

# 分层区域穷举的中文嵌套命名实体识别方法

余诗媛<sup>1,2</sup>, 郭淑明<sup>2</sup>, 黄瑞阳<sup>2</sup>, 张建朋<sup>2</sup>, 胡楠<sup>1,2</sup>

(1. 郑州大学软件学院, 河南 郑州 450001;

2. 国家数字交换系统工程技术研究中心, 河南 郑州 450002)

**摘要:**嵌套命名实体之间蕴含着丰富的语义关系与结构信息,开发能够准确识别嵌套命名实体的算法具有重要研究意义。针对现有的中文嵌套命名实体数据集中存在错标漏标以及现有识别方法大多忽略嵌套实体内部信息关联关系而导致准确性下降的问题,结合自动生成与手动标注的方法构建新的中文嵌套命名实体数据集 NEPD,在此基础上,设计一种利用分层区域穷举的中文嵌套命名实体识别模型。该模型通过遍历文本组合实体,获取低层编码层的词嵌入信息;其次,为使邻接编码层之间实现信息交换,将低层编码层的词嵌入信息融入高层编码层;最后,利用多层解码层使长度为 L 的命名实体仅在第 L 层预测,有效防止错误传播现象发生从而提高识别准确度。实验结果表明,在没有外部知识资源的情况下,LREM 模型在嵌套命名实体与非嵌套命名实体上的识别 F1 值分别达到 87.19% 和 86.27%,其中非嵌套命名实体识别的 F1 值比传统的 BiLSTM+CRF 模型提升 1.18%,验证了该模型的可靠性。

**关键词:**嵌套命名实体识别;分层区域穷举;卷积神经网络;双向长短期记忆网络;信息抽取

**中图分类号:**TP18

**文献标识码:**A

**文章编号:**1673-629X(2022)09-0161-06

doi:10.3969/j.issn.1673-629X.2022.09.025

## Layered Regional Exhaustive Model for Chinese Nested Named Entity Recognition

YU Shi-yuan<sup>1,2</sup>, GUO Shu-ming<sup>2</sup>, HUANG Rui-yang<sup>2</sup>,

ZHANG Jian-peng<sup>2</sup>, HU Nan<sup>1,2</sup>

(1. School of Software, Zhengzhou University, Zhengzhou 450001, China;

2. National Digital Switching System Engineering and Technological R&D Center, Zhengzhou 450002, China)

**Abstract:** Nested named entities contain rich semantic relationships and structural information among them, and it is essential to develop algorithms that can accurately identify nested named entities. To address the problems of mislabeling and omission in the existing Chinese nested named entity dataset, and the problem that most of the existing recognition methods ignore the internal information association relationship of nested entities, a new Chinese nested named entity dataset NEPD is constructed by combining automatic generation and manual annotation methods, based on which a Chinese nested named entity recognition model is designed using hierarchical region exhaustive. The model obtains the word embedding information of the lower coding layer by traversing the text combination entities. Furthermore, the word embedding information of the lower coding layer is incorporated into the higher coding layer to exchange data between neighboring coding layers. Finally, the named entities of length L are predicted only in the L layer by using multiple decoding layers, which effectively prevents the occurrence of error propagation and thus improves the recognition accuracy. The experimental results show that without external knowledge resources, the F1 values of the LREM model reach 87.19% and 86.27% for the recognition of nested named entities and non-nested named entities, respectively, with the F1 value of non-nested named entities recognition improving 1.18% compared with the traditional BiLSTM+CRF model. The experiments verify the reliability of the model in this paper.

**Key words:** nested named entity recognition; layered regional exhaustive model; convolutional neural network; bi-directional long short term memory network; information extraction

收稿日期:2021-09-16

修回日期:2022-01-17

**基金项目:**国家自然科学基金青年基金项目(62002384);中国博士后科学基金面上项目(47698);郑州市协同创新重大专项(162/32410218)

**作者简介:**余诗媛(1996-),女(彝族),硕士研究生,研究方向为自然语言处理、知识图谱;通信作者:郭淑明(1977-),男,副研究员,研究方向为无线通信、大数据技术。

## 0 引言

命名实体识别 (Name Entity Recognition, NER) 是信息抽取的标准任务之一,其主要目的是抽取出自语言文本中具有特定含义的命名实体,例如:组织名、地名、人名等。命名实体识别往往被视为一个序列标注任务,即对于指定字符序列,命名实体识别模型需要预测每个字符对应的命名实体标签。嵌套命名实体是一类特殊形式的命名实体,其内部包含一个或者多个命名实体,最外层的命名实体称为外部命名实体,嵌套在其内部的实体称为内部命名实体。由于单个词条可能拥有两个及以上的标签,传统的命名实体识别模型无法完整地识别出具有嵌套结构的内部命名实体,从而难以精确捕捉文本中存在的细粒度语义信息及结构信息。

当前,只有少数研究针对英文嵌套命名实体识别。Ju 等人<sup>[1]</sup>提出 Layered-BiLSTM-CRF 模型,该模型动态堆叠平面 NER 层以识别内部命名实体,并运用充分的内部命名实体编码信息识别外部命名实体,最终在 ACE 2005、GENIA 数据集上分别获得了 72.2%、74.7% 的 F1 值。但该方法存在层与层之间的错误传播,若无法识别外部命名实体,则无法检测到内部命名实体。Sohrab 等人<sup>[2]</sup>提出了神经穷举模型,其关键思想是列举所有可能的区域或跨度作为潜在的实体提及,并用深层神经网络对它们进行分类,该方法简单有效,但是外部命名实体信息和内部命名实体信息完全没有交互。中文嵌套命名实体识别任务目前缺少被广泛认可的数据集,《人民日报》数据集虽然含有部分嵌套命名实体的标注信息,但是这些命名实体信息不完整且存在错标漏标现象,如“[中共中央/nt 顾问/n 委员会/n]nt”转化为“[中共中央]nt”及“[中共中央顾问委员会]nt”,漏标了“[中共]nt”。

结合自动生成及手工标注方法,首先,基于《人民日报》数据集构建了新的中文嵌套命名实体识别数据集 NEPD (Nested Entity of the People's Daily);其次,提出了一种结合动态堆叠平面及神经穷举的嵌套命名实体识别方法,即根据实体长度分层枚举所有可能的区域或组合,使用卷积神经网络将短组合实体的词嵌入与相邻字符的词嵌入聚合形成长组合实体的词嵌入;然后,利用 BiLSTM 预测出每一个组合的标签;最后,在中文嵌套命名实体识别数据集上对该方法进行实验验证。

## 1 相关工作

嵌套命名实体具有结构复杂多变,嵌套颗粒度与嵌套层数缺乏规律性等特点,例如,“中共北京市委宣

nt 宣传部]nt”包含“[中共]nt”、“[北京]ns”和“[中共北京市委]nt”三个内部命名实体。

当前,嵌套命名实体识别研究工作主要基于英文基准数据集 (ACE 语料<sup>[3-4]</sup>、GENIA 语料<sup>[5]</sup>、NNE 数据集<sup>[6]</sup>、KBP2015 语料库等)。ACE 语料用于自动内容提取技术评估,标注了包含 7 种主要的命名实体类型:地理命名实体 (GPE)、组织 (ORG)、人 (PER)、地点 (LOC)、车辆 (VEH)、设施 (FAC) 和武器 (WEA); GENIA 语料是最早标注的生物医学文献集合,主要有五种命名实体类型:DNA、RNA、蛋白质、细胞系和细胞类型;NNE 数据集标注了华尔街日报的细粒度、嵌套命名实体,包含 114 种实体类型,嵌套深度高达 6 层。

针对英文嵌套命名实体识别,文献[1]充分运用内部命名实体编码信息识别外部命名实体,提出了一种动态堆叠平面 NER 层以识别嵌套命名实体的模型,若检测到命名实体,该模型的平面 NER 层首先获取 LSTM 层的上下文表示,其次将该表示作为输入传递到新的平面 NER 层,最后将上下文表示提供给 CRF 层进行标签预测,当没有检测到命名实体时,模型停止堆叠,直至完成命名实体识别;文献[2]提出了神经穷举模型,首先从参数共享的 BiLSTM 中获取区域表示,将每个区域表示输入到修正线性单元中作为激活函数,并把激活层的输出传递到 softmax 输出层,以识别该区域是否为特定的命名实体类型或非命名实体;文献[7]提出了一种基于超图表示的模型,称为提及超图 (Mention Hypergraph, MH),该模型首先使用节点和有向超边共同对命名实体及其组合进行表示,紧凑地将一个句子中不同类型且无限长度的嵌套命名实体表示出来,以解决嵌套命名实体检测难的问题,在 ACE 04 和 ACE 05 数据集上的 F1 值达到 62.8% 及 62.5%;文献[8]提出了一种利用边界预测命名实体分类标签的边界感知模型。该模型将嵌套命名实体识别分为了两个任务,首先用序列标注模型来发现命名实体的位置;其次,用基于跨度的模型对序列标注发现的候选命名实体跨度进行实体类型的分类;最后采用了多任务损失算法同时训练两个任务,以获取边界检测模块和命名实体分类预测模块的底层依赖关系;文献[9]提出了一种基于机器阅读理解 (Machine Reading Comprehension, MRC) 的方法统一解决非嵌套和嵌套命名实体识别问题,在 ACE04 和 MSRA 等 8 个中英数据集上均取得了显著的识别成果。

由于缺乏规范统一的语料库,当前中文嵌套命名实体识别工作缺乏横向比较的基准。文献[10]引入实体语素概念,基于机器学习方法构建汉语命名实体的双层模型,很大程度上解决了嵌套命名实体边界识

别错误问题;文献[11]提出并设计了一种联合模型对嵌套命名实体进行识别,可联合处理分词问题、命名实体的边界确定问题、类别确定问题。

## 2 中文嵌套命名实体识别数据集构建

### 2.1 数据集构建

《人民日报》语料库是对一九九八年一月至六月出版的人民日报纯文本语料进行分词和词性标注后制作而成,该文分析了语料库中嵌套命名实体统计分布规律,该语料中共 106 430 个名词,主要包含人名实体、地名实体以及组织机构名实体;其中,嵌套命名实体总数为 7 993, 占所有命名实体总数的 18.5%。

为了减少标注工作复杂度,该文自动抽取语料中已经标注的部分嵌套命名实体,然后人工标注漏标的命名实体,并修订标注错误的命名实体,保证原始数据的清洁与专业性,主要步骤如下:

(1) 自动抽取:从《人民日报》语料库中抽取命名实体,保留原有的命名实体标注,例如:“[中共/j 上海/ns 市委/n]nt”提取后变成“[中共[上海]ns 市委]nt”;

(2) 人工调整:基于自动抽取的结果,人工增添未标注命名实体,并对部分错误标注命名实体进行调整,例如:“[中共[上海]ns 市委]nt”经人工调整后为“[[中共]nt[上海]ns 市委]nt”;

(3) 一致性验证:使用一致性分数衡量标注的一致性,一致性分数通过标注一致的标签除以标签总数得到;

(4) 人工验证:人工对标注数据复查,核查数据一致性与准确性。

### 2.2 数据集统计

表1、表2分别列出了修订后的《人民日报》嵌套命名实体数据集的层次结构统计及实体比例分布情况,表中括号内为原数据集的实体比例分布。由表1、

表1 嵌套命名实体层次结构统计

类型	层次	数量	所有命名实体 比例/%
基本命名实体	1	44 605(35 124)	79.1(81.5)
	2	8 388(6 864)	14.9(16.0)
嵌套命名实体	3	3 251(1 046)	5.8(2.4)
	4	118(83)	0.2(0.1)

表2 嵌套命名实体比例分布情况

类型	嵌套命名 实体数量	本类型所有 命名实体数	最大嵌套 层数
地名	2 771(1 131)	23 937	4
组织机构名	8 986(6 862)	12 445	4

表2可知,调整后的数据集包含 19 980 个人名实体、23 937个地名实体、12 445 个组织机构名实体;其中,嵌套命名实体总数为 11 757, 占所有命名实体总数的 20.9%, 最深层次达到四层,如:“[[[[长沙市]ns 公安局]nt 交警支队]nt 党委]nt”。由于人名中含嵌套命名实体的数量非常少,可以忽略不计,所以含有嵌套结构的命名实体主要存在于地名、机构名之中,结构多为地名+地名+其他+后缀词、组织名+组织名+其他+后缀词,如:“[[中国]ns 驻[南非]ns 大使馆]nt”、“[[中共]nt 中央]nt 统战部]nt”。

## 3 中文嵌套命名实体识别方法

针对现有的嵌套命名实体识别方法忽略嵌套实体内部信息关联关系而导致准确性下降的问题,提出了分层区域穷举模型(Layered Regional Exhaustive Model, LREM), 分层区域穷举模型利用神经网络详尽地列举了句子中所有的区域或组合。该模型建立在多层内部编码层以及解码层之上, 编码层由卷积神经网络(Convolutional Neural Networks, CNN)构成, 解码层由双向长短时记忆网络(Bi-directional Long Short Term Memory Network, BiLSTM)构成, 本节将详细描述 LREM 的体系结构, 如图1所示。

### 3.1 词嵌入

给定输入的句子序列  $X = \{x_1, x_2, \dots, x_n\}$ , 其中  $x_i$  为第  $i$  个字符,  $n$  为句子序列中的字符数。该文参考文献[12]中的方法, 使用 Word2Vec<sup>[13]</sup> 在自动分词的 Chinese Giga-Word 数据集上对词嵌入进行预训练, 并使用预训练后的词嵌入初始化输入句子序列的词嵌入  $W = \{w_1, w_2, \dots, w_n\}$ , 其中  $w_i$  为第  $i$  个字符的词嵌入。

### 3.2 编码层

分层区域穷举模型的编码层由  $n$  层内部编码层构成, 每一层内部编码层由卷积神经网络构成, 用于构建固定长度的区域表示, 即第  $L$  层编码层用于构建字符长度为  $L$  的组合实体的区域表示,  $1 \leq L \leq n$ 。首先, 利用卷积神经网络聚合两个相邻的区域表示及词嵌入, 将聚合得到的区域表示传递给相应长度的解码层; 同时, 将聚合得到的区域表示与邻接词嵌入拼接得到新的嵌入序列, 将该序列传递至更高编码层。通过卷积神经网络, 分层区域穷举模型可以遍历文本中所有的组合实体, 获取低层编码层的词嵌入信息融入高层编码层, 使邻接编码层之间实现信息交换。

设  $H_{i,i+l}$  为以  $i$  为起点且长度为  $l$  的组合实体的区域表示, 则其计算如公式(1)所示。

$$H_{i,i+l} = \text{conv1}(H_{i,i+l-1}, w_{i+l}) \quad (1)$$

### 3.3 解码层

传统的层叠模型<sup>[14-16]</sup>在嵌套命名实体识别过程

中容易产生层迷失问题,即模型在错误的嵌套层输出嵌套命名实体,例如:从第一层识别出嵌套命名实体“[[中共]nt 中央]nt[台湾]ns 工作办公室]nt”。虽然命名实体边界与类别均正确,但这会导致模型趋向于不预测内部的嵌套命名实体,从而影响召回率。分

层区域穷举模型通过使用多层解码层,使长度为  $L$  的命名实体仅在第  $L$  层预测来防止层迷失现象以及层与层之间的错误传播现象产生。将经过编码层获得的词向量输入到 BiLSTM 中进行进一步处理,得到其预测标注序列。

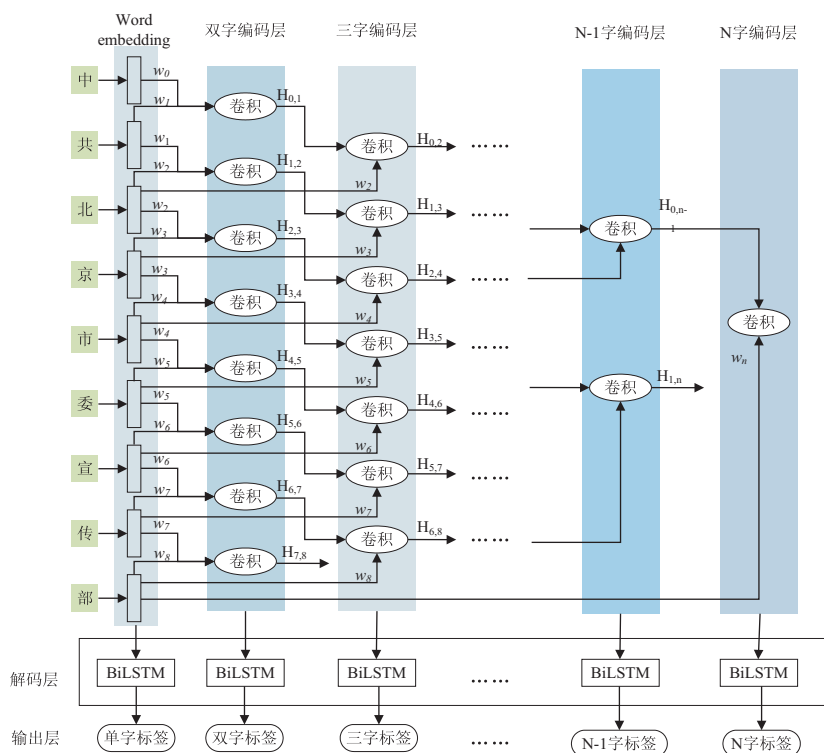


图 1 LREM 的体系结构

长短期记忆网络 (Long Short-Term Memory, LSTM) 是一种特殊的循环神经网络 (Recurrent Neural Network, RNN), 能够更好地学习到较长距离的依赖关系, 非常适合基于时间序列数据的分类问题。LSTM 由  $t$  时刻的输入词  $x_t$ 、细胞状态  $C_t$ 、临时细胞状态  $\tilde{C}_t$ 、隐层状态  $h_t$ 、遗忘门  $f_t$ 、记忆门  $i_t$ 、输出门  $o_t$  构成, 通过遗忘门  $f_t$ 、记忆门  $i_t$ 、输出门  $o_t$  结构选择性的遗忘历史信息 and 记忆输入的新信息, 最终整合到当前状态并在每个时间步输出隐层状态。

遗忘门将细胞状态中的信息选择性遗忘, 其计算过程如公式(2)所示。

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2)$$

其中,  $h_{t-1}$  表示  $t-1$  时刻的隐层状态,  $x_t$  表示当前时刻  $t$  的输入词,  $\sigma$  表示 sigmoid 函数。

记忆门决定将在细胞状态中存储的新信息范畴, 首先使用记忆门的 sigmoid 层决定需要更新的信息, 然后利用 tanh 层创建包含新候选值的向量, 最后, 将这两部分联合更新细胞状态, 其计算过程如公式(3)、(4)、(5)所示。

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (4)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (5)$$

输出门将基于细胞状态决定输出值, 首先使用 sigmoid 层确定是否将细胞状态的某个部分输出, 然后, 将细胞状态通过 tanh 进行处理并将其与 sigmoid 的输出相乘, 从而仅输出决定输出的部分, 其计算过程如公式(6)、(7)所示。

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (7)$$

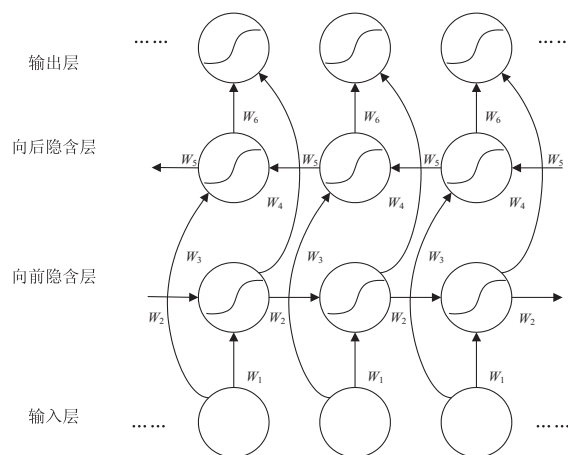


图 2 BiLSTM 网络结构



BiLSTM 由前向 LSTM 与后向 LSTM 构成,能够更好地学习双向的语义依赖。BiLSTM 网络结构如图 2 所示。向前隐含层计算并保存输入序列中  $t$  时刻以及之前时刻的信息,向后隐含层计算并保存输入序列中  $t$  时刻以及之后时刻的信息,最后将每个时刻向前隐含层和向后隐含层输出的结果向量经过连接、相加或平均值等方式处理得到最终的隐层表示。

## 4 实验结果与分析

### 4.1 实验环境配置

实验基于 Tensorflow 平台搭建,实验硬件环境配有 Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10GHz 的浪潮服务器,同时装载 8 块型号为 NVIDIA Corporation GP102 的 GPU。软件环境为 Debian 10 操作系统、Cuda10.2、Python3.6.5 以及 Teansorflow Keras 2.2.5 版本。

### 4.2 实验参数设置

根据以往命名实体识别相关研究<sup>[17]</sup>以及参数调优过程,具体的参数值如下:词向量维度大小设置为 50;卷积核尺寸为 2;Dropout 设置为 0.5,学习率采用 warmup 与余弦退火策略动态调整,采用随机下降梯度(SGD)进行模型参数优化,设置初始学习率为  $5e^{-4}$ 。

### 4.3 实验结果

#### 4.3.1 嵌套命名实体识别效果

该文采用准确率(Precision)、召回率(Recall)、F1 值(F1-Measure)三项基本测评指标来评估嵌套命名实体识别的效果。将修订后的《人民日报》数据集的 70% 作为训练集,30% 作为测试集,对嵌套命名实体的识别结果如表 3 所示。实验结果显示:LREM 模型对嵌套组织机构名的识别召回率比较低,导致整体组织机构名 F 值下降,而对地名的识别效果在准确率和召回率均保持稳定的效果,最后地名、组织机构名的 F 值分别是 89.05%、77.82%,LREM 模型在没有外部知识资源的情况下,全部嵌套命名实体识别的 F1 值达到 87.19%。

表 3 嵌套命名实体识别结果

类型	数量	嵌套命名实体		
		P / %	R / %	F1 / %
LREM(地名)	7 767	91.58	86.65	89.05
LREM(组织机构名)	1 875	80.79	75.06	77.82
LREM	9 642	89.81	84.72	87.19

#### 4.3.2 层数分析

表 4 显示了不同层数  $L$  对不同长度命名实体识别结果的影响。由统计结果可知,2-3 字符长度的命名实体数量占全部命名实体的 79.91%,这些命名实体

的识别结果对整体命名实体识别的效果有较大的影响。随着  $L$  的增长,模型识别效果逐渐增强,当  $L$  大于 5 时,出现过拟合现象,导致模型识别效果有所下降,但总体上依然优于普通命名实体识别效果。

表 4 不同层数对不同长度命名实体的识别结果

字符长度	数量	%		
		$L=16$ F1	$L=8$ F1	$L=5$ F1
全部	9 642	84.45	86.61	87.19
2	5 185(53.78%)	91.95	93.35	95.97
3	2 519(26.13%)	87.59	89.63	93.29
4	667(6.92%)	79.27	82.43	87.42
5	479(4.97%)	66.28	79.44	82.86
6	233(2.42%)	57.62	59.63	-
7	212(2.20%)	45.87	51.88	-
8	115(1.19%)	26.31	48.39	-
9-16	232(2.41%)	22.19	-	-

表 5 显示了在不同层数  $L$  下的推理速率。实验结果显示:推理速率随层数的增加而下降。这是因为 LREM 模型的时间复杂度为  $O(mn)$ ,其中  $m$  为字符数量, $n$  为层数。当需要识别出长字符的命名实体时,意味着需要进行更多层次的堆叠,这需要更多的计算成本。

表 5 不同层数  $L$  的推理速率对比

推理速率	Batch = 32
$L=5$	729 ms
$L=8$	856 ms
$L=16$	945 ms

#### 4.3.3 消融实验

为了证明利用短组合实体构建长组合实体表示方法的有效性,该文设置了消融实验,结果如表 6 所示。实验证明,利用低层编码层的词嵌入信息,使邻接编码层之间实现信息交换的有交互卷积方法优于直接将邻接字符卷积的无交互卷积方法。

表 6 消融实验结果

类型	利用短组合实体表示			不利用短组合实体表示		
	P / %	R / %	F1 / %	P / %	R / %	F1 / %
地名	91.58	86.65	89.05	91.67	81.69	86.39
组织机构名	80.79	75.06	77.82	80.89	66.67	73.09
全部	89.81	84.72	87.19	89.98	79.18	84.24

#### 4.3.4 非嵌套命名实体识别效果

为了对文中模型做出更客观的评价,在通用的《人民日报》数据集上进行实验,该数据集不包括嵌套命名实体,采用 LSTM-CRF、BiLSTM、BiLSTM+CRF

和 LREM 模型进行性能分析,实验结果如图 3 所示。

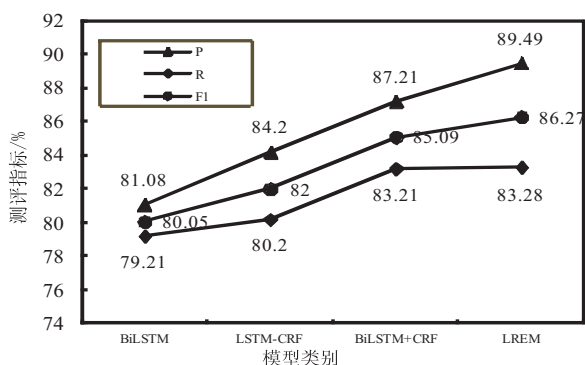


图 3 非嵌套命名实体识别结果

实验结果显示:BiLSTM+CRF 识别结果的 F1 值比 LSTM-CRF 高出 3.09%,这是因为 BiLSTM 能够利用双向结构获取上下文的语义信息,所以 BiLSTM+CRF 优于 LSTM-CRF 的识别效果;通过比较 BiLSTM 与 BiLSTM-CRF 的实验结果,增加 CRF 模块后,F1 值提高了 5.04%,这归因于 CRF 能够充分考虑标签间的联系,避免不准确标签的出现,例如:预测的标签序列为“B-Organization I-Person”,由于 LREM 模型考虑的是片段标签,无需考虑标签间的联系,所以无需增加 CRF 模块进行实验比较;与以上方法相比,该文提出的分层区域穷举模型的非嵌套命名实体识别也取得较好的效果,准确率、F 值分别比 BiLSTM-CRF 提升了 2.28%、1.18%。

## 5 结束语

基于《人民日报》数据集构建了新的嵌套命名实体数据集,提出了一种分层区域穷举模型,该模型利用多层内部解码层遍历文本中所有长度的组合实体,结合短组合实体的词嵌入信息构建长组合实体的区域表示,实现了邻接编码层的交互;使用多层解码层分别预测长度为  $L$  的命名实体防止层迷失现象发生。最后,利用分层区域穷举模型在嵌套命名实体识别数据集上进行识别。实验表明,该模型能够从所有可能的区域中识别出嵌套命名实体,且对于非嵌套命名实体识别也取得了较好的效果。

对于未来的工作有两方面考虑:(1)考虑使用字词融合信息来提高嵌套命名实体识别性能;(2)考虑嵌套命名实体与关系联合抽取,利用嵌套命名实体中存在的额外信息作为辅助特征以增强关系抽取的效果。

## 参考文献:

[1] JU M, MIWA M, ANANIADOU S, et al. A neural layered model for nested named entity recognition[C]//North American chapter of the association for computational linguistics.

tics. New Orleans: Association for Computational Linguistics, 2018: 1446-1459.

[2] SOHRAB M G, MIWA M. Deep exhaustive model for nested named entity recognition[C]//Empirical methods in natural language processing. Brussels: Association for Computational Linguistics, 2018: 2843-2849.

[3] ALEXIS M, STEPHANIE S, SHUDONG H, et al. ACE 2004 multilingual training corpus LDC2005T09 [EB/OL]. 2004. <https://catalog.ldc.upenn.edu/LDC2005T09>.

[4] WALKER C, STRASSEL S, MEDERO J, et al. ACE 2005 multilingual training corpus LDC2006T06 [EB/OL]. 2005. <https://catalog.ldc.upenn.edu/LDC2006T06>.

[5] KIM J D, OHTA T, TATEISI Y, et al. GENIA corpus a semantically annotated corpus for bio-text mining[J]. Bioinformatics, 2003, 19(suppl\_1): 180-182.

[6] RINGLAND N, DAI X, KARIMI S, et al. NNE: a dataset for nested named entity recognition in english newswire [C]//Meeting of the association for computational linguistics. Florence: Association for Computational Linguistics, 2019: 5176-5181.

[7] MUIS A O, LU W. Labeling gaps between words: recognizing overlapping mentions with mention separators[C]//Proceedings of the 2017 conference on empirical methods in natural language processing. Copenhagen: Association for Computational Linguistics, 2017: 2608-2618.

[8] ZHENG C, CAI Y, XU J, et al. A boundary-aware neural model for nested named entity recognition[C]//International joint conference on natural language processing. Hong Kong: Association for Computational Linguistics, 2019: 357-366.

[9] LI X, FENG J, MENG Y, et al. A unified MRC framework for named entity recognition [C]//Proceedings of the 58th annual meeting of the association for computational linguistics. [s. l.]: Association for Computational Linguistics, 2020: 5849-5859.

[10] FU Chunyuan, FU Guohong. Morpheme-based Chinese nested named entity recognition [C]//9th international conference on fuzzy system and knowledge discovery. Chengdu: IEEE, 2012: 2546-2550.

[11] 尹迪, 周俊生, 曲维光. 基于联合模型的中文嵌套命名实体识别[J]. 南京师大学报: 自然科学版, 2014, 37(3): 29-35.

[12] ZHANG Yue, YANG Jie. Chinese NER using lattice LSTM [C]//Proceedings of the 56th annual meeting of the association for computational linguistics. Melbourne: Association for Computational Linguistics, 2018: 1554-1564.

[13] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [C]//NIPS'13: proceedings of the 26th international conference on neural information processing systems. Red