小 型 微 型 计 算 机 系 统
Journal of Chinese Computer Systems
DOI:10.20009/j.cnki.21-1106/TP.2025-0119
2026 年 4 月 第 4 期
Vol.47 No.4 2026

一种提示微调驱动的模型知识校准关系抽取方法

郑滋辉[1],黄玉娇[2],杨旭华[1],徐新黎[1]

[1](浙江工业大学 计算机科学与技术学院,杭州 310023)
[2](浙江工业大学之江学院 理学院,浙江 绍兴 312030)
E-mail:huangyujiao@zjut.edu.cn

摘 要: 近年来,提示微调在自然语言处理任务中展现出卓越的性能,其核心是设计与任务匹配的提示模板和答案空间.然而,在关系抽取任务中,构建合适的模板依赖专家知识,且难以有效利用关系标签语义.同时,模型预测易受实体与关系虚假关联的影响.为此,本文提出了一种基于提示微调的模型知识校准关系抽取方法.首先,将关系标签知识注入可微提示,通过注意力机制聚合模型多层表征.采用特征匹配策略和结构约束方法增强模型对关系语义和三元组结构知识的感知,协同优化提示模版.最后,通过因果分析消除实体知识带来的预测偏差.实验结果表明,在 4 个关系抽取基准数据集的全监督和小样本场景下,模型仍展现出具有竞争力或更优的性能.

关 键 词: 关系抽取;提示微调;注意力机制;特征匹配;协同优化
中图分类号: TP391 **文献标识码:** A **文 章 编 号:** 1000-1220(2026)04-0852-07

Model Knowledge Calibration Knowledge for Relation Extraction Driven by Prompt-tuning

ZHENG Zihui[1], HUANG Yujiao[2], YANG Xuhua[1], XU Xinli[1]

[1](College of Computer and Technology, Zhejiang University of Technology, Hangzhou 310023, China)
[2](School of Science, Zhijiang College of Zhejiang University of Technology, Shaoxing 312030, China)

Abstract: In recent years, prompt-tuning has demonstrated excellent performance in natural language processing tasks, which aims to design prompt templates and answer spaces that match the task. However, building appropriate templates relies on expert knowledge in relation extraction tasks, and it is difficult to effectively utilize the semantics of relation labels. Moreover, the predictions are susceptible to the false associations between entities and relations. To this end, we propose a model knowledge calibration method based on prompt-tuning for relation extraction. Firstly, we inject relation label knowledge into differentiable prompts and aggregate the multi-layer representations of the model through an attention mechanism. Feature matching and structural constraints strategies enhance the model's understanding of relation semantics and triple structures, jointly optimizing prompt templates. Additionally, causal analysis is used to eliminate the prediction bias caused by entity knowledge in the reasoning stage. Experimental results show that the model still performs competitively or better in standard and low-resource scenarios on four relation extraction benchmark datasets.

Keywords: relation extraction; prompt-tuning; attention mechanism; representation matching; jointly optimizing

0 引 言

关系抽取(Relation Extraction,RE)是信息抽取领域的重要任务,旨在给定的上下文中识别实体对之间的关系,在知识图谱构建[1]、知识问答系统[2]、智能对话系统[3]等多个领域具有广泛应用.近年来,基于预训练语言模型(Pretrained Language Models,PLMs)的微调方法在多个关系基准数据集上取得了显著进展,但仍有一定的局限性.这些方法依赖额外的分类层,并需要大量标注数据来调整模型参数,导致模型复杂度增加,且对标注数据的依赖性较强.

为解决这个问题,研究者们提出了基于提示的微调方法[4],通过设计任务特定模版将原始任务重新形式化为与预训练任务更接近的结构,并利用候选答案词映射关系标签,使下游任务与预训练目标更好地对齐.具体而言,提示微调通过将原始输入与包含掩码令牌([MASK])标记的提示模版结合,并将其输入到 PLMs 中,利用掩码语言模型进行学习和推理[5].

对于标签较为简单的自然语言处理任务,如情感分类,通过提示可以有效地将答案词映射到标签空间.但在 RE 任务中,由于关系标签蕴含丰富的语义信息,传统的离散答案空间方法难以充分利用这些知识.现有 RE 提示方法主要依赖手动设计和搜索模板,既费时费力,又难以找到最优解.此外,实体信息可能干扰模型预测,导致错误关系识别.

针对上述问题,本文提出了一种提示微调驱动的模型知识校准的关系抽取方法,通过增强关系标签的语义信息,优化提示模板构建,提高答案词与标签空间的对齐效率.同时,该

收稿日期:2025-03-14 收修改稿日期:2025-05-07 基金项目:"尖兵领雁 + X"研发攻关项目(2024C03274)资助;国家自然科学基金项目(62176236)资助;浙江省自然科学基金项目(LZ24F030011)资助. 作者简介:**郑滋辉**,男,1998 年生,硕士研究生,研究方向为自然语言处理;**黄玉娇**(通信作者),女,1985 年生,博士,副教授,研究方向为深度学习、文本数据分析;**杨旭华**,男,1971 年生,博士,教授,CCF 杰出会员,研究方向为机器学习、网络科学;**徐新黎**,女,1977 年生,博士,副教授,CCF 会员,研究方向为信息抽取、网络嵌入、医学图像计算.

方法强化实体与关系间的隐含结构关联,协同优化提示模板,并在推理阶段消除实体信息带来的偏差,从而提升模型性能.本文的主要贡献可归纳为以下几点:

1)利用提示学习构建知识注入的提示和答案空间,并结合注意力机制聚合关键知识,增强掩码的表征.

2)采用特征匹配策略增强模型对关系标签语义感知,结合三元组结构约束强化实体与关系的隐含结构关联.

3)借助因果分析方法消除实体信息给 PLMs 预测带来的潜在偏差,提升模型预测准确性.

1　相关研究

1.1　句子级关系抽取

句子级关系抽取是自然语言处理领域的一项核心任务,旨在从上下文中识别给定实体对之间的特定语义关系.这项任务对于知识图谱构建、问答系统以及信息检索等应用具有重要意义.句子级关系抽取的一个关键挑战是如何充分利用上下文信息来准确识别实体间的关系.传统的关系抽取方法主要依赖于规则模版、统计模型以及早期的神经网络架构(如 CNN、RNN).近年来,随着 PLMs 的兴起,基于 PLMs 的微调方法在关系抽取任务中取得了显著进展.同时,研究者还开发了多种知识增强的 PLMs,例如 KNOWBERT[6]、SPanBERT[7]以及 LUKE[8]等,通过引入外部知识进一步提升模型性能.

在小样本环境中,关系抽取面临标注数据稀缺的挑战.为此,基于度量学习的方法(如原型网络)通过计算嵌入空间中的语义相似性来实现关系分类,展现出较强的适应性和泛化能力,为低资源场景下的关系抽取提供了新的解决方案.

1.2　提示微调

提示微调是一种通过利用自然语言文本作为提示来微调 PLMs 适应下游任务的方法.自 GPT-3[9]、LLMA[10]等大型语言模型问世以来,提示学习能够有效将预训练模型中的知识迁移到情感分析[11]、自然语言推理和关系抽取等下游任务中,在小样本环境[12]中展现出显著优势.传统方法[13]依赖领域知识人工构建与任务相关的提示模板和标签词映射(Verbalizer),但这种方式成本较高且缺乏灵活性.为降低人工构建提示的成本,Li 等人[14]提出了自动搜索方法,Gao 等人[15]和 Schick 等人[16]首次探索了自动生成答案词和模板的方法.Shin 等人[17]进一步提出了基于梯度引导的搜索方法,自动生成模板和词汇表中的标签词.近年来,提示学习通过进一步融合外部知识以增强模型性能.例如,Han 等人[18]通过逻辑规则构造出具有多个子提示的提示模版,Zhang 等人[19]引入虚拟提示和协同优化策略,通过可学习的虚拟类型词和答案词构建提示,减少对预定义规则的依赖.总体而言,提示学习通过结合外部知识、自动化提示设计和连续优化策略,不仅降低了人工干预成本,还增强了模型对复杂语义任务的理解能力,为自然语言处理领域提供了一种高效且灵活的解决方案.

2　模型设计

本文提出了一种基于提示微调的模型知识校准关系抽取方法(整体框架如图 1 所示),主要由特征提取模块、知识增强模块、以及偏差移除模块构成.

2.1　特征提取模块

类型标记方法认为引入额外的实体类型,可以增强 PLMs 的 RE 能力.关系标签通常具有固定的头、尾实体类型.因此,可利用关系标签集 $Y=\{y_1,y_2,\cdots,y_n\}$ 分析头、尾实体的可能类型,将其平均嵌入来初始化头、尾实体的标识[sub]和[obj]以构建可微提示 P1 =[sub]entity1 [MASK][obj]entity2,其中 entity1 表示头实体,entity2 表示尾实体,掩码令牌([MASK])在文本序列中代表需 PLMs 预测的给定实体间的关系,并根据上下文动态调整.例如,可将"per:cities_of_residence"去除标点并恢复缩写,扩展为"person cities of residence"作为关系描述输入,与提示 P1 构成新的输入序列"[CLS] person cities of residence. [SEP] [sub] entity1 [MASK][obj]entity2 [SEP]",以关系描述词通过提示得到的掩码表征来初始化答案空间 $Q=\{r_1,r_2,\cdots,r_n\}$ 中与真实标签对应的答案词.模型的输入为 $X=\{x_1,x_2,\cdots,x_n\}$,将输入的原始文本序列与提示 P1 拼接成新的文本序列"[CLS] X [SEP] P1 [SEP]"输入到 PLMs 中.

BERT 类模型输出的每一层都有不同的语义特征,因此本文提出结合注意力机制聚合关键知识.经过初步实验,选择 PLMs 输出的第 2 层 ~ 第 M(M 为隐藏层大小)层([MASK])位置输出作为实体间关系源,如公式(1)所示:

$$\chi_{mask}=\begin{bmatrix} e_{mask\_2} \\ e_{mask\_3} \\ \cdots \\ e_{mask\_M} \end{bmatrix},\chi_{mask}\in R^{(M-1)\times d_{PLM}} \tag{1}$$

其中,d_{PLM}是嵌入维数的大小.由于集合 χ_{mask} 中关系表征对于输入表示具有不同的权重.因此,本文利用注意力机制为每个关系特征计算不同的权重.具体地说,首先计算 χ_{mask} 的 *tanh* 值,计算过程如公式(2)所示:

$$\chi_i=tanh(\chi_{mask\_i}) \tag{2}$$

其中,i 为当前隐藏层层数.计算 χ_i 和输入表征 e_{CLS} 之间的相似性,并通过 *Softmax* 将其转换为概率分布,如公式(3)所示:

$$\alpha_i=\frac{exp(\chi_i^T e_{CLS})}{\sum_{j=2}^{M}\chi_j^T e_{CLS}} \tag{3}$$

α_i 可用来度量每层知识特征相对输入的重要程度,利用 α_i 作为 χ_{mask} 上的权重计算得到模型预测的最终关系 x_{out},如公式(4)所示:

$$x_{out}=\sum_i \alpha_i x_{mask\_i} \tag{4}$$

2.2　知识增强模块

直接采用模型预测结果作为最终输出可能导致其过度依赖训练数据的表面特征,而忽略深层语义信息,尤其难以区分语义相似的关系.为此,本文引入特征匹配机制,通过对比学习增强模型对关系标签语义特征的理解.选择答案空间 $Q=\{r_1,r_2,\cdots,r_n\}$ 中真实答案词 r_i 作为正样本,随机选择除 r_i 外的任意答案词作为负样本 r'_i,通过计算余弦相似度,使得模型预测结果与真实答案词更加接近,如公式(5)所示:

$$L_{rel}=max\{cos(x_{out},r_i)-cos(x_{out},r'_i)+\gamma,0\} \tag{5}$$

其中 $cos(\cdot)$ 表示余弦相似度,γ 为边缘参数.此外,本文通过增强三元组间的结构约束使模型在预测关系时更加关注三元

组结构上的关联,减轻无关上下文对模型预测结果的影响. 计算过程如公式(6)所示:

$$L_{struct} = -log[cos(e_{[sub]} + x_{out}, e_{[obj]})] - log[cos(e'_{[sub]} + r_i, e'_{[obj]})] \quad (6)$$

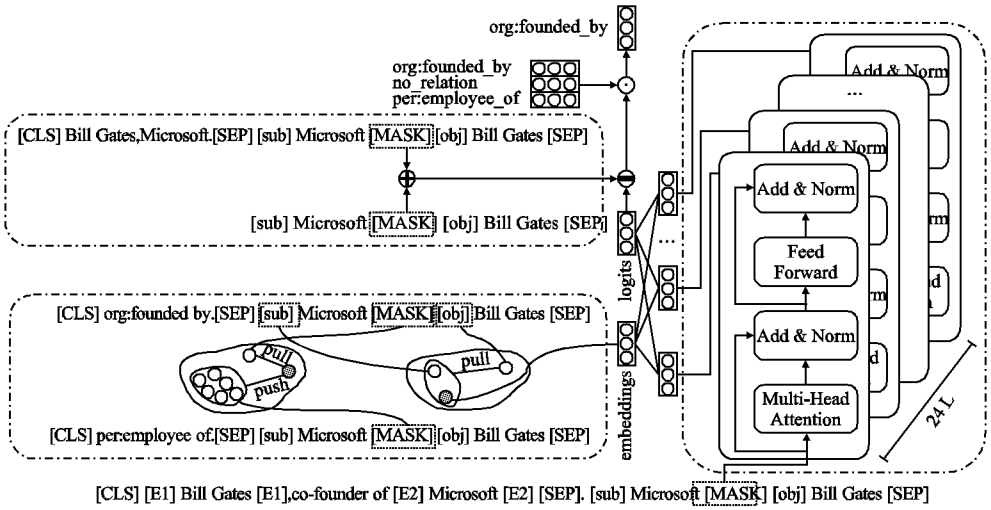

图 1 模型架构

Fig. 1 Model architecture

其中 $e_{[sub]}$ 和 $e_{[obj]}$ 分别为 2.1 节构建的可微提示模板中头、尾实体的类型标识,x_{out} 为模型预测的实体间的关系嵌入,构成正三元组结构($e_{[sub]}, x_{out}, e_{[obj]}$). 将提示模板中掩码令牌"[MASK]"位置填入实体对(E_1, E_2)间真实关系标签 y 所对应的答案词 r_i,并从给定上下文 X 中随机截取不同片段的平均嵌入分别作为负样本中的主、客实体 $e'_{[sub]}$ 和 $e'_{[obj]}$,从而构造出被扰动的负样本三元组($e'_{[sub]}, r_i, e'_{[obj]}$),其中该实体的类型与当前关系不匹配.

2.3 偏差移除模块

通过实验发现,模型可能会过拟合到局部的上下文或部分实体知识,导致它过于依赖这些表面的线索,而忽略了更全局的语义信息. 本文将仅含实体的文本和空白文本分别与 2.1 节中提示模版 P1 拼接成新的文本序列,将模版 P1 中掩码令牌"[MASK]"嵌入分别作为实体偏差 e_{entity} 和上下文偏差 $e_{contextual}$,使用简单的减法计算减轻来自预测结果中偏差,如公式(7)所示:

$$x_{result} = x_{out} - \lambda_1 * e_{entity} - \lambda * e_{contextual} \quad (7)$$

其中,λ_1 和 λ_2 是两个独立的超参数,分别用于平衡减轻偏差的项. 本文使用网格搜索方法在一个限定的二维空间中为不同数据集动态设置 λ_1 和 λ_2 的取值,如公式(8)所示:

$$\lambda_1^*, \lambda_2^* = \arg\max_{\lambda_1,\lambda_2}(F_1(\lambda_1, \lambda_2)) \quad \lambda_1, \lambda_2 \in [a, b] \quad (8)$$

其中,F_1 为打分函数,a, b 为搜索范围的边界. 在训练集中搜索 λ_1, λ_2 的取值,并在所有的验证、测试样本的推理过程中使用固定的值. 将去偏结果输出到全连接层,如公式(9)所示:

$$Z = w(ReLU(x_{result})) + c \quad (9)$$

其中,w 表示权重系数,c 为偏置,y 为真实关系标签. 最后,计算 y_i 与 Z_i 间的交叉熵损失,如公式(10)所示:

$$L_{MASK} = -\sum_{i=1}^{n} y_i log(\frac{Z_i}{\sum_{j=1}^{n} e^{Z_j}}) \quad (10)$$

其中,n 表示数据集中标签类别的数量.

2.4 训练说明

本文的方法有两个阶段的优化过程,首先,以较大的学习率 lr_1 优化可学习提示词、答案词得到最优的提示模版和答案空间,如公式(11)所示:

$$L = L_{MASK} + \alpha * L_{rel} + \beta * L_{struct} \quad (11)$$

其中,α, β 为超参数. 接着,在已优化的提示基础上,使用较小的学习率 lr_2 通过目标函数 L_{MASK} 对 PLMs 参数进行微调.

3 实 验

本章将详细介绍实验部分,包括数据集、基线方法、实验环境和结果,并在 4 个数据集上使用 $F1$ 指数评估方法性能.

3.1 数据集

本节对 4 个广泛使用的关系抽取数据集:TACRED、TACREV、ReTACRED 和 SemEval 的统计详情见表 1.

表 1 数据集描述

Table 1 Data description

Dataset	#Train	#Dev	#Test	#Rel
TACRED	68124	22631	15509	42
TACREV	68124	22631	15509	42
ReTACRED	58465	19584	13418	40
SemEval	6507	1493	2717	19

TACRED 是一个包含 106264 个实例的广泛使用的关系抽取数据集,涵盖了 41 种预定义的关系类型,对于不属于这些预定义关系的实例,数据类型标注为"no_relation".

TACREV 是 TACRED 的改进版本,修正了原始数据集的标注噪声和模糊性问题,保留了 41 种关系类型和"no_relation"类别,提供了更高质量的关系抽取基准.

ReTACRED 是对 TACRED 数据集的重新标注版本,保留了 41 种关系类型和"no_relation"类别,并通过更严格的标

注标准提升了数据质量,成为关系抽取任务中更可靠的评估基准.

SemEval 是一个广泛使用的自然语言处理评测系列,其数据质量高、标注规范,常被用于评估关系抽取模型的泛化能力和性能.

3.2　实验设置与评价指标

本实验基于 RoBERTa 模型,并在全监督和小样本场景下实验,以评估模型在数据稀缺时的泛化能力及充足数据下的性能上限.

全监督环境:严格遵循数据集划分标准,使用完整的训练集和验证集进行实验,采用统一评估指标和超参数设置,验证方法在充足数据条件下的性能上限.

小样本环境:参考 LM-BFF 的设置,分别进行 8-shot、16-shot 和 32-shot 的实验. 为了确保实验结果的稳定性和可靠性,基于固定随机种子集合 seed,从初始训练集和验证集中为每个类别随机抽取 k 个实例,构建少样本训练集和验证集. 每种实验设置均重复 5 次,最终结果取 5 次实验的平均值,以消除随机采样带来的偏差,全面评估模型在数据稀缺场景下的泛化能力.

本实验采用 Python 语言进行模型构建,并基于 Facebook 开源的 PyTorch 深度学习框架实现. 实验环境配置详见表 2.

表 2　实验设置

Table 2　Experiment settings

操作系统	Windows10
内存	24GB
CPU	AMD EPYC 7320
GPU	NVIDIA GeForce RTX 3090Ti
Python	3.8
PyTorch	1.8.2

对于网格搜索方法中的超参数,本实验将公式(8)中超参数的搜索范围限定为[-1,1],搜索步长为 0.05,其他超参数设置详见表 3.

表 3　超参数设置

Table 3　Parameter settings

参数名称	参数值
词嵌入维度	1024
隐藏层大小	1024
最大序列长度	256
Batch Size	8/16/32
α	0.1
β	0.01
lr_1,lr_2	[3e-5,4e-5]/3e-5
Epoch	7/10
γ	0.3
a,b	-1/1

本实验选取分类任务常用的 F1-score 作为模型性能的评价指标,F1-score 是一种平衡精确率 P 和召回率 R 的度量方式,特别适用类别不均衡的任务. 具体计算方式如公式(12)~公式(14)所示:

$$F1=\frac{2\times P\times R}{P+R} \tag{12}$$

其中,精确率和召回率的计算公式分别为:

$$P=\frac{TP}{TP+FP} \tag{13}$$

$$R=\frac{TP}{TP+FN} \tag{14}$$

其中,公式(13)、公式(14)中的 TP 表示被正确预测为正类的样本数,FP 表示被错误预测为正类但实际为负类的样本数,FN 表示被错误预测为负类但实际为正类的样本数. $F1$ 值的取值范围在 0~1 之间,值越接近 1,说明模型在精确率和召回率之间取得了更好的平衡,分类效果更优.

3.3　对比实验

为验证本文方法的有效性,本实验通过在测试集上多次实验,并选取以下基线模型对比. 鉴于复杂语义关系抽取包含微调与提示微调,选取近年具代表性的分类方法进行对比.

a) GDPNet[20]:Xue 等人于 2020 年提出,通过构建和动态优化多角度图结构,解决了长文本关系抽取任务中难以捕捉关键词语和复杂关系的问题.

b) TYP Marker[21]:Zhou 等人于 2022 年提出,通过引入带类型标记的实体表示和改进数据集标签质量,提出了一种改进的句子级关系抽取基线方法,解决了现有模型中实体表示不充分和噪声标签影响性能的问题.

c) NLI[22]:Sainz 等人于 2021 年提出,将关系抽取转化为文本蕴含任务,结合预训练模型与手工关系描述,在低资源场景下有效减少对标注数据的依赖并提升性能.

d) KNOWBERT[6]:Peters 等人于 2019 年提出,将知识库嵌入预训练模型,增强上下文表示,解决模型缺乏显式知识和对实体记忆能力不足的问题.

e) SpanBERT[7]:Joshi 等人于 2020 年提出,通过屏蔽连续随机跨度而非随机标记,并训练边界表示预测完整跨度内容,提高 BERT 在问答和共指消解等跨度预测任务的表现.

f) LUKE[8]:Yamada 等人于 2020 年提出,该方法使用一个新的基于双向 transformer 的预训练的上下文表示的词和实体的模型.

g) PTR[18]:Han 等人于 2021 年提出,该方法应用逻辑规则来构建 prompts 和一些 sub-promtps,可以在 prompt tuning 时从各个类中表征先验知识.

h) KnowPrompt[19]:Zhang 等人于 2022 年提出,该方法把关系标签之间的知识整合到关系提取的提示模版中,并提出一种使用协同优化的知识感知提示微调方法.

i) FPC[23]:Yang 等人于 2022 年提出,通过逐步增加难度的提示学习课程进行微调,使模型适应复杂的多任务设置.

j):GenPT[24]:Han 等人于 2022 年提出,该方法提出了一种新颖的生成式提示调优方法,将关系分类重新定义为填充问题,摆脱了当前基于提示方法的限制.

本模型与上述对比方法在各数据集上的实验结果汇总于表 4、表 5. 表中采用粗体标注对比模型中性能最佳结果,并使用下划线标注次优结果. 观察实验结果做出分析:

1)从表 4 可见,本文方法在少样本环境下表现更优. 知识校准方法在所有数据集上均显著优于普通微调方法,尤其在 16-shot 实验中取得最优性能. 与提示微调方法相比,本文方法在不同样本规模下均表现出竞争力或领先优势. 随着 k 从 8 增

加到32,相对其他方法的提升幅度有所减少.本文认为当标注数据量较大时,进一步增强模型对关系和三元组结构知识的利用,对性能的提升有限.这也表明,本文方法能够有效结合关系语义和上下文特征,构建适用于RE任务的最优提示.

表4 低资源关系抽取在不同测试集上的$F1$分数(%)性能表现

Table 4 Low-resource RE performance of $F1$-scores(%) on different test sets

ModeL	TACRED			TACREV			ReTACRED			SemEval		
Shot	8	16	32	8	16	32	8	16	32	8	16	32
SpanBERT	8.4	17.5	17.9	5.2	5.7	18.6	14.2	29.3	43.9	–	–	–
LUKE	9.5	21.5	28.7	9.8	22.0	29.3	14.1	37.5	52.3	–	–	–
GDPNet	11.8	22.5	28.8	8.3	20.8	28.1	18.8	48.0	54.8	42.0	67.5	81.2
TYP Marker	28.9	32.0	32.4	27.6	31.2	32.0	44.8	54.1	60.0	–	–	–
PTR	28.1	30.7	32.1	28.7	31.4	32.4	51.5	56.2	62.1	70.5	81.3	84.2
KnowPrompt	32.0	35.4	36.5	32.1	33.1	34.7	55.3	63.3	65.0	74.3	82.9	84.8
FPC	33.6	34.7	35.8	33.1	34.3	35.5	57.9	60.4	65.3	–	–	–
GenPT	**35.7**	**36.6**	37.4	**34.4**	34.6	36.2	57.1	60.4	65.2	–	–	–
Ours	34.0	36.3	**40.6**	32.7	**36.2**	**39.4**	**60.5**	**66.0**	**68.4**	**78.5**	**82.6**	**85.0**

2)如表5所示,在全监督环境下,基于外部知识库增强的PLMs性能明显优于传统微调模型,表明引入任务相关的外部知识有助于提升模型表现.值得注意的是,本文的方法在部分任务上甚至超越了依赖外部知识进行数据或架构增强的模型.此外,尽管PLMs具备一定任务知识,传统微调方法在激发其下游任务能力方面仍然效率较低且存在挑战.相比基于离散答案空间的提示微调方法,本文的方法在大多数数据集上表现更优,说明在连续答案空间中能达到更好的优化状态.此外,与当前最先进的生成式提示微调方法GenPT[24]相比,本文的方法无需人工设计答案,并且在实体类型不可知的情况下仍能取得相当的性能.尽管KnowPrompt[19]同样采用连续空间进行答案搜索,但本文的模型在挖掘关系语义方面表现更佳,进一步证明了其在关系抽取任务上的优势.

经综合实验分析发现,本模型可有效用于复杂语义关系抽取任务中.

表5 全监督关系抽取在不同测试集上的$F1$分数(%)性能表现

Table 5 Standard RE performance of $F1$-scores(%) on different test sets

Model	Answer space	Extra Data	TACRED	TACREV	ReTACRED	SemEval
Fine-tuning pre-trained models						
Roberta-large	–	W/o	68.7	76.0	84.7	87.6
GDPNet	–	W/o	70.5	80.2	–	–
TYP Marker	–	W/o	74.6	83.2	91.1	–
NLI	–	W/o	71.0	–	–	–
KNOWBERT	–	W	71.5	79.3	89.1	89.1
SpanBERT	–	W	70.8	78.0	85.3	–
LUKE	–	W/o	72.7	80.6	90.3	–
Prompt-tuning pre-trained models						
PTR	discreet	W/o	72.4	81.4	90.9	89.9
FPC	discreet	W/o	72.9	82.9	91.3	90.4
GenPT	discreet	W/o	**74.7**	**83.4**	91.1	–
KnowPrompt	continuous	W/o	72.4	82.4	91.3	90.2
Ours	continuous	W/o	73.5	82.7	**91.4**	**90.8**

3.4 消融实验

本节将深入探究各模块对模型性能的贡献,分别验证了特征提取、知识增强和偏差移除模块在全监督和小样本环境下的效果,具体见表6.

a)特征提取模块效果:对于-attention设置,本文只采用隐藏层最后一层中[MASK]位置的表征进行关系预测.结果表明,利用注意力机制融合模型输出不同特征时,在小样本和中等样本情况下对模型性能的提升作用较为显著,而在样本较多情况下,其作用相对减弱.

b)知识增强模块的效果:本文还进行了消融研究以验证知识增强模块设计的有效性.对于-knowledge,直接利用特征提取模块的结果做交叉熵损失去优化模型,-rel of knowledge设置表示去除用于增强模型对关系语义感知的特征匹配模块,对于-struct of knowledge,本文选择忽略实体和关系之间的结构关联.如表6所示,在16-shot设置中,缺少特征增强模块的模型性能从66.0下降到62.4,而移除特征匹配模块的性能下降到63.0,忽略实体与关系间结构关联模块性能下降到63.5.结果表明,通过增强模型对关系知识的感知,以及结构化约束确实可以提高模型的性能,可能是因为它们可以通过优化可微提示模版以激活模型中与其任务相

适应的知识.

表 6　消融实验结果

Table 6　Results of ablation experiment

Model	F1-score			
	8-shot	16-shot	32-shot	full
Ours	**60.5**	**66.0**	**68.4**	**91.4**
-attention	59.1	63.4	67.6	90.7
-knowledge	59.2	62.4	68.1	89.4
-rel of knowledge	59.4	63.0	65.1	90.0
-struct of knowledge	60.0	63.5	65.7	90.6
-bias	58.0	61.7	67.8	91.0

c)偏差移除模块的效果:此外,-bias 指忽略实体在预测时给模型带来的影响. 在 8-shot 设置中,模型性能为 58.0,是所有设置中最低的,可能是在数据量非常少时,模型没办法从外部数据中获得足够的知识,因此很容易产生对实体特征的病理性依赖.

经综合实验分析发现,各功能组件均对模型整体表现产生正向影响.

4　分析与讨论

4.1　手工设计的提示模版是否影响模型性能

本节比较了不同提示模版对模型性能的影响. 本方法中默认使用的提示模板是“P1 = [sub] entity1 [MASK] [obj] entity2”. 本实验尝试与两个手工设计的提示模版进行比较:“P2 = The meaning of [label] is [MASK]”,“[Label] means [MASK]”. 在 SemEval 数据集上进行了多次实验,结果如图 2 所示.

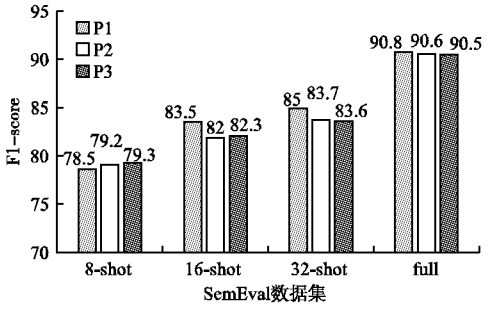

图 2　模型效果对比

Fig. 2　Experimental results of base classifiers

结果表明,以知识注入的可学习提示模版展现出更有效的性能,其他不同提示模版对模型性能的影响几乎一致. 鉴于方法中多处使用了提示模版,且 P1 性能更优,为了更好的优化可微提示,本文选择 P1 作为默认提示模版.

4.2　注意力机制是如何作用于知识特征提取

不同的自然语言处理任务对具有决定性注意力值的层关注度是不同的,可以通过动态权重分配机制,实现对多力度知识表征的层级化捕捉. PLMs 的编码层在 RE 任务中展现出 3 阶段特性:浅层主要捕捉词法特征,中层强调上下文依赖,而深层则集中处理高阶语义推理. RE 任务对 PLMs 输出知识的每一层表现出不同的关注程度,这说明,不同层次的知识对 RE 任务具有不同的意义,突出了 RE 任务的注意力机制的有效性.

4.3　RE 任务中因果推理分析偏差

为了更好地理解关系抽取任务中的因果关系,如图 3 所示. 传统关系抽取模型通过实体识别(X→E)构建实体节点,往往将实体跨度作为独立语义单元处理,这种建模方式隐含了实体特征与上下文解耦的假设. 在因果图框架下,X→E揭

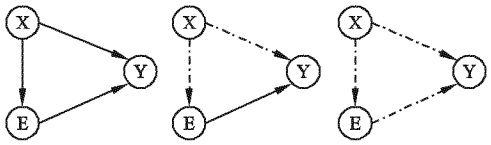

图 3　因果图

Fig. 3　Causal graph

示了上下文对预测结果的直接因果效应,而实体节点 E 在此结构中实际充当了混淆变量,当模型过度拟合实体表面特征,会形成 E 到 Y 的伪相关路径,导致模型决策偏离真实因果链(X→Y). 为解耦这种偏差效应,本文通过固定 X 的取值构建反事实预测场景,此时模型若仍能通过 E→Y 路径产生显著预测置信度,则证明其决策机制中存在对实体特征的病理性依赖. 从而揭示了传统管道式实体-关系建模的局限,即实体边界的硬性切割可能破坏文本内在的语义连贯性.

5　结　论

本文提出了一种基于提示微调的模型知识校准关系抽取方法,主要利用特征匹配及结构约束策略矫正模型知识来联合优化知识注入的可微提示,并利用因果推理方法去除预测结果偏差. 实验证明本文提出的方法对提高提示微调的有效性和泛化能力具有重要意义和实用价值. 未来将继续探索可提高模型性能的外部及其他可导致偏差的因素.

References:

[1] Mirtaheri M. Relational learning to capture the dynamics and sparsity of knowledge graphs[C]//Proceedings of the 35th AAAI Conference on Artificial Intelligence,2021:15724-15725.

[2] Lan Y S,Jiang J. Modeling transitions of focal entities for conversational knowledge base question answering[C]// Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing,2021:3288-3297.

[3] Madotto A,Wu C S,Fung P. Mem2seq:effectively incorporating knowledge bases into end to end task oriented dialog systems [C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics,2018:1468-1478.

[4] Liu P F,Yuan W Z,Fu J L,et al. Pretrain,prompt,and predict:a systematic survey of prompting methods in natural language processing[J]. ACM Computing Surveys,2023,55(9):1-35.

[5] Chen Y,Shi B W,Xu K. PTCAS:prompt tuning with continuous answer search for relation extraction[J]. Information Science,2024,659:120060,doi:10.1016/j.ins.2023.120060.

[6] Peters M E,Neumann M,Logan Iv R L,et al. Knowledge enhanced contextual word representations[C]//Proceedings of the Confer-

ence on Empirical Methods in Natural Language Processing and the International Joint Conference on Natural Language Processing, 2019:43-54.

[7] Joshi M, Chen D Q, Liu Y H, et al. SpanBERT: improving pretraining by representing and predicting spans[J]. Transactions of the Association for Computational Linguistics, 2020, 8:64-77, doi:10.48550/arXiv.1907.10529.

[8] Yamada I, Shindo H, Takeda H, et al. LUKE: deep contextualized entity representations with entity-aware self-attention[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2020:6442-6454.

[9] Brown T B, 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.

[10] Petroni F, Lewis P, Bakhtin A, et al. Language models as knowledge bases[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, 2019:2463-2473.

[11] LIU X Y, GUO Y. Aspect-aware sentiment classification model based on multi-task joint training[J]. Journal of Chinese Computer Systems, 2024, 45(7):1545-1551.

[12] Zhang P Y, Lu W. Better few-shot relation extraction with label prompt dropout[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2022:6996-7006.

[13] Schick T, Schütze H. Exploiting cloze questions for few shot text classification and natural language inference[C]//Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics, 2021:255-269.

[14] Li X L, Liang P. Prefix-tuning: optimizing continuous prompts for generation[C]//Proceedings of the Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, 2021:4582-4597.

[15] Gao T Y, Fisch A, Chen D Q. Making pre-trained language models better few-shot learners[C]//Proceedings of the Annual Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing, 2021:3816-3830.

[16] Schick T, Schmid H. Automatically identtifying words that can serve as labels for few-shot text classification[C]//Proceedings of the 28th International Conference on Computational Linguistics, 2020:5569-5578.

[17] Shin T, Razeghi Y, Wallace E, et al. AutoPrompt: eliciting knowledge from language models with automatically generated prompts[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2020:4222-4235.

[18] Han X, Zhao W L, Ding N, et al. PTR: prompt tuning with rules for text classification[J]. AI Open, 2022, 3:182-192, doi:10.1016/j.aiopen.2022.11.003.

[19] Chen X, Zhang N Y, Xie X, et al. KnowPrompt: knowledge-aware prompt-tuning with synergistic optimization for relation extraction[C]//Proceedings of the ACM Web Conferencec, 2022:2778-2788.

[20] Xue F Z, Sun A X, Zhang H, et al. GDPNet: refining latent multi-view graph for relation extraction[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2021:14194-14202.

[21] Zhou W X, Chen M H. An improved baseline for sentence-level relation extraction[C]//Proceedings of the 2nd Conference of the Asia Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing, 2022:161-168.

[22] Sainz O, Labaka G, Barrena A, et al. Label verbalization and entailment for effective zero and few-shot relation extraction[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2021:1199-1212.

[23] Yang S C, Song D D. Fpc: fine-tuning with prompt curriculum fot relation extraction[C]//Proceedings of the 2nd Conference of the Asia Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing, 2022:1065-1077.

[24] Han J L, Zhao S, Cheng B, et al. Generative prompt tuning for relation classification[C]//Proceedings of the Association for Computational Linguistics, 2022:3170-3185.

附中文参考文献：

[11] 刘欣怡，过弋. 基于多任务联合训练的属性感知情感分类模型[J]. 小型微型计算机系统，2024，45(7)：1545-1551.