计 算 机 研 究 与 发 展
Journal of Computer Research and Development
https://crad.ict.ac.cn
62(12): 2983−2992, 2025

基于多尺度边框融合的实体语义增强方法

吴 灿 陈艳平 扈 应 黄瑞章 秦永彬
(文本计算与认知智能教育部工程研究中心(贵州大学) 贵阳 550025)
(公共大数据国家重点实验室(贵州大学) 贵阳 550025)
(贵州大学计算机科学与技术学院 贵阳 550025)
(ypench@gmail.com)

Entity Semantic Enhancement Method via Multi-Scale Box Fusion

Wu Can, Chen Yanping, Hu Ying, Huang Ruizhang, and Qin Yongbin
(*Text Computing & Cognitive Intelligence Engineering Research Center* (*Guizhou University*), *Ministry of Education*, *Guiyang* 550025)
(*State Key Laboratory of Public Big Data* (*Guizhou University*), *Guiyang* 550025)
(*College of Computer Science and Technology*, *Guizhou University*, *Guiyang* 550025)

Abstract Named entity recognition (NER) is a traditional task in natural language processing. The mainstream approach for nested named entity recognition is the span-based classification method, which involves concatenating representations of entity boundaries to obtain spans in general. However, long entities tend to lead to weakened semantic associations between two entity boundaries. In addition, single-scale spans cannot completely capture how entities behave in different contexts. To address this, an entity semantic enhancement method based on multi-scale box fusion is proposed. The method represents spans as boxes with boundary position information. At first, multi-scale boxes will be obtained by fusing features of different scales to enhance the semantic features in the boxes and make the boxes more context-dependent. Then, the boundary positions of the boxes are further refined by a position-weighted attention mechanism to make the box information more accurate. Finally, the entity category of the boxes and the offset of the boxes relative to the true entities are simultaneously predicted to effectively support the recognition and localization of nested named entities. The method achieves a new state-of-the-art *F*1 scores of 88.63% on the ACE2004 English dataset, 88.53% on ACE2005 English dataset and 73.86% on Weibo Chinese dataset, which demonstrates the effectiveness of the model.

Key words named entity recognition; span classification; multi-scale boxes fusion; entity semantic enhancement; position-weighted attention mechanism

摘 要 命名实体识别(named entity recognition, NER)是自然语言处理中的一项传统任务. 基于跨度分类的方法是用来解决嵌套命名实体识别的主流方法. 该方法通常是拼接实体边界的表示来获得跨度. 然而,长实体容易导致2个实体边界之间的语义关联被弱化,并且单一尺度的跨度无法完整地捕捉实体在不同上下文中的表现. 对此提出了一种基于多尺度边框融合的实体语义增强方法. 该方法将跨度表示为带有边界位置信息的边框. 首先,将通过融合不同尺度实体特征得到多尺度边框以增强边框中的语义特征,使边框的上下文依赖性更强. 然后,通过基于位置权重的注意力机制进一步细化边框的边界位置使得边框信息更准确. 最后,同时预测边框的实体类别和相对于真实实体的位置偏移量,有效支持嵌套命名实体的

收稿日期: 2024-08-13; **修回日期**: 2025-04-07
基金项目: 贵州省重大科技专项项目([2024]003); 国家自然科学基金项目(62066007, 62066008); 国家重点研发计划项目(2023YFC3304500)
This work was supported by the Major Science and Technology Project of Guizhou Province ([2024]003), the National Natural Science Foundation of China (62066007,62066008), and the National Key Research and Development Program of China (2023YFC3304500).
通信作者: 陈艳平(ypench@gmail.com)

识别和定位. 该方法在 ACE2004 英文数据集、ACE2005 英文数据集和 Weibo 中文数据集上分别取得了 88.63%, 88.53%, 73.86% 的 $F1$ 值, 证明了模型的有效性.

关键词 命名实体识别; 跨度分类; 多尺度边框融合; 实体语义增强; 位置权重注意力机制

中图法分类号 TP391.1

DOI: 10.7544/issn1000-1239.202440665 **CSTR:** 32373.14.issn1000-1239.202440665

命名实体识别(named entity recognition, NER)[1-2]是指从给定文本中识别特定类别的实体信息, 例如人名、机构名和地点等. 命名实体识别作为自然语言处理(natural language processing, NLP)的核心任务之一, 广泛应用于理解文本中的实体和关系. 在关系抽取[3]、信息检索[4]、实体链接[5]等许多下游任务中发挥了至关重要的作用.

目前, 基于深度学习的命名实体识别的方法可以分为基于序列、基于超图和基于跨度的方法. 基于序列标注的方法主要对句中的每个元素进行标签标记, 以指示其在组成命名实体中的角色. 例如, 在“B-I-O”编码中, “B”“I”“O”分别表示命名实体的“开始”“内部”“外部”. 由于每个单词只能分配一个标签, 只能解决非嵌套结构, 所以该方法的主要问题是很难识别嵌套的命名实体.

基于超图的方法设计了专有结构来处理嵌套命名实体识别困难的问题[6]. 这种方法允许 1 条边连接到多个节点, 以代表不同的实体, 使用神经网络对这些边和节点进行编码, 最后从超图标签中恢复实体. 尽管这些方法解决了嵌套结构, 但不可避免地需要部署复杂的转换过程和昂贵的解码过程, 会引入额外的错误.

基于跨度的方法是支持嵌套命名实体识别的常用方法. 该方法通常通过边界拼接从一个句子中生成实体跨度, 然后为每个可能的跨度输出一个类标签. 该方法通过枚举句子中的所有候选跨度并对其进行分类来识别实体. 具有解决嵌套结构和充分利用跨度中字符特性的优点. 虽然枚举在理论上是可行的, 但是高计算复杂度仍是需要解决的问题. 其次, 这些方法主要集中在学习跨度表示. 没有实体边界的监督[7], 边界可能不准确. 此外, 这些方法在面对较长实体的时候未能有效利用跨度间的特征, 从而导致语义被弱化.

一些研究[8]已经注意到目标检测与命名实体识别任务的形式一致性. 通过对候选实体边界位置进行回归计算可以更好地利用监督信息, 也有效解决了实体嵌套问题. 但边界回归模型[9]需要预先设置固定长度和数量的密集文本边界框(简称边框), 用于定位实体并进行回归运算. 该方法通过单一尺度映射边框, 无法考虑到所有边框情况. 并且边框的确定需要考虑上下文信息, 特别是存在歧义或复杂语境的情况下. 单一尺度的边框可能无法完整地捕捉实体在不同上下文中的表现. 这限制了实体识别任务对所有尺度实体的准确识别.

针对上诉问题, 本文提出了一种基于多尺度边框融合的实体语义增强方法. 将实体跨度表征为带有边界位置信息的边框, 通过结合多尺度句子表示, 生成多尺度的边框. 这一策略有效地融合了不同尺度的实体特征, 从而提升了边框内部的语义信息和表征能力. 为了得到语义完整的文本边界框, 融合不同尺度实体特征以得到多尺度边框, 使得边框表示语义更加完整. 这些边框不仅具有丰富的语义信息, 还具有高重叠的跨度和类别信息, 提升了边框的上下文依赖性. 而为了更好地利用实体边界的监督, 通过边界回归运算对边框进行微调. 同时使用基于位置权重的多头注意力机制进一步学习微调后的边框, 作为模型识别实体的监督信息使得边框的信息准确. 通过一系列实验证明, 本文方法在多个数据集上实现了显著的性能提升.

本文的贡献有 2 点:

1)提出一种基于多尺度边框融合的实体语义增强方法. 将跨度表示为带有边界位置信息的边框, 并融合边框内的多尺度实体特征以增强边框中的语义特征. 提高了模型对不同长度实体尤其是长实体的识别.

2)引入基于位置权重的多头注意力机制, 加强边界位置信息和实体分类之间的联系, 有效提升了性能.

1 相关工作

1.1 常用的实体识别方法

由于传统的序列标记的方法已经得到了很好的研究, 许多研究致力于将序列标记方法扩展到嵌套命名实体识别, 对序列模型进行了重新设计. 探索最

多的方法之一就是分层方法. 基于分层的方法是将嵌套结构转化为平面结构来预测实体类型. Wang 等人[10]将顺序深度神经网络堆叠成金字塔形状，用于嵌套命名实体识别. 此外，Li 等人[11]基于预定义规则将嵌套图元填充到边界提示模板中，将其与原始文本相结合也取得了较好的性能. 虽然这些方法已经取得了先进的性能，但由于需要人为设计标记方案，仍然不够灵活.

在解决嵌套问题上，基于跨度的方法通过对候选跨度进行分类，也引起了广泛的关注. 该方法的核心是枚举和验证句子中所有可能的实体跨度，具有解决嵌套结构和充分利用跨度中的字符特性的优点. 为此，Xu 等人[12]利用局部检测方法对每个可能的跨度进行分类. Sohrab 等人[13]提出了一种简单的嵌套 NER 深度神经模型，该模型列举了所有可能的跨度再对其进行分类. 此外，Tan 等人[7]除了对跨度进行分类外，还加入了一个额外的边界检测任务来预测实体边界.

1.2 目标检测技术在实体识别中的应用

基于跨度的方法在形式上类似计算机视觉中的目标检测任务. 又根据长期以来将图像与自然语言相关联的想法[14]，一些研究提出了一种 2 阶段标识符[8]. 该标识符充分利用了部分匹配的实体提议，并引入回归运算的过程. 同时也有人将边框提议和边框回归阶段纳入到一个端到端框架[9]中，以解决 2 阶段模型导致的错误传播问题. Yu 等人[15]还提出了一种支持全跨度命名实体识别的边界回归模型. 该模型采用回归操作来细化句子中实体跨度的空间位置，从而找到句子中所有可能的命名实体. 尽管将计算机视觉中已证明的方法迁移到命名实体识别任务的洞察力具有指导意义，但基于边框的方法需要人为设置锚框尺寸. 存在边框适应能力差导致的对长实体识别困难问题和单一尺度映射出来的边框对不同长度实体识别困难的问题. 在本文中，我们将采用多尺度边框融合的方法来增强边框的实体语义，并基于位置权重对边框的特征进行修正，以提高模型的识别性能.

2 方法介绍

本文模型的整体结构如图 1 所示，主要分为基本网络、多尺度边框融合、边界回归、注意力机制以及实体分类 5 个部分. 在基本网络部分，将单词或字符标记转化为上下文特征映射，从而捕捉句子的语义和句法信息. 在多尺度边框融合部分，得到边框以便后续进行分类的过程. 通过对文本边框的特征进行多尺度的特征融合，得到更为丰富语义信息的边框. 边界回归部分主要通过捕获外部语义来更新边框的位置. 然后基于位置权重的多头注意力机制，使得调整以后的位置信息可以作为监督信息辅助后续分类. 实体分类部分为类别判断层. 下面分别对模型各个部分进行详细分析.

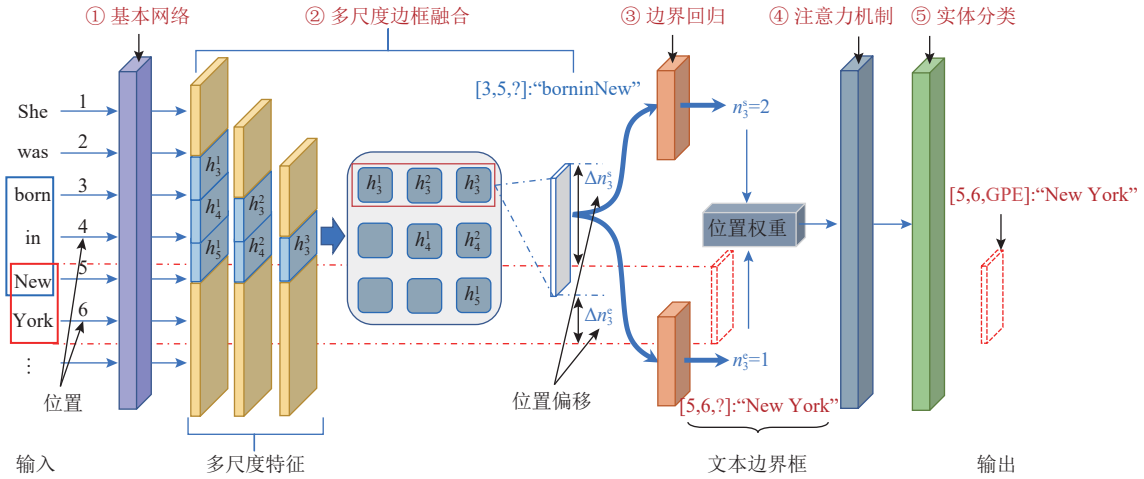


Fig. 1 Framework diagram of model

图 1 模型框架图

2.1 基本网络

基本网络是将句子抽象表示的核心部分，将单词或字符标记转化为上下文特征映射，从而捕捉句子的语义和句法信息. 设句子 $\boldsymbol{T}=(t_1,t_2,\cdots,t_i)$，$1\leqslant i\leqslant N$. 其中，$t_i$表示单词或者字符的标记，$N$ 表示是输入句子的长度. 为避免在实体边界远离任何真实实体边

界时, 找到这些边界变得困难. 在基本网络中使用多粒度句子表示, t_i经过以下方式得到抽象表示$\boldsymbol{h}_i$:

$$\boldsymbol{h}_i=[\boldsymbol{h}_i^{\text{char}};\boldsymbol{h}_i^{\text{BERT}};\boldsymbol{h}_i^{\text{word}};\boldsymbol{h}_i^{\text{pos}}], \tag{1}$$

字符级嵌入$\boldsymbol{h}_i^{\text{char}}$是将每个字符$t_i$嵌入为向量, 然后通过 Bi-LSTM(bidirectional long short-term memory)转换为高维度向量. $\boldsymbol{h}_i^{\text{BERT}}$是通过 BERT-Large[16] 网络转换为向量, 有效地编码单词的语义信息. 词级别嵌入$\boldsymbol{h}_i^{\text{word}}$是$t_i$经过 Pennington 等人[17] 的预训练词向量映射的. 为引入词性的语义信息, 每个t_i使用词性标记工具生成词性标记(part-of-speech tagging, POS), 然后映射成向量$\boldsymbol{h}_i^{\text{pos}}$. [;] 表示对几种表示进行拼接, 拼接得到标记t_i的抽象表示$\boldsymbol{h}_i$. 整个句子的抽象表示由所有字符抽象表示组成, 表示为$\boldsymbol{H}=[\boldsymbol{h}_1,\boldsymbol{h}_2,\cdots,\boldsymbol{h}_i]$.

2.2 多尺度边框融合

本阶段是为每个字符生成一个边框. 边框包含边框特征$\boldsymbol{h}'_j$以及边框位置$\boldsymbol{n}_j$, 边框位置包括开始位置和结束位置. 本节所有特征都是分别实现卷积核大小为 1, 2, 3 的卷积层从第 1 层直接生成的. 第 1 层使用的是基本网络的输出结果$\boldsymbol{H}$, 第 k 个特征层表示为

$$\boldsymbol{H}^k=CNN(\boldsymbol{H}^{k-1})=[\boldsymbol{h}_1^k,\boldsymbol{h}_2^k,\cdots,\boldsymbol{h}_{n-k+1}^k], \tag{2}$$

在卷积过程中, 生成了多尺度的多个特征映射层. 每个层级表示为$\boldsymbol{H}^k$. 这种多尺度句子表示方法有助于更好地捕捉句子中不同尺度的语义信息, 从而增强了命名实体识别的性能.

$$\begin{aligned}&\boldsymbol{S}_i^k=MLP(\boldsymbol{h}_i^k)=[\boldsymbol{S}_{i,\text{s}}^k,\boldsymbol{S}_{i,\text{e}}^k],\\&\boldsymbol{\delta}_{i,j}^k=\frac{\exp(\boldsymbol{S}_i^k)}{\sum\limits_{(i',k')\in\varphi_j}\exp(\boldsymbol{S}_{i'}^{k'})}.\end{aligned} \tag{3}$$

为提升命名实体识别任务在表示上面的合理性, 引入了多尺度边框融合机制. 如式(3)所示, 通过$\boldsymbol{h}_i^k$计算每个特征的得分向量$\boldsymbol{S}_i^k$. 其中$\boldsymbol{S}_{i,\text{s}}^k$和$\boldsymbol{S}_{i,\text{e}}^k$分别表示开始和结束位置的向量表示. 后续用来计算每个边框的权重向量. $\boldsymbol{\delta}_{i,j}^k$表示$t_i$的第 k 层特征映射层的第 j 个跨度的权重向量. 包含t_j特征所对应的边框的位置表示为$\varphi_j=\{(i,k)|j\in n_i^k\}$. 例如包含第 1 个字符特征所对应的位置是$\varphi_1=\{(1,1),(1,2),(1,3)\}$, 即融合这 3 个位置的向量生成新的融合边框.

$$\begin{aligned}&\boldsymbol{h}_{i,j}^k=\frac{\exp(\boldsymbol{h}_i^k)}{\sum\limits_{(i',k')\in\varphi_j}\exp(\boldsymbol{h}_{i'}^{k'})},\\&\boldsymbol{h}'_j=\sum_{(i,k)\in\varphi_j}\boldsymbol{h}_{i,j}^k,\end{aligned} \tag{4}$$

如式(4)所示, 为了得到更为丰富的语义信息, 计算得到的边框权重向量和边框的特征会被融合, 以产生全局的权重表示. 这种融合方式是将包含t_j的特征以及特征对应位置进行融合, 目的是综合考虑不同尺度的语义信息, 对边框起到语义信息增强作用. 然后利用权重向量$\boldsymbol{\delta}_{i,j}^k$对边框位置进行初始化, 如下所示:

$$\boldsymbol{n}_j=\sum_{(i,k)\in\varphi_j}\boldsymbol{\delta}_{i,j}^k\cdot p_i^k, \tag{5}$$

其中“$\cdot$”代表逐元素乘积, p_i^k代表对应于特征$\boldsymbol{h}_i^k$的位置, 使得初始边框的位置, 可以理解成是包含字符t_j的开始位置n_j^{s}和结束位置n_j^{e}. 这种限制是有助于模型理解并让后续边界回归方法进行合理的调整.

2.3 基于位置权重的多头自注意力机制

2.3.1 边界回归

很多边框表示虽然没有被模型识别为负样本, 但它们之间相互重叠, 共享相同的上下文, 从而导致了实体识别任务的边界不匹配问题. 为此设计了边界回归模块, 通过捕获外部语义来更新边框的位置, 可以有效提升跨度的准确率, 改善实体定位任务.

在本模块, 对边框开始位置和结束位置进行调整. 应用 *MLP* 来学习它们与最近的真实边框相关的位置偏移. 将边框边界的开始位置和结束位置对应的特征$\boldsymbol{h}'_j$输入边界回归模块, 计算边框边界相对于真实实体位置的偏移. 边框位置的更新公式为:

$$\begin{aligned}&\Delta\boldsymbol{n}_j=\boldsymbol{W}\boldsymbol{h}'_j+\boldsymbol{b}=[\Delta n_j^{\text{s}};\Delta n_j^{\text{e}}],\\&\boldsymbol{d}_j=\boldsymbol{n}_j+\Delta\boldsymbol{n}_j,\end{aligned} \tag{6}$$

其中 [;] 表示拼接操作, $\Delta\boldsymbol{n}_j$由跨度开始边界的偏移量Δn_j^{s}和跨度结束边界的偏移量Δn_j^{e}组成. $\boldsymbol{W}\in\mathbb{R}^{3D\times2}$, $\boldsymbol{b}\in\mathbb{R}^2$, 是可学习参数.

2.3.2 注意力机制

先前基于边界回归的实体识别模型在经边界回归调整后, 使用原本的边框特征得到的类别对数会导致模型错误传播. 而基于位置权重的多头注意力机制可以使得调整以后的位置信息融入到边框特征里. 为此我们通过类高斯分布为每个注意力头的每个边框生成一个位置权重, 类高斯分布的公式为:

$$G(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})=\exp((\boldsymbol{x}-\boldsymbol{\mu})^{\text{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})), \tag{7}$$

其中$\boldsymbol{x}\in\mathbb{R}^2$属于自变量; $\boldsymbol{\mu}$ 属于期望向量, 也即分布的中心; $\boldsymbol{\Sigma}^{-1}\in\mathbb{R}^{2\times2}$表示协方差矩阵的逆, 可知它是一个正定矩阵, 满足高斯分布的精度矩阵要求. 缺少的归一项将在后续 *softmax* 操作中补足. 生成一系列参数以完成高斯分布的构建, 方法为:

$$\begin{aligned}&\boldsymbol{\Sigma}_j=\boldsymbol{\eta}_i\boldsymbol{\eta}_i^{\text{T}},\\&\boldsymbol{\eta}_j=MLP(\boldsymbol{h}'_i).\end{aligned} \tag{8}$$

可逆矩阵$\boldsymbol{\Sigma}_j$使用$\boldsymbol{h}'_i$来构造. $\boldsymbol{h}'_j$, $\boldsymbol{n}_j$, $\boldsymbol{d}_j$, $\boldsymbol{\Sigma}_j$进行线性变换分别得到$\boldsymbol{h}'_{j,m}$, $\boldsymbol{n}_{j,m}$, $\boldsymbol{d}_{j,m}$, $\boldsymbol{\Sigma}_{j,m}$. 使用$\boldsymbol{n}_{j,m}$和$\boldsymbol{d}_{j,m}$的差值表示为边框位置和期望的差值. 故位置权重的结果为$\boldsymbol{w}_{i,m}$, 计算公式为:

$$\boldsymbol{w}_{j,m}=G(\boldsymbol{n}_{j,m}|\boldsymbol{d}_{j,m},\boldsymbol{\Sigma}_{j,m}). \tag{9}$$

基于位置感知的多头自注意力机制计算如式(10)所示.

$$\begin{aligned}\boldsymbol{A}&=[\boldsymbol{head}_1;\boldsymbol{head}_2;\cdots;\boldsymbol{head}_m]\boldsymbol{W}^{\mathrm{a}},\\ \boldsymbol{head}_m&=softmax\left(\frac{(\boldsymbol{h}^{\mathrm{q}}_{j,m})^{\mathrm{T}}\boldsymbol{h}^{\mathrm{k}}_{j,m}}{\sqrt{d_k}}+\log\boldsymbol{w}_{j,m}\right)\boldsymbol{h}^{\mathrm{v}}_{j,m},\end{aligned} \tag{10}$$

其中[;]表示拼接操作, $\boldsymbol{h}^{\mathrm{q}}_{j,m}$, $\boldsymbol{h}^{\mathrm{k}}_{j,m}$, $\boldsymbol{h}^{\mathrm{v}}_{j,m}$分别代表注意力机制中的查询向量, 键向量和值向量的初始化, 都是通过$\boldsymbol{h}'_j$进行线性变换得到的. $\boldsymbol{head}_m$表示第i个注意力头计算结果, m表示注意力头的数量. d_k是超参数. $\boldsymbol{W}^{\mathrm{a}}$是注意力头计算结果拼接后的权重矩阵. 注意力结果送入到下一层进行最后分类预测.

2.4 实体分类模块

基于位置的多头自注意力机制增加了边框的灵活性. 将融合边框位置的多头自注意力结果对边框特征$\boldsymbol{h}'_j$进行更新, 使得调整过的位置信息很好地监督后续的分类任务. 然后输入分类器计算该跨度为命名实体的概率$\boldsymbol{p}^{\mathrm{c}}_i$, 计算公式为:

$$\begin{aligned}\boldsymbol{h}^{\mathrm{a}}_j&=\boldsymbol{W}_1\boldsymbol{a}_j+\boldsymbol{b}_1,\\ \boldsymbol{h}^{*}_j&=\boldsymbol{h}'_j+\boldsymbol{h}^{\mathrm{a}}_j,\\ \boldsymbol{p}^{\mathrm{c}}_j&=softmax(MLP(\boldsymbol{h}^{*}_j)),\end{aligned} \tag{11}$$

其中$\boldsymbol{W}_1\in\mathbb{R}^{D_1\times D_2}$, $\boldsymbol{b}_1\in\mathbb{R}^{D_2}$都是可学习参数, $\boldsymbol{a}_j\in\boldsymbol{A}$表示前一阶段注意力机制的结果, D_1为注意力头的个数和$\boldsymbol{head}_m$维度的乘积, D_2表示边框特征的向量维度.

模型的输出结果为$\boldsymbol{d}_j$和$\boldsymbol{p}^{\mathrm{c}}_j$, 是一组包含实体边界位置和实体分类概率的$N$个预测, 其中$N$设置为明显大于文本中命名实体个数的值, 本模型的$N$值设置为句子长度大小. y表示对象的真实实体集合, N个预测的集合表示为$\hat{y}=\{\hat{y}_i\}_{i=1}^{N}$. 将$y$视为用$\varnothing$(无对象)填充到大小为$N$的集合. 为了找到这2个集合之间的二分匹配, 以最低成本搜索N个元素$\sigma\in\Psi_N$的排列:

$$\hat{\sigma}=\arg\min_{\sigma\in\Psi_N}\sum_{i}^{N}\Gamma_{\mathrm{match}}(y_i,\hat{y}_{\sigma(i)}), \tag{12}$$

其中$\Gamma_{\mathrm{match}}(y_i,\hat{y}_{\sigma(i)})$是真实值$y_i$和索引为$\sigma(i)$的预测之间的成对匹配成本. 根据先前的工作[18], 可以使用匈牙利算法有效地计算出这种最佳分配. 匹配成本综合考虑了类别预测及其边界预测. 真实实体集合的每个元素i可以看作是$y_i=(c_i,g_i)$, c_i表示实体标签(可能是$\varnothing$), g_i包括真实实体的起始位置g^{s}_i和结束位置g^{e}_i. 对于索引为$\sigma(i)$的预测, 将类别c_i概率定义为$\hat{\boldsymbol{p}}^{c_i}_{\sigma(i)}$, 预测的边框定义为$\boldsymbol{d}_{\sigma(i)}$. 匹配定义如下:

$$\Gamma_{\mathrm{match}}(y_i,\hat{y}_{\sigma(i)})=-\hat{\boldsymbol{p}}^{c_i}_{\sigma(i)}-\mathbb{I}_{\sigma(i)\neq\varnothing}(L_{\mathrm{boundry}}(g_i,\boldsymbol{d}_{\sigma(i)})). \tag{13}$$

边界损失L_{boundry}定义如下:

$$L_{\mathrm{boundry}}(g_i,\boldsymbol{d}_{\sigma(i)})=\frac{1}{N}\sum_{i}\sum_{h\in\{\mathrm{s},\mathrm{e}\}}Smooth_{L_1}(g^h_i,\boldsymbol{d}^h_{\sigma(i)}), \tag{14}$$

$Smooth_{L_1}$是一个稳健的L_1损失, 可量化g_i和$\boldsymbol{d}_{\sigma(i)}$之间的差异[19]. 损失函数的计算, 由类预测的负对数似然和边界损失线性组合:

$$L_{\mathrm{Hungarian}}(y,\hat{y})=\sum_{1}^{N}(-\log\hat{\boldsymbol{p}}^{c_i}_{\sigma(i)}-L_{\mathrm{boundry}}(g_i,\boldsymbol{d}_{\sigma(i)})). \tag{15}$$

3 实　　验

为了验证本文提出模型的效果, 本文在ACE2004和ACE2005数据集进行了实验. 并且选取其他基于深度学习的命名实体识别模型进行对比, 使用精确率(precision, P)、召回率(recall, R)和$F1$值($F1$ score)作为主要评估指标, 验证本模型的各项性能.

3.1 实验数据集

ACE2004和ACE2005语料库是从广播、新闻专线和博客中收集的. 本文使用英语语料库. 数据集定义7种实体类型: 交通工具(VEH)、地名(LOC)、武器(WEA)、地理政治实体(GPE)、人名(PER)、组织名(ORG)、基础设施(FAC). 我们采用了与Katiyar等人[20]和Lin等人[21]相同的设置, 以8∶1∶1的比例划分数据集进行训练、验证和测试.

Weibo数据集是从微博采样的中文数据集, 共有4种实体类型, 分别为人名(PRE)、地名(LOC)、机构名(ORG)和地理-政治(GEP), 每种类型又分为明确实体(NAM)、泛指实体(NOM)两种. 本文使用Li等人[22]相同的设置. 使用这个中文数据集来评估模型的跨语言能力, 数据集信息如表1所示.

Table 1 Dataset Statistics

表1 数据集统计

数据集	训练集	验证集	测试集	实体类型数
ACE2004	6 200	745	812	5
ACE2005	7 194	969	1 047	7
Weibo	1 350	270	270	4

3.2 评价指标

为了评估本文模型的性能, 我们使用了精确率、

召回率和 $F1$ 值. 在这种评估中，只有当模型正确识别了实体的位置和类型时，我们才认为结果是正确的. 具体来说，评估采用以下计算公式：

$$P=\frac{TP}{TP+FP}, \tag{16}$$

$$R=\frac{TP}{TP+FN}, \tag{17}$$

$$F1=\frac{2PR}{P+R}, \tag{18}$$

其中TP表示正确识别出的实体个数，FP表示实际上是假实体但是模型识别为真实体的个数，FN表示实际为真实实体但模型并未识别出的实体个数.

3.3 实验环境与参数设置

本文提出模型的实验在 Python3.10 和 Pytorch2.1 环境中进行. 训练中使用 BERT-Large 作为数据集的预训练模型，采用 Adaw 优化器，参数设置如表 2 所示.

Table 2 Parameter Setting

表 2 参数设置

模型参数	取值
CNN 卷积核大小	(1，2，3)
多头注意力头数	4
Dropout	0.5
学习率	0.000 4
Batch size	20
Epoch	40

3.4 实验结果分析

为了验证本文模型的有效性，将本文模型与近年提出的模型进行对比. 不同模型在 ACE2004 和 ACE2005 数据集上的性能如表 3 所示. 所列性能均来自相应论文的实验结果.

从表 3 可以看出，本文模型在 ACE2004，ACE2005，Weibo 数据集上与其他基线模型相比精确率和 $F1$ 值都达到了最优，召回率略低于 W2Ner 但是和最新的实体识别模型(DiffusionNER)相比提升了 1.08 个百分点.

Table 3 Experimental Results Comparison of Datasets

表 3 数据集实验结果对比 %

模型	ACE2004			ACE2005			Weibo		
	P	*R*	*F*1	*P*	*R*	*F*1	*P*	*R*	*F*1
Layered[23]	87.01	86.55	86.78	74.20	70.30	72.20			
W2Ner[24]	87.33	87.71	87.52	85.03	**88.62**	86.79	70.84	**73.87**	72.32
BS[25]	88.43	87.53	87.98	86.25	88.07	87.15			
Full-span[16]				87.23	87.46	87.35			
Locate-and-Label[8]	87.44	87.38	87.41	86.09	87.27	86.67	70.11	68.12	69.16
DiffusionNER[26]	88.11	**88.66**	88.39	86.15	87.72	86.93			
本文	**88.87**	88.40	**88.63**	**88.46**	88.60	**88.53**	**75.19**	72.59	**73.86**

注：黑体数值表示最优结果.

与 Layered 相比，通过动态叠加平面命名实体识别层来识别嵌套实体. 本文模型减少了错误传播，并提供了更为丰富的特征表示，增强了性能.

与 Locate-and-Label 相比，虽然 Locate-and-Label 提出的通过文本边框采样的 2 阶段模型可以先定位实体边界再进行实体分类任务，并细化不准确的实体边界，但在实体类型特征表示上并未充分利用边界信息. 而本文通过引入位置权重使得调整后的位置信息能够有效地融入到边框特征中，从而提高了模型的识别精度.

Full-span 是支持全跨度命名实体识别的边界回归模型. 用回归操作细化句子中实体跨度的空间位置，找到句子中所有可能的命名实体. 该模型在单一的尺度特征上对实体位置进行调整，无法完整地捕捉实体在不同上下文中的表现. 本文模型采用的多尺度边框融合的方法很好地解决了该问题. 综合考虑不同尺度的语义信息，对边框起到语义信息增强作用. BS 受标签平滑的启发，提出了边界平滑作为基于跨度的神经命名实体识别模型的正则化技术. 但通过拼接实体边界的表示来获得跨度. 在面对长实体时，容易导致 2 个实体边界之间的语义关联被弱化. 而基于位置权重的多头自注意力机制可以有效地结合边界位置信息来提升模型识别实体尤其是长实体的能力.

DiffusionNER 是最新提出的一种边界扩散方法，通过模拟边界标签的扩散过程来提升命名实体识别的精度. 该方法在捕捉边界信息上有其独特的优势. 然而，在面对复杂的嵌套实体和长实体时仍存在一定的局限性. 与其相比，本文模型通过多尺度边框融合和位置权重机制，能够更好地捕捉不同尺度和复杂结构的实体信息，从而在多种数据集上表现出色.

此外，在 Weibo 数据集的对比实验中可知，本文模型的表现不局限于英文数据集，在中文数据集上依然表现出色，表明了该模型具有出色的跨语言能力. 更重要的是，由于 Weibo 数据集是一个少样本数据集，该数据集的成功证明了训练本文模型的迁移能力强，本文模型在 2 个英文数据集和 1 个中文数据集的实验中展现出显著的性能提升，从而验证了其有效性. 通过融合边框不同尺度的特征表示，模型能够合理生成可能存在的候选实体的左右边界，使得

边框的边界位置表示更为合理. 这种多尺度的特征融合不仅有助于捕获不同尺度的特征，还能有效支持后续的边界回归操作，使其在面对嵌套语义问题时更具优势. 此外，本文模型还通过引入位置权重，使模型在分类时能够更加关注实体的边界细节信息. 这一改进在细化实体边界的基础上，进一步提高了实体分类的效果.

3.5 消融实验

为了验证模型各组成部分的有效性，本节设计了以下消融实验. 从表 4 的结果可以看出，本文提出的完整模型在 $F1$ 值上显著优于去掉任何一个模块的模型. 说明设计的 2 个模块（MB 模块和 PA 模块）对命名实体识别任务都有正向的贡献，并且两者协同工作有效提升了模型的整体性能.

Table 4 Ablation Experiment Results of Datasets

表 4 数据集消融实验结果 %

模型	ACE2004			ACE2005			Weibo		
	P	*R*	*F*1	*P*	*R*	*F*1	*P*	*R*	*F*1
w/o MB	87.62	87.38	87.78	87.33	87.71	87.52	73.48	71.85	72.65
w/o PA	87.28	88.00	87.64	88.43	87.53	87.98	74.55	71.60	73.04
w/o MB+PA	87.84	86.65	87.24	86.27	88.40	87.32	74.03	71.11	72.54
本文	88.87	88.40	88.63	88.46	88.60	88.53	75.19	72.59	73.86

注：w/o MB 表示模型去掉不同尺度的特征；w/o PA 表示模型中去掉基于位置权重的多头注意力机制；w/o MB + PA 表示同时去掉不同尺度和基于位置权重的多头注意力机制 2 个模块.

其中多尺度边框融合模块（MB）能够生成高质量的文本边框，并通过融合边框内部的多尺度实体特征，增强了边框内特征的语义完整性. 这一特性在提升模型召回率方面尤为突出. 以 ACE2005 数据集为例，去掉 MB 模块后，模型的 $F1$ 值下降了 0.56 个百分点，说明 MB 模块对长实体和复杂实体的识别起到了关键作用. 而基于位置权重的多头注意力机制（PA）通过利用位置权重机制，精准捕捉与边框相关的特征，显著增强了边框的语义表示能力. 在提升模型的精确度上表现出色. 而去掉 PA 模块后，模型的 $F1$ 值下降了 1.11 个百分点，表明 PA 模块在准确定位实体边界和区分相似实体上具有重要作用. 当这 2 个模块同时被移除时，模型的 $F1$ 值下降了 1.31 个百分点，表明 MB 模块和 PA 模块在命名实体识别任务中是相互协同的.

具体来说，MB 模块提供了高质量的边框基础，而 PA 模块在此基础上进一步优化边框语义的准确性和完整性. 这种协同作用使得模型在处理复杂嵌套实体和长实体时展现了更强的鲁棒性. 其他数据集的实验结果也显示出类似的趋势，进一步验证了 2 个模块在不同领域数据中的通用性和有效性. 同时，MB 模块对模型性能的影响略大，这是因为在生成边框和捕捉长距离依赖方面对模型性能影响更大. 然而，PA 模块同样不可或缺，其在优化边框语义表示和边界精确度方面起到了关键补充作用.

3.6 不同实体长度的分析

本文提出的 MB 通过整合不同粒度的语义特征，使模型能够同时捕捉局部特征以精准识别短实体，并通过全局特征建模复杂的长实体语义依赖；而 PA 则有效提升了边界预测的精确性，尤其在处理嵌套实体时表现出色. 为了验证这 2 个模块对不同长度命名实体的影响，我们将本文模型与 Full-span 模型在 ACE2005 数据集上进行了对比实验，其结果如图 2 所示.

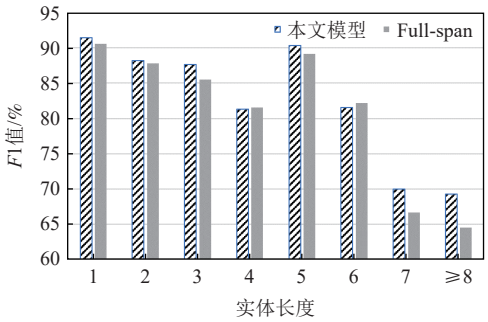


Fig. 2 Performance of named entities with different lengths

图 2 不同长度的命名实体性能

从图 2 中可以看出，随着命名实体长度的增加，模型性能大体上呈逐渐下降趋势. Full-span 在长实体上的性能下降尤为显著，而本文模型在所有长度范围内均表现出稳定优势. 具体而言：

1）短实体（长度为 1~3）. 本文模型和 Full-span 均能达到较高的 $F1$ 值. 但本文模型因 PA 模块能够捕捉边界相关特征，略高于对比模型，展现了边界精确度的优势.

2）中等长度实体（长度为 4~7）. 本文模型也展现出较为不错的性能优势. 这归因于 MB 模块能够有效融合多尺度上下文特征，增强语义完整性，降低了对中等长度实体识别的误差率.

3）长实体（长度大于等于 8）. 本文模型通过 MB 模块捕捉长距离依赖特征，结合 PA 模块对边界位置的精确建模，缓解了对比模型因梯度消失导致的性能下降，表现出更高的鲁棒性.

实验结果表明，本文提出的多尺度边框融合方

法和位置权重多头注意力机制能够有效提升模型在不同长度命名实体上的性能表现，尤其在长实体的识别中展现了显著的优势. 这验证了多尺度特征融合和注意力机制在应对长距离语义依赖和边界精确识别方面的有效性.

3.7 案例分析

为了进一步分析本文模型中 MB 模块和 PA 模块的作用，本文通过案例分析将本文方法与 Full-span 的方法进行对比. 我们从 ACE2005 数据集中选取了 2 句例句. 例 1 和例 2 都包含嵌套的命名实体，例 1 还包含一个很长的命名实体. 这 2 句分别输入到本文模型和 Full-span 中，输出如图 3 所示.

例 1	[1 many members on that tape]$^{1}_{\mathrm{PER}}$ working together through [2 a cell in [3 Germany]$^{3}_{\mathrm{GPE}}$ known as [4the hamburg cell]$^{2}_{\mathrm{ORG}}$]$^{4}_{\mathrm{ORG}}$seen in [5 the same room]$^{5}_{\mathrm{FAC}}$at the same time.
本文模型	[many members on that tape, PER]，[a cell in Germany known as the hamburg cell, ORG]，[Germany, GPE]，[the hamburg cell, ORG]，[the same room, FAC]
Full-span	[members on that tape, PER]，[Germany, GPE]，[the hamburg cell, ORG]，[the same room, FAC]
例 2	[1Anwar]$^{1}_{\mathrm{ORG}}$is expected to appeal again with [2 the Federal Coure]$^{2}_{\mathrm{ORG}}$,[4[3 Malaysia]$^{3}_{\mathrm{ORG}}$'s highest court]$^{4}_{\mathrm{ORG}}$.
本文模型	[Anwar, PER]，[the Federal Coure, ORG]，[Malaysia, GPE]，[Malaysia's highest court, ORG]
Full-span	[Anwar, PER]，[with the Federal Coure, ORG]，[Malaysia, GPE]，[highest court, ORG]

注：方括号里面的内容表示标注的命名实体，并以上标指示实体类型.

Fig. 3 Case analysis

图 3 案例分析

例 1 中包含 3 个嵌套实体，且包含一个长命名实体. 本文模型准确识别了所有嵌套实体和长实体，而 Full-span 因梯度消失问题，无法捕捉最长命名实体的边界. 而本文模型使用的多尺度边框融合方法，能够捕捉到不同长度和复杂度的实体. 通过在不同尺度上融合上下文信息，能够更全面地理解实体的语义并对长实体进行有效识别. 有效捕捉了长实体的全局语义特征，确保了“a cell in Germany known as the hamburg cell”这样的长实体能够被完整识别.

例 2 中“Malaysia's highest court”与“Malaysia”部分边界重叠. 本文模型能够准确识别这些嵌套实体并提供精确边界，而 Full-span 在边界处理上表现较弱. 基于位置权重的多头注意力机制通过将边界调整后的位置信息与语义特征相融合，显著提升了对嵌套实体的区分能力. 例如“Malaysia's highest court”与“Malaysia”共用部分边界，仍被模型正确区分. 该注意力机制还使得模型能够灵活关注边界调整后的特征，展现了更强的鲁棒性. Full-span 未能有效融合位置信息，导致“with the Federal Coure”与“highest court”等边界识别错误.

案例分析表明，本文模型在命名实体识别任务上至少有 2 个优势. 一是多尺度边框融合方法能够捕捉长距离语义依赖，并结合上下文信息全面建模长命名实体的语义特征，显著提升了长命名实体的识别性能. 二是本文模型通过基于位置权重的多头注意力机制，通过融合边界调整后的位置信息与语义特征，增强了对嵌套实体的区分能力，确保了复杂实体的边界和类别预测的准确性.

4 结 论

本文针对嵌套命名实体识别任务，提出了一种基于多尺度边框融合的实体语义增强方法. 通过同时预测边框的实体类别和相对真实实体的位置偏移量，模型利用多粒度句子表示生成包含边界位置信息的多尺度边框. 多尺度边框融合机制能够有效捕捉不同粒度的上下文特征，增强边框的语义完整性，从而显著提升对复杂语义和嵌套结构的建模能力. 同时，模型结合边界回归与基于位置权重的多头注意力机制，对边框进行进一步的微调和语义增强，有效提升了对长实体和嵌套实体的识别. 实验结果表明，本文方法在 3 个公共数据集上的表现均优于现有基线模型，充分验证了其有效性和鲁棒性.

尽管本文提出的方法在处理长实体和嵌套实体任务中展现了显著的性能优势，但仍然存在一些局限性需要进一步研究. 例如，在极度嵌套实体的场景中，嵌套层级的增加可能导致上下文特征之间的冲突，边界信息难以全面捕捉；而在超长实体的识别中，单一边框可能不足以完整建模其全局语义. 未来的研究可以从以下 3 个方面展开：

1)增强层次化建模能力. 通过引入层次化注意力机制或图神经网络，以更好地捕捉嵌套层级关系.

2)优化超长实体的建模. 设计更高效的边界回归机制，并尝试将超长实体分解为子实体进行建模，从而缓解长距离语义依赖问题.

3)结合数据增强和迁移学习.通过生成更多复杂结构实体的训练样本,提升模型在多领域任务中的泛化能力.

作者贡献声明:吴灿完成实验并撰写论文;陈艳平提出了算法思路和实验方案;扈应、黄瑞章、秦永彬提出指导意见并修改论文.

参 考 文 献

[1] Li Dongmei, Zhang Yang, Li Dongyuan, et al. Review of entity relation extraction methods[J]. Journal of Computer Research and Development, 2020, 57(7): 1424−1448 (in Chinese)
(李冬梅,张扬,李东远,等.实体关系抽取方法研究综述[J].计算机研究与发展,2020,57(7):1424−1448)

[2] Li Li, Xi Xuefeng, Sheng Shengli, et al. Research progress on named entity recognition in Chinese deep learning[J]. Computer Engineering and Applications, 2023, 59(24): 46−69 (in Chinese)
(李莉,奚雪峰,盛胜利,等.深度学习中文命名实体识别研究进展[J].计算机工程与应用,2023,59(24):46−69)

[3] Wang Qingbang, Zhang Qingchuan, Zuo Min, et al. A entity relation extraction model with enhanced position attention in food domain[J]. Neural Processing Letters, 2022, 54(2): 1449−1464

[4] Zhang Li, Zhang Shuo, Krisztian Balog. Table2vec: Neural Word and Entity Embeddings for Table Population and Retrieval[D]. Norway: University of Stavanger, 2018

[5] Chen Hui, Wei Baogang, Liu Yonghuai, et al. Bilinear joint learning of word and entity embeddings for entity linking[J]. Neurocomputing, 2018, 294: 12−18

[6] Huang Heyan, Lei Ming, Feng Chong. Hypergraph network model for nested entity mention recognition[J]. Neurocomputing, 2021, 423: 200−206

[7] Tan Chuanqi, Qiu Wei, Chen M, et al. Boundary enhanced neural span classification for nested named entity recognition[C]//Proc of the 34th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2020: 9016−9023

[8] Shen Yongliang, Ma Xinyin, Tan Zeqi, et al. Locate and label: A two-stage identifier for nested named entity recognition[C]//Proc of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int Joint Conf on Natural Language Processing (Volume 1: Long Papers). Stroudsburg, PA : ACL, 2021: 2782−2794

[9] Chen Yanping, Wu Lefei, Zheng Qinghua, et al. A boundary regression model for nested named entity recognition[J]. Cognitive Computation, 2023, 15(2): 534−551

[10] Wang Jue, Shou Lidan, Chen Ke, et al. Pyramid: A layered model for nested named entity recognition[C]//Proc of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2020: 5918−5928

[11] Li Zhun, Song Mei, Zhu Yi et al. Chinese nested named entity recognition based on boundary prompt[C]//Proc of the 20th Int Conf on Web Information Systems and Applications. Beilin: Springer, 2023: 331−343

[12] Xu Mingbin, Jiang Hui, Watcharawittayakul S. A local detection approach for named entity recognition and mention detection [C]//Proc of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg, PA: ACL, 2017: 1237−1247

[13] Sohrab M G, Miwa M. Deep exhaustive model for nested named entity recognition[C]//Proc of the 23rd Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2018: 2843−2849

[14] Pont-Tuset J, Uijlings J, Changpinyo S, et al. Connecting vision and language with localized narratives[C]//Proc of the 16th European Conf on Computer Vision. Berlin: Springer, 2020: 647−664

[15] Yu Junhui, Chen Yanping, Zheng Qinghua, et al. Full-span named entity recognition with boundary regression[J]. Connection Science, 2023: 2181483

[16] Devlin J, Chang Mingwei, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[C]//Proc of the 2019 Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long and Short Papers). Stroudsburg, PA: ACL, 2019: 4171−4186

[17] Pennington J, Socher R, Manning C D. Glove: Global vectors for word representation[C]//Proc of the 18th Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2014: 1532−1543

[18] Stewart R, Andriluka M, Ng A Y. End-to-end people detection in crowded scenes[C]//Proc of the 33rd IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 2325−2333

[19] Ren Shaoqing, He Kaiming, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]//Proc of the 28th Conf on Neural Information Processing Systems. Montreal, Canada: Curran Associates, 2015: 91−99

[20] Katiyar A, Cardie C. Nested named entity recognition revisited[C]//Proc of the 2018 Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL, 2018: 794−805

[21] Lin Hongyu, Lu Yaojie, Han Xianpei, et al. Sequence-to-nuggets: Nested entity mention detection via anchor-region networks [C]//Proc of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2019: 5182−5192

[22] Li Xiaonan, Yan Hang, Qiu Xipeng, et al. FLAT: Chinese NER using flat-lattice Transformer[C]//Proc of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2020: 6836−6842

[23] Ju Meizhi, Miwa M, Ananiadou S. A neural layered model for nested named entity recognition[C]//Proc of the 2018 Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, (Volume 1: Long Papers). Stroudsburg, PA: ACL, 2018: 1446−1459

[24] Li Jingye, Fei Hao, Liu Jiang, et al. Unified named entity recognition as word-word relation classification[C]//Proc of the 36th AAAI Conf

on Artificial Intelligence. Palo Alto, CA: AAAI, 2022: 10965−10973

[25] Zhu Enwei, Li Jinpeng. Boundary smoothing for named entity recognition[C]//Proc of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg, PA: ACL, 2022: 7096−7108

[26] Shen Yongliang, Song Kaitao, Tan Xu, et al. DiffusionNER: Boundary diffusion for named entity recognition[C]//Proc of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg, PA: ACL, 2023: 3875−3890



Wu Can, born in 1999. Master. His main research interests include natural language processing and information extraction.

吴 灿, 1999 年生. 硕士. 主要研究方向为自然语言处理、信息抽取.



Chen Yanping, born in 1980. PhD, professor. His main research interests include artificial intelligence and natural language processing.

陈艳平, 1980 年生. 博士, 教授. 主要研究方向为人工智能、自然语言处理.



Hu Ying, born in 1996. PhD. His main research interests include natural language processing and information extraction.

扈 应, 1996 年生. 博士. 主要研究方向为自然语言处理、信息抽取.



Huang Ruizhang, born in 1979. PhD, professor, PhD supervisor. Member of CCF. Her main research interests include natural language processing, data mining, and machine learning.

黄瑞章, 1979 年生. 博士, 教授, 博士生导师, CCF 会员. 主要研究方向为自然语言处理、数据挖掘、机器学习.



Qin Yongbin, born in 1980. PhD, professor, PhD supervisor. Senior member of CCF. His main research interests include intelligent computing, machine learning, and algorithm design.

秦永彬, 1980 年生. 博士, 教授, 博士生导师. CCF 高级会员. 主要研究方向为智能计算、机器学习、算法设计.