计算机科学
COMPUTER SCIENCE

基于分割注意力与边界感知的中文嵌套命名实体识别算法

张汝佳, 代璐, 郭鹏, 王邦

引用本文

张汝佳, 代璐, 郭鹏, 王邦.基于分割注意力与边界感知的中文嵌套命名实体识别算法[J]. 计算机科学, 2023, 50(1): 213-220.

ZHANG Rujia, DAI Lu, GUO Peng, WANG Bang. Chinese Nested Named Entity Recognition Algorithm Based on Segmentation Attention andBoundary-aware [J]. Computer Science, 2023, 50(1): 213-220.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.211100257

基于分割注意力与边界感知的中文嵌套命名实体识别算法

张汝佳　代　璐　郭　鹏　王　邦
华中科技大学电子信息与通信学院　武汉 430074
(m201971992@hust.edu.cn)

摘　要　由于中文文本缺少天然分隔符,中文嵌套命名实体识别(Chinese Nested Named Entity Recognition,CNNER)任务极具挑战性,而嵌套结构的复杂性和多变性更增添了任务的难度。文中针对CNNER任务提出了一种新型边界感知层叠神经网络模型(Boundary-aware Layered Nerual Model,BLNM)。首先通过构建了一个分割注意力网络来捕获潜在的分词信息和相邻字符之间的语义关系,以增强字符表示;然后通过动态堆叠扁平命名实体识别层的网络,由小粒度到大粒度逐层识别嵌套实体;最后为了利用被预测实体的边界信息和位置信息,构建了一个边界生成式模块,用于连接相邻的扁平命名实体识别层以及缓解错误传递问题。基于ACE 2005中文嵌套命名实体数据集的实验结果表明,该模型具有较好的性能。

关键词:中文嵌套命名实体识别;分割注意力;边界生成式;层叠神经网络
中图法分类号　TP391.1

Chinese Nested Named Entity Recognition Algorithm Based on Segmentation Attention and Boundary-aware

ZHANG Rujia,DAI Lu,GUO Peng and WANG Bang
School of Electronic Information and Communications,Huazhong University of Science and Technology,Wuhan 430074,China

Abstract　Chinese nested named entity recognition(CNNER) is a challenging task due to the absence of natural delimiters in Chinese and the complexity of the nested structure. In this paper,we propose a novel boundary-aware layered neural model(BLNM) with segmentation attention for the CNNER task. To exploit some semantic relation among adjacent characters,we first design a segmentation attention network to capture the potential word information and enhance character representation. Next,we model the nested structure with dynamically stacked Flat NER networks to detect entities in an inner to outer manner. We also design a boundary generative module to connect adjacent Flat NER layers,which can mark the boundary and position of detected entities and greatly alleviate the error propagation problem. Experiment results on ACE 2005 Chinese nested NE dataset show that the proposed model achieves superior performance than the state-of-the-art methods.

Keywords　Chinese nested named entity recognition,Segmentation attention,Boundary generative,Layered neural network

1　引言

命名实体识别(Named Entity Recognition,NER)是自然语言处理(Natural Language Processing,NLP)领域的一项基础任务,为实体链接[1-2]、关系抽取[3-4]、共指消解[5]等各种下游NLP任务提供了实体信息。命名实体识别任务的主要目的在于确定文本中命名实体的边界,并将实体分类到预先定义的类别中。在实际场景中还会出现有重叠结构的实体,即嵌套命名实体。如图1所示,中文文本"中华人民共和国国务院侨务办公室"属于二层嵌套结构,其包含一个大粒度实体"[中华人民共和国国务院侨务办公室]ORG"和两个小粒度实体"[中华人民共和国]GPE"和"[国务院]ORG"。目前大多数命名实体识别模型都能够较好地识别出结构相对简单的大粒度命名实体,但却很难完整、准确地识别出结构复杂的嵌套命名实体(Nested Named Entity)中的小粒度命名实体,导致无法在深层次文本理解中捕获更多层次粒度的语义信息。

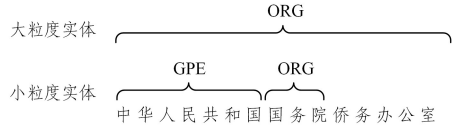

图1　中文嵌套命名实体示例图

Fig.1　Example of Chinese nested named entities

传统的嵌套命名实体识别(Nested Named Entity Recognition,NNER)方法[6-7]主要是基于规则的,但这些规则对

到稿日期:2021-11-25　返修日期:2022-06-17
基金项目:国家自然科学基金面上项目(62172167)
This work was supported by the National Natural Science Foundation of China(62172167).
通信作者:王邦(wangbang@hust.edu.cn)

领域知识的依赖性很强，通常需要领域专家和语言学者进行手动构造和修改，存在耗时耗力、灵活性差、可移植性差等问题。近年来，随着深度学习(Deep Learning)在众多 NLP 任务上的成功应用，很多学者开始尝试使用神经网络模型来识别嵌套实体。Ju 等[8]设计了一种层叠序列标注模型，通过对多个扁平命名实体识别(Flat Named Entity Recognition，Flat NER)层进行动态堆叠，实现了由内到外、由小粒度到大粒度逐层检测实体。但是，该方案中的层与层之间的连接方式存在局限性，容易出现误差传递，例如，当小粒度实体边界识别错误时，大粒度实体也很难被正确识别。尽管 Li 等[9]于 2020 年提出了一种多层联合学习模型，结合自注意力机制来提升实体聚合效果，但仍然没有解决误差传递问题。此外，现有的中文嵌套命名实体识别模型都是基于字符的，未利用相邻字符之间的语义信息，忽略了显式的词边界信息和词序信息。例如“键盘鼠标”和“仓鼠”中都包含字符“鼠”，但其携带的语义信息完全不一样，为其生成相同的字符嵌入表示是不恰当的。

针对上述问题，本文提出了一种边界感知层叠神经网络模型，该模型通过动态堆叠扁平 NER 层的方式，由小粒度到大粒度逐层地识别中文文本中的嵌套命名实体。一方面，本文设计了一个新型边界感知模块(Boundary Aware Module)，用于标记、传递识别出的实体的边界信息和位置信息，在不压缩句子长度的情况下大大缓解了误差传递问题。另一方面，本文通过结合 Jieba 分词工具与注意力机制，将词信息融入基于字符的模型中，有效利用了局部上下文语义信息。目前已在 ACE 2005 中文数据集上验证了 BLNM 模型的有效性，实验结果表明，该模型实现了最优性能。

本文第 2 节介绍了相关工作；第 3 节主要介绍了 BLNM 的模型结构；第 4 节评估了该模型在中文嵌套命名实体数据集上的性能，并对识别效果进行了可视化；最后总结全文并展望未来。

2 相关工作

当前解决命名实体识别的思路主要分为序列标注[10-12](Token-based)和基于分类[13](Span-based)两类。其中，序列标注方法最为常见。

Token-based 方法先将文本转化为特征序列并提取关键特征信息，然后对其进行标签序列预测。Huang 等[14]首次将 BiLSTM-CRF 模型用于 NER 任务，并发现 CRF 模块可以获取句子级别的标注信息，取得了不错的识别效果。Liu 等[15]提出了一种新型多任务序列标记模型 LM-LSTM-CRF，通过结合词级信息与字符级信息来提升命名实体识别性能。

然而，Token-based 模型往往存在着耗时、实体边界预测不准确等缺点，因此有学者提出了 Span-based 的方案来提升任务性能。该方案将潜在实体的所有可能区域或范围枚举出来，然后用深度神经网络对其进行分类。Xia 等[7]设计了 MGNER 模型架构，该模型由一个探测器和一个分类器组成，前者检测所有可能的实体片段，后者对候选实体进行分类。Luan 等[16]提出了一个通用框架 DyGIE，用于捕获实体片段间的交互关系并动态构建实体片段图(Span Graph)，实现实体信息的共享。

相比英文文本，中文文本没有空格作为显式边界标识符，难以直接确定词语边界。Dong 等[17]首次将字符级 BiLSTM-CRF 网络结构应用在中文命名实体识别任务上，并率先引入偏旁部首特征，通过拆解中文字符并对字符进行编码的方式来增强字符本身的特征，然后利用 BiLSTM-CRF 模型进行序列标注。Zhang 等[18]设计了 Lattice LSTM 结构，并首次将词典信息引入 CNER 任务中。Liu 等[19]提出了 WC-LSTM 模型，通过优化词典融合模式，不仅解决了 Lattice LSTM 训练时无法 batch 并行化的问题，大大提升了计算效率，还适用于各种应用场景。但 RNN 模型存在难以长期保持整体语义信息、梯度爆炸等缺点，因此有不少学者[20-23]尝试采用其他神经网络模型来解决这一问题。Gui 等[24]结合 CNN 结构与 Rethinking 机制，提出了 LR-CNN 模型，有效解决了 Lattice LSTM 模型无法有效处理词汇信息冲突的问题。Sui 等[25]率先将 GAT(Graph Attention Network)网络和自动构建的语义图引入 CNER 任务，不仅提升了命名实体识别的性能，而且大大降低了时间成本。近两年，也有一些学者提出了不同的思路[26-30]来提升中文扁平命名实体的识别效果。

上述工作只解决了中文扁平命名实体识别问题，而忽略了结构复杂的中文嵌套结构。传统的中文嵌套命名实体识别模型是基于规则和传统机器学习[31]的。Zhou 等[32]于 2004 年提出的方案和 Zhou 等[33]于 2006 年提出的方案都是先识别扁平命名实体，再通过基于规则的方法检测嵌套命名实体。Fu 等[34]通过抽取中文语素特征，提升了基于 CRF 框架的 CNNER 模型性能。近年来，随着深度学习的研究热潮袭来，开始有学者尝试将神经网络模型应用于 CNNER 任务中。Li 等[9]首次提出了一种多层联合学习模型，结合自注意力机制实现实体特征聚合，逐层识别嵌套实体。

3 边界感知层叠神经网络模型

本文提出了一种基于动态层叠融合边界特征信息的中文嵌套命名实体识别方法，以端到端的方式进行实体检测，模型的整体结构如图 2 所示。

具体来说，该模型主要由 3 个基础模块组成。1)分割注意力模块(Segmentation Attention Module，SAM)。该模块结合了中文分词(Chinese Word Segmentation，CWS)工具和注意力机制，将文本中每个中文字符转化为包含词信息的字符向量，并输入到第一层扁平命名实体识别层中。2)扁平命名实体识别模块(Flat NER Module)。每层 Flat NER 层由 BiLSTM 和 CRF 组成，前者进行文本序列的特征提取，后者对上下文标注进行约束，并输出序列标注结果。模块的主体结构由数个扁平命名实体识别层动态堆叠而成，通过由小粒度到大粒度的方式逐层识别实体。3)边界生成式模块(Boundary Generative Module，BGM)。作为 Flat NER 层之间的连接部分，BGM 模块利用边界编码(Boundary Encodings)和位置编码(Position Encodings)对上一层 Flat NER 层中被识别的实体的边界和位置进行标识，然后整合特征并传入下一层 Flat NER 层，直到 Flat NER 层不再识别出新的实体，模型停止堆叠，识别过程结束。

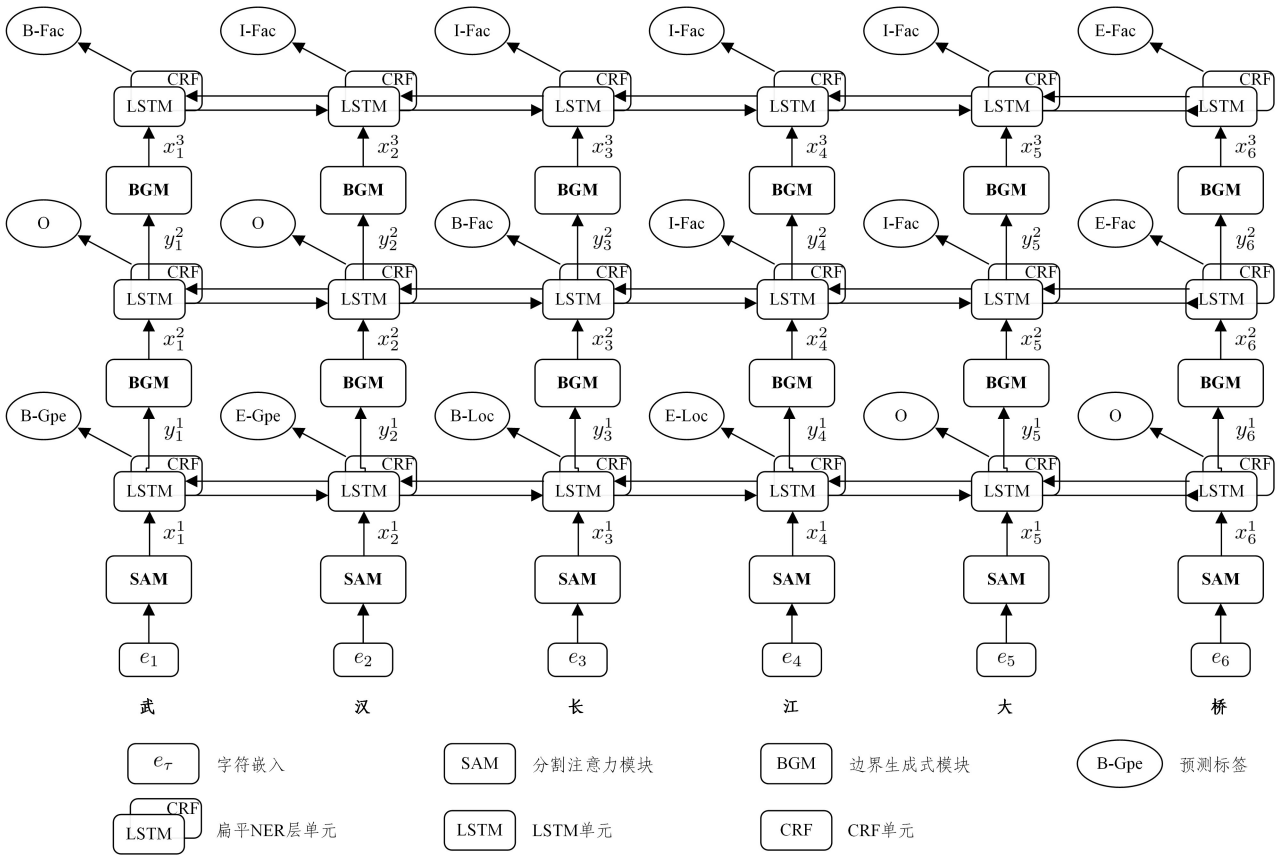

图 2 BLNM 模型的总体结构

Fig. 2 Overall structure of BLNM model

3.1 分割注意力模块

基于字符的中文嵌套命名实体识别模型无法捕获相邻字符之间的语义关系，忽略了显式的词边界信息和词序信息，为此，我们设计了分割注意力模块，如图 3 所示，通过聚合局部上下文中有语义关系的字符来增强字符表示，其中分词信息被用作 NER 任务的软特征(Soft Features)。

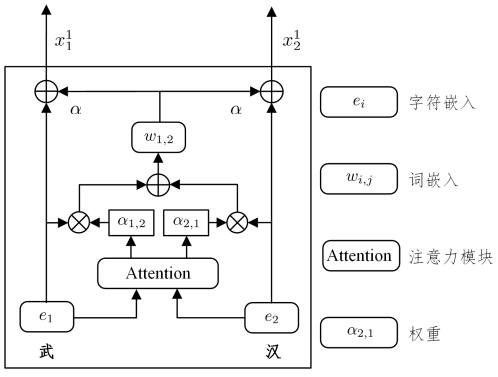

图 3 分割注意力模块结构

Fig. 3 Structure of segmentation attention module

对于输入的句子 $s=\{c_1,c_2,\cdots,c_N\}$，SAM 模块通过查找字符向量矩阵来为第 i 个字符c_i分配对应的字符嵌入$\boldsymbol{e}_i$。

$$\boldsymbol{e}_i=\text{Lookup}(c_i) \tag{1}$$

其中，Lookup$\in\mathbb{R}^{N\times d_e}$是随机初始化的字符嵌入查找表，可在训练过程中学习变化，d_e表示字符嵌入的维度，SAM 模块使用 Jieba 工具对文本进行分词，即引入大小可变且有意义的滑动窗口对文本中语义相关的字符进行分组，再根据分词结果利用局部自注意力(Local Attention)网络计算每个词语对应的词向量。以句中第 m 个字符为例，我们计算其所属词范围内任一字符c_n对c_m的影响权重$\alpha_{m,n}$，计算过程如下：

$$\alpha_{m,n}=\frac{\exp score(\boldsymbol{e}_m,\boldsymbol{e}_n)}{\sum\limits_{\tau\in\{i,i+1,\cdots,j\}}\exp score(\boldsymbol{e}_m,\boldsymbol{e}_\tau)} \tag{2}$$

其中，$1\leqslant i\leqslant m,n\leqslant j\leqslant N$，$i$ 和 j 分别表示词语中首字符和尾字符对应的索引值，$\boldsymbol{e}_n$对应第 n 个字符的字符嵌入。

分数的计算方式如下：

$$score(\boldsymbol{e}_m,\boldsymbol{e}_n)=\boldsymbol{v}^{\mathrm{T}}\tanh(\boldsymbol{W}_1\boldsymbol{e}_m+\boldsymbol{W}_2\boldsymbol{e}_n) \tag{3}$$

其中，$\boldsymbol{W}_1,\boldsymbol{W}_2\in\mathbb{R}^{d_h,d_e}$是可训练参数矩阵，$\boldsymbol{v}\in\mathbb{R}^{d_h}$。

文本中从第 i 个字符开始、到第 j 个字符结束的词向量$\boldsymbol{w}_{i,j}$的计算式如下：

$$\boldsymbol{w}_{i,j}=\sum_{n=i}^{j}\alpha_{m,n}\boldsymbol{e}_n \tag{4}$$

最终，将字符嵌入$\boldsymbol{e}_m$与其对应分词的词向量$\boldsymbol{w}_{i,j}$相加，得到字符向量$\boldsymbol{x}_m^1$：

$$\boldsymbol{x}_m^1=\boldsymbol{e}_m+\boldsymbol{w}_{i,j} \tag{5}$$

我们将分割注意力模块的输出$\boldsymbol{X}^1=\{\boldsymbol{x}_1^1,\boldsymbol{x}_2^1,\cdots,\boldsymbol{x}_N^1\}$作为第一层扁平命名实体识别层的输入。

3.2 扁平命名实体识别模块

通过 SAM 模块得到包含分词信息的字符特征序列后，将其输入到动态堆叠的扁平命名实体识别模型中进行多层序列标注。一旦检测到新的实体，就会在当前模型顶部引入一个新的 Flat NER 层，并通过连接层连接；否则，

模型终止堆叠，完成实体检测。

每个 Flat NER 层[35]由一个双向长短期记忆网络(BiLSTM)和一个条件随机场(CRF)组成。图 2 给出了扁平命名实体识别模块的结构。

由于 LSTM 编码器可以有效学习字符的隐藏层状态和每一步的序列信息，因此我们将第 i 层的特征序列 $\boldsymbol{X}^i=\{\boldsymbol{x}_1^i,\boldsymbol{x}_2^i,\cdots,\boldsymbol{x}_N^i\}$ 输入 BiLSTM 计算得到隐藏层状态，再将前向隐藏层状态序列 $(\overrightarrow{h_1^i},\overrightarrow{h_2^i},\cdots,\overrightarrow{h_N^i})$ 与后向隐藏层状态序列 $(\overleftarrow{h_1^i},\overleftarrow{h_2^i},\cdots,\overleftarrow{h_N^i})$ 进行拼接，得到完整的隐藏层特征序列 $H^i=(\boldsymbol{h}_1^i,\boldsymbol{h}_2^i,\cdots,\boldsymbol{h}_N^i)$。

接着我们利用标准的 CRF 网络[36]捕获相邻标签之间的依赖关系，以做出更好的决策。对于给定的特征序列 $\boldsymbol{H}^i$，生成标签预测序列 $T^i=(t_1^i,t_2^i,\cdots,t_N^i)$ 的概率为：

$$P(T^i\mid\boldsymbol{H}^i)=\frac{\exp\left(\sum_{n=0}^{N}\boldsymbol{A}_{t_n^i,t_{(n+1)}^i}+\sum_{n=1}^{N}U_{n,t_n^i}\right)}{\sum_{\overline{T}\in T_{H^i}}\exp\left(\sum_{n=0}^{N}\boldsymbol{A}_{\overline{t}_n,\overline{t}_{(n+1)}}+\sum_{n=1}^{N}U_{n,\overline{t}_n}\right)} \tag{6}$$

其中，$T_{\boldsymbol{H}^i}$ 是第 i 层所有可能生成的标签序列的集合；$\boldsymbol{A}$ 是标签转移矩阵，$\boldsymbol{A}_{t_n^i,t_{(n+1)}^i}$ 表示从标签 t_n^i 转移到标签 $t_{(n+1)}^i$ 的分数；U_{n,t_n^i} 表示第 n 个字符被分类为标签 t_n^i 的分数。在解码过程中，CRF 网络使用维特比算法[37](Viterbi Algorithm)得到最佳预测标签序列。

给定训练样本 $\{(\boldsymbol{X}^i,T^i)\}\mid_{i=1}^{K}$，若模型一共有 K 层 Flat NER 层，那么总损失值的计算式如下：

$$L=-\log(P(T\mid\boldsymbol{X}))=-\sum_{i=1}^{K}\log(P(T^i\mid\boldsymbol{H}^i)) \tag{7}$$

根据 CRF 输出的标签序列，我们对被识别出的实体所包含的字符的隐藏层特征进行加和求平均，得到实体的特征表示。若一个被识别出的实体从第 j 个字符开始，到第 k 个字符结束，那么该实体对应的特征表示为：

$$y_{j,k}=\frac{1}{k-j+1}\sum_{\tau=j}^{k}\boldsymbol{h}_\tau \tag{8}$$

3.3 边界生成式模块

文献[8]在两个 Flat NER 层之间传递被识别出的实体信息时，是通过对实体内的字符特征向量加和求平均实现的，即被识别出的实体的特征序列被压缩为一个字符特征大小。

$$\boldsymbol{m}_j=\frac{1}{\text{end}-\text{start}+1}\sum_{k=\text{start}}^{\text{end}}\boldsymbol{n}_k \tag{9}$$

其中，start 和 end 分别代表被识别实体的始末位置索引，$\boldsymbol{n}_k$ 表示第 k 个字符对应的特征向量，$\boldsymbol{m}_j$ 表示压缩之后的特征向量。

然而，这种压缩方法存在误差传播问题，不仅丢失了边界信息，也忽略了实体原本的长度信息，造成了文本内容的改变，进而容易导致后续模型对文本语义的理解出错，而这个损失过程是不可逆的。为了解决这个问题，本文设计了一种边界生成模块，该模块通过拼接字符特征、边界编码(Boundary Encodings)和实体位置编码(Entity Position Encodings)，来标识被识别的实体的边界信息和位置信息，边界生成式模块的结构如图 4 所示。

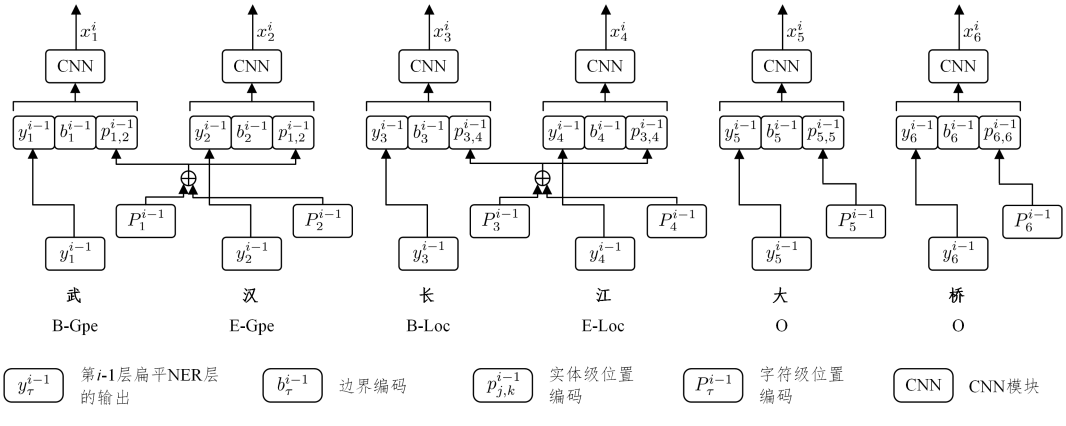

图 4 边界生成式模块的结构

Fig. 4 Structure of boundary generative module

参照 BIEO 标注规则，我们将被识别的实体分为 3 部分，即起始字符分配标签“B”、末尾字符分配标签“E”、实体内非边界字符分配标签“I”，分别对应边界向量值 1.0，0.9 和 0.6，非实体字符分配标签“O”，对应向量值为 0.1。

依据上述规则，得到第 $i-1$ 层被识别出的实体的边界表示 $\boldsymbol{B}^{i-1}=(\boldsymbol{b}_1^{i-1},\boldsymbol{b}_2^{i-1},\boldsymbol{b}_3^{i-1},\cdots,\boldsymbol{b}_N^{i-1})$。

与此同时，我们注意到词之间的顺序关系通常会影响到整个句子的含义。我们引入了实体位置编码，将被识别实体包含的字符的位置编码的平均值作为实体的位置表示。

$$\boldsymbol{p}_{j,k}^{i-1}=\frac{1}{k-j+1}\sum_{\tau=j}^{k}P_\tau^{i-1} \tag{10}$$

其中，j 和 k 分别表示实体始于字符 y_j、结束于字符 y_k，$\boldsymbol{p}_{j,k}^{i-1}$ 表示实体位置编码，$\boldsymbol{p}_\tau^{i-1}$ 代表句子中第 τ 个字符对应的字符位置编码，$\boldsymbol{p}_\tau^{i-1}\in\mathbb{R}^{N\times d_e}$ 是一个二维矩阵。我们选择正余弦函数[38]来编码绝对位置信息，将每一维的数值限制在一定范围内，便于后续进行数据处理，同时加快模型训练收敛。

如图 4 所示，BGM 模块先将特征序列 $\boldsymbol{Y}^{i-1}=(\boldsymbol{y}_1^{i-1},\boldsymbol{y}_2^{i-1},\boldsymbol{y}_3^{i-1},\cdots,\boldsymbol{y}_N^{i-1})$ 与边界编码 $\boldsymbol{B}^{i-1}$ 和位置编码 $\boldsymbol{p}^{i-1}$ 拼接得到 $\widetilde{\boldsymbol{y}}_\tau^{i-1}=[\boldsymbol{y}_\tau^{i-1};\boldsymbol{b}_\tau^{i-1};\boldsymbol{p}_\tau^{i-1}]$，再利用 CNN 网络对其进行通道压缩，得到下一层 Flat NER 层的输入 $\boldsymbol{X}^i=\{\boldsymbol{x}_1^i,\boldsymbol{x}_2^i,\cdots,\boldsymbol{x}_N^i\}$，计算过程如下：

$$\boldsymbol{X}^i=\text{CNN}(\widetilde{\boldsymbol{Y}}^{i-1}) \tag{11}$$

其中，$\widetilde{\boldsymbol{Y}}^{i-1}=(\widetilde{\boldsymbol{y}}_1^{i-1},\widetilde{\boldsymbol{y}}_2^{i-1},\cdots,\widetilde{\boldsymbol{y}}_N^{i-1})$，$\widetilde{\boldsymbol{Y}}^{i-1}\in\mathbb{R}^{N\times 3d_e}$。

4 实验

本文在 ACE 2005 中文嵌套命名实体数据集[39]上对

BLNM 模型的有效性进行评估。

4.1 实验设置

4.1.1 数据集

本文将 ACE 2005 中文数据集[39]按照 9∶1 的比例划分为训练集和测试集[6,8,20,34]。该数据集主要来源于广播新闻(Broadcast News)、新闻专线(Newswire)和网页博客(Weblog)，由 6000 多条句子、30000 余个实体组成。其中，实体主要包含七大类，分别为人名(person，Per)、组织(organization，Org)、位置(location，Loc)、设施(facility，Fac)、武器(weapon，Wea)、车辆(vehicle，Veh)和地缘政治实体(geo-political entity，Gpe)。该数据集在每一层的数据分布如表 1 所列。

表 1 ACE 2005 中文嵌套命名实体数据集的实体分布

Table 1 Entity distribution of ACE 2005 Chinese nested NE dataset

层数	Per	Org	Loc	Fac	Wea	Veh	Gpe	Ne	
								Total	百分比/%
第一层	10750	4726	1050	1156	324	512	7761	26279	77.62
第二层	2831	1737	405	390	42	127	815	6347	18.74
第三层	524	231	79	104	9	22	78	1047	3.09
第四层	89	28	12	20	0	2	7	158	0.47
第五层	17	1	2	3	0	1	2	26	0.08
总计	14211	6723	1548	1673	375	664	8663	33857	100

注：Ne 代表实体数量

从表 1 中不难看出，ACE 2005 中文嵌套命名实体数据集中的高层嵌套实体分布稀疏，第 4 层和第 5 层一共仅有不到 1%的实体。同时，我们还能从表 1 中看出，不同类型实体的分布不均衡，其中 Per 类型实体最多，其次是 Gpe 类型实体和 Org 类型实体。

4.1.2 评估指标

本文在评估中文嵌套命名实体识别性能时遵循精确匹配模式(Exact-match Evaluation)，即实体的预测边界和预测类型与人工标注的结果均完全相同时，才能判定该实体被正确识别。

评估过程中使用真正例(True Positives，TP)、假正例(False Positives，FP)和假反例(False Negatives，FN)这 3 个参数计算 CNNER 任务的评价指标，即精确率(Precision)、召回率(Recall)和 F1 值(F1-score)，计算式如下：

$$Precision=\frac{TP}{TP+FP} \tag{12}$$

$$Recall=\frac{TP}{TP+FN} \tag{13}$$

$$F1=2\times\frac{Precision\times Recall}{Precision+Recall} \tag{14}$$

4.1.3 超参数设置

在模型超参数方面，本文参考文献[8]，设置 LSTM 隐藏层状态维度和字符嵌入维度均为 200，批次大小(Batch Size)为 16，学习率(Learning Rate)初始值为 0.001，可在训练过程中随预定义规则进行动态调整。为避免模型出现过拟合，本文设置了早停参数(Earlystopping)、L2 正则化参数(L2-regularization)和 Dropout 率，其中 Dropout 设置为 0.5。

本文通过多组对比实验，在一定范围内选取最佳超参数。如表 2 所列，每组对比实验中，只有一个超参数取值发生改变，其余超参数取值均保持相同。为公平起见，所有实验都是基于 PyTorch 框架[40]，并且在同一块 NVIDIA GeForce GTX 1080 Ti GPU 上运行的。结合表 2 可知，当批次大小取 16，Dropout 率取 0.5，初始学习率取 0.001，隐藏层状态维度取 200 时，模型性能最好。

表 2 ACE 2005 中文嵌套命名实体数据集上超参数的调整范围和最佳取值

Table 2 Value range and best value of tuned hyper parameters in ACE 2005 Chinese nested NE dataset

超参数	范围	最优值
批次大小(Batch size)	8，16，32，64	16
Dropout 率(Dropout rate)	0.1，0.3，0.5，0.7	0.5
初始学习率(Learning rate)	0.0001，0.001，0.003，0.01	0.001
隐藏层状态维度(No. of hidden units)	100，200，300	200

4.2 整体评估

为了验证本文模型的有效性，本文选取了与本文模型相关的具有代表性的模型作为对比模型。Ju 等[8]提出了一个能聚合实体特征的神经层叠模型(Neural Layered Model，NLM)，用于检测英文数据集中的嵌套实体。考虑到中英文之间语言特性不同，我们将 NLM 模型的嵌入层(Embedding Layer)替换为适用于中文字符的结构。基础模型(Baseline)是一个层叠神经网络(Stacked Neural Model，SNM)，相比 NLM 模型，其未在 Flat NER 层之间对被识别出的实体进行压缩。所有模型均在 ACE 2005 中文数据集上进行训练和测试。

表 3 列出了整体实验结果，不难看出，本文模型 BLNM 在 Precision，Recall 和 F1-score 上都优于现有模型。这归功于本文模型能够检测出更为准确的实体边界。一方面，SNM 和 NLM 模型都忽略了重要的分词信息和词序列信息；另一方面，SNM 模型没有向下一层传递预测信息，虽然 NLM 模型弥补了这一不足，但它同时也不可避免地造成了错误传递问题。

表 3 ACE 2005 中文嵌套命名实体测试集上的主要实验结果

Table 3 Main results on ACE 2005 Chinese nested NE test set

(单位：%)

模型	*P*	*R*	*F1*
SNM	75.27	70.73	72.93
NLM	75.52	72.21	73.84
本文模型(BLNM)	**77.21**	**73.71**	**75.42**

如表 4 所列，我们对比了 BLNM 模型与 SNM 模型在不同层检测实体上的效果，其中 BLNM 模型的性能均优于 SNM 模型，且分别在底层和高层达到了 76.50%和 56.55%的 F1-score。这里，我们将第一层的嵌套实体统称为底层实体(Bottom-level Entities)，将其余层的实体统称为高层实体(High-level Entities)。从表中我们也可以看出，随着层数增加，模型性能急剧下降，发生此现象主要有两个原因：1)仅有 19.73%的实体分布在高层，而底层实体的数量高达 80.27%，实体数量的急剧减少导致模型的高层无法得到充分训练；2)高层嵌套实体通常比底层实体更长，也更为复杂，这就导致

模型更难识别出高层实体。此外，层与层之间的连接方式非常重要，这决定了有效信息传递的质量。基于以上分析，BLNM 模型识别高层实体的性能理应优于 SNM 模型。

表 4 本文模型与 SNM 模型不同层识别效果的对比

Table 4 Results comparison of different layers of the proposed model and SNM

实体类别	BLNM			SNM			实体数量
	P/%	R/%	F1/%	P/%	R/%	F1/%	
底层实体	77.96	75.01	76.50	76.20	72.68	74.40	2449
高层实体	58.78	54.49	56.55	53.53	50.33	51.88	602

4.3 SAM 模块有效性验证

我们针对分割注意力模块 SAM 进行实验，以证明该模块能辅助模型更精确地识别实体。表 5 中 SNM+SAM 达到了 74.42%的 F1-score，相比基于字符的基础模型有 1.49%的提升。这表明对于中文嵌套命名实体识别任务来说，词信息和字符信息都是有效信息，其中词信息可以帮助模型更加准确地定位实体。此外，该模块同时具有可迁移性。由表 5 可知，将 SAM 模块应用于 NLM 模型时，同样能辅助模型更好地利用潜在有效词信息，将 F1-score 从 73.87%提升到 74.32%。

表 5 SAM 的性能

Table 5 Performance of SAM

（单位：%）

模型	P	R	F1
SNM	75.27	70.73	72.93
+SAM	75.30	73.55	74.42
NLM	75.52	72.21	73.87
+SAM	75.76	72.93	74.32

4.4 BGM 模块有效性验证

本文通过在 ACE 2005 中文嵌套命名实体测试数据集上进行对比实验，验证了边界生成式模块 GBM 的有效性，结果如表 6 所列。BE 代表边界编码，相比 BGM 模块，其没有实体位置信息。从表中可以看出，SNM+BE 与 SNM+BGM 分别实现了 0.55%和 0.88%的 F1-score 提升。我们认为，本文提出的 BGM 模块通过将边界编码和位置编码集成到字符特征中为每个字符构造边界感知表示，可以在相邻层之间更好地传递信息并有效缓解错误累积问题。作为独立于嵌套结构的连接层，BGM 也可以灵活地迁移到其他模型中。

表 6 BGM 性能

Table 6 Performance of BGM

（单位：%）

模型	P	R	F1
SNM	75.27	70.73	72.93
+BE	75.36	71.68	73.48
BGM	75.86	71.88	73.81

4.5 案例学习

我们从 ACE 2005 中文测试数据集中选取了两个具有代表性的例子来分析 BLNM 模型的提升效果，一个是“韩国工会联合会主席李南成发表了演讲。”，另一个是“上周一撤换陆军及海军参谋长。”。

图 5 给出了本文模型 BLNM 与对比模型 SNM 和 NLM 在这两个例子上的识别效果。在第一个示例中，实体“韩国工会联合会”嵌套在大实体“韩国工会联合会主席”中。本文模型 BLNM 能够精确识别这两个实体的边界并进行准确分类，然而 SNM 和 NLM 都遗漏了大粒度实体“韩国工会联合会主席”，并误将“韩国工会”识别为实体。此外，NLM 模型还误检出了两个多余的实体“韩国”和“联合会”。

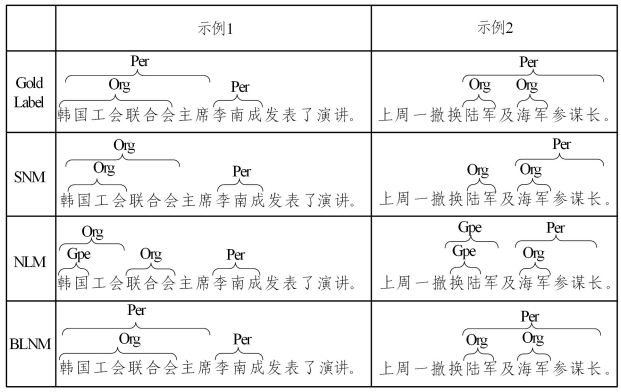

图 5 ACE 2005 中文测试集上预测结果的对比

Fig. 5 Prediction result comparison on ACE 2005 Chinese test set

在另一个示例中，大粒度实体“陆军及海军参谋长”包含两个小粒度实体“陆军”和“海军”。本文模型 BLNM 能够准确识别出这些实体，但是 SNM 和 NLM 模型不能。SNM 和 NLM 模型在识别大粒度实体时，同时遗漏了“陆”“军”“及”这 3 个字符。不仅如此，NLM 模型还在底层模型错误地将“换陆”识别为实体，并在高层将“换陆军”也识别为实体，这归咎于 NLM 模型层与层之间不合理的连接方式导致了错误边界信息的累积和传播。以上两个示例表明，本文设计的模型能准确定位大小粒度实体的边界，并进行精确分类。

结束语 本文提出了一种边界感知层叠神经网络模型(BLNM)，提升了中文嵌套命名实体识别性能。该模型结合了中文分词和注意力机制，将潜在的分词信息整合到字符特征中。同时，为了在相邻层之间更好地传递预测信息，本文设计了一个边界生成模块来标记被识别出的实体的边界和位置。基于通用数据集的实验结果表明，本文设计的模型在 Precision，Recall 和 F1-score 上均优于现有模型。下一步工作将考虑针对实体类型分布不均和高层实体分布稀疏问题来改进模型。

参 考 文 献

[1] GUPTA N, SINGH S, ROTH D. Entity linking via joint encoding of types, descriptions, and con-text[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Denmark: ACL, 2017: 2681-2690.

[2] JI Z, SUN A, CONG G, et al. Joint recognition and linking of fine-grained locations from tweets[C]// Proceedings of the 25th International Conference on World Wide Web. Montréal: WWW, 2016: 1271-1281.

[3] LIN Y, SHEN S, LIU Z, et al. Neural relation extraction with selective attention over in-stances[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Germany: ACL, 2016: 2124-2133.

[4] ZHENG S，WANG F，BAO H，et al. Joint extraction of entities and relations based on a novel tagging scheme[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver：ACL，2017：1227-1236.

[5] CHANG K W，SAMDANI R，ROTH D. A constrained latent variable model for coreference resolution [C]// Proceedings of the 2013 Conference on Empirical Methodsin Natural Language Processing. Seattle：EMNLP，2013：601-612.

[6] SHEN D，ZHANG J，ZHOU G，et al. Effective adaptation of hidden markov model-based named entity recognizer for biomedical domain[C]// Proceedings of the ACL 2003 Workshop on Natural Language Processing in Biomedicine. Japan：ACL，2003：49-56.

[7] XIA C，ZHANG C，YANG T，et al. Multi-grained named entity recognition[C]// Proceedings of the 57th Annual Meeting of the Association for Compu-tational Linguistics. Italy：ACL，2019：1430-1440.

[8] JU M，MIWA M，ANANIADOU S. A neural layered model for nested named entity recognition[C]// Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics. New Orleans：ACL，2018：1446-1459.

[9] LI H，XU H，QIAN L，et al. Multi-layer Joint Learning of Chinese Nested Named Entity Recognition Based on Self-attention Mechanism[C]// Proceedings of the 9th CCF International Conference on Natural Language Processing and Chinese Computing. Cham：Springer International Publishing，2020：144-155.

[10] KURU O，CAN O A，YURET D. CharNER：Character-level named entity recognition[C]// Proceedings of COLING 2016，the 26th International Conference on Computational Linguistics：Technical Papers. Osaka：COLING，2016：911-921.

[11] TRAN Q，MACKINLAY A，YEPES A J. Named entity recognition with stack residual lstm and trainable bias decoding[J]. arXiv：1706. 07598，2017.

[12] GRIDACH M. Character-level neural network for biomedical named entity recognition[J]. Journal of biomedical informatics，2017，70：85-91.

[13] EBERTS M，ULGES A. Span-based joint entity and relation extraction with transformer pre-training[J]. arXiv：1909. 07755，2019.

[14] HUANG Z，XU W，YU K. Bidirectional LSTM-CRF models for sequence tagging[J]. arXiv：1508. 01991，2015.

[15] LIU L，SHANG J，REN X，et al. Empower sequence labeling with task-aware neural language model[C]// Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence. New Orleans：AAAI Press，2018：5253-5260.

[16] LUAN Y，WADDEN D，HE L，et al. A general framework for information extraction using dynamic span graphs[C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics. Minnesota：ACL，2019：3036-3046.

[17] DONG C，ZHANG J，ZONG C，et al. Character-based LSTM-CRF with radical-level features for Chinese named entity recognition[C]// International Conference on Computer Processing of Oriental Languages. Cham：Springer International Publishing，2016：239-250.

[18] ZHANG Y，YANG J. Chinese NER using lattice LSTM[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne：ACL，2018：1554-1564.

[19] LIU W，XU T，XU Q，et al. An encoding strategy based word-character LSTM for Chinese NER[C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics. Minneapolis：ACL，2019：2379-2389.

[20] WU Y，JIANG M，LEI J，et al. Named entity recognition in Chinese clinical text using deep neural network[J]. Studies in health technology and informatics，2015，216：624-628.

[21] GUI T，ZOU Y，ZHANG Q，et al. A Lexicon-Based Graph Neural Network forChinese NER[C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Hong Kong：ACL，2019：1040-1050.

[22] LI X，YAN H，QIU X，et al. FLAT：Chinese NER Using Flat-Lattice Transformer[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online：ACL，2020：6836-6842.

[23] MENGGE X，YU B，LIU T，et al. Porous Lattice Transformer Encoder for Chinese NER[C]// Proceedings of the 28th International Conference on Computational Linguistics. Barcelona：International Committee on Computational Linguistics，2020：3831-3841.

[24] GUI T，MA R，ZHANG Q，et al. CNN-Based Chinese NER with Lexicon Rethinking[C]// Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence Main track. Macao：IJCAI，2019：4982-4988.

[25] SUI D，CHEN Y，LIU K，et al. Leverage lexical knowledge for Chinese named entity recognition via collaborative graph network[C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language. Hong Kong：EMNLP，2019：3830-3840.

[26] LI J H，CHEN M M，WANG H J，et al. Chinese Named Entity Recognition Method Based on ALBERT-BGRU-CRF[J]. Computer Engineering，2022，48(6)：89-94，106.

[27] ZHONG S S，CHEN X，ZHAO M H，et al. Incorporating word-set attention into Chinese named entity recognition Method[J]. Journal of Jilin University(Engineering and Technology Edition)，2022，52(5)：1098-1105.

[28] GUO X R，LUO P，WANG W L. Chinese named entity recognition based on Transformer encoder[J]. Journal of Jilin University(Engineering and Technology Edition)，2021，51(3)：989-995.

[29] SI Y C，GUAN Y Q. Chinese Named Entity Recognition Model Based on Transformer Encoder[J]. Computer Engineering，2022，48(7)：66-72.

[30] HU X B，YU X Q，LI S M，et al. Chinese Named Entity Recog-

nition Based on Knowledge Enhancement[J]. Computer Engineering, 2021, 47(11): 84-92.
[31] FU C, ZHAO Y, FU G. Exploiting entity-level morphology to Chinese nested named entity recognition[J]. International Journal on Asian Language Processing, 2012, 22(1): 33-48.
[32] ZHOU G, ZHANG J, SU J, et al. Recognizing names in biomedical texts: a machine learning approach[J]. Bioinformatics, 2004, 20(7): 1178-1190.
[33] ZHOU G D. Recognizing names in biomedical texts using mutual information independence model and SVM plus sigmoid[C]// Proceedings of the International Joint Workshop on Natural Language Processing in Bio-medicine and its Applications. Geneva, Switzerland: COLING, 2004: 1-7.
[34] FU C, FU G. Morpheme-based chinese nested named entity recognition[C]// Proceedings of the 2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery. Shanghai: FSKD, 2011: 1221-1225.
[35] LAMPLE G, BALLESTEROS M, SUBRAMANIAN S, et al. Neural architectures for named entity recognition[C]// Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics. San Diego: NAACL, 2016: 260-270.
[36] LAFFERTY J, MCCALLUM A, PEREIRA F C N. Conditional random fields: Probabilistic models for segmenting and labeling sequence data[C]// Proceedings of the 8th International Conference on Machine Learning. Evanston: ICML, 1991: 282-289.
[37] VITERBI A. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm [C]// Proceedings of IEEE Transactions on Information Theory. IEEE, 1967: 260-269.
[38] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. arXiv: 1706.03762, 2017.
[39] DODDINGTON G R, MITCHELL A, PRZYBOCKI M A, et al. The automatic content extraction(ace) program-tasks, data, and evaluation[C]// Proceedings of the Fourth International Conference on Language Resources and Evaluation. Portugal: European Language Resources Association, 2004: 837-840.
[40] PASZKE A, GROSS S, MASSA F, et al. Pytorch: An imperative style, high-performance deep learning library[J]. arXiv: 1912.01703, 2019.

ZHANG Rujia, born in 1997, postgraduate. Her main research interests include natural language processing and nested name dentity recognition.

WANG Bang, born in 1975, professor, Ph.D supervisor, is a member of China Computer Federation. His main research interests include recommendation algorithm, knowledge graph and so on.

(责任编辑:喻藜)