

基于多重卷积神经网络模型的命名实体识别

赵建强¹, 朱万彤², 陈 诚¹

(1. 厦门市美亚柏科信息股份有限公司, 福建 厦门 361008;

2. 西安电子科技大学 计算机科学与技术学院, 陕西 西安 710071)

摘 要:针对命名实体识别任务, 该文以可通过参数调节感受野范围的 IDCNN 为基础, 提出了 BERT-Duplex CNN-Self Attention-CRF 模型, 在不引入其他辅助特征的前提下, 采用 BERT 预训练模型来提供具有丰富语义信息的字嵌入, 将字嵌入分别输入 IDCNN 和 CNN 中进行特征提取。IDCNN 在提取长距离语义信息的前提下可以充分利用 GPU 的并行能力, CNN 在不损失并行能力的前提下可以弥补 IDCNN 对于局部上下文信息的缺失, 将提取到的特征融合, 通过引入自注意力机制在众多特征中选取对命名实体识别更有效的特征, 最后通过 CRF 层提升实体标签预测的准确性。为验证模型的有效性, 该文在常用的 MSRA 数据集以及 Resume 数据集上进行实验, 实验结果表明: 该模型在 MSRA、Resume 数据集上, 结果超越了 Lattice LSTM、BERT-Tagger、LR-CNN、PLET 等几个目前较优模型, 相对于结果最好的 PLET 模型, 该模型的 F1 值分别提高了 1.28 百分点、0.15 百分点。

关键词:命名实体识别; BERT; 卷积神经网络; 膨胀卷积; 自注意力机制

中图分类号: TP391

文献标识码: A

文章编号: 1673-629X(2023)01-0187-06

doi:10.3969/j.issn.1673-629X.2023.01.028

Named Entity Recognition Based on Duplex Convolution Neural Network Model

ZHAO Jian-qiang¹, ZHU Wan-tong², CHEN Cheng¹

(1. Xiamen Meiya Pico Information Co., Ltd., Xiamen 361008, China;

2. School of Computer Science and Technology, Xidian University, Xi'an 710071, China)

Abstract: For the task of named entity recognition, we present BERT-Duplex CNN-Self Attention-CRF model based on IDCNN, which can adjust the receptive field range through parameters. Without introducing other auxiliary features, the BERT pre-training model is used to provide word embedding with rich semantic information, and the word embedding is input into IDCNN and CNN respectively for feature extraction. IDCNN can make full use of the parallel ability of GPU on the premise of extracting long-distance semantic information. Meanwhile, CNN can make up for IDCNN's lack of local context information without losing parallel ability. The self-attention mechanism can be used to select the feature which is more effective for named entity recognition task, next. Finally, the CRF layer is used to improve the accuracy of entity label prediction. Experiments are carried out on MSRA dataset and Resume dataset. The F1-score of the model on MSRA and Resume datasets are improved respectively than several current better models such as Lattice LSTM, BERT Tagger, LR-CNN and PLET, with 1.28%, 0.15% increased than the best baseline.

Key words: named entity recognition; BERT; convolutional neural network; dilated convolution; self-attention mechanism

0 引 言

命名实体识别 (Named Entity Recognition, NER) 是自然语言处理 (Natural Language Processing, NLP) 中一项非常基础的任务, 其目标是识别文本中具有特定意义实体并对其进行分类, 主要包括人名、地名、组织机构名、专有名词等。NER 任务是信息抽取、问答

系统、句法分析、机器翻译、知识图谱构建、信息检索等众多 NLP 任务的重要基础。

早期对命名实体识别的研究多采用基于规则和字典的方法以及基于统计机器学习的方法, 近年来, 对命名实体识别的研究主要集中于深度学习方法。命名实体识别任务首先在英文数据集上展开研究, 相比于英

收稿日期: 2021-12-29

修回日期: 2022-04-29

基金项目: 广州科技攻关重大专项 (201903007)

作者简介: 赵建强 (1979-), 男, 高级工程师, CCF 会员 (12330M), 研究方向为人工智能、自然语言处理、电子数据取证; 通讯作者: 朱万彤 (1996-), 女, 硕士, 研究方向为自然语言处理。

文数据集单词之间通过空格分开便于研究,中文数据集中汉字字符排列紧密且一词多义,增加了命名实体识别研究的难度^[1]。英文 NER 针对输入的文本大多采用单词级别的分布式表示,这是由于英文单词可以看作是英文中的最小语义单位,而对于中文文本,一个字往往就有其本身的语义,由字组成的词语更是存在一词多义的现象,因此,选取何种方式进行输入文本的分布式表示也会对中文 NER 的效果产生很大影响。目前对于中文 NER 的特征提取大都基于 BiLSTM 模型^[2],BiLSTM 可以整合序列前后项信息,获取更大范围的上下文依赖关系,但由于其采用循环处理序列的方式,不能充分利用 GPU 的并行计算能力。相比 BiLSTM 处理输入序列的方式,卷积神经网络(CNN)可以一次性并行处理整个输入序列,但 CNN 在进行卷积操作时,不能有效地提取到输入序列的全局特征,Strubell 等人^[3]提出的 IDCNN 通过膨胀卷积获得范围更大的感受野,并且相比 BiLSTM 具有更好的并行能力。近年来,随着 BERT^[4]等预训练模型的陆续提出,其对多义词进行建模、整合句法和语义信息的能力,为多项自然语言处理任务带来了优异的性能提升。

针对以上情况,在不引入外部特征的前提下,以 IDCNN 模型为基础,该文提出了 BERT-DuCNNS-CRF 模型,通过多特征提取器获取输入文本的局部和全局特征,充分利用 GPU 的并行性,并在 MSRA 和 Resume 数据集上进行实验。

1 相关工作

近年来,随着硬件计算能力的发展及词的分布式表示的提出,神经网络可以有效处理许多 NLP 任务。与基于特征的机器学习方法相比,深度学习方法有助于自动发现隐藏的特征。在深度学习方法中,NER 任务的一般流程主要包括:输入的分布式表示、上下文编码器、标签解码器^[5]。

输入的分布式表示通常包含字符及单词级别的嵌入,以及在基于特征的方法中附加其他有效特征(如词典、词汇相似度等^[6-7])。中文的词内是包含大量信息的,直接用基于词的方法会受到分词效果的影响,因此目前很多模型都是基于字嵌入的,词级别的信息则依赖于通过好的向量嵌入和模型的上下文依赖来补充。Ma 等人^[8]首先使用卷积神经网络将单词的字符级信息编码成其字符级表示形式,之后将字符级和单词级表示形式进行组合,并将它们输入到双向 LSTM 中,以对每个单词的上下文信息进行建模,在解码层使用顺序 CRF 联合解码整个句子的标签。Y. Zhang 和 J. Yang^[9]提出通过 Lattice LSTM 方法将词信息加入到基于字的 LSTM+CRF 模型中,减轻模型受到分词错

误的影响。除了单词级和字符级的表示法外,一些研究还将其他信息纳入最终的表示中,J. Devlin 等人^[4]提出了来自 Transformer 的双向编码器表示模型—BERT,使用屏蔽语言模型来启用预训练的深度双向表示。对于给定的令牌,其输入表示是通过对相应的位置、片段和令牌嵌入求和而构成的。经过预训练的语言模型嵌入通常需要大规模的语料库进行训练,并且本质上并入了辅助信息的嵌入(如位置和片段嵌入)。

上下文编码器使用 CNN、RNN 或其他网络捕获上下文依赖关系。在使用 CNN 网络结构作为上下文编码器的方法中,Collobert 等人^[10]提出了一种基于 CNN 的句子逼近网络。在输入表示阶段之后,将输入序列中的每个单词嵌入到 N 维向量中,使用卷积层在每个单词周围产生局部特征,卷积层的输出大小取决于句子中单词的数量;利用卷积层提取的局部特征向量来构造全局特征向量。通过对句子中的位置进行最大值或平均运算来提取全局特征。最后,将这些固定大小的全局特征反馈到标签解码器,计算网络输入中单词的所有可能标签。继 Collobert 的工作之后,Zhu Y 和 Wang G^[11]提出一种卷积注意网络 CAN,它由具有局部注意力层的基于字符的卷积神经网络(CNN)和具有全局自注意力的门控递归单元组成,用于从局部的相邻字符和全局的句子上下文中获取信息。对于普通的 CNN,在卷积之后,末层神经元可能只是得到了原始输入数据中一小块的信息。而对于命名实体识别,整个句子中的每个字都可能对当前需要标注的字产生影响。为了覆盖更多的信息,加入更多的卷积层,从而导致层数更深、参数更多,为防止过拟合要加入更多的 Dropout 之类的正则化,带来更多的超参数,整个模型将会变得庞大和难以训练。E. Strubell 等人^[3]提出迭代膨胀卷积神经网络(IDCNN),由于能够处理更大的上下文和结构化预测,因此计算效率更高。在以 RNN 网络结构为基础编码上下文的方法中,Huang 等人^[2]率先提出 BiLSTM-CRF 架构应用于序列标记任务(POS、NER)。一方面,BiLSTM 使得模型可以获得前后文的特征信息,另一方面,CRF 使得模型能够获取句子级别的标注信息。由于 CRF 层的引入可以有效解决预测标签之间的强语法依赖的问题,因此有效避免了预测标签冲突的情况,尤其是对于 NER 这种标签带有强约束的任务来说。该模型在各项 NLP 序列标注任务中取得了当时的 SOTA 表现,之后,大量的研究都应用了 BiLSTM 作为编码序列上下文信息的基本架构^[12-13]。基于神经网络的序列标注模型通常采用复杂的卷积或循环网络,由 Vaswani 等人^[14]提出的 Transformer,利用堆叠的自注意力和点乘注意力,使用全连接层来建立编码器和解码器的基本块,完全避免

了循环和卷积,提升了训练的效率。Transformer 的优势是并行性和对长程上下文语义建模,但它在 NER 中的表现却不尽如人意,Yan 等人^[15]提出一种采用自适应 Transformer 的 NER 架构—TENER,对字符级特征和词级特征进行建模,通过结合方向感知、距离感知和 Unscaled 的 Attention,使得改进的 Transformer 在 NER 中获得了良好的表现。BERT 是 Transformer 的双向编码表示,通过在所有层中对左右文本进行联合条件化来预训练深度双向 Transformer。2019 年,Guan Guoliang^[16]基于 BERT 模型,在后面加上了 BiLSTM+CRF 层,进行命名实体识别。在人民日报等语料库上进行训练和测试,实验表明,该模型在性能上超过了以往的模型。

标签解码器是 NER 模型的最后步骤,它以上下文相关的表示形式作为输入,并生成与输入序列相对应的标签序列。条件随机场(CRF)是一个以观察序列为整体条件的随机场。CRF 已广泛用于基于特征的监督学习方法中。许多基于深度学习的 NER 模型使用 CRF 层作为标签解码器,如在双向 LSTM 层顶部^[2,17]和 CNN 层顶部^[10]。

2 模型

该文提出了 BERT-DuCNNS-CRF 模型,结构如图 1 所示。分为三个模块:(1)输入层:BERT 对文本的原始输入进行字嵌入;(2)特征提取层:分别使用 IDCNN 和 CNN 模型对字向量进行特征提取,而后进行特征融合,随后进行 Attention 操作;(3)输出层:计算从特征提取层输入序列的预测标签。

2.1 输入层

BERT 的网络架构如图 2 所示,BERT 使用 Transformer 中的 Encoder 部分,每个 Encoder 由 6 个相

同的层组成,每层又由 2 个子层组成,分别是多头自注意力机制和全连接前馈神经网络。多头注意力机制让模型去关注不同方面的信息,有助于网络捕捉到更丰富的特征信息。全连接前馈神经网络提供非线性变换。图中 Trm 代表一个 Encoder。

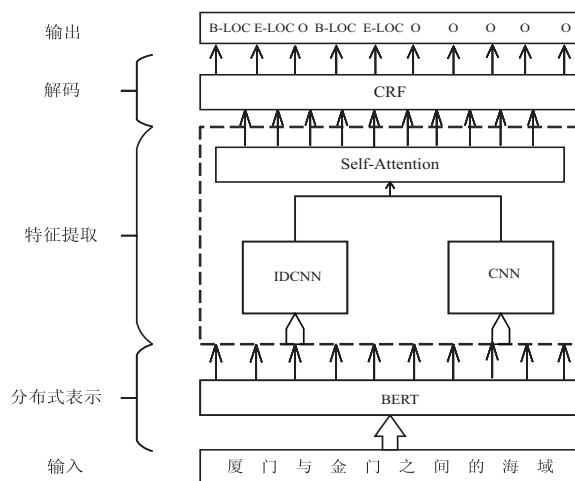


图1 模型结构

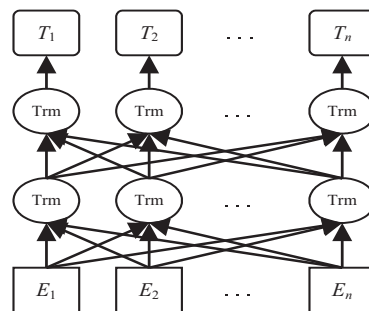


图2 BERT 网络架构

BERT 将输入的句子转换为 Embedding,由三个部分相加得到,如图 3:Token Embedding (单词嵌入/字符嵌入)、Segment Embedding (分隔嵌入) 和 Position Embedding (位置嵌入)。

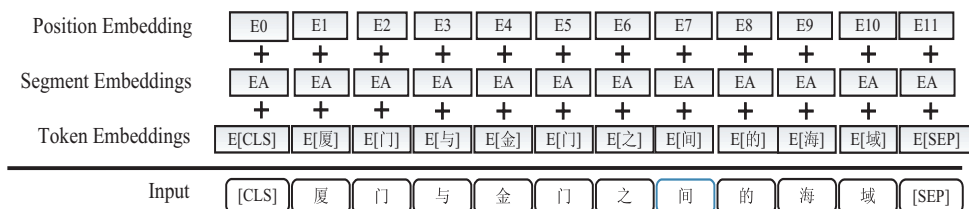


图3 BERT 输入表示

BERT 是多任务模型,通过预训练的方式训练模型,预训练任务包括 Masked Language Model (MLM) 和 Next Sentence Prediction (NSP)。MLM 随机屏蔽部分输入单词,然后训练模型来预测被屏蔽的单词,模型为了预测正确,就需要保持每个输入的分布式上下文表示,从而实现上下文进行特征提取的目的。NSP 是判断句子间的上下文关系,并通过标签进行标注,从而提取句子级别的特征。

2.2 特征提取层

普通 CNN 在输入矩阵一片连续的位置上滑动进行卷积操作,通过池化操作整合多尺度的上下文信息,这种方式会损失精度,但是不加池化层会使感受野变小,学习不到全局的特征。而通过加入更多卷积层、去掉池化层、扩大卷积核等操作来获取上下文信息,势必导致参数增多,计算量增大,容易发生过拟合。膨胀卷积在不做池化损失信息的情况下,通过加入卷积核的

间隔加大了感受野,让每个卷积输出都包含较大范围的信息。膨胀卷积在处理需要全局信息或者较长范围的信息依赖的问题中,都能很好地应用。但膨胀卷积在增大感受野的同时,却损失了局部语义信息。基于以上问题,该文使用 IDCNN+CNN 对字嵌入进行联合特征提取,在提取全局信息的同时弥补局部语义信息的缺失,同时使用自注意力机制学习联合特征之间的关系。

2.2.1 IDCNN+CNN

普通 CNN 结构通常包括输入层、卷积层、池化层和全连接层,如图 4 所示。对一个 filter 来说,可以认为滑动窗口从输入矩阵的第一个字开始不断往后移动,对于某个时刻的窗口,通过非线性变换将窗口内的输入值转换为某个特征值,随着窗口的不断向后移动,得到该 filter 下的特征向量。池化层对 filter 提取的特征进行降维操作,该模型中采用最大池化操作。在池化层后接全连接层,得到最终的特征。

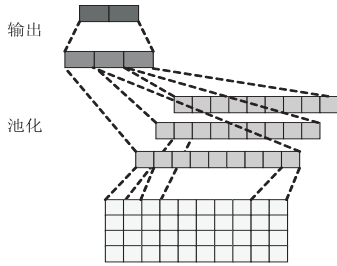


图 4 CNN 结构

IDCNN 将 4 个结构相同的 Dilated CNN 模块拼接在一起,每个模块里是膨胀宽度分别为 1、1、2 的三层 DCNN。IDCNN 对输入句子的每一个字生成一个 logits,然后放入 CRF 层解码出标注结果,每个 Dilated CNN 模块如图 5 所示。IDCNN 与经典的序列标注模型 BiLSTM 都具有获取远距离上下文信息的能力,但是相对于 BiLSTM, IDCNN 可以充分利用 GPU 并行,提升训练效率。

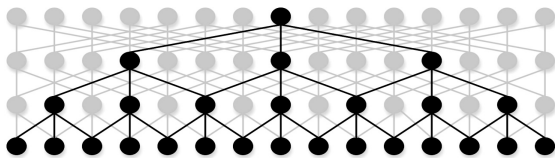


图 5 膨胀卷积块

输入特征提取层的字向量表示为 $X = [x_1, x_2, \dots, x_i]$, $D_\delta^{(j)}$ 表示第 j 层、膨胀度为 δ 的膨胀卷积神经网络,该文设置 $\delta = 1, 1, 2$, $c_i^{(j)}$ 表示通过第 j 层卷积网络得到的特征, IDCNN 过程如下:

$$c_i^{(0)} = D_1^{(0)} X \quad (1)$$

$$c_i^{(j)} = \text{relu}(D_\delta^{(j-1)}(c_i^{(j-1)})) \quad (2)$$

CNN 层的输入也为 $X = [x_1, x_2, \dots, x_i]$, 卷积核为 H_c , 偏置为 b_c , 经过非线性变换得到 $C_c = \{c_1, c_2, \dots,$

$c_i\}$ 特征序列:

$$c_i = \text{relu}(x_i \cdot H_c + b_c) \quad (3)$$

通过 IDCNN 和 CNN 得到的特征序列进行拼接, C_I 、 C_C 分别代表 IDCNN、CNN 提取的特征序列:

$$C = C_I \oplus C_C \quad (4)$$

2.2.2 多头自注意力机制

自注意力机制可以用来学习句子中任意两个字符之间的依赖关系并捕捉句子内部的结构信息,并且具有并行计算的能力。自注意力机制通过 (K=key、Q=query、V=value) 的三元组,提供了一种有效的捕捉全局上下文信息的建模方式,可用公式表示为:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (5)$$

Q 、 K 、 V 三个矩阵都来自同一个输入,文中 Self-Attention 层的输入来源于 IDCNN 与 CNN 输出的拼接 C , 设置 $Q = K = V = C$ 。Self-Attention 使用缩放点积,在使用点积运算进行相似度计算的基础上,缩小了 \sqrt{d} 倍 (d 为输入向量的维度)。

多头注意力是指重复做 h 次 attention 之后进行拼接,即首先对 query、key、value 进行线性变换,然后输入到缩放点积 attention,每一次算作一个头,每次使用不同的线性变换参数 W ,最后,将拼接后的结果做一线性变换作为多头注意力的结果。多头注意力可以从不同的维度和表示子空间里学习到相关的信息,公式表达如下:

$$\text{head} = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (6)$$

$$C' = \text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W \quad (7)$$

2.3 输出层

经过特征提取层可以获得语义特征及当前标签的预测结果,但不考虑标签之间的依赖关系,而 CRF 层可以训练数据中学习到标签之间的约束条件,从而保证预测标签的有效性。对于给定一个句子 $s = [s_1, s_2, \dots, s_T]$, 对应的预测标签序列为 $y = [y_1, y_2, \dots, y_T]$, y 的概率计算公式如下:

$$p_i = WC_i + b \quad (8)$$

$$f(y_i, y_{i+1}) = \sum_{t=0}^n A_{y_i, y_{i+1}} + \sum_{t=1}^n p_i y_i \quad (9)$$

$$P(y | s) = \frac{\sum_{t=1}^T f(y_{t-1}, y_t)}{\sum_y \sum_{t=1}^T e^{f(y_{t-1}, y_t)}} \quad (10)$$

其中, p 表示句中第 t 个字的第 y_t 个标签的得分, A 为状态转移矩阵, $f(y_{t-1}, y_t)$ 为 y_{t-1} 到 y_t 的转移分数,目标是使 $P(y | s)$ 最大。在解码时,采用 Viterbi 算法来寻找达到最大概率的路径。

3 实验设计

3.1 数据集

该文使用 MSRA^[18]、Resume^[9] 数据集来对模型效果进行评估。表 1 提供了这两个数据集的详细情况,

表 1 数据集

数据集	类型	训练集	验证集	测试集	来源	实体类别
MSRA	句子/条	46.4k	-	4.4k	新闻	PER, LOC, ORG
	实体/个	74.7k	-	6.2k		
Resume	句子/条	3.8k	0.5k	0.5k	简历	NAME, LOC, ORG, CONT, RACE, TITLE, EDU, PRO
	实体/个	13.4k	1.5k	1.6k		

3.2 实验设置

所用数据集均采用 BIOES 标注方法,采用精确率 (Precision, P)、召回率 (Recall, R) 和 F1 值 (F1-score, F1) 进行实验结果评估,见表 2。

表 2 实验设置

参数	值
learning_rate	0.000 1
decay_rate	0.9
dropout	0.1
sequence_length	128
batch_size	128 (MSRA) 32 (Resume)
epoch	100

3.3 实验结果及分析

为验证所提模型的性能,在 MSRA 和 Resume 数据集上,与 Lattice LSTM^[9]、BERT-Tagger^[4]、LR-CNN^[19]、PLET^[20] 等几个模型进行对比。

Lattice LSTM 模型通过门控单元获取词汇,避免了实体分割错误的情况。BERT-Tagger 使用 BERT 最后一层的输出作为字符级的上下文表示来进行序列标注。LRCNN 使用 Rethinking 机制来处理词汇冲突问题,并且加速了模型训练。PLET 以 transformer encoder 为基础,能够并行地批量处理对所有字符和匹配的词典词信息的建模。除此之外,它还添加了位置关系表示;引进了一种多孔机制增强局部性的建模和维持捕捉长期依赖的能力。为保证实验的公平性,以上模型均使用 BERT 预训练词嵌入。

测试结果如表 3 和表 4 所示。

通过表 3 和表 4 可知,所提模型在 MSRA 数据集和 Resume 数据集上的 F1 值分别达到了 95.81%、96.60%,与目前效果较好的 PLET 模型在 MSRA 数据集上相比,F1 值提升了 1.28 百分点,在 Resume 数据集上 F1 值提升 0.15 百分点。实验结果表明,所提模型在不引入其他外部特征的情况下,通过构造融合的特征提取器,分别针对全局语义信息和局部语义信息进行提取,对不同特征进行融合、扩展,有效提升了模型在 NER 任务上的预测能力,证明了该方法的优

越性。MSRA 数据集来源于新闻文本,Resume 数据集来源于简历文本。其中 MSRA 数据集包含实体数目,种类较多,具有较大的识别难度。

表 3 MSRA 数据集测试结果 %

模型	P	R	F1
Lattice-LSTM[BERT] (2018)	93.99	92.86	93.42
BERT-Tagger[BERT] (2019)	94.43	93.86	94.14
LR-CNN[BERT] (2019)	94.68	94.03	94.35
PLET[BERT] (2020)	94.91	94.15	94.53
BERT-DuCNNS-CRF	96.08	95.53	95.81

表 4 Resume 数据集测试结果 %

模型	P	R	F1
Lattice-LSTM[BERT] (2018)	95.79	95.03	95.41
BERT-Tagger[BERT] (2019)	96.12	95.45	95.78
LR-CNN[BERT] (2019)	95.68	96.44	96.06
PLET[BERT] (2020)	96.16	96.75	96.45
BERT-DuCNNS-CRF	96.24	96.96	96.60

3.4 消融实验

为验证所提模型的有效性,在 MASR 和 Resume 数据集上针对各模块进行了消融实验,结果如表 5、表 6 所示。其中 Baseline 设置为 IDCNN+CRF 模型。

表 5 MSRA 数据集消融实验 %

模型	P	R	F1
Baseline	89.39	84.64	86.95
BERT+Baseline	95.67	95.38	95.52
BERT+Baseline+CNN	95.69	95.72	95.71
BERT-DuCNNS-CRF	96.08	95.53	95.81

由表 5 和表 6 可以看出,与 Baseline 相比,引入 BERT 预训练模型在两个实验数据集上的准确率和召回率均有大幅提升,证明了通过 BERT 获取的字向量相比于传统的词向量表示的有效性。通过引入 CNN 来弥补 IDCNN 对于局部语义信息的缺失,在实体种类不同、数目不同的两个数据集上均验证了有效性。Self-attention 的引入进一步提升了模型对于字符之间语义信息的捕捉,这一点也可在实验中得到证明。此结果表明,对模型的改进可以进一步提升命名实体识别的效果。

表6 Resume数据集消融实验 %

模型	P	R	F1
Baseline	91.82	93.90	92.65
BERT+Baseline	95.73	96.03	95.88
BERT+Baseline+CNN	95.58	96.53	96.05
BERT-DuCNNS-CRF	96.24	96.96	96.60

4 结束语

针对目前常用NER模型中,BiLSTM-CRF模型并行性不高,IDCNN-CRF模型损失局部上下文信息等问题,提出了BERT-DuCNNS-CRF模型。模型使用BERT预训练字向量,通过融合CNN模型弥补IDCNN对局部语义信息的缺失,加入自注意力机制学习联合特征之间的关系。实验证明,该模型的性能在以上测评数据集上相比其他模型有一定的优势。

在未来的工作中,将继续研究如何将模型拓展到特定领域的命名实体识别任务,同时保持模型的性能。

参考文献:

- [1] 陈曙东,欧阳小叶.命名实体识别技术综述[J].无线电通信技术,2020,46(3):251-260.
- [2] HUANG Z,XU W,YU K. Bidirectional LSTM-CRF models for sequence tagging[J]. arXiv: 1508.01991,2015.
- [3] 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,2017:2670-2680.
- [4] DEVLIN J,CHANG M W,LEE K,et al. BERT:pre-training of deep bidirectional transformers for language understanding [C]//Proceedings of the conference of the North American chapter of the association for computational linguistics: human language technologies. Minneapolis: Association for Computational Linguistics,2019:4171-4186.
- [5] LI J,SUN A,HAN J,et al. A survey on deep learning for named entity recognition[J]. IEEE Transactions on Knowledge and Data Engineering,2020,PP(99):1-1.
- [6] LIU Tianyu,YAO Jing, LIN Chin-Yew. Towards improving neural named entity recognition with gazetteers [C]// Proceedings of the 57th annual meeting of the association for computational linguistics. Florence: Association for Computational Linguistics,2019:5301-5307.
- [7] GHADDAR A, LANGLAIS P. Robust lexical features for improved neural network named-entity recognition [C]// Proceedings of the 27th international conference on computational linguistics. Santa Fe: Association for Computational Linguistics,2018:1896-1907.
- [8] MA X,HOOVY E. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF [C]//Proceedings of the 54th annual meeting of the association for computational linguistics. Berlin: Association for Computational Linguistics,2016: 1064-1074.
- [9] ZHANG Y, YANG J. 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.
- [10] COLLOBERT R, WESTON J, BOTTOU L, et al. Natural language processing (almost) from scratch [J]. Journal of Machine Learning Research,2011,12(1):2493-2537.
- [11] ZHU Y,WANG G,KARLSSON B F. CAN-NER:convolutional attention network for Chinese named entity recognition [C]//Proceedings of the conference of the North American Chapter of the association for computational linguistics: human language technologies. Minneapolis: Association for Computational Linguistics,2019:3384-3393.
- [12] JIE Z, LU W. Dependency-guided LSTM-CRF for named entity recognition [C]//Proceedings of conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing. Hong Kong, China: Association for Computational Linguistics,2019:3862-3872.
- [13] LUO L,YANG Z,YANG P,et al. An attention-based BiLSTM-CRF approach to document-level chemical named entity recognition [J]. Bioinformatics, 2018, 34 (8) : 1381 - 1388.
- [14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. arXiv:1706.03762,2017.
- [15] YAN H,DENG B, LI X, et al. TENER: adapting transformer encoder for named entity recognition [J]. arXiv: 1911.04474,2019.
- [16] GUAN G,ZHU M. New research on transfer learning model of named entity recognition [J]. Journal of Physics Conference Series,2019,1267:012017.
- [17] 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: Association for Computational Linguistics,2017:1227-1236.
- [18] LEVOW G. The third international Chinese language processing bakeoff: word segmentation and named entity recognition [C]//Proceedings of the fifth SIGHAN workshop on Chinese language processing. Sydney: Association for Computational Linguistics,2006:108-117.
- [19] GUI T,MA R T,ZHANG Q, et al. CNN-based Chinese NER with lexicon rethinking [C]//Proceedings of twenty-eighth international joint conference on artificial intelligence (IJCAI-19). Macao, China: [s. n.], 2019:4982-4988.
- [20] XUE M,YU B,LIU T, et al. Porous lattice-based transformer encoder for Chinese NER [J]. arXiv:1911.02733. 2019.