

# 基于BERT的多特征融合中文命名实体识别

孙璐冰<sup>1</sup>, 康怡琳<sup>1</sup>, 王俊<sup>1</sup>, 朱容波<sup>2\*</sup>

(1 中南民族大学 a. 计算机科学学院; b. 国家民委信息物理融合智能计算重点实验室, 武汉 430074; 2 华中农业大学 信息学院, 武汉 430070)

**摘要** 针对中文命名实体识别往往需要引入外部知识获取深层语义信息, 以及基于RNN结构的模型对空间特征提取效果不佳等问题, 提出了基于BERT的多特征融合中文命名实体识别模型. 通过BERT模型获取输入文本序列的深层次语义信息, 利用MHSA模块和IDCNN模块增强特征提取能力. 前者利用相对位置编码和多头自注意力机制来捕获输入序列的隐藏特征, 使模型能够考虑到字符间的距离方向信息; 后者则可以对空间特征建模, 获得输入序列的全局信息. 通过将两个模块的输出特征进行连接, 增强模型性能. 实验结果表明: 模型在MSRA、Resume和Weibo三个公共数据集上的F1值分别达到了95.12%、95.45%和66.14%, 优于其它最新模型, 验证了模型在中文命名实体识别上的有效性.

**关键词** 自然语言处理; 中文命名实体识别; BERT模型; 迭代膨胀卷积神经网络; 自注意力

**中图分类号** TP389.1 **文献标志码** A **文章编号** 1672-4321(2025)01-0068-07

**doi**: 10.20056/j.cnki.ZNMDZK.20250109

## BERT-based multi-feature fusion for Chinese named entity recognition

SUN Lubing<sup>1</sup>, KANG Yilin<sup>1</sup>, WANG Jun<sup>1</sup>, ZHU Rongbo<sup>2\*</sup>

(1 South-Central Minzu University, a. School of Computer Science; b. Key Laboratory of Cyber-Physical Fusion Intelligent Computing of State Ethnic Affairs Commission, Wuhan 430074, China; 2 School of Information, Huazhong Agricultural University, Wuhan 430070, China)

**Abstract** Given the problems that external knowledge is introduced into most Chinese named entity recognition methods to obtain deep semantic information, and that models based on RNN structure are weak in extracting spatial features, a BERT-based multi-features fusion model for Chinese named entity recognition is proposed. The deep semantic information of the input sequence is obtained based on BERT, and the feature extraction capability is enhanced by using MHSA and IDCNN module. The former employs relative position encoding and multi-head self-attention mechanism to capture potential features of the input sequence, so that the distance and direction information between characters can be obtained. The latter module can model spatial features to get the global information of the input sequence. The output features of the two modules will be concatenated to improve the performance of the model. Experimental results show that F1 scores of the proposed model on MSRA, Resume and Weibo can reach 95.12%, 95.45%, and 66.14% respectively, which outperforms other latest methods and validates the effectiveness of the proposed model for Chinese named entity recognition.

**Keywords** natural language processing; Chinese named entity recognition; BERT model; iterated dilated convolutional neural network; self-attention

命名实体识别(Named Entity Recognition, NER)作为自然语言处理领域中的关键技术之一, 主要目标是从文本中找到潜在的实体, 识别出诸如人名、

地名、机构名和专有名词等实体类型. 由于其在关系抽取<sup>[1]</sup>、信息检索和问答系统等自然语言处理(Natural Language Processing, NLP)任务中具有重要

收稿日期 2022-09-28

\* 通信作者 朱容波(1978-), 男, 教授, 博士, 研究方向: 人工智能物联网、边缘计算、大数据智能, E-mail: rbzhu@mail.hzau.edu.cn

基金项目 国家重点研发计划资助(2020YFC1522900)

作用,它受到了广泛研究.

近年来,随着深度学习的发展,具有时序特征的循环神经网络(Recurrent Neural Network, RNN)模型广泛应用于NLP领域.长短记忆网络在英文NER任务中取得了出色的表现.MA等<sup>[2]</sup>提出了一种双网络NER模型,该模型结合双向长短记忆网络(Bi-directional Long Short-Term Memory, BiLSTM)和卷积神经网络(Convolutional Neural Networks, CNN),使用CNN训练词向量,将训练后结果作为BiLSTM网络的输入,并利用条件随机场(Conditional Random Field, CRF)建立标签关系,在两个英文数据集上取得了较好的性能.QIN等<sup>[3]</sup>构建了一个与特征模板相结合的NER模型,并使用人工特征模板提取文本的局部特征,在大规模网络安全数据上取得了良好的效果.在大语料库训练的背景下,预训练模型BERT<sup>[4]</sup>由于其出色的文本表示与特征提取能力,在NER以及其他自然语言处理任务中取得了较好的效果.

然而,由于中文NER比英文发展晚,可用的语料少,语料库领域比较单一,其发展受到限制.与英文NER任务相比,中文NER的难点主要集中在英文语料中,单词之间有空格作为分隔,而中文的词语之间没有分割符号,会对实体边界的识别造成困难,如果主动进行分词,则可能导致分词错误传播.为了解决上述问题,许多研究将深度学习技术应用到了中文命名实体识别任务中<sup>[5-6]</sup>.通过深度学习对文本数据进行特征提取,不仅可以避免繁琐的手工特征提取,还可以有效提高模型的实体识别能力<sup>[7]</sup>.

ZHANG等<sup>[8]</sup>提出了基于格结构的长短记忆网络模型,该模型利用外部词典,通过设计一种格结构引入句子中的潜在词汇,为模型提供词边界信息,结合BiLSTM模型,提高了实体识别的性能.LIU等<sup>[9]</sup>将单词的语义信息添加到整个字符的开头或结尾,以增强语义表示,通过BiLSTM模型编码,在公开的微软语料库MSRA上获得了93.74%的F1值.WU等<sup>[10]</sup>通过CNN网络提取汉字拆分后的最小字根特征,并利用汉语文字的部首信息来提高模型的性能.尽管上述方法通过引入词汇、词根等外部信息,增强了句子嵌入的表示能力,在中文NER任务上取得较好的效果,但同时增加了特征的维数和计算量,并且庞大的外部词典通常只对公共词汇有帮助,有一定局限性.王蕾等<sup>[11]</sup>基于BiLSTM和双向RNN提出组合网络,应用分段深度学习结构提取文本深度特征,模型得到较好的识别效果.张召武等<sup>[12]</sup>

对向量表示层进行改进,结合双向门控循环单元(Bi-directional Gate Recurrent Unit, BiGRU)模型进行特征提取,提高了实体识别准确率.而BiLSTM、BiGRU和BiRNN等RNN结构的网络模型虽然能够很好地建模文本序列,但对整体的空间特征感知能力较弱,容易忽略掉句子中一些潜在信息.

针对中文NER任务引入外部信息会增加计算复杂度且RNN结构的模型空间特征提取能力较弱等问题,本文提出了一种基于BERT的多特征融合中文命名实体识别模型(BERT-Multi-Head Self-Attention-Iterated Dilated Convolutional Neural Network-CRF, BERT-MHSA-IDCNN-CRF).首先,为了保证在不引入外部信息的情况下能够得到较好的语义表示,本文利用BERT模型获得输入序列的字符嵌入.其次,将自适应多头自注意力机制(Multi-Head Self-Attention, MHSA)与迭代膨胀卷积神经网络(Iterated Dilated Convolutional Neural Network, IDCNN)的特征输出相融合,增强模型实体识别能力.其中,自适应多头自注意力机制把相对位置编码融入字符向量,在输入中增加了字符间距离方向等时序信息,通过自适应模块使模型更好地学习提取到的特征.迭代膨胀卷积网络则用来对输入文本序列的空间特征进行建模.结果表明,BERT-MHSA-IDCNN-CRF模型在多个中文命名实体识别数据集上取得了良好的效果.

## 1 基于BERT的多特征融合模型

模型主要分为三部分,从输入到标签输出分别是嵌入层、编码层和解码层.嵌入层主要由BERT组成,文本序列输入后,经过BERT得到蕴含深层语义信息的嵌入表示.编码层由自适应MHSA和IDCNN模块组成,在自适应MHSA模块中,字符嵌入将和相对位置编码进行连接,先后通过多头自注意力机制和前馈神经网络提取特征,自适应模块和层正则化分别用来选择性融合浅层特征和优化训练;在IDCNN模块中,字符嵌入将直接输入IDCNN网络,通过多层膨胀卷积块对空间特征建模,提取到的特征将与自适应MHSA模块的输出进行融合.解码层由CRF组成,编码层融合后的输出通过CRF解码,得到最终预测标签序列.模型架构如图1所示.

### 1.1 字符嵌入

对于输入句子 $S = [c_1, c_2, \dots, c_n]$ ,  $c_i$ 表示句子中

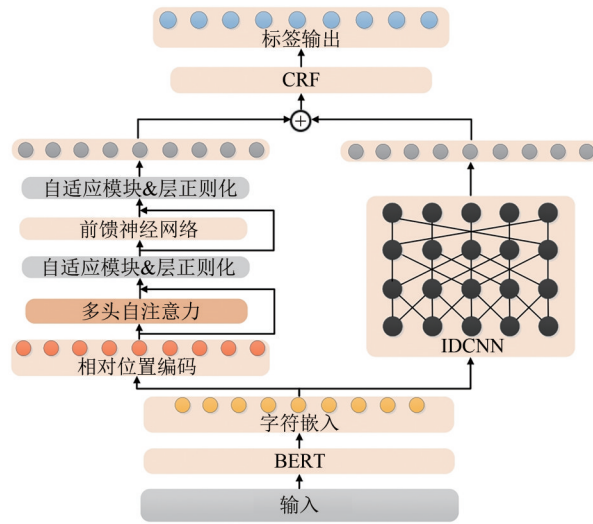


图1 模型总体架构

Fig. 1 Overall architecture of the model

单个字符,通过BERT模型得到字符级别的嵌入表示  $E$ ,公式如下:

$$\mathbf{x}_i = B^c(c_i), \mathbf{x}_i \in \mathbb{R}^d, \quad (1)$$

$$\mathbf{E} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{n \times d}, \quad (2)$$

公式(1)中  $B^c$  表示 BERT 嵌入,  $\mathbf{x}_i$  为单个字符嵌入表示.

### 1.2 自适应多头自注意力机制

嵌入层输出字符嵌入后,首先与位置嵌入结合.因为相对位置信息具有方向感知的能力,优于绝对位置编码.所以该模块通过融入 Transformer-XL<sup>[13]</sup> 中相对位置编码,得到文本序列的位置嵌入,计算方式如下:

$$d_{i,t+k} = \text{pos}(t) - \text{pos}(t+k), \quad (3)$$

$$\mathbf{R}_{i,j} = [\dots \mathbf{PE}_{d,2i} \mathbf{PE}_{d,2i+1} \dots]^\top, \quad (4)$$

其中:  $d_{i,t+k}$  为字符  $c_i$  和  $c_{t+k}$  之间的距离,  $\mathbf{R}_{i,j}$  为相对位置嵌入.编码内部对应维度  $\mathbf{PE}$  的计算方式为:

$$\mathbf{PE}_{d,2i} = \sin(d/10000^{2i/d_{\text{model}}}), \quad (5)$$

$$\mathbf{PE}_{d,2i+1} = \cos(d/10000^{2i+1/d_{\text{model}}}), \quad (6)$$

其中:  $d$  等同于公式(3)中  $d_{i,t+k}$ ,  $i$  为位置嵌入的维度.对句子中字符位置进行编码,能够把字符间距离和方向等相对位置信息融入字符嵌入.

通过多头自注意力机制,含有位置信息的字符嵌入  $\mathbf{H} \in \mathbb{R}^{l \times d}$  将获得序列中每个单词之间的相关性并捕捉整个序列的隐藏特征.其中  $l$  为序列长度,  $d$  为输入.通过三个可学习的参数矩阵  $\mathbf{W}_q, \mathbf{W}_k, \mathbf{W}_v$ , 矩阵的维度是  $\mathbb{R}^{d \times d_k}$ , 投射到三个不同的空间,得到查询向量  $\mathbf{Q}$ , 键向量  $\mathbf{K}$  和值向量  $\mathbf{V}$ , 计算方式如下:

$$\mathbf{Q}, \mathbf{K}, \mathbf{V} = \mathbf{H}\mathbf{W}_q, \mathbf{H}\mathbf{W}_k, \mathbf{H}\mathbf{W}_v. \quad (7)$$

而单头自注意力  $\text{head}$  可以通过  $\mathbf{Q}, \mathbf{K}$  和  $\mathbf{V}$  三个

向量,经由特征缩放和 softmax 函数计算得到,公式如下:

$$\text{head} = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}, \quad (8)$$

其中  $d_k$  是输出的维度.多头注意力 MultiHead 通过多组  $\mathbf{W}_q, \mathbf{W}_k$  和  $\mathbf{W}_v$  来增强自注意力的特征提取能力,计算公式为:

$$\text{MultiHead}(\mathbf{H}) = [\text{head}^{(1)}; \dots; \text{head}^{(n)}]\mathbf{W}_o, \quad (9)$$

其中:  $n$  为注意力头的数量;  $[\text{head}^{(1)}; \dots; \text{head}^{(n)}]$  为不同头的注意力按照最后一个维度相结合;  $\mathbf{W}_o$  是线性映射的可学习的权重矩阵,维度是  $\mathbb{R}^{d \times d}$ .

自适应模块将融合当前层的输出特征  $A^h$  和前一层的特征  $A^l$ , 相较于直接相加的残差方法,更好地结合深层特征与浅层特征,计算方式如下:

$$\mathbf{A} = \tanh(\mathbf{W}_q[\mathbf{A}^h; \mathbf{A}^l] + \mathbf{b}_q), \quad (10)$$

其中  $\mathbf{A}$  为融合后的输出,之后将进行归一化处理.接着经过前馈神经网络和相同的自适应模块操作,得到特征输出  $h_s$ .

### 1.3 IDCNN 网络

IDCNN 网络<sup>[14]</sup> 是由膨胀卷积神经网络迭代而成.而膨胀卷积神经网络是在 CNN 网络里面加入膨胀宽度,使得在卷积核大小一致的前提下能够有更大的感受范围,获得更多的上下文信息.

IDCNN 模块第一层的膨胀系数为 1, 嵌入层的输出  $E$  经过卷积操作后,得到输出序列  $c_t^1$ , 计算方法如下:

$$c_t^1 = D_1^{(0)}\mathbf{E}, \quad (11)$$

其中:  $D_k^{(j)}$  为第  $j$  层, 膨胀系数为  $k$  的卷积操作;  $t$  为不同的输入文本序列.随后可以得到每一层的膨胀卷

积的输出  $c_i^{(j)}$ , 计算方法如下:

$$c_i^{(j)} = \text{ReLU}(D_{k_j}^{(j-1)} c_i^{(j-1)}), \quad (12)$$

其中:  $k_j$  为第  $j$  层的膨胀系数, 将多个卷积层叠加, 得到一个膨胀卷积块  $B$ . 为避免深层网络导致的过拟合, 同时保证更大的感受野, 卷积块  $B$  会迭代使用, 第  $l$  次迭代后的输出  $b_i^{(l)}$  为:

$$b_i^{(l)} = B(b_i^{(l-1)}). \quad (13)$$

通过  $W_b$  矩阵进行映射变换, 可以得到输入序列的标签得分  $h_i$ . 最后, 将自适应多头自注意力机制和 IDCNN 网络得到的输出进行连接操作得到的融合输出  $h$ , 计算方式如下:

$$h_i = W_b b_i^l, \quad (14)$$

$$h = (h_s \oplus h_i). \quad (15)$$

#### 1.4 条件随机场(CRF)

解码层采用条件随机场(CRF)<sup>[15]</sup>. 对于给定的输入序列  $x = \{x_1, x_2, \dots, x_m\}$ , 预测标注序列  $y = \{y_1, y_2, \dots, y_m\}$ , 通过编码层输出的状态矩阵  $T$  以及转移矩阵  $M$ , 能够得到输入序列对应标签的分数矩阵  $Score$ :

$$Score(x, y) = \sum_{i=1}^n M_{y, y_{i+1}} + \sum_i T_{i, y_i}, \quad (16)$$

其中:  $n$  为标签的种类数量,  $T_{i, y_i}$  为  $x_i$  标记为  $y_i$  的概率,  $M_{y, y_{i+1}}$  为标签  $y$  后续标签为  $y_{i+1}$  的概率即标签间的转移概率. 对于输入序列  $x = \{x_1, x_2, \dots, x_m\}$ , 输出标签序列  $y = \{y_1, y_2, \dots, y_m\}$  概率  $P(y|x, \theta)$  为:

$$P(y|x, \theta) = \frac{e^{Score(x, y)}}{\sum_{y' \in Y} e^{Score(x, y')}}, \quad (17)$$

其中:  $y'$  表示真实的标签序列输出,  $Y$  为所有可能的输出序列. CRF 模型使用极大似然法, 最大化正确标签的概率, 模型参数  $\theta$  计算方式如下:

$$\theta^* = \underset{\theta}{\operatorname{argmax}} \sum_i \log [P(y^i|x^i, \theta)]. \quad (18)$$

最后, 利用维特比算法进行解码, 得到整体输出概率最高的标签序列, 计算如下:

$$y^* = \underset{y \in Y(x)}{\operatorname{argmax}} P(y|x, \theta). \quad (19)$$

## 2 实验结果

### 2.1 实验环境及主要参数设置

本文实验环境配置如表1所示.

本文采用BERT基准模型作为嵌入层, 模型共12层, 隐藏层大小为768维, 最大序列长度128, 批处理大小为8, 学习率  $5e-5$ , dropout 设置0.5防止过

表1 实验环境

Tab. 1 Experimental environment	
环境	配置
操作系统	ubuntu
CPU	Intel® Xeon(R) Gold 5218R
GPU	RTX3090(24G×2)
Python	3.7
Pytorch	1.8.1

拟合. IDCNN 网络的卷积核大小为  $3 \times 3$ , 卷积核个数为100, 膨胀系数为1、1、2, 共4个大小相同的膨胀卷积块叠加. 本文还对多头自注意力中头的数量进行了讨论, 在 Weibo 数据集上的实验结果如图2所示.

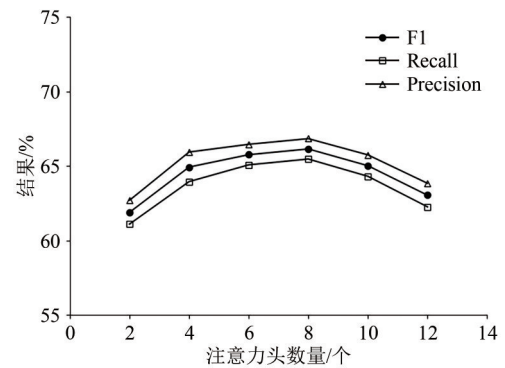


图2 不同自注意力头实验结果

Fig. 2 Experimental results of different self-attention heads

由图2可见: 在自注意力头数量较少的时候, 识别效果随着自注意力头的递增而增加; 在数量为8的时候达到最高; 继续增加头的数量, 结果反而开始下降. 其原因可能是头的数量过多将会感知到更多无用信息, 使得累计误差扩大, 影响识别效果. 在后续实验中, 统一将自注意力头的数量设置为8.

### 2.2 实验数据及评价指标

本文采用3个公共中文命名实体识别数据集对模型进行评估. 其中, MSRA是微软公开的新闻领域数据集, 包含3种实体类型. Resume是中文简历领域数据集, 包含8种实体类型. Weibo是中文社交媒体领域的数据集, 包含4种实体. 数据集全部采用BMES方法标注, 其中B、M、E、S分别表示实体的开始字符、实体内部字符、实体的结束字符和单字符实体, 非实体字符均用O表示. 其与BIESO方法区别为实体内部字符表示符号不一样, 其余均相同. 训练集、测试集和验证集划分, 及其中语句数量详细信息如表2所示.

本文采用精确率、召回率和F1值三种指标对模型进行评价. 精确率  $P$  指识别出来的所有实体中, 正

表2 数据集信息

Tab. 2 Information of the datasets

划分方式	不同数据下的句子数量/k		
	MSRA	Resume	Weibo
训练集	46.40	3.80	1.40
测试集	4.40	0.48	0.27
验证集	—	0.46	0.27

确实体数量的占比;召回率  $R$  指正确识别出来的实体占总的实体数量的比重;F1 综合了精确率和召回率,可以全面评价模型的性能,具体定义为:

$$P = \frac{TP}{TP + FP} \times 100\%, \quad (20)$$

$$R = \frac{TP}{TP + FN} \times 100\%, \quad (21)$$

$$F1 = \frac{2 \times P \times R}{P + R} \times 100\%, \quad (22)$$

其中:TP 为准确识别出来的实体总数,FP 为将非实体识别为实体的总数,FN 为未识别到的实体总数.

### 2.3 实验结果

本文将 BERT-BiLSTM-CRF 作为基线模型进行对比,在 MSRA 公共数据集上进行了实验,实验对比结果如表 3 所示.

由表 3 可见:本文提出的 BERT-MHSA-IDCNN-CRF 模型在精确率、召回率和 F1 值三个指标上取得了最优,其中 F1 值提高了 0.47%.这是由于多头自注意力本身可以无偏地学习到远距离依赖,在输入的字符嵌入中增加了相对位置嵌入后,使得 Self-Attention 也能像 BiLSTM 等 RNN 结构的网络一样具

表3 与基线模型对比结果

Tab. 3 Comparison results with the baseline model

模型	不同指标下结果			Time (epoch) /s
	$P$	$R$	F1	
BERT-BiLSTM-CRF	94.38	94.92	94.65	584
BERT-MHSA-IDCNN-CRF	94.86	95.38	95.12	479

有很强的时序特征提取能力,感知字符间的距离和方向等信息.自适应模块可以增强浅层特征的利用,有选择地进行特征融合.此外,模型通过 IDCNN 可以提取输入序列的全局特征.这些由不同网络得到的多角度特征经过聚合后,进一步提高了模型的性能.

从每轮训练时间上来看,相较于 BERT-BiLSTM-CRF,本文模型训练时间也有所减少.因为 BiLSTM 网络在计算单元特征输出时,依赖前一个单元的输出,只能顺序处理文本,无法并行计算,导致训练时间过长.而 IDCNN 模块和 MHSA 模块都可以并行计算,充分利用 GPU 资源进行,有效地加速了模型训练.本文虽采用了两个模块,但整体训练时间仍小于 BERT-BiLSTM-CRF 模型.

此外,将本文与最近几年提出的模型在 Weibo、Resume 和 MSRA 三个公共数据集上进行了对比.Weibo 数据集的实验分为三种,分别是命名实体 (Named Entity, NE)、标称实体 NM (Nominal Entity, NM) 和二者总体 (Overall) 的实验,结果如表 4 所示.

表4 与其他先进模型对比结果

Tab. 4 Comparison results with other advanced models

年份	模型	3个数据集下的实验结果								
		Weibo			Resume			MSRA		
		NE	NM	Overall	$P$	$R$	F1	$P$	$R$	F1
2020	CAN-NER	55.28	62.98	59.31	95.05	94.82	94.94	93.53	92.42	92.97
2020	IDC-HSAN	55.40	63.13	59.35	95.12	94.87	94.97	93.60	92.65	93.08
2021	ACNN	55.98	64.09	59.79	94.69	95.21	94.95	93.51	92.51	93.01
2022	BERT-BGC-NER	60.32	66.87	63.60	—	—	—	95.00	94.96	94.97
本文	BERT-MHSA-IDCNN-CRF	63.46	69.74	66.14	95.96	95.43	95.67	94.86	95.38	95.12

其中,ZHU 等<sup>[4]</sup>提出一种融合注意力机制的卷积神经网络的 CAN-NER 模型.利用 CNN 捕获窗口范围内相邻字符的依赖关系,分别采用 Bi-GRU 和自注意力来获取句子级别的全局信息以及来自相邻字符和句子上下文的信息,在不使用外部词汇资源的情况下得到了更好的效果.陈茹等<sup>[14]</sup>提出了 IDC-HSAN 模型,通过 CNN 提取局部特征,采用多头自注意力机制提取上下文信息,该模型同样融入了

IDCNN 网络,取得了不错的效果.KONG 等<sup>[16]</sup>提出了一种完全基于卷积神经网络的高效模型 ACNN,通过构建多层次的 CNN 网络,有效地提取句子中短期和长期的上下文信息,并设计了一种简单的注意力机制获取全局的上下文信息,提高了模型的性能.LI 等<sup>[17]</sup>提出了 BERT-BGC-NER 模型,把依赖句法分析树的自注意力方法嵌入到 BERT 模型中,同时利用 BERT 将训练好的词向量嵌入全局注意力机制

中,得到句子中最重要的语义信息,有效地提高了模型的实体识别能力。

基于BERT强大的语义信息表达能力,本文模型可以得到较好的嵌入表示,在不添加外部词汇、句法等先验知识的情况下,取得了较好的效果。其中,在Weibo数据集下,NE、NM和总体情况下的F1值均为最高,总体的F1值达到了66.14%,提高了2.54%。Resume数据集下,模型的精确率、召回率、F1值同样取得了最优,F1值为95.65%,提高了0.21%。在MSRA数据集下,虽然精确率低于BERT-BGC-NER模型,但召回率和整体的F1值仍为最高,F1值为92.12%,提高了0.15%。与BERT-BGC-NER模型相比,虽然该模型同样使用了BERT作为嵌入层,但本文中自适应多头自注意力和IDCNN网络可以从多角度对输入序列特征进行建模,提高了模型的性能。而对同样使用了IDCNN和多头自注意力的IDC-HSAN模型来说,虽然该模型引入了多粒度信息,增强了输入,但其语义表示能力弱于BERT。此外,本文对多头自注意力模块进行了改进,输入中增加了相对位置特征,让模型学习到字符间的距离和方向信息,自适应模块也能有效的利用每层提取的特征。因此,本文模型能获得更好的识别效果。

#### 2.4 消融实验

本文将通过实验验证MHSA和IDCNN融合的有效性。实验分为只使用MHSA的BERT-MHSA-CRF模型和只用IDCNN的BERT-IDCNN-CRF模型,以及同时去掉两个模块的BERT-CRF模型,数据集为MSRA。其中,单独使用BERT模型时,内部参数都会参与训练,学习率为 $3e-5$ ,batch\_size设置为16,采用Adam优化策略,结果如表5所示。

表5 MSRA数据集上的消融分析  
Tab. 5 Ablation analysis on MSRA

模型	不同指标下结果		
	P	R	F1
BERT-MHSA-IDCNN-CRF	94.86	95.38	95.12
BERT-MHSA-CRF	94.65	94.98	94.81
BERT-IDCNN-CRF	94.19	95.06	94.62
BERT-CRF	93.87	94.96	94.41

由表5可见:经过训练的BERT-CRF模型识别效果可以达到94.41%。在此基础上,固定训练好的BERT模型参数,结合MHSA模块和IDCNN模块,在参数设置与前文保持一致地情况下,训练后F1值分别为94.81%和94.62%。可见两个模块都能带来性能的改进,验证了各自的有效性。而BERT-MHSA-

IDCNN-CRF模型同时利用MHSA和IDCNN,融合两个模块的特征输出,相较于只用其中一个模块取得更高的实体识别准确度,能够给模型带来进一步的提升,证明了融合特征输出的可行性。

### 3 结语

本文提出了一种基于BERT的多特征融合中文命名实体识别模型,该模型利用BERT获得深层语义表示能力,在多头自注意力机制中加入相对位置编码和自适应模块,并融合IDCNN网络建模输出,增强了对输入序列的特征提取能力。在3个公共数据集上的结果表明:本文方法优于其他相关模型,提升了实体识别效果。下一步将尝试在细粒度中文命名实体识别方向继续开展研究,进一步优化模型的计算开销,为在现实复杂的语言环境中提供更大的帮助。

#### 参 考 文 献

- [1] 毛养勤,覃俊,尹帆,等.基于新型关系注意力机制的实体关系抽取[J].中南民族大学学报(自然科学版),2021,40(3):286-291.
- [2] MA X, HOVY E. End-to-end sequence labeling via bidirectional LSTM-CNNs-CRF [C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin: ACL, 2016: 1064-1074.
- [3] QIN Y, SHEN G W, ZHAO W B, et al. A network security entity recognition method based on feature template and CNN-BiLSTM-CRF [J]. Frontiers of Information Technology & Electronic Engineering, 2019, 20(6): 872-884.
- [4] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: ACL, 2019: 4171-4186.
- [5] ZHU Y, WANG G. 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: ACL, 2019: 3384-3393.
- [6] 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. Macao: IJCAI, 2020: 4982-4988.
- [7] 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.
- [8] 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.
- [9] 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: Human Language Technologies. Minneapolis: ACL, 2019: 2379-2389.
- [10] WU S, SONG X, FENG Z, et al. 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. Online: ACL, 2021: 1529-1539.
- [11] 王蕾, 谢云, 周俊生, 等. 基于神经网络的片段级中文命名实体识别[J]. 中文信息学报, 2018, 32(3): 84-90, 100.
- [12] 张召武, 徐彬, 高克宁, 等. 面向教育领域的基于SVR-BiGRU-CRF中文命名实体识别方法[J]. 中文信息学报, 2022, 36(7): 114-122.
- [13] DAI Z, YANG Z, YANG Y, et al. Transformer-XL: Attentive language models beyond a fixed-length context [J]. arXiv Preprint arXiv:1901.02860, 2019.
- [14] 陈茹, 卢先领. 融合空洞卷积神经网络与层次注意力机制的中文命名实体识别[J]. 中文信息学报, 2020, 34(8): 70-77.
- [15] 梁文桐, 朱艳辉, 詹飞, 等. 基于BERT的医疗电子病历命名实体识别[J]. 湖南工业大学学报, 2020, 34(4): 54-62.
- [16] KONG J, ZHANG L, JIANG M, et al. Incorporating multi-level CNN and attention mechanism for Chinese clinical named entity recognition[J]. Journal of Biomedical Informatics, 2021, 116: 103737.
- [17] LI D, YAN L, YANG J, et al. Dependency syntax guided BERT-BiLSTM-GAM-CRF for Chinese NER [J]. Expert Systems with Applications, 2022, 196: 116682.

(责编&校对 刘钊)