

文章编号: 1003-0077(2026)06-0001-13

基于检索增强思维提示的汉语框架语义解析方法

李迎旭, 陈涛, 李斌阳, 黎议泽

(国际关系学院 网络空间安全学院, 北京 100091)

摘要: 汉语框架语义解析基于框架语义学理论,旨在通过识别句子中词语所激活的语义框架,分析句子中各个成分的语义角色,从而揭示语言背后的深层语义结构,进一步更好地抽取事件关系和语境信息。大语言模型出现后,其强大的通用文本理解与生成能力被广泛应用于各种自然语言处理任务中。然而,当前大语言模型在汉语框架语义解析任务中存在推理路径简单、准确率过低的不足,尤其在思维链的逻辑连贯性和检索增强生成的深度应用上存在欠缺。为此,该文提出了一种面向汉语框架语义解析的思维提示方法。该方法结合检索增强生成(RAG)与链式思维(CoT)技术,引导大语言模型完成汉语框架语义解析任务。该文在 CFN 2.1 数据集上的实验结果表明,与 SOTA 方法相比,该方法的框架识别准确率提升 13.52%,论元识别 F_1 值提升 2.24%,角色识别 F_1 值提升 5.09%。

关键词: 汉语框架语义解析;提示词工程;检索增强生成;链式思维

中图分类号: TP391

文献标识码: A

Retrieval-augmented Chain-of-thought Prompting Method for Chinese Frame Semantic Parsing

LI Yingxu, CHEN Tao, LI Binyang, LI Yize

(College of Cyberspace Security, University of International Relations, Beijing 100091, China)

Abstract: Chinese Frame Semantic Parsing (CFSP) seeks to identify the semantic frames evoked by lexical units and assign roles to sentence constituents. To improve large language models (LLMs) in CFSP, we propose a method that integrates RAG with chain-of-thought (CoT) reasoning. Experiments on the CFN2.1 dataset show improvements of 13.52% in frame identification accuracy, 2.24% in argument F_1 , and 5.09% in role F_1 over the best existing methods, respectively.

Keywords: Chinese frame semantic parsing; prompt engineering; retrieval-augmented generation; chain-of-thought

0 引言

命名实体识别 (Named Entity Recognition, NER) 是自然语言处理 (NLP) 中的关键任务,在信息抽取^[1-2]、知识图谱构建^[3]和智能问答^[4]等应用中发挥着重要作用。近年来,NER 任务逐渐向细粒度发展,不仅关注实体类别,还涉及复杂语境中的实体关系和语义角色。常见的基准数据集有 PropBank^[5]、FrameNet^[6]、OntoNotes^[7]等。

传统的细粒度 NER 方法依赖 BiLSTM-CRF^[8]或预训练模型微调^[9],尤其是在细粒度识别任务中^[10-11],其面临数据依赖性强、计算成本高和泛化能力有限等问题。检索增强生成(RAG)^[12]作为微调替代方案,

通过动态检索外部知识增强模型能力,降低计算成本并提升模型泛化性。其检索方式包括:稀疏检索(如 BM25^[13],基于词频、逆文档频率、倒排索引^[14]等传统技术);密集检索(如 DPR^[15]、BGE^[16]),利用向量表示实现语义匹配。RAG 在问答系统、知识增强生成等任务^[15,17]中表现优异,尤其擅长处理长尾知识、低资源场景和开放域问题,成为提升大模型推理能力的关键策略。其他技术还有提示词工程^[18]和链式思维 (CoT)^[19]。

汉语框架语义解析是一种基于框架语义学理论的细粒度的 NER 任务,旨在通过识别句子中词语所激活的语义框架,分析句子中各个成分的语义角色,从而揭示语言背后的深层语义结构,进一步更好地抽

收稿日期: 2025-10-18 定稿日期: 2025-12-31

基金项目: 北京市自然科学基金 (4262075); 国际关系学院中央高校基本科研业务费专项基金 (3262026T23)

取事件关系和语境信息。当前还存在推理路径简单、准确率不足等问题,调研发现结合 RAG 与 CoT 等技术的大语言模型方法仍较少应用于汉语框架语义解析,且提示词工程在该任务中的优化探索较少。模块化提示词的引入使得模型能够高效利用已有知识,以生成式方式执行 NER 任务;RAG 结合外部检索信息,有助于增强模型对复杂实体及语义框架的理解;CoT 推理进一步提升模型的逻辑推理能力,使其能够更准确地解析实体间的层次关系。

本文的主要创新点如下:

(1) 提出了模块化设计的汉语框架语义解析任务提示词,以提高汉语语义解析任务的可扩展性和适用性。

(2) 首次将 RAG 和 CoT 有效整合,提升了大语言模型在汉语语义解析任务中的推理能力。

(3) 在提示词设计中调整任务表述方式,以降低大语言模型的推理难度。

1 相关工作

框架语义解析是语义角色标注^[20]的一种更细粒度的形式,它基于预定义的语义框架,对句子中的谓词及其论元进行建模。框架语义解析不仅识别谓词-论元结构,还要求识别具体的语义框架(Frame Identification, FI)和框架成分(Frame Elements, FEs)。

研究主要分为基于神经网络和大语言模型两类方法。采用基于神经网络的方法如通过旋转位置编码建模实体间注意力机制^[21],或将框架识别转化为基于词的图解析任务,以提升目标词框架与论元的联合识别能力^[22]。近年来,此类研究开始涉及大语言模型领域,相关研究借助微调小模型帮助大模型完成推理^[23]。在没有依赖微调模型的情况下^[21],通过结合思维链设计不同的提示词,并使用 ChatGPT(gpt-3.5-turbo-1z6k)对 CFSP 任务进行测试;还有通过稀疏检索技术有效筛选相关数据来提升模型性能^[24]的研究。其缺点是思维链构建过于简单,没有深度应用检索增强生成,导致推理结果不佳。针对这一问题,本研究构建了融合 RAG 与 CoT 的模块化提示词,引导模型根据精准的检索内容进行深度思考,同时通过重构简化任务范式,显著提升大语言模型在汉语框架语义解析中的推理性能。

2 方法

2.1 问题描述

本文提出结合提示词工程、RAG 和 CoT 的新范式,以汉语框架语义解析为研究场景,探索大语言模型在细粒度 NER 任务中的应用潜力。当前汉语框架语义解析^[25]包含三个核心子任务:框架识别、论元识别和角色识别。以句子“各种社会保险仍需不断完善”为例,目标词“完善”可激活“使发展”语义框架。

整体任务流程如图 1 所示,在将目标词和原始文本作为输入提供给模型后,要求模型分别输出三种不同的识别结果。

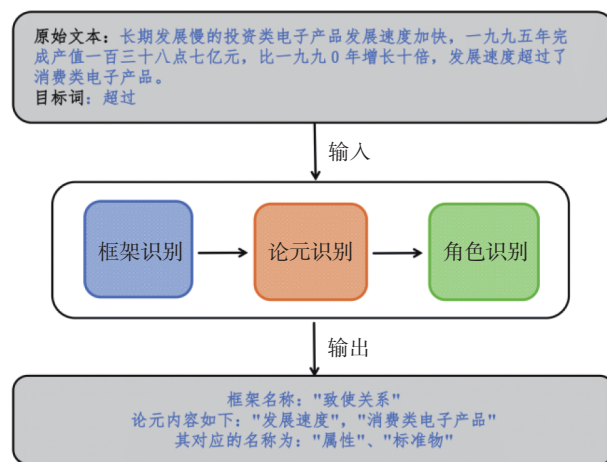


图 1 汉语框架语义解析总体任务流程

(1) 框架识别(Frame Identification, FI)旨在根据目标词在句子中的语境含义,从所有预定义的框架中选择最匹配的框架。框架成分(Frame Elements, FEs),亦称论元,表示与已识别框架相关的语义角色。

(2) 论元识别(Argument Identification, AI)的目标是定位句子中的所有论元,并确定它们的边界。论元可以是句子中的单个词,也可以是一个跨越多个词的短语。

(3) 角色识别(Role Identification, RI)旨在为已识别的论元分配适当的语义角色标签,以明确其在目标框架中的功能。

其中,框架识别的结果信息同时作为论元识别的输入信息,论元识别的结果信息同时作为角色识别的输入信息,构成汉语语义框架识别 pipeline。

2.2 框架识别

针对大语言模型在框架识别任务中对专业概念

理解不足的问题,我们提出了模块化的提示词,在提示词知识补充模块重点优化以下几方面:明确定义框架、论元等核心概念;在其他模块里增加思维引导机

制,提供典型示例。图2是框架识别任务的模块化提示词构建过程,具体公式在附录A.1中,完整提示词及模型输出在附录D中。

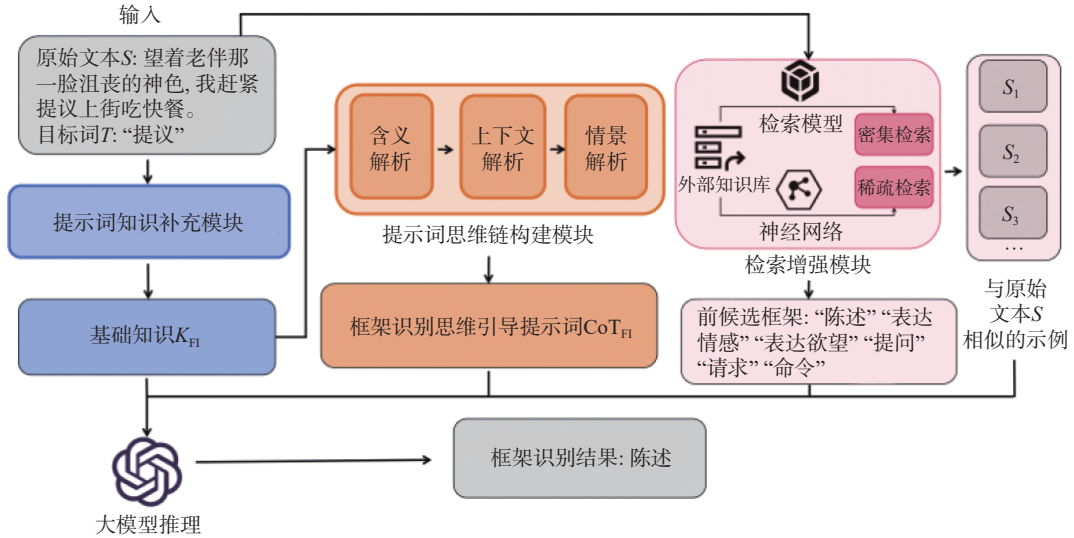


图2 框架识别的模块化提示词

在获取原始文本和目标词后,我们将其输入提示词知识补充模块以获得模型完成任务需要的基础知识 K_{FI} (Knowledge with Frame Identification)。

(1) 提示词知识补充模块: 本文创新性地设计了提示词知识补充模块,通过通俗化的自然语言解释框架、论元等专业概念,有效降低了模型的理解难度。实验证明,该设计显著提升了任务准确率。

然后,将提示词知识补充模块的输出即基础知识 K_{FI} 、原始文本 S 和目标词 T 作为输入提供给提示词思维链构建模块,得到框架识别的思维引导提示词 CoT_{FI} (CoT with Frame Identification)。

(2) 提示词思维链构建模块: 提出了提示词思维链构建模块,通过明确设计完整推理路径,引导模型进行逐步的逻辑推理。该模块显著提升了框架匹配的准确性和可解释性。同时,将同一数据集中的部分样本作为外部知识库,将原始文本 S 和目标词 T 作为检索增强模块的输入,得到与原始文本 S 相似的示例作为 few-shots 内容,并将出现过的框架作为前候选框架让模型在框架中选择一个作为推理结果。

(3) 检索增强模块: 框架种类数量存在不平衡的问题^[26],示例选择的三个维度: 筛选策略、数量控制和语义相关度都会直接影响框架匹配的准确率。示例选择策略对框架识别效果具有重要影响: 虽然多样性优先和频率优先策略相比随机选择略有提升,但效果有限;增加示例数量虽能丰富上下文信息,但受限于模型

的上下文窗口容量并可能引发知识遗忘问题;而通过稀疏检索获取的统计相关示例可显著提升准确率^[24],稠密检索提供的语义相关示例能带来额外性能提升。一方面基于 BM25 算法^[13]实现稀疏检索,通过传统信息检索技术高效匹配关键词;另一方面采用 BGE 模型^[16]进行密集检索,利用深度语义向量捕捉上下文关联性。两种检索方式的公式详见附录C。

(4) 前候选框架: 设计了基于检索增强的候选框架筛选机制,首先通过语义检索获取与输入文本相关的框架集合作为候选答案,有效缩小匹配范围。该方法将框架识别转化为受限的 N 分类问题,显著降低了模型生成非法框架的概率。尽管数据集不平衡可能导致低频框架遗漏,但实验证实候选框架机制仍能提升整体准确率。

最终将提示词知识补充模块输出的基础知识 K_{FI} 、提示词思维引导提示词 CoT_{FI} 、检索增强示例和前候选框架拼接作为输入交由大语言模型推理,得到框架识别结果。

2.3 论元识别

我们同样为论元识别构建了模块化的提示词,如图3所示,在输入原始文本 S 、目标词 T 和框架识别结果后将三者作为知识补充模块的输入。详细公式在附录A.2中,完整提示词和模型输出在附录E中。

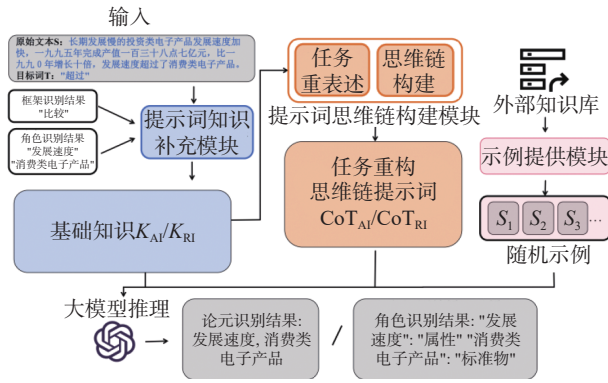


图3 论元识别和角色识别的模块化提示词

(1) 提示词知识补充模块: 提示词知识补充模块采用两阶段增强策略, 首先基于框架识别结果, 从知识库中精准检索该框架的完整定义及所属论元的详细说明; 然后将这些结构化知识整合到提示词中, 为模型提供完备的语义信息。

得到提示词知识补充模块输出的基础知识 K_{AI} (Knowledge with Argument Identification) 后, 将其与原始文本 S 和目标词 T 一并输入思维引导模块得到论元识别的任务重构思维链提示词 CoT_{AI} (CoT with Argument Identification)。

(2) 提示词思维链构建模块: 大语言模型在 NER 任务中表现不及微调小模型的核心原因在于其预训练目标与序列标注任务存在本质差异, 具体包括两点: 缺乏显式的边界标注学习, 以及 token 级生成机制导致实体边界模糊。为此, 引入 GPT-NER^[27] 和 BERT-MRC^[28] 的任务重构思想, 将论元识别转化为文本生成任务: 通过在论元两侧插入 “[start]” “[end]” 特殊标记, 并配合思维链引导, 有效解决了边界识别难题。在 GPT-NER 的研究^[27]中, 通常采用 “## 二元组” 对命名实体进行标记, 我们尝试使用 “[开始重新输出句子][start][end]” 三元组进行标记, 使标记本身具备一定的语义。在生成过程中有效减少了幻觉现象, 对比只使用#输出内容更加完整。

(3) 示例提供模块: 在示例模块中, 我们从同一数据集的知识库随机选取 5~10 个示例为模型提供范例。未采用检索增强的原因是: 首先, 框架识别阶段已通过检索增强获取相关信息, 其输入文本和检索逻辑与角色识别相同; 其次, 若为检索结果都编写思维链需对整个数据集进行细粒度标注, 则工作量过大。

随后将基础知识 K_{AI} 、任务重构思维链提示词 CoT_{AI} 、随机示例作为提示词输入大语言模型进行推理, 大语言模型推理的原始结果如图3所示, 用正则表达式匹配出“亚洲金融危机”“对我们”和“不良影响”。

2.4 角色识别

角色识别任务是汉语框架语义解析中的最终环节。该任务基于前置框架识别提供的语义框架标签和论元识别提取的句内成分, 要求模型在受限语义空间内建立精确的对应关系, 将传统的 NER+N 分类复合任务解耦为分步实施的 N 分类序列操作。其核心挑战在于跨框架的语义角色对齐, 需借助框架/论元识别的先验知识缩小候选范围。这种渐进式分类流程实现了从细粒度 NER 到结构化生成任务 (N 分类-文本生成任务- N 分类) 的系统性优化。

其构建流程如图3所示, 将框架与论元识别结果结合原始文本 S 与目标词 T , 经提示词知识补充模块输出为基础知识 K_{RI} (Knowledge with Role Identification); 后续的思维链构建与示例提供模块继承自 2.2 节, 对应的输出为任务重构思维链提示词 CoT_{RI} (CoT with Role Identification)。将三个模块的输出作为大语言模型的输入, 得到相应的推理结果, 即论元所属的角色。具体公式在附录 A.3 中, 详细的提示词和模型输出在附录 F 中。其中, 提示词知识补充模块与 2.3 节的不同在于其额外整合论元识别结果输入大模型, 而在思维链构建中, 将任务重表述为对识别论元到可能论元名称的 N 分类任务。实验表明, 子任务缺失将显著削弱角色识别性能。该设计通过缩小推理范围、降低难度, 有效提升整体表现。多阶段任务分解可降低推理复杂度并提升完成度。

3 实验

为了验证方法的有效性, 本文在 CFN 2.1^[29] 上进行实验, 涵盖多语境实体、多类论元及复杂语义框架, 支持框架识别、论元识别与角色识别三任务。对比方法包括基于 ChatGPT (gpt-3.5-turbo-1z6k) 与思维链的 baseline^[29], 以及采用稀疏检索增强生成与输入格式化技术的 PKU 方法^[24]。

我们使用 api 进行 GPT 系列模型的测试, 使用一张 NVIDIA A800-SXM4 80GB 作为算力支持本地部署 Qwen 系列^[30]模型, BGE-M3^[16]+Deepseek-R1-distill-Qwen-32B^[31]推理约占用 65 GB 显存, Temperature=0.7, top_p=0.8。

检索增强部分的实验设置如下:

(1) 外部知识库的构建: 随机抽取 CFN 2.1 的 20% 作为推理测试数据, 80% 数据作为外部知识库, 按照每一句文本及其对应论元作为一个检索分片。

(2) 检索实现细节: BM25 为默认超参数($K_1=1.2$, $b=0.75$), BGE^[6]为 BAAI 的未微调 BGE-M3(MaxLength=8 192)。

实验设置涵盖模型选择、RAG 检索库构建与提示词设计, 系统评估各技术的有效性。框架识别任务采用准确率(ACC), 论元与角色识别任务采用精确率(P)、召回率(R)与 F_1 分数, 并辅以定性分析, 评估大模型在语义解析中的表现与优化潜力。各任务运行三次取均值作为最终结果。评估指标的具体计算方式见附录 B。

如表 1 所示, 在三种大语言模型上, 对比评估了 baseline、PKU 两种方法。在框架识别(FI)、论元识别(AI)与角色识别(RI)三个任务指标上, 均展现出显著的性能优势。本文方法在框架识别任务中, 在所有模型上均显著优于 Baseline 与 PKU 方法, Qwen1.5-32B 模型准确率(ACC)最高达 67.47%。在论元识别任务中 F_1 值领先三种模型, GPT-3.5turbo 模型达 59.89%, 优于 Baseline 的 32.90% 和 PKU 的 58.99%。在角色识别任务中, 本文方法同样表现优越, 其 F_1 值在所有模型上显著高于对比方法。

表 1 各个模型下的任务完成情况 (单位: %)

模型	方法	FI	AI			RI		
		ACC	P	R	F_1	P	R	F_1
GPT-3.5-turbo	Baseline	53.00	60.98	22.52	32.90	6.38	7.59	6.93
	PKU	52.54	52.18	67.85	58.99	14.53	19.52	16.66
	ours	64.88	78.42	48.45	59.89	20.95	16.50	18.46
Qwen-1.5-32B	Baseline	25.60	45.50	19.67	27.46	2.83	1.94	2.30
	PKU	53.95	44.20	60.35	51.02	11.45	16.56	13.53
	ours	67.47	72.20	42.20	53.26	21.87	16.20	18.62
DeepSeek-R1-distill-Qwen-32B	Baseline	29.54	47.15	42.25	44.56	7.94	3.65	5.00
	PKU	52.16	42.11	62.33	50.26	17.23	21.40	18.99
	ours	65.04	54.56	56.56	55.54	29.54	23.97	26.46

本文方法在各项指标上稳定且具一定泛化能力, 特别是在 QWEN 与 Deepseek -R1-distill-Qwen-32B 等开源模型上的出色表现, 验证了其在汉语框架语义解析任务中的有效性与优越性。

我们评估了不同 few-shot 选择策略对 Qwen1.5-32B 模型完成 FI 任务性能的影响, 如图 4、图 5 所示, 性能指标为准确率(accuracy)。随机选择 few-shot 示例提升模型表现, 提供 10 个随机示例提高 5.13%, 100 个示例提升至 38.87%, 效果逐渐平稳。BM25 检索可以在 10-shots 下直接达到 38.59, 加入候选框架和 COT 后进一步提升至 42.89。单独使用并分别计算 COT 和候选框架的贡献时, COT 使随机与 BGE-10shots 分别提升 3.27 和 4.66。而候选框架使 BGE-10shots 性能提升至 62.77(+19.52), 组合使用后达到 67.43。COT 引入增强了模型能力, 但不及候选框架带来的提升, 可能是因为 COT 依赖示例质量, 不相关示例的思维引导无法带来效果的本质提升。

如图 5 所示, 检索选择优于随机选择。BM25-10shots 结合候选框架 COT 性能达 42.89, BGE-10shots 结合候选框架 COT 后达 67.43, 明显优于 BM25(63.95%)。

进一步增加示例数量(bge 20shots)仅带来 0.04% 的微小增益, 验证示例质量提升后数量增加对效果影响减小。实验表明, few-shot 示例能有效提升准确率, 基于检索的方法优于随机选择, 且密集检索相较稀疏检索更具优势。结合候选框架的示例筛选机制显著增强模型性能, CoT 在示例质量较低时有助于推理能力提升。增加示例数量性能虽有提升, 但 10 shots 结合合适的检索与筛选已接近最优效果。

我们进行了如表 2、表 3 所示的对论元识别和角色识别的消融实验, 在提示词中提供 10 个随机示

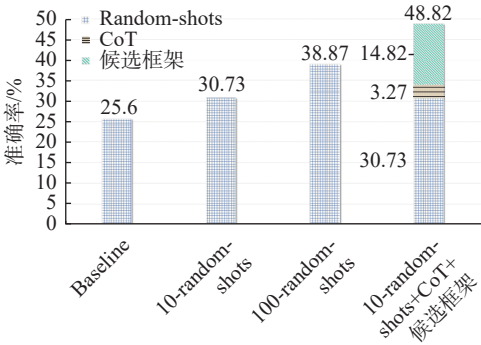


图 4 在不使用检索的情况下各方法对 FI 任务的影响

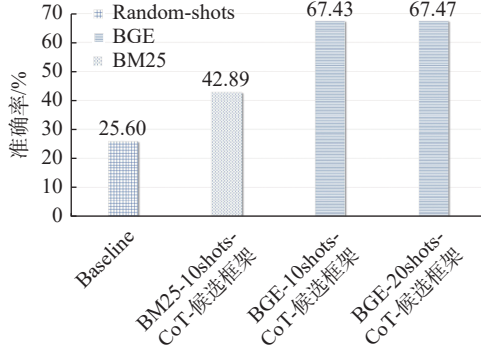


图 5 在使用检索的情况下各方法对 FI 任务的影响

例。其中, ours 是我们模块化提示词的方法。FI、AI 的增加是提供论元识别和角色识别的结果。

表 2 方法(Qwen1.5-32B)性能对比(AI)

(单位: %)

方法	AI		
	P	R	F_1
Baseline	45.50	19.67	27.46
ours	72.20	42.20	53.26
ours+FI	65.02	35.96	46.30

表 3 方法(Qwen1.5-32B)性能对比(RI)

(单位: %)

方法	RI		
	P	R	F_1
Baseline	2.83	1.94	2.30
ours	3.54	1.86	2.43
ours+AI	5.25	1.88	2.76
ours+FI	23.16	13.00	16.65
ours+FI+AI	21.87	16.20	18.62

我们发现,任务重表述对 AI 任务产生了较大影响, P 、 R 和 F_1 分数均大幅提升,特别是 F_1 提升至 53.26%,比 Baseline 高约 25.8%。然而,结合 FI 结果后, F_1 下降至 46.30%,原因包括:错误累积:FI 结果中的错误传递到 AI 任务,由于框架识别准确率较低,反而负面影响了论元识别;任务表述兼容性问题:句子框架及论元名称信息未被模型采纳,且内容过多占用上下文窗口,影响模型性能。

提供 FI 和 AI 结果相比 Baseline 显著提升 RI 任务的 F_1 分数。仅提供 AI 结果时,模型仍需在大量标签中分类,效果与未提供子任务结果相近。这表明,将框架语义解析任务拆解为有序子任务,并优化前两个子任务对 RI 任务具有提升作用,尤其是 FI 任务。这

是因为输入信息不会被完全依赖,而是基于自身语义理解进行调整,校正错误的论元识别结果,生成更合理的输出。相比之下,角色识别任务更依赖框架识别结果提供的筛选候选论元,框架识别对角色识别任务更为关键。

4 结论

本文深入探索细粒度 NER 领域的汉语框架语义解析,在提示词工程的角度提出了结合 RAG 和 CoT 的模块化方法,以提升大模型面对这类问题的推理能力。首先,提出了针对汉语框架语义解析的模块化的提示词,这一设计显著提高了任务的可扩展性和适用性。其次,首次结合了 RAG 和 CoT 技术,有效地增强了大模型在推理过程中的能力,还对任务类型进行了转化,并在提示词设计中调整了任务的重表述方式,这一改进显著降低了大模型的推理难度。同时,本文进行了大量的实验工作验证方法的有效性,为大模型在汉语框架语义解析提供了解决方案。

附录 A 子任务公式

A.1 框架识别子任务公式

框架识别提示词知识补充模块在大语言模型推理时的公式如下:

$$K_{FI} = \operatorname{argmax}_y P(y | T, S, \Pi) \quad (\text{A-1})$$

其中, S 是原始文本, T 是目标词, K_{FI} 是基础知识, “ Π ”表示大语言模型。

框架识别提示词思维链构建模块大语言模型推理时的思维链公式如下:

$$c_1 = \operatorname{argmax}_y P(y | T, S, K_{FI}, \Pi)$$

$$c_i = \operatorname{argmax}_y P(y | c_j, T, S, K_{FI}, \Pi), 0 < j < i < 4 \quad (\text{A-2})$$

$$\text{CoT}_{FI} = \operatorname{argmax}_y P(y | c_1, c_2, c_3, K_{FI}, \Pi)$$

其中, S 为原始文本, T 为目标词, c_1 是对目标词 T 的初步语义解释, c_2 是对 T 在上下文 S 中的角色语义抽取, c_3 是整个句子的情景及事件结构分析结果。 K_{FI} 是提示词知识补充模块输出的基础知识, CoT_{FI} 是框架识别提示词, “ Π ”表示大语言模型。

A.2 论元识别子任务公式

论元识别提示词知识补充模块在大语言模型推

理时的公式如下：

$$K_{AI} = \underset{y}{\operatorname{argmax}} P(y | T, S, FI, \Pi) \quad (A-3)$$

其中, S 是原始文本, T 是目标词, K_{AI} 是基础知识, FI 是框架识别结果, “ Π ” 是大语言模型。

得到知识补充模块输出的基础知识后, 将其与原始文本 S 和目标词 T 一并输入思维引导模块得到思维链提示词。

论元识别提示词思维引导模块公式如下：

$$\begin{aligned} c_1 &= \underset{y}{\operatorname{argmax}} P(y | T, S, K_{AI}, \Pi) \\ c_i &= \underset{y}{\operatorname{argmax}} P(y | c_j, T, S, K_{AI}, \Pi), 0 < j < i < 4 \quad (A-4) \\ CoT_{AI} &= \underset{y}{\operatorname{argmax}} P(y | c_1, c_2, c_3, K_{AI}, \Pi) \end{aligned}$$

其中, K_{AI} 是提示词知识补充模块输出的基础知识, T 是目标词, S 是原始文本, CoT_{AI} 是任务重表述提示词, c_1 是所给论元和 T 的初步语义解释, c_2 是提取 T 在 S 中的上下文信息, c_3 是情景分析和角色定位, “ Π ” 是大语言模型。

A.3 角色识别子任务公式

角色识别提示词知识补充模块在大语言模型推理时的公式如下：

$$K_{RI} = \underset{y}{\operatorname{argmax}} P(y | T, S, FI, AI, \Pi) \quad (A-5)$$

其中, S 是原始文本, T 是目标词, K_{RI} 是基础知识, FI 是框架识别结果, AI 是论元识别结果, “ Π ” 是大语言模型。

附录B 评估计算公式

框架识别任务旨在判断给定句子中的目标词 (Target Word) 所对应的语义框架。我们采用 ACC 作为评估指标：

$$ACC(FI) = \frac{\text{correct}}{\text{total}} \quad (B-1)$$

其中, correct 表示模型预测正确的框架数量, total 表示待识别的框架总数量。

论元识别任务的目标是预测句子中属于特定框架的论元 (Argument), 并正确识别其文本范围 (Span)。我们采用精确率 (Precision, P)、召回率 (Recall, R) 和 F_1 分数作为评估指标, 计算公式如下：

$$\text{Precision} = \frac{\text{InterSec}(\text{gold}, \text{pred})}{\text{Len}(\text{pred})} \quad (B-2)$$

$$\text{Recall} = \frac{\text{InterSec}(\text{gold}, \text{pred})}{\text{Len}(\text{gold})} \quad (B-3)$$

$$F_1(AI) = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (B-4)$$

其中, gold 表示真实的论元范围字符串集合, pred 表示模型预测的论元范围字符串集合。InterSec(gold, pred) 表示真实结果和预测结果之间的重叠 token 数, Len($\cdot$) 表示计算 token 数量。

论元角色识别任务的目标是为已识别的论元范围分配正确的语义角色标签。由于每个论元可能对应多个语义角色, 我们采用精确率 (Precision, P)、召回率 (Recall, R) 和 F_1 分数作为评估指标, 计算公式如下：

$$\text{Precision} = \frac{\text{Count}(\text{gold} \cap \text{pred})}{\text{Count}(\text{pred})} \quad (B-5)$$

$$\text{Recall} = \frac{\text{Count}(\text{gold} \cap \text{pred})}{\text{Count}(\text{gold})} \quad (B-6)$$

$$F_1(RI) = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (B-7)$$

其中, gold 和 pred 分别表示真实结果与预测结果的角色集合, Count($\cdot$) 计算集合元素的数量, 即正确匹配的角色数量。

附录C 检索的核心公式

BM25 是信息检索领域的一种经典打分函数, 能够有效衡量文档与查询之间的相关性, 核心公式如下：

$$BM25(D, Q) = \sum_{t \in Q} IDF(t) \cdot \frac{f(t, D) \cdot (k_1 + 1)}{f(t, D) + k_1 \cdot \left(1 - b + b \cdot \frac{|D|}{\text{avgdl}}\right)} \quad (C-1)$$

其中,

Q 是将待查询文本拆分成词项的集合;

D 是根据同一数据集的部分作为外部知识库的文档;

t 是 Q 集合中的一个词项;

$f(t, D)$ 是词 t 在文档 D 中的出现次数;

$|D|$ 是文档 D 的长度;

avgdl 是整个文档集合的平均文档长度;

k_1 和 b 是超参数;

IDF(t) 是逆文档频率, 计算方式如下:

$$IDF(t) = \log \left(\frac{N - n_t + 0.5}{n_t + 0.5} + 1 \right) \quad (C-2)$$

其中, N 是外部知识库的总文档数量, n_t 是包含词 t 的文档数量, +0.5 是平滑项, 可以缓解小样本偏差。

BGE 检索的核心公式如下, 包含编码-相似度计算-排序的过程:

$$D^* = \underset{D_i \in \mathcal{D}}{\operatorname{argmax}} \frac{\text{Encoder}(q) \cdot \text{Encoder}(D_i)}{\|\text{Encoder}(q)\| \cdot \|\text{Encoder}(D_i)\|} \quad (C-3)$$

其中, q 是查询文本; $\mathcal{A} = \{D_1, D_2, \dots, D_n\}$ 是候选文档集合; Encoder 是 BGE 模型, 用于将查询和文档转换为嵌入向量 embeddings; D^* 是 D_i 的相似度结果, 可以取 top_k 。

附录 D 框架识别提示词及模型输出

提示词知识补充模块结果

语义框架用来描述事件里的人或事、动作, 以及相关的属性, 语义框架就像是给句子贴上标签, 它把句子分类, 也能概括句子的信息。用语义框架, 我们可以快速了解句子的主要意思, 知道谁做了什么事, 以及这件事有什么属性。例如我会给出句子中的一个词, 需要首先分析这个词在句子中上下文的语境并理解这个词的谓语都有什么, 它和这些谓语共同组成句子的主干, 解释了句子中发生的事件。

提示词思维链构建模块结果

举个例子, 假如我们有句子: “此外, 伦敦近年来因家庭暴力引发的凶杀案件数量也有增无减, 从 2014 年的 18 起上升到 2019 年的 23 起。” 我给定“到”, 你可以在句子中看到某个数值的变化, 你发现了发生量变的实体, 实体的什么属性发生了量变, 量变的初始值和最终值等, 在候选框架中发现“量变”框架与句子匹配, 因为它描述了事件参与者中的实体: “伦敦近年来因家庭暴力引发的凶杀案件”, 属性: “数量”, 初状态: “2014 年的 18 起”, 终状态: “2019 年的 23 起”等属性。

再举个例子, 假如我们有句子: “传统的宣纸生产工艺, 既是前人生产经验的总结, 也体现了前人对客观世界的认知。” 我给定“是”, 这个词在句子中的含义是什么、属于哪个类”。在候选框架中“属于某类”与句子匹配, 因为它描述了事件参与者中的实体: “传统的宣纸生产工艺”以及范畴: “前人生产经验的总结”。表明“传统的宣纸生产工艺”这个实体属于“前人生产经验的总结这个范畴”

你需要理解句子中发生了什么事件, 重点关注我给出词语的意思, 在我所给定的候选框架中匹配一个最能描述目标词语义以及概括句子发生的事件。

提供句子、目标词和前候选框架结果

现在有一个句子:

“化学工业部部长×××在会议所作的主旨报告中

指出, 中国今年化工行业尽管面临多项减利因素、效益下滑的严峻形势, 但以深化国有企业改革为突破口, 较好完成了全年生产和建设任务。”

你需要分析上下文语境的词是:

“作”

候选框架为:

[“文本创作”, 有意行为, “属于某类”]

在这些框架内匹配一个最合适的框架名。只需要回答框架的名称, 不要添加其他内容。

输出结果:

框架名称

检索增强模块结果

example:

句子:

“国家委员会副主任、外经贸部部长×××在会议上就中国与联合国发展业务系统之间的合作作了发言。”

我给定词语: “作”

输出结果:

文本创作

example:

句子:

“《决定》总结了二十年农村改革的基本经验; 确定了从现在起到 2010 年中国农村经济、政治、文化建设的目标和任务; 阐述了社会主义初级阶段农村尤其不发达的主要特征; 对中国农村基本经营制度作了更科学、更准确的新概括; 提出坚持以公有制为主体、多种所有制经济工作的领导提到了一个新的高度。”

我给定词语: “作”

输出结果:

有意行为

example:

句子:

“9 美国田径名将(花蝴蝶)×××X 9 月因心脏病发作逝世, 她曾于 8 月访问香港, 并在香港体育学院作技术辅导;”

我给定词语: “作”

输出结果:

属于某类

请使用这些示例来帮助理解, 按照示例中的输出格式选择候选框架中最有可能的框架名称, 不要输出不属于候选框架的结果, 不要添加额外内容, 只输出框架名称。

模型输出(DeepSeek-distill-qwen-32B)

好的,我现在需要分析句子中的词语“作”在给定的句子中的上下文,并选择合适的语义框架。句子是:“化学工业部部长×××在会议所作的主旨报告中指出,中国今年化工行业尽管面临多项减利因素、效益下滑的严峻形势,但以深化国有企业改革为突破口,较好完成了全年生产和建设任务。”

首先,我注意到“作”出现在“会议所作的主旨报告”中。这里的“作”是指×××部长在会议上做了主旨报告。主旨报告通常是一种正式的演讲或陈述,属于文本创作的范畴。

接下来,我查看候选框架: ['有意行为', '属于某类', '文本创作']。根据例子,当“作”涉及撰写报告或演讲时,通常会被归类为“文本创作”。例如,之前的例子中,当“作”用于报告或演讲时,选择“文本创作”框架。

因此,这里“作”指的是撰写或发表主旨报告,属于文本创作的行为。

</think>

文本创作

附录E 论元识别提示词及模型输出

提示词知识补充模块结果

你需要从我提供的句子中,使用指定的特殊标识符,标注出句子中目标词的论元。论元是指谓词或动词所依赖的必要成分,使句子完整且符合语法。目标词是句子中的核心行为词,如动词或谓词。论元和目标词在语义上相关,共同构成句子的关键信息,论元通常内部是完整的,如果用了从句,论元可以是一句话中的一个小短句。它们同属于一个语义框架。

提示词思维链构建结果

根据给定的句子和其中的目标词在句子中的语义场景,将句子重新输出,句子开始的标志为“[开始重新输出句子]”,标记目标词的每个论元时使用[start]作为初始位置,[end]作为结束位置。[start]与[end]作为一对标识符标记目标词的一个论元。

让我们一步步思考,给出你对句子含义的理解,句子中发生了什么事情,目标词在句子中是什么成分,与目标词有关系的成分有哪些,分析在句子中能

找到的论元有哪些,然后把论元的完整内容用[start]与[end]标记出来,标记出来的论元在语义上必须完整,尽量保证找到的论元准确性以及找到全部的论元,这里论元的范围非常广泛,通常论元数量不少于3个。

下面的被标记的句子内容应该包括完整的[start]和[end]对,你要在给出最终结果前检查。

输出格式:

目标词的含义是什么,在提供的可能论元中哪些出现了,选择这些论元的理由。

[开始重新输出句子]被标记的句子内容。

提供句子、目标词和框架识别结果

现在有一个句子:

“化学工业部部长×××在会议所作的主旨报告中指出,中国今年化工行业尽管面临多项减利因素、效益下滑的严峻形势,但以深化国有企业改革为突破口,较好完成了全年生产和建设任务。”

它的目标词是:

“作”

这个句子的框架名是文本创作,其中可能包含以下论元,我将解释论元及其定义:

- “作者”: 作者是创造具体文本的人。
- “文本”: 由写或说的行为而产生的实体。
- “信息接收者”: 指传达信息的对象。
- “成分”: 成分是作者用来构建文本的小文本。
- “形容”: 描述参与者的短语。
- “形式”: 文本被表达的形式。
- “受众”: 受众表明文本为谁而写。
- “工具”: 作者创作文本所使用的工具,如用钢笔写/用墨水写。

...(省略全部的论元)

注意,你找到的论元必须是符合以上定义的论元,如果句子中出现了符合以上定义的部分内容,一定要找到。

示例提供模块结果

这里有几个示例:

example:

句子:

“八五”期间,广东电子工业投资七十亿元进行企业改造和发展,目前已有二十个重点项目竣工投产,使四大类产品形成了规模经济,其中程控交换机、微

机已上规模,成为电子工业新的支柱产品。”

目标词:

“形成”

总结一下,根据目标词的上下文语境,是某个事件导致形成了某个情况,目标词“形成”的论元应该包含时间“八五”期间、实体“规模经济”和致因“广东电子工业投资七十亿元进行企业改造和发展,目前已有二十个重点项目竣工投产”,在“八五”期间“这个时间由于”广东电子工业投资七十亿元进行企业改造和发展,目前已有二十个重点项目竣工投产的原因导致实体“规模经济”的形成。

输出格式:

[开始重新输出句子][start]“八五”期间[end], [start]广东电子工业投资七十亿元进行企业改造和发展[end],目前已有二十个重点项目竣工投产,使四大类产品形成了[start]规模经济[end],其中程控交换机、微机已上规模,成为电子工业新的支柱产品。

请使用这些示例来帮助理解,发挥你的想象力,尽可能精准识别目标词的全部论元,按照示例中的输出格式输出,不要添加额外内容。

模型输出(DeepSeek-distill-qwen-32B)

好的,我现在需要分析用户提供的句子,找出目标词“作”的论元。首先,句子是:“化学工业部部长×××在会议所作的主旨报告中指出,中国今年化工行业尽管面临多项减利因素、效益下滑的严峻形势,但以深化国有企业改革为突破口,较好完成了全年生产和建设任务。”目标词是“作”。

首先,我需要明确“作”在这里的意思,它是指“制作”或“撰写”报告。接下来,我要根据提供的论元列表,看看句子中有哪些部分符合这些论元的定义。

1. **作者**: 谁在作报告? 显然是“化学工业部部长×××”,所以她就是作者。

2. **文本**: 作的是什么? 是“主旨报告”,所以文本是“主旨报告”。

3. **处所**: 作报告是在哪里进行的? 句子中提到是在“会议”上,所以处所是“在会议”。

4. **目的**: 作报告的目的是什么? 是为了“指出”后面的情况,即“中国今年化工行业尽管面临多项减利因素、效益下滑的严峻形势,但以深化国有企业改革为突破口,较好完成了全年生产和建设任务。”所以这部分是目的。

5. **方式**: 是否有说明作的方式? 句子中没有

提到工具或方法,所以可能没有方式论元。

6. **时间**: 是否有明确的时间? 句子提到“今年”,但“今年”是化工行业的背景,不是作报告的时间,所以可能没有时间论元。

7. **其他论元**: 是否有其他如工具、材料等? 句子中没有提到,所以暂时不考虑。

总结一下,论元包括作者、文本、处所和目的。现在需要将这些部分用[start]和[end]标出。

检查句子结构,确保每个论元都完整且准确。特别是“目的”部分,虽然比较长,但确实是为了传达后面的信息,所以应包含在内。

最后,确保输出格式正确,使用[开始重新输出句子]开始,并正确标注每个论元。

</think>

总结一下,根据目标词的上下文语境,重点分析“作”这个动作的作者是谁,作的是什么,以及在什么场合下作的。目标词“作”的论元应该包含作者“化学工业部部长×××”、文本“主旨报告”和处所“在会议”

[开始重新输出句子][start]化学工业部部长×××[end][start]在会议[end]所作的[start]主旨报告[end]中指出,中国今年化工行业尽管面临多项减利因素、效益下滑的严峻形势,但以深化国有企业改革为突破口,较好完成了全年生产和建设任务。

附录F 角色识别提示词及模型输出

提示词知识补充模块结果

你是一个框架语义学家。框架语义学认为,语言中的词义依赖于认知框架,即一组与该词相关的情境和语义角色(论元)。论元是框架中的核心成分,表示特定的语义角色,如事件的施事、受事、时间、地点等。给定一个句子和一个目标词,你需要完成以下任务:确定目标词所激活的框架。识别该框架的论元及其语义定义。为句子中对应的论元标明它们的对应关系。

提示词思维链构建模块结果

举个例子,假如我提供句子为:由于由签署契约至把契约送交该处登记,通常相隔一段时间,因此,六月份的注册数字,一般显示送交登记前约四个星期内签署契约的物业交易。与句子中“送交”这个词有联系的成分的论元为[‘契约’,‘该处’,‘登记’],然后我提供

的所属框架论元名称有['文件','官方','意图目的',等等],你需要将论元贴上所属框架论元名称的标签最终得到'契约'是'文件','该处'是'官方','登记'是'意图目的'的结果,然后按照格式输出。

需要按照输出格式使用 python 元组输出成分的名称,不要添加其他内容。

输出格式:

[[句子中论元1的内容,论元1的名称],[句子中论元2的内容,论元2的名称],[句子中论元3的内容,论元3的名称]...]

需要按照输出格式使用 python 元组输出成分的名称,不要添加其他内容。

提供句子、目标词,框架识别结果和论元识别结果

给定一个句子:

“化学工业部部长×××在会议所作的主旨报告中指出,中国今年化工行业尽管面临多项减利因素、效益下滑的严峻形势,但以深化国有企业改革为突破口,较好完成了全年生产和建设任务。”

目标词为

作

其所属的框架和论元信息如下:

这个句子的框架名是文本创作,其中可能包含以下论元,我将解释论元及其定义:

- “作者”: 创造具体文本的人。
- “文本”: 由写或说的行为而产生的实体。
- “信息接收者”: 指传达信息的对象。
- “成分”: 作者用来构建文本的小文本。
- “形容”: 描述参与者的短语。
- “形式”: 文本被表达的形式。
- “受众”: 文本为谁而写。

...(省略全部论元)。

已识别的文本中论元如下:“化学工业部部长×××”,“主旨报告”,“在会议”。

对于这些论元在其所属的框架和论元信息中找到与其语义对应的论元名称与之对应。

示例提供模块结果

这里有几个示例帮助你理解任务工作:

example:

句子:

“八五”期间,广东电子工业投资七十亿元进行企业改造和发展,目前已有二十个重点项目竣工投产,

使四大类产品形成了规模经济,其中程控交换机、微机已上规模,成为电子工业新的支柱产品。

目标词:

“形成”

根据目标词的上下文语境,目标词“形成”的论元应该包含时间““八五”期间”、实体“规模经济”和致因“广东电子工业投资七十亿元进行企业改造和发展,目前已有二十个重点项目竣工投产”,在“八五”期间“这个时间由于”广东电子工业投资七十亿元进行企业改造和发展,目前已有二十个重点项目竣工投产”的原因导致实体“规模经济”的形成。

输出格式:

[[“八五”期间,“时间”],[“规模经济”,“实体”],[“广东电子工业投资七十亿元进行企业改造和发展,目前已有二十个重点项目竣工投产”,“致因”]]

example:

句子:

“由于由签署契约至把契约送交该处登记,通常相隔一段时间,因此,六月份的注册数字,一般显示送交登记前约四个星期内签署契约的物业交易。”

目标词:

“送交”

“根据目标词的上下文语境,目标词“送交”的论元应该包含文件“把契约”、官方“该处”和意图目的“登记”即为了登记把契约送交给该处。

输出格式:

[[“把契约”,“文件”],[“该处”,“官方”],[“登记”,“意图目的”]]

模型输出(DeepSeek-distill-qwen-32B)

好的,我现在要解决的问题是根据给定的句子和目标词“作”,确定其所属的框架,并识别该框架中的论元及其对应的语义角色。目标词“作”在句子中的使用情况如下:

句子:“化学工业部部长×××在会议所作的主旨报告中指出,中国今年化工行业尽管面临多项减利因素、效益下滑的严峻形势,但以深化国有企业改革为突破口,较好完成了全年生产和建设任务。”

目标词:“作”

首先,我需要确定目标词“作”所属的框架。根据框架语义学,每个词都有其激活的框架。在这个句子中,“作”出现在“在会议所作的主旨报告中”,这里“作”指的是“制作”或“撰写”报告。因此,目标词

“作”激活的框架应该是“文本创作”。

接下来,我需要识别该框架中的论元及其语义定义。根据提供的论元列表,可能的相关论元包括:

- “作者”: 创造具体文本的人。
- “文本”: 由写或说的行为而产生的实体。
- “信息接收者”: 传达信息的对象。
- “工具”: 作者创作文本所使用的工具。
- “方式”: 描述创作活动的特征或状态。
- “目的”: 作者创作文本的目的。
- “时间”: 文本被创作的时间。
- “地点”: 作者创作文本的地点。

现在,我需要分析句子中与“作”相关的部分,找出对应的论元。

句子中的相关部分是“化学工业部部长×××在会议所作的主旨报告”。这里,“化学工业部部长×××”是制作报告的人,即“作者”。“主旨报告”是被制作的文本,即“文本”。“在会议”表示制作报告的地点,即“处所”。

因此,句子中的论元如下:

1. “化学工业部部长×××”对应“作者”。
2. “主旨报告”对应“文本”。
3. “在会议”对应“处所”。

其他可能的论元如“信息接收者”“工具”“方式”等在句子中没有明确提及,因此不需要考虑。

最后,按照输出格式,将这些论元及其对应的名称列出。

</think>

[[化学工业部部长×××, 作者], [主旨报告, 文本], [在会议, 处所]]

参考文献

- [1] HUANG J, LI C, SUBUDHI K, et al. Few-shot named entity recognition: An empirical baseline study[C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2021: 10408-10423.
- [2] QIAN J, LIU Y, LIU L, et al. Fine-grained entity typing without knowledge base[C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2021: 5309-5319.
- [3] HU S, WANG Z, ZHANG B, et al. Data augmentation with knowledge graph-to-text and virtual adversary for specialized-domain Chinese NER[C]// Proceedings of the International Joint Conference on Neural Networks. 2024: 1-8.
- [4] LIU A, XIAO W, ZHU H, et al. QaNER: Prompting question answering models for few-shot named entity recognition[J/OL]. arXiv preprint. arXiv: 2203.01543, 2022.
- [5] PALMER M, GILDEA D, KINGSBURY P. The proposition bank: An annotated corpus of semantic roles[J]. Computational Linguistics, 2005, 31(1): 71-106.
- [6] COLLIN F. baker. FrameNet: A knowledge base for natural language processing[C]//Proceedings of Frame Semantics in NLP: A Workshop in Honor of Chuck Fillmore, Baltimore, MD, USA, 2014: 1-5.
- [7] PRA DHAN S, HOVY E, MARCUS M, et al. OntoNotes: A unified relational semantic representation[C]// Proceedings of the International Conference on Semantic Computing, 2007: 517-526.
- [8] HUANG Z, XU W, YU K. Bidirectional LSTM-CRF models for sequence tagging[J/OL]. arXiv preprint. arXiv: 1508.01991, 2015.
- [9] JIAO X, YIN Y, SHANG L, et al. TinyBERT: Distilling BERT for natural language understanding[C]// Proceedings of the Association for Computational Linguistics: EMNLP, 2020: 4163-4174.
- [10] 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.
- [11] LAI Y, LIU Y, FENG Y, et al. Lattice-BERT: Leveraging multi-granularity representations in Chinese pre-trained language models[C]// Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021: 1716-1731.
- [12] LEWIS P, PEREZ E, PIKTUS A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems, 2020: 9459-9474.
- [13] ROBERTSON S E, WALKER S, JONES S, et al. Okapi at TREC-3[M]//HARMAN, D.K NIST special publication. Gaithersburg: National Institute of Standards and Technology, 1995: 109-123.
- [14] BAEZA YATES R, RIBEIRO NETO B. Modern information retrieval: The concepts and technology behind search[M]. 2nd ed. Harlow: Addison-Wesley; 2011:913.
- [15] KARPUKHIN V, OGUZ B, MIN S, et al. Dense passage retrieval for open-domain question answering[C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2020: 6769-6781.
- [16] CHEN J, XIAO S, ZHANG P, et al. M3-Embedding: Multi-linguality, multi-functionality, multi-granularity text embeddings through self-knowledge distillation[C]// Proceedings of the Association for Computational Linguistics: ACL, 2024: 2318-2335.

- [17] ZHANG T, PATIL S, JAIN N, et al. RAFT: Adapting language model to domain specific RAG[J/OL]. arXiv preprint. arXiv: 2403.10131, 2024.
- [18] BROWN T, MANN B, RYDER N, et al. Language models are few-shot learners[C]// Proceedings of the 34th Conference on Neural Information Processing Systems, 2020: 1877-1901.
- [19] WEI J, WANG X, SCHUURMANS D, et al. Chain-of-thought prompting elicits reasoning in large language models[C]// Proceedings of the 36th Conference on Neural Information Processing Systems, 2022: 24824-24837.
- [20] CHEN H, ZHANG M, LI J, et al. Semantic role labeling: A systematic survey[J/OL].arXiv preprint. arXiv: 2502.08660, 2025.
- [21] LI Z, GUO X, QIAO D, et al. System report for CCL23-eval task 3: Application of entity classification model based on rotary position embedding in Chinese frame semantic parsing[C]// Proceedings of the 22nd Chinese National Conference on Computational Linguistics, 2023: 94-104.
- [22] LIU Y, LI Z, ZHANG M. System report for CCL23-eval task3: SUDA CFSP system[C]// Proceedings of the 22nd Chinese National Conference on Computational Linguistics, 2023: 84-93.
- [23] LIU Y, GONG C, ZHANG M. Leveraging LLMs for Chinese frame semantic parsing[C]// Proceedings of the 23rd Chinese National Conference on Computational Linguistics, 2024: 21-31.
- [24] HUANG F, GUO Z, LI W, et al. Construction of CFSP model based on non-finetuning large language model[C]// Proceedings of the 23rd Chinese National Conference on Computational Linguistics, 2024: 1-9.
- [25] YANG P, LI J, YAN Z, et al. Chinese frame semantic parsing evaluation[C]// Proceedings of the 23rd Chinese National Conference on Computational Linguistics, 2024: 32-42.
- [26] GUAN Y, LIU X, ZHANG L, et al. System report for CCL23-eval task 3: UIR ISC pretrained language model for Chinese frame semantic parsing[C]// Proceedings of the 22nd Chinese National Conference on Computational Linguistics, 2023: 124-138.
- [27] WANG S, SUN X, LI X, et al. GPT-NER: Named entity recognition via large language models[C]// Proceedings of the Association for Computational Linguistics: NAACL, 2025: 4257-4275.
- [28] LI X, FENG J, MENG Y, et al. A Unified MRC framework for named entity recognition[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020: 5849-5859.
- [29] LI J, YAN Z, SU X, et al. Overview of CCL23eval task 1: Chinese FrameNet semantic parsing[C]// Proceedings of the 22nd Chinese National Conference on Computational Linguistics, 2023: 113-123.
- [30] BAI J, BAI S, CHU Y, et al. Qwen technical report[J/OL]. arXiv preprint. arXiv: 2505.09388, 2025.
- [31] GUO D, YANG D, ZHANG H, et al. DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning[J]. Nature, 2025, 645: 633-638.



李迎旭(2000—), 硕士, 主要研究领域为自然语言处理、提示词工程、大模型微调 and 检索增强生成。

E-mail: liyingxu@uir.edu.cn



李斌阳(1982—), 通信作者, 博士, 教授, 主要研究领域为自然语言处理和社会计算。

E-mail: byli@uir.edu.cn



陈涛(2000—), 硕士, 主要研究领域为自然语言处理、大语言模型的推理机制、模型在复杂问答任务中的逻辑推理与生成能力。

E-mail: taochen@uir.edu.cn