基于全局指针的司法命名实体识别模型

李林瑛[1] 曲云平[1] 付子轩[1] 刘美汐[1]
LI Linying QU Yunping FU Zixuan LIU Meixi

摘 要 涉毒案件法律文书包含大量非规范地点描述、细粒度涉案要素及高度语境依赖的司法实体，其命名实体识别面临实体边界模糊、语义角色歧义等挑战。文章提出了一种基于全局指针网络的司法命名实体识别方法，通过显式建模文本中任意位置间的全局依赖，以增强模型对长距离语义线索的捕获能力。该方法首先以 BERT 为上下文编码器，随后结合双向长短记忆网络优化实体边界识别，最终通过全局指针机制生成候选实体并完成类别判定。基于涉毒案件法律文书的实验结果表明，该方法能够更加准确地抽取关键司法要素，为智能化辅助办案系统提供技术支撑。

关键词 司法命名实体识别；全局指针网络；法律文书；预训练语言模型；双向长短记忆网络

doi：10.3969/j.issn.1672-9528.2026.05.007

0 引言

当前，人工智能在司法场景中的深入应用已成为提升司法效率的关键支撑。其中，从海量非结构化法律文书中精准抽取具有司法领域语义属性的关键要素，是构建司法智能系统的基础性任务。在此背景下，法律文书的命名实体识别作为连接原始文本与结构化知识的桥梁，被视为支撑智慧司法诸多上层应用的前提条件，近年来受到广泛关注。然而，不同于通用命名实体识别主要关注人名、地名、组织名等自然语言中的基本语义类别，司法领域的命名实体识别旨在识别承载特定法律角色与功能的细粒度语义单元（如“被告人”“涉案毒品数量”“作案时间”等）。这类实体的语义角色无法仅凭局部词汇或表面形式确定，而必须依赖对案件上下文中法律语义的深层建模。在司法领域，命名实体识别是指从法律文书中抽取特定司法属性的名词和短语，包括案件时间、被盗金额、盗窃获利等。

尽管现有研究在司法命名实体识别方面取得了一定进展，其性能仍受制于深层次语义建模能力的缺乏。一方面，高质量标注语料的稀缺制约了模型的泛化；另一方面，主流方法或沿用通用实体类别体系，难以准确刻画司法实体所承载的细粒度司法语义角色，或受限于局部上下文建模机制，既难以缓解非规范实体表达所引发的实体边界定位偏差，也难以捕捉词语之间隐含的法律语义关联。上述局限反映出模型未能深入理解案件整体语境中实体所承载的司法语义角色，因而难以有效应对语义歧义、嵌套结构及非规范实体表达等挑战。针对以上问题，本文提出一种基于全局指针网络的命名实体识别方法，通过显式建模文本中任意两位置间的全局依赖关系，增强模型对长距离语义线索的感知能力，并在涉毒案件法律文书上开展系统性实验，以精准抽取关键司法要素，为智能化辅助办案提供技术支撑。

1 相关工作

命名实体识别方法总体可分为基于规则、基于统计和基于深度学习三类。基于规则的方法依赖人工构建的规则与词典，具有可解释性强的特征，但在新领域和新实体类型下的迁移能力有限，维护成本高，难以满足实际应用中实体类型动态扩展的需求；基于统计的方法通过隐马尔可夫模型（HMM）、条件随机场（CRF）等模型学习标注规律，但在数据稀缺或数据分布不均衡时识别性能显著下降，难以获得鲁棒结果。相比之下，深度学习方法通过端到端学习可获得更高的召回率与更优的特征表达能力，成为近年主流趋势[1]。

随着大语言模型（LLM）与多智能体（Multi-Agent）架构的发展，相关研究逐渐开始探索在 NER 上的迁移与泛化潜力。例如，CMAS[2] 将 NER 任务分解为命名实体识别与类型相关特征提取两个子任务，并通过带有自反思机制的示例判别器增强识别性能；FewFMNER[3] 通过概念过滤智能体确定实体边界，并结合双路径推理与自一致性策略提升类型分类的准确性；DiffusionNER[4] 则针对扩散模型难以建模离散边界的问题，通过机器阅读理解任务引导去噪过程，实现更

1. 大连外国语大学软件学院 辽宁大连 116044
[基金项目] 辽宁省教育厅高等学校基本科研项目（LJKMZ20221549）

精细的实体边界定位。然而，上述方法多面向通用语言环境，对领域知识的依赖较弱，难以充分捕捉司法文本中的高密度法条引用、语义隐喻与概念嵌套关系，因而在法律场景下表现仍存在不足。

在法律领域，研究者提出了多种面向特定任务的法律NER方法，为司法大数据分析与智能审判提供支持。毛星亮等[5]通过融合全局分类信息与局部分词信息，实现关键案情要素识别，提高了法律文本中要素抽取的完整性；Lu等[6]采用并行实例查询网络实现司法NER，并通过线性标签分配机制实现法律实体与查询向量的精准匹配；杨长春等[7]针对低频、长实体与句法信息捕捉不足的问题，提出GLYCE-ONLSTM-CRF模型以增强语义表示能力；Mao等[8]将NER视为序列生成问题，基于BART框架引入拷贝机制提升Decoder对原文实体的复制能力；张天宇等[9]则利用大规模盗窃罪文书预训练司法领域词向量，并结合Adapter结构与边界指针网络提升表示能力与实体定位；彭晗等[10]通过上下文感知的嵌入式标注策略，将实体识别重建为条件生成任务，从而提升复杂语境下的识别效果。

总体来看，法律领域的NER研究在实体类型建模、领域语义注入和长实体处理方面已有一定进展。但现有方法难以有效应对语义歧义、嵌套结构及非规范实体表达仍显不足。因此如何有效地将深度学习等技术应用到司法命名实体识别中，仍需要进一步深入研究和探索。

2 基于全局指针网络的命名实体识别方法

给定一条文本序列 $S = \{X_1, X_2, \cdots, X_n\}$，其中 X_i 表示文本序列中的第 i 个词，n 表示序列长度。命名实体识别为每个 X_i 赋予一个标签 Y_i，这些标签共同构成了标签序列 $L = \{Y_1, Y_2, \cdots, Y_n\}$，其长度与输入文本序列相同。在标签体系中，$Y_i$ 取自集合 $\{B, I, O\}$，其中“B”（Begin）表示一个命名实体的起始，“I”（Inside）表示该实体内部的连续部分；“O”（Outside）则表示那些不属于任何命名实体的token。通过这样的标注方式，命名实体识别任务能够精准划分输入文本中所有命名实体的边界，并赋予它们对应的类别标签。

为了理解不同语境中的一词多义，解决静态词向量存在分词误差传递问题，本文提出基于全局指针网络的命名实体识别方法。首先利用预训练语言模型和BiLSTM，以便获得并加强句子的上下文语义表示。接着为了降低实体边界模糊引起的边界识别错误率，提高实体分类的准确度，本文模型引入全局指针来识别备选司法实体片段，并通过实体分类器得到最终的实体类别，具体结构如图1所示。

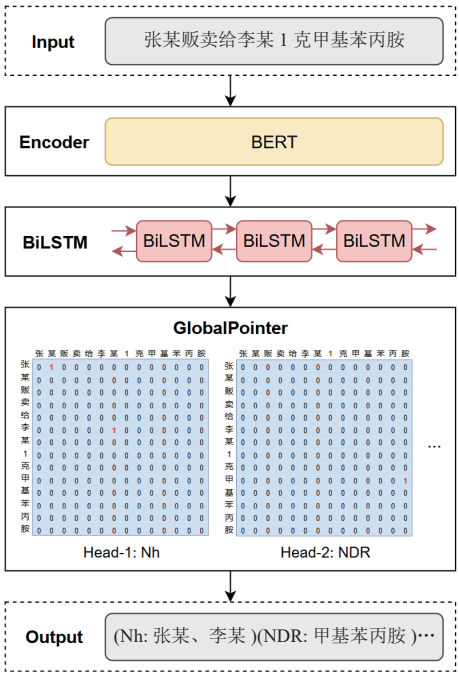

图1 基于全局指针网络的命名实体识别模型结构图

2.1 预训练语言模型

在本文的命名实体识别任务中，本文采用了BERT作为编码器，对法律文书的词汇和上下文信息进行编码，将其映射为包含深层语义信息的向量表征。BERT作为预训练语言模型，使用了双向Transformer结构，能够同时处理输入文本的上下文信息，从而捕捉到更丰富的语义特征，适用于多种自然语言处理任务。

BERT的输入表示由三部分组成，每个输入token x_i 都会被转换为嵌入向量 $\boldsymbol{E}(x_i)$，用公式表示为：

$$\boldsymbol{E}(x_i) = \boldsymbol{W}_T + \boldsymbol{W}_S + \boldsymbol{W}_P \tag{1}$$

式中：$\boldsymbol{W}_T$ 是token嵌入（token embedding）；$\boldsymbol{W}_S$ 是片段嵌入（segment embedding）；$\boldsymbol{W}_P$ 是位置嵌入（position embedding）。

BERT使用掩码语言模型（masked language model, MLM）任务进行训练。对于输入序列中的每个被遮蔽的[MASK]token，模型会预测其原始token，预测概率计算公式为：

$$P(x_i|\text{MASK}) = \frac{\exp\left(\boldsymbol{E}(x_i)^{\top}\boldsymbol{H}_i\right)}{\sum_{v\in V}\exp\left(\boldsymbol{E}(v)^{\top}\boldsymbol{H}_i\right)} \tag{2}$$

式中：$\boldsymbol{E}(x_i)$ 表示token x_i 的嵌入；$\boldsymbol{H}_i$ 表示Transformer编码器最后一层的输出；v表示词典，包含所有可能的token；$\exp(\boldsymbol{E}(x_i)^{\top}\boldsymbol{H}_i)$ 表示将标记嵌入 $\boldsymbol{E}(x_i)$ 和隐状态 $\boldsymbol{H}_i$ 进行点积后计算出的指数；$\sum_{v\in V}\exp\left(\boldsymbol{E}(v)^{\top}\boldsymbol{H}_i\right)$表示词典中所有token的未归一化概率总和，用于归一化以确保所有标记的概率之和为1。

2.2 BiLSTM 语义理解层

在通过 BERT 得到文本序列的向量表示之后，输入 BiLSTM 层进一步提取特征，获取上下文语义信息。BiLSTM 是一种基于长短期记忆网络的双向循环神经网络，BiLSTM 由两个独立的 LSTM 单元组成，分别用于处理从左到右和从右到左的输入序列。每个时间步的输入信息都通过这两个 LSTM 单元进行处理，最终的输出是这两个 LSTM 单元的隐藏状态的拼接或组合，表示当前时间步前后文的完整信息。LSTM 单元的结构主要由 5 个部分组成。

（1）遗忘门：遗忘门决定当前时间步应遗忘多少前一时间步的信息，它的输出值 $\boldsymbol{f}_t$ 介于 0~1 之间，遗忘门根据当前输入 $\boldsymbol{x}_t$ 和前一时间步的隐藏状态 $\boldsymbol{h}_{t-1}$ 生成一个控制信号 $\boldsymbol{f}_t$。当 $\boldsymbol{f}_t$ 越接近 1，前一时间步的记忆状态会被保留更多；当 $\boldsymbol{f}_t$ 越接近 0，前一时间步的记忆状态会被舍弃更多。遗忘门的激活公式为：

$$\boldsymbol{f}_t = \sigma(\boldsymbol{W}_f \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_f) \tag{3}$$

（2）输入门：输入门控制要写入记忆单元中的当前时间步的新信息数量。输入门生成两个信号：一个是输入门的输出 $\boldsymbol{i}_t$，范围在 0~1 之间，控制写入新信息的比例；另一个是当前时间步的候选记忆 $\widetilde{\boldsymbol{C}}_t$，代表新的可能存储在 LSTM 单元中的信息。输入门激活公式和候选记忆生成公式分别用公式表示为：

$$\boldsymbol{i}_t = \sigma(\boldsymbol{W}_i \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_i) \tag{4}$$

$$\widetilde{\boldsymbol{C}}_t = \tanh(\boldsymbol{W}_C \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_C) \tag{5}$$

（3）记忆单元：记忆单元的作用是存储序列的长期信息，通过遗忘门和输入门的组合控制来更新状态，遗忘门决定保留多少前一时间步的记忆状态，输入门决定将多少新的信息写入记忆单元中，用公式表示为：

$$\boldsymbol{C}_t = \boldsymbol{f}_t \cdot \boldsymbol{C}_{t-1} + \boldsymbol{i}_t \cdot \widetilde{\boldsymbol{C}}_t \tag{6}$$

式中：$\boldsymbol{C}_t$ 表示当前时间步的记忆状态，由前一时间步的记忆状态 $\boldsymbol{C}_{t-1}$ 和当前时间步的候选记忆 $\widetilde{\boldsymbol{C}}_t$ 通过遗忘门和输入门的输出加权求和得到。

（4）输出门：输出门控制当前时间步的输出，即如何将记忆单元中的信息映射为隐藏状态 $\boldsymbol{h}_t$。输出门根据记忆单元的状态决定当前时间步应该输出多少信息。结合输出门的激活结果和记忆单元的非线性变换，生成最终的隐藏状态，输出门激活公式为：

$$\boldsymbol{o}_t = \sigma(\boldsymbol{W}_o \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_o) \tag{7}$$

（5）隐藏状态：隐藏状态 $\boldsymbol{h}_t$ 是 LSTM 的输出，包含记忆单元的部分信息，经过输出门的筛选和非线性变换后，作为当前时间步的输出，并传递给下一时间步。隐藏状态更新的公式为：

$$\boldsymbol{h}_t = \boldsymbol{o}_t \cdot \tanh(\boldsymbol{C}_t) \tag{8}$$

BiLSTM 的结构中有两个 LSTM 单元，通过从前后两个方向同时处理输入序列，增强了对上下文信息的捕捉能力，有效解决了传统 RNN 在处理长距离依赖问题上的不足。

（1）前向和后向 LSTM 的计算公式为：

$$\overrightarrow{\boldsymbol{h}_t} = \mathrm{LSTM}(\boldsymbol{x}_t, \overrightarrow{\boldsymbol{h}_{t-1}}, \overrightarrow{\boldsymbol{C}_t}) \tag{9}$$

$$\overleftarrow{\boldsymbol{h}_t} = \mathrm{LSTM}(\boldsymbol{x}_t, \overleftarrow{\boldsymbol{h}_{t+1}}, \overleftarrow{\boldsymbol{C}_t}) \tag{10}$$

式中：$\overrightarrow{\boldsymbol{h}_t}$和$\overleftarrow{\boldsymbol{h}_t}$分别是前向 LSTM 在时间步 t 的隐藏状态；$\overrightarrow{\boldsymbol{h}_{t-1}}$是前向 LSTM 在时间步 t-1 的隐藏状态；$\overleftarrow{\boldsymbol{h}_{t+1}}$是后向 LSTM 在时间步 t+1 的隐藏状态；$\overrightarrow{\boldsymbol{C}_t}$和$\overleftarrow{\boldsymbol{C}_t}$分别是前向和后向 LSTM 在时间步 t 的记忆状态；$\boldsymbol{x}_t$ 是当前时间步的输入。

（2）最终隐藏状态：BiLSTM 的输出为前向与后向 LSTM 隐藏状态的拼接结果计算公式为：

$$\boldsymbol{h}_t = [\overrightarrow{\boldsymbol{h}_t}; \overleftarrow{\boldsymbol{h}_t}] \tag{11}$$

通过拼接前向和后向的隐藏状态，BiLSTM 能够捕捉到每个时间步前后文的完整信息。$\boldsymbol{h}_t$ 表示 BiLSTM 在时间步 t 的最终隐藏状态。

2.3 全局指针预测层

全局指针网络是一种基于跨度的命名实体识别框架，利用实体的起始和结束位置进行预测。当词向量经过 BiLSTM 层进一步特征提取之后输入全局指针预测层，全局指针网络矩阵结构如图 2 所示。

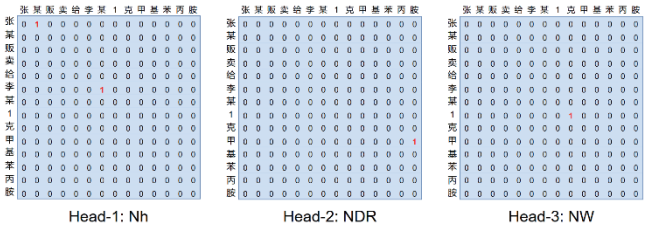

图 2 全局指针网络邻接矩阵

假设文本中可能的跨度集合为 $S = [s_1, s_2, \cdots, s_n]$，每个跨度 $s[i:j]$ 表示从位置 i 到 j 的片段。模型目标是识别所有属于实体类型集合 E 的片段 $s \in E$。给定一个文本序列 $X = [x_1, x_2, \cdots, x_n]$，在经过预训练语言模型 BERT 和 BiLSTM 特征提取层之后生成了向量表示矩阵 $\boldsymbol{H} = [h_1, h_2, \cdots, h_n]$，$\boldsymbol{H} \in \mathbb{R}^{n \times v}$，其中 n 表示句子中的 token 数量，v 表示每个 token 的向量维度。在得到文本表示 $\boldsymbol{H}$ 之后，使用两个前馈神经网络分别生成实体类型 α 对应的 token 作为起始和结束位置时的向量表示 $\boldsymbol{q}_{i,\alpha}$ 和 $\boldsymbol{k}_{i,\alpha}$，用公式表示为：

$$\boldsymbol{q}_{i,\alpha} = \boldsymbol{W}_{q,\alpha}\boldsymbol{h}_i + \boldsymbol{b}_{q,\alpha} \tag{12}$$

$$\boldsymbol{k}_{i,\alpha} = \boldsymbol{W}_{k,\alpha}\boldsymbol{h}_i + \boldsymbol{b}_{k,\alpha} \tag{13}$$

再通过计算这些向量的点积得到分数 $s_\alpha(i, j)$，从而预测跨度 $s[i:j]$ 属于某一特定的实体类型 α 的可能性。分数越高的

片段，属于实体类型 α 的可能性越高。用公式表示为：

$$s_\alpha(i,j)=\boldsymbol{q}_{i,\alpha}^{\top}\boldsymbol{k}_{j,\alpha} \tag{14}$$

通过将旋转位置编码（rotary position embedding, ROPE）应用到实体的起始和结束位置的向量表示中，使模型能够利用相对位置信息来增强对实体边界的感知能力。该编码机制满足 $\boldsymbol{R}_i^{\top}\boldsymbol{R}_j=\boldsymbol{R}_{j-i}$，即 token 对之间的相对位置可通过位置编码表示。这种方式将语义信息和位置信息进行了有效结合，能够帮助模型更准确地识别实体的起始和结束位置，从而提高命名实体识别的精度。融合了相对位置信息的评分函数计算公式为：

$$s_\alpha(i,j)=(\boldsymbol{R}_i\boldsymbol{q}_{i,\alpha})^{\top}(\boldsymbol{R}_j\boldsymbol{k}_{j,\alpha}) \tag{15}$$

由于在使用全连接层生成起始和结束位置的表示时，使用的权重矩阵 $\boldsymbol{W}_{q,\alpha}$ 和 $\boldsymbol{W}_{k,\alpha}$ 的大小为 $\mathbb{R}^{v\times d}$（v 是输入向量的维度，d 是全连接层输出的维度）。因此，每个新的实体类型都会引入额外的 $2\times v\times d$ 个参数（分别来自起始位置和结束位置的表示生成）。当实体类型增多时，模型参数会线性增长。为了减少参数的数量，模型引入了一种参数共享机制，并将任务分为两个子任务：实体片段提取（Extraction of Entity Spans）和实体分类（Classification of Entity Types）。

在实体片段提取任务中，模型只关注实体的边界，即确定哪些跨度 $s[i{:}j]$ 是实体，无需考虑它们的具体类型。因此，这个子任务的权重矩阵可以对所有实体类型共享。提取片段的分数计算公式为：

$$s_{\text{span}}(i,j)=(\boldsymbol{W}_q\boldsymbol{h}_i)^{\top}(\boldsymbol{W}_k\boldsymbol{h}_j) \tag{16}$$

在确定实体片段的边界（起始和结束位置）之后，需要对这些片段进行分类，实体分类任务会判断每个跨度属于哪个实体类型，分类打分函数的计算公式为：

$$s_{\text{class}}(i,j)=\boldsymbol{W}_\alpha^{\top}[\boldsymbol{h}_i;\boldsymbol{h}_j] \tag{17}$$

式中：$\boldsymbol{W}_\alpha$ 是与实体类型 α 相关的权重向量，大小为 $\mathbb{R}^{2v}$，即输入是拼接后的起始和结束位置的 token 表示向量，新的评分函数组合计算公式为：

$$s_\alpha(i,j)=\left(\boldsymbol{W}_q\boldsymbol{h}_i\right)^{\top}\left(\boldsymbol{W}_k\boldsymbol{h}_j\right)+\boldsymbol{w}_\alpha^{\top}[\boldsymbol{h}_i;\boldsymbol{h}_j] \tag{18}$$

通过这种拆分，模型的参数增长速度显著降低。每个新的实体类型增加的参数从 $2\times v\times d$ 个减少到 $2v$ 个。最后，为了解决多个标签样本间数量差异过大所导致的类别不平衡问题，模型采用了经过改进的交叉熵损失函数，利用指数函数放大了目标类别与其他类别的分数差值，使得正确类别的分数 s_t 显著高于其他类别，从而提高分类准确性，公式为：

$$\mathcal{L}=\log\left(1+\sum_{i=1,i\neq t}^{n}e^{s_t-s_t}\right) \tag{19}$$

式中：s_t 表示目标类别（即正确类别）的分数；s_i 表示其他类别的分数；n 表示总类别数；$\sum_{i=1,i\neq t}^{n}$ 表示对除了目标类别 t 以外的所有类别的分数进行求和；log 表示对整个损失值取对数，通常用于平滑梯度，避免梯度爆炸。

3 实验结果及分析

3.1 样本数据和实验设置

本文基于真实司法公开文书，人工标注构建了面向涉毒类刑事案件的细粒度实体关系语料库，经脱敏处理后，共标注五类关键实体：人名实体（Nh）、地名实体（Ns）、时间实体（NT）、毒品类型实体（NDR）和毒品重量实体（NW）。训练集与验证集按 8:2 比例划分，实体分布如表 1 所示。值得注意的是，毒品类型与重量实体具有高度结构化特征（如“甲基苯丙胺 30 克”），而地名常以模糊表述出现（如“某市某区”“外地”），这对模型泛化能力构成挑战。

表 1 涉毒类刑事案件语料库实体标签数量统计

实体标签类型	标签数量		
	训练集	验证集	合计
Nh	13 877	3 399	17 276
Ns	3 293	854	4 147
NT	3 952	1 058	5 010
NDR	6 340	1 728	8 068
NW	2 538	725	3 263

3.2 实验分析

本文模型与下列四种模型进行对比实验，代表了当前命名实体识别领域中较为先进的技术路线，通过此项对比实验，能够全面评估本文命名实体识别方法的表现。各模型的实验环境与外部超参数均保持一致，对比模型简介如下：

（1）BERT+CRF：将命名实体识别任务视为序列标注问题，利用 BERT 捕获句子中词语的深层语义和上下文信息，并通过 CRF 层建模标签之间的依赖关系，从而优化整个序列的标注概率。该方法特别适用于处理实体边界模糊和标签间强相关性的场景。

（2）BERT+MRC：将命名实体识别任务转化为机器阅读理解（MRC）问题。模型利用 BERT 提取上下文特征，并通过构造问题模板定位文本中的实体。不同问题用于抽取不同类别的实体，能够有效处理多类别和复杂嵌套的实体结构，提高实体识别的灵活性和准确性。

（3）BERT+SPAN：将命名实体识别任务视为实体边界预测问题，通过 BERT 生成句子的上下文表示，并预测实体的起始和结束位置。模型避免了序列标注方法对标签定义的依赖，更适用于长文本和实体边界复杂的场景。

（4）BERT+TPLinker Plus：利用 BERT 提取句子特征，并通过构造全局矩阵来建模实体的起始和结束位置对应关系，通过全局优化同时识别实体的类别和边界，特别适用于处理多实体重叠的情况，提升了复杂实体识别的性能。

本文提出的基于全局指针网络的命名实体识别方法表现出了优秀的性能，在与其他主流模型的对比实验中，本文模型在多个指标上均取得了较为显著的优势，在召回率和 F_1 值上，均超越了其他模型。其原因主要在于，与传统的 CRF 和 SPAN 方法相比，本文模型有两项改进：（1）双通道指针机制，通过头指针与尾指针的全局交互，计算所有词对构成实体的共现概率，从而避免 CRF 对局部标签转移规则的依赖；（2）类别感知解码器，在指针矩阵基础上引入轻量级分类头，同步预测实体边界及其类别，避免了 SPAN 类方法依赖后处理进行边界 - 类别匹配的问题。尽管本文模型在准确率上略低于 BERT+SPAN，但在召回率（95.89%）与 F_1 值（95.05%）上显著领先，表明其在减少漏识别方面更具优势——毒品重量等关键要素的漏识别可能引发量刑偏差甚至司法误判，因此高召回率在法律应用场景中具有优先级更高的价值。实验结果如表 2 所示。

表 2 各模型实验结果

单位：%

模型	准确率	召回率	F_1 值
BERT + CRF	94.46	95.39	94.92
BERT + MRC	94.40	95.38	94.89
BERT + SPAN	**94.87**	94.58	94.73
BERT+TPLinker Plus	92.62	93.13	92.87
BERT + Our Model	94.22	**95.89**	**95.05**

本文模型在命名实体识别任务中的各类标签识别结果如表 3 所示。根据实验结果可得，模型在人名、时间、毒品类型和毒品重量这四种实体类型的识别上表现出色，F_1 值均达到了 95% 以上，说明这些类型的信息在法律文书中具有较强的规范性和格式化特征，模型能够准确且全面地识别出这些实体。与之相比，模型在地名实体的识别上相对较弱，F_1 值为 76.13%，虽然识别结果仍较为准确，但地名的表达通常更加复杂和模糊，导致其识别效果相对较差。总体而言，本文模型在命名实体识别任务中表现优秀，特别是在人名、时间、毒品类型和毒品重量的识别上，展现了较高的精确度。虽然地名实体的识别效果略差，但整体 F_1 值较高，证明了模型在法律文书的实体识别任务上的有效性。

表 3 各标签实验结果

单位：%

实体标签类型	准确率	召回率	F_1 值
Nh	97.25	98.68	97.96
Ns	75.90	76.36	76.13
NT	94.91	96.36	95.63
NDR	96.91	99.07	97.98
NW	94.28	98.07	96.14

本文模型的训练 Loss 值数据变化如图 3 所示，从图中可以得出，模型的 Loss 值在训练的前 3 轮迅速下降，从初始的高值 7.961 9 快速收敛，表明模型在初始阶段对数据的拟合能力较强，模型参数快速调整以学习到有效的特征。在第 4 轮训练后，Loss 值下降的速率显著减缓，逐渐趋于平稳，并在第 10 轮后基本维持在一个较低的水平，说明模型在此阶段已逐渐接近最优解，训练的损失不再有显著变化，持续下降并最终稳定在 0.114 2 附近， 没有出现显著的反弹现象，表明模型在训练集上的表现较为良好，未出现明显的过拟合现象。

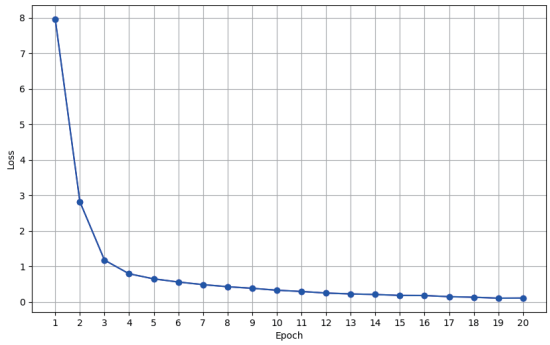

图 3 命名实体识别模型训练 Loss 值变化折线图

4 结论

本文提出了一种基于全局指针网络的命名实体识别方法，用于法律文书中的实体识别任务。经过在自建的涉毒类刑事案件语料库上的实验验证，该方法表现优异。实验结果表明，相较于传统模型，本模型在召回率和 F_1 值上均取得显著提升，分别达到了 95.89% 与 95.05%，有效提高了实体识别的准确性和鲁棒性。未来将进一步优化该方法，以应对其他类型案件文书中复杂实体的识别挑战，提升模型的适应性和泛化能力。

参考文献：

[1] 卢睿 , 李林瑛 . 一种面向法律文书的命名实体识别模型 [J]. 信息网络安全 ,2024,24(11):1783-1792.

[2]Wang Zihan, Zhao Ziqi, Lü Yougang, et al. A cooperative multi-agent framework for zero-shot named entity recognition[PP/OL].(2025-02-25)[2026-04-27].https://doi.org/10.48550/arXiv.2502.18702.

[3]Lu Hengyang, Liu Xintong, Shang Xingda, et al. A multi-agent framework for fine-grained multimodal named entity recognition through check and reasoning[C]//Proceedings of the 7th ACM International Conference on Multimedia in Asia. NewYork: ACM, 2025: 1-7.

[4]Wang Mengying, Lü Wenxiu, Zhang Yukun, et al. MDNER: integrating MRC with diffusion models for enhanced named entity recognition[J]. Neural Computing and Applications, 2025, 37(30): 25077-25094.

基于改进 YOLOv8n 的雾天行人车辆检测算法

安 旭[1] 胡蓉华[1]
AN Xu HU Ronghua

摘 要 针对雾天场景下，行人车辆检测算法存在检测精度低、参数量大的问题，文章提出了一种基于 YOLOv8n 改进的轻量化检测算法 YOLOv8n-TPDM。首先，在传统 C2f 模块中引入三重注意力机制，强化模型对空间和通道的关系理解和复杂数据的处理能力；同时在检测头位置增加小目标检测层，解决模型对小目标易漏检问题；将原始损失函数 CIoU 替换为 MPDIoU，提升锚框的精度，并加快收敛速度；最后，采用深度可分离卷积降低模型的参数量和计算量，并在雾天数据集 RTTS 上完成训练和验证。实验结果表明，与基准模型 YOLOv8n 相比，改进后的算法平均精度提高了 2.1%，模型参数量减少约 55.5%，能够更精准地完成雾天目标检测任务。

关键词 YOLOv8；轻量化；三重注意力机制；MPDIoU；检测头；深度可分离卷积

doi：10.3969/j.issn.1672-9528.2026.05.008

0 引言

当前行人车辆检测在智能交通以及自动驾驶等方面的应用十分广泛。但在雾天环境下，光照强度低，相机拍摄图片对比度下降、辨识度较低，算法对行人车辆检测的准确性下降，进而容易引发安全事故。因此，针对雾天交通图像的目标检测研究，对精确感知道路交通信息，减少交通事故发生，具有重要的现实意义 [1-3]。

随着技术不断发展，基于深度学习的行人车辆检测算法取得了诸多进展，在大规模数据集上训练达到较好的泛化能力。在雾天环境下，进一步提升目标检测精度仍具挑战，赖镜安等 [4] 提出基于 YOLOv5 轻量级雾天目标检测算法，采用感受野注意力模块，并使用 Slimneck 网络优化颈部结构提升准确率；Li 等 [5] 基于 YOLOv3 设计了一种图像处理模块，虽然在检测精度上有所提升，但计算量也有一定的增长；苏彤等 [6] 在 YOLOv5 的基础上，提出了 YOLOv5-SGE 网络，为提升模型的特征提取能力而在相应模块中引入了 SimAM 三维注意力机制，使用 GSConv 卷积替换标准卷积，减少参数量，实现模型的轻量化，最后替换传统损失函数为 EIoU，

1. 长江大学计算机科学学院 湖北荆州 434100

[5] 毛星亮 , 陈晓红 , 宁肯 , 等 . 全局和局部信息融合的案情关键要素识别 [J]. 软件学报 , 2023, 34(12):5724-5736.
[6]Lu Rui , Li Linying .Named entity recognition method of Chinese legal documents based on parallel instance query network[J].International Journal of Digital Crime and Forensics (IJDCF), 2024, 16(16):1-19.
[7] 杨长春 , 严鑫杰 , 顾晓清 , 等 . 融合字形特征的法律文书命名实体识别方法 [J]. 中文信息学报 , 2025, 39(9):91-99.
[8]Mao Xingliang, Jiang Jie, Zeng Yongzhe, et al. Generative named entity recognition framework for Chinese legal domain[J]. PeerJ Computer Science, 2024, 10: e2428.
[9] 张天宇 , 孙媛媛 , 杜文玉 , 等 . 基于语义边界增强的司法命名实体识别 [J]. 清华大学学报 (自然科学版), 2024, 64(5): 749-759.
[10] 彭晗 , 阮日青 , 胡颖 , 等 .JURIS: 基于理解增强型指令微调的司法命名实体识别方法 [J]. 电子学报 , 2025, 53(9): 3117-3133.

【作者简介】

李林瑛（1975—），男，辽宁大连人，博士，教授，研究方向：自然语言处理。

曲云平（1999—），男，辽宁大连人，硕士，研究方向：自然语言处理。

付子轩（2001—），女，辽宁鞍山人，硕士研究生，研究方向：自然语言处理。

刘美汐（2002—），女，辽宁沈阳人，硕士研究生，研究方向：自然语言处理。

（收稿日期：2026-01-21 修回日期：2026-05-06）