

基于自适应注意力与边界增强的中文命名实体识别方法

唐瑞雪, 吴利琴, 钱清

引用本文

唐瑞雪, 吴利琴, 钱清. 基于自适应注意力与边界增强的中文命名实体识别方法[J]. 计算机科学, 2026, 53(5): 309-318.

TANG Ruixue, WU Liqin, QIAN Qing. [Named Entity Recognition for Chinese Based on Adaptive Attention and Boundary Enhancement](#) [J]. Computer Science, 2026, 53(5): 309-318.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[多尺度聚合协同轴向语义引导的实体关系联合抽取方法](#)

Joint Entity and Relation Extraction Method with Multi-scale Collaborative Aggregation and Axial-semantic Guidance

计算机科学, 2026, 53(3): 97-106. <https://doi.org/10.11896/jsjcx.250500095>

[聚焦边界和多尺度特征融合的脑卒中病灶分割](#)

Boundary-focused Multi-scale Feature Fusion Network for Stroke Lesion Segmentation

计算机科学, 2026, 53(2): 264-272. <https://doi.org/10.11896/jsjcx.250300137>

[融合多尺度特征和注意力机制的时间序列预测模型](#)

Time Series Forecasting Model Integrating Multi-scale Features and Attention Mechanism

计算机科学, 2026, 53(2): 180-186. <https://doi.org/10.11896/jsjcx.250100113>

[基于生成式数据增强与Faster-RCNN改进的发动机打刻面缺陷检测](#)

Defect Detection of Engine Engraved Surface Based on Generative Data Augmentation and Improved Faster-RCNN

计算机科学, 2025, 52(11A): 241200025-7. <https://doi.org/10.11896/jsjcx.241200025>

[YOLO-BFEPS:一种高效注意力增强的跨尺度YOLOv10火灾检测模型](#)

YOLO-BFEPS:Efficient Attention-enhanced Cross-scale YOLOv10 Fire Detection Model

计算机科学, 2025, 52(6A): 240800134-9. <https://doi.org/10.11896/jsjcx.240800134>

基于自适应注意力与边界增强的中文命名实体识别方法

唐瑞雪 吴利琴 钱清

贵州财经大学信息学院 贵阳 550025

贵州省高等学校区块链与金融科技重点实验室 贵阳 550025

摘要 命名实体识别(NER)是自然语言处理中的核心任务之一,被广泛应用于信息抽取、问答系统、知识图谱构建等领域。然而,现有方法在处理中文文本中的嵌套实体和边界模糊问题时,仍面临多尺度特征利用不足、实体边界识别不准确等问题。为此,提出了一种面向中文命名实体识别的模型。该模型基于自适应注意力(AAM)与边界增强(BEM)机制,专门针对中文词汇无显式分隔、语义结构复杂等语言特性设计。模型通过自适应注意力机制动态融合局部与全局上下文特征,增强对中文复杂语义结构的建模能力;同时引入边界增强模块,利用深度卷积强化实体边界感知,有效缓解中文文本中嵌套实体与歧义边界带来的识别误差。实验结果表明,模型在 ACE2005-Chinese 和 Cnerta 两个中文嵌套数据集上的 F1 值分别达到 94.39% 和 83.72%,在 Weibo, Ontonotes 和 Resume 这 3 个中文非嵌套数据集上的 F1 值分别为 77.75%, 84.88% 和 96.36%,均优于现有主流中文命名实体识别方法,验证了其在复杂中文文本场景下的有效性与泛化能力。

关键词: 中文命名实体识别; 自适应注意力; 边界增强; 嵌套实体; 多尺度特征

中图分类号 TP391

Named Entity Recognition for Chinese Based on Adaptive Attention and Boundary Enhancement

TANG Ruixue, WU Liqin and QIAN Qing

School of Information, Guizhou University of Finance and Economics, Guiyang 550025, China

Key Laboratory of Blockchain and Fintech of Department of Education of Guizhou Province, Guiyang 550025, China

Abstract NER(Named Entity Recognition) is a fundamental task in natural language processing, with extensive applications in information extraction, question answering systems, and knowledge graph construction. However, existing approaches still struggle with inadequate multi-scale feature utilization and inaccurate boundary identification when processing nested entities and ambiguous entity boundaries in Chinese text. To tackle these challenges, this paper proposes a Chinese NER model incorporating an AAM(Adaptive Attention Mechanism) and a BEM(Boundary Enhancement Module), specifically designed to handle the absence of explicit word delimiters and complex semantic structures in Chinese. The AAM dynamically integrates local and global contextual features to enhance the modeling of intricate Chinese semantic patterns, while the BEM employs depthwise convolution to strengthen boundary perception, effectively reducing recognition errors caused by nested entities and ambiguous spans. Experimental results demonstrate that the proposed model achieves F1 scores of 94.39% and 83.72% on the nested Chinese datasets ACE2005-Chinese and Cnerta, and 77.75%, 84.88%, and 96.36% on the flat Chinese datasets Weibo, Ontonotes, and Resume, consistently surpassing existing mainstream Chinese NER methods and validating its effectiveness and generalization capability across diverse Chinese text scenarios.

Keywords Chinese named entity recognition, Adaptive attention, Boundary enhancement, Nested entity, Multi-scale features

命名实体识别(Named Entity Recognition, NER)^[1]是自然语言处理(Natural Language Processing, NLP)中的一个核心任务,旨在从文本中识别出具有特定意义的实体,如人名、地名、组织名等,在关系抽取^[2]、机器翻译^[3]、问答系统^[4]等众多领域发挥着重要作用。近年来,随着深度学习技术的快速

发展,命名实体识别方法已从传统的隐马尔可夫模型(Hidden Markov Model, HMM)^[5]、条件随机场(Conditional Random Field, CRF)^[6]逐步转向基于神经网络的架构,在性能上取得了显著提升。

对比英文等具有显式词分隔符的语言,中文 NER 由于

到稿日期:2025-09-11 返修日期:2025-12-09

基金项目:贵州省科技计划项目(黔科合基础-ZK[2022]一般 027);贵州省高等学校区块链与金融科技重点实验室建设项目(黔教技[2023]014);贵州财经大学 2025 年度在校学生科研项目(2025BAZYSY016)

This work was supported by the Science and Technology Projects of Guizhou Province(ZK[2022]027), Key Laboratory Program of Blockchain and Fintech of Department of Education of Guizhou Province([2023]014) and 2025 Student Research Projects in Guizhou University of Finance and Economics(2025BAZYSY016).

通信作者:唐瑞雪(trx_0401@163.com)

缺乏显式词边界,且存在一词多义和复杂的实体嵌套现象,面临着更为严峻的挑战。为解决这些特有难题,研究者主要从模型结构与特征表示两个方向进行研究。在模型结构方面,基于分层结构的方法将实体识别任务分解为多个层级,但通常受限于层级数,难以灵活适应不同长度的实体;基于超图的方法则借助超图结构来表示嵌套实体,在理论上具有可行性,但实际应用中的计算复杂度较高;基于跨度枚举的方法虽能有效处理嵌套结构,但随着文本长度的增加,计算效率会下降,且往往缺乏对局部信息与全局语义的协同建模。在特征表示方面,基于特征增强的方法通过引入外部知识或多模态信息辅助实体识别,但在处理边界模糊问题上仍存在不足。

当前中文 NER 主要面临两大核心挑战:一是实体边界模糊,无显式分隔符,嵌套实体与歧义短语加剧了边界混淆,如“中国科学院自动化研究所”易被误判为单一或割裂实体,“北京大学生”存在两种解读且边界无显式标识;二是多尺度特征利用不足,中文实体依赖“局部细节”与“全局上下文”协同,现有模型单一尺度特征难提取局部字符关联,也难挖掘全局语义依存,易致实体类别误判。

针对上述问题,本文提出基于自适应注意力与边界增强模块的中文命名实体识别方法。其创新点主要体现在 3 个方面:设计了自适应注意力模块(Adaptive Attention Module, AAM),通过 3×3 分组卷积捕获中文实体局部细节,通过 5×5 膨胀卷积挖掘全局上下文,并结合空间-通道注意力自适应调节不同尺度特征权重,从而缓解中文场景下单一尺度信息不足的问题;引入边界增强模块(Boundary Enhancement Module, BEM),基于深度卷积强化每个通道的中文实体边界特征,针对性解决中文嵌套、歧义场景下的边界模糊问题,降低边界识别错误率;同时,在 ACE2005-Chinese 和 Cnerta 等中文嵌套类数据集,以及 Weibo, Ontonotes, Resume 等中文非嵌套类数据集上开展实验,进一步验证该模型在中文复杂文本场景下的有效性。

1 相关工作

中文命名实体识别任务因文本缺乏显式分隔符且实体结构复杂而极具挑战性。近年来,研究者围绕结构建模与特征增强两大方向提出了多种方法,主要可分为以下 4 类。

1.1 基于分层结构的方法

基于分层的方法将实体识别任务分解为多个层级,通过逐层识别由内向外或由外向内的实体来解析嵌套结构。Duan 等^[7]提出了一种融合位置嵌入与多级 CRF 的层次模型,在每一层预测当前层级的实体,并利用路径屏蔽机制排除已识别实体对后续层的干扰,以此提升嵌套实体的识别准确率。Jia 等^[8]基于 ERNIE 预训练模型构建了分层识别框架,将任务拆解为多个平面的序列标注子任务。Li 等^[9]将掩码 Swin Transformer 与边界平滑策略相结合,在分层结构中引入更强的空间特征感知与边界约束。Wang 等^[10]建立多层次 CNER 模型,在 BiLSTM 层后引入软注意力掩码模块,自适应地处理不同复杂度的实体。这类方法在精度上有所提升,但其性能受限于预设的固定层级数,难以动态适配不同长度和复杂度的实体嵌套结构。

1.2 基于超图的方法

基于超图的方法通过超图结构来表示和处理嵌套实体。Lu 等^[11]首次提出用超图模型来表示可能的实体提及,这种方法可以指数级地表示所有可能的提及组合。后续研究在此基础上进行拓展:Muis 等^[12]将该方法扩展到不连续实体的识别;Wang 等^[13]利用深度神经网络增强超图模型,通过学习超图的结构来识别嵌套实体。尽管超图方法在理论上能够处理复杂的实体结构,但在实际应用中,它们面临着虚假结构和结构歧义的问题,这会导致推理过程中出现错误,并且计算复杂度较高。

1.3 基于跨度枚举的方法

为了突破分层方法结构的限制,基于跨度枚举的方法应运而生。这类方法通过枚举文本中所有可能的连续片段来识别实体,天然支持嵌套与重叠结构。Huang 等^[14]提出的 Levitated Control Attention,在不扩大候选量的前提下建立跨度间的高阶依赖,将复杂度从 $O(n^2)$ 降至 $O(n \log n)$ 。He 等^[15]提出的边界交互网络,通过双仿射结构建模开始-结束位置的配对置信度,显著减少了低质量跨度的召回量。Hu 等^[16]设计了边界掩码预训练,在跨度表示中显式加入 [MASK] 与 [B]/[E] 标记,以增强模型对左右边界的敏感度。Zhu 等^[17]利用图神经网络融合句法依存弧与跨度表示,在多任务框架下联合优化实体识别与关系抽取,进一步细化边界。这类方法在理论上能处理任意嵌套,但其本质是对所有可能片段进行独立评估,对局部线索与全局上下文信息利用不足,面对严重依赖上下文信息的文本时,区分度有限。

1.4 基于特征增强的方法

基于特征增强的方法聚焦于引入外部知识和多模态信息,以增强语义与边界表示能力。Liu 等^[18]提出的 USAF 模型创新性地引入了合成语音特征,利用声学边界信息来增强文本片段的边界感知能力。Zhu 等^[19]提出了融合词性特征的 CTE 模型,通过外部注意力机制捕获样本间的潜在语法关联,提升边界判别的准确性。Zhang 等^[20]提出的 VisPhone 模型整合了汉字的视觉字形特征与拼音特征,增强了对谐音字、形近字的判别能力。Yang 等^[21]提出的 LM-CNER 模型,通过融合字符与词向量、语序上下文及句法依存等内部多阶段特征,增强了中文 NER 的语义表示。Zhang 等^[22]提出的 MFLC 模型融合了汉字的字形特征、拼音特征及词性特征,构建了丰富的多模态汉字表示。Liu 等^[23]将文本-语音多模态学习引入农作物病虫害领域,利用音频中的语句停顿信息辅助中文分词。这些方法丰富了特征的多样性,但在解决中文边界模糊、上下文相关的难题时,表现仍不理想。

综上所述,现有研究在结构建模与特征增强方面均取得了实质性进展,但仍面临多尺度特征利用不充分、实体边界模糊等挑战。本文在此基础上,提出一种基于自适应注意力与边界增强的中文 NER 模型,旨在协同解决上述问题。

2 基于自适应注意力与边界增强模块的 NER 模型

2.1 任务定义

给定输入文本序列 $T = \{\omega_1, \omega_2, \dots, \omega_n\}$,命名实体识别模型 f 的核心目标是输出实体集合 $E = \{(e_i, c_i)\}$, e_i 是具有独

立语义的实体片段,片段为文本连续子序列,须满足起始与结束位置约束; c_i 是预定义的语义类别,类别除 Person(人名)、Location(地名)、Organization(机构名)等通用类外,还可依领域扩展 Disease(疾病名)、StockCode(股票代码)等专业类。模型须保证实体片段语义完整、类别归属准确,为信息抽取、知识图谱构建等提供结构化语义基元,本质上是实现非结构化文本到语义标注序列的映射。

2.2 模型架构

本文模型的整体架构如图1所示,由语义关联提取模块、自适应注意力模块(AAM)、边界增强模块(BEM)和实体类别

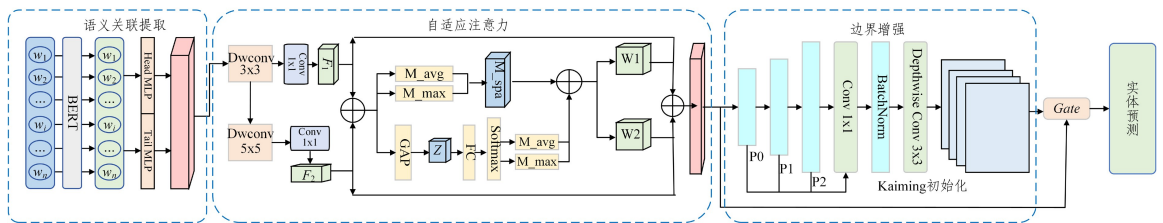


图1 本文模型框架

Fig. 1 Framework of the proposed model

2.2.1 语义关联提取模块

语义关联提取模块以 Hugging Face 的预训练模型 bert-base-chinese 作为初始编码器。预训练涵盖了中文维基百科等大规模语料数据,能够有效捕获中文词汇的语义关联和上下文依赖。bert-base-chinese 采用 Transformer 架构同时捕获文本的前向和后向语义信息,生成的向量更能够反映词汇在特定上下文中的独特含义。针对长度为 n 的句子 $T = \{\omega_1, \omega_2, \dots, \omega_n\}$,借助 bert-base-chinese 为每个单词生成表示向量,最终形成句子的表示向量集合 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ 。其中, \mathbf{x}_i 是第 i 个单词的表示向量,这些向量共同构成句子的初始二维语义表示基础。

根据 Geng 等^[24]提出的平面化句子表示方法,在预训练模型的基础上,为捕捉头尾位置的语义关联信息,通过带 dropout 的 MLP 分别对词级特征进行变换,具体如式(1)和式(2)所示:

$$\mathbf{H}_{\text{head}} = \text{LeakyReLU}(\text{Dropout}(\mathbf{W}_h \mathbf{H}_{\text{word}} + \mathbf{b}_h)) \quad (1)$$

$$\mathbf{H}_{\text{tail}} = \text{LeakyReLU}(\text{Dropout}(\mathbf{W}_t \mathbf{H}_{\text{word}} + \mathbf{b}_t)) \quad (2)$$

其中, $\mathbf{W}_h, \mathbf{W}_t \in \mathbf{R}^{d \times s}$ 为可训练的权重矩阵(d 为隐藏层维度, s 为双仿射特征维度); $\mathbf{b}_h, \mathbf{b}_t \in \mathbf{R}^s$ 是偏置向量。然后,通过双仿射变换计算头尾特征之间的关系得分,如式(3)所示:

$$\mathbf{S}_1 = \begin{cases} \text{MultiHeadBiaffine}(\mathbf{H}_{\text{head}}, \mathbf{H}_{\text{tail}}), & n_{\text{head}} > 0 \\ \mathbf{H}_{\text{head}} \mathbf{U} \mathbf{H}_{\text{tail}}^T, & \text{else} \end{cases} \quad (3)$$

其中, MultiHeadBiaffine 为多头部双仿射函数,当 $n_{\text{head}} > 0$ 时,默认头数 $n_{\text{head}} = 5$; $\mathbf{U} \in \mathbf{R}^{c \times s \times s}$ 是仿射变换的权重矩阵(c 为特征通道数)。为了进一步增强特征表示,拼接头尾特征与尺寸嵌入,并通过线性变换进行特征融合,如式(4)和(5)所示:

$$\mathbf{F}_{\text{cat}} = [\mathbf{H}_{\text{head}} \oplus \mathbf{H}_{\text{tail}} \oplus \mathbf{E}_{\text{size}}] \quad (4)$$

$$\mathbf{S}_2 = \text{Linear}(\mathbf{F}_{\text{cat}}) \quad (5)$$

其中, \oplus 表示通道维度拼接; \mathbf{E}_{size} 维度由可学习嵌入矩阵生成,输入为实体跨度长度; Linear 层权重采用 Xavier 正态分布初始化。最终,将双仿射分数 \mathbf{S}_1 和线性变换后的特征 \mathbf{S}_2 相加,得到融合后的基础特征矩阵:

预测模块组成。语义关联提取模块利用预训练模型生成上下文的词嵌入,并通过双仿射机制变换捕捉头尾语义关系,生成句子的二维表示。自适应注意力模块通过双路径特征提取局部细节和全局语义,然后利用空间-通道注意力机制动态融合多尺度特征,增强关键特征的表达能力。边界增强模块构建特征金字塔,通过不同膨胀率的卷积提取多尺度信息,并利用深度卷积强化实体边界特征。实体类别预测模块将增强后的特征映射到实体类别空间,通过线性变换和 ReLU 激活进行分类,采用带掩码的二元交叉熵损失指导训练,实现精准的实体类别预测。

$$\mathbf{S}_{\text{base}} = \mathbf{S}_1 + \mathbf{S}_2 \quad (6)$$

该矩阵即为增强后的二维语义表示,为后续模块提供了更丰富、更精准的语义信息。

2.2.2 自适应注意力模块

传统固定尺度的卷积操作难以同时捕获实体内部的局部细节与其在句子中的全局上下文,导致嵌套、歧义实体的关键特征被削弱。为解决这些问题,提升特征的质量和模型的性能,引入了自适应注意力模块。该模块的核心在于,其自适应注意力机制能够根据输入特征动态调整权重,自动平衡局部细节与全局上下文的关注度。

具体而言,模块首先通过双路径特征提取来捕捉不同尺度的信息。路径一通过 3×3 分组卷积捕捉局部细节特征,分组数等于输入通道数(即 $\text{cnn_dim} = 200$),确保通道间特征的独立提取。随后,通过 1×1 卷积调整通道维度至原始大小。路径二首先通过 3×3 分组卷积(同样采用分组数为 200)进行初步特征提取,后接 5×5 膨胀卷积(膨胀率 = 2)扩大感受野,以捕捉全局上下文信息,最终也通过 1×1 卷积统一通道维度。两条路径的输出特征在通道维度上进行拼接,形成融合后的多尺度特征表示。双路径特征提取的计算式如下:

$$\mathbf{F}_1 = \text{Conv}_{1 \times 1}(\text{GroupConv}_{3 \times 3}(\mathbf{S}_{\text{base}})) \quad (7)$$

$$\mathbf{F}_2 = \text{Conv}_{1 \times 1}(\text{GroupConv}_{5 \times 5, d=2}(\text{Conv}_{3 \times 3}(\mathbf{S}_{\text{base}}))) \quad (8)$$

$$\mathbf{F}_{\text{path}} = \mathbf{F}_1 \oplus \mathbf{F}_2 \quad (9)$$

其中, $\text{GroupConv}_{k \times k, d}$ 表示核大小为 $k \times k$ 、膨胀率为 d 的分组卷积,分组数均为输入通道数,确保通道间特征独立提取。

在自适应注意力机制的作用下,模块进一步通过空间和通道两个维度对特征进行增强。空间注意力机制通过全局平均池化和最大池化聚合空间信息,生成空间注意力权重 M_{spa} 。空间注意力机制的相关计算式如下所示:

$$\mathbf{M}_{\text{avg}} = \text{Mean}(\mathbf{F}_{\text{path}}) \quad (10)$$

$$\mathbf{M}_{\text{max}} = \text{MaxPool}(\mathbf{F}_{\text{path}}) \quad (11)$$

$$\mathbf{M}_{\text{spa}} = \text{Concat}(\mathbf{M}_{\text{avg}}, \mathbf{M}_{\text{max}}) \quad (12)$$

其中, $\mathbf{M}_{\text{avg}} \in \mathbf{R}^{b \times 1 \times n \times n}$, $\mathbf{M}_{\text{max}} \in \mathbf{R}^{b \times 1 \times n \times n}$ (b 为批次大小, n 为序列长度), 拼接后 $\mathbf{M}_{\text{spa}} \in \mathbf{R}^{b \times 2 \times n \times n}$ 。

通道注意力机制通过全局平均池化压缩空间维度, 并结合全连接层生成通道注意力权重 \mathbf{A} 。通道注意力机制的计算式如下:

$$\mathbf{Z} = \text{AdaptiveAvgPool}(\mathbf{F}_{\text{path}}) \quad (13)$$

$$\mathbf{A} = \text{Softmax}(\text{FC}_2(\text{ReLU}(\text{FC}_1(\mathbf{Z})))) \quad (14)$$

其中, $\mathbf{Z} \in \mathbf{R}^{b \times c \times 1 \times 1}$ 为通道全局特征, $\mathbf{A} \in \mathbf{R}^{b \times 2 \times c/2 \times 1}$ 为通道注意力权重。这些权重不仅反映了特征的重要性, 而且能够根据输入数据的特性进行动态调整, 凸显了 AAM 模块的自适应能力。

特征融合过程中, 空间与通道注意力权重协同作用于双路径特征, 通过逐元素加权融合生成增强特征。这一过程不仅完整保留了输入特征的空间和通道结构, 还使模型能够聚焦于对实体识别最关键的特征部分。特征融合的计算式如下:

$$\mathbf{W}_1 = \mathbf{A}_1 \cdot \mathbf{M}_{\text{spa}}[:, 0, :, :] \quad (15)$$

$$\mathbf{H}_2 = \mathbf{A}_2 \cdot \mathbf{M}_{\text{spa}}[:, 0, :, :] \quad (16)$$

$$\mathbf{F}_{\text{attn}} = \mathbf{F}_1 \cdot \mathbf{H}_1 + \mathbf{F}_2 \cdot \mathbf{H}_2 \quad (17)$$

$$\mathbf{S}_{\text{AAM}} = \mathbf{S}_{\text{base}} \cdot \text{Sigmoid}(\text{Conv}_{1 \times 1}(\mathbf{F}_{\text{attn}})) \quad (18)$$

其中, $\mathbf{A}_1, \mathbf{A}_2$ 为 \mathbf{A} 的两组子权重; $\mathbf{W}_1, \mathbf{W}_2$ 为融合后的空间-通道权重; Sigmoid 函数将注意力值归一化至 $[0, 1]$ 。这些权重并非预设常量, 而是完全基于当前输入 \mathbf{S}_{base} 即时计算所得。 $\mathbf{A}_1, \mathbf{A}_2$ 的 Softmax 分布随文本中实体长度的变化实时调整, 确保模型在不同场景下均能聚焦于最关键特征。特别地, AAM 模块的自适应特性体现在权重 $\mathbf{W}_1, \mathbf{W}_2$ 的动态调整上, 当遇到长实体时, 全局路径权重 \mathbf{A}_2 自动升高; 而面对短实体时, 局部路径权重 \mathbf{A}_1 则占据优势。同时, 空间注意力权重 \mathbf{M}_{spa} 的每个值直接由特征统计量决定, 能够在遇到歧义边界时即时聚焦于可能的切分位置, 无歧义时则保持均匀响应。

2.2.3 边界增强模块

边界增强模块结合特征金字塔捕捉并融合多尺度信息, 通过深度卷积强化实体边界特征, 解决实体边界模糊问题, 提升模型对实体边界的感知能力。

为捕获不同尺度的实体信息, 构建特征金字塔结构。首先, 通过 3 个不同膨胀率的卷积操作提取多尺度特征, 3 个膨胀卷积分支的输入/输出通道数均为 cnn_dim (默认 200), 拼接后通过 1×1 卷积将总通道降回 cnn_dim , 保证后续模块维度一致。

$$\mathbf{P}_i = \text{Conv}_{3 \times 3, d=2^i}(\mathbf{S}_{\text{AAM}}) (i=0, 1, 2) \quad (19)$$

其中, \mathbf{S}_{AAM} 是自适应注意力模块的输出特征; $d=2^i$ 为膨胀率, 通过递增的膨胀率扩大感受野, 覆盖不同尺度的实体范围。

然后, 将这些特征进行融合, 融合后的特征通过边界增强模块进一步强化边界信息。

$$\mathbf{P}_{\text{fuse}} = \text{BatchNorm}(\text{Conv}_{1 \times 1}([\mathbf{P}_0 \oplus \mathbf{P}_1 \oplus \mathbf{P}_2])) \quad (20)$$

其中, \oplus 表示通道维度拼接。 1×1 卷积将 3 个尺度的特征融合为统一通道数, BatchNorm 后接 LeakyReLU (负斜率 = 0.1), 引入非线性变换, 增强特征表达能力。通过深度卷积对融合后的多尺度特征进行边界增强, 每个输入通道独立卷积以强化通道内的边界信息。

$$\mathbf{S}_{\text{BEM}} = \text{DepthwiseConv}_{3 \times 3}(\mathbf{P}_{\text{fuse}}) \quad (21)$$

具体地, 深度卷积对每个通道进行单独操作。

$$\mathbf{S}_{\text{out}} = \sum_{k=1}^K \mathbf{W}_k * \mathbf{S}_{\text{in}, k} \quad (22)$$

其中, K 为输入通道数, \mathbf{W}_k 为第 k 个通道的卷积核, $*$ 表示单通道卷积操作, $\mathbf{S}_{\text{in}, k}$ 为输入特征的第 k 个通道。

为充分利用 AAM 与 BEM 各自的优势, 引入可学习门控融合机制。将 BEM 输出 \mathbf{S}_{BEM} 与 AAM 输出 \mathbf{S}_{AAM} 按元素级加权融合。

$$\mathbf{P}_{\text{enhance}} = \alpha \cdot \mathbf{S}_{\text{AAM}} + (1 - \alpha) \cdot \mathbf{S}_{\text{BEM}} \quad (23)$$

其中, $\alpha \in [0, 1]$ 为可学习标量参数, 初始值为 0.5, 随训练过程自动更新。该机制使模型能够自适应地平衡多尺度语义特征与边界强化特征, 弥补单一模块可能引入的偏差。

2.2.4 实体类别预测模块

实体类别预测模块负责将增强后的特征映射到实体类别空间, 输出每个位置对的实体类别得分。将增强后的特征 $\mathbf{P}_{\text{enhance}}$ 进行线性变换, 映射到实体类别空间。

$$\mathbf{Y}_{\text{pred}} = \mathbf{W}_{\text{cls}} \cdot \text{ReLU}(\mathbf{P}_{\text{enhance}}) \in \mathbf{R}^{B \times N \times N \times C} \quad (24)$$

其中, B 是批量大小, N 是序列长度, C 是实体类别数, $\mathbf{W}_{\text{cls}} \in \mathbf{R}^{c \times C}$ 是分类权重矩阵 (c 为特征通道数), ReLU 激活函数引入非线性变换以增强特征表达能力。

为了训练模型, 采用带掩码的二元交叉熵损失函数计算预测值与真实值之间的差异。

$$L = \frac{1}{\sum \mathbf{M}_{i,j}} \sum \mathbf{M}_{i,j} \cdot \text{BCEWithLogits}(\mathbf{Y}_{\text{pred}, i,j}, \mathbf{Y}_{\text{true}, i,j}) \quad (25)$$

其中, $\mathbf{M}_{i,j} \in \{0, 1\}$ 为有效位置掩码 (1 表示位置 (i, j) , 为有效实体候选, 0 表示无效); $\mathbf{Y}_{\text{true}, i,j}$ 为真实标签; BCEWithLogits 为带 Sigmoid 激活的二元交叉熵函数, 自动将预测分数归一化至 $[0, 1]$ 并计算损失。

通过上述步骤, 实体类别预测模块能够准确预测实体类别, 并利用损失函数指导模型训练, 从而在命名实体识别任务中取得优异性能。

3 实验与分析

为了验证所提方法的有效性, 在 ACE2005-Chinese 和 Cnerta 两个嵌套类数据集, 以及 Weibo, Ontonotes 和 Resume 这 3 个非嵌套类数据集上进行了实验, 并将本文模型与其他基于深度学习的命名实体识别模型进行了对比。数据集的基本信息如表 1 所列。

表 1 数据集的基本信息

Table 1 Basic information of datasets

数据集	训练集	验证集	测试集	实体类型
ACE2005-Chinese	11 601	1 450	1 450	7
cnerta	34 102	4 440	4 445	3
Weibo	1 349	268	270	4
Ontonotes	37 557	6 217	4 293	18
Resume	3 819	463	477	8

3.1 数据集

实验所采用的数据集可根据其标注的实体结构, 划分为嵌套类与非嵌套类两大类, 以全面评估模型在不同实体结构场景下的性能。

3.1.1 嵌套类数据集

ACE2005-Chinese:本数据集由 LDC(语言数据联盟)发布,语料来源多样,涵盖了新闻报道、广播新闻及新闻组对话等多种文体。其标注体系较为丰富,包括了人名(PER)、组织名(ORG)、地名(LOC)和设施名(FAC)等实体类型。

Cnerta:作为一个广泛使用的中文命名实体识别基准,Cnerta的语料同样来源于新闻报道和社交媒体等多种渠道。它标注了 PER, LOC, ORG 等常见实体类型。与 ACE2005-Chinese 类似,Cnerta 中也普遍存在实体嵌套结构。

3.1.2 非嵌套类数据集

Weibo:构建自微博平台,反映了网络口语化文本的特点。标注的实体类型主要包括 PER, LOC, ORG 等常见实体类型,常被用于检验模型在非正式文本上的泛化能力。

Ontonotes:一个大规模的多语言命名实体识别数据集,包含多种文本类型,如新闻报道、广播新闻等。其提供了广泛的实体类型标注,除常见的 PER, LOC, ORG 外,还涵盖了设施名(FAC)、产品名(PROD)等。

Resume:基于真实的中文简历文本构建。其标注的实体类型与简历内容高度相关,包括人名(PER)、公司名(ORG)、职位名(POS)以及教育机构名(EDU)等。

3.2 实验设置

本文实验所用的数据集均采用 Yan 等^[25]发布的统一预处理脚本完成数据标准化,对“格式统一性”进行处理,直接采用 Hugging Face 官方 bert-base-chinese 权重,模型基于 PyTorch 2.3.0 深度学习框架构建,运行于 Ubuntu 22.04 操作系统,依赖 Python 3.12 版本。训练阶段选用 AdamW 作为优化器,全局统一核心参数设置如下:初始学习率为 2×10^{-6} ,Dropout 正则化率为 0.1,学习率 warmup 比例为 0.1(采用线性调度策略),权重衰减系数为 1×10^{-2} 。考虑到不同数据集的文本复杂度与实体分布差异,设置数据集专属差异化超参数(最大训练轮数、批次大小),具体如表 2 所列。

表 2 实验超参数设置

Table 2 Experimental hyperparameter settings

数据集	最大训练轮数	批次大小
ACE2005-Chinese	50	24
cnerta	50	24
Weibo	50	24
Ontonotes	5	8
Resume	30	24

3.3 评价指标

本文采用准确率(Precision, P)、召回率(Recall, R)和 F1 值(F1-score)作为主要评估指标,以全面评估本模型的各项性能。其中,P 用于衡量模型预测为实体的样本中真正实体的比例,R 用于衡量模型能够识别出的真正实体的比例,F1 值则为准确率和召回率的调和平均值,综合反映了模型的性能。

$$P = \frac{TP}{TP + FP} \quad (26)$$

$$R = \frac{TP}{TP + FN} \quad (27)$$

$$F1 = 2 \cdot \frac{P \cdot R}{P + R} \quad (28)$$

3.4 实验结果

为全面评估本文方法的有效性,在多个中文命名实体识别数据集上进行了实验,包括嵌套实体数据集(ACE2005-Chinese, Cnerta)和非嵌套实体数据集(Weibo, Ontonotes, Resume)。表 3—表 5 列出了本文模型与多个主流方法的对比结果。

表 3 ACE2005-Chinese 数据集上的对比结果

Table 3 Comparison results on ACE2005-Chinese dataset

	(%)		
Model	P	R	F1
本文模型	93.74	95.06	94.39
Jia 等 ^[8]	84.40	87.50	85.90
Huang 等 ^[14]	93.02	88.44	90.67
He 等 ^[15]	88.65	90.94	89.78
Hu 等 ^[16]	93.39	89.49	91.40

表 4 Cnerta 数据集上的对比结果

Table 4 Comparison results on Cnerta dataset

	(%)		
Model	P	R	F1
本文模型	84.02	83.42	83.72
Duan 等 ^[7]	78.13	60.54	68.21
Liu 等 ^[18]	75.59	80.92	78.16
Liu 等 ^[23]	80.20	81.93	81.05
Li 等 ^[9]	86.08	80.63	83.27

表 3 与表 4 的结果表明,本文模型在两个嵌套实体数据集上均取得了最优的 F1 值,验证了其在处理复杂嵌套结构时的有效性。

在 ACE2005-Chinese 数据集上,本文模型相较于 Jia 等^[8]的分层 ERNIE 模型在 F1 值上提升了 8.49 个百分点,表明 AAM 能够通过动态融合多尺度特征,有效缓解固定分层结构可能造成的信息割裂问题;相比于 He 等^[15]引入边界交互信息的分层方法,本文模型的 F1 值提升了 4.61 个百分点,体现了 BEM 在强化边界感知方面的优势。

在 Cnerta 数据集上,本文模型相较于 Duan 等^[7]基于位置嵌入与多级预测的跨度方法,在 F1 值上取得了 15.51 个百分点的显著提升。这表明在不依赖于显式枚举所有可能候选跨度的情况下,通过 AAM 与 BEM 的协同作用,依然能实现对嵌套结构的精准识别。

在非嵌套实体识别任务上的性能对比如表 5 所列。本文模型在 3 个数据集上均达到了领先的 F1 值,展现了良好的泛化能力。在口语化程度高、噪声明显的 Weibo 数据集上,本文模型的 F1 值较基线模型最高提升了 14.51 个百分点,体现出 AAM 在捕捉局部关键特征方面的能力,以及 BEM 模块在噪声环境下的边界鲁棒性。在实体类型复杂、语境多样的 Ontonotes 数据集上,模型凭借自适应注意力机制,实现了对多种实体的准确判别。在实体边界清晰、基线模型性能趋近饱和的 Resume 数据集上,本文模型仍能通过精细的特征权重分配与边界优化,实现性能的提升。模型通过 AAM 中局部与全局路径的协同设计,并结合 BEM 的边界强化机制,有效缓解了传统注意力方法在中文场景下对局部边界特征关注不足的问题。

表5 非嵌套数据集上的对比结果

Table 5 Comparison results on non-nested datasets

Model	(%)								
	Weibo			Ontonotes			Resume		
	P	R	F1	P	R	F1	P	R	F1
本文模型	77.01	78.50	77.75	83.27	86.55	84.88	96.78	95.96	96.36
Zhu 和 Cheng ^[17]	—	—	63.65	76.96	76.94	76.95	96.04	95.86	95.95
Yang 等 ^[21]	68.67	58.61	63.24	—	—	—	95.41	95.52	95.66
Zhang 等 ^[22]	67.93	71.13	69.48	77.82	82.21	79.95	94.41	96.38	95.41
Zhu 等 ^[19]	68.32	63.75	65.95	75.94	78.61	77.25	95.61	96.05	96.11

综上所述,本文提出的基于自适应注意力与边界增强的中文NER模型,在嵌套与非嵌套任务上均表现出色。AAM通过多尺度特征提取与自适应融合,增强了对中文复杂语义结构的建模能力;BEM则通过金字塔融合与深度卷积,针对性地强化了实体边界表征,验证了AAM与BEM模块协同作用的有效性。本研究为提升中文命名实体识别的性能,特别是解决边界

模糊与多尺度特征利用问题,提供了一种有效的方案。

3.5 消融实验分析

为了验证本文模型各组成部分的有效性,本节设计了消融实验。“w/o AAM”表示去掉自适应注意力模块;“w/o Pyramid”表示去掉特征金字塔结构;“w/o BEM”表示去掉边界增强模块。消融实验在5个数据集上的实验结果如表6所列。

表6 消融实验结果

Table 6 Results of ablation experiment

Model	(%)														
	Ace2005-Chinese			Cnerta			Weibo			Ontonotes			Resume		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
本文模型	93.74	95.06	94.39	84.02	83.42	83.72	77.01	78.50	77.75	83.27	86.55	84.88	96.78	95.96	96.36
w/o AAM	94.24	95.24	94.23	82.99	82.46	82.75	75.23	79.95	77.52	82.93	86.23	84.55	95.96	96.26	96.11
w/o Pyramid	94.63	94.31	93.97	82.72	81.95	82.33	64.57	66.91	65.72	83.91	85.54	84.71	95.73	96.26	95.99
w/o BEM	94.12	94.53	94.32	83.2	83.09	83.15	74.77	80.19	77.39	82.43	87.15	84.73	95.05	96.56	95.80

从表6可以发现,移除AAM模块后,所有数据集的F1值均出现下降,这表明AAM能有效整合空间与通道维度的特征信息,增强模型对多层次语义特征的捕捉能力;当去除特征金字塔结构后,模型性能下降最为显著,证明特征金字塔对处理实体长度差异大、边界模糊的复杂场景具有关键作用;移除BEM模块导致各数据集F1值普遍下降,证明边界增强模块能有效强化实体边界特征感知,提升边界判别的准确性。其中,在实体类型单一、边界清晰的Resume数据集上,各消融模型的性能降幅相对较小,表明在此类简单场景中,模型对特定模块的依赖性有所降低。而在实体边界模糊、嵌套严重的Cnerta和Weibo数据集上,各模块的贡献更为突出。

识别表现良好,尤其在样本充足的实体类型上效果显著,展现出良好的实用性与稳定性。

综上,自适应注意力机制、金字塔结构与边界增强模块均对模型性能提升具有重要价值,三者协同工作共同保障了模型在不同复杂度场景下的实体识别能力。

3.6 实体类型识别性能分析

为验证模型在不同实体类型上的识别能力与泛化性能,在ACE2005-Chinese数据集上对7类实体进行了实验,结果如图2所示。

结果表明,模型在不同实体类型上均展现出稳健的识别性能。具体而言,在PER、GPE、ORG等数据量充足(样本量为828,532,437)的实体类型上,模型表现最为优异,其中PER更是达到94.6%,这充分体现了模型对常见实体具备较强的识别能力。对于FAC和LOC等中等规模(样本量为145,123)的实体,模型仍能维持79.0%~81.5%的较高F1分数,性能波动较小。值得关注的是,在样本量仅为35的VEH低频实体上,模型仍能取得87.2%的F1分数,不仅优于部分中等样本实体,也显示出模型对低频实体具备一定的泛化能力。总体而言,模型在该数据集上对不同实体类型的

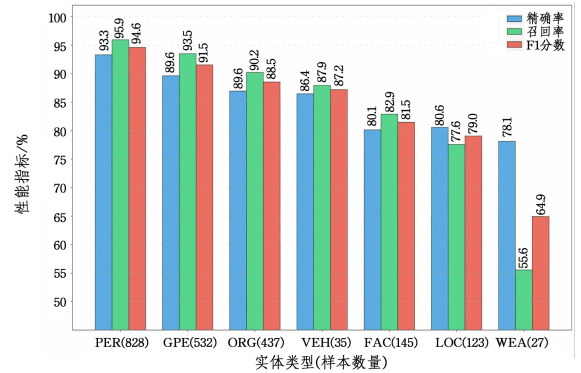


图2 不同实体类型的性能对比

Fig. 2 Performance comparison of different entity types

3.7 AAM的有效性分析

为了验证AAM的有效性,设计了3种AAM的变体模型在ACE2005-Chinese和Resume数据集上进行对比实验,以探究AAM各核心组件的具体作用。“变体1”将AAM空间与通道协同注意力机制改为分离式设计,用于验证“协同融合”机制的有效性;“变体2”将自适应权重固定为0.5,用于对比“自适应”与固定融合方式的性能差异;“变体3”仅保留 3×3 单尺度卷积,用来验证“多尺度”设计的必要性。3种变体与完整AAM的性能对比结果如表7所列。

结果表明,完整AAM在两个数据集上均取得最优性能。在ACE2005-Chinese数据集上,“变体1”的F1值下降0.25个百分点,表明空间与通道注意力的分离式设计会削弱特征能力,验证了协同融合机制的有效性;“变体2”的F1值下降0.45个百分点,说明固定权重分配会显著限制模型对实体

边界特征的动态感知能力;“变体 3”的 F1 值下降 0.35 个百分点,证明多尺度感受野对于处理嵌套实体结构具有重要价值。在 Resume 数据集上,3 种变体的 F1 值分别下降 0.83 个百分点、0.56 个百分点和 0.91 个百分点,进一步验证了 AAM 完整设计的必要性。即使在实体边界相对清晰的场景中,任一核心组件的缺失也都会导致性能明显下降,从而充分证明了 AAM 设计的合理性与有效性。

表 7 AMF 各组件的影响
Table 7 Impact of AMF components

Model	Impact of AMF components (%)					
	ACE2005-Chinese			Resume		
	P	R	F1	P	R	F1
AAM	93.74	95.06	94.39	96.78	95.96	96.36
变体 1	93.10	95.21	94.14	95.3	95.77	95.53
变体 2	92.44	95.48	93.94	95.71	95.89	95.80
变体 3	93.29	94.79	94.04	95.07	95.83	95.45

3.8 超参数敏感性分析

为验证模型对学习率与 Dropout 正则化率的敏感程度,以 ACE2005-Chinese 为例,固定其余超参数,采用 4×4 网格搜索;学习率 $lr \in \{1 \times 10^{-6}, 2 \times 10^{-6}, 5 \times 10^{-6}, 1 \times 10^{-5}\}$, Dropout 率 $p \in \{0.05, 0.1, 0.2, 0.3\}$,结果如图 3 所示。

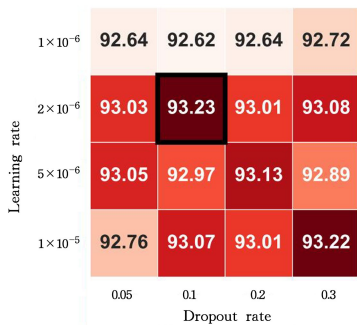


图 3 lr-p 组合性能热力图

Fig. 3 Heatmap of lr-p combination-based performance

从图 3 能够观察到:当学习率为 2×10^{-6} 且 Dropout 率为 0.1 时,模型取得了相对更优的性能表现。综合各方面因素考量,后续实验采用 $lr=2 \times 10^{-6}$, $p=0.1$ 这一稳健的最优超参数组合。

3.9 极端场景鲁棒性分析

为了验证所提模型在极端输入场景下的边界性能与稳定性,基于 ACE2005-Chinese 原始测试集,针对性地构造 3 类极端测试子集,以模拟实体识别任务中可能面临的特殊数据分布场景。

1) 超短实体子集:筛选含文本字符跨度小于等于 2 (含边界值)的实体样本,此类实体因字符长度极短、语义信息密度高,易出现边界定位偏差,如“韩国”等。

2) 超长实体子集:选取含文本字符跨度大于等于 10 的实体样本,此类实体通常包含多层修饰成分,结构复杂,易因信息过载而识别不完整,如“国防科学技术工业委员会副主任”等。

3) 深层嵌套实体子集:提取同一句内存在至少一个实体被另外两层实体完全包裹的样本,形成字符跨度嵌套层级大于等于 3 的结构。此类场景需模型精准捕捉实体间的包含关系,如“[上海 [浦东 [川沙工业园区]]]”(地名嵌套)等。

为确保实验结果的可比性,所有测试均保持原始训练流程(含参数设置、优化器、迭代次数等)不变,仅将标准测试集替换为上述 3 类极端子集。最终性能评估结果如表 8 所列。

表 8 极端测试子集性能

Table 8 Extreme-test subset performance

极端子集	P	R	F1
超短实体(小于等于 2 字符)	94.12	95.03	94.57
超长实体(大于等于 10 字符)	93.31	93.72	93.52
深嵌套(大于等于 3 层)	93.81	94.01	93.91

由表 8 数据可知,模型在 3 类场景下性能优异且稳定:超短实体的 F1 值为 94.57% (精确率为 94.12%, 召回率为 95.03%),这得益于 AAM 的 3×3 局部卷积与 BEM 的边界区分能力;超长实体的 F1 值为 93.52% (精确率为 93.31%, 召回率为 93.72%),这是由于 AAM 的 5×5 膨胀卷积有效扩大了感受野;深层嵌套实体的 F1 值为 93.91% (精确率为 93.81%, 召回率为 94.01%),这是因为 AAM 与 BEM 协同区分层级边界。3 类场景下的 F1 值均大于等于 93.5%,无显著的性能波动,证明 AAM 与 BEM 使模型对复杂文本场景具备强鲁棒性,验证了其在极端任务中的可靠性。

3.10 模型复杂度与效率分析

为全面评估本文模型的实际效率,采用 Gou 等^[26]提出的基于潜在词汇化选区解析的嵌套命名实体识别模型作为基线模型,从参数量、计算复杂度和推理速度等方面进行系统对比,并详细分析了 AAM 与 BEM 引入的计算开销,结果如表 9 所列。

表 9 模型复杂度与效率对比分析

Table 9 Model complexity and efficiency comparison

	可训参数	计算量	单句延迟	吞吐量
基线模型	1.0399×10^8	1.390×10^{10}	43.97	23
本文模型	1.0378×10^8	1.302×10^{10}	7.19	139
本文模型相对基本模型	-0.2%	-6.3%	-82%	+6.1×
本文模型去掉 AAM	-4.0×10^5	-1.02×10^9	+0.2	147
本文模型去掉 BEM	$+2.4 \times 10^5$	$+9.4 \times 10^8$	+33.5	24

注:“+6.1×”表示“吞吐量提高到原来的 6.1 倍”;可训参数单位为个,计算量单位为 FLOPs,单句延迟单位为 ms,吞吐量单位为 qps。

结果表明,本文模型在保持高性能的同时,具备了良好的效率特性。与基线模型相比,本文模型在参数量轻微减少、计算量显著降低的基础上,实现了推理速度的大幅提升,单句延迟缩短 82%,吞吐量提升至原来的 6.1 倍,充分满足了实际应用中模型效率的要求。

综合分析表明,AAM 以极低的计算开销为模型提供了自适应注意力;BEM 则作为核心效率优化模块,大幅降低了计算量,缩短了推理延迟,并提升了吞吐量。AAM 与 BEM 模块协同作用在模型效率优化中发挥了重要作用,使得本文模型在性能与效率之间实现了良好平衡,为后续模型在实际场景中的部署与应用提供了基础。

3.11 模型可解释性分析

为验证模型对嵌套实体结构的感知机制,本节从

ACE2005-Chinese 测试集中抽取两个典型样例进行可解释性

可视化,如图 4 所示。

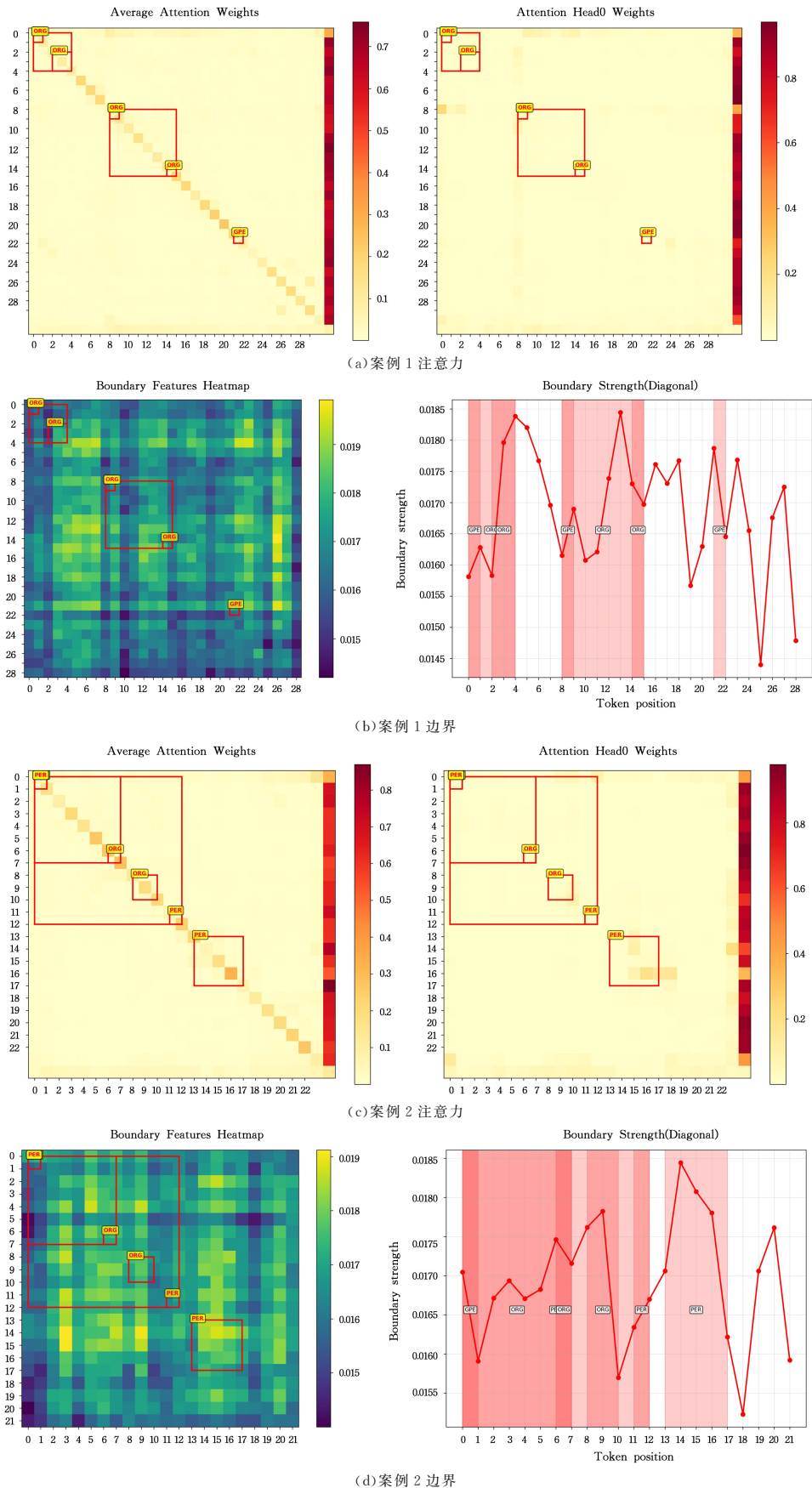


图 4 嵌套实体可解释性可视化

Fig. 4 Visualization of nested entity interpretability

如图 4(a) 所示,模型对完整实体“中国司法部”(位置 0-5)及其内部子实体“中国”(0-2)和“司法部”(2-5)均赋

予高权重(峰值大于等于 0.8),体现了 AAM 模块的分层关注能力;图 4(b)的对角线峰值与实体起止位置一致(最大边界强度为 0.019),表明 BEM 模块有效增强了边界信号;图 4(c)和图 4(d)进一步给出“日本在野第一大党民族党代表”这一 4 层嵌套结构的分析结果,外层实体与所有内层子实体均出现显著响应(注意力权重大于等于 0.7,边界强度大于等于 0.018),且无漏检或误合现象,验证了模型在复杂中文场景下的可解释性。

3.12 案例分析

为进一步量化评估模型在真实样本上的性能优势,直观体现 AAM 与 BEM 对中文嵌套实体边界识别的修正效果,本节选取两个典型嵌套案例进行深入分析,结果如表 10、表 11 所列。

表 10 案例 1 分析
Table 10 Study of case 1

文献	识别结果
Gold label	GPE: 中国(0,2);ORG: 中国司法部(0,5);ORG: 司法部(2,5);GPE: 乡镇(8,10);ORG: 乡镇司法调解中心(8,16);ORG: 中心(14,16);GPE: 中国(21,23)
文献[16]	GPE: 中国(0,2);ORG: 司法部(2,5);GPE: 乡镇(8,10);ORG: 中心(14,16);GPE: 中国(21,23)
Ours	GPE: 中国(0,2);ORG: 中国司法部(0,5);ORG: 司法部(2,5);GPE: 乡镇(8,10);ORG: 乡镇司法调解中心(8,16);ORG: 中心(14,16);GPE: 中国(21,23)

表 11 案例 2 分析
Table 11 Study of case 2

文献	识别结果
Gold label	ORG: 日本在野第一大党(0,8);GPE: 日本(0,2);PER: 日本在野第一大党民族党代表(0,13);ORG: 大党(6,8);ORG: 民族党(8,11);PER: 代表(11,13);PER: 赵三尤寄夫(13,18)
文献[16]	ORG: 日本在野第一大党(0,8);GPE: 日本(0,2);ORG: 大党(6,8);ORG: 民族党(8,11);PER: 代表(11,13);PER: 赵三尤寄夫(13,18)
Ours	ORG: 日本在野第一大党(0,8);GPE: 日本(0,2);PER: 日本在野第一大党民族党代表(0,13);ORG: 大党(6,8);ORG: 民族党(8,11);PER: 代表(11,13);PER: 赵三尤寄夫(13,18)

案例 1 中,基线模型未能完整识别“中国司法部”与“乡镇司法调解中心”两个嵌套实体,仅抽取其内部子实体(如“司法部”“中心”)。这表明基线模型在实体边界共享(如“院”“司法”等字)的复杂场景下,难以感知完整的实体语义跨度。案例 2 中,基线模型未能识别最外层的 PER 实体“日本在野第一大党民族党代表”,仅将其拆解为多个 ORG 与 GPE 实体,表明其在处理深层嵌套与长距离语义依赖时存在局限。而本文模型可以完整识别所有嵌套实体并分类。综合分析表明,AAM 与 BEM 的协同设计使模型能够同时处理中文嵌套实体的边界模糊与结构层次问题,在复杂场景下表现出更强的鲁棒性。

结束语 针对中文命名实体识别任务中实体边界模糊和多尺度特征利用不足的挑战,本文提出了一个基于自适应注意力与边界增强的中文 NER 模型。该模型通过双路径卷积结构分别捕捉中文实体的局部细节与全局上下文信息,并结合空间-通道注意力机制动态聚焦关键语义特征;同时,设计

的边界增强模块利用特征金字塔与深度卷积,强化实体边界感知。在 ACE2005-Chinese 和 Cnerta 等中文嵌套数据集以及 Weibo,Ontonotes,Resume 等中文非嵌套数据集上的实验结果表明,本文模型在多场景下的 F1 值均优于现有主流方法,展现了其卓越的准确度、鲁棒性与中文泛化能力。未来工作将聚焦于模型在中文垂直领域的应用,如“医疗”领域 NER(采用 BC5CDR 等数据集)和“金融”领域 NER(如股票名称识别),并通过引入句法特征与多模态信息,为实体边界的判定提供更丰富的信号,以期进一步提升模型在复杂场景下的鲁棒性与可解释性。

参考文献

- [1] QUAN Y Q,ZHANG H T,YANG B,et al. A review of named entity recognition based on deep learning[J]. *Microelectronics & Computer*,2026,43(2):8-21.
- [2] PENG B,LI Y D,GONG X F,et al. A method of entity relation extraction based on heterogeneous graph neural network and text semantic enhancement [J]. *Computer Science*,2024,51(S1):268-272.
- [3] WANG X,XU Y,HE X,et al. Reinforced negative sampling over knowledge graph for recommendation[C]// *Proceedings of The Web Conference 2020*. 2020:99-109.
- [4] KHALID N,MOULAY A. Transformer models used for text-based question answering systems[J]. *Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies*,2023,53(9):10602-10635.
- [5] LI Y H,SONG L,ZHANG C. Sparse conditional hidden Markov model for weakly supervised named entity recognition[C]// *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining(KDD'22)*. New York:ACM,2022:978-988.
- [6] PANCHENDRARAJAN R,AMARESAN A. Bidirectional LSTM-CRF for named entity recognition[C]// *The 32nd Pacific Asia Conference on Language, Information and Computation (PALCLIC 32)*. 2019.
- [7] DUAN J Y,ZHU Y F,WANG H,et al. Chinese nested named entity recognition based on position embedding and multi-level prediction[J]. *Computer Engineering*,2023,49(12):71-77.
- [8] JIA L R Z,LIU S Q,LIU Y,et al. Chinese nested named entity recognition based on hierarchical ERNIE model[J]. *Journal of Northeast Normal University(Natural Science Edition)*,2023,55(1):97-103.
- [9] LI X,ZHANG J W. Chinese nested named entity recognition based on masked SwinTransformer and boundary smoothing [J]. *Journal of North China University of Technology*,2024,36(5):39-48.
- [10] WANG X H,XU Y B. Chinese named entity recognition algorithm with soft attention mask embedding[J]. *Journal of Jilin University(Engineering and Technology Edition)*,2026,56(1):231-238.
- [11] LU W,ROTH D. Joint mention extraction and classification

- with mention hypergraphs[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015;857-867.
- [12] MUIS A O, LU W. Learning to recognize discontinuous entities-supplementary[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016;75-84.
- [13] WANG B, LU W. Neural segmental hypergraphs for overlapping mention recognition[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. ACL, 2018; 204-214.
- [14] HUANG R, CHEN Y, HUANG R. A Levitated Controlled Attention for Named Entity Recognition[J]. Cognitive Computation, 2025, 17(1):1-15.
- [15] HE A K, CHEN Y P, HU Y, et al. A named entity recognition method fusing boundary interaction information[J]. Journal of Guangxi Normal University (Natural Science Edition), 2025, 43(3):1-11.
- [16] HU J L, CHEN Y P, QIN Y B, et al. A Chinese named entity recognition method based on boundary mask[J/OL]. Computer Engineering, 1-12 [2026-03-16]. <https://doi.org/10.19678/j.issn.1000-3428.0070482>.
- [17] ZHU P, CHENG D, YANG F, et al. Improving Chinese named entity recognition by large-scale syntactic dependency graph[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2022, 30:979-991.
- [18] LIU Y, HUANG S B, LI R S, et al. USAF: Multimodal Chinese named entity recognition using synthesized acoustic features [J]. Information Processing & Management, 2023, 60(3): 103290.
- [19] ZHU H H, ZHANG K, LIU Y D, et al. A Chinese Named Entity Recognition Model Fusing Part-of-Speech Features[J]. Computer and Digital Engineering, 2025, 53(6):1669-1674, 1703.
- [20] ZHANG B H, CAI J H, ZHANG H P, et al. VisPhone: Chinese named entity recognition model enhanced by visual and phonetic features [J]. Information Processing & Management, 2023, 60(3):103314.
- [21] YANG X F, FAN Y, LI Z Q, et al. Chinese Named Entity Recognition Model Integrating Multi-stage Features [J]. Computer Engineering and Design, 2025, 46(1):37-43.
- [22] ZHANG H, QIN D H, BAI F B, et al. Chinese Named Entity Recognition Integrating Multi-level Chinese Character Features and Text Local Features [J]. Journal of Chinese Information Processing, 2024, 38(9):93-107.
- [23] LIU R, GUO X, ZHU H, et al. A text-speech multimodal Chinese named entity recognition model for crop diseases and pests [J]. Scientific Reports, 2025, 15:5429.
- [24] GENG R, CHEN Y, HUANG R, et al. Planarized sentence representation for nested named entity recognition[J]. Information Processing & Management, 2023, 60:103352.
- [25] YAN H, SUN Y, LI X, et al. An embarrassingly easy but strong baseline for nested named entity recognition [C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. ACL, 2023;1442-1452.
- [26] GUO X, CHEN Y, TANG R, et al. Camouflaged named entity recognition in 2D sentence representation [J]. Expert Systems with Applications, 2024, 257:125096.



TANG Ruixue, born in 1987, Ph.D, is a member of CCF (No. P2389M). Her main research interests include natural language processing, digital media forensics and multimedia signal processing.

(责任编辑:柯颖)