 展望了提示学习与知识相结合的发展趋势。

关键词: 提示学习; 知识抽取; 知识提示

中图分类号: TP391

文献标识码: A

A Survey of Prompt Learning Combined with Knowledge

BAO Chenlong, LYU Mingyang, TANG Jintao, LI Shasha, WANG Ting

(College of Computer Science and Technology, National University of Defense Technology,

Changsha, Hunan 410073, China)

Abstract: In recent years, prompt learning methods have attracted more attention from researchers for exploiting the pre-trained language models. To optimize the performance of prompt learning, researchers explored the template engineering and the answer engineering which are both based on knowledge. In this paper, the researches on prompt learning combined with knowledge are reviewed systematically, with a focus on the prompt learning methods in knowledge extraction. We also reveal the constraints in these method and discuss the possible future developments.

Keywords: prompt learning; knowledge extraction; knowledge-based prompt

0 引言

知识抽取, 是通过识别、理解、筛选、格式化, 把文本中的各个知识点抽取出来, 以一定形式存入知识库中的过程^[1], 是构建知识图谱的关键一环, 在自然语言处理领域备受瞩目, 其方法也随着自然语言处理方法的发展而发展。早期的自然语言的处理方法大多是完全监督学习方法, 这类方法严重依赖特征工程^[2-3]; 随着神经网络的出现, 自然语言处理出现了新范式——基于神经网络的深度学习^[4-5], 虽然不需要手动设置特征, 但仍需要精心设计最适配下游任务的网络架构。近几年, 随着预训练语言模型 (Pre-trained Language Model, PLM), 如 GPT^[6]、BERT^[7]、BART^[8]等的出现, “预训练”+“微调”成

为了自然语言处理的标准范式, 并以统一的模型架构显著提升了大量任务的性能^[9]。然而, 在预训练和微调时, 不同阶段的优化目标不一致, 在预训练语言模型去适配任务的过程中, 会给模型的性能带来损失; 并且预训练语言模型的体量越来越大、训练越来越繁琐, 导致很多时候无法对所使用的预训练语言模型进行“微调”。因此, 预训练模型存在适配下游任务造成性能损失与较难“微调”的问题, 需要探索出一种更小巧轻量、更普适高效的方法。

因为提示学习可以充分激发预训练语言模型的潜能, 所以成为当下研究热点。提示学习是指对输入文本信息按照特定模板进行转换, 把任务重构成一个更能充分利用预训练语言模型处理的形式。在构造提示模板和标签映射的过程中, “提示学习”将预训练和微调的优化目标进行统一, 充

收稿日期: 2022-06-13 定稿日期: 2022-08-30

基金项目: 科技创新 2030-“新一代人工智能”重大项目(2021ZD0110700); 湖南省自然科学基金(2022JJ30668)

分激发预训练语言模型的潜力,提升了任务的性能。表 1 展示了不同任务的示例。对于不同任务,通过设计特定的模板,无论是分类任务、序列标注任务还是生成任务,均可以被重构为掩码语

言模型任务,这使得下游任务与预训练语言模型无需借助“微调”的方式来适配。提示学习拓展了预训练语言模型的适用性,并且打破了不同任务之间统一建模的困难。

表 1 不同任务的提示学习示例

类型	任务	输入	模板	答案
分类	情感分类	今天天气真好。	[X]我的心情是[Z]。	开心、沮丧……
	主题分类	今天是晴天。	[X]这段文字是关于[Z]的。	天气、运动……
	意图分类	去北京的车票多少钱?	[X]这个问题是关于[Z]的。	价格、城市……
序列标注	命名实体识别	[X1]: 北京是中国首都。 [X2]: 北京	[X1][X2]是[Z]实体。	地点、机构……
文本生成	文本摘要	据新华社报道……	[X]: [Z]	受害人是……
	机器翻译	我爱你	中文: [X]英文: [Z]	I love you. I love you. ……

目前提示学习相关工作尚处于初期探索阶段,还面临巨大挑战,Liu 等人^[10]已指出提示模板设计是影响提示学习效果的一个重要问题,并指出与知识相结合可能是解决该问题的一条路径。近两年间,提示学习与知识抽取相结合的方向已经逐步成为新的研究热点,基于提示学习的知识抽取在各具体任务中取得了领先水平,在提示学习中引入知识的相关研究也崭露头角。

提示学习的动机是帮助预训练语言模型回忆“知识”;对于知识抽取任务而言,其目的是抽取知

识,与此同时,利用抽取的知识也能提升提示学习的效果,二者相辅相成。因此,开展提示学习与知识相结合的研究将成为一个重要趋势。对该领域进行全面总结和系统梳理,将有助于研究者了解相关研究前沿和主要挑战。区别于 Liu 等人^[10]对整个提示学习从基本概念到各种任务应用的系统综述,本文主要关注基于提示学习的知识抽取任务和基于知识的提示学习相关文献。因此,本文在归纳整理了表 2 所列文献的基础上,尝试归纳总结拓展该方向的概念与方法,旨在为感兴趣的研究人员提供参考。

表 2 提示学习与知识相结合的任务/方法分类

任务/方法	核心做法	优缺点	相关文献
知识抽取之 实体抽取	提示分解,通过掩码(mask)转换为完型填空任务	存在错误传播;构造提示代价高	Ding 等人 ^[11] ;Cui 等人 ^[12]
	前缀提示,在输入文本前插入可训练向量	轻量高效	Zhou 等人 ^[13] ;Chen 等人 ^[14]
	提示增强,构造演示来挖掘模型能力	存在演示的排序和选择问题	Lee 等人 ^[15] ;Ma 等人 ^[16]
知识抽取之 关系抽取	标记提示,标记实体跨度和类型	方法简单有效	Zhou 等人 ^[17]
	人工提示,利用人工设计的模板,在答案工程部分进行优化	手工设计的模板限制了模型的性能,小样本场景下表现欠佳	Chen 等人 ^[18] ;Chen 等人 ^[19] ;Chia 等人 ^[20]
	连续提示,利用连续模板和连续答案搜索	在小样本场景下表现较好	Zhang 等人 ^[21]
知识抽取之 事件抽取	要素抽取,利用模板将任务转换为完型填空任务	提升了预测性能	Ye 等人 ^[22]
	流水线模式事件抽取,将抽取任务转换为生成任务	转换为生成任务更好地利用了预训练语言模型	Hsu 等人 ^[23] ;Si 等人 ^[24]
	提示融合事件抽取,在事件检测和事件要素提取过程中利用多种提示	多提示的互补有时使模型更好地理解答案和上下文的联系	Lin 等人 ^[25]

续表

任务/方法	核心做法	优缺点	相关文献
知识约束	在构建提示、标签映射过程中注入知识	通过知识约束降低了搜索空间	Han 等人 ^[26] ; Hu 等人 ^[27] ; Chen 等人 ^[28] ; Ye 等人 ^[29]
选择性知识注入	使用可见矩阵缓解知识注入的可能负面影响	解决知识噪音和异构性问题	Ye 等人 ^[30]

1 提示学习基本概念

本节对提示学习中涉及的概念进行阐述,以 Liu 等人^[10]的规范为主要参考。提示学习的核心是重构下游任务,例如情感分类任务,原有任务的输入为“今天的天气真好”,输出则是“正面情绪”类别。如果运用提示学习,输入通过提示函数被改写为“今天的天气真好,我的心情是[MASK]的。”,输出则是“开心”,然后通过标签词映射,将“开心”识别为“正面情绪”标签。

具体而言,在提示学习中,需要以下三个步骤:

① 模板构建:用提示函数 $f_{\text{prompt}}(x)$ 将输入的文本 x 修改为提示 $x' = f_{\text{prompt}}(x)$;

② 答案搜索:将 x' 传给预训练模型,得到“最高分”的输出;

③ 标签词映射:在第二步中得到的输出是满足预训练任务形式的输出,可能会与下游任务的输出形式存在差别,需要一个函数映射来进行转换。

模板构建步骤通常被称为提示工程,而标签词映射通常被称为答案工程。

1.1 提示工程

提示工程(也可称为模板工程)是创建提示函数 $f_{\text{prompt}}(x)$ 的过程,这会影响到下游任务的表现。对于提示工程而言,首先要考虑提示的模式,然后再考虑采用人工或者自动化的方式创建所需模式的提示。

提示的模式主要分为完型填空式提示和前缀提示两种。完型填空式提示就是表 1 中的模板给出的示例,用于填充文本字符串中的空白部分;而前缀提示则是在输入文本前添加一组可训练的向量作为提示^[31],训练时固定预训练语言模型中的其余参数,对这些向量进行优化;在最新研究中,前缀提示的概念也在不断扩充,在特定字符串周围使用一些特殊的标记来进行标注作为提示,这种方法也纳入了前缀提示的范畴,例如 @ Bill @ was born in # Seattle #.,其中 Bill 实体和 Seattle 实体周围都有特

殊标记。

传统的提示工程主要依靠人工构建模板^[11-12,18],过于依赖专家知识;同时研究者还探索了自动构建提示模板的方法,主要包括离散提示^[32]和连续提示^[21]两种。离散提示是指让计算机自行搜索添加文本字符串;考虑到构建提示的目的是找到让语言模型能够有效执行任务的方式,而不是供人类阅读或使用,没有必要将提示限制为人类可读的文本,因此衍生出了由连续向量构成的连续提示,避免了语言模型参数对模板的限制;此外,还有人工设计和自动生成相结合的混合提示,采取在人工设计的提示中插入一些可训练的向量^[33]。在上述自动构建提示的探索中,研究者们已经发现在提示中引入知识,对构建过程进行约束能够取得更好的效果^[28]。

1.2 答案工程

答案工程旨在设计合适映射,即从预训练语言模型输出到答案空间中所对应标签词的过程,这同样需要考虑模式和设计方法。

答案的模式有标记^[34](Token)、跨度^[35](Span)、句子^[31](Sentence)这三种。在实践中,答案模式的选择取决于执行的任务,例如,标记、跨度模式可以用于实体抽取和关系抽取任务,跨度、句子模式可以用于文本生成任务。

答案模式设计方法分为手动设计、离散答案搜索、连续答案搜索这三种。手动设计即人为定义映射函数^[35];离散答案搜索常用的方法有:①扩充原答案空间,使用回译的方法,将答案翻译成另一种语言,然后返回生成多个释义答案的列表^[36];②先剪枝再进行搜索,先利用模型生成可能的答案,组成初始答案空间,即对所有答案的搜索空间进行剪枝,然后再使用一个具体的算法在这个剪枝后的空间内进一步搜索以选择最终答案^[37];③标签分解^[14],将标签分解为组成的词,并将这些词作为答案。目前,连续答案搜索在具体工作中运用较少,使用梯度下降的方式能够优化可变长答案标记^[21]。

1.3 多提示学习

多提示学习是指在模板构建的过程中使用多个提示,这能进一步增强提示学习的效果。常用的多提示学习有提示融合^[25]、提示增强^[15]、提示合成^[26]、提示分解^[12]这几种方式。

提示融合是在预测时使用多个未回答的提示作为输入,最终使用均值、加权平均、投票等多种方案得到输出。这种提示融合的方式可以很好利用不同提示之间的互补优势,降低搜索出表现最好提示的成本,避免不同提示的差异对模型表现产生的影响,稳定下游任务的性能。

提示增强又可以称作演示学习,通过为语言模型提供带有答案的示例作为提示,相当于一种数据增强,在一定程度上能解决小样本问题^[38]。例如,给模型输入“湖南的省会是[Z]”作为提示时,在提示前面加上“江苏的省会是南京”等示例;这种方式利用了预训练模型学习重复模式的能力,可以很好处理小样本问题,但存在示例选择和示例排序^[39]问题,后文中会探讨示例选择的解决方案^[15]。

提示合成是指在处理一些可以分解成更基础的子任务时,使用多个子提示,每个子提示对应一个子任务,然后将这些子提示组合为一个完整的提示。这种方法常用于关系抽取任务中,利用子提示处理实体识别和关系分类,然后根据标签映射完成关系抽取任务。

提示分解是将完整的提示分解为不同的子提示,然后分别对每个子提示进行预测。这种方法往往用于序列标注任务,例如命名实体识别任务。在输入中可能存在多个实体,同时预测出所有实体的类型极其困难,因此需要为可能是实体的跨度创建子提示,分别对每个跨度进行预测。

1.4 提示学习的应用

提示学习在自然语言处理的各类任务中应用广泛,如 Wang 等人^[40]利用提示学习重新建模了对话和推荐任务,用一个预训练语言模型就可以实现两个任务的统一;Rajagopal 等人^[41]利用提示学习,引入特定域的概念,实现了序列到序列模型的跨域推理;Zhao 等人^[42]利用提示学习在一定程度上解决了对话摘要任务中的领域迁移问题,减少了对大量标注数据的依赖。此外,提示学习还运用于机器翻译^[43]、语义解析^[44]等任务。

2 基于提示学习的知识抽取

知识抽取的主要任务有实体抽取、关系抽取和事件抽取,这些任务在具体场景中会受到不同目标、异构结构和特定需求的影响,存在迁移性弱、域适应受限、重新训练代价高的问题。通过提示学习,可以进一步挖掘大规模预训练语言模型的能力。实际上,因为知识抽取任务的形式与完型填空任务较为接近,提示学习的方法可以有效指导模型自适应地完成抽取任务,减轻人工标注大量数据的负担,提升模型在小样本与领域迁移场景下的表现。

2.1 基于提示学习的实体抽取

传统的实体抽取采用序列标注的方式^[45],而基于提示学习的方法则将其通过掩码很自然地表述为完型填空问题,能更充分地利用预训练语言模型蕴含的知识,在实验中取得不错的效果。

基于提示学习方法进行实体抽取的基本模型框架如图 1 所示,将输入文本和提示输入预训练语言模型,再通过标签词映射将[MASK]预测为城市实体。通过完型填空式提示的方式实现对实体的预测,并将实体发现、实体分类两个子任务分解为两个子提示,但提示分解方法跨度检测时会受错误传播影响,且候选跨度过多增加了构造提示的代价。研究者们尝试采用前缀提示或提示增强的方式来解决这个问题。

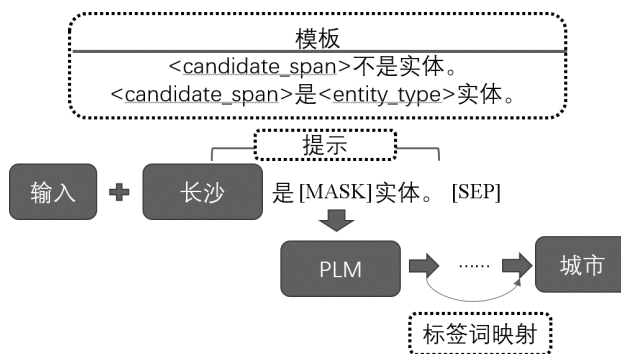


图 1 基于提示学习的实体抽取基本模型框架

2.1.1 基于提示分解的实体抽取方法

在实体抽取任务中,对提示学习最早的探索来自于 Ding 等人^[11]将提示学习用于实体分类任务的研究。提出 PLET 框架来处理细粒度的实体分类问题,通过人工创建了三种模板对实体分类进行建模,并提出了带有特殊标记的模板,证明了不同长

度的模板会影响提示学习的效果。PLET 框架使用的模板如图 2 所示,图 2 左半部分给出了 3 个定长

模板和一个带有特殊标记的不定长模板,右半部分解释了图中符号的含义,并阐述了模板的设置效果。

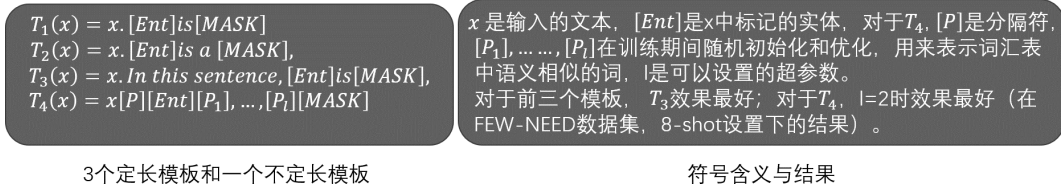


图 2 PLET 的几种模板

TemplateNER^[12] 最先将提示学习应用于序列标注任务,采用提示分解的方式,将输入转化为文本跨度,然后利用 BART 模型^[8]为每一个子提示进行预测打分。在枚举文本跨度时,为保证效率,设定了跨度上限为 8,但为句子中的不同跨度生成提示仍会带来巨大的开销,而且这种方式在预测同一句子中不同跨度的实体时忽略了实体间的相互依赖性。

2.1.2 基于前缀提示的实体抽取方法

为了解决枚举跨度的问题,Zhou 等人^[13]提出了插件标记器(plug-tagger),利用前缀提示来使用预训练语言模型解决序列标注问题。在输入文本前添加一系列可训练的向量,输入变为 $X' = [\theta_x; x'_1, \dots, x'_n]$,这里的 θ_x 就是前缀提示, x'_n 是原输入。为了增强这些向量的影响,这些向量还被添加到模型的每一层中作为额外输入,在训练时,冻结插入向量之外的参数。

LightNER^[14] 同样也是使用前缀提示解决实体抽取问题的探索,在编码和解码层都加入了可调节的参数作为前缀提示,同样冻结其余参数对模型进行微调,这样就不需要通过枚举所有可能的跨度和构建模板的方式来促使预训练模型抽取实体。

2.1.3 基于提示增强的实体抽取方法

基于提示增强的实体抽取方法通过构造任务演示让模型学习上下文,不需要使用完形填空式提示进行实体抽取,避免了利用提示分解进行抽取产生的问题。

Lee 等人^[15]提出的提示增强方法沿用了传统序列标注模式,直接将任务演示拼接到原输入后作为输入。该方法的关键是选择合适的实体示例,以两种不同的方式构建任务演示:

① 面向实体类型的演示,从训练数据中为每个实体类型选择(三种选择策略:随机选择、频率最高、搜索策略)一个实体示例来构建演示,使得模型通过学习其实体示例来更好地了解实体类型;

② 面向实例的演示,在训练数据中检索(SBERT^[46]

或 BERTScore^[47]方法)与输入句子相似的实例示例,使得模型通过学习相似的实例及其实体来更好地了解任务。

其中,面向实体类型的演示优于面向实例的演示,且搜索策略的表现最好。

类似于面向实体类型的演示构建,Ma 等人^[16]提出 EntLM,但无须拼接演示,直接通过预训练语言模型检索实体类型来挖掘模型能力,大幅提升模型的效率。

2.2 基于提示学习的关系抽取

对于关系抽取任务,提示学习通过引入前缀标记提示,让模型更多地关注句子中的实体信息,充分激发模型的性能。

2.2.1 基于标记提示的关系抽取方法

Zhou 和 Chen^[17]采用类型化实体标记的方式,先标记实体跨度和实体类型,然后将处理后的句子输入模型获得上下文输入,最后使用分类器完成任务,这种方法简单却能取得较好结果。图 3 介绍了 5 种标记方法并给出样例,展示了不同方法之间的区别。

2.2.2 基于人工提示的关系抽取方法

AdaPrompt^[18]同样使用标记提示,在实体两侧加入标签来强调实体,与上文工作不同的是,使用离散模板和标签词映射将任务转化为掩码语言模型任务。

模板主要依赖专家知识进行人工设计,例如,“[实体 1]是[实体 2]的[MASK]”或者“[实体 1]和[实体 2]之间的关系是[MASK]”。在实际输入时,模型分别掩码关系词和实体名称来构造两个提示,这样能更充分使用储存在预训练模型中的知识。

在答案工程部分,采用自适应标签选择机制进行映射。通过将关系标签分解,仅留下名词并补全属性词,将搜索空间缩减到词汇表,从而可以处理多关系标签的映射。

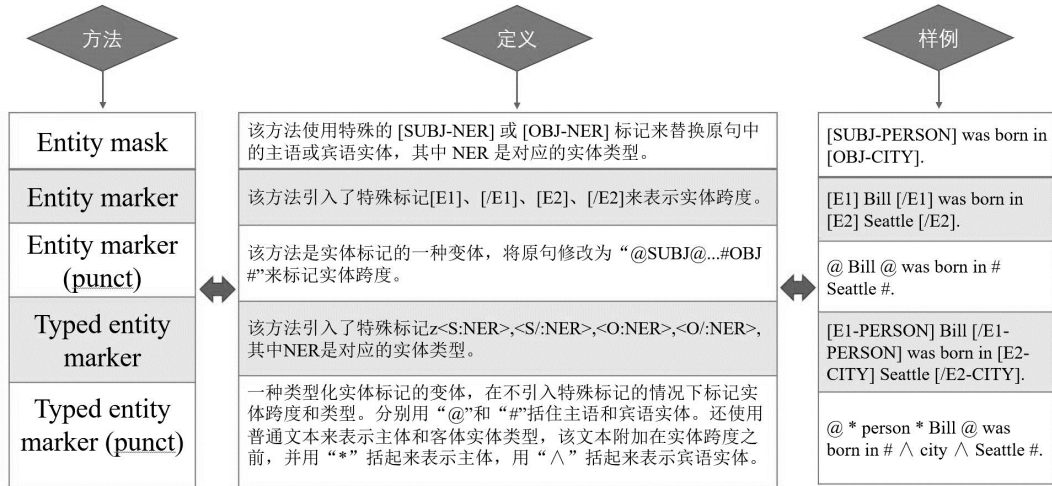


图3 类型化实体标记

Chen 等人^[19]在提示工程部分的做法与 AdaPrompt 做法相似，但对答案工程进行优化，在标签词映射时引入从训练集检索的关系实例辅助决策，从而提高模型的鲁棒性和通用性。

RelationPrompt^[20]使用人工设计的模板，通过一个关系生成模型和关系抽取模型为零样本关系抽取任务提供新的研究思路。具体而言，利用关系名称当作语言模型的提示，生成未见过的关系类别的新句子样本，类似于远程监督方法生成数据^[48]，然后利用这些伪数据去训练模型完成抽取。但由于手工设计的提示过于简单，存在生成数据质量不高、抽取语义关系不符的情况。

2.2.3 基于连续提示的关系抽取方法

使用提示标记和离散答案搜索，AdaPrompt 可在数据充足的情况下取得理想效果；但手工设计的模板依赖专家知识且无法应对小样本情况，因此研究者们尝试使用连续模板和连续答案搜索挖掘模型的性能，提升在小样本情况下的表现。

DART 方法^[21]使用一类特殊的非语义标记作为模板和标签词映射，并使用反向传播在连续空间内进行优化。模型仍以 BERT 为基础，优化目标是关系分类。

对于模板 $T = \{[T_{0:i}], [\text{MASK}][T_{i+1:j}]\}$ ，不同于离散模板的语义标记嵌入表示 $\{\omega([T_{0:i}]), \omega([\text{MASK}]), \omega([T_{i+1:j}])\}$ ，DART 采用可训练的参数替换语义标记，模板变为 $\{h_0, \dots, h_i, \omega([\text{MASK}]), h_{i+1}, \dots, h_m\}$ ，这些可训练参数 h_i 能在训练过程中学习到最优的表示效果；对于标签映射，DART 不同于 AdaPrompt 分解关系标签的方法，而是采取连续映射，即 $M(Y_j) = \{h_{m+j}\}$ ；另外，

为了减少外部参数的使用，来增强方法的迁移性， $\{h_1, \dots, h_m, \dots, h_{m+n}\}$ 作为嵌入参数，是语言模型中未使用的标记或词表中的特殊标记。

2.3 基于提示学习的事件抽取

从自然语言文本中抽取事件实例是一个关键又十分具有挑战性的工作。目前对事件抽取的任务研究大多数都遵守 ACE 2005 评测会议^[49]中的规范进行抽取。

目前主流的事件抽取方法分为基于序列标注的方法^[50-51]、基于问答的方法^[52-54]和基于生成的方法 (TANL^[55]，Text2Event^[56])，这些方法依赖大量的训练数据或精心设计的高质量问题，无法适应事件抽取的实际需要，研究者们希望通过引入提示学习，减轻数据处理的负担，完成在小样本情况下的事件抽取任务。

当前，基于提示学习的事件抽取主要有以下研究方向：面向要素抽取的提示学习、面向流水线模式事件抽取的提示学习、基于提示融合的事件抽取。这些方法都是建立在编码器-解码器模型的基础上的，基本模型架构如图 4 所示。

- (1) 将文本和提示输入编码器 (Encoder)；
- (2) 解码层 (Decoder) 负责输出文本；
- (3) 对输出文本进行处理得到结构化数据。

2.3.1 面向要素抽取的提示学习

L2A^[22]将要素抽取任务视为问题的完型填空任务，并用陈述句作为问题模板，以增加跨上下文语言的一致性并提高预测性能。利用模板将输入句子转换为 [CLS]question[SEP]sentence[SEP] 的形式。对于模板的构建，文中提出了手动和自动构建

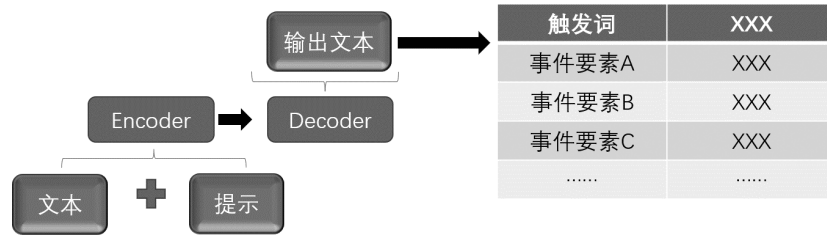


图4 基于提示学习的事件抽取

的方法。手动构建是将要素角色替换为[MASK]，并将必要的信息(例如事件类型和要素跨度标记)添加到问题模板中；自动构建是引入了未占用的标记 $[u1] \sim [u8]$ 并利用反向传播搜索最佳伪问题模板。

2.3.2 面向流水线模式事件抽取的提示学习

DEGREE^[23]通过在输入文本中添加包含触发词和要素的结构化句子提示，将事件抽取任务转换为生成任务。为了生成结构化的句子，设计了特定事件类型的模板来编码每种事件类型。模型的输入是将输入段落与设计好的提示用特殊分隔符[SEP]连接，提示是由对应事件的语义信息描述和事件类

型的预定义模板组成的。该模型可以按照特定事件类型的模板生成输出文本，这些模板包含用于填充触发词或要素的占位符，最终通过比较模板和输出文本来得到抽取结果。

GDAP^[24]使用三个编码器-解码器模型，每个模型分别对应事件类型检测、触发词抽取、要素抽取模块，其中触发词抽取和要素抽取两个模块均采用提示学习的方式。先对句子进行事件类型检测，根据事件的类型，分别进行触发词抽取和要素抽取，并根据模板生成相应的序列。图5介绍了触发词抽取和要素抽取相应的模板和预期输出。

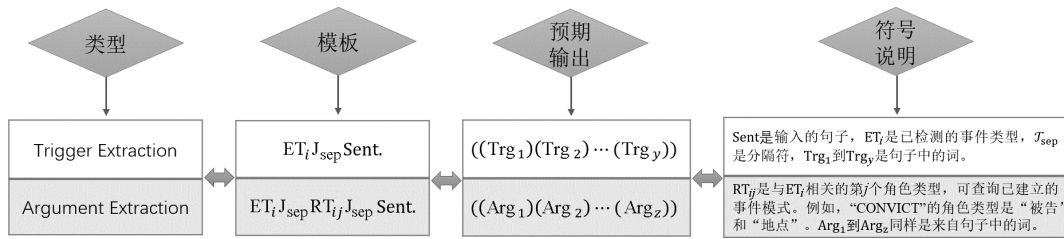


图5 GDAP提示模板

2.3.3 基于提示融合的事件抽取

PoKE^[25]利用多种提示的互补优势使模型更好地理解答案和理解上下文的联系，在流水线模式的事件抽取任务中取得了较好效果。

在事件检测任务中，将任务分成了两步，先检测出主要类型，再映射为子类型。为提取主要类型的触发词，设计了两个联合触发提示，对事件进行交互建模。对于外部联合触发提示，在输入文本后添加8种主要类型的提示，使模型共同生成触发词，这样就能检测出不同事件类型的差异。对于内部联合触发提示，掩码句中的触发词，学习事件的内在联系。

为了提取事件要素，文中提出了单一要素提示和联合要素提示。对于前者，在ACE注释的描述中进行掩码，促使模型生成相应单词，并将事件类型和相应的触发词附在句尾。另外，由于在描述事件时，相同的要素可能在多处出现，为使模型学习到这一点，文中提出联合要素提示，对事件内要素和事件

间要素的交互进行建模。对于每个包含事件的句子都生成提示，并掩码句子中的要素。

2.4 小结

提示学习的动机就是通过提示使预训练语言模型回忆起相关“知识”，因此提示学习在知识抽取相关任务中都获得了较好的效果。提示学习可用于复杂的知识抽取任务，但存在模板工程和答案工程过于复杂、可迁移性不强的问题，如何结合知识抽取任务特点，进一步提高提示学习的性能将成为研究者下一步探索的方向。

3 基于知识的提示学习

基于提示学习的方法已经在知识抽取等任务上取得较好效果，但如何针对具体任务，构造出最合适的提示模板和标签映射，从而进一步提高任务性能

成为研究者面临的一个重大挑战。从 GPT-3^[38]、AutoPrompt^[32] 到 P-tuning^[33], 模板构造经历了从人工构造到自动生成离散或连续模板三个阶段。然而在其快速发展的背后, 提示学习仍然面临构造提示成本高、无法应对专业领域场景、缺乏常识性知识等问题。针对上述问题, 研究者尝试引入外部知识辅助模板生成, 通过知识约束提升模型对任务和领域的感知。

3.1 知识约束方法

知识约束方法是指在构建提示和标签映射的过程中注入外部知识, 压缩搜索空间。主要介绍 PTR^[26]、KPT^[27]、KnowPrompt^[28] 等; 其中, PTR 是利用逻辑规则构建提示; KPT 是通过知识库来扩展标签映射; KnowPrompt 在模板构建和标签映射过程中均注入了知识, 上述方法的示意图见图 6 和图 7。

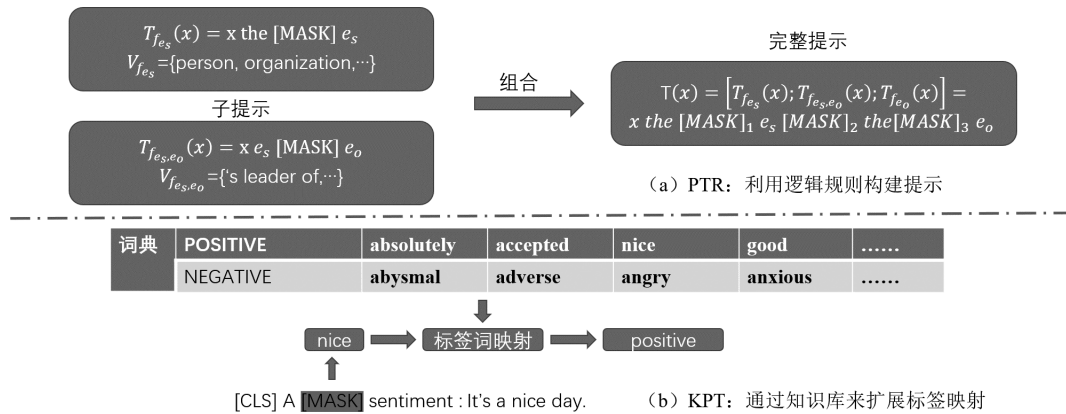


图 6 PTR 和 KPT 方法示意图

PTR^[26] 等工作关注利用知识对提示工程进行约束, 主要采用提示组合策略, 过程如图 6(a) 所示, 手动设计一些基本的子提示, 然后利用逻辑规则制定知识约束, 能自动将子提示构建与任务相关的提示, 降低了构造提示的代价。

具体而言, PTR 设计了本质是一阶谓词逻辑的条件函数, 为条件函数手动设计子提示; 采用类似的方法设计适合于分类任务的二元函数, 并将这些一元和二元函数的设计方法组合为多变量函数, 提供更强大的提示。当需要为任务组合子提示时, 采取具有合取范式的逻辑规则, 直接连接与规则相关的函数的子提示形成模板。PTR 引入的逻辑规则知识并不复杂, 但在关系分类任务上进行的实验, 超过了现有的最佳基线模型, 这证明了利用知识指导模板构建的有效性。

KPT^[27] 主要关注利用知识对答案工程进行约束; 通过注入知识扩充答案空间, 使模型的更多输出能被映射到标签词, 能对专业领域的小样本问题进行数据增强。

具体而言, 先使用外部知识为每个标签生成一组相关词, 这组相关词涵盖了不同粒度和角度的同义词及近义词; 再使用预训练语言模型对扩展的相关词集合进行去噪。例如, 对于主题分类, 利用 ConceptNet^[57] 与 WordNet^[58] 等知识库引入相关

词, 通过边缘表示相关性筛选; 对于情绪判断这种二分类问题, 步骤如图 6(b) 所示, 引入情感字典获得扩展词, 构建扩展标签词词汇表, 通过映射完成预测。

KnowPrompt^[28] 在 PTR 和 KPT 的思想更进一步, 弥补提示学习缺乏常识性知识的缺点。

对于 Typed entity marker 这种标记提示(图 3), 需要给出具体的实体类型, 而虚拟类型词就是使用潜在实体类型的聚合嵌入作为标记, 这种标记是可学习上下文的向量, 这样向量中就有了与实体有关的知识。

在标签词映射时, 通过构建虚拟答案词来实现连续答案搜索, 将关系知识注入。离散答案搜索只能自动生成在词汇表中的一个标签词, 并和一个任务标签之间建立一对一的映射; 而使用向量作为虚拟答案词, 可以得到关系语义词在候选集上的概率分布, 利用这个概率分布初始化这些向量, 然后使用三元组 (s, r, o) 描述关系事实进行约束, 并调整损失函数以便感知上下文优化向量, 这样可以利用预训练语言模型的语义知识完成连续答案搜索。

Ye 等人提出的 PL-Marker^[29] 采用前缀标记提示, 类似于 KnowPrompt 的知识注入方法, 仅对提示的位置进行了调整与优化, 在关系抽取任务上达到目前最佳模型的性能。

3.2 选择性知识植入方法

由于知识噪声和异构性问题,并不是所有的外部知识都能有助于提示学习模型性能提升。针对这一问题,OntoPrompt^[30]基于知识线性转化和注意力掩码矩阵实现了本体知识的选择性注入。

OntoPrompt 利用任务相关的本体知识生成文本作为辅助的标记提示。其主要思路就是从外部知识图谱中提取本体的描述文本作为辅助提示,并将其直接附加到输入序列模板中。如图 7 所示,箭头指向的是输入序列模板,其中[CLS]和[SEP]是开始符号与分隔符,[Input Text]代表原始输入,

[Template]代表处理后的输入,[Ontology Text]和[Meta-Relation Text]代表图中对应方框的文字。但不加选择的知识植入可能会给原始文本引入一些噪声,并导致性能下降。为了解决这一问题,OntoPrompt 基于 K-BERT^[59],提出了跨片段选择性知识植入方法。使用一个可见矩阵调节“Ontology Text”和“Meta-relation Text”的权重,减缓知识注入对输入文本可能造成的负面影响。另外,由于从本体库植入的知识应与输入文本序列相关联,因此采用类似于前缀提示的做法,将“ $u_1 u_2 u_3 u_4$ ”置于[MASK]标签周围,并使用实体词嵌入来初始化和优化,能学习外部的本体序列表示和输入文本序列的表示。

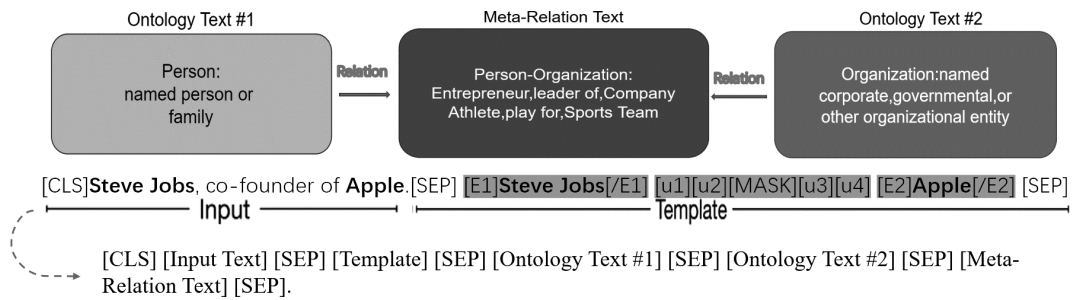


图 7 本体知识注入示意图

3.3 小结

目前,基于知识的提示学习方法在文本分类、关系抽取、事件抽取等任务中验证了模型的性能,并在小样本场景下取得了很好的效果,证明了知识注入、知识选择可以对模板构建、标签映射进行指导,减轻人工设计的负担,提升小样本环境下的迁移能力。

追本溯源,提示学习主要从预训练语言模型里面获取和任务相关的知识,这对预训练模型的规模、预训练数据的来源等都提出了要求^[60]。而提示工程、答案工程都是为了能更好的指导如何从预训练模型中获取任务相关知识,人工设计模板的方法主要也是利用了专家知识;因此,基于知识约束及辅助的提示工程、答案工程必然会在提示学习研究中发挥越来越重要的作用。

4 当前挑战与未来研究趋势

4.1 当前挑战

提示学习的方法不仅仅在知识抽取任务上取得了较好表现,并对小样本场景下的运用进行了探索,

还在自然语言处理的各个方向上有了一定的突破,但从构建、原理和选择三个具体角度来看,仍面临提示学习的模板设计问题以及预训练模型的选择问题。这些问题制约了提示学习的深入研究和广泛应用,亟待后续研究加以解决。

4.1.1 提示学习的模板设计问题

目前应用提示学习的工作大多集中于分类工作和生成工作,其他工作则较少。另外,“模板”和“标签词”的分割也亟待解决。模型的效果同时依赖于应用“模板”和“标签词”的映射,仅自动生成最优的连续模板或者搜索出最优的答案空间就是一项非常复杂的工作,同时让二者达到最优极富挑战性。此外,提示学习的模板设计效果存在可解释性差的问题,部分意义相近的模板会对模型性能产生巨大影响。例如,Liu 等人对 LAMA-TREx P17 案例进行分析^[33],对哥伦比亚阳光海岸的所属国家进行提问,模板中的“In”使准确率存在 20% 的波动,见表 3。结合知识开展提示工程和答案工程的一体化设计,并通过知识约束使提示学习的模型更稳定、效果更可解释,这可能是解决提示学习设计问题的可行之道。

表 3 提示模板影响研究

模板	准确率/%
[X] is located in which country? [Y].	31.40
[X] is located in which country? In [Y].	51.08

4.1.2 提示学习的预训练模型选择问题

提示学习的效果会受预训练模型选择的影响,目前各种自动生成提示和自动搜索答案的方法层出不穷,研究者们忽略对预训练语言模型的选择,并没有探究不同的自动生成方法和自动搜索方法在不同预训练模型下的表现。此外,面向大规模知识图谱的图预训练模型中蕴含了更为密集的知识,而当前提示学习方法忽略了图预训练模型中知识的利用;对于知识密集型任务,提示学习也没做针对性的预训练模型优化。结合基于大规模知识图谱的图表示和图预训练模型^[61-63],开展提示学习以及结合提示学习进行知识抽取,可能会取得更好的效果。

4.2 未来研究趋势

4.2.1 联合知识抽取的探索

在基于提示学习的知识抽取方面,对于关系抽取和事件抽取,目前基于提示学习的方法都是流水线模式的,流水线模式存在固有的错误传播问题,会影响模型的性能。近期,百度与中科院的 UIE^[64]基于提示学习,统一了抽取任务,并在 13 个数据集的全监督、低资源、少样本设置下均达到目前最佳模型的性能,证明了基于提示学习的联合抽取方法是极具价值的研究方向。

4.2.2 更复杂的知识注入

在结合知识的提示学习方面,目前注入的知识以三元组中的实体关系为主,注入的方法仅为简单的谓词逻辑或前缀提示;但知识图谱蕴含的知识结构信息、推理信息尚未得到充分利用,如何设计更好的知识约束方法,将这些复杂的知识通过提示学习的方式引入模型中,是基于知识的提示学习方法未来发展的重点方向。

5 总结

本文对提示学习方法在知识抽取任务上的运用做了综述,对具体方法中的提示工程、答案工程和多提示学习进行了分类和对比分析,已有的研究表明提示学习适用于知识抽取任务。针对提示学习中模

板和标签映射的构造问题,阐述了基于知识约束的解决方案与知识选择注入方法,并分析了与知识相结合的提示学习研究面临的问题和未来发展趋势。从目前形势看,在基于提示学习的知识抽取方面,联合抽取方法是极具价值的研究方向;在基于知识的提示学习方面,更复杂的知识注入与针对图模型的提示学习可能是未来的重点发展方向。相信随着更深入研究,与知识相结合的提示学习方法将会成为自然语言处理的一个新范式。

参考文献

- [1] 图书馆·情报与文献学名词审定委员会. 图书馆·情报与文献学名词 [M]. 北京: 科学出版社, 2019:103-358.
- [2] SUTTON C, ROHANIMANESH K, MCCALLUM A. Dynamic conditional random fields: Factorized probabilistic models for labeling and segmenting sequence data [J]. Journal of Machine Learning Research, 2007, 8(3): 693-723.
- [3] GILDEA D, KHUDANPUR S, SARKAR A, et al. A smorgasbord of features for statistical machine translation [C]//Proceedings of HLT-NAACL, 2004: 161-168.
- [4] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [J]. arXiv preprint arXiv:1409.0473, 2014.
- [5] KIM Y. Convolutional neural networks for sentence classification [C]//Proceedings of EMNLP, 2014: 1746-1751.
- [6] Alec R, Karthik N, Tim S, Ilya S, et al. Improving language understanding by generative pre training [EB/OL]. https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf [2022-7-17].
- [7] DEVLIN J, CHANG M, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [C]//Proceedings of NAACL, 2019: 4171-4186.
- [8] LEWIS M, LIU Y, GOYAL N, et al. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension [C]//Proceedings of ACL, 2020: 7871-7880.
- [9] QIU X, SUN T, XU Y, et al. Pre-trained models for natural language processing: A survey [J]. Science China. Technological Sciences, 2020, 63(10): 1872-1897.
- [10] LIU P, YUAN W, FU J, et al. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing [C]//Proceedings of the ACM Computing Surveys, 2023, 55(9): 1-35.
- [11] NING D, YULIN C, XU H, et al. Prompt-learning

- for Fine-grained entity typing[C]//Proceedings of the EMNLP, 2022: 6888-6901.
- [12] LEYANG C, YU W, JIAN L, et al. Template-Based Named Entity Recognition Using BART[C]//Proceedings of the ACL-IJCNLP, 2021: 1835-1845.
- [13] ZHOU X, MA R, GUI T, et al. Plug-tagger: A pluggable sequence labeling framework using language models[J].arXiv preprint arXiv: 2110.07331, 2021.
- [14] CHEN X, ZHANG N, LI L, et al. LightNER: A lightweight generative framework with prompt-guided attention for low-resource NER [J]. arXiv preprint arXiv:2109.00720, 2021.
- [15] LEE D, KADAKIA A, TAN K, et al. Good examples make a faster learner: Simple demonstration-based learning for low-resource NER[C]//Proceedings of ACL, 2022: 2687-2700.
- [16] MA R, ZHOU X, GUI T, et al. Template-free prompt tuning for few-shot NER[C]//Proceedings of NAACL, 2022:5721-5732.
- [17] ZHOU W, CHEN M. An improved baseline for sentence-level relation extraction [C]//Proceedings of AACL-IJCNLP, 2022:161-168.
- [18] CHEN X, XIE X, ZHANG N, et al. AdaPrompt: Adaptive prompt-based finetuning for relation extraction[J]. arXiv e-prints, 2021: arXiv: 2104.07650.
- [19] CHEN X, LI L, ZHANG N, et al. Relation extraction as open-book examination: Retrieval-enhanced prompt tuning [C]//Proceedings of SIGIR, 2022: 2443-2448.
- [20] CHIA Y K, BING L, PORIA S, et al. Relation-Prompt: Leveraging prompts to generate synthetic data for zero-shot relation triplet extraction[C]//Proceedings of ACL, 2022: 45-57.
- [21] ZHANG N, LI L, CHEN X, et al. Differentiable prompt makes pre-trained language models better few-shot learners[C]//Proceedings of ICLR, 2022.
- [22] YE H, ZHANG N, BI Z, et al. Learning to ask for data-efficient event argument extraction [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2022.
- [23] HSU I, HUANG K, BOSCH E, et al. DEGREE: A data-efficient generative event extraction model [C]//Proceedings of NAACL, 2022:1890-1908.
- [24] SI J, PENG X, LI C, et al. Generating disentangled arguments with prompts: A simple event extraction framework that works[C]//Proceedings of ICASSP, 2022: 6342-6346.
- [25] LIN J, JIAN J, CHEN Q. Eliciting knowledge from language models for event extraction[J].arXiv preprint arXiv: 2109.05190, 2021.
- [26] HAN X, ZHAO W, DING N, et al. PTR: Prompt tuning with rules for text classification[J].arXiv preprint arXiv: 2105.11259, 2021.
- [27] HU S, DING N, WANG H, et al. Knowledgeable prompt-tuning: Incorporating knowledge into prompt verbalizer for text classification[C]//Proceedings of ACL, 2022: 2225-2240.
- [28] 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.
- [29] YE D, LIN Y, LI P, et al. Packed levitated marker for entity and relation extraction[C]//Proceedings of ACL, 2022: 4904-4917.
- [30] 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.
- [31] LI X L, LIANG P. Prefix-tuning: Optimizing continuous prompts for generation [C]//Proceedings of ACL-IJCNLP, 2021:4582-4597.
- [32] SHIN T, RAZEGHI Y, LOGANIV R L, et al. AutoPrompt: Eliciting knowledge from language models with automatically generated prompts[C]//Proceeding of EMNLP, 2020:4222-4235.
- [33] LIU X, ZHENG Y, DU Z, et al. GPT understands, too[J].arXiv preprint arXiv:2103.10385, 2021.
- [34] PETRONI F, ROCKTÄSCHEL T, LEWIS P, et al. Language models as knowledge bases[C]//Proceedings of EMNLP-IJCNLP, 2019:2463-2473.
- [35] SCHICK T, SCHÜTZE H. It's not just size that matters: Small language models are also few-shot learners [C]//Proceedings of NAACL, 2021: 2339-2352.
- [36] JIANG Z, XU F, et al. How can we know what language models know[J]. Transactions of the Association for Computational Linguistics, 2020, 8: 423-438.
- [37] SCHICK T, SCHMID H, SCHÜTZE H. Automatically identifying words that can serve as labels for few-shot text classification [C]//Proceedings COLING, 2020:5569-5578.
- [38] 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, 33: 1877-1901.
- [39] LU Y, BARTOLO M, MOORE A, et al. Fantastically ordered prompts and where to find them: overcoming few-shot prompt order sensitivity[C]//Proceedings of ACL, 2022: 8086-8098.
- [40] WANG X, ZHOU K, WEN J, et al. Towards unified conversational recommender systems via knowledge-enhanced prompt learning [C]//Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022: 1929-1937.
- [41] RAJAGOPAL D, KHETAN V, SACALEANU B, et al. Cross-domain reasoning via template filling[J]. arXiv preprint arXiv: 2111.00539, 2021.

- [42] ZHAO L, ZHENG F, ZENG W, et al. ADPL: Adversarial prompt-based domain adaptation for dialogue summarization with knowledge disentanglement [C]//Proceedings of SIGIL, 2022: 245-255.
- [43] TAN Z, ZHANG X, WANG, S, et al. MSP: Multi-stage prompting for making pre-trained language models better translators [C]//Proceedings of ACL, 2022: 6131-6142.
- [44] SCHUCHER N, SIVA R, HARM V. The power of prompt tuning for low-resource semantic parsing [C]//Proceedings of ACL, 2022: 148-156.
- [45] HUANG Z, XU W, YU K. Bidirectional LSTM-CRF models for sequence tagging [J]. arXiv preprint arXiv: 1508.01991, 2015.
- [46] REIMERS N, GUREVYCH I. Sentence-BERT: Sentence embeddings using siamese BERT-networks [C]//Proceedings of EMNLP-IJCNLP, 2019: 3982-3992.
- [47] ZHANG T, KISHORE V, WU F, et al. BERTScore: Evaluating text generation with BERT [C]//Proceedings of ICLR, 2020.
- [48] RU C, TANG J, LI S, et al. Using semantic similarity to reduce wrong labels in distant supervision for relation extraction [J]. Information Processing & Management, 2018, 54(4): 593-608.
- [49] GRISHMAN R, WESTBROOK D, MEYERS A. NYU's English ACE system description [EB/OL]. <http://nlp.cs.nyu.edu/pubs/papers/ACE05-NYUChinese.pdf> [2022-5-13].
- [50] LIN Y, JI H, HUANG F, et al. A joint neural model for information extraction with global features [C]//Proceedings of ACL, 2020: 7999-8009.
- [51] WADDEN D, WENNERBERG U, LUAN Y, et al. Entity, relation, and event extraction with contextualized span representations [C]//Proceedings of EMNLP-IJCNLP, 2019: 5784-5789.
- [52] LI F, PENG W, CHEN Y, et al. Event extraction as multi-turn question answering [G]//Proceedings of the Association for Computational Linguistics: EMNLP, 2020: 829-838.
- [53] LIU J, CHEN Y, LIU K, et al. Event extraction as machine reading comprehension [C]//Proceedings of EMNLP, 2020: 1641-1651.
- [54] DU X, CARDIE C. Event extraction by answering (Almost) natural questions [C]//Proceedings of EMNLP, 2020: 671-683.
- [55] PAOLINI G, ATHIWARATKUN B, KRONE J, et al. Structured prediction as translation between augmented natural languages [C]//Proceedings of ICLP, 2021: 1-26.
- [56] LU Y, LIN H, XU J, et al. Text2Event: Controllable sequence-to-structure generation for end-to-end event extraction [C]//Proceedings of ACL, 2021: 2795-2806.
- [57] SPEER R, CHIN J, HAVASI C. ConceptNet5.5: An open multilingual graph of general knowledge [C]//Proceedings of AAAI, 2017: 4444-4451.
- [58] PEDERSEN T, PATWARDHAN S, MICHELIZZI J. WordNet: Similarity—measuring the relatedness of concepts [C]//Proceedings of AAAI, 2004: 25-29.
- [59] LIU W, ZHOU P, ZHAO Z, et al. K-BERT: Enabling language representation with knowledge graph [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(03): 2901-2908.
- [60] ZHU C, ZENG M. Impossible triangle: What's next for pre-trained language models? [J]. arXiv preprint arXiv: 2204.06130, 2022.
- [61] YU D, ZHU C, YANG Y, et al. JAKET: Joint pre-training of knowledge graph and language understanding [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2022.
- [62] SUN T, SHAO Y, QIU X, et al. CoLAKE: Contextualized language and knowledge embedding [C]//Proceedings of COLING, 2020: 3660-3670.
- [63] SHEN T, MAO Y, HE P, et al. Exploiting structured knowledge in text via graph-guided representation learning [C]//Proceedings of EMNLP, 2020: 8980-8994.
- [64] LU Y, LIU Q, DAI D, et al. Unified structure generation for universal information extraction [C]//Proceedings of the ACL, 2022: 5755-5772.



鲍琛龙(1998—), 硕士研究生, 主要研究领域为信息抽取和知识图谱。

E-mail: baochenlong@nudt.edu.cn



吕明阳(1999—), 硕士研究生, 主要研究领域为自然语言处理和文本生成。

E-mail: lvmingyang17@nudt.edu.cn



唐晋韬(1981—), 通信作者, 博士, 教授, 主要研究领域为信息抽取和自然语言处理。

E-mail: tangjintao@nudt.edu.cn