# 多特征中文命名实体识别

胥小波[1]，王　涛[2]，康　睿[3]，周　刚[3]，李天宁[3]
(1. 中国电子科技网络信息安全有限公司，成都 610041；
2. 四川省科学技术信息研究所，成都 610016；3. 四川大学计算机学院，成都 610065)

**摘　要**：命名实体识别任务是对文本中的实体进行定位，并将其分类至预定义的类别中. 目前主流的中文命名实体识别的模型是基于字符的命名实体识别模型. 该模型在使用句法特征之前，需先进行分词，不能很好的引入句子的句法信息. 另外，基于字符的模型没有利用词典中的先验词典信息，以及中文偏旁部首蕴含的象形信息. 针对上述问题，论文提出了融合句法和多粒度语义信息的多特征中文命名实体识别模型. 实验证明论文模型相对目前主流模型有了较大的提高，同时论文还通过实验分析了各种特征对模型识别效果的影响.

**关键词**：命名实体识别；中文；多特征；自然语言处理
**中图分类号**：TP391　　**文献标识码**：A　　DOI：10.19907/j.0490-6756.2022.022003

# Multi-feature Chinese named entity recognition

*XU Xiao-Bo*[1], *WANG Tao*[2], *KANG Rui*[3], *ZHOU Gang*[3], *LI Tian-Ning*[3]
(1. China Electronic Technology Cyber Security Company Limited, Chengdu 610041, China;
2. Institute of Science and Technology Information of Sichuan, Chengdu 610016, China;
3. College of Computer Science, Sichuan University, Chengdu 610065, China)

**Abstract**: The task of named entity recognition is to locate the entities in the text and classify them into predefined categories. The current mainstream Chinese named entity recognition models are character-based named entity recognition models which word segmentation is required before using syntactic features, syntactic information of sentences cannot be well utilized as a result. In addition, the character-based models cannot make use of the prior dictionary information and the pictographic information contained in Chinese radicals. To solve the above problems, this paper proposes a multi-feature Chinese named entity recognition model combining syntax and multi-granularity semantic information. The experiments demonstrate that the proposed model is better than the current mainstream Chinese named entity recognition models, the influence of various features on the Chinese entity recognition effect is analyzed through experiments as well.

**Keywords**: Name entity recognition; Chinese; Multi-feature; Natural language processing

## 1　引　言

命名实体识别任务旨在对文本中特定类别的实体进行定位并且将其分类为预先定义的类别，是信息抽取领域的一个子任务. 由于中文本身的一些特点，中文命名实体识别任务在方法上与英文命

**收稿日期**：2021-11-15
**基金项目**：国家自然科学基金(62137001，JG2020125)
**作者简介**：胥小波(1985－)，男，博士，研究方向为网络安全. E-mail：xxb202111@163.com
**通讯作者**：周刚. E-mail：zhougang@scu.edu.cn

名实体识别任务有所区别. 其中,中文的词之间没有天然的边界,如果以词作为基本单位,需要先对中文进行分词,但是分词工具会发生错误,并且有些错误是不可能通过之后模型的训练来缓解的. 例如,在"南京市长江大桥"这句话中,正确的实体标记是"南京"为地名(LOC),"江大桥"是人名(PER),如果分句工具将句子分为"南京市|长江大桥",那么模型就不可能识别出正确的命名实体. 为了避免上述的情况发生,如今主流的中文命名实体识别模型都是将字作为基本单位,但是基于字的命名实体识别模型也存在一些问题. 首先,由于中文句法分析需要先进行分词,所以基于字符的模型并不能很好地引入句子的句法信息,显式的引入句法信息的方法跟基于字的中文命名实体识别模型并不契合;然后,从向量嵌入的角度,字嵌入包含的特征信息没有词嵌入丰富,例如对于"长"这个字,在基于字的命名实体识别模型中,无论是在"市长"还是在"长江大桥",它最初的表示都是相同的,没有引入能够区分的先验信息. 但是,如果是基于词的命名实体识别模型"市长"和"长江大桥"拥有不同的表示,在最初就包含了更多的先验信息,同时还引入了一定的边界信息. 所以,基于字的命名实体识别模型缺少词的先验信息,需要通过一些方法将词典的信息融入模型中. 最后,中文属于象形文字,文字的偏旁部首体现了文字的含义,例如"江"和"河"都包含三点水,暗示了这两个字都与水有关,所以汉字的结构中暗含了这个字的属性. 通过对文字的结构进行编码,不仅能够提取部分字形方面的特征,也可以解决一些 OOV 问题.

## 2　相关工作

命名实体识别方法主要归为 3 类:基于规则[1]、基于传统机器学习[2]和基于深度学习的方法. 其中,由于具有自动捕获输入句子特征,实现端到端的命名实体识别的优点,基于深度学习的方法已经成为了近年的研究热点.

近年来,在命名实体识别任务中,一些研究工作很好地应用了基于深度学习的方法. Huang 等[3]将双向长短期记忆网络(Bi-directional Long Short-Term Memory, BiLSTM)和条件随机场网络应用于命名实体标记. 但由于 BiLSTM 对长序列的编码能力有限,并且在计算速度方面表现不佳. Strubell 等[4]将卷积神经网络(Convolutional Neural Network, CNN)用于命名实体识别. 卷积神经网络相比 BiLSTM 之类的循环神经网络具有更快的计算速度. 然而,卷积神经网络虽然具有很好的局部捕捉能力,但是会损失大量的全局信息. 已有研究对 Transformer 编码器进行改进,补充了方向信息,提升了 Transformer 编码器的编码能力[5]. 但是浅层的 Transformer 编码器具有编码层级结构能力不足的缺点.

由于中文本身的特殊性,中文命名实体识别相比起步更早的英文命名实体识别更具有挑战性. Yang 等[6]在对单词序列进行标注之前用分词工具对中文句子序列进行分词. 然而,分词工具不可避免地会出现单词的错误划分,造成实体边界的错误识别. 这是因为不同于英文中采用分隔符来标识词与词之间的边界,中文句子中的词没有天然的边界,在分词方面更加困难. 词增强的方法可以减少分割错误并增加中文语义和边界信息,但是 Wu 等[7]表示这种方法忽略了汉字结构中的信息,并提出了一种将汉字特征和部首信息相结合的多元数据嵌入进行改善. 为了解决这一边界问题,已有研究表明字符级别的命名实体识别的效果比单词级别的更好[8,9]. 但是,基于字符的命名实体识别存一个很明显的缺点就是损失了单词中丰富的信息. 因此,将词典信息充分融入字符模型中是中文命名识别的一大研究热点. 文献[10]提出的 Lattice-LSTM 模型结合词典信息提升了模型识别能力. 具体来说,该模型利用长短期神经网络的门控机制来自动匹配句子中每个字符对应的单词,将词典中与句子语义最匹配的单词信息融入句子表示中. Li 等[11]采用"五笔画"编码方式表示汉字结构模式改进汉字的字形嵌入,提高在中文命名实体识别的整体表现. Xu 等[12]表示中文字符部首蕴含的特征信息也能够帮助提升识别命名实体的能力. 该工作提出在模型中同时使用字、词和部首 3 种不同粒度级别的嵌入能够丰富句子中的字符表示,并且验证了部首信息的有效性. 这些工作的成功也印证了中文多级别特征信息的有效性. 虽然,现已有研究将中文的语法信息用于命名实体识别任务中,例如 Nie 等[13]研究句法信息对命名实体识别模型的影响,通过键-值记忆网络来提取词性、句法成分信息和句法依存信息,然后通过门控的方式对提取出来的信息进行筛选和融合. 由于该方法需要先对输入文本进行句法分析,对中文进行句法分析首先需要进行分词,与基于字符的模型并不契合.

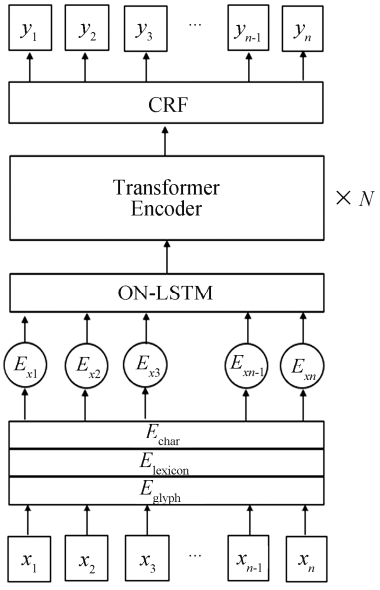


图 1　模型整体框架图

Fig. 1　Model overall framework diagram

## 3　本文方法

多特征中文命名实体识别模型在表示层增加了词典特征和字形特征来融入先验的词信息和中文的象形特征；在编码层的底层首次使用 ON-LSTM 对表示层的输出进行编码来隐式的捕获层次结构信息，引入语法归纳偏置. 本节具体介绍在表示层引入词典特征和字形特征的方法以及对 ON-LSTM 的介绍.

### 3.1　表示层

表示层由 3 部分表示组成，分别是字符表示、词表示和字形表示. 其中，字符表示使用预训练的词嵌入，通过查表的方式匹配到每个字符对应的向量表示，维度为$\mathbb{R}^d$. 词典嵌入采用 Ma 等[14]提出的 SoftLexicon 方法，该方法不仅引入了字符在对应词中的位置信息，同时还引入了其在各个位置对应词的嵌入向量. 具体步骤为：首先统计文本序列中每个字符对应的{B,M,E,S}4 个集合，这 4 个集合即包含了字符在词中的位置信息还包含了对应词的嵌入表示；在统计了每个字符对应的 4 个集合表示后，通过加权平均的方式计算每个集合对应的集合向量表示. 如式(1)和式(2)所示，是集合向量表示的计算方法，其中 $z(w)$表示词 $w$ 出现的频率，$e^w(w)$是词 $w$ 对应词向量，$Z$ 表示 4 个集合中词 $w$ 出现频率的总和；最后，如式(3)所示，将 4 个集合的向量表示进行拼接组成该字符对应的词典嵌入表示.

$$v(S)=\frac{4}{Z}\sum_{w\in S}z(w)\,e^w(w) \tag{1}$$

$$Z=\sum_{w\in B\cup M\cup E\cup S}z(w) \tag{2}$$

$$e(B,M,E,S)=[v(B);v(M);v(E);v(S)] \tag{3}$$

为了引入字形特征，本章分别采用了两种模型结构对字符的偏旁序列进行编码. 第 1 种方式如图 2 所示，使用双向 LSTM 对字符的偏旁序列进行编码，然后取前向和后向的最后一个隐藏向量进行拼接，来表示这个字符的字形特征.

$$e=\mathrm{Em\,b}^{\mathrm{radical}}(x) \tag{4}$$

$$h^{FW}=\mathrm{LST\,M}^{FW}(e) \tag{5}$$

$$h^{BW}=\mathrm{LST\,M}^{BW}(e) \tag{6}$$

$$x^r=[h_{-1}^{FW};h_{-1}^{BW}] \tag{7}$$

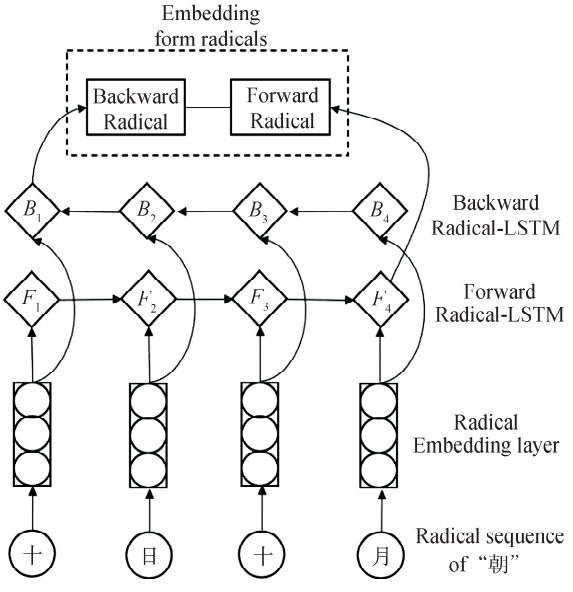


图 2　双向循环神经网络编码偏旁序列

Fig. 2　Bidirectional recurrentneural network coding radical sequence

第 2 种方式如图 3 所示，使用多尺度的卷积神经网络对字符的偏旁序列进行编码. 分别使用跨度为 2 和 3 的两种卷积核对偏旁序列进行卷积，然后将两种尺度的卷积核卷积出来的特征向量进行拼接，最后采用最大池化获得这个字符的象形特征.

$$h^2=\mathrm{Con}\,v^2(e) \tag{8}$$

$$h^3=\mathrm{Con}\,v^3(e) \tag{9}$$

$$x^r=\mathrm{Max}p([h^2;h^3]) \tag{10}$$

在分别得到字符表示、词典表示和字形表示后，模型对这 3 种表示进行拼接，如式(11)所示，其中 $x^c$ 是字符表示，$e(B,M,E,S)$是词典表示，$x^r$ 是字形表示.

$$x \leftarrow [x^c; e(B,M,E,S); x^r] \tag{11}$$

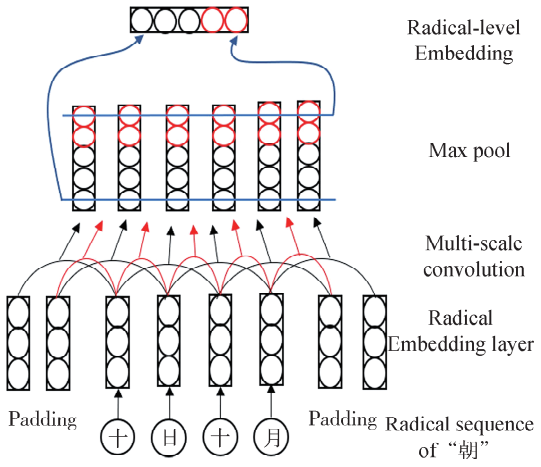


图 3　多尺度卷积神经网络编码偏旁序列

Fig. 3　Multi-scale convolutional neural network coding radical sequence

### 3.2　编码层

编码层分为两个部分，分别是 ON-LSTM 和改进后的 Transformer 编码器，其中 ON-LSTM 用来归纳语法信息；Transformer 编码器用来捕获长距离的依赖，从多个特征子空间提取更丰富的特征.

ON-LSTM 是 Shen 等[15]对 LSTM 网络的改进，由于文本序列并不是从左到右的链式结构，而是一种树形层级结构，从顶层到底层粒度不断减小. 编码器对这种树形结构信息的捕获对命名实体的识别有辅助作用，因为命名实体一般都是一个词或者多个词组成的短语. 如图 4 所示，ON-LSTM 可以学习这种树形层级结构，ON-LSTM 在编码时对神经元进行了排序，并且划分了不同的层级，高层级表示粗粒度，低层级表示细粒度. 在编码过程中，高层级的信息更新的频率低，所以保留了更多的历史信息，但是低层级的信息容易受到当前输入的影响，更新的频率高.

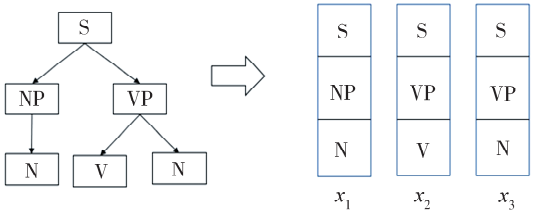


图 4　ON-LSTM 表示的树形结构

Fig. 4　The tree structure represented by ON-LSTM

如图 5 所示，ON-LSTM 的网络结构和原始的 LSTM 大体相同，不同的是记忆单元 $\hat{c}_t$ 到 $c_t$ 的更新机制. $c_t$ 在初始化时是一个 0 向量，代表没有存储任何记忆信息，在对 $c_t$ 进行更新的过程中，模型需要权衡历史信息和当前信息的层级. 在每个时间步对 $c_t$ 进行更新时，首先需要预测历史信息 $h_{t-1}$ 的层级 $d_f$ 和当前输入 $x_t$ 的层级 $d_i$. 如果 $d_f \leqslant d_i$，则表示当前输入 $x_t$ 的层级要高于历史记录 $h_{t-1}$ 的层级，这表示需要将当前输入的信息整合到高于等于 $d_f$ 的层级中，如式(3)～式(12)所示. 由于当前输入的层级更高，对 $[d_f, d_i]$ 区间内的历史信息产生了影响，所以在更新时需要同时考虑当前输入信息和历史信息，其更新公式与普通 LSTM 的更新公式相同；对于小于 $d_f$ 的区间，直接使用当前输入 $\hat{c}_t$ 对应的区间来覆盖；大于 $d_i$ 的区间，需要保留原来的历史信息，即 $c_{t-1}$ 对应的区间.

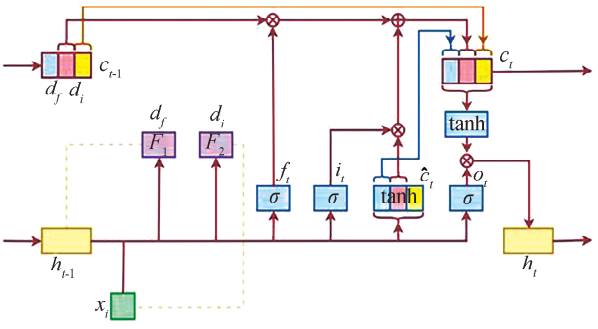


图 5　ON-LSTM 结构

Fig. 5　Structure of ON-LSTM

$$c_t = \begin{pmatrix} \hat{c}_t[0:d_f] \\ f_t[d_f:d_i]^\circ c_{t-1}[d_f:d_i] + i_t[d_f:d_i]^\circ \hat{c}_t[d_f:d_i] \\ c_{t-1}[d_i:] \end{pmatrix} \tag{12}$$

当 $d_f > d_i$，代表当前输入 $x_t$ 的层级低于历史记录 $h_{t-1}$ 的层级，那么对于 $(d_f, d_i)$ 的区间就会保持初始状态；低于 $d_f$ 的区间 $[0, d_f]$ 写入当前的输入信息，高于 $d_i$ 的区间 $[d_i:]$ 保留历史信息不变.

其中，$d_i$ 和 $d_f$ 如果直接输出整数可能导致模型不可导，所以，ON-LSTM 通过分段软化的方式来表示 $d_i$ 和 $d_f$. 模型通过 softmax 函数来近似两个 one-hot 向量 $1_{d_f}$ 和 $1_{d_i}$，然后通过式(13)～式(16)来实现上述的工作原理，其中 $\overrightarrow{cs}$ 和 $\overleftarrow{cs}$ 是两个方向上的 cumsum 函数.

$$\tilde{f}_t = \overrightarrow{cs}(1_{d_f}) \tag{13}$$

$$\tilde{i}_t = \overleftarrow{cs}(1_{d_i}) \tag{14}$$

$$w_t = \tilde{f}_t^\circ \tilde{i}_t \tag{15}$$

$$c_t = w_t^\circ (f_t^\circ c_{t-1} + i_t^\circ \hat{c}_t) + (\tilde{f}_t - w_t)^\circ c_{t-1} + (\tilde{i}_t - w_t)^\circ \tilde{c}_t \tag{16}$$

模型可以通过使用 ON-LSTM 来编码这种层次结构信息，从神经元底层的细粒度到顶层的粗粒度，学习层级句法结构来辅助命名实体识别.

## 4　实验结果

### 4.1　数据集与评价指标

本实验在两个领域的中文命名实体识别数据集进行验证，分别是网络社交领域的数据集 Weibo 和新闻领域的数据集 CLUENER. 同时，这两个数据集有不同的粒度. CLUENER 是细粒度命名实体识别数据集. 该数据集共有 10 种不同的实体类别，除了人名、组织等常见类别外，还有一些在常见类别基础上进行了细粒度划分得到的类别. 这就造成了细粒度类别与划分前的常见类别之间混淆程度较高. 比如从“组织”中划分出了“政府”和“公司”. 相比起区分“人名”和“组织”，区分“组织”、“政府”和“公司”会更加困难. 此外，CLUENER 数据集中的同一实体所属类别在不同语境下也会有变化. 如某个属于“游戏”的实体在其他语境中可能指的是改编的某个“书籍”或者“电影”. 如表 1 所示，实体“《黑暗之塔》”在第 2 个句子中是一个“游戏”，然而第 2 个和第 3 个句子中代表的是一本“书籍”.

Weibo 数据集包括人名、地名、机构和地缘政治实体，从社交网络数据中获得，存在一些非规范用词和语法不规范的情况，如表 2 所示. 微博用户为了加强情感经常使用叠字、叠词的方式写评论，增加了文章的不规范性.

**表 1　CLUENER 数据集中实体在不同的语境属于不同类别的样例**

Tab. 1　Examples of entities in the CLUENER data set belonging to different categories in different contexts

| 句子 | 标签 |
|---|---|
| 《黑暗之塔》改编游戏将在 2013 年 5 月随电影版同步上市 | 游戏:《黑暗之塔》 |
| 斯蒂芬金的《黑暗之塔》小说共有七卷本，这个系列是斯蒂芬金最负盛名的小说 | 书籍:《黑暗之塔》<br>姓名：斯蒂芬金 |
| 蒂芬金《黑暗之塔》将改编成游戏 | 书籍:《黑暗之塔》<br>姓名:蒂芬金 |

### 4.2　参数设置

本实验在 Colab pro p100 16 GB 内存的环境下进行. 在该实验环境下，本文模型参数设置如表 3 所示，其中优化器使用动量梯度下降的优化器，该优化器相对于普通的梯度下降优化器具有减少震荡和加速收敛的作用，动量参数 momentum 设置为默认值 0.9. 同时在计算梯度时，对梯度值进行裁剪，限制梯度值在[−5.0，5.0]之间，防止梯度爆炸.

**表 2　Weibo 数据集用词和语法不规范的样例**

Tab. 2　Examples of irregular words and grammar in the Weibo dataset

| 不规范示例 | 标签 |
|---|---|
| 我也觉得只有你，会想我花心花心花心花心小的妈咪无时无刻都在想你爱你爱你爱你暗英小英转发微博 | PER:妈咪 |
| 必须转啊啊宇多田光之樱流泪泪泪明日香啊丽 | PER:宇多田光;明日香 |

**表 3　参数取值**

Tab. 3　Parameter value

| 参数 | 取值 |
|---|---|
| Epoch | 100 |
| Batchsize | 16 |
| 嵌入层 dropout | 0.3 |
| 优化器 | SGD+momentum |
| 梯度裁剪 | 5.0 |
| warmup_proportion | 0.01 |

本文模型本实验采用分层优化的方式，对不同层使用不同的学习率进行参数训练，由于训练条件随机场层需要比其他层更大的学习率才能充分的学习各个标签之间的约束关系，即转移矩阵中各个标签之间的转移概率. 所以，本实验将条件随机场层的学习率设置的大于其他层的学习率，减少一些不可能的标签序列的出现.

### 4.3　实验结果

本文模型在 CLUENER 数据集上的训练过程如图 6 所示，图中横坐标为模型的迭代轮次，纵坐标为模型的 $F_1$ 值. 该图展示了在验证集上加入句法信息和不加入句法信息的两个模型的 $F_1$ 值的变化情况. 很明显可以看出加入语法信息的模型在训练的前期收敛更快，同时模型的最终收敛的效果也要好于无语法信息的模型.

表 4 和表 5 分别是论文模型与近年多个模型的在两个数据集上的实验结果对比. 由表 4 的实验结果可以看出，论文模型在 Weibo 数据集上取得了 63.61%的 $F_1$ 值，分别超过 TENER 和 SoftLexicon(LSTM)模型 5.44%和 2.19%. 其主要原因是，论文模型同时结合了这两个模型在嵌入层和编码层的优势，在嵌入层增加了词典特征，在编码层使用改进后的 Transformer 编码器直接建立任意

位置两两字符之间的联系. 另外,论文模型还在编码器的底层加入 ON-LSTM 弥补 Transformer 编码器编码树形层次结构信息能力不足的缺点,进一步提升了模型的识别效果.

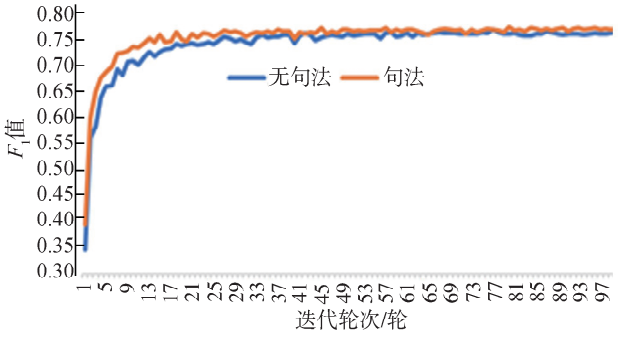


图 6　模型在验证集上的 $F_1$ 值对比

Fig. 6　Comparison of $F_1$ values of the model on the validation set

**表 4　各个模型在 Weibo 数据集上的结果对比**

Tab. 4　Comparison of the results of each model on the Weibo data set

| 模型 | 测试集($F_1$) |
|---|---|
| CAN-NER[16] | 59.31 |
| TENER[5] | 58.17 |
| SoftLexicon(LSTM)[14] | 61.42 |
| Our | 63.61 |

**表 5　各个模型在 CLUENER 数据集上的结果对比**

Tab. 5　Comparison of the results of each model on the CLUENER data set

| 模型 | 测试集($F_1$) |
|---|---|
| LSTM[15] | 70.0 |
| TENER[5] | 72.49 |
| Our | 76.93 |

表 6 是多特征模型的消融实验表,对比表中的第 1 行和第 2 行,可以得出在底层加入 ON-LSTM,引入句法偏置可以提高模型的识别效果;对比表中的第 1 行和第 3 行,可以得出增加词典特征可以提升命名实体识别的效果,说明字典提供的先验的语义信息的有效性;对比表中的第 3 行和第 4 行,可以的得出融合 ON-LSTM 编码的语法信息和嵌入层的词典信息有相互促进的作用,可以提高命名实体识别模型的识别效果. 表 7 在 Weibo 数据集上进行实验,加入了字形特征的命名实体识别模型的实验结果.

**表 6　模型消融对比实验表**

Tab. 6　Comparison experiment tableof model ablation

| 模型 | Weibo | | CLUNER | |
|---|---|---|---|---|
| | $P$ | $R$ | $F_1$ | $F_1$ |
| Transformer(baseline)[5] | 62.60 | 58.85 | 60.67 | 72.49 |
| Transformer+ON-LSTM | 66.20 | 57.17 | 61.36 | 74.52 |
| Transformer+lexicon | 69.86 | 56.69 | 62.28 | 76.77 |
| Transformer+lexicon+ON-LSTM | 70.23 | 58.13 | 63.61 | 76.93 |

**表 7　字形特征对实验结果的影响**

Tab. 7　The influence of glyphfeatureon the experimental results

| 模型 | 验证集($F_1$) | 测试集($F_1$) |
|---|---|---|
| Transformer | 61.52 | 54.99 |
| Transformer+radical(CNN) | 59.55 | 55.57 |
| Transformer+radical(LSTM) | 59.97 | 56.56 |
| Transformer+bichar[5] | 63.15 | 60.58 |
| Transformer+bichar+radical(CNN) | 63.61 | 60.80 |
| Transformer+bichar+radical(LSTM) | 62.34 | 59.28 |

从表 7 可知,字形特征在基于 Transformer 编码器的命名实体识别模型上的作用,以及不同网络结构编码的字形特征的效果. 表 7 中 radical(CNN)是使用多尺度的卷积神经网络编码偏旁序列,radical(LSTM)是使用双向 LSTM 来编码偏旁序列,bichar 是引入了 n-gram 特征. 从表 7 前 3 行可以看出,加入字形特征在一定程度上可以增加模型的鲁棒性. 从表 7 后 3 行可以看出,在加入 n-gram 特征后,使用多尺度卷积神经网络编码的字形特征对模型的识别效果有提升,但是效果并不明显.

## 5　结　论

论文主要分析了 Transformer 模型编码能力的不足,由于句子拥有树形的语法结构,从单独的字组成词,再从词组成短语,最后形成一个句子,这种分层级的组成结构无法通过单层的 Transformer 模型来表示,甚至浅层的 Transformer 模型也存在编码树形结构能力不足的情况. 论文通过在编码器的底层增加有序神经元的 LSTM 来编码句子的层级结构,弥补了 Transformer 编码器编码句法信息能力不足的缺点. 另外,论文研究了多特征对中文命名实体识别的影响,包括先验的字典特征,句法成分特征以及字形特征,以及各种特征叠加后的模型效果,从本文的实验可以看出词典信息和句法信息可以同时辅助模型进行命名实体识别,字形特征对模型的效果影响微弱,甚至有时会产生负面效果.

## 参考文献：


[1] 向晓雯，史晓东，曾华琳. 一个统计与规则相结合的中文命名实体识别系统[J]. 计算机应用，2005，25：2404.

[2] 张传岩，洪晓光，彭朝晖，等. 基于 SVM 和扩展条件随机场的 Web 实体活动抽取[J]. 软件学报，2012，23：2612.

[3] Huang Z，Xu W，Yu K. Bidirectional LSTM-CRF models for sequence tagging [EB/OL]. [2015-08-09]. https://arxiv.org/abs/ 1508.01991.

[4] Strubell E，Verga P，Belanger D，*et al*. Fast and accurate entity recognition with iterated dilated convolutions [C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen：Association for Computational Linguistics，2017.

[5] Yan H，Deng B C，Li X N，*et al*. TENER：adapting transformer encoder for name entity recognition [EB/OL]. [2019-12-10]. https:// arxiv. org /abs/ 1911.04474.

[6] Yang J，Teng Z Y，Zhang M S，*et al*. Combining discrete and neural features for sequence labeling [C]//Proceedings of the 17th International Conference on Intelligent Text Processing and Computational Linguistics. Konya：Springer，2016.

[7] Wu S，Song X，Feng Z. MECT：Multi-Metadata embedding based cross-transformer for Chinese named entity recognition [C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. Bangkok：Association for Computational Linguistics，2021.

[8] Liu Z X，Zhu C H，Zhao T J. Chinese named entity recognition with a sequence labeling approach：based on characters，or based on words [C]//Proceedings of the 6th International Conference on Intelligent Computing. Changsha：Springer，2010.

[9] Lu Y N，Zhang Y，Ji D H. Multi-prototype chinese character embedding [C]//Proceedings of the Tenth International Conference on Language Resources and Evaluation. Portorož：European Language Resources Association，2016.

[10] Zhang Y，Yang J. Chinese ner using lattice lstm [C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL). Melbourne：Association for Computational Linguistics，2018.

[11] Li J，Meng K. MFE-NER：Multi-feature fusion embedding for Chinese named entity recognition [EB/OL]. [2021-09-16]. https:// arxiv. org /abs/ 2109.07877.

[12] Xu C W，Wang F Y，Han J L，*et al*. Exploiting multiple embeddings for Chinese named entity recognition [C]//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. New York：Association for Computing Machinery，2019.

[13] Nie Y，Tian Y，Song Y，*et al*. Improving named entity recognition with attentive ensemble of syntactic information [C]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Punta Cana：Association for Computational Linguistics：Findings，2020.

[14] Ma R T，Peng M L，Zhang Q，*et al*. Simplify the usage of lexicon in Chinese NER[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Seattle：Association for Computational Linguistics，2020.

[15] Shen Y，Tan S，Sordoni A，*et al*. Ordered neurons：integrating tree structures into recurrent neural networks [C]//Proceedings of International Conference on Learning Representations. Miyazaki：European Language Resources Association，2019.

[16] Zhu Y，Wang G，Karlsson B F. CAN-NER：convolutional attention network for Chinese named entity recognition [C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics：Human Language Technologies. Minneapolis：Association for Computational Linguistics，2019.



**引用本文格式：**

中　文：胥小波，王涛，康睿，等. 多特征中文命名实体识别[J]. 四川大学学报：自然科学版，2022，59：022003.

英　文：Xu X B，Wang T，Kang R，*et al*. Multi-feature Chinese named entity recognition[J]. J Sichuan Univ：Nat Sci Ed，2022，59：022003.