JURIS:基于理解增强型指令微调的司法命名实体识别方法

彭 晗[1,2],阮日青[3],胡 颖[4],刘琼林[4],张 震[2,4*]

(1. 湖南工商大学智能机器人学院,湖南长沙 410000;2. 湘江实验室,湖南长沙 410000;
3. 湖南工商大学计算机学院,湖南长沙 410000;4. 湖南工商大学人工智能与先进计算学院,湖南长沙 410000)

摘 要: 命名实体识别(Named Entity Recognition,NER)是法律文本结构分析和语义理解的基础任务,能够极大提高司法效率,促进司法公正. 然而,受限于法律文本的高度复杂性与专业性,传统 NER 方法难以充分理解法律文书中的上下文关联,较多依赖于浅层的词级预测,缺乏实体角色解析与深层语境推理能力,尤其在面对司法文本中频繁出现的嵌套实体、细粒度实体以及模糊的实体边界时存在明显的局限性. 为解决上述问题,本文基于理解增强的建模范式,提出了一种面向中文法律场景的新型命名实体识别框架——JURIS(Judicial Understanding-enhanced Reasoning via Instruction-tuned Strategies for named entity recognition). 该框架将实体识别重新建模为基于语境理解的条件生成任务,通过采用创新性的上下文感知的嵌入式标注策略,在保留文本原始语义结构的同时有效增强上下文信息建模能力,从而提升复杂语境下的实体识别效果. 同时,JURIS 构建了一个由规范模块、知识引导模块和类比学习模块组成的三元理解增强模块(Tri-aspect Understanding Enhancement Module,Tri-UEM),分别从输出一致性、领域知识注入与语境类比迁移 3 个维度协同提升模型对法律领域实体语义的深层理解与判别能力. 实证结果表明,JURIS 在 CAIL2021、Drug 和 CSKS2019 等多个领域数据集上均超过现有强基线模型,取得了当前最佳性能,改善了嵌套实体处理与细粒度识别表现,并展现出其在垂直领域信息抽取任务中的广泛适用性与推广潜力.

关键词: 司法命名实体识别;理解增强;指令微调;信息抽取

基金项目: 湘江实验室重大项目(No.25XJ01001)

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112(2025)09-3117-17

电子学报 URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20250656

JURIS: Judical Understanding-Enhanced Reasoning via Instruction-Tuned Strategies for Named Entity Recognition

PENG Han[1,2], RUAN Ri-qing[3], HU Ying[4], LIU Qiong-lin[4], ZHANG Zhen[2,4*]

(1. *School of Intelligent Robotics*, *Hunan University of Technology and Business*, *Changsha*, *Hunan* 410000, *China*;
2. *Xiangjiang Laboratory*, *Changsha*, *Hunan* 410000, *China*;
3. *School of Computer Science*, *Hunan University of Technology and Business*, *Changsha*, *Hunan* 410000, *China*;
4. *School of AI and Advanced Computing*, *Hunan University of Technology and Business*, *Changsha*, *Hunan* 410000, *China*)

Abstract: Named entity recognition (NER) serves as a fundamental task in the structural analysis and semantic understanding of legal texts, with the potential to greatly enhance judicial efficiency and promote fairness. However, due to the high complexity and domain specificity of legal language, traditional NER methods struggle to adequately capture contextual dependencies in legal documents. They often rely on shallow token-level predictions, lacking both role-based entity interpretation and deeper contextual reasoning. These limitations are particularly pronounced when dealing with nested entities, fine-grained entity categories, and ambiguous boundaries that frequently occur in judicial texts. To address these challenges, this paper introduces a novel NER framework for Chinese legal scenarios, termed JURIS (judicial understanding-enhanced reasoning via instruction-tuned strategies for named entity recognition). JURIS reformulates entity recognition as a context-driven conditional generation task and adopts an innovative context-aware embedded annotation strategy, which preserves

收稿日期:2025-07-28;录用日期:2025-09-24;责任编辑:覃怀银
*通讯作者:张震

the original semantic structure of the text while effectively enhancing contextual modeling. In addition, JURIS incorporates a tri-aspect understanding enhancement module (Tri-UEM), consisting of a standardization module, a knowledge-guided module, and an analogy-based learning module. These components jointly strengthen the model's semantic understanding and discrimination ability in the legal domain by improving output consistency, injecting domain-specific knowledge, and enabling contextual analogy transfer. Experimental results demonstrate that JURIS consistently outperforms strong baseline models on multiple datasets, including CAIL2021, Drug, and CSKS2019, achieving state-of-the-art performance. It significantly improves recognition of nested and fine-grained entities while showing strong generalizability and applicability in domain-specific information extraction tasks.

Key words: judicial named entity recognition; understanding enhancement; instruction tuning; information extraction

Foundation Item(s): Major Project of Xiangjiang Laboratory (No.25XJ01001)

1 引言

截至2025年,中国裁判文书网累计公开裁判文书逾1.5亿篇,呈现出持续的指数级增长趋势. 据《最高人民法院工作报告》[1]统计,截至2025年9月,全国法官年人均办案量为354件,司法资源紧张与案件激增的矛盾日益尖锐. 在此背景下,如何从海量法律文书中高效、精确地抽取关键信息,已成为当前司法智能化建设中的核心技术瓶颈.

命名实体识别(Named Entity Recognition,NER)作为法律文本结构化处理与语义理解的基础任务,近年来受到广泛关注. 尽管已有多种NER技术在通用领域取得显著进展,如基于特征工程[2]的传统方法、基于深度学习的BiLSTM-CRF[3]等模型,但是面对司法领域特有的语义复杂性与结构多样性,这些方法仍面临诸多挑战. 具体来说,法律文本不仅高度专业化、语义表达多样,而且实体之间常存在角色歧义、同形异义及边界模糊等问题. 如图1所示,"××元"在不同上下文中可能分别表示"被盗金额"或"盗窃获利",而传统方法难以通过局部特征完成语义区分. 同时,法律文书中常使用嵌套的命名实体来明确地描述法律事实,而现有的主流NER模型主要面向平面结构,难以适应复杂结构下的实体识别需求. 此外,司法文本中许多人名或地名都是较长的复合表达式的一部分,没有明确的分隔符.

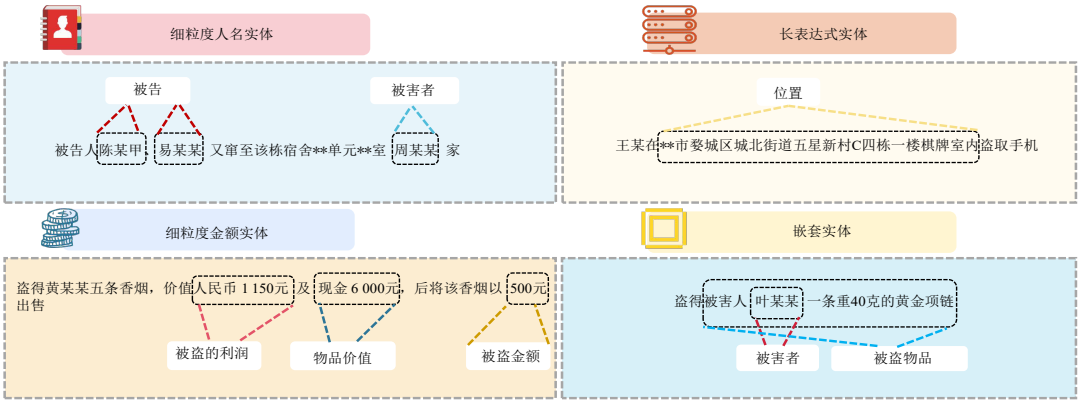


图1 司法命名实体识别中4种关键实体类型的例子

当前命名实体识别在司法文本中的应用面临诸多挑战,其根本原因不仅在于实体表征的复杂性,更在于对实体在语境中语义角色的理解能力不足. 与通用场景下的实体识别不同,法律领域的实体识别任务更依赖于上下文语义的精准建模,需要借助语境信息以消解角色歧义并捕捉法律术语中的语义微差. 因此,传统以浅层特征分类为主的"识别式"范式难以满足法律NER任务对高阶语义建模的需求. 为此,本文构建了一个以理解驱动为核心的司法命名实体识别框架,即JURIS(Judicial Understanding-enhanced Reasoning via Instruction-tuned Strategies for named entity recognition),将实体识别建模为面向语境理解的条件生成任务,并通过引入嵌入式标注机制与多维语义推理模块,在结构建模与语义对齐两个层面协同提升识别性能.

具体而言,JURIS设计了基于上下文感知的嵌入式标注策略,通过将实体标签以自然语言形式嵌入原始文本,在保持语义完整性的同时增强模型的上下文感知能力. 此外,本文提出三元理解增强模块(Tri-aspect Understanding Enhancement Module,Tri-UEM),由规范引导、知识提示与类比学习3个子模块组成,旨在实现跨类别、跨场景和跨结构的深层语义对齐与泛化建模.

本文的主要贡献总结如下.

(1)针对传统标注策略难以保留上下文语义的问题,本文提出一种新颖的基于上下文感知的嵌入式标注策略,通过将实体标签自然嵌入文本流中,为模型提供更丰富的上下文语义信息,以增强其对实体边界和语义角色的理解与推理能力.

(2)为缓解现有模型在语义理解方面不足的问题,本文设计三元理解增强模块,集成结构规范、知识引导与类比推理3类机制,分别从形式约束、先验知识注入与语义迁移3个维度协同增强模型的语境建模能力,有效应对传统方法在深层语义建构方面的局限.

(3)针对现有生成式NER方法中语义建模浅层、结构生成主导的问题,本文提出一个以理解为导向的司法NER框架——JURIS,将实体识别任务建模为基于语境理解的条件生成过程,突破传统生成式方法仅关注目标结构生成的局限.

(4)在CAIL2021与Drug两个法律领域NER数据集与CSKS2019医学数据集上进行大量实验,结果表明所提出的JURIS优于多种强基线模型,取得当前最优结果,展现出良好的通用性与扩展潜力.

2　相关工作

本节将从3个方面对已有研究工作进行综述:(1)基于特征工程与深度学习架构的传统方法;(2)利用大型语言模型的生成式命名实体识别框架;(3)针对法律领域特点设计的实体识别方法. 上述分类有助于揭示技术发展的脉络,同时突出法律文本处理中所面临的独特挑战.

2.1　传统命名实体识别方法

命名实体识别(NER)的早期研究主要依赖于手工制作的规则[4,5]和传统的基于机器学习的方法[6]. 随后,基于深度学习的NER方法出现并迅速成为主流. BiAffineNER[7]采用双仿射机制进行跨度分类,BERT-MRC[8]将NER任务转化为阅读理解问题. BOCNER[9]将BERT与有序神经元LSTM相结合,用于处理法律文本中的实体识别任务,而W^{2}NER[10]则将NER重构为词与词之间的关系分类问题. RoBERTa-BiLSTM-CRF[11]模型融合了上下文建模与序列建模能力. 尽管这些方法在实体识别方面取得了一定成效,但大多仍侧重于词级分类,缺乏对法律文本中复杂实体语义进行上下文推理的能力.

2.2　生成式命名实体识别方法

大语言模型已在命名实体识别中展现出强大的迁移与泛化能力. 其中,GLiNER[12]基于双向Transformer设计了一种轻量化的开放式NER框架. 该方法通过将实体类型与文本编码映射到统一的潜在空间,实现类型与文本Span的匹配. Lee等人[13]利用T5模型在金融文本中进行生成式实体识别. DiffusionNER[14]将命名实体识别建模为边界去噪扩散过程,通过对跨度进行去噪以生成实体. InstructUIE[15]进一步拓展了指令范式在信息抽取中的应用,提出一种基于指令微调的统一抽取框架,能够在多个子任务间共享知识并统一建模. 然而,上述方法多聚焦于通用领域任务,在面对复杂司法文本时,仍普遍面临语境理解不充分与语义歧义处理能力有限的问题.

2.3　司法领域命名实体识别方法

在法律领域,已有多种面向特定任务的模型被提出,以应对受领域约束的命名实体识别问题. Chen等人[16]介绍了一个利用BERT与BILSTM、注意力机制和CRF结合的模型,以识别涉罪报告中的关键实体,在识别常见实体(如犯罪时间戳、受害者姓名和案件处理方法)方面展示了卓越的准确性. Gu等人[17]利用BERT-BILSTM-CRF架构从盗窃案叙述中提取实体,在专有数据集上获得了89.12%的F_1分数. Hu等人[18]将ELECTRA[19]与CRF集成,对来自CAIL-2018语料库的与电信网络欺诈案件相关的文本执行NER. 虽然这些特定于任务的模型在其目标实体类型上表现良好,但它们通常将NER视为一个简单的序列标记问题. Mao等人[20]提出了一个生成式命名实体识别框架,将NER任务视为序列到序列的文本生成问题. 该方法基于BART结构,同时引入拷贝机制增强Decoder对原文实体的复制能力. 尽管上述方法在特定法律场景下表现良好,但仍主要依赖于平面序列标注方式与序列建模,忽略了在复杂法律情境中所需的深层语义推理与嵌套结构建模能力,限制了其在复杂司法场景中的适应能力.

3　JURIS

3.1　理论基础

在法律文本这一高度专业化且语言结构多样的语境中,仅依赖单一方法难以同时满足对实体边界精确性、细粒度实体类别区分及嵌套结构识别的综合需求. 本文提出的理解增强范式不仅是技术机制的整合,更建立在系统性的理论支撑之上,其整体架构如图2所示.

首先,深层语义理解的必要性. 司法命名实体识别对深层语义理解具有高度依赖性,同一表层词汇在不同司法语境中可能承担差异化的法律角色. 因此,命名实体识别不仅是词级分类问题,更是语义推理与角色判定问题,实体的界定往往不能凭借表层特征,而必须结合上下文逻辑和语义角色完成推理.

其次,语义建模的合理性. 传统的离散化标注方式往往割裂了实体与上下文的连续性,导致语义信息损

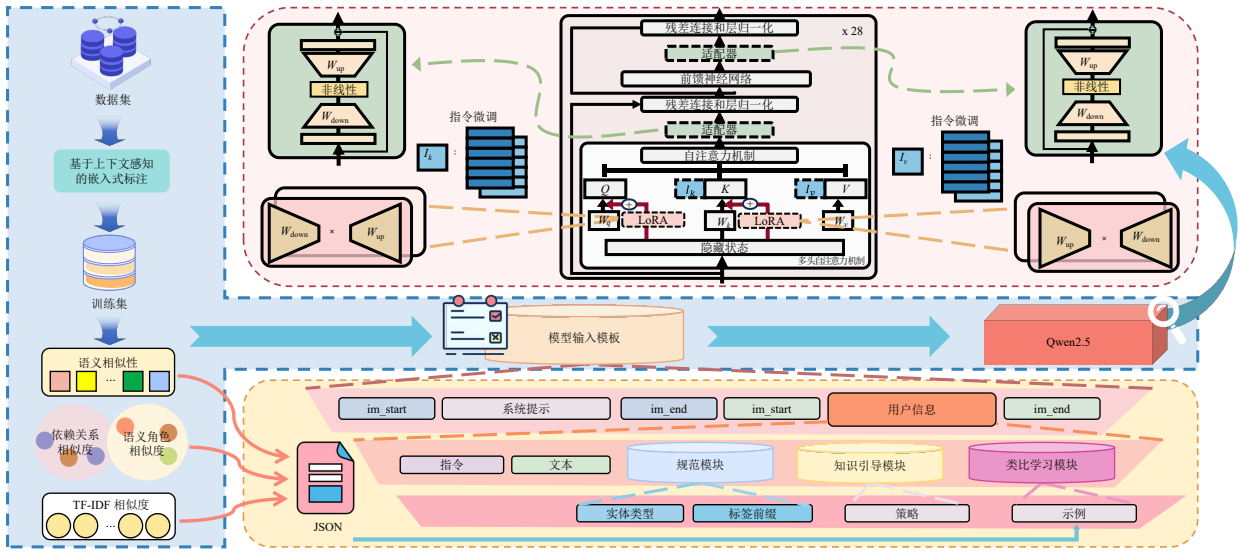


图2 JURIS模型的整体结构(该框架集成了规范模块、知识引导模块和类比学习模块)

失. 对于包含嵌套实体和边界模糊的法律文本而言,语义一体化建模能够更好地保持语义的完整性,并在生成过程中能够动态感知实体与上下文的依赖关系,从而强化对复杂语篇结构的理解.

最后,多维度互补建模的系统性. 司法文本的复杂性决定了单维度的语义建模视角往往难以覆盖全部需求,需要在理论层面引入多元化的理解框架. 不同的理论取向在输出一致性、知识利用以及语义迁移等方面各有优势,将其有机结合能够形成互补效应,进而在整体上提升模型的稳健性与适应性. 这种多维度协同的思路,为"理解增强"提供了系统而全面的理论根基.

基于上述范式,本文提出了以理解驱动为核心的命名实体识别框架JURIS,其将语义理解任务解构为多个互补的建模维度,具体包括:(1)上下文语义建模(通过上下文感知的嵌入式标注策略);(2)输出结构规范约束(规范模块);(3)领域知识建构(知识引导模块);(4)跨场景语义迁移建模(类比学习模块).

3.2 基于上下文感知的嵌入式标注策略

现有主流标注方法如span[21]或BIO[22]在命名实体识别任务中被广泛应用,但在生成式大模型框架下存在明显的适配障碍. BIO标注需要对每个标记附加标签,显著增加了输入长度,同时引入非自然语言符号(如B-、I-、O),削弱语言模型的结构建模能力与上下文理解能力. Span方法虽结构简洁,但仅提供起止位置信息,缺乏实体类型与上下文之间的显式语义关系表达,尤其在中文法律文本中,面对嵌套结构与边界模糊等现象时,表现出建模能力的不足. 此外,部分研究尝试采用实体列表形式进行输出,即将预测实体按类别组织为列表结构. 尽管在形式上简化了输出格式,但在处理实体位置信息、标签层级关系与上下文歧义等问题上仍面临显著挑战,尤其在多层嵌套与标签精细化场景下,易造成语义缺失或匹配偏差.

为克服上述限制,本文提出一种基于上下文感知的嵌入式标注策略,充分结合大语言模型的自回归生成特性,将实体类型信息直接嵌入到自然语言输出序列中,构建具有上下文一体化的标注形式. 该策略通过特殊格式【LABEL:Entity】标记,将实体类型与实体内容嵌入原始文本,使得模型在生成过程中能够动态感知实体与上下文之间的语义依赖关系,如表1所示.

表1 上下文感知的嵌入式标注示例

ID	f414e3564d99d0f9853cb786e3636965
原始文本	2017年8月10日左右的一天13时许,被告人黄某某在浙江**领带服饰有限公司宿舍一楼,从郭某某停放的电动自行车储物箱内,窃得"lovme"牌手机一部,计价值285元人民币.
输出	【NT:2017年8月10日左右的一天13时许】,被告人【NHCS:黄某某】在【NS:浙江**领带服饰有限公司宿舍一楼】,从【NHVI:郭某某】停放的电动自行车储物箱内,窃得【NASI:"lovme"牌手机一部】,计价值【NCGV:285元人民币】.

通过聚焦于完整的上下文语义,模型在训练阶段能够在连续文本中直接感知并建模实体及其语境关联,不必依赖孤立的Span结构或预定义的实体列表进行推断,从而显著增强了语义建模能力. 同时,该策略也为后续的三元理解增强模块提供了更为连续、自然的语义输入表示,如图3所示. 形式上,注释文本是通

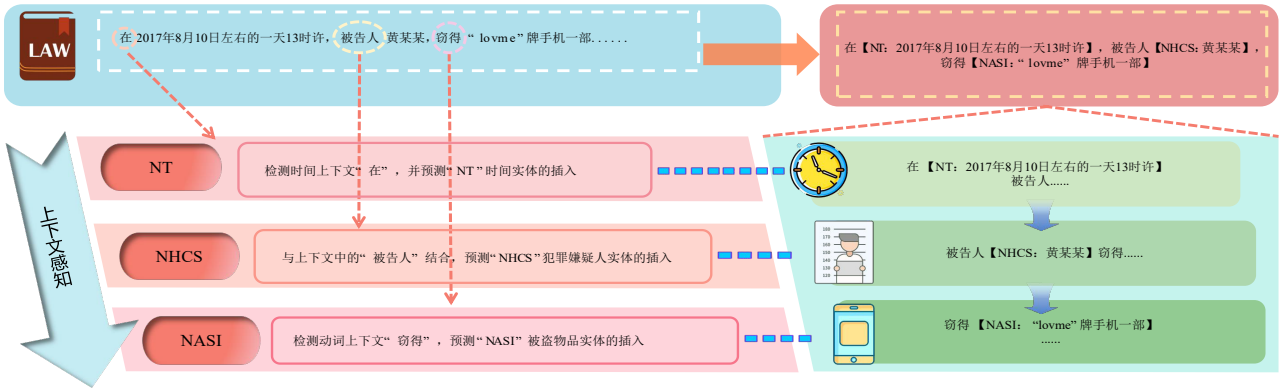
过将实体集嵌入到原始文本中来构造的,表示为

$$T'=\sum_{i=1}^{|T|}t_i+\sum_{j=1}^{|E|}[\text{LABEL}_j:e_j] \tag{1}$$

其中,$T=\{t_1,t_2,\cdots,t_{|T|}\}$表示原始文本中的标记序列,$E=\{e_1,e_2,\cdots,e_{|E|}\}$表示需要标注的实体集合;$[\text{LABEL}_j:e_j]$表示文本中每个实体特定的标注格式,表示嵌入实体标签后的标注文本,用于训练输入. 例如,原始文本T=“被告人黄某某盗窃手机”,嵌入后的文本为“被告人【NHCS:黄某某】盗窃【NASI:手机】”.

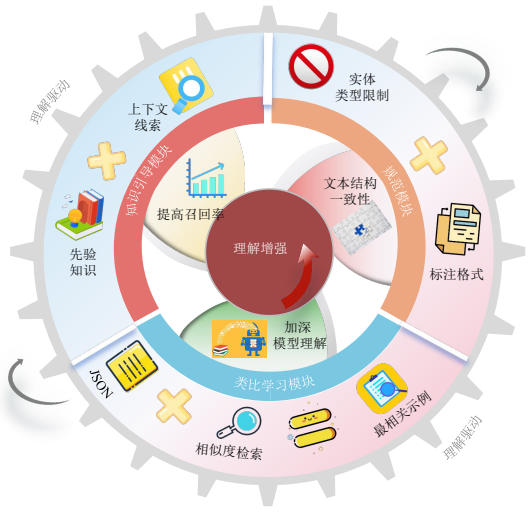
3.3　三元理解增强模块(Tri-UEM)

三元理解增强模块(Tri-UEM)构成以理解为核心的建模范式中的关键组件,如图4所示. 该模块集成了3种互补的子模块:规范模块、知识指导模块和类比学习模块. 三者并非独立执行任务,而是在统一理解增强目标下协同运行,通过形式一致性约束、先验知识注入及类比迁移机制,从不同维度加强对实体的上下文推理与语义对齐,以提升模型对语境的综合建模能力.



注:基于LLM自回归特性,充分利用每个实体周围的上下文来生成标签序列.

图3　基于上下文感知的嵌入式标注策略



注:由3个关键的子模块组成,基于“理解增强”的概念协同工作.

图4　三元理解增强模块(Tri-UEM)

3.3.1　规范模块

法律实体的识别高度依赖于统一且规范的标注方式,生成式大语言模型在输出方面具有高度灵活性,若缺乏有效约束,易导致实体标注格式不统一、实体类型错误等问题,从而降低实体识别的精度. 为此,本文设计规范模块用以强制模型遵循预定义的实体类型集合与标注格式规范,最大限度地降低了产生不相关或错误信息的风险. 该模块基于两个基本原则运行.

(1)实体类型限制:设$\mathcal{E}=\{\tau_1,\tau_2,\cdots,\tau_n\}$为预定义的法律实体类型集合,在模型生成输出时强制约束模型只识别和输出集合中预定义的实体类型,杜绝生成超出任务范围的无效实体类别. 可形式化表示为约束识别函数R_{Entity}:

$$R_{\text{Entity}}(T)=\{(\tau_i,e_i)\}(i=1,2,\cdots,k),\quad \tau_i\in\mathcal{E},e_i\in T \tag{2}$$

其中,T表示输入文本,e_i为模型从文本T中识别的实体,(τ_i,e_i)为模型在文本上输出的“类型—实体”对.

(2)标注格式一致性:所有输出实体均须符合上下文感知嵌入式标注格式,即使用统一结构【LABEL:Entity】,可由一致性函数F_{label}表示:

$$F_{\text{label}}(T,\mathcal{E})=\left\{[\tau:e]\ \middle|\ (\tau,e)\in R_{\text{Entity}}(T)\right\} \tag{3}$$

其中,$[\tau:e]$为统一标注格式,即将识别得到的每个$(\tau:e)$转换为统一的嵌入式标注片段$[\tau:e]$,形成一致的标注输出. 该约束机制显著降低了输出噪声,确保实体结构清晰、风格统一,从而提高模型的生成精度与解释性.

3.3.2　知识引导模块

法律实体通常具备强语境依赖性与领域特定语义,不同类型实体之间的区分不仅依赖表层词汇差异,更依赖于其在上下文中的语义功能. 例如,“物品价值”和“盗窃利润”之间虽在字面上相似,但其法律意义和

句法功能却存在本质差异. 为了增强模型在复杂语境下对实体的识别与召回能力,本文引入知识引导模块,注入不同实体类型的语义先验.

本文基于GPT-4o生成覆盖多实体类别的识别策略集. 例如,针对CAIL2021数据集,构造Prompt如下.

“请为入室盗窃案件的司法文本中涉及的10种实体类型(嫌疑人、被盗物品、地点、受害者、时间、物品价值、被盗货币、机构、犯罪手段、盗窃利润)分别制定识别策略和方法,需要明确每种实体在文本中的语法位置、上下文关键词特征及典型表述模式,最终输出可直接用于模型训练的实体识别规则,辅助模型区分司法场景下的实体语义角色.”

所生成的策略涵盖了实体的常见语法位置、上下文搭配模式、典型词汇提示等内容. 实体类型与其对应的引导策略详见表2与表3. 通过将此先验知识与上下文线索引入模型的训练过程,有效增强了模型对实体类型的深层次理解能力,可更准确建模法律语境下实体的功能角色与语义表达形式,从而实现面向理解的实体识别.

表2 知识引导策略(Drug)

实体类型	知识引导策略
被告人(NH)	通常指与案件相关的个人,经常出现在“被告”或“吸毒者”等词汇之后
毒品类型(NDR)	案件中涉及的毒品类型,如“海洛因”“甲基苯丙胺”等
地点(NS)	通常是与案件发生时相关的地理位置,在描述案件发生地时,通常以“在”或“位于”引导
时间(NT)	描述案件发生的时间点,例如 “于××年××月××日××时”
毒品重量(NW)	案件中涉及的毒品重量或数量,通常表示为数值后跟“克”

表3 知识引导策略(CAIL2021)

实体类型	知识引导策略
犯罪嫌疑人(NHCS)	通常是与案件有关的个体,经常出现在句子的主语或宾语位置以及“被告人”等词汇之后
被盗物品(NASI)	案情中提到的、被失窃的物品,通常出现在“被盗”“窃得”等动词的宾语位置,例如“手机”
地点(NS)	通常是与案件发生时相关的地理位置,在描述案件发生地时,通常以“在”或“位于”引导
受害人(NHVI)	通常是案件中受到伤害或损失的个体或组织
时间(NT)	通常描述案件发生的日期或具体时间点,紧随“在”“发生于”“当天”等时间定位词
物品价值(NCGV)	描述被盗物品的金额,通常为数值和货币单位的组合
被盗货币(NCSM)	通常是指被窃取的具体金额,一般以数值和货币单位出现
机构(NO)	通常是与案件相关的组织或部门单位名称,可能以专有名词形式出现
作案工具(NATS)	通常指的是犯罪过程中使用的具体工具,一般在描述犯罪行为的动词后的工具名词
盗窃获利(NCSP)	犯罪分子通过盗窃获得的金额或非法收益,一般是数值+货币单位的形式

3.3.3 类比学习模块

为进一步提升模型在复杂法律语境中对模糊实体和长依赖实体的语义理解能力,本文设计了一种基于情境映射的类比学习模块. 该模块改进了传统基于表层相似度的示例检索机制,通过结构对齐与语义迁移策略,从训练语料中自动选取功能相似的类比样本,用于增强提示,辅助模型完成当前输入的语义解析与实体识别任务. 所有候选样本均来自训练集,确保测试集样本不被用于检索,避免数据泄露.

具体而言,本文构建了一个多维结构融合的情境映射相似度函数,首次将命名实体中的上下文语义角色纳入跨样本结构对齐的核心评价指标. 该相似度函数综合考虑以下4个方面:

(1)语义嵌入相似度(S_sem):衡量句子在预训练模型空间中的语义接近程度;

(2)TF-IDF特征相似度(S_tfidf):反映文本在词项分布层面的关键词重合程度;

(3)依存句法结构相似度(S_dep):考察句法依赖图中主要成分间的结构一致性;

(4)语义角色相似度(S_role):衡量实体在句中所承担的语义功能是否一致.

令输入文本为x,候选类比样本为x',则其情境映射相似度函数定义为

$$A(x,x')=\alpha\cdot \mathrm{S\_sem}(x,x')+\beta\cdot \mathrm{S\_tfidf}(x,x') +\gamma\cdot \mathrm{S\_dep}(x,x')+\delta\cdot \mathrm{S\_role}(x,x') \tag{4}$$

其中,$\alpha,\beta,\gamma,\delta$分别为各自对应的权重,各维度融合权重分别设定为$\alpha$=0.2,$\beta$=0.2,$\gamma$=0.2,$\delta$=0.4,此配置在验证集上通过网格搜索得到. 在实际推理过程中,模型从训练语料中检索出与当前输入x相似度最高的类比样本,并将其与当前输入进行拼接形成增强Prompt,以显式提供结构与语义指导,提升模型对复杂结构与低频实体的泛化能力. 下面对各个相似度指标的定义进行说明.

①语义嵌入相似度

本文使用预训练的语言模型(RoBERTa)对文本进行编码,以生成高维嵌入.

$$\mathrm{S\_sem}(x,x')=\frac{\sum_{k=1}^{d}\boldsymbol{e}_{x_k}\cdot\boldsymbol{e}_{x'_k}}{\sqrt{\sum_{k=1}^{d}\boldsymbol{e}_{x_k}^2}\cdot\sqrt{\sum_{k=1}^{d}\boldsymbol{e}_{x'_k}^2}} \tag{5}$$

其中,$\boldsymbol{e}_x$和$\boldsymbol{e}_{x'}$分别表示输入文本x与候选样本x'的语义嵌入向量,维度为d. 该指标反映了两段文本在预训练语义空间中的余弦相似性,值越大表示语义越接近.

②TF-IDF特征相似度

为进一步量化输入文本与训练示例文本的特征相似度,本文采用TF-IDF加权向量表示法. 对训练集中的所有示例使用jieba库对文本进行中文分词处理,并基于词频构建词袋模型. 令词汇表为V,输入文本的TF-IDF表示为$\boldsymbol{v}(x)$,则相似度计算公式为

$$\boldsymbol{v}(x)=\left[\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_{|V|}\right] \tag{6}$$

$$\mathrm{S\_tfidf}\left(x,x'\right)=\frac{\boldsymbol{v}(x)\cdot\boldsymbol{v}(x')}{\|\boldsymbol{v}(x)\|\|\boldsymbol{v}(x')\|} \tag{7}$$

③依存句法结构相似度

依存句法结构相似度用于衡量输入文本与候选样本在整体依存句法结构上的一致性. 本文使用Spacy工具抽取输入文本与候选样本的全局依存句法树,分别记为$D(x)$和$D(x)$,并通过图编辑距离(GED)计算其差异度:

$$\mathrm{S\_dep}(x,x')=1-\frac{\mathrm{GED}\left(D(x),D(x')\right)}{\max\left(|D(x)|,|D(x')|\right)} \tag{8}$$

其中,$\mathrm{GED}\left(D(x),D(x')\right)$表示依存树间的编辑距离,$|D(x)|$和$|D(x')|$分别为依存树的节点数.

传统相似度计算多侧重文本层级的表面相似性,难以刻画句法与语义结构在实体识别任务中的功能一致性,其在法律推理中具有高度区分性. 本文提出的语义角色相似度指标为该模块的核心创新,结合实体类型匹配、依存路径相似性及句法位置一致性建模实体在法律场景中所承载的语义角色. 其定义如下:

$$\begin{aligned}\mathrm{S\_role}\left(x,x'\right)=&\mathrm{I\_type}\left(\boldsymbol{e},\boldsymbol{e}'\right)+\mathrm{Sim\_dep}\left(\boldsymbol{e},\boldsymbol{e}'\right)\\&+\mathrm{Sim\_pos}\left(\boldsymbol{e},\boldsymbol{e}'\right)\end{aligned} \tag{9}$$

最终,S_role被纳入情境映射总相似度$A(x,x')$的融合计算中,作为模型进行类比示例选择的关键依据. 下面对S_role中的各个相似度指标的定义进行说明.

①实体类型一致性

I_type为实体类型一致性指标,若两者命名实体类型相同则为1,否则为0. 其定义如下:

$$\mathrm{I\_type}(\boldsymbol{e},\boldsymbol{e}')=\begin{cases}1, & \mathrm{type}(\boldsymbol{e})=\mathrm{type}(\boldsymbol{e}')\\0, & \text{otherwise}\end{cases} \tag{10}$$

②句法路径相似度

句法路径相似度Sim_dep用于衡量特定实体在句中与其核心谓词之间局部依存路径的一致性. 与全局的S_dep不同,Sim_dep更关注局部实体在句法层面的功能角色,其核心是精准抓取实体与关键谓词的局部关联,而非关注整句结构. 具体而言,本研究采用LTP工具的依存句法分析模块,提取输入文本中实体e与其依赖谓词之间的路径Path($\boldsymbol{e}$),以及候选样例中对应实体的依存路径Path($\boldsymbol{e}'$). 其定义如下:

$$\mathrm{Sim\_dep}(\boldsymbol{e},\boldsymbol{e}')=1-\frac{\mathrm{GED}\left(\mathrm{Path}(\boldsymbol{e}),\mathrm{Path}(\boldsymbol{e}')\right)}{\max\left(|\mathrm{Path}(\boldsymbol{e})|,|\mathrm{Path}(\boldsymbol{e}')|\right)} \tag{11}$$

其中,GED表示两条依存路径之间的最小编辑距离.

③句法位置相似度

句法位置相似度Sim_pos用于衡量两个实体在句中是否处于相同的语法位置或语义角色位置,更注重实体在句中所承担的功能角色是否一致. 具体而言,本研究采用LTP工具的语义角色标注(SRL)模块,识别句中谓词及其论元关系,并结合句法分析结果判断实体在句中所处的位置类别(如主语、宾语). 其定义如下:

$$\mathrm{Sim\_pos}(\boldsymbol{e},\boldsymbol{e}')=\frac{\left|\mathrm{Roles}(\boldsymbol{e})\cap\mathrm{Roles}(\boldsymbol{e}')\right|}{\left|\mathrm{Roles}(\boldsymbol{e})\cup\mathrm{Roles}(\boldsymbol{e}')\right|} \tag{12}$$

其中,Roles(·)表示实体的语法/语义角色集合.

需要指出的是,依存句法分析在不同相似度指标中的功能定位存在差异,因此本文在实现上结合使用了Spacy与LTP两种工具. 前者用于抽取全局依存句法树并支持基于图编辑距离(GED)的整体结构对齐,更契合句法结构相似度的全局衡量需求;后者则在中文法律文本的局部依存关系与语义角色标注任务中具有更高的准确性,因而被用于句法路径相似度与语义角色相似度的计算. 二者的结合兼顾了全局结构与局部语义两个层面的建模要求,从而在保证方法鲁棒性的同时提升了类比学习模块的判别精度与适用性.

此外,为了进一步优化训练效率,本文在实践中预先计算了所有训练样本之间的相似性分数,并通过ID映射将其存储在JSON文件中. 在训练阶段,模型只需根据样本ID快速读取预存分数,而不必重复进行依存分析与相似度计算. 该设计有效避免了冗余开销,使得类比学习模块在保证准确性的同时仍具有良好的实用性.

4　任务定义

为优化基于理解驱动的条件生成过程,JURIS旨在从司法文本中准确识别命名实体,并生成符合结构规范的输出序列. 模型的输出序列由一系列标记$Y=(y_1,y_2,\cdots,y_n)$组成,其中每个标记y_i包含文本内容t_i及其对应的实体标签,格式为【LABEL:Entity】,对于非实

体标记，则保留原始文本而不需要添加标签前缀，即 $y_i=t_i$. 训练过程的目标是学习条件概率分布 $P(y_i|T,I,\mathcal{T}_{\text{Tri-UEM}})$，其表示在给定输入文本 T、指令 I 以及来自三元理解增强模块的上下文信息 $\mathcal{T}_{\text{Tri-UEM}}$ 的条件生成正确标签序列的概率.

训练过程中采用负对数似然(NLL)作为优化目标，以衡量预测输出序列与参考标注序列之间的匹配程度. 具体而言，JURIS通过生成正确实体的标签前缀，同时忽略非实体内容的标注，实现对条件分布的高效建模. 其训练目标为最小化以下损失：

$$\mathcal{L}_{\text{NLL}}=-\sum_{i=1}^{n}\log P(y_i|T,I,\mathcal{T}_{\text{Tri-UEM}},y_1,y_2,\cdots,y_{i-1}) \quad (13)$$

4.1 指令微调

本文引入指令微调机制作为JURIS框架中的核心组成部分，与上下文感知的嵌入式标注、Tri-UEM模块形成互补，通过明确、结构化的任务指令，共同支撑了面向司法理解增强框架的构建. 模型的输入由司法文本 T、指令 I 以及三元理解增强模块提供的增强信息 $T_{\text{Tri-UEM}}$ 共同组成，如图5所示. 这些组件协同构建模型的输入序列，用于引导模型生成结构化的实体标注结果.

其中，任务指令 I 用于明确生成目标并提供格式约束. 考虑到大语言模型以自回归方式逐标记生成输出，本文充分利用其边生成边理解的特性，引导模型逐个token进行语义判断与实体识别，从而有效地辅助模型在理解实体语义角色的基础上完成准确标注. 具体地，本文将指令设计如下：

“请分析提供的句子，逐个标记地识别每个实体的类型. 请使用【LABEL:Entity】格式标注实体，如果某个词不属于任何实体类别，请不要为其添加实体标签.”

在模型结构方面，本文采用Qwen 2.5系列作为JURIS框架的基础模型. 该模型是专为中文场景优化的大型预训练语言模型，具备优异的指令响应能力与上下文生成能力，特别适用于面向中文司法任务的上下文建模与结构生成. 为了以有限的计算开销实现高效的任务适应，本文应用了基于LoRA的参数高效微调，如图2所示，将轻量级LoRA模块插入Transformer架构的多头自注意力和前馈层，以更好地学习嵌入式标注序列的方式.

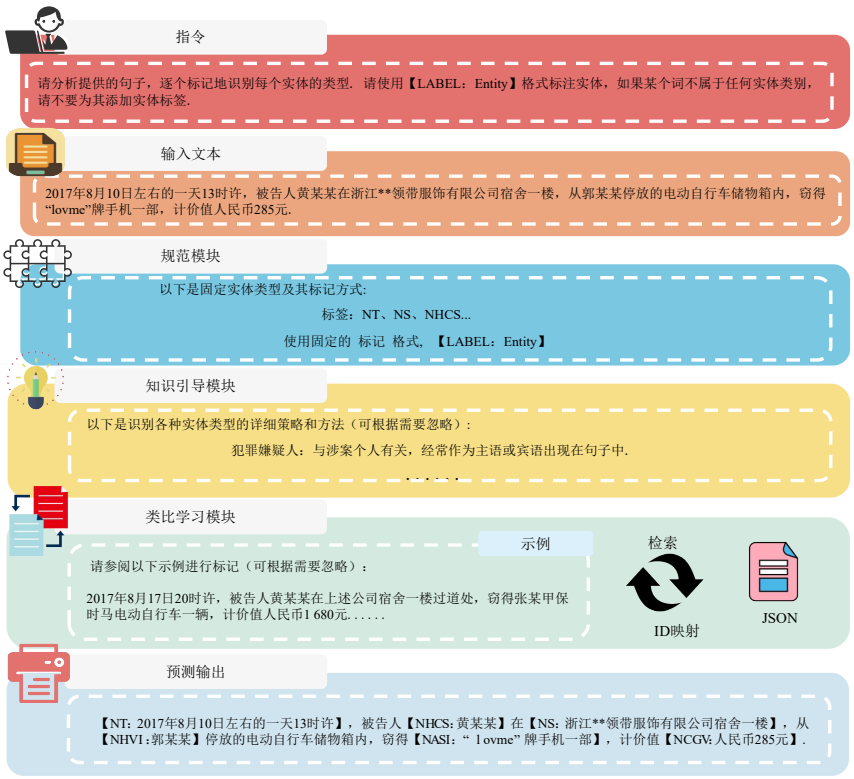


图5 模型的输入示例(主要由指令、输入文本以及三元理解增强模块拼接成prompt)

4.2 基于Tri-UEM的解码处理

JURI采用基于Transformer的解码器结构，负责生成输出标签序列. 与传统解码不同，本文并未重新实现新的解码器，而是在大模型的标准自回归解码流程中引入了Tri-UEM提供的上下文信息. 它在每个时间步使用以下递归更新隐藏状态 h_t：

$$h_t=f(h_{t-1},y_{t-1},T,I,\mathcal{T}_{\text{Tri-UEM}}) \tag{14}$$

其中,h_t表示解码器在时间步t的隐藏状态,输入包括历史隐状态h_{t-1}、前一时刻输出y_{t-1}、输入文本T、指令I,以及Tri-UEM生成的上下文增强表示$\mathcal{T}_{\text{Tri-UEM}}$.

随后,解码器使用softmax函数为当前令牌y_i生成可能实体标签的概率分布:

$$P(y_i|h_i)=\text{softmax}(Wh_i+b) \tag{15}$$

在每个时间步骤中,解码器使用T、I和$T_{\text{Tri-UEM}}$提供的上下文来预测当前文本最可能的标签. 此过程确保模型为实体生成所需的标注格式,并保留未标记的非实体文本.

5　实验及结果分析

本节在两个中文法律领域的命名实体识别(NER)数据集上对提出的JURIS框架进行了评估,验证了其在法律语境理解与实体提取方面的有效性. 本文将所提方法与一系列具有代表性的基线模型加以比较,包括传统模型(例如BiAffineNER[7]、BERT-MRC[8])、结构增强型模型(例如BOCNER[9]、W²NER[10]、RoBERTa-CRF[11])以及生成式方法(例如T5Based[13]、Diffusion-NER[14]). 本文还对比了两类代表性生成式NER方法:GLiNER[12]和GenerativeNER[20],以探究当前主流生成范式在法律领域的泛化能力. 此外,在医学领域的补充实验验证了JURIS在结构多样化的NER任务中的泛化能力.

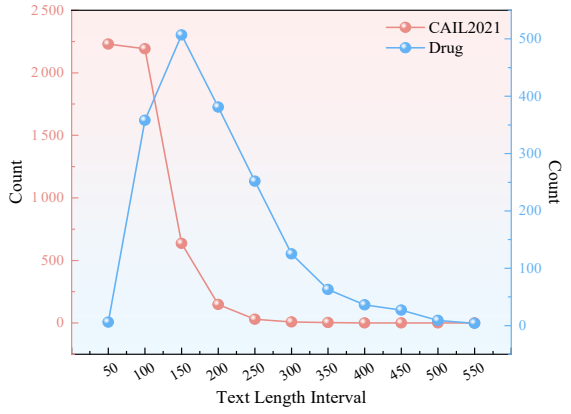


图6　CAIL2021数据集与Drug数据集文本长度分布

原始数据集多以扁平化的Span标注或面向实体对的三元组结构存在,如图7所示,缺乏任务指令和跨句上下文,难以直接用于指令微调训练. 为提升大模型对法律实体与关系的理解能力,本文对原始数据集进行了统一的结构重构与上下文感知重标注,将每条样本转换为结构明确的输入-输出对,如表1所示,以契合指令微调的生成范式.

5.1　数据集

本文在两个中文法律领域的命名实体识别数据集上评估所提出的JURIS框架,以验证其在法律语境理解和实体识别任务中的有效性. 这两个数据集均来自真实司法文书,并经过严格的脱敏与人工标注,具有代表性和挑战性,文本长度分布如图6所示,实体类型分布如表4所示.

CAIL2021[23]是由中国法律智能技术评测提供的专用于命名实体识别任务的标注数据集. 该数据集重点关注刑法第264条下的盗窃案件,包含5 247条数据、343 640个字符和25 466个实体,涵盖了10种不同细粒度的实体类型. 其中,最显著的特点是系统性地引入了嵌套命名实体结构,这是法律文书中广泛存在但对传统NER方法极具挑战性的现象. 为进一步验证模型在不同语义复杂度场景下的迁移能力,所使用的第二个数据集为Drug[24]数据集,专注于分析《中华人民共和国刑法》第三百四十七条至第三百五十七条所涉及的"毒品罪"相关案件,主要来自网络公开的若干涉毒类罪名法律文书,总计1 768条数据,总字符数为299 516个,实体数量达到19 321个以及5种不同实体类型.

表4　CAIL2021与Drug数据集实体类型分布

CAIL2021					Drug			
实体标签	实体类型	数量	嵌套实体数量	嵌套比例/%	实体标签	实体类型	数量	比例/%
NHCS	犯罪嫌疑人	6 463	67	1.04	NH	被告人	8 067	41.86
NHVI	受害人	3 108	1 017	32.72	NDR	毒品类型	4 221	21.90
NCSM	被盗货币	915	49	5.36	NW	毒品重量	2 054	10.66
NCGV	物品价值	2 090	49	2.35	NT	时间	2 693	13.97
NCSP	盗窃获利	481	15	3.12	NS	地点	2 237	11.61
NASI	被盗物品	5 781	432	7.47				
NATS	作案工具	735	20	2.72				
NT	时间	2 765	20	0.72				
NS	地点	3 517	545	15.50				
NO	组织	806	37	4.59				

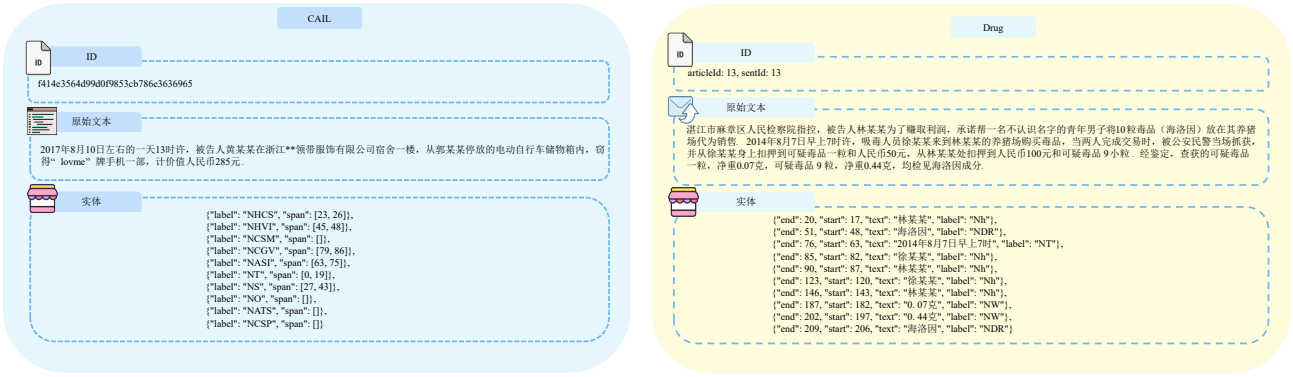


图7 CAIL2021数据集与Drug数据集原始结构

5.2 实验参数设置

本节概述了用于训练和微调司法命名实体识别的JURIS模型的超参数设置. 该模型利用预训练的基于Transformer的体系结构Qwen2.5-7B,并对法律文本处理进行了专门的增强,包括LoRA微调[25],这些设置经过优化,以平衡模型复杂性、计算效率和特定于任务的性能. 实验环境为Ubuntu 22.04操作系统,NVIDIA显卡为GeForce RTX A6000,CUDA版本为12.4,使用PyTorch框架来搭建JURIS模型,训练超参数包括学习率为,批大小为4,epoch为15,权衰减为0.01,LoRA低秩矩阵的秩为8.

5.3 对比实验分析

5.3.1 标注策略对比

在生成式命名实体识别任务中,传统的抽取式标注策略通常将实体直接输出为结构化列表(如:[“黄某某”“lovme手机”“浙江公司”]),而不关注实体周围的上下文,在基于生成式语言模型进行指令微调时,通常会存在表达结构与模型语义建模能力不匹配的问题. 为进一步验证本文所提出的基于上下文感知的嵌入式标注策略的有效性,采用相同的预训练模型(Qwen2.5-7B)进行LoRA微调且不添加Tri-UEM模块,分别比较了“基于上下文感知的嵌入式标注”与“实体列表抽取式”的训练效果,实验结果如表5所示,Pr表示精度,Rec表示召回率,后文相同. 可观察到使用传统抽取式结构时,模型往往忽略细粒度类别信息或输出不完整的实体短语,而嵌入式标注不仅能提供更清晰的语义边界表达,更能提升语言模型在复杂上下文下的实体抽取能力.

表5 不同标注策略的性能对比分析

标注策略	CAIL2021			Drug		
	Pr	Rec	F_1	Pr	Rec	F_1
上下文感知的嵌入式标注	0.877 9	0.879 5	0.878 7	0.862 5	0.861 3	0.861 9
实体列表抽取式标注	0.783 0	0.744 2	0.764 6	0.756 3	0.739 6	0.747 9

5.3.2 嵌套实体识别任务(CAIL2021)

CAIL2021是一个专门面向司法案件中嵌套实体识别的中文数据集,具有显著的结构复杂性. 表6中显示,所有的Qwen2.5+JURIS系列模型在所有参数规模下均表现优异. 其中,Qwen2.5-7B+JURIS实现了94.37%的F_1分数,已显著优于如W^{2}NER(92.76%)、GLiNER(91.79%)和GenerativeNER(93.30%)等当前SOTA方法,同时在精度(94.47%)与召回率(94.28%)两个子指标上亦表现最优,展现出出色的综合能力与稳定性. Qwen2.5-0.5B +JURIS作为最轻量模型,在召回率达到92.64%,F_1达到92.87%,也已经逼近GenerativeNER的整体性能. 结果表明,即便在资源有限的条件下,本文提出的上下文感知的标注策略和三元理解增强模块也能有效释放小模型的潜力. 同时,相较于BiAffineNER(F_1 = 89.44%)、BERT-MRC(F_1 = 92.28%)等基于平面标注或MRC结构的方法,本文的方法在处理嵌套实体、不同语义角色时具有明显优势. 传统方法仅能识别浅层语义,而本文的设计通过将实体识别任务转化为上下文驱动的语言理解问题,显著提升了模型的抽取准确性和泛化能力. 此外,从不同规模模型在CAIL2021上的表现来看,可观察到一个清晰的趋势:随着模型参数规模的提升(从0.5 B到7 B),F_1分数呈现出单调上升态势. 这一趋势在结构复杂任务中尤为显著,说明了在嵌套结构复杂、语义角色交织的任务中,越大的模型具备越强的上下文建模与指令响应能力,越能充分激发本文所提出的“理解增强”理念所提供的提示信息,从而取得越优的性能.

表6　CALI2021数据集下的命名实体识别模型的对比

模型	Pr	Rec	F_1
BiAffineNER[7]	0.912 4	0.877 2	0.894 4
BERT-MRC[8]	0.927 6	0.918 2	0.922 8
BOCNER[9]	0.902 1	0.904 4	0.903 2
W²NER[10]	0.934 7	0.920 6	0.927 6
RoBERTa-BiLSTM-CRF[11]	0.888 9	0.940 4	0.913 9
GLiNER[12]	0.921 7	0.914 2	0.917 9
T5Based[13]	0.925 5	0.916 3	0.920 8
Diffusion NER[14]	0.938 7	0.925 3	0.932 0
GenerativeNER[20]	0.941 4*	0.924 8	0.933 0
PUnifiedNER[26]	0.928 6	0.917 1	0.922 8
Qwen2.5-0.5B+JURIS	0.931 1	0.926 4	0.928 7
Qwen2.5-1.5B+JURIS	0.939 0	0.929 4	0.934 2
Qwen2.5-3B+JURIS	0.937 3	0.936 3*	0.936 8*
Qwen2.5-7B+JURIS	**0.944 7**	**0.942 8**	**0.943 7**

注:加粗数据代表最优,*代表次优.

5.3.3　平面实体识别任务(Drug)

Drug数据集涉及更为平面化的领域实体抽取任务,结构相对简单但语义表述复杂. 在该数据集上,Qwen2.5+JRUIS系列依旧展现出极强的稳健性与适应性,并在所有参数规模下均优于现有方法,如表7所示.

表7　Drug数据集下的命名实体识别模型的对比

模型	Pr	Rec	F_1
BiAffineNER[7]	0.891 6	0.858 7	0.874 8
BERT-MRC[8]	0.893 5	0.881 9	0.887 6
BOCNER[9]	0.882 1	0.873 6	0.877 8
W²NER[10]	0.908 3	0.777 9	0.892 9
RoBERTa-BiLSTM-CRF[11]	0.886 9	0.872 3	0.879 5
GLiNER[12]	0.902 8	0.884 2	0.893 4
T5Based[13]	0.889 5	0.881 6	0.885 5
Diffusion NER[14]	0.910 7	0.883 1	0.896 7
GenerativeNER[20]	0.913 7	0.885 9	0.899 6
PUnifiedNER[26]	0.898 6	0.875 3	0.886 8
Qwen2.5-0.5B+JURIS	0.916 4	0.911 8	0.914 1
Qwen2.5-1.5B+JURIS	0.903 7	0.904 7	0.904 2
Qwen2.5-3B+JURIS	**0.921 6**	**0.929 2**	**0.925 4**
Qwen2.5-7B+JURIS	0.920 5*	0.927 7*	0.924 1*

注:加粗数据代表最优,*代表次优.

具体而言,Qwen2.5-3B+JURIS实现最高F_1为92.54%,优于目前最强生成式模型GenerativeNER(89.96%),也超越了结构增强模型DiffusionNER(89.67%)和GLiNER(89.34%). 尤其在Recall指标上差距更为明显,表明JURIS在处理隐含或变体表达的实体方面有更强的语义捕捉能力.

值得注意的是,与CAIL2021数据集不同,在Drug数据集中并未观察到性能随模型规模单调提升的趋势. 在该任务上3B模型略优于7B(92.54% VS 92.41%). 这一现象表明,在任务结构相对简单、数据规模较小的任务中,模型参数规模对性能影响减弱,导致大模型难以发挥全部能力,同时可能出现轻微的过拟合或注意力稀释,而3B模型则在能力与泛化之间取得更优权衡,说明在结构不同的任务场景下,理解增强机制的最优发挥形式存在差异,模型容量的适配性亦需结合任务属性综合考量.

5.3.4　zero-shot与few-shot

为了消除模型本身参数规模对实验结果的干扰,同时进一步评估JURIS框架是否真实提升了模型理解能力,本文选取了Qwen2.5[27]系列的4种不同尺寸模型(0.5B、1.5B、3B、7B)作为基座,分别在zero-shot与few-shot两种推理设置下进行对比实验,评估其在CAIL2021与Drug两个法律领域数据集上的表现.

实验中,所有模型均采用统一的输入输出结构. 在few-shot设置下,本文为每类实体选择3条训练样本作为上下文提示拼接于任务前,以增强模型对不同实体标签结构的记忆能力,形成更贴近人类提示工程的提示范式,实验结果如表8所示. 在法律领域的命名实体识别上,尤其是在CAIL2021数据集上处理复杂或嵌套的实体时,在未进行领域适配的情况下,性能仍然有限,F_1分数在不同规模的模型间差异较为显著. 在zero-shot设置下,Qwen2.5系列模型的F_1值整体随参数规模增长而稳定上升,说明基础语言模型具备一定的指令理解与实体抽取能力. 相比zero-shot,加入3-shot上下文示例后,各模型F_1值均明显提升,表明上下文示例有助于激发大模型对任务理解与实体角色的结构化建模能力. 进一步地,本文在图8中绘制了不同模型规模下3种设置(zero-shot、few-shot、JURIS)的F_1表现折线图. 可观察到,在所有规模设定中,JURIS始终保持领先,且其优势并不依赖于大规模模型,验证了该方法在理解范式上的独立有效性.

5.3.5　实验小结

与当前代表性的生成式司法命名实体识别方法(如GenerativeNER、GLiNER、T5Based、DiffusionNER等)相比,JURIS在两个数据集上均表现出显著优势,如图9所示. 在CAIL2021中,Qwen2.5-7B+JURIS实现了94.37%的F_1分数,超过当前SOTA方法GenerativeNER(93.30%). 尤其在召回率上提升了1.80个百分点,这表明JURIS能更完整地覆盖复杂嵌套实体. 在Drug数据集中,性能表现上展现出更显著的领先优势,Qwen2.5-3B+JURIS模型取得了92.54%的F_1,与GenerativeNER的89.96%相比提升了2.58个百分点,召回率提升近4.33个百分点. 这种差异的核心原因在于这些模型虽然采用了生成范式,但其解码过程仍以生成目标结构

表 8 Qwen2.5 系列模型在 zero-shot 与 few-shot 设置下的性能对比分析

模式	模型	CAIL2021			Drug		
		Pr	Rec	F_1	Pr	Rec	F_1
zero-shot	Qwen2.5-0.5B	0.452 1	0.436 7	0.444 3	0.483 4	0.467 3	0.475 2
	Qwen2.5-1.5B	0.510 1	0.488 0	0.498 8	0.547 1	0.509 9	0.527 8
	Qwen2.5-3B	0.557 2	0.595 3	0.575 6	0.604 2	0.584 6	0.594 8
	Qwen2.5-7B	0.613 8	0.620 9	0.617 3	0.642 7	0.611 2	0.622 6
few-shot	Qwen2.5-0.5B	0.491 0	0.483 6	0.487 3	0.553 8	0.502 1	0.526 7
	Qwen2.5-1.5B	0.545 7	0.527 4	0.536 4	0.590 8	0.544 9	0.566 9
	Qwen2.5-3B	0.615 7	0.590 2	0.602 7	0.633 5	0.610 8	0.621 9
	Qwen2.5-7B	0.691 4	0.632 5	0.660 6	0.682 7	0.696 5	0.689 5

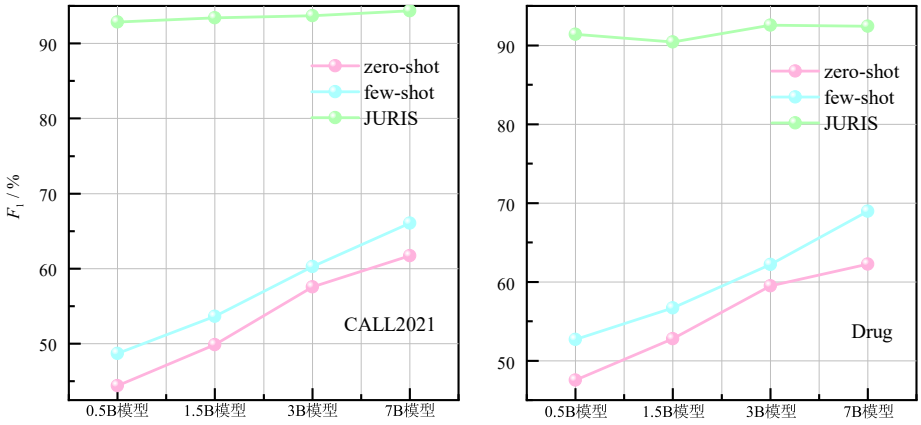


图 8 不同模型规模下 3 种设置(zero-shot、few-shot、JURIS)

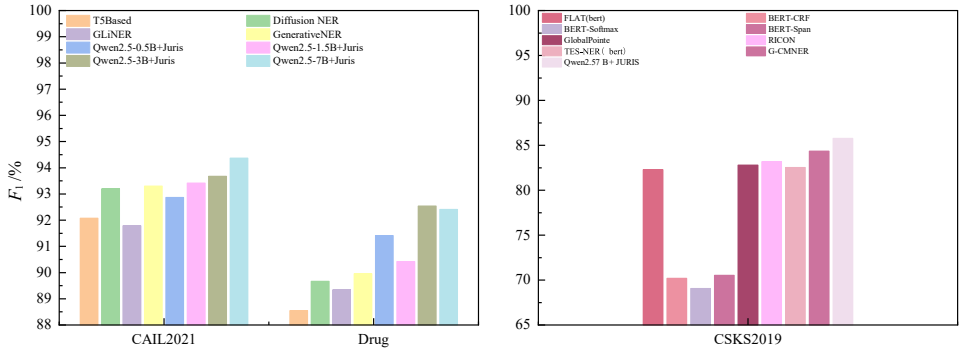


图 9 JURIS 与主流命名实体识别模型在 3 个数据集上的 F_1 值对比柱状图

文本为导向,模型须学习如何组织结构,而非理解文本语义,而 JURIS 通过上下文感知的嵌入式标注策略,不再从文本外部组织结构,而是在文本内部完成语义生成与识别,使得实体预测过程更加自然. 此外,本文提出的三元理解增强模块与嵌入式标注方式共同协作,使得模型不仅在性能指标上领先,而且在泛化能力与推理可解释性上更具优势.

5.4 详细评估结果

为了进一步展示 JURIS 的性能,本文提供了两个数据集在 Qwen2.5-7B+JURIS 模型上每种实体类型的精度、召回率和 F_1 分数的细分情况,如表 9 和表 10 所示. 本文给出了 Drug 和 CAIL2021 数据集上每种实体类型的精度、召回率和 F_1 分数. 此外,为了更直观地展示模型的分类效果,分别为 CAIL2021 和 Drug 数据集绘制了混淆矩阵,以显示模型对每种实体类型的预测性能,如图 10 所示,后续工作可以针对少量低评分实体类型优化特定策略和数据扩展技术.

5.5 消融实验

为了评估每个 Tri-UEM 子模块的贡献与组合效果,本文通过在 Qwen2.5-7B+JURIS 框架中每次移除一个组件来对 CAIL2021 和 Drug 进行消融研究,并保持所有其

表9　CAIL2021数据集的各个实体类型所对应的精度、召回率、F_1

实体类型	Pr	Rec	F_1
NT	0.961 2	0.973 5	0.967 3
NHCS	0.987 4	0.986 3	0.986 8
NATS	0.974 7	0.866 7	0.914 5
NS	0.924 9	0.908 3	0.916 5
NASI	0.974 6	0.916 8	0.944 8
NCSP	0.891 3	0.865 4	0.878 1
NHVI	0.955 4	0.939 4	0.947 3
NCGV	0.982 0	0.975 1	0.978 5
NCSM	0.881 2	0.904 6	0.892 7

表10　Drug数据集的各个实体类型所对应的精度、召回率、F_1

实体类型	Pr	Rec	F_1
NT	0.918 9	0.922 5	0.920 6
NH	0.979 4	0.988 9	0.984 1
NS	0.875 8	0.849 0	0.858 1
NDR	0.980 7	0.965 3	0.972 9
NW	0.945 0	0.979 0	0.960 2

他设置固定以进行公平比较,在Qwen2.5-7B模型基础上,仅采用基于上下文感知的嵌入式标注策略进行LoRA微调且不引入任何增强模块作为基线设置,表11与表12中使用精度、召回率和F_1展示了结果.

实验结果表明,完整的Tri-UEM设置可以在两个数据集上实现最佳性能,Tri-UEM的加入带来了6个百分点左右的F_1提升,删除任何模块都会降低结果.移除类比学习模块后,精度和召回率均有下降,CAIL2021的F_1降至91.61%,Drug的F_1降至90.66%,突出了其通过基于类比的推理识别复杂或低频实体的作用,尤其在低频类别与法律术语歧义处理中表现出良好增益;移除知识引导模块显著降低了召回率,CAIL2021的召回率降至91.19%,Drug的召回率降至90.12%,反映了特定领域先验知识的重要性,该模块在提升模型召回能力方面具有关键作用.移除规范模块,精度降幅最为显著,CAIL2021的精度降至90.32%,Drug的精度降至89.97%,这表明统一的输出风格与一致的标记格式对于边界和类型识别至关重要.实验结果验证了该策略在提升模型语义表达能力方面的重要性,并为Tri-UEM模块提供了稳定而高质量的语义信息,使得Tri-UEM能够在信息结构清晰的基础上有效建模理解增强任务,二者协同构成了JURIS框架的核心.

表11　不同模块组合在CAIL2021数据集上的消融结果

Module	Pr	Rec	F_1
上下文感知 + 无模块	0.877 9	0.879 5	0.878 7
规范模块 + 知识引导模块	0.912 1	0.920 2	0.916 1
规范模块 + 类比学习模块	0.915 7	0.911 9	0.913 8
知识引导模块 + 类比学习模块	0.903 2	0.928 3	0.915 6
Tri-UEM(三元理解增强模块)	0.944 7	0.942 8	0.943 7

表12　不同模块组合在Drug数据集上的消融结果

Module	Pr	Rec	F_1
上下文感知 + 无模块	0.862 5	0.861 3	0.861 9
规范模块 + 知识引导模块	0.904 9	0.908 4	0.906 6
规范模块 + 类比学习模块	0.909 0	0.901 2	0.905 1
知识引导模块 + 类比学习模块	0.899 7	0.918 8	0.909 1
Tri-UEM(三元理解增强模块)	0.920 5	0.927 7	0.924 1

为验证类比学习模块中语义角色相似度S_role的独立贡献,本文同样使用Qwen2.5-7B+JURIS框架在消融实验中新增了移除的对比组,即在相似度函数$A(x, x')$中仅使用S_sem、S_tfidf和S_dep这3个指标.实验结果如表13和表14所示.

实验结果表明,移除S_role后F_1分数在两个数据集

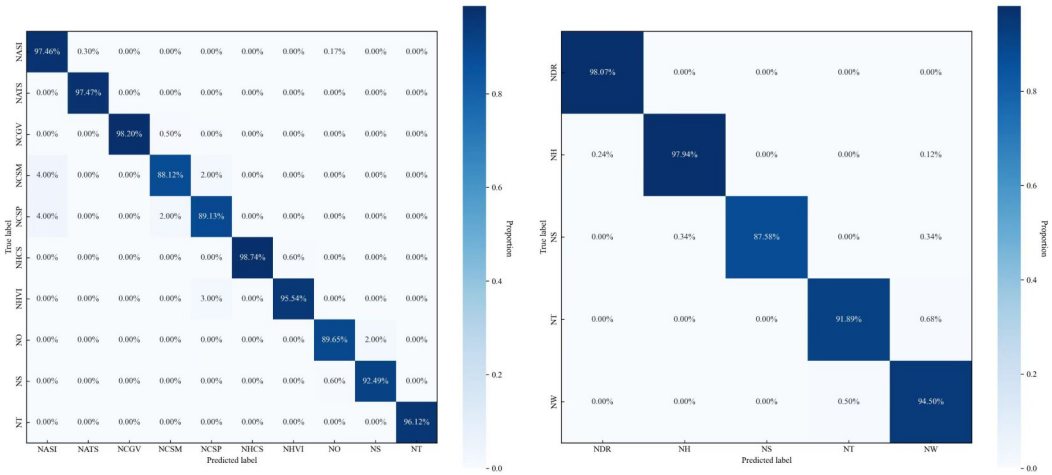


图10　CAIL数据集与Drug数据集上各个实体类型的混淆矩阵

表 13 类比学习模块组合在 CAIL2021 数据集上的消融结果

模型配置	Pr	Rec	F_1
类比学习模块(含 S_role)	0.944 7	0.942 8	0.943 7
类比学习模块(移除 S_role)	0.928 5	0.914 2	0.921 3

表 14 类比学习模块组合在 Drug 数据集上的消融结果

模型配置	Pr	Rec	F_1
类比学习模块(含 S_role)	0.920 5	0.927 7	0.924 1
类比学习模块(移除 S_role)	0.912 3	0.905 6	0.908 9

上均显著下降. Drug 数据集上 F_1 值下降了 1.52%. CAIL2021 数据集 F_1 值降至 92.13%,较完整 Tri-UEM 下降 2.24%,尤其在嵌套实体识别中召回率下降 3.1%,充分证明了语义角色相似度在类比学习模块中发挥的关键作用,能够有效提升实体在复杂语境下的功能对齐能力,有效验证了其必要性.

5.6 案例研究

本文将 JURIS 的性能与传统方法进行了比较,如图 11~图 13 所示,展示了 JURIS 在处理各种类型的实体时的输出情况,包括嵌套实体、金额和地理位置等.

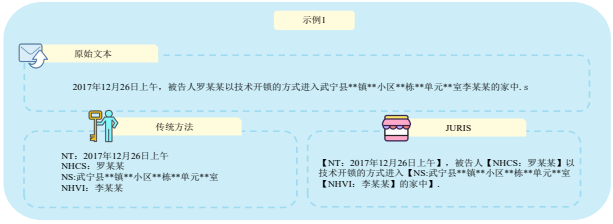


图 11 嵌套实体示例输出

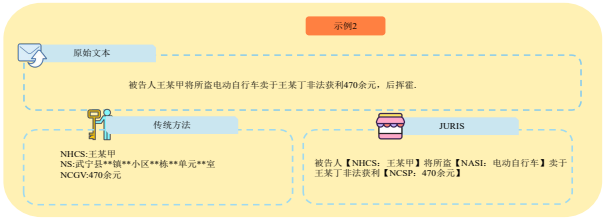


图 12 金额实体示例输出

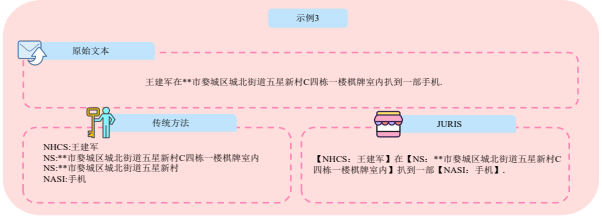


图 13 长表达式实体示例输出

图 11 展示了 JURIS 在司法文本中准确处理嵌套实体的能力. 在法律文书中,实体嵌套结构广泛存在,例如地理位置与人物、事件之间常构成语义上的从属关系. 然而,传统 NER 模型往往倾向于将嵌套实体各自孤立建模,忽略其语义层级,从而破坏整体结构的完整性. 例如,在“NS:武宁县镇小区栋单元室”与“NHV:李某某”的复合表达中,传统方法通常将前者视为单一地点实体,将后者识别为独立人物实体,未能捕捉“人物位于地点之中”的从属关系. 相比之下,JURIS 成功识别出嵌套结构“【NS:武宁县镇小区栋单元室【NHVI:李某某】的家中】”,准确建模了人物实体在地点实体中的嵌套位置,显著提升了语义结构的表达能力. 该结果表明,JURIS 在法律文本中复杂实体层级关系的建模方面具备更强的表达能力与结构敏感性.

图 12 进一步展示了 JURIS 在金额实体识别任务中的语义理解能力. 在司法语境中,同样的金额表达可能代表不同的法律含义,例如“470 余元”既可能表示物品价值,也可能指向非法获利. 传统方法多依赖正则匹配或关键字触发,常将“××元”统一标注为财物类实体,忽略其在上下文中的法律角色定位. JURIS 通过上下文语义建模,结合“犯罪分子非法所得”等语境提示,准确识别“470 余元”在此处应标注为盗窃获利类实体. 该能力的获得归因于系统中集成的知识引导模块,该模块可根据语义上下文自动调整实体解释策略,提升模型在法律语义角色层面的区分与表达能力,从而实现对金额实体更精细的语义判定.

图 13 评估了 JURIS 对地理位置实体的识别与归并能力. 在法律文书中,地理信息往往涉及多层次的细化描述,包括城市、街道、社区及具体建筑物名称. 传统方法由于缺乏结构建模能力,常将多段连续的地理信息切分为多个独立实体,造成实体碎片化与表达不完整. 例如,“NS:**市婺城区北街道五星新村 C 四栋一楼棋牌室”与“NS:**市吴县区北街五星新村”在传统模型中可能被割裂为多个片段式地名,影响案件空间信息的准确解析与后续法律推理. JURIS 通过上下文感知机制与结构规范模块联合建模,能够将上述描述整合为连续、完整的地点实体:“【NS:**市婺城区北街道五星新村 C 四栋一楼棋牌室】”,有效保持地理信息的结构完整性与语义一致性. 这一能力对于准确还原法律案件发生地点及相关证据链条具有重要意义.

5.7 跨领域评估:医学领域

为在法律领域之外进一步评估所提 JURIS 框架的跨领域适应性,本文在公开的 CCKS2019 数据集上开展了初步实验. 该数据集包含 1 379 份中文临床病历,标注了多种医学实体类型,包括“手术”“解剖部位”“药物”“疾病和诊断”“影像检查”和“实验室检验”.

本文沿用了法律领域中的方法论,采用上下文感知嵌入式标注策略,保留语义完整性. 模型训练采用了与法律领域相同的Tri-UEM架构和超参数设置,同时针对医学领域特点,调整了实体类型规范、知识引导方法和类比推理示例. 本文将JURIS与近3年来在通用领域上表现较好的模型和基于BERT的模型等标准基线模型进行了对比,如图9与表15所示. 实验结果表明,即便在结构迥异的领域,JURIS也能保持强大的性能,在F_1分数上领先多个强基线模型1.3 ~ 3.5个百分点,表明其嵌入式标注与Tri-UEM结构对跨领域任务具有较好的迁移能力. 结果表明,JURIS具有良好的跨领域适应性,能够有效迁移至其他对结构化语义理解有较高要求的专业场景,为推动自然语言处理驱动的垂直领域研究提供了一个可复现的范式.

表15　CCKS2019数据集下的不同命名实体识别模型的对比

Module	Pr	Rec	F_1
FLAT(bert)[28]	0.814 4	0.831 9	0.823 0
BERT-CRF[28]	0.726 4	0.679 0	0.701 9
BERT-Softmax[28]	0.709 6	0.672 6	0.690 6
BERT-Span[29]	0.750 5	0.665 4	0.705 4
GlobalPointe[28]	0.833 5	0.822 6	0.828 1
RICON[28]	0.834 7	0.828 9	0.831 8
TES-NER(bert)[28]	0.819 7	0.831 0	0.825 4
G-CMNER[28]	0.856 1	0.832 8	0.843 6
Ours(Qwen2.5-7B + JURIS)	0.861 0	0.854 7	0.857 8

6　结论

针对司法命名实体识别任务中存在的实体边界模糊、上下文依赖强、类别细粒等挑战,本文提出了一种基于理解驱动为核心的命名实体识别新框架——JURIS. 该框架从标注策略与推理流程两个层面协同设计,融合上下文感知的嵌入式标注策略与三元理解增强模块,充分挖掘了大语言模型在指令推理、上下文建模与类比生成方面的潜力. 在实验评估方面,本文在多个真实法律数据集上验证了JURIS在zero-shot、few-shot与全监督设置下的稳定性能. 此外,本文还在医疗领域平面实体识别任务上开展迁移实验,进一步验证了JURIS在不同专业语境下的可迁移性与一致性. 与多种强基线相比,本文方法在精确度、召回率和F_1分数上均表现出显著优势,为应对复杂法律文本中的信息抽取任务提供了一种具备理解驱动能力的新型解决方案. 未来研究将聚焦于不同基础模型容量对性能的影响,深入理解大模型语义建模能力的边界以及探索JURIS在更多垂直专业语境中的迁移适应验证.

参考文献

[1] 张军. 最高人民法院工作报告: 2024年3月8日在第十四届全国人民代表大会第二次会议上[R]. 北京: 最高人民法院, 2024.
ZHANG J. Report on the Work of the Supreme People's Court: Delivered at the Second Session of the 14th National People's Congress on March 8, 2024[R]. Beijing: Supreme People's Court, 2024. (in Chinese)

[2] GHOSH S, MAITRA P, DAS D. Feature based approach to named entity recognition and linking for tweets[C]//6th Workshop on Making Sense of Microposts. Montréal: Microposts, 2016: 74-76.

[3] ARSLAN S. Application of BiLSTM-CRF model with different embeddings for product name extraction in unstructured Turkish text[J]. Neural Computing and Applications, 2024, 36(15): 8371-8382.

[4] DHARVIYANTI N A D, WILANTIKA N. Rule-based NER for crime information extraction through online news site[C]//2024 International Conference on Information Technology Research and Innovation. Piscataway: IEEE, 2024: 99-104.

[5] MUMTAZ R, QADIR M A. CustNER: A rule-based named-entity recognizer with improved recall[J]. International Journal on Semantic Web and Information Systems, 2020, 16(3): 110-127.

[6] RUDNER T, POROD W, CSABA G. Design of oscillatory neural networks by machine learning[J]. Frontiers in Neuroscience, 2024, 18: 1307525.

[7] YU J T, BOHNET B, POESIO M. Named entity recognition as dependency parsing[EB/OL]. (2020-06-13) [2025-07-27]. https://arXiv.org/abs/2005.07150.

[8] LI X Y, FENG J R, MENG Y X, et al. A unified MRC framework for named entity recognition[EB/OL]. (2022-11-22)[2025-07-27]. https://arXiv.org/abs/1910.11476.

[9] LU H W, PENG Y N. Named entity recognition of Chinese legal text based on BERT[C]//2022 4th International Conference on Applied Machine Learning. Piscataway: IEEE, 2023: 1-4.

[10] LI J Y, FEI H, LIU J, et al. Unified named entity recognition as word-word relation classification[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(10): 10965-10973.

[11] SHI J W, ZHENG K, ZHANG Z H, et al. A named entity recognition method based on deep learning for Chinese legal documents[C]//2022 7th International Conference on Im-

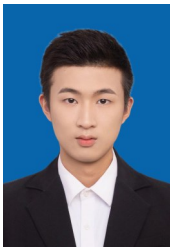
age, Vision and Computing. Piscataway: IEEE, 2022: 65-68.
[12] ZARATIANA U, TOMEH N, HOLAT P, et al. GLiNER: Generalist model for named entity recognition using bidirectional transformer[EB/OL]. (2023-11-14)[2025-07-27]. https://arXiv.org/abs/2311.08526.
[13] LEE J Y, PHAM L H, UZUNER O. MNLP at fincausal2022: Nested NER with a generative model[C]//Proceedings of the 4th financial narrative processing workshop@ LREC2022. Paris: ELRA, 2022: 135-138.
[14] SHEN Y L, SONG K T, TAN X, et al. DiffusionNER: Boundary diffusion for named entity recognition[EB/OL]. (2023-05-22)[2025-07-27]. https://arXiv.org/abs/2305.13298.
[15] WANG X, ZHOU W K, ZU C, et al. InstructUIE: Multi-task instruction tuning for unified information extraction[EB/OL]. (2023-04-17)[2025-07-27]. https://arXiv.org/abs/2304.08085.
[16] CHEN Y, ZHANG Y, WANG J, et al. YNU-HPCC at SemEval-2023 task 6: LEGAL-BERT based hierarchical BiLSTM with CRF for rhetorical roles prediction[C]//Proceedings of the The 17th International Workshop on Semantic Evaluation. Stroudsburg: ACL, 2023: 2075-2081.
[17] GU L, ZHANG W J, WANG Y, et al. Named entity recognition in judicial field based on BERT-BiLSTM-CRF model[C]//2020 International Workshop on Electronic Communication and Artificial Intelligence. Piscataway: IEEE, 2020: 170-174.
[18] HU Z Y, YUAN Z A. URF4CCT: A text understanding framework for Chinese telecom fraud cases[C]//2023 IEEE 9th International Conference on Cloud Computing and Intelligent Systems. Piscataway: IEEE, 2023: 121-125.
[19] CLARK K, LUONG M T, LE Q V, et al. ELECTRA: Pre-training text encoders as discriminators rather than generators[EB/OL]. (2020-03-23)[2025-07-27]. https://arXiv.org/abs/2003.10555.
[20] MAO X L, JIANG J, ZENG Y Z, et al. Generative named entity recognition framework for Chinese legal domain[J]. PeerJ Computer Science, 2024, 10: e2428.
[21] LI F, LIN Z C, ZHANG M S, et al. A span-based model for joint overlapped and discontinuous named entity recognition[EB/OL]. (2021-06-28)[2025-07-27]. https://arXiv.org/abs/2106.14373.
[22] LOUKACHEVITCH N, MANANDHAR S, BARAL E, et al. NEREL-BIO: A dataset of biomedical abstracts annotated with nested named entities[J]. Bioinformatics, 2023, 39(4): btad161.
[23] CAO Y, SUN Y Y, XU C, et al. CAILIE 1.0: A dataset for challenge of AI in law - information extraction V1.0[J]. AI Open, 2022, 3: 208-212.
[24] CHEN Y G, SUN Y Y, YANG Z H, et al. Joint entity and relation extraction for legal documents with legal feature enhancement[C]//Proceedings of the 28th International Conference on Computational Linguistics. International Committee on Computational Linguistics, 2020: 1561-1571.
[25] DEVALAL S, KARTHIKEYAN A. LoRa technology-an overview[C]//2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA). Piscataway: IEEE, 2018: 284-290.
[26] LU J H, ZHAO R, NAMEE B MAC, et al. PUnifiedNER: A prompting-based unified NER system for diverse datasets[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(11): 13327-13335.
[27] HUI B Y, YANG J, CUI Z Y, et al. Qwen2.5-coder technical report[EB/OL]. (2024-11-12) [2025-07-27]. https://arXiv.org/abs/2409.12186.
[28] 孟伟伦, 郭景峰, 邢珂萱, 等. 基于字形特征的中文医学命名实体识别方法[J]. 电子学报, 2024, 52(6): 1945-1954.
MENG W L, GUO J F, XING K X, et al. A Chinese medical named entity recognition method based on glyph features[J]. Acta Electronica Sinica, 2024, 52(6): 1945-1954. (in Chinese)
[29] YANG P, CONG X, SUN Z Y, et al. Enhanced language representation with label knowledge for span extraction[EB/OL]. (2021-11-01) [2025-07-27]. https://arXiv.org/abs/2111.00884.

作者简介

彭 晗 男,1992年8月出生于湖南省湘西土家族苗族自治州. 现为湖南工商大学智能机器人学院讲师. 主要研究方向为大语言模型优化、多模态融合、具身智能等.
E-mail: Han.Peng@hutb.edu.cn

阮日青 男,1999年9月出生于福建省莆田市. 现为湖南工商大学计算机学院硕士研究生. 主要研究方向为大语言模型优化.
E-mail: 230720835001@stu.hutb.edu.cn

胡　颖　女,2004年3月出生于湖北省咸宁市. 现为湖南工商大学人工智能与先进计算学院本科生. 主要研究方向为大语言模型优化.
E-mail: 2223030063@stu.hutb.edu.cn



刘琼林　男,2002年10月出生于湖南省岳阳市. 现为湖南工商大学人工智能与先进计算学院硕士研究生. 主要研究方向为大语言模型优化.
E-mail: 240120854118@stu.hutb.edu.cn

张　震　男,1994年8月出生于内蒙古自治区呼伦贝尔市. 现为湖南工商大学人工智能与先进计算学院副教授. 主要研究方向为算法分析与设计、人工智能理论基础.
E-mail: zz@hutb.edu.cn