

## 关系抽取方法研究综述

张勇<sup>1,2\*</sup>, 纪伟<sup>1,2</sup>

(1. 深圳大学 电子与信息工程学院, 广东 深圳 518060;

2. 广东省智能信息处理重点实验室, 广东 深圳 518060)

**摘要:**随着信息技术的不断演进与自然语言处理技术的深入探索,关系抽取作为链接海量文本数据中实体间隐含联系的关键技术,在智能问答、知识图谱构建、个性化信息服务等领域扮演着至关重要的角色。在大语言模型的浪潮下,关系抽取技术迎来了前所未有的发展机遇,其效能与智能化程度得到了显著提升。本文首先简要介绍了关系抽取发展的背景,接着系统性回顾了关系抽取技术的发展脉络,从早期依赖手工规则与知识库的方法,到传统机器学习的技术变迁,再到基于深度学习的关系抽取模型的发展。此外,本文核心讨论了当今研究前沿,即大语言模型在关系抽取领域的应用。最后,本文不仅反思了当前关系抽取研究面临的挑战,如数据噪声处理能力有限、跨领域泛化能力不足以及重叠关系抽取难题等,还展望了未来可能的研究趋势,包括多模态信息融合、文档级关系抽取以及无监督学习策略的创新,旨在为推动关系抽取技术的持续进步提供参考。

**关键词:**自然语言处理;关系抽取;大语言模型;多模态;文档级

**中图分类号:**TP391 **文献标志码:**A **文章编号:**0253-2395(2025)06-1059-21

## Review on Relation Extraction Methods Research

ZHANG Yong<sup>1,2\*</sup>, JI Wei<sup>1,2</sup>

(1. College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, China;

2. Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen 518060, China)

**Abstract:** With the ongoing evolution of information technology and deepening exploration in natural language processing (NLP), relation extraction, as a pivotal technique linking implicit connections among entities in vast textual data, plays a crucial role in intelligent question answering, knowledge graph construction, personalized information services, and etc. Amidst the surge of large language models, relation extraction technology has encountered unprecedented development opportunities, displaying substantial enhancements in efficiency and intelligent level. In this paper, we first briefly introduce the background of the development of relation extraction, and then systematically review the development of relation extraction techniques, from the early methods relying on manual rules and knowledge bases, to the technological changes in traditional machine learning, and then to the development of deep learning-based relation extraction models. In addition, this paper mainly discusses the fusion application of large language models in the field of relational extraction. Finally, this paper not only discusses the challenges faced by current relation extraction research, such as limited data noise processing capability, insufficient capability for cross-domain generalization, and overlapping relation extraction, but also prospects the possible future research trends, including multimodal information fusion, document-level relation extraction, and innovations in unsupervised learning strategies, providing references to push forward the continuous advancement of the relation extraction technology.

**Key words:** natural language processing; relation extraction; large language models; multi-modal; document-level

收稿日期:2024-12-16;接受日期:2025-03-08

基金项目:深圳市科技计划项目(KJZD20230923114405012)

\* 通信作者:张勇(1976-),男,江苏泗洪人,教授,研究方向为人工智能技术研究与应用。E-mail:yzhang@szu.edu.cn

引文格式:张勇,纪伟.关系抽取方法研究综述[J].山西大学学报(自然科学版),2025,48(6):1059-1079. DOI:10.13451/j.sxu.ns.2025015.

## 0 引言

随着互联网和自然语言处理技术的快速发展,在信息爆炸的数字化时代,海量文本数据构成了人们理解世界的基础,这些数据背后隐藏着错综复杂的信息网络。信息抽取(Information Extraction, IE),作为自然语言处理(Natural Language Processing, NLP)领域内的核心组成部分,旨在从非结构化文本中挖掘出易于理解和利用的结构化知识,其中涵盖了两大关键技术:命名实体识别(Named Entity Recognition, NER)和关系抽取(Relation Extraction, RE)。关系抽取作为信息抽取任务的一项核心技术,其目的是从无序文本中自动检测、提取实体之间预定义的关系,最终实现从非结构化数据转换为结构化数据的过程,为知识图谱的构建、智能问答系统以及个性化推荐服务提供有力的支持,例如在智能问答系统中关系抽取所得到的关系实例可以作为后台的知识以此支撑用户问题的回答。

本文第1节介绍了关系抽取发展的相关背景;第2节系统介绍了传统关系抽取方法,分析了早期的关系抽取方法和基于机器学习的关系抽取方法的优点和局限性;第3节进一步地综述了基于深度学习的关系抽取方法;第4节探讨了大语言模型对关系抽取技术的革新影响;第5节对现存挑战进行深刻反思,同时对未来研究趋势进行展望;最后总结全文。本文旨在为关系抽取的学术研究提供一个全面的视角,为新的研究道路提供坚实的理论基础。

## 1 关系抽取的发展

1998年召开的第七届消息理解会议(Message Understanding Conference, MUC)标志着信息抽取领域的一个重要里程碑,尤其在实体关系抽取的界定与发展中。作为最后一届MUC会议,MUC-7不仅沿袭了MUC-6的评测任务,例如命名实体识别、模板填充等,还开创性地引入了模板关系确定任务<sup>[1]</sup>,旨在探索并定义实体间普遍存在的关系模式。具体而言,其目标为针对飞机事故与航天发射两大场景,开发出能够识别如 location\_of、employee\_of、prod-

uct\_of等关系的系统<sup>[2]</sup>。正是在MUC-7的舞台上,关系抽取作为一个明确的研究课题首次亮相,为后续的科研工作奠定了基础。

紧随其后,计算自然语言学习会议(Conference on Computational Natural Language Learning, CoNLL),同样也是信息抽取领域中具有重要意义的会议,涵盖了更多的实体类型和关系类型,进一步引领信息抽取技术向着更广泛的语言应用、更深入的语义理解迈进。

再到之后召开的语义评估(Sem-Eval)会议,其致力于解决面向语义问题,同样对关系抽取技术贡献显著,尤其是在语义角色标注、指代消解、情感分析等任务上,不断推动技术前沿,例如在Sem-Eval 2007和Sem-Eval 2010中对名词和名词短语间关系类型的探索<sup>[3]</sup>,Sem-Eval 2017还涉及从科学文章中提取实体和关系的任务,极大促进了实体关系抽取研究的发展。

2000年召开的自动内容抽取(Automatic Content Extraction, ACE)会议,由美国国家安全局和国家标准与技术研究所联合主导,旨在发展文本抽取技术以支持对语料库的自动化处理。ACE会议中大量的实体关系数据集均由语言资源联盟(Linguistic Data Consortium, LDC)<sup>[4]</sup>提供,显著扩展了语种覆盖范围,从单一的英语扩展至阿拉伯语、西班牙语等,涵盖了广播新闻、报纸新闻等多个维度,大大丰富了数据的多样性和实用性。ACE会议不仅延续了对实体间显性和隐性语义联系的探索,而且系统性地预设了五大关系类别:位置、组织、成员、部分-整体、人物-社会,这些类别构成了关系抽取的框架基础。

传统上,MUC、CoNLL、Sem-Eval、ACE会议所使用的实体关系语料依赖于人工标注,这一过程繁琐且成本高昂,限制了语料的扩展性。随着时代的发展与技术的进步,学者们开始利用维基百科、Freebase、DBpedia等大型高质量知识库作为数据<sup>[5]</sup>支撑,有效缓解了高质量语料获取的难题,这些知识库不仅覆盖领域广泛,而且关系类型更为多样,极大地推动了关系抽取技术的创新与发展。关系抽取相关会议如表1所示。

表1 关系抽取相关会议

Table 1 Meetings related to relation extraction

会议	首次召开时间	常用数据集	涉及语言	关系类别
MUC	1987	MUC-7	英语	3
CoNLL	1997	CoNLL 2004	英语、西班牙语	6
Sem-Eval	1998	Sem-Eval 2010	英语	9
ACE	2000	ACE 2005	英语、汉语、阿拉伯语	7

## 2 传统关系抽取方法

### 2.1 早期的关系抽取方法

早期的关系抽取方法主要依赖于语言学专家和领域专家手工编制的规则。该方法围绕触发词、依存关系构建规则,通过模式匹配技术,结合语义特征、词汇特征进而人工挖掘因果关系或其他语义关系。具体流程是:先对文本进行预处理,接着将规则集与文本进行匹配,从而识别并提取含有特定关系的实体对。

Riloff<sup>[6]</sup>通过建立抽取规则开发了 AutoSlog 系统,该系统能够自动构建特定领域的字典,用于抽取特定领域的信息。Appelt 等<sup>[7]</sup>提出了面向真实世界文本信息抽取的有限状态处理器 (A Finite-state Processor for Information Extraction from Real-world Text, FASTUS) 抽取系统,该系统允许用户根据特定任务需求调整和扩展系统,通过添加或修改规则适应不同的领域,其灵活性使得系统在处理关系抽取任务时节省更多的时间成本。Brin<sup>[8]</sup>所构建的双迭代模式关系扩展 (Dual Iterative Pattern Relation Expansion, DIPRE) 系统采用了迭代扩展的策略,通过少量的实例作为种子,就能从大量的非结构化文本中自主抽取新的关系实例,同时学习新的抽取规则模板,证明了结合种子实例与自动模板学习的潜力。Gravano<sup>[9]</sup>在 Brin 的研究基础上通过完善关系描述模式与实例置信度评估机制,推出了 Snowball 系统,提高了关系抽取的准确性和可靠性。

早期关系抽取方法的核心机制在于模式匹配,该方法高度依赖专家手动创建的词典与规则。然而,该方法存在一定的缺陷:首先是专家需对某个特定领域深度了解,这个过程繁琐耗时容易出错;其次是可移植性较差,即在面对另一个领域或新的数据类型时,需要重新构建规则,无法进行跨领域迁移扩展。因此,尽

管早期基于规则与模板的方法为关系抽取任务奠定了基础,但其效率与准确性较低的缺点导致其不得不朝更加智能的方向发展。

### 2.2 基于机器学习的关系抽取方法

基于机器学习的关系抽取方法是 NLP 领域的一大进展,它代表了从传统的规则与模板匹配的方法向数据驱动自动化的转变。这种方法的核心思想是利用统计学习算法,通过模式识别和数据分析,从文本中自动检测并抽取实体间的关系。该方法通常将关系抽取问题转换为分类问题。其流程大致可以概括为以下几个关键步骤:首先,从大规模标注语料中提取特征,这些特征可能包括语法、词法、上下文等信息,然后将这些特征转化为高维向量表示;接着,应用诸如最大熵 (Maximum Entropy, ME)、支持向量机 (Support Vector Machine, SVM)、条件随机场 (Conditional Random Field, CRF) 等分类器对特征向量进行处理,以识别并分类实体对之间的关系。在该过程中,机器学习模型的性能高度依赖于所选取特征的质量与多样性。

Kambhatla<sup>[10]</sup>在其研究中开创性地使用 ME 模型进行关系抽取,通过整合多种文本特征,包括实体类型、依存关系等,显著增强了特征向量的表达能力,为模型提供了更为丰富的上下文理解基础,有效降低了对密集语义特征提取的依赖,同时展示了模型在处理实际数据时的强大性能。在 ACE RDC 2003 英文语料库上的实验中,该模型达到了 52.8% 的  $F$  值,证实了其在关系抽取任务中的有效性。Tratz 等<sup>[11]</sup>则是以 2007 年 Sem-Eval 数据集为测试语料库,采用 ME 模型实现 77.57% 的  $F$  值。树核 (Tree Kernel, TK) 函数的原理是根据相同子树计算相似度,最早由 Collins 等<sup>[12]</sup>将其引入 NLP 领域。Zhou 等<sup>[13]</sup>采用 SVM 进行关系抽取,在 5 种 ACE 关系类型的提取上比基于 TK 的方法高出大约 20% 的  $F$  值。Zelenko 等<sup>[14]</sup>通过将 TK

与SVM结合,从大量文本资料中提取出“人-隶属”与“组织-位置”两类关系。该方法在多个新闻文章上进行了验证,处理了总计200篇新闻文章的语料库,证明了该方法在实际应用场景中的有效性和泛用性。Culotta等<sup>[15]</sup>构造线性链CRF从传记文本中提取关系,在置信度为0.5的CRF模型中取得最佳F1值,优于ME模型。Bundschuh等<sup>[16]</sup>采用级联的CRF在抽取特定领域关系类型时取得不错的效果。

通常情况下,RE任务的核心在于从非结构化文本中识别并提取实体之间的关联,这一过程因应用领域的不同而展现出不同的特性与需求。基于机器学习的RE方法覆盖了一系列专业领域,比如生物、医学、农业、法律、军事等。在生物领域,Warikoo等<sup>[17]</sup>提出语言模式感知依存树核的方法,通过基于语法关系的依存树对句子进行修剪进而增强语法关系和词汇信

息,为TK生成更多的相关信息,并与SVM结合形成分类器,最终对PubMed摘要中的生物实体关系进行抽取。在医学领域,Ben等<sup>[18]</sup>采用UMLS(统一医学语言系统)提取句法、词汇等特征,将这些特征输入SVM进行关系抽取。在农业领域,Liu等<sup>[19]</sup>将词嵌入特征、依存解析信息和核心谓词位置信息进行结合,采用朴素贝叶斯、SVM和决策树模型进行关系抽取。在法律领域,Thomas等<sup>[20]</sup>则是通过定义一组特定于领域知识的规则识别关系,然后采用半监督的机器学习方法对刑事案件相关的电子判决书进行关系分类,该方法在面对不同长度的句子时提取准确率较高。同样在军事领域,Liang等<sup>[21]</sup>提出基于规则和SVM结合的方法,首先定义一组领域规则识别关系,然后采用SVM分类器对关系进行分类。上述方法如表2所示。

表2 不同领域基于机器学习的关系抽取方法总结

Table 2 Summary of relation extraction methods based on machine learning in various fields

领域	关系类型	方法	应用	局限性
生物	化学物-蛋白质	改进的TK-SVM	药物发现、疾病诊断	分类器特征依赖单一参数、跨类分类问题
医学	药物-疾病、疾病-基因	UMLS-SVM	临床决策、药物控制	采用NLP技术预处理可能引入噪声
农业	作物-营养物质	多特征融合-ML模型	作物管理、疾病控制	关系实例数量不平衡
法律	原告-证据、被告-犯罪	规则匹配-TK	案件分析、合规监测	文本包含冗长、复杂结构的句子
军事	设备-使用、部队-任务	规则匹配-SVM	危险分析、任务规划	训练数据少、语义关系分布不均匀

虽然基于机器学习的关系抽取方法在跨领域的广泛应用中展现出了巨大的潜力,但同时也暴露出了一系列固有的局限性,并且这些局限性因应用领域不同而不同。

显然,机器学习方法在关系抽取领域带来了显著的进展,减少了对领域知识和人工规则的依赖,同时提高了模型的通用性和适应性,但它仍然存在局限性:首先,机器学习模型的训练高度依赖于大量高质量的标注数据,而这些数据的获取和标注往往成本昂贵且耗时;其次,特征工程过程相对复杂,需要领域专家的深入参与,进行设计和验证有效的特征,这对于小众语言或低资源的领域有所约束;再者,预处理(如词性标注、句法分析)的误差可能会累积并传递到后续阶段,影响最终的抽取准确率;最后,模型是否能准确处理未见过的文本并做出有效预测,即机器学习模型的泛化能力无法得到保证,

例如在跨领域适应性上,机器学习模型在不同领域和数据集之间迁移的挑战,包括领域特异性词汇、关系类型差异和标注标准的变化,除此之外,在面对零样本或者少样本情况时应该考虑如何缓解这一问题。因此,学者们开始将研究重心转向自动化特征选择方法,增强模型的鲁棒性。表3对上述方法进行了总结。

### 3 基于深度学习的关系抽取方法

传统关系抽取方法依赖人工设计特征,包括词汇和句法结构,不仅耗时费力,且容易忽略重要信息,扩展性差、泛化能力受限,导致关系抽取进展缓慢。因此,从传统方法转向深度学习进行自动特征学习,成为关系抽取领域的一大转变。深度学习模型不仅减轻了人工设计负担,提高了特征提取的效率和准确性,还能在大规模数据上展现更好的性能,应用最为广

表3 传统关系抽取方法总结

Table 3 Summary of traditional relation extraction

方法	数据集	F1/%	特点
FASTUS	MUC-4	55.00	运行速度快
ME	ACE RDC 2003	52.80	结合文本特征提高上下文理解能力
ME	Sem-Eval 2010	77.57	抽取OTHER类性能低
SVM	ACE RDC	55.50	采用基于短语分块策略
TK-SVM	Corpus comprises 200 articles	86.8/83.3	转换为浅解析分类问题
线性CRF	Wikipedia	61.36	使用置信度限制噪声
级联CRF	PubMed摘要	78.00	易于合并丰富的特征

泛的深度学习关系抽取方法包括卷积神经网络 (Convolutional Neural Networks, CNN)、循环神经网络 (Recurrent Neural Network, RNN)、图神经网络 (Graph Neural Network, GNN)、Transformer 以及基于远程监督策略的关系抽取方法, 这些方法能自动捕获文本复杂特征, 深度挖掘关系信息, 推动 RE 的进一步发展。如图 1 所示为基于深度学习的关系抽取框架。流程包括: 首先将文本通过静态方法或动态方法转换为词向量, 随后输入神经网络, 经多层传播提取深层语义特征, 最后根据这些特征在输出层对实体间关系进行分类, 实现从文本到关系类型的自动抽取。

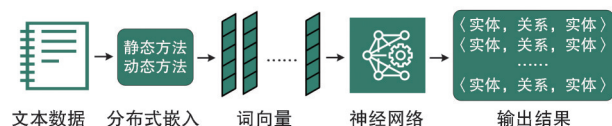


图1 基于深度学习的关系抽取框架

Fig. 1 Deep learning-based relation extraction framework

### 3.1 基于卷积神经网络的关系抽取方法

CNN 最初应用在图像处理领域, 随后被 Yoon<sup>[22]</sup> 研究者引入文本分类任务, 标志着其在 NLP 领域的广泛应用。CNN 在文本处理中的优势在于其能够有效捕获文本局部的关键信息。在关系抽取流程中, CNN 通过在卷积层采用一个或多个卷积核进行卷积操作, 逐层提取局部语义特征, 随后通过池化层进行特征选择, 重复进行得到多个向量, 最终汇总形成一个固定长度的向量, 即整个句子的语义表示。该向量随后输入到全连接层进行分类或预测, 完成了从原始文本到高层语义理解的转化。基于 CNN 的关系抽取框架如图 2 所示。

Liu 等<sup>[23]</sup> 首次将 CNN 应用到关系抽取任

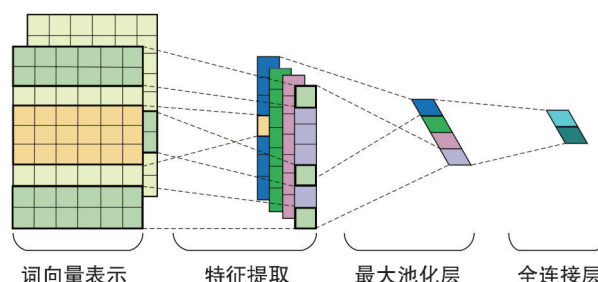


图2 基于CNN的关系抽取框架(图片参考文献[22], 版权所有 ACL)

Fig. 2 CNN-based relation extraction framework (See Ref. [22], Copyright ACL)

务, 通过依赖外部 NLP 工具的同义词字典对输入词进行编码, 结合词汇特征, 经过卷积层、max 层, 最后在 softmax 层进行关系抽取, 该方法在 ACE 2005 语料库上较 TK 方法提升了 9% 的 F1 值。验证了 CNN 在关系抽取任务中的可行性。然而由于外部 NLP 工具的使用, 可能引入了噪声。因此, Zeng 等<sup>[24]</sup> 则摒弃了复杂的预处理步骤, 直接通过词嵌入将文本转为词向量, 提取词级和句子级特征同时融合位置特征 (Position Feature, PF), 将特征连接起来送入 softmax 分类器抽取关系。该方法在 Sem-Eval 2010 的语料库中训练, 取得 82.7% 的 F1 值, 解决了传统关系抽取方法过于依赖预处理的问题。Nguyen 等<sup>[25]</sup> 提出的关系抽取方法, 通过预训练的词嵌入结合相对位置特征并采用多窗口的过滤器提取更大范围的 n-gram 特征, 较 Zeng 等的模型, 其性能提升 0.1%。Santos 等<sup>[26]</sup> 提出基于排名的 CNN 分类方法 (Classification by Ranking with Convolutional Neural Networks, CR-CNN), 采用成对排序的损失函数, 实现了 84.1% 的 F1 值。在提升鲁棒性方面, Xu 等<sup>[27]</sup> 提出从对象到主题的最短依存路径作为负样本

的采样方法对关系中的主客体进行准确的分配,在 Sem-Eval 2010 Task 8 数据集上取得了最佳效果。为了减少对特征工程的依赖, Sahu 等<sup>[28]</sup>提出基于 CNN 关系抽取新框架,采用 6 个离散特征表示句子中的词汇,即词向量、两个位置向量、词性特征、词干特征以及实体类型,在 i2b2-2010 数据集上取得 71.16% 的  $F1$  值。在 Sahu 的工作基础上, Gu 等<sup>[29]</sup>融合最大熵模型与 CNN,有效地提取出句内和句间的关系。在结合注意力机制方面, Shen 等<sup>[30]</sup>提出基于词级注意力机制的 CNN 方法,充分利用词嵌入、词性标签嵌入和位置嵌入信息来选择关系语义表示的关键信息。Wang 等<sup>[31]</sup>则是在 CNN 的基础上,结合两层注意力机制,即输入注意力层和关系注意力层,分别捕获实体及目标关系的主要特征。Asada 等<sup>[32]</sup>基于 Wang 等提出的输入注意力层进行优化,在该层增加偏差项来调节平滑参数,提高了关系抽取的性能。Zhu 等<sup>[33]</sup>则提出了关系分类方法,只采用词层面的注意力机制直接提取词汇与关系的相关性,能够更有效地捕获句子的关键部分。

尽管 CNN 在关系抽取任务中存在固定的卷积核的局限性,导致其在捕获长距离依赖问题上的性能不足。针对该问题,目前研究主要有两种处理方法:一是通过增大卷积核尺寸以覆盖更广泛的上下文信息;二是利用树结构,如依存句法树,间接减小实体间的距离。然而,在实践中卷积核尺寸通常较小,并且这些方法往往存在较高的计算成本。而循环神经网络在长序列建模方面正好能够避免这个问题。

## 3.2 基于循环神经网络的关系抽取方法

### 3.2.1 循环神经网络

RNN 引入了循环连接机制,不仅允许网络权重在时间序列的每个步骤中共享,还使得模型能够捕捉并保留序列中的短期依赖信息。RNN 通过维持一种“记忆”状态,能够基于过往输入和当前数据动态更新,适应处理不同长度的序列数据。它通过隐藏状态储存累积信息,并利用反馈机制整合历史信息与当前输入,进而输出预测。RNN 的结构如图 3 所示。

在 RNN 被提出之后,具有相似结构的递归神经网络(Recursive Neural Network)也能处理

序列数据。Socher 等<sup>[34]</sup>最早将递归神经网络应用在关系抽取领域,提出了矩阵向量 RNN(MV-RNN)模型,旨在克服传统词向量模型在理解长句和短语深层意义上的局限性。MV-RNN 通过双层表示法:矩阵代表词汇间的上下文关联信息,向量代表词汇内在含义,实现了对句子语义的深度捕捉,尤其擅长处理含有复杂结构和长距离依赖的语句。该模型建立在解析树与 RNN 的结合之上,首先定位待抽取关系的实体间路径,并选取路径上的最高节点,利用该节点的复合特征向量(融合了词汇的向量和矩阵表示)执行关系抽取。MV-RNN 在抽取名词间语义关系的实验中达到 82.4% 的  $F1$  值。而 RNN 被视为递归神经网络的特殊情况。Zhang 等<sup>[35]</sup>提出采用位置指示器(Position Indicators, PI)的方法,较 PF 的方法更为简单有效,实验中通过结合双向 RNN 与 PI 得到优于 Zeng 等提出的 CNN 模型的结论。

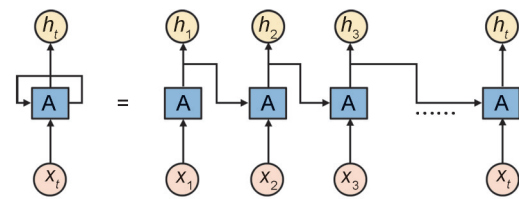


图3 RNN结构

Fig. 3 The structure of RNN

实际上,尽管 RNN 擅长处理序列数据,但在面对超长序列时,存在梯度消失和梯度爆炸的问题,这限制了其学习长期依赖的能力。

### 3.2.2 长短时记忆网络

长短时记忆网络(Long Short Term Memory Network, LSTM)通过“门控”机制,包括遗忘门、输入门、输出门以及细胞状态,使得其在处理超长序列时能够避免 RNN 梯度消失和梯度爆炸的问题。LSTM 的结构如图 4 所示。

其中遗忘门负责调控先前记忆的保留程度  $C_{t-1}$ ,输入门控制新信息  $x_t$  的接纳量并保存到细胞状态中  $C_t$ ,而输出门则决定了哪些信息应被当前的隐藏状态  $C_t$  输出  $h_t$ 。这种设计利用 tanh 和 sigmoid 激活函数的特性,前者保证了信息流动范围的稳定性,后者则帮助实现信息的二值化,即保留(值接近 1)或遗忘(值接近 0)。在每个时间步, LSTM 综合考虑上一时刻的信

息、当前输入以及细胞状态,更新其信息状态和隐藏状态,以连续反映序列信息。

Xu 等<sup>[36]</sup>最先将 LSTM 应用于关系抽取任务,其提出的 SDP-LSTM 方法利用了文本中两个实体间最短依存路径(Shortest Dependency Path, SDP)的信息,该路径在句法依存树中定义为连接两个实体到它们最近公共祖先节点的最短路径。模型通过 LSTM 沿着 SDP 的左、右路径捕获上下文特征,并采用最大池化技术整合路径上关键节点的特征表示,以强化模型对关系特征的识别能力。证明了 LSTM 与最短依存路径结合的有效性,显著提升了在句子层面理解复杂关系的能力。在 Sem-Eval 2010 的数据集中,SDP-LSTM 模型实现了 83.7% 的  $F1$  值。Ningthoujam 等<sup>[37]</sup>在临床医学领域也采用 SDP-LSTM 方法进行关系抽取,在部分 i2b2 2010 数据集上取得了不错的效果。在生物医学领域,Zhou 等<sup>[38]</sup>融合基于特征的模型、基于树核的模型、神经网络进行化学-疾病关系(Chemical-disease Relations, CDR)抽取,其中基于特征的模型提取词汇特征,基于树核的模型采用 SDP 提取句法结构特征,神经网络采用 LSTM 生成语义表示。Lim 等<sup>[39]</sup>提出二叉树 LSTM 模型结合叶子节点的位置特征、内部节点的相对位置特征和上下文特征的方法,在药物相互作用(Drug-drug Interactions, DDI)数据集上的关系检测和关系分类任务上分别比当时最先进的模型高出 4.4% 和 2.8%。实际上,单向 LSTM 只能获取某一时间步之前的信息而忽略了后续的上下文,限制了对序列数据中依赖关系的全面理解。然而,双向 LSTM(Directional Long Short Term Memory Network, BiLSTM)采用了双流向的训练机制,存在前向隐藏层和后向隐藏层。前向 LSTM 捕捉序列从左至右的上文信息,即每个时间点之前的上下文,而后向 LSTM 则逆序工作,捕捉每个时间点之后的下文信息。通过将这两个方向的隐藏状态合并,BiLSTM 不仅整合了序列的过去和未来上下文信息,还大大提升了模型对整个文本环境的理解深度和关系捕获精度。Zhang 等<sup>[40]</sup>仅采用词嵌入作为输入,通过 BiLSTM 模型提取上下文特征时,结果就超过 Zeng 等的 CNN 方法和 Santos 等的 CR-CNN 方

法,当结合位置特征、相对依赖特征(Relative-dependency Feature, Relative-Dep)等其他特征时效果提升更加显著。在结合注意力机制方面,Zhou 等<sup>[41]</sup>率先提出基于注意力机制的 BiLSTM 网络进行关系抽取,利用 BiLSTM 捕获句子中的上下文,再借词汇级别的注意力机制,自动加权句子中对关系判断重要的词汇,合并为句子级向量,采用 dropout 正则化,实现 Sem-Eval 2010 Task 8 数据集上的关系抽取。Xu 等<sup>[42]</sup>提出了用于关系分类的深度循环神经网络,通过堆叠多层结构解决了浅层网络无法探索不同抽象层次特征表示的问题。该方法基于最短依存路径,并结合创新的数据增强技术,在 SemEval-2010 Task 8 任务上达到 86.1%  $F1$ -score,显著提升当时最优性能。Wang 等<sup>[43]</sup>采用剪枝策略在减少输入句子大小的同时最大限度保留句子中的相关信息,提出双向 SDP 注意力机制,并结合 BiLSTM 捕获句子的重要信息,最终通过突出关系语义词和指向性词进行关系抽取。

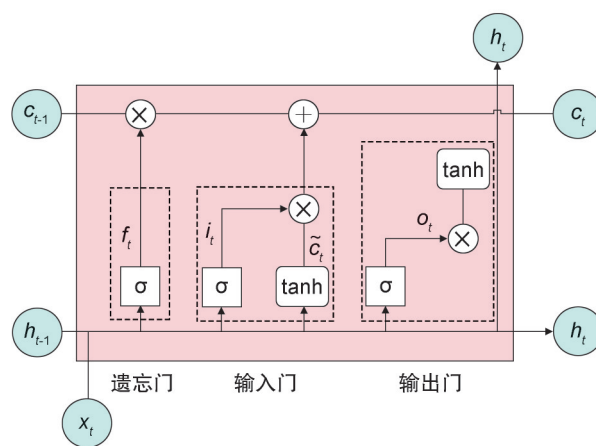


图4 LSTM结构

Fig. 4 The structure of LSTM

### 3.2.3 门控循环单元网络

门控循环单元(Gated Recurrent Unit, GRU)网络是对RNN的一种简化和优化,基于输入门和遗忘门的互补性将LSTM模型中的输入门和遗忘门整合为单一的更新门,调节过往状态信息如何流入当前状态,同时省去了独立的细胞状态设计。GRU利用更新门控制新旧信息的保留与遗忘,配合重置门共同作用,类似于LSTM的遗忘门和输入门的机制,分别控制旧信息的释放与新信息的接纳。GRU的结

构如图5所示。GRU模型的简化不仅体现在门结构上,其较少的参数量也使得模型在训练时更为高效,能够更快达到收敛,同时在许多任务上实现与LSTM相媲美的性能。

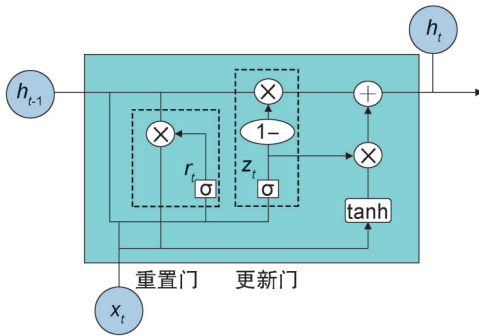


图5 GRU结构

Fig. 5 The structure of GRU

Nguyen等<sup>[44]</sup>将GRU和CNN分别通过集成方法、堆叠方法和投票方法进行组合,实验表明,在关系抽取任务中采用集成方法进行网络组合表现不佳,堆叠方法效率较低,而投票方法是表现最佳的组合方法。GRU是LSTM的一种变体,与LSTM一样,GRU也有双向结构的模型BiGRU。为了充分利用实体对的信息,Qin等<sup>[45]</sup>提出基于实体对的注意力机制,通过实体对信息作为先验知识生成注意力权重向量,结合BiGRU的输出进行加权,结果表明基于实体对的注意力机制较原始注意力机制有更强的关系抽取能力。Luo等<sup>[46]</sup>将Highway网络融入基于注意力机制的BiGRU模型中,其中前向GRU层学习历史信息,后向GRU层则用来学习未来信息,Highway网络学习如何通过门控单元控制信息在神经网络中的流动,能够捕获词汇之间更多的语义特征,接着采用词级注意力机制和句子级注意力机制分配网络结构中关键信息的权重,能够有效抽取地质资料中的关系。在生物医学领域,Shen等<sup>[47]</sup>提出不同于SDP仅使用部分依存树信息的方法,而是利用整个依存树的信息,采用BiGRU作为编码器,将注意力机制应用于池化层,以提取疾病和症状之间的关系。Zhang等<sup>[48]</sup>则是采用基于字符级和句子级注意力机制的BiGRU网络构建关系抽取模型,以提取疾病、症状和检测之间的关系。

### 3.3 基于图神经网络的关系抽取方法

图数据结构以其独特的形式,通过节点和

边的相互连接,为表示实体及其关系提供了一种直观且强大的模型。由于GNN擅长从这样的网络结构中挖掘和理解复杂的关系模式,GNN在最近几年逐渐受到学者们关注。相较于传统的CNN和RNN,GNN在处理非欧几里得数据时,能够展现出更高的效能和全局理解能力。在NLP领域,句子和文档常被转换为图结构,比如依存关系树和句法树,这些结构不仅编码了词序信息,还蕴含了丰富的语法和语义联系,极大提高关系的抽取准确率。

Zhu等<sup>[49]</sup>提出了带有生成参数的图神经网络(Graph Neural Networks with Generated Parameters, GP-GNN)。GP-GNN首先通过文本序列中的实体构建全连通图,然后采用三个模块来处理关系推理:首先是编码模块,能够对自然语言中的信息进行编码;其次是传播模块,能在各个节点之间传播关系信息;最后是分类模块,能够利用节点表示进行预测。与传统GNN相比,GP-GNN能够从自然语言中学习边的参数。此外,实验结果还表明GP-GNN能够表现出多跳关系推理的能力以实现丰富文本信息的关系传递。实际上,GNN还包括图卷积网络(Graph Convolutional Network, GCN)、图注意力网络(Graph Attention Network, GAT)。Schlichtkrull等<sup>[50]</sup>提出将关系图卷积网络(R-GCN)应用于知识库补全任务,并且认为GCN能够有效应用在关系抽取任务中。Zhang等<sup>[51]</sup>首次将GCN应用于关系抽取任务,提出一种以依存路径为中心的剪枝策略,删除不相关的内容而保留关键信息,然后将剪枝后的树送入GCN,不仅能够捕获更多的全局信息,还能够并行化进行关系抽取,加速抽取进度。Guo等<sup>[52]</sup>提出基于注意力机制的GCN(AGGCN)模型,采用的模型直接将完整的依存树转化为全连接的图结构。在注意力引导层(如图6所示)引入“软剪枝”策略为所有的边分配权重,并允许模型通过自注意力机制端到端地学习每一条边的权重,即节点间的关联程度,接着将结果矩阵输入到 $N$ 个密集连接层生成新的表示,最后在线性连接层组合成隐藏表示,实现选择性地保留和丢弃信息。在生物医学领域,Park等<sup>[53]</sup>提出基于注意力机制的剪枝策略,通过

自注意力机制为完整依存图中的所有边分配权重,以忽略无关的信息,输入到GCN模型中进行药物相互作用(DDI)关系抽取。由于密集地利用依存信息可能引入噪声,因此,Tian等<sup>[54]</sup>提出依存驱动的基于关注图卷积网络(A-GCN)的关系抽取方法,首先构建依存树,

然后在依存树的基础上构建图,将注意力机制应用于依存树的上下文,对依存连接和类型都分配权重,以区分不同词汇依存关系的重要性。该方法在ACE 2005和Sem-Eval 2010 Task 8两个数据集上的实验结果证明了A-GCN进行RE的有效性。

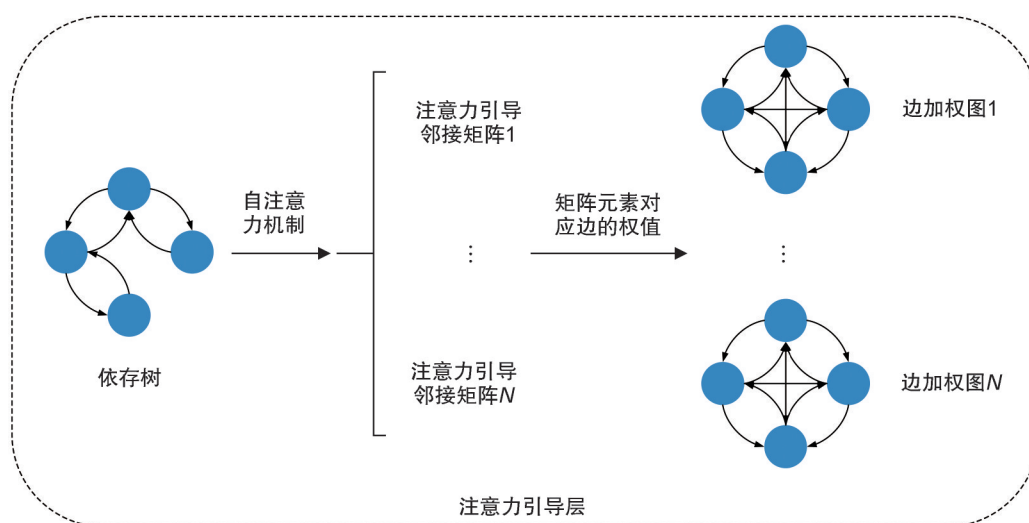


图6 AGGCN注意力引导层(图片参考文献[52],版权所有arXiv.org)

Fig. 6 AGGCN Attention-Guided Layer (see original pictures in Refs. [52], Copyright arXiv.org)

然而,图数据通常包含噪声,例如缺失边、错误的连接等,GNN在处理这些情况时可能不如预期。此外,在某些情况下,GNN可能因过度关注高频关系而忽视了长尾或稀有关系,进而导致模型偏向。

### 3.4 基于Transformer的关系抽取方法

2017年谷歌推出的Transformer模型<sup>[55]</sup>改变了序列数据处理的格局,特别是在NLP领域,成为了目前的主流架构。其核心在于引入了自注意力机制(Self-attention Mechanism),这一机制使得模型能够直接、灵活地处理序列中的各个部分之间的相互依赖,无需像RNN那样按顺序逐步传递信息,也无需像CNN那样在固定范围内局部滑动提取特征。Transformer通过这种机制捕捉全局信息,并在模型的每一层中重复这一过程,有效融合了序列的上下文信息。标准Transformer架构如图7所示,采用编码器-解码器框架,其中编码器和解码器均是由多层Transformer模块堆叠而成。每个模块包括多头自注意力层(Multi-headed Self-Attention),它能够并行处理输入序列的不同表示子

空间,捕捉不同位置间的复杂关系,紧随其后的全连接前馈网络层(Position-wise Feed-forward Networks)进一步提取这些信息。在编码器中,这些组件通过求和(残差连接)以及归一化确保了信息流动的连贯性和稳定性。解码器模块在编码器的基础上添加了交叉注意力层(Cross-attention Layer),以利用编码器的输出作为条件,指导解码过程,从而在生成序列时考虑到了源序列的上下文信息。

预训练语言模型(Pre-trained Language Model, PLM)通过“预训练-微调”的学习范式构建,采用自监督方法学习非标注文本中可迁移的特征,然后采用有监督方法“微调”下游任务,这从根本上改变了NLP任务的处理方式。随着Transformer的提出,序列建模能力显著提升,出现了诸如GPT<sup>[56]</sup>和BERT<sup>[57]</sup>等的基于Transformer架构的预训练语言模型,这些模型具备通用且强大的语义表达能力,只需要根据具体的下游任务对PLM进行“微调”就可以显著提升性能。

#### 3.4.1 GPT

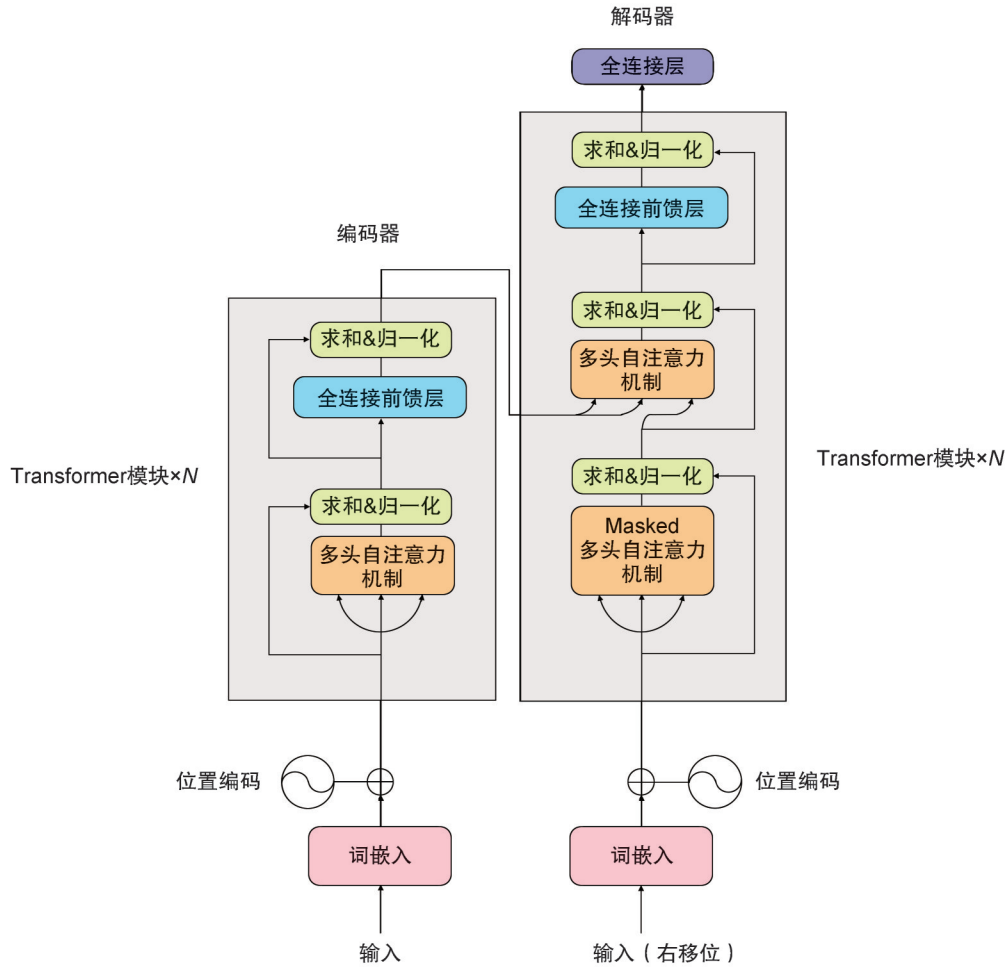


图7 Transformer结构

Fig. 7 The structure of Transformer

GPT 由 12 个经过修改的 Transformer 解码器模块构成,这些模块仅包含 Masked 多头自注意力机制网络和前馈网络,能够实现基于历史内容预测后续文本的生成式任务。GPT 的结构如图 8 所示。

由于 GPT 采用生成式预训练方法,主要用于语言生成任务,所以其在关系抽取领域通常用于数据增强。Alt 等<sup>[58]</sup>使用 GPT 进行关系抽取。在输入表示方面使用字节对编码(Byte Pair Encoding, BPE)对文本进行编码,提高了模型学习语义特征的能力。Papanikolaou 等<sup>[59]</sup>通过 GPT-2 对每个关系类型进行微调,然后使用微调模型生成新的训练数据,以提高训练数据的质量,接着通过采样,最后微调关系分类器。实验结果表明该方法适合处理关系类型不平衡的场景,该方法在 3 个生物医学领域关系抽取数据集上实现 SOTA 效果。然而,该方法

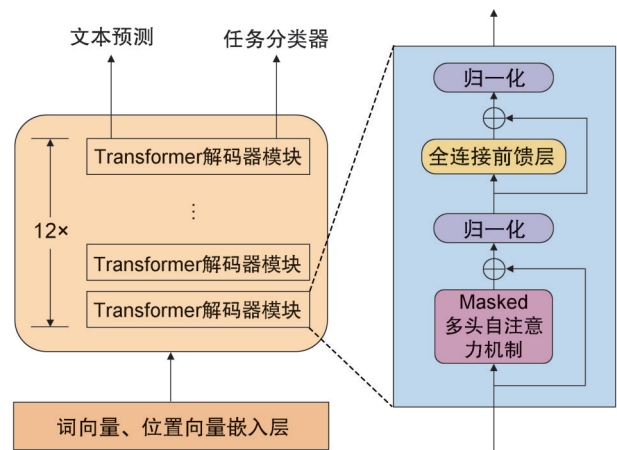


图8 GPT结构

Fig. 8 The structure of GPT

仅在小规模数据量下(50 万条 PubMed 摘要)进行预训练。Luo 等<sup>[60]</sup>采用 GPT-2 模型结构作为主干网络提出 BioGPT 模型,在 1 500 万条 PubMed 摘要语料库上进行预训练,然后再将预

训练的BioGPT应用于下游任务,在3个端到端关系抽取任务中与基线模型相比具有更好的性能。由于GPT采用单向的Transformer结构,因此每个输入文本只能依赖其左侧的信息。

### 3.4.2 BERT

现阶段关系抽取领域应用最广泛的PLM是BERT模型,相较于GPT,BERT采用的双向Transformer结构能够快速生成更深层次的语义表征,此外,其预训练策略包括MLM(Masked Language Model)和NSP(Next Sentence Prediction)。其中,MLM遮盖部分词汇让模型依据上下文推测,提升语境理解能力;NSP则增强句间关系学习能力。BERT的提出使得关系抽取准确率得到进一步提高。

Wu等<sup>[61]</sup>首次将BERT应用于关系抽取任务,提出了一种将目标实体信息整合到BERT模型中的方法(R-BERT),在Sem-Eval 2010 Task 8数据集中取得89.25%的F1值,大幅度领先其他基线模型。Soares等<sup>[62]</sup>提出BERT-MTB方法,使用ENTITY MARKERS作为输入编码、ENTITY START作为输出关系表示进一步提高了BERT的性能,采用匹配空格的策略提高了关系抽取的准确性。Zhao等<sup>[63]</sup>提出EEPGNN模型,通过“\$”“#”标记实体以增强BERT捕获实体的能力,采用GCN提取实体对图的拓扑结构,并与BERT生成的句子语义信息结合,在Sem-Eval 2010 Task 8数据集和ACE 2005数据集中超过Wu等提出的R-BERT模型,分别取得90.2%和77.1%的F1值。Peters等<sup>[64]</sup>提出KnowBERT-W+W模型,通过轻量级的KAR机制将知识库集成到BERT的中间两层中,使模型能够结合外部知识库的信息,实现知识增强的上下文表示,进而提高关系抽取的准确率。在结合其他神经网络方面,Shi等<sup>[65]</sup>提出BERT-BiLSTM的关系抽取框架,如图9所示。首先通过WordPiece分词器进行分词并将结果输入到BERT中获得上下文表示,然后将其与位置信息进行结合输入单层BiLSTM中进行关系抽取,在TACRED数据集的F1值高于所有基线模型。在缓解重叠关系问题上,Li等<sup>[66]</sup>通过BERT在关系计算层计算相似度提取实体位置信息,采用sigmoid激活函

数代替softmax激活函数进行归一化,最后通过计算所有实体对、关系类型是否存在某种关系的概率进行关系抽取。

实际上,根据下游任务对预训练语言模型进行“微调”的目的是对预训练的通用语义理解能力进行特化,使之能够适应具体应用场景的需求。这一过程使用带有标签的数据集,针对性地调整模型参数以优化特定任务的表现。但是,该过程也存在模型可能会过度专注于提升在狭窄任务上的性能的局限性,进而“遗忘”预训练期间积累的通用知识,即灾难性遗忘问题。在研究实践中,通常采用结合预训练损失和下游任务微调损失的方法缓解该问题。

### 3.5 基于远程监督策略的关系抽取方法

为了解决关系抽取领域数据集短缺、数据标注成本高的问题,Mintz等<sup>[67]</sup>于2009年首次将远程监督策略引入关系抽取领域,并在ACL会议上发布。其核心假设是:如果知识库中两个实体存在某关系,则任何提及这两个实体的文本可能表达了相同的关联。通过将大规模无标注文本与知识库对齐,远程监督能够自动生成大量训练样本,从而减轻人工标注负担,加速数据集构建过程,并扩大知识库规模。远程监督过程如图10所示。但是该假设容易导致错误标注问题引起噪声。2010年,Riedel等<sup>[68]</sup>针对远程监督关系抽取中的严格假设问题,提出了“at-least-one”新假设。该假设认为,如果两个实体之间存在某种关系,那么所有包含这两个实体的句子集合里,至少存在一个句子表达了这种关系。相较于Mintz原先的严格假设,这一改变旨在减少因实体共现而产生的错误标注,提高数据集的准确性和实用性。为解决由此带来的复杂性,即在一组句子中确定哪一句真正表达了关系的难题,Riedel等引入了多实例学习(Multi-instance Learning, MIL)策略。在MIL框架下,含有相同实体对的所有句子被组合成“包”,模型不是对单个句子分类,而是预测整个包的关系标签。只要包内至少有一个句子表达了包的正确关系,该包就被视为正例。该方法提高了模型在有错误标注数据上的鲁棒性。

Zeng等<sup>[69]</sup>提出分段卷积神经网络(Piece-

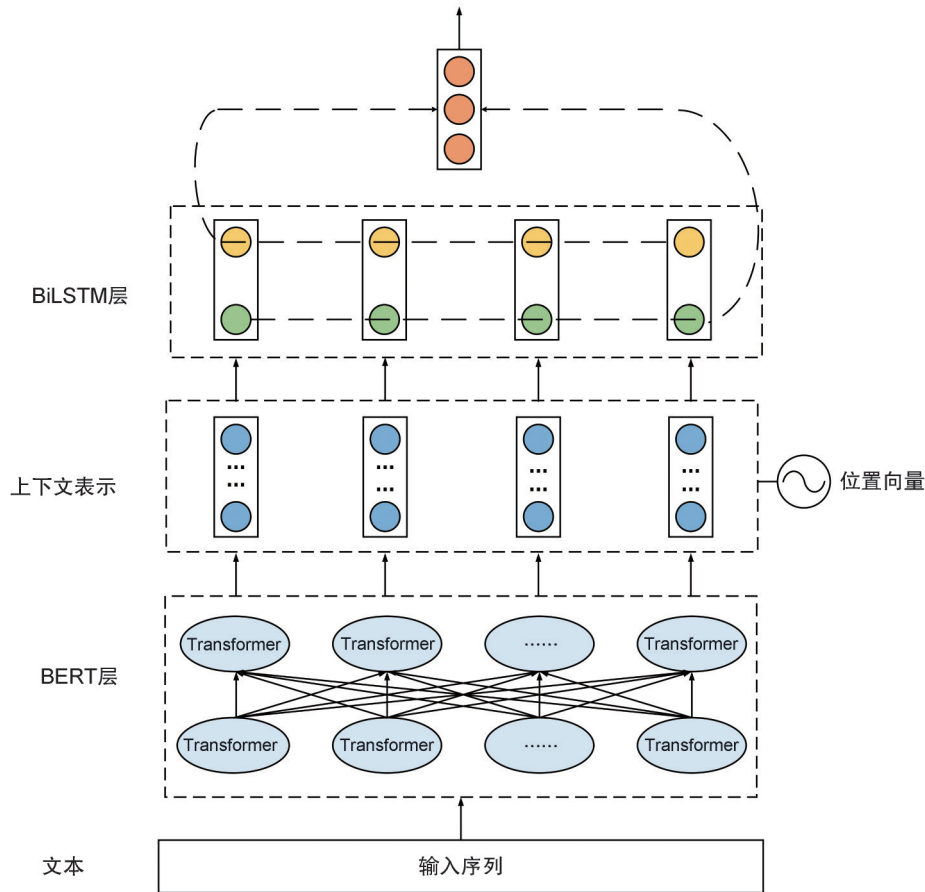


图9 BERT-BiLSTM结构

Fig. 9 The structure of BERT- BiLSTM

wise Convolutional Neural Networks, PCNN), 通过将句子根据实体位置划分为三个部分,即头实体前、两实体之间、尾实体后,并在这些区域应用平均或最大池化操作,来针对实体的上下文特征进行抽取,提高了模型对关系的敏感性,允许模型从句子的不同部分捕捉信息,增强对复杂关系的理解。除此之外,为了应对错误标注问题,将PCNN与多实例学习策略相结合,在这一框架下,包含相同实体对的所有句子被视为一个包,而不仅仅是选取概率最高的句子作为包的代表,而是通过多实例学习策略考虑包内所有句子的信息,尽管实际操作中往往依赖于最高置信度的句子来简化学习任务,但是这在一定程度上仍然限制了模型对包内全面信息的利用。

Lin等<sup>[70]</sup>通过在多实例学习框架中融入基于注意力的机制来改善远程监督策略的性能。这种方法使得模型能够识别并重视包内对关系标签具有较大影响力的句子,同时降低噪声数

据的干扰。首先利用CNN对句子进行编码得到向量表示,随后通过注意力机制为包中每个句子分配一个权重,反映其在预测特定关系时的重要性。加权后的句子向量汇总成为包的表示,进而用于关系抽取。该方法能够有效缓解噪声问题。同样在采用句子级注意力机制方面,Ji等<sup>[71]</sup>则是结合PCNN与知识库中的实体描述信息,提出APCNN模型,通过句子级注意力机制有效利用多实例数据,增强实体表示并优化关系抽取,减少噪声影响。

在多关系建模方面,Jiang等<sup>[72]</sup>提出了一种多实例多标签的远程监督正则化卷积神经网络,采用了跨句子的最大池化策略捕获句子间的关联信息以建模多关系,每种关系将被建模为二分类任务,通过判断关系概率是否超过阈值作为预测标签。Han等<sup>[73]</sup>提出基于知识和基于语义的相互注意力机制,结合知识图谱和文本相互指导以增强模型性能。

在结合强化学习(Reinforcement Learning,

RL)方面,Feng等<sup>[74]</sup>提出了一种基于噪声数据的句子级关系分类模型,该模型有两个模块:一个实例选择器和一个关系分类器。实例选择器通过强化学习选择高质量的句子,并将选择的句子输入到关系分类器中,关系分类器进行句子级预测,并向实例选择器提供奖励。Qin等<sup>[75]</sup>则是采用RL生成假阳性样本(错误地将实例中实体之间关系标注为某种关系类型),通过动态选择策略自动识别每种关系类型的假阳性,并将它们重新分配到负例中,最大化提升关系抽取性能的奖励。Feng等方法的奖励是基于预测概率,而该方法的奖励是基于关系分类器的性能变化,能够显著提高远程监督策略的性能。

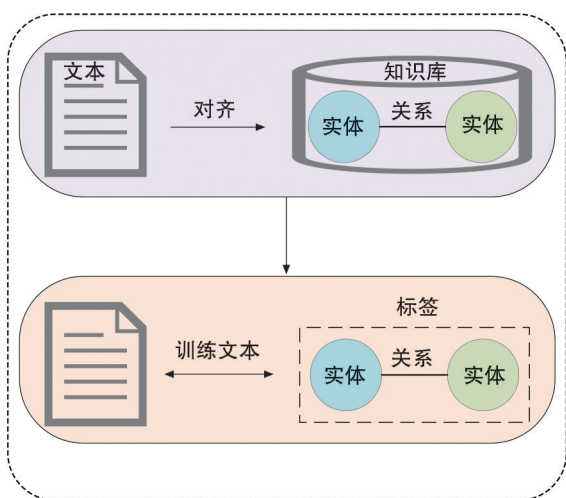


图10 远程监督框架

Fig. 10 The framework of distant supervision

在结合对抗学习(Adversarial Learning, AL)方面,Qin等<sup>[76]</sup>受生成式对抗网络的启发,提出远程监督生成式对抗网络(Distant Supervision Generative Adversarial Networks, DSGAN)方法,将生成器生成的正样本作为负样本来训练判别器,直到判别器的性能下降到最大时,得到最优生成器。采用该生成器对远程监督训练集进行去噪,将假阳性实例重分配到负例中,为关系抽取提供干净的数据集。实验结果显示,基于DSGAN的方法在关系抽取精度方面超越了当时的最佳模型,证实了DSGAN方法是一种具有鲁棒性的对抗学习策略。

在结合预训练语言模型方面,Alt等<sup>[77]</sup>提出DISTRE方法如图11所示,将GPT扩展到远程监督数据集,在NYT 10数据集上对其进行微调,实验表明该方法能够以高置信度预测更多的不同的关系类型。Yu等<sup>[78]</sup>提出基于BERT的关系抽取框架BRE,采用BERT作为特征提取器,接着采用位置增强的CNN(PE-CNN)处理实体之间的相对位置信息以弥补预训练任务和关系抽取任务之间的语义差距,随后将文本表示输入到时间衰减的选择性注意力机制模块(TDSAM),根据时间衰减掩盖低置信度的实例,以减少干扰,有效缓解了MIL框架下的噪声问题。

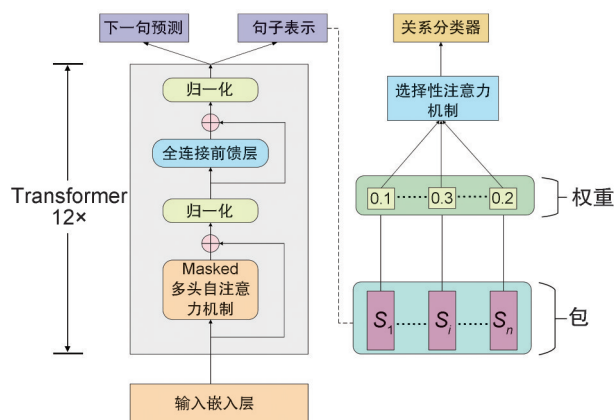


图11 DISTRE框架(图片参考文献[77],版权所有arXiv.org)

Fig. 11 The framework of DISTRE (see original pictures in Refs. [77], Copyright arXiv.org)

远程监督策略作为关系抽取领域的一项重要技术,通过自动将知识库中的事实信息投射到未标注文本上,实现了大规模训练数据的快速生成,极大促进了该领域的发展。这种方法的优势主要体现在数据规模的扩大、标注效率的提升以及实体关系类型的广泛覆盖上,为模型提供了丰富的学习资源,加速了模型训练与迭代过程,并有助于模型泛化能力的构建。然而,错误标注问题是关系抽取中引入噪声的主要问题,知识库与自然语言文本间的简单匹配忽略了语境复杂性,导致出现大量不准确的标签,这些噪声数据不仅降低了模型学习的质量,还可能引发模型对错误模式的过度学习。表4对上述基于深度学习的关系抽取方法进行了总结。

表4 基于深度学习的关系抽取方法总结

Table 4 Summary of relation extraction methods based on deep learning

方法	优点	缺点	适用性分析
卷积神经网络 (CNN)	局部特征捕捉能力强,通过卷积操作识别文本中的关键n-gram结构	固定大小的卷积核无法有效处理长距离依赖关系	短文本或中等长度文本的处理,局部依赖性强的关系抽取任务
循环神经网络 (RNN)	能够保留可变长度序列的上下文信息;通过隐藏状态动态更新,适应不同长度的序列	面对超长序列时,存在梯度消失和梯度爆炸问题;学习长期依赖能力有限	较短的序列或局部依赖性强的关系抽取任务
图神经网络 (GNN)	有效处理非欧几里得数据;对于句子和文档转换为图结构(如依存关系树和句法树)有出色的表现	在处理含噪图数据时,如缺失边或错误连接,性能降低;过度关注高频关系,忽视长尾或稀有关系,导致模型偏向	关系复杂的文本数据例如知识图谱补全中的关系抽取任务
Transformer	引入自注意力机制,直接处理序列中各部分的相互依赖,不需要逐元素传递信息;捕捉全局信息,有效融合序列上下文	可能因长序列计算资源限制表现不佳;过度关注提升狭窄任务的性能上限,存在灾难性遗忘问题	需要全局上下文理解的关系抽取任务
远程监督策略	自动生成大规模训练数据,降低人工标注成本;加速数据集构建,扩大知识库规模	错误标注问题,导致数据噪声	大规模训练集、数据稀缺的关系抽取任务

#### 4 基于大语言模型的关系抽取方法

2020年Open AI发布了1750亿参数的GPT-3<sup>[79]</sup>模型,开启了大语言模型(large language models, LLM)(大模型)的新时代。2022年底ChatGPT的出现,使得LLM的能力得到充分的展现,激发了大模型的研究热潮。2023年3月,GPT-4的出现使ChatGPT具备更强的理解能力。2024年5月,GPT-4o(o指全方位)能够实时对音频、图像和文本进行推理,与现有大模型相比在图像和音频方面表现更出色,使得人机交互更加流畅。

大模型中的提示(Prompting)技术是指在输入中采用预设计的提示来指导大模型,使其在特定任务上展现出更优的性能。该技术利用模型的泛化能力,不需要对模型进行微调就能够完成各种任务。提示技术主要分为两种:指令提示和思维链。

指令提示(提示学习)起源于GPT-3,通过在输入中加入提示,为模型提供关于如何处理当前任务的上下文信息。这些提示能够重塑任务的形式,让模型将其视为与预训练阶段相似的问题来解决,而不是直接要求模型适应新的任务框架。这种方法展示了即使在少量样本的情况下,也能显著提升模型的表现。提示学习主要是指上下文学习(In-context Learning, ICL),也称为情境学习。思维链(Chain-of-thought, CoT)<sup>[80]</sup>则是一种促进大模型进行多步推理的策略,模仿人类解决问题时的思考过程。思维链促使模型

不仅输出最终答案,还要展示达到答案的逻辑路径,不仅帮助提升模型的推理能力和输出的合理性,还增强了模型决策的透明度。这种方法通常要求模型具备一定的参数量,因为较小的模型可能难以理解复杂的思维链。如图12所示为上下文学习和思维链的提示过程。

尽管上下文学习擅长捕捉文本中的语境信息和模式,而CoT策略则在处理需要多步骤推理的数学问题和常识推理中显示出其优越性,这两种方法在关系抽取任务中都有各自的优点和局限性。对于涉及多跳逻辑和长距离依赖的关系,上下文学习可能面临难以解决复杂的多步推理问题的情况,而CoT策略可能导致推理路径变得过于复杂,从而增加推理时间和资源消耗,同时降低模型的稳定性。此外,在应对文本中的模糊性和歧义性时,两者都需要具备较高的理解能力和歧义消解技巧,这是目前技术领域的一个共同挑战。因此,未来的研究应致力于结合这两种方法的优势,以实现更高效、准确的关系抽取。

提示技术是挖掘和应用大型语言模型潜力的高效策略。通过巧妙设计的提示词,能激活模型潜在能力,提高回答的针对性和准确性,同时降低成本,提升效率,有利于模型决策过程的解释。

提示调优的核心思想是将文本片段(模板)插入输入,并将分类任务转换为MASK语言建模问题,Chen等<sup>[81]</sup>将关系标签中的知识融入关系抽取的提示调优中,并提出基于协同优化的知识感知提示调优方法(KnowPrompt)。具体

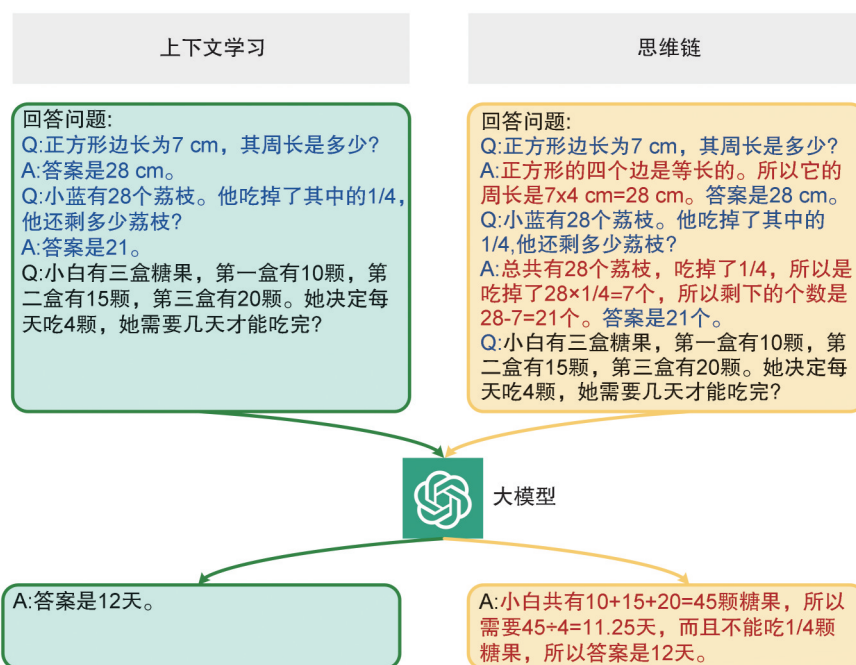


图12 上下文学习提示过程vs思维链提示过程(图片参考文献:《中国人工智能系列白皮书》,版权所有中国人工智能学会)

Fig. 12 ICL prompting process vs. CoT prompting process (see original pictures in *China AI Series White Papers*, Copyright CAAI)

来说,通过构建虚拟答案词(关系)和可以根据上下文动态调整的虚拟类型词(实体),用于关系抽取任务的知识注入,将这一步骤作为提示的构建,再通过上下文感知校准和隐式结构化约束协同优化虚拟答案词和虚拟类型词,这一步骤则作为提示的优化。

为了克服数据稀缺、弥补标注数据不足的问题,Xu等<sup>[82]</sup>提出上下文学习和数据生成策略,其中上下文学习策略通过设计包含关系抽取任务相关指令的提示,将其添加到提示符中,向提示符添加实体类型作为模式约束,帮助LLM准确识别和生成关系;数据生成策略则通过使用特定的提示引导LLM生成更多符合关系抽取任务需求的训练数据。最后通过这两个策略实现少样本关系抽取。实验结果表明,使用GPT3.5仅需一次示例进行上下文学习即可达到微调预训练语言模型相当的性能,采用数据生成策略能够提高关系抽取的性能。

由于ICL中实体和关系相关性较低,Wan等<sup>[83]</sup>提出GPT-RE,其中采用两种策略解决这个问题,即任务感知检索和事实标签诱导推理策略。任务感知检索策略采用实体对信息重建上下文和微调关系表示的方法强调实体和关系信息,事实标签诱导推理策略丰富了每个论证的

推理证据,通过实验验证结合这两种策略的上下文学习框架能够在关系抽取任务中取得很好的效果,弥补了基于GPT-3框架的局限性。

最常见的提示策略是Vanilla提示策略,该策略使LLM在一个步骤完成三个重要的推理过程,即提取句子中主语和宾语之间的关系语义、了解每个关系标签的语义以及匹配实体与给定关系标签之间的关系语义,但这种提示策略无法获得高性能。Li等<sup>[84]</sup>提出总结和提问的提示策略,将RE分解为两个子任务,一个是文本摘要任务,另一个是问答任务,使用LLM递归地将RE输入转换为有效的问答格式,能够引导LLM更好地理解 and 推理文本。

为了解决ICL方法易受概念偏差的影响,在新场景下可能出现性能下降的情况,Pang等<sup>[85]</sup>提出指导学习(Guideline Learning, GL)框架(如图13所示),在训练阶段从少量错误实例中学习高质量的指导方针,在推理阶段中收集一组指导方针,检索与输入实例相关的规则,构造查询以询问LLM代理关于实例类别的原因,对于每个实例,如果来自LLM代理的预测答案与注释(正确的规则)不同,则另一个LLM代理生成新规则并更新现有指导方针,该方法能够帮助ICL更准确地识别和分类实体和关

系。在关系抽取任务上的实验结果表明, GL 框架能够显著提升 ICL 的性能。

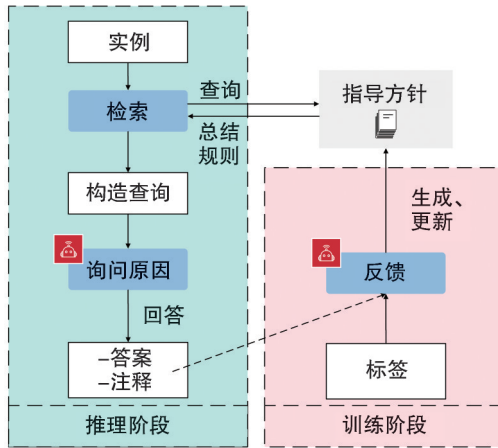


图13 指导学习框架(图片参考文献[85], 版权所有 arXiv.org)

Fig. 13 The guideline learning framework (see original pictures in Refs. [85], Copyright arXiv.org)

Wadhwa 等<sup>[86]</sup>提出 LLM+CoT+fine-tuning 的方法应用于关系抽取, 其中 CoT 是采用蒸馏技术从 GPT-3 中得到的, 以微调增强关系抽取, 在 CoNLL、ADE、NYT 三个关系抽取数据集中产生 SOTA。

在深度使用 LLM 如 ChatGPT 进行自然语言理解 (Natural Language Understanding, NLU) 任务如关系抽取任务时, 存在提示工程繁琐的问题, Yu 等<sup>[87]</sup>提出序列理解增强的 GPT (SeqGPT), 将 NLU 转换成抽取任务和分类任务, 即统一任务范式, 提供一致的输入和输出格式, 能够应用在模型没见过的任务上。

LLM 生成的回答是基于概率的, 导致生成的关系缺乏确定性, 不具有高置信度。为了进一步提高生成输出的置信度, Ling 等<sup>[88]</sup>设计了不确定性量化模块来过滤低置信度的预测。

实际上, 基于大模型的关系抽取方法仍面临不少挑战, 例如在提示学习中构建提示模板时需要专业的领域知识, 需要大量的数据进行验证, 以及标签词的检索过程复杂度高, 在面对大量关系标签时难以获得合适的标签词。

## 5 面临的挑战及未来研究趋势

### 5.1 面临的挑战

#### (1) 数据噪声

关系抽取正面临着数据噪声带来的挑战,

噪声来源多样, 比如同义词、同音字误用、信息遗漏以及错误标注等。其中错误标注造成的噪声最为显著, 最终导致错误传播, 严重影响了模型的训练过程, 降低了关系抽取的准确率和可靠性, 而错误标注问题是由远程监督策略引起的, 因此, 目前大多数研究都是针对远程监督策略进行改善, 从而降低错误标注带来的影响, 进而缓解数据噪声问题。

#### (2) 跨领域泛化能力不足

当前关系抽取通常在单一领域内通过人工标注数据集进行模型训练, 这种方法虽然能够确保高度准确的训练过程, 产生性能优异的模型, 但其局限性在于生成的数据集规模有限、扩展困难, 并且受限于特定领域。尽管基于 Transformer 的预训练语言模型具备一定的跨领域学习能力, 但是这种间接的泛化策略并未从根本上解决领域适应问题, 模型在面对跨领域时未能充分利用目标领域的特性, 导致泛化能力受限。

#### (3) 重叠关系

重叠关系是关系抽取领域中一个亟待解决的主要问题。根据实体重叠程度可以将重叠关系划分为两类: 单实体重叠即某个实体与多个实体之间存在多重关系, 以及实体对重叠即某个实体与另一个实体之间存在多重关系。目前, 基于参数共享的方法虽然能够缓解该问题, 但无法避免曝光偏差。此外, 学者们还针对该问题改进解码器, 例如指针网络和片段分类, 但仍未能从根本上解决关系抽取中的重叠关系问题。

## 5.2 未来研究趋势

#### (1) 多模态信息融合

多模态数据涵盖了文本、图像、语言等自然语言, 尽管当前研究多集中在文本数据, 但是已有学者将图像、语音等信息结合文本进行研究, 虽然能够解决单模态歧义问题, 但其抽取准确率不高, 因此需要更有效的策略挖掘多模态数据潜力。换言之, 关系抽取领域不应仅限于文本, 需探索图像、语音、视频等多模态, 以实现更多元的认知智能。随着人工智能从感知迈向认知, 更全面的数据层次是多模态信息融合未来的发展趋势。

## (2) 文档级关系抽取

目前大多数的研究都以句子内部作为切入点,然而文档级关系抽取的复杂性通常由于不同信息分布在不同的句子中,需要通过跨句子的推理挖掘深层关系,要求关系抽取模型具备较高的逻辑推理能力。因此,如何整合跨句甚至跨段落的关系是目前关系抽取领域亟待解决的问题。通过调研发现,当前文档级关系抽取的研究绝大多数基于图结构,而图注意力卷积神经网络的提出,为建立多层次的图结构模型提供了可能。此外,跨文档的关系抽取也成为新的研究热点,它要求模型在更宽泛的文本范围内捕捉关系信息,进一步扩展关系抽取的边界。

## (3) 无监督学习关系抽取

无监督学习关系抽取采用未进行人工标注的训练语料进行训练,目前大多数无监督学习研究基于聚类的思想,因此,改进聚类算法以提高关系抽取的准确率至关重要,包括动态调整聚类阈值、采用自适应聚类策略,以适应不同领域和数据规模的需求。此外,鉴于无监督学习关系抽取缺乏统一的评估标准,构建客观、全面的性能评价体系,是推进该领域发展的关键一环。尽管当前无监督学习的研究程度不如监督学习,尚处于探索阶段,但是其在领域适应性、数据规模处理潜力及减轻标注依赖方面的独特优势,使其成为未来的研究热点。

## 6 结论

关系抽取作为数据挖掘、信息检索及NLP领域的核心研究内容之一,从早期基于规则与模板的初步实践,到机器学习的应用,尽管这些方法在当前学术界已非主流,但是它们奠定的基础为后续技术的发展提供了宝贵的经验。随着深度学习技术在NLP领域的不断成熟,尤其是近年来大语言模型的创新高潮,其在RE任务上正展现出前所未有的潜力与广泛的应用前景,这一进展已成为国内外学术界的研究热点。

本文围绕关系抽取的研究进展进行系统性论述,从关系抽取的发展背景出发,详细综述了关系抽取的各种方法,总结了研究者在关系

抽取领域中的贡献,最后分析了关系抽取领域依然面临的挑战与未来研究趋势以推进这一领域的发展。

## 参考文献:

- [1] CHINCHOR N, MARSCH E. MUC-7 Information Extraction Task Definition[C]//Proceedings of a Seventh Message Understanding Conference (MUC-7). Fairfax: the MIT Press, 1998.
- [2] CHINCHOR N A. Overview of MUC-7[C]//Proceedings of a Seventh Message Understanding Conference (MUC-7). Fairfax: The MIT Press, 1998.
- [3] 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 Workshop on Semantic Evaluations: Recent Achievements and Future Directions-SEW 2009. Boulder: ACL, 2009: 94-99. DOI:10.3115/1621969.1621986.
- [4] TAYLOR A, MARCUS M, SANTORINI B. The Penn Treebank: An Overview[M]//Treebanks. Dordrecht: Springer Netherlands, 2003: 5-22. DOI:10.1007/978-94-010-0201-1\_1.
- [5] 刘绍毓,李弼程,郭志刚,等. 实体关系抽取研究综述[J]. 信息工程大学学报, 2016, 17(5): 541-547. DOI: 10.3969/j.issn.1671-0673.2016.05.006.  
LIU S Y, LI B C, GUO Z G, *et al.* Review of Entity Relation Extraction[J]. *J Inf Eng Univ*, 2016, 17(5): 541-547. DOI: 10.3969/j.issn.1671-0673.2016.05.006.
- [6] RILOFF E M. Automatically Constructing a Dictionary for Information Extraction Tasks[C]//National Conference on Artificial Intelligence. Cambridge: AAAI Press, 1993.
- [7] APPELT D E, HOBBS J R, BEAR J, *et al.* FASTUS: A Finite-state Processor for Information Extraction from Real-world Text[C]//Proceedings of the 13th International Joint Conference on Artificial Intelligence. San Mateo: Morgan Kaufmann, 1993.
- [8] BRIN S. Extracting Patterns and Relations from the World Wide Web[M]//The World Wide Web and Databases. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999: 172-183. DOI:10.1007/10704656\_11.
- [9] AGICHTEIN E, GRAVANO L. Snowball: Extracting Relations from Large Plain-text Collections[C]//Proceedings of the Fifth ACM Conference on Digital Libraries. New York: Association for Computing Machinery, 2000. DOI: 10.1145/336597.336644.
- [10] KAMBHATLA N. Combining Lexical, Syntactic, and Semantic Features with Maximum Entropy Models for Extracting Relations[C]//Proceedings of the 42nd Annual

- Meeting of the Association for Computational Linguistics. Barcelona: ACL, 2004. DOI: 10.3115/1219044.1219066.
- [11] TRATZ S, HOVY E. ISI: Automatic Classification of Relations Between Nominals Using a Maximum Entropy Classifier[C]//Proceedings of the 5th International Workshop on Semantic Evaluation. Uppsala: ACL, 2010.
- [12] COLLINS M, DUFFY N. New Ranking Algorithms for Parsing and Tagging: Kernels over Discrete Structures, and the Voted Perceptron[C]//Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Philadelphia: ACL, 2002: 263–270. DOI: 10.3115/1073083.1073128.
- [13] ZHOU G D, SU J, ZHANG J, *et al.* Exploring Various Knowledge in Relation Extraction[C]//Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. Ann Arbor: ACL, 2005: 427–434. DOI:10.3115/1219840.1219893.
- [14] ZELENKO D, AONE C, RICARDELLA A. Kernel Methods for Relation Extraction[J]. *J Mach Learn Res*, 2003, **3**(6): 1083–1106.
- [15] CULOTTA A, MCCALLUM A, BETZ J. Integrating Probabilistic Extraction Models and Data Mining to Discover Relations and Patterns in Text[C]//Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics. New York: ACL, 2006: 296–303. DOI:10.3115/1220835.1220873.
- [16] BUNDSCHUS M, DEJORI M, STETTER M, *et al.* Extraction of Semantic Biomedical Relations from Text Using Conditional Random Fields[J]. *BMC Bioinformatics*, 2008, **9**: 207. DOI:10.1186/1471-2105-9-207.
- [17] WARIKOO N, CHANG Y C, HSU W L. LPTK: a Linguistic Pattern-aware Dependency Tree Kernel Approach for the BioCreative VI CHEMPROT Task[J]. *Database*, 2018, **2018**: bay108. DOI:10.1093/database/bay108.
- [18] BEN ABDESSALEM KARAA W, ALKHAMMASH E H, BCHIR A. Drug Disease Relation Extraction from Biomedical Literature Using NLP and Machine Learning[J]. *Mob Inf Syst*, 2021, **2021**(1): 9958410. DOI: 10.1155/2021/9958410.
- [19] LIU Z H, CHEN Y Y, DAI Y F, *et al.* Syntactic and Semantic Features Based Relation Extraction in Agriculture Domain[M]//Web Information Systems and Applications. Cham: Springer International Publishing, 2018: 252–258. DOI:10.1007/978-3-030-02934-0\_23.
- [20] THOMAS A, SIVANESAN S. An Adaptable, High-performance Relation Extraction System for Complex Sentences[J]. *Knowl Based Syst*, 2022, **251**: 108956. DOI:10.1016/j.knosys.2022.108956.
- [21] LIANG C, ZAN H, LIU Y, *et al.* Research on Entity Relation Extraction for Military Field[C]//Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation. Hong Kong: ACL, 2018.
- [22] KIM Y. Convolutional Neural Networks for Sentence Classification[EB/OL]. (2014-09-03) [2024-12-10]. <https://arxiv.org/abs/1408.5882>.
- [23] LIU C Y, SUN W B, CHAO W H, *et al.* Convolution Neural Network for Relation Extraction[C]//Advanced Data Mining and Applications. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013: 231–242. DOI: 10.1007/978-3-642-53917-6\_21.
- [24] ZENG D, LIU K, LAI S, *et al.* Relation Classification via Convolutional Deep Neural Network[C]//International Conference on Computational Linguistics. the 25th International Conference on Computational Linguistics. Dublin: ACL, 2014.
- [25] NGUYEN T H, GRISHMAN R. Relation Extraction: Perspective from Convolutional Neural Networks[C]//Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing. Denver: ACL, 2015: 39–48. DOI:10.3115/v1/w15-1506.
- [26] DOS SANTOS C N, XIANG B, ZHOU B W. Classifying Relations by Ranking with Convolutional Neural Networks[J]. *Computer Science*, 2015, **86**(86):132–137. DOI:10.3115/v1/P15-1061.
- [27] XU K, FENG Y S, HUANG S F, *et al.* Semantic Relation Classification via Convolutional Neural Networks with Simple Negative Sampling[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon: ACL, 2015: 536–540. DOI: 10.18653/v1/d15-1062.
- [28] SAHU S, ANAND A, ORUGANTY K, *et al.* Relation Extraction from Clinical Texts Using Domain Invariant-convolutional Neural Network[C]//Proceedings of the 15th Workshop on Biomedical Natural Language Processing. Berlin: ACL, 2016: 206–215. DOI: 10.18653/v1/w16-2928.
- [29] GU J H, SUN F Q, QIAN L H, *et al.* Chemical-induced Disease Relation Extraction via Convolutional Neural Network[J]. *Database*, 2017, **2017**(1): bax024. DOI: 10.1093/database/bax024.
- [30] SHEN Y, HUANG X. Attention-based Convolutional Neural Network for Semantic Relation Extraction[C]//Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. Osaka: the COLING 2016 Organizing Commit-

- tee. 2016.
- [31] WANG L L, CAO Z, DE MELO G, *et al.* Relation Classification via Multi-level Attention CNNs[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Berlin: ACL, 2016. DOI:10.18653/v1/P16-11.
- [32] ASADA M, MIWA M, SASAKI Y. Extracting Drug-drug Interactions with Attention CNNs[C]//Proceedings of the 16th BioNLP Workshop. Vancouver: ACL, 2017: 9-18. DOI:10.18653/v1/w17-2302.
- [33] ZHU J Z, QIAO J Z, DAI X X, *et al.* Relation Classification via Target-concentrated Attention CNNs[M]//Neural Information Processing. Cham: Springer International Publishing, 2017: 137-146. DOI: 10.1007/978-3-319-70096-0\_15.
- [34] SOCHER R, HUVAL B, MANNING C D, *et al.* Semantic Compositionality through Recursive Matrix-Vector Spaces[C]//Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Jeju Island: ACL, 2012: 1201-1211.
- [35] ZHANG D X, WANG D. Relation Classification via Recurrent Neural Network[EB/OL]. (2015-12-25) [2024-12-10]. <https://arxiv.org/abs/1508.01006>.
- [36] XU Y, MOU L L, LI G, *et al.* Classifying Relations via Long Short Term Memory Networks along Shortest Dependency Paths[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon: ACL, 2015: 1785-1794. DOI: 10.18653/v1/d15-1206.
- [37] NINGTHOUJAM D, YADAV S, BHATTACHARYYA P, *et al.* Relation Extraction Between the Clinical Entities Based on the Shortest Dependency Path Based LSTM[EB/OL]. (2019-03-24) [2024-12-10]. <https://arxiv.org/abs/1903.09941>.
- [38] ZHOU H W, DENG H J, CHEN L, *et al.* Exploiting Syntactic and Semantics Information for Chemical-Disease Relation Extraction[J]. *Database*, 2016, **2016**: baw048. DOI:10.1093/database/baw048.
- [39] LIM S, LEE K, KANG J. Drug Drug Interaction Extraction from the Literature Using a Recursive Neural Network[J]. *PLoS One*, 2018, **13**(1): e0190926. DOI: 10.1371/journal.pone.0190926.
- [40] ZHANG S, ZHENG D, HU X, *et al.* Bidirectional Long Short-Term Memory Networks for Relation Classification[C]//Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation. Shanghai: Shanghai Jiaotong University, 2015: 73-78.
- [41] ZHOU P, SHI W, TIAN J, *et al.* Attention-based Bidirectional Long Short-term Memory Networks for Relation Classification[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Berlin: ACL, 2016: 207-212. DOI:10.18653/v1/p16-2034.
- [42] XU Y, JIA R, MOU L L, *et al.* Improved Relation Classification by Deep Recurrent Neural Networks with Data Augmentation[EB/OL]. (2016-10-13) [2024-12-10]. <https://arxiv.org/abs/1601.03651v2>.
- [43] WANG H L, QIN K, LU G M, *et al.* Direction-sensitive Relation Extraction Using Bi-SDP Attention Model[J]. *Knowl Based Syst*, 2020, **198**: 105928. DOI: 10.1016/j.knsys.2020.105928.
- [44] NGUYEN T H, GRISHMAN R. Combining Neural Networks and Log-linear Models to Improve Relation Extraction[EB/OL]. (2015-11-18) [2024-12-10]. <https://arxiv.org/abs/1511.05926>.
- [45] QIN P D, XU W R, GUO J. Designing an Adaptive Attention Mechanism for Relation Classification[C]//2017 International Joint Conference on Neural Networks (IJCNN). New York: IEEE, 2017: 4356-4362. DOI: 10.1109/IJCNN.2017.7966407.
- [46] LUO X, ZHOU W W, WANG W P, *et al.* Attention-based Relation Extraction with Bidirectional Gated Recurrent Unit and Highway Network in the Analysis of Geological Data[J]. *IEEE Access*, 2017, **6**: 5705-5715. DOI:10.1109/ACCESS.2017.2785229.
- [47] SHEN Y, ZHANG L Z, ZHANG J, *et al.* CBN: Constructing a Clinical Bayesian Network Based on Data from the Electronic Medical Record[J]. *J Biomed Inform*, 2018, **88**: 1-10. DOI:10.1016/j.jbi.2018.10.007.
- [48] ZHANG Y L, LI X M, ZHANG Z. Disease-pertinent Knowledge Extraction in Online Health Communities Using GRU Based on a Double Attention Mechanism [J]. *IEEE Access*, 2020, **8**: 95947-95955. DOI:10.1109/ACCESS.2020.2995739.
- [49] ZHU H, LIN Y K, LIU Z Y, *et al.* Graph Neural Networks with Generated Parameters for Relation Extraction[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: ACL, 2019: 1331-1339. DOI:10.18653/v1/p19-1128.
- [50] SCHLICHTKRULL M, KIPF T N, BLOEM P, *et al.* Modeling Relational Data with Graph Convolutional Networks[M]//The Semantic Web. Cham: Springer International Publishing, 2018: 593-607. DOI: 10.1007/978-3-319-93417-4\_38.
- [51] ZHANG Y H, QI P, MANNING C D. Graph Convolu-

- tion over Pruned Dependency Trees Improves Relation Extraction[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels: ACL, 2018: 2205 – 2215. DOI:10.18653/v1/d18-1244.
- [52] GUO Z J, ZHANG Y, LU W. Attention Guided Graph Convolutional Networks for Relation Extraction[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: ACL, 2019: 241–251. DOI:10.18653/v1/p19-1024.
- [53] PARK C, PARK J, PARK S. AGCN: Attention-based Graph Convolutional Networks for Drug-drug Interaction Extraction[J]. *Expert Syst Appl*, 2020, **159**: 113538. DOI:10.1016/j.eswa.2020.113538.
- [54] TIAN Y H, CHEN G M, SONG Y, *et al.* Dependency-driven Relation Extraction with Attentive Graph Convolutional Networks[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). Virtual, Online: ACL, 2021: 4458–4471. DOI:10.18653/v1/2021.acl-long.344.
- [55] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention is All You Need[EB/OL]. (2023-08-02)[2024-12-10]. <https://arxiv.org/abs/1706.03762>.
- [56] RADFORD A, NARASIMHAN K, SALIMANS T, *et al.* Improving Language Understanding by Generative Pre-training[EB/OL]. (2018) [2024-12-10]. <https://api.semanticscholar.org/CorpusID:49313245>.
- [57] DEVLIN J, CHANG M W, LEE K, *et al.* Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding[EB/OL]. (2019-05-24) [2024-12-10]. <https://arxiv.org/abs/1810.04805>.
- [58] ALT C, HÜBNER M, HENNIG L. Improving Relation Extraction by Pre-trained Language Representations [EB/OL]. (2019-06-07) [2024-12-10]. <https://arxiv.org/abs/1906.03088>.
- [59] PAPANIKOLAOU Y, PIERLEONI A. DARE: Data Augmented Relation Extraction with GPT-2[EB/OL]. (2020-04-06) [2024-12-10]. <https://arxiv.org/abs/2004.13845>.
- [60] LUO R Q, SUN L A, XIA Y C, *et al.* BioGPT: Generative Pre-trained Transformer for Biomedical Text Generation and Mining[J]. *Brief Bioinform*, 2022, **23**(6): bbac409. DOI:10.1093/bib/bbac409.
- [61] WU S C, HE Y F. Enriching Pre-trained Language Model with Entity Information for Relation Classification[C]//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. New York: ACM, 2019: 2361–2364. DOI: 10.1145/3357384.3358119.
- [62] BALDINI SOARES L, FITZGERALD N, LING J, *et al.* Matching the Blanks: Distributional Similarity for Relation Learning[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: ACL, 2019: 2895–2905. DOI:10.18653/v1/p19-1279.
- [63] ZHAO Y, WAN H, GAO J, *et al.* Improving Relation Classification by Entity Pair Graph[C]//Asian Conference on Machine Learning. Nagoya: PMLR, 2019: 1156–1171.
- [64] PETERS M E, NEUMANN M, LOGAN I R, *et al.* Knowledge Enhanced Contextual Word Representations [C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Hong Kong: ACL, 2019: 43–54. DOI: 10.18653/v1/D19-1005.
- [65] SHI P, LIN J J. Simple BERT Models for Relation Extraction and Semantic Role Labeling[EB/OL]. (2019-04-10) [2024-12-10]. <https://arxiv.org/abs/1904.05255>.
- [66] LI C, YE T. Downstream Model Design of Pre-trained Language Model for Relation Extraction Task[EB/OL]. (2020-04-08) [2024-12-10]. <https://arxiv.org/abs/2004.03786>.
- [67] MINTZ M, BILLS S, SNOW R, *et al.* Distant Supervision for Relation Extraction without Labeled Data[C]//Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-ACL-IJCNLP '09. Suntec: ACL, 2009: 1003–1011. DOI:10.3115/1690219.1690287.
- [68] RIEDEL S, YAO L M, MCCALLUM A. Modeling Relations and Their Mentions without Labeled Text[C]//Machine Learning and Knowledge Discovery in Databases. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010: 148–163. DOI:10.1007/978-3-642-15939-8\_10.
- [69] ZENG D J, LIU K, CHEN Y B, *et al.* Distant Supervision for Relation Extraction via Piecewise Convolutional Neural Networks[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon: ACL, 2015: 1753–176. DOI: 10.18653/v1/d15-1203.
- [70] LIN Y K, SHEN S Q, LIU Z Y, *et al.* Neural Relation Extraction with Selective Attention over Instances[C]//Proceedings of the 54th Annual Meeting of the Associa-

- tion for Computational Linguistics (Volume 1: Long Papers). Berlin: ACL, 2016: 2124–2133. DOI: 10.18653/v1/p16-1200.
- [71] JI G L, LIU K, HE S Z, *et al.* Distant Supervision for Relation Extraction with Sentence-level Attention and Entity Descriptions[C]//Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco: AAAI press, 2017: 3060–3066. DOI: 10.1609/aaai.v31i1.10953.
- [72] JIANG X, WANG Q, LI P, *et al.* Relation Extraction with Multi-instance Multi-label Convolutional Neural Networks[C]//Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. Osaka: the COLING 2016 Organizing Committee, 2016.
- [73] HAN X, LIU Z Y, SUN M S. Neural Knowledge Acquisition *via* Mutual Attention Between Knowledge Graph and Text[C]//Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans: AAAI press, 2018: 4832–4839. DOI:10.1609/aaai.v32i1.11927.
- [74] FENG J, HUANG M L, ZHAO L, *et al.* Reinforcement Learning for Relation Classification from Noisy Data [C]//Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans: AAAI press, 2018, **32** (1): 5780–5786. DOI:10.1609/aaai.v32i1.12063.
- [75] QIN P, XU W, WANG W Y. Robust Distant Supervision Relation Extraction via Deep Reinforcement Learning [C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne: ACL, 2018: 2137–2147. DOI: 10.18653/v1/P18-1199.
- [76] QIN P D, XU W R, WANG W Y. DSGAN: Generative Adversarial Training for Distant Supervision Relation Extraction[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne: ACL, 2018: 496–505. DOI:10.18653/v1/p18-1046.
- [77] ALT C, HÜBNER M, HENNIG L. Fine-tuning Pre-trained Transformer Language Models to Distantly Supervised Relation Extraction[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: ACL, 2019: 1388–1398. DOI:10.18653/v1/p19-1134.
- [78] YU H T, CAO Y, CHENG G, *et al.* Relation Extraction with BERT-based Pre-trained Model[C]//2020 International Wireless Communications and Mobile Computing (IWCMC). New York: IEEE, 2020: 1382–1387. DOI:10.1109/iwcmc48107.2020.9148384.
- [79] BROWN T, MANN B, RYDER N, *et al.* Language Models are Few-Shot Learners[EB/OL]. (2020-07-22) [2024-12-10]. <https://arxiv.org/abs/2005.14165>.
- [80] WEI J, WANG X, SCHUURMANS D, *et al.* Chain-of-Thought Prompting Elicits Reasoning in Large Language Models[EB/OL]. (2023-01-10) [2024-12-10]. <https://arxiv.org/abs/2201.11903>.
- [81] 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 Conference 2022. New York: ACM, 2022: 2778 – 2788. DOI:10.1145/3485447.3511998.
- [82] XU X, ZHU Y Q, WANG X H, *et al.* How to Unleash the Power of Large Language Models for Few-shot Relation Extraction?[C]//Proceedings of The Fourth Workshop on Simple and Efficient Natural Language Processing (SustainLP). Toronto: ACL, 2023: 190–200. DOI:10.18653/v1/2023.sustainlp-1.13.
- [83] WAN Z, CHENG F, MAO Z, *et al.* GPT-RE: In-context Learning for Relation Extraction Using Large Language Models[EB/OL]. (2023-12-09) [2024-12-10]. <https://arxiv.org/abs/2305.02105>.
- [84] GUOZHENG LI, PENG WANG, WENJUN KE. 20. Revisiting Large Language Models as Zero-shot Relation Extractors[EB/OL]. (2023-11-24) [2024-12-10]. <https://arxiv.org/abs/2310.05028>.
- [85] PANG C X, CAO Y X, DING Q, *et al.* Guideline Learning for In-context Information Extraction[EB/OL]. (2023-10-21) [2024-12-10]. <https://arxiv.org/abs/2310.05066>.
- [86] WADHWA S, AMIR S, WALLACE B C. Revisiting Relation Extraction in the Era of Large Language Models[C]//Proceedings of the Annual Meeting of the Association for Computational Linguistics, Toronto: ACL, 2023, **1**: 15566–15589. DOI: 10.18653/v1/2023.acl-long.868.
- [87] YU T Y, JIANG C Y, LOU C, *et al.* SeqGPT: an Out-of-the-box Large Language Model for Open Domain Sequence Understanding[C]//Proceedings of the AAAI Conference on Artificial Intelligence, Washington: AAAI Press, 2024, **38**(17): 19458–19467. DOI: 10.1609/aaai.v38i17.29917.
- [88] LING C, ZHAO X J, ZHANG X C, *et al.* Improving Open Information Extraction with Large Language Models: A Study on Demonstration Uncertainty[EB/OL]. (2023-09-07) [2024-12-10]. <https://arxiv.org/abs/2309.03433>.