

融合上下文的残差门卷积实体抽取

苏丰龙[†] 孙承哲 景宁

国防科技大学电子科学学院, 长沙 410073; [†] E-mail: xueshu2021@qq.com

摘要 基于传统卷积框架的实体抽取方法, 由于受到卷积感受野大小的控制, 当前词与上下文的关联程度有限, 对实体词在整个句子中的语义欠考虑, 识别效果不佳。针对这一问题, 提出一种基于残差门卷积的实体识别方法, 利用膨胀卷积和带残差的门控线性单元, 从多个时序维度同步考虑词间的语义关联, 借助门控单元调整流向下一层神经元的消息量, 缓解跨层传播的梯度消失问题, 同时结合注意力机制捕捉词间的相关语义。在公开命名实体识别数据集和专业领域数据集上运行结果表明, 与传统的实体抽取框架相比, 基于残差门卷积命名实体算法的速度和精度都有较强的竞争优势, 体现出算法的优越性和强鲁棒性。

关键词 实体抽取; 残差门卷积; 梯度消失; 注意力机制

A Context-Fusion Method for Entity Extraction Based on Residual Gated Convolution Neural Network

SU Fenglong[†], SUN Chengzhe, JING Ning

School of Electronic Science, National University of Defense Technology, Changsha 410073; [†] E-mail: xueshu2021@qq.com

Abstract Due to the convolutional receptive field size, the current word has a limited relevance to the context. It brings about a problem, that is, the semantics of the entity words in the whole sentence is under-considered. The Residual Gated Convolution Neural Network (RGCNN) uses dilated convolution and residual gated linear units to simultaneously consider the associations between words from different dimensions, which adjusts the amount of information flowing to the next layer of neurons. And then by this way the vanishing gradient can be alleviated in cross-layer propagation. At the same time, RGCNN combines the attention mechanism to calculate the semantics between words in the last layer. The results on datasets show that RGCNN has a competitive advantage in speed and accuracy, which reflects the superiority and robustness of the algorithm.

Key words entity extraction; residual gated convolution; vanishing gradient; attention mechanism

实体抽取, 也称为命名实体识别(named entity recognition, NER), 作为自然语言处理领域的基础工作, 主要是通过序列标注方式联合识别实体边界和确定实体类型(如人物、地理位置和组织机构等)。近年来, 在知识图谱构建、信息检索、机器翻译、自动问答以及舆情监测等任务中都有广泛的应用。

传统的方法主要通过手工来设计特征, 耗时费力, 代价巨大, 并且高度依赖专家知识, 领域迁移较为困难, 所以基于神经网络的实体抽取方法成为

主流。通过连续实值向量表示和非线性处理的语义合成, 形成端到端的输入和输出, 能有效地避免手工设计特征。主要的研究方向包括输入的分布式表示、上下文编码器、标签解码以及其他辅助信息添加等4个方面。一般来说, 基于深度学习的NER模型几乎都依赖神经网络提取特征, 然后加上标签解码, 基本框架主要依赖于经典的循环神经网络(Recurrent Neural Network, RNN)序列模型以及局部语义感知的卷积神经网络(Convolutional Neural Network, CNN)模型。然而, 这两类模型有各自的缺陷,

RNN 计算速度较慢,无法并行,CNN 无法处理长期语义依赖,很少有研究者能同时考虑和避免上述问题。

为此,本文提出一种能同时兼顾速度和精度的实体抽取模型,综合考虑文本局部空间语义和长期序列的组合特征,在多个连续词间编码潜在的关联信息,使用注意力机制捕捉序列标注中的重要语块,可有效地解决前人研究的一些不足之处。

1 相关工作

NER 任务最早在第六届语义理解会议中提出,当时仅定义一些通用实体类别。随着应用的发展,NER 任务已深入各种垂直领域,如生物、医疗、金融、军事和地理等。该任务一直作为序列标注问题来解决,求解目标是使得其中的实体边界和类别标签被联合预测。

2015 年,Huang 等^[1]提出使用带有 CRF 层的双向 LSTM 模块,可以有效地使用过去和将来的输入,减少外部资源依赖,在命名实体识别和词性标注任务中都取得较好的效果。随后, Ma 等^[2]提出一个结合双向 LSTM, CNN 和 CRF 的端到端模型,可以同时利用字符和单词的表征信息,在英文词性标注任务和 NER 识别任务中均取得不错的效果。Lattice LSTM 利用词加字的经典模型^[3],加入词和词胞元,将潜在词信息整合到基于字符的 LSTM 中,能更好地提升 NER 效果。针对远程监督会产生嘈杂标签的问题,Shang 等^[4]提出在传统序列标记框架下使用改进的模糊 CRF 层,以便处理具有多个可能标记的标签。在基准数据集上的大量实验表明,仅使用字典,模型即可达到最佳性能。Li 等^[5]参考 Transformer-XL 模型,利用相对位置编码,把词的信息融入 Transformer 中,实验表明新的相对位置编码有利于定位实体边界。Ma 等^[6]提出一种融入词信息的方法,把 BMES 的 4 个特征和字嵌入拼接,考虑了字在词中的作用,实验结果比 FLAT 和 Lattice LSTM 模型都更好,并且计算更简单,可移植性较强。最近, BERT-MRC 模型^[7]将 NER 转化为阅读理解问题,把“文本-实体-实体类型”转换成“上下文-答案-问题”,可以同时实现普通实体抽取和嵌套实体抽取。此外,借助词典和小样本的标注来完成在低资源场景下的 NER 也是近年来的研究热点^[8-12]。

综上所述,NER 研究任务大体上可分为 Encoder 和 Decoder 两部分。前者获取输入句子的向量化表

示,后者得到对每个字符的分类结果。其中,有关实体抽取的上下文编码器研究较多,下面主要介绍基于神经网络架构的实体编码。

1.1 基于循环神经网络的实体识别编码

RNN 在序列标注任务中应用普遍,主要包括 LSTM 和 GRU 及其各种变体等。由于其特殊的循环结构可以很好地拟合句子中的字符序列,同时可以高效地利用序列中过去和未来的输入特征,较为符合语言模型的相关特性,所以这方面的研究较多。

Luo 等^[13]通过在 BI-LSTM+CRF 模型中注入注意力机制,用于文档级化学名词实体抽取,可以在文档中同一语块的多个实例之间实施统一标记,较好地解决了以往句子之间实体标记不一致的问题。针对复杂疾病名称识别困难以及标记不一致的问题, Xu 等^[14]提出使用融合疾病词典和注意力机制的 BI-LSTM+CRF 模型来识别疾病实体名,在领域数据集上取得较好的效果。Lample 等^[15]借鉴基于转移的句法分析模型,搭建 stack-LSTM 与 CRF 结合的神经网络,在语料库 CoNLL-2003 取得较高的 F 值。Zhang 等^[16]使用 Lattice LSTM 模型,编码输入字序列以及词典,显式地利用字符和词序列信息,避免分词错误的传递,同时门控循环胞元可以使模型从句子中选择最相关的字符和词,实验结果表明 Lattice LSTM 优于基于词和基于字符的 LSTM 模型。Liu 等^[17]认为 Lattice LSTM 模型中的快捷路径可能导致模型退化,带来分词错误,故提出 WC-LSTM 模型,把词信息加入整个字符的开头或末尾,用来增强语义信息,在 MSRA 语料上取得较好的成绩。Cui 等^[18]提出一种基于标签嵌入的序列标注方法,通过分层的注意力机制,更好地利用长程依赖关系,同时抛弃 CRF 解码器,当标签数目较多时,提速明显。

1.2 基于卷积神经网络的实体识别编码

由于 RNN 在序列数据建模中需要考虑整个句子的所有字符,其循环特性导致 GPU 的并行计算能力无法充分利用,计算效率较低。CNN 在计算机视觉问题中取得巨大成功,在文本处理方面,近年来有部分学者尝试将其应用到 NER 工作中。

Zhu 等^[19]提出一种新颖的生物学 NER 模型 GRAM-CNN,将单词中每个字母的字符嵌入与单词嵌入与 POS 标签嵌入连接起来,然后利用 CNN 提取 N-gram 特征和单词嵌入的局部上下文,生成每个单词的最终表示输入 CRF 层,整个模型不需要

人为制定特征,取得较好的效果。为了解决基于双向 LSTM 的 NER 模型效率低,无法并行实现的问题,Strubell 等^[20]提出使用迭代膨胀卷积来进行实体识别,同时大幅度地提高训练速度。Cho 等^[21]提出一种模型,可通过组合特征嵌入,有效地表示生物学单词,并通过集成 CNN 和双向 LSTM 中提取的不同字符表示进行增强,将注意力机制应用于减轻 LSTM 模型的长期依赖性问题,评测结果表明该模型具有竞争优势。Chiu 等^[22]提出使用双向 LSTM 和 CNN 混合体系结构自动检测单词和字符特征,在 CoNLL-2003 和 OntoNotes 数据集上的 F1 得分超过采用复杂特征工程和丰富实体链接信息的系统。针对 LSTM 不能充分利用 GPU 并行与候选词典冲突的问题,Gui 等^[23]提出一个融合词典与再思考机制的 CNN 框架,实验结果表明其比基线快 3.21 倍。针对中文 NER 训练数据缺乏,分词困难和标注代价大等实际问题,Wu 等^[24]提出一种 CNN-LSTM-CRF 架构来捕获本地和长距离的上下文,并在模型中加入分词同步训练,产生伪标签来增强训练样本。针对中文分词错误和词汇不足的问题,Zhu 等^[25]提出在字符卷积模块中加入本地注意力机制,在门控 GRU 模块中加入全局注意力机制,用来捕获相邻字符和句子上下文信息,大量的实验结果表明该方法优于无需单词嵌入的方法和外部词典添加的方法。针对 RNN 训练和测试时间成本较高的实际问题,以及 CNN 在处理文本序列方面的劣势,Chen 等^[26]提出在基于 CNN 的 NER 模型中引入门控关系层,通过对关系得分,与其对应的局部上下文特征向量进行加权,组成全局上下文特征向量,增强 NER 的识别效果。

2 模型介绍

2.1 模型整体架构

在用于实体抽取的时候,RNN 按顺序处理输入的序列,可以很好地学习序列信息,但不能很好地学习词之间关联信息,尤其是语义依赖比较复杂的时候,模型较难捕捉,并且运行速度较慢。CNN 通过缩放窗口获取 N-gram 特征,学习词间复杂的语义信息,并且可以多尺度并行计算加速。本文以此为切入点,为了更好地融合上下文信息,在卷积模型的嵌入层引入位置向量,设计一种融合上下文的残差门卷积实体抽取模型,其整体框架如图 1 所示。

2.2 融合上下文的残差门控卷积编码

本文模型在特征编码阶段考虑字符上下文,同时引入位置向量 position embedding 信息^[27],与字符向量拼接,实现混合编码 context embedding,序列经过编码之后,将其送入带残差的门控卷积单元,计算结构如图 2 所示。

$$GLU(X) = f_{w_1}(X) \otimes \sigma(f_{w_2}(X)), \quad (1)$$

可将式(1)单独视为门控线性单元 GLU(Gated Linear Units),本文在一维膨胀卷积的基础上加入门控机制,可以选择性地更好地提取文本高维特征。

进一步地,在 GLU 的基础上引入残差机制:

$$Y = X + f_{w_1}(X) \otimes \sigma(f_{w_2}(X)). \quad (2)$$

当 $f_{w_1}(X)$ 做卷积操作时, $f_{w_1}(X)$ 和 $f_{w_2}(X)$ 形式一样,但权值不共享, $f_{w_2}(X)$ 使用 sigmoid 函数激活(取值在 0~1 之间), $f_{w_1}(X)$ 不加激活函数,因此 $f_{w_1}(X) - X$ 的运算效果与 $f_{w_1}(X)$ 等同,式(3)与式(2)等价。

$$Y = X + (f_{w_1}(X) - X) \otimes \sigma(f_{w_2}(X)), \quad (3)$$

变换得到

$$Y = X \otimes [1 - \sigma(f_{w_2}(X))] + f_{w_1}(X) \otimes \sigma(f_{w_2}(X)). \quad (4)$$

可以解释为,输入信息 X 以 $1 - \sigma(f_{w_2}(X))$ 的概率直接通过,以 $\sigma(f_{w_2}(X))$ 的概率经过 $f_{w_1}(X)$ 变换后才通过,该信息在多通道内传输,可有效地降低模型梯度消失的风险。

2.3 注意力机制

注意力机制可以综合考虑句中的语义信息,利用计算后的注意力向量,同时捕捉词之间的依赖关系,已经在多个 NLP 任务中发挥显著的效果。假设输入文本序列 $X = (\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_n)$, 其中 \mathbf{x}_i 代表第 i 个词的词向量(行向量),维度为 d , $\mathbf{x}_i \in R^d$, 注意力机制就是从文本序列 X 中选择与查询向量 \mathbf{q} 相关的信息。

$$\alpha_i = \text{softmax}(s(\mathbf{x}_i, \mathbf{q})), \quad (5)$$

$$s(\mathbf{x}_i, \mathbf{q}) = \frac{\mathbf{x}_i \mathbf{q}^\top}{\sqrt{d}}, \quad (6)$$

其中, $\alpha_1, \dots, \alpha_i, \dots, \alpha_n$ 是关于 $\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_n$ 的各个向量的相关性度量,也称为 \mathbf{q} 关于 $\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_n$ 的注意力分布, $s(\mathbf{x}_i, \mathbf{q})$ 是注意力打分函数。

根据注意力分布,可以通过加权的方式,计算

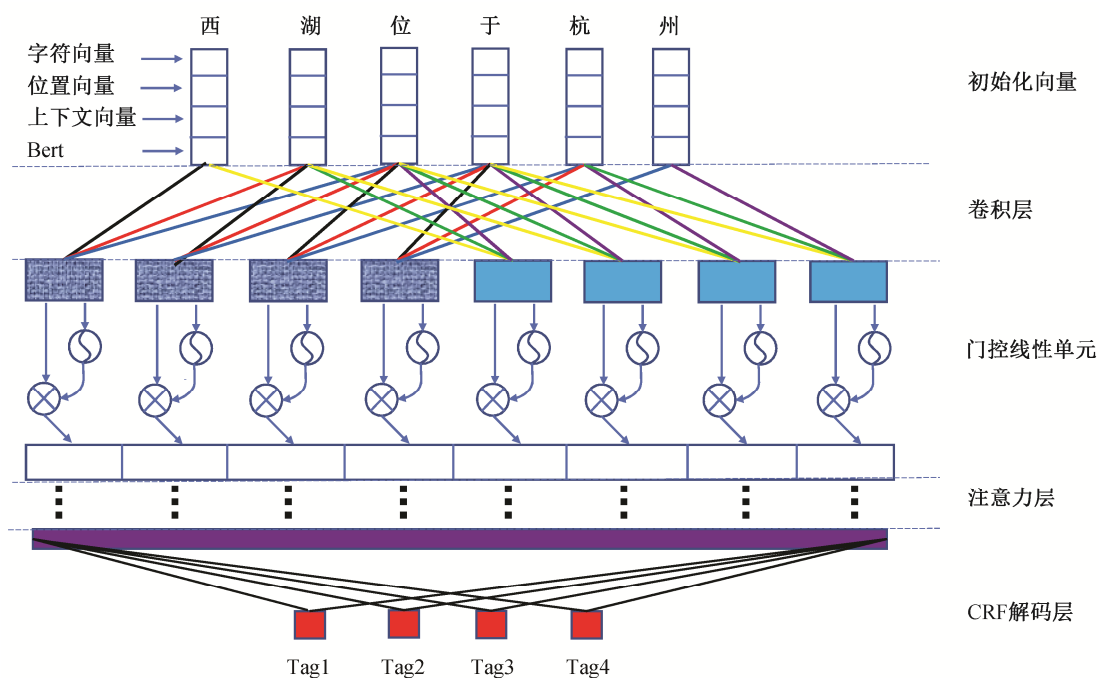


图 1 模型整体框架
Fig. 1 Model framework

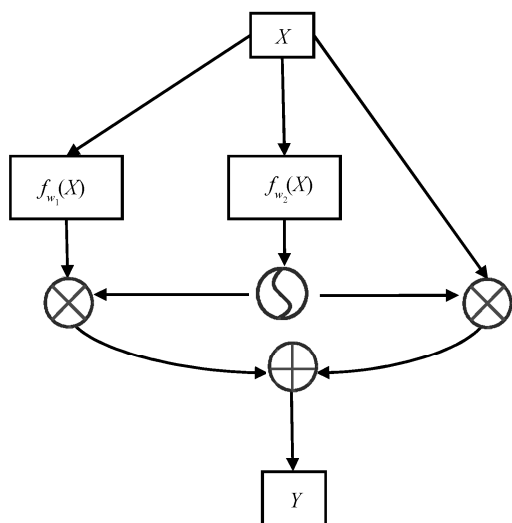


图 2 残差门控线性单元
Fig. 2 Residual gated linear unit

q 对输入信息 $x_1, \dots, x_i, \dots, x_n$ 的最终选择:

$$\text{Attention}(q, [x_1, \dots, x_i, \dots, x_n]) = \sum_{i=1}^n \alpha_i x_i \quad (7)$$

最后, 将多次缩放的点积注意力结果拼接, 再做线性变换, 得到多头注意力的结果。

2.4 输出层

序列经过编码层计算出的是每个词分类的概率, 而 CRF 层引入序列的转移概率, 通过考虑标签

之间的相邻关系来获得全局最优标签序列, 最终计算出 loss 反馈网络。对于输入的文本序列 $X = (x_1, \dots, x_i, \dots, x_n)$, 输出层就是求解概率最大的一组标注序列 $Y = (y_1, \dots, y_i, \dots, y_n)$ 。

文本序列通过编码层后, 可得到对应的分数矩阵 $H \in R^{n \times L}$, $H_{i,j}$ 表示输入序列中第 i 个词对应第 j 个标记的分数, 标签之间的关系可以通过转移矩阵 $T \in R^{L \times L}$ 来刻画, L 为标签数量, 采用负对数似然作为损失函数, Y^* 为正确的标注序列。得分如下式所示:

$$S(X, Y) = \sum_{i=1}^n H_{i,y_i} + \sum_{i=0}^n T_{y_i,y_{i+1}} \quad (8)$$

假设 U 为所有可能的标注序列所构成的集合, 则标注序列概率计算公式为

$$P(Y | X) = \frac{e^{S(X,Y)}}{\sum_{Y' \in U} e^{S(X,Y')}} \quad (9)$$

在训练阶段, 通过反向传播最小化损失函数:

$$L(X, Y^*) = -\log P(Y^* | X); \quad (10)$$

在测试阶段, 使用维特比算法, 将条件概率最大的一组标注序列作为最终的序列标注结果。

3 实验

3.1 数据集与实验环境

本文主要验证基于残差门卷积的实体抽取方法

表 1 数据集说明
Table 1 Details of datasets

数据集名称	类型	训练集	验证集	测试集
MSRA	句子	46.4k	4.4k	4.4k
	字符	2169.9k	172.6k	172.6k
	实体	74.8k	6.2k	6.2k
地理百科	句子	112.3k	5k	5k
	字符	2155.1k	208.4k	208.4k
	实体	110.4k	8.9k	8.9k

的先进性和鲁棒性,选取NER常用任务中的基于循环神经网络BI-LSTM^[1]以及基于卷积架构的IDCNN模型^[20]和GRN模型^[26]作为对比。实验中使用 P 准确率、 R 召回率和 $F1$ 值作为评价指标,分别采用MSRA中文数据集(包含3种实体,分别是人名、机构名和地名)和实验室自行标注的地理百科语料(包含近215万个汉字和1700多个文档中的11万多个地名),相关数据集的说明见表1,所有数据集均采用BIO标注样式。实验环境为Ubuntu18.04操作系统,1080ti显卡,64G内存,其余参数如表2所示。

3.2 模型与评价

3.2.1 模型运算效率分析

为了分析模型的运算效率,在MSRA数据集上选择模型处理同批次样本更新一次权重的单步耗时作为比较基准。由于基于预训练的Bert模型耗时严重,本实验未做比较。

从表3可以看出,基于CNN的模型运行速度普遍比基于RNN的模型快,并且取得的 $F1$ 值也较高。其中,IDCNN的运行速度比BI-LSTM快近3倍,RGCNN的 $F1$ 值比BI-LSTM提升近6.1个百分点,GRN模型的准确率与RGCNN接近,但是单步耗时和召回率不如RGCNN。基于CNN的模型运行速度有明显优势,主要是由于RNN模型要逐步递归才能获得全局信息,而卷积神经网络通过层叠来增大感受野获取信息,并且每层的运算是并行的,

表 2 模型参数
Table 2 Model parameters

模型	参数	值
RGCNN	Kernel size	3
	Filter number	128
	Dilated rate	1, 2, 4, 1, 1
通用	Optimizer	Adam
	Learning rate	0.001
	Dropout	0.5
	Batch size	128
	Epoch	400
	Position embedding	128
	Character embedding	128
Context embedding	256	

因此可以大幅度提升模型的速度。

3.2.2 注意力机制对于模型的影响

注意力机制可以不依赖任何外部的知识或NLP系统,能够自动发现对序列标注起关键作用的语块和标签,使得NER模型能够进一步受到语义约束,实现句中非相邻语块与标签之间的关联计算。

在RGCNN模型中对注意力机制进行消融实验,分别使用自注意力机制Self-attention和多头注意力机制Multi-attention,结果如表4所示,可见RGCNN模型在两组数据集上加入attention机制后效果都有一定程度的提升。其中,在MSRA数据集上3组实

表 3 模型运算效率
Table 3 Model calculation efficiency

模型	单步耗时/ms	P /%	R /%	$F1$ /%
Bi-LSTM	199	85.6	84.6	85.1
IDCNN	82.4	87.1	86.5	86.8
GRN	163.1	91.1	88.7	89.9
RGCNN	124.2	91.9	90.5	91.2

表 4 RGCNN 加上 Attention 的效果
Table 4 Effect of different attentions to RGCNN

模型	MSRA			地理百科		
	P/%	R/%	F1/%	P/%	R/%	F1/%
RGCNN	88.3	86.9	87.6	75.8	77.4	76.6
RGCNN +Self-attention	93.0	90.8	91.9	76.3	78.1	77.2
RGCNN +Multi-attention	93.9	91.5	92.7	76.6	79.6	78.1

验准确率都大于召回率,在地理百科数据集上 3 组实验准确率都小于召回率,可能是由于专业领域百科数据中地名实体较为集中,模型比较容易召回。另外,RGCNN 模型添加 Multi-attention 的效果好于 Self-attention,可能是由于 Multi-attention 采用的是多头注意力机制,与 Self-attention 的单头自注意力机制相比,捕捉到的语块和标签关联特征更加细致。

3.2.3 先验语义对于模型的影响

本文在基础模型 RGCNN 中引入词向量来验证先验语义对模型的影响,分别引入字符向量 character embedding、位置向量 position embedding、上下文向量 context embedding 以及预训练语言模型 Bert 进行对比实验。

从表 5 可以看出,RGCNN 模型加上先验语义后,F1 值均有提升。其中,加入 position embedding 之后 MSRA 的 F1 值提升 1.2%,地理百科数据集提升 2%。可能是因为 position embedding 将每个位置编号,并对应一个向量,帮助 RGCNN 模型中更好地捕捉序列的位置信息,减弱过拟合,从而提高了实体抽取的准确率。在 RGCNN 模型中加入 context embedding 之后,MSRA 提升 4.3%,地理百科数据集提升 3.6%,说明 context embedding 可以通过覆盖上下文语义的编码方式来提高实体抽取的准确率。在 RGCNN 中加入预训练语言模型 Bert 之后,F1 值提

升效果明显,MSRA 提升 4.6%,地理百科数据集提升 4.7%,说明预训练语言模型能够为本文的实体抽取模型提供强大的先验语义信息,但训练时间较长,开销成本较大,本文最终还是采用 RGCNN 架构来添加 context embedding 方案。

3.3 实验结论

在序列标注问题上的 CNN 架构运行效率整体上好于 RNN 架构。本文通过带残差的膨胀门控卷积来控制信息的流动,同时用少量的参数实现更大的感受野,最后利用覆盖上下文语义的编码和基于注意力机制的后处理模块来提升模型对序列数据局部空间特征和标签依赖的联合建模能力。在定量实验中,模型训练速度比 RNN 架构提升约 3 倍,精度也好于 RNN 架构。通过添加 context embedding 和 attention 机制来验证模型实体抽取的能力,实验表明,在计算资源有限的情况下,可以在 RGCNN 的基础上通过添加 context embedding 和 attention 来做 NER 任务,能够在较少的开销下达到接近使用训练语言模型 Bert 的效果。

4 总结与展望

本文针对现有的 CNN 实体抽取框架的缺陷,提出改进方案,在相关数据集上取得具有竞争力的效果,验证了方法的可行性。本文的主要贡献在以下几方面。

表 5 RGCNN 添加不同词向量的效果
Table 5 Effect of adding different word embeddings to RGCNN

模型	MSRA			地理百科		
	P/%	R/%	F1/%	P/%	R/%	F1/%
RGCNN	88.3	86.9	87.6	75.8	77.4	76.6
RGCNN+character embedding	90.8	85.4	88.1	76.1	78.3	77.2
RGCNN+position embedding	90.2	87.4	88.8	77.7	79.5	78.6
RGCNN+context embedding	93.4	90.4	91.9	79.3	81.1	80.2
RGCNN +Bert	92.9	91.5	92.2	79.8	82.8	81.3

说明: position embedding 构造参考文献[27], Bert 模型使用 chinese_L-12_H-768_A-12 版本。

1) 针对传统的实体抽取框架编码方法效率低、精度差等问题, 提出使用基于残差门控卷积的实体抽取方法, 强化有效信息, 降低无效信息的影响, 缓解梯度消失风险以及跨词语义依赖差等难题。

2) 针对传统 CNN 实体抽取框架编码位置编码能力弱、上下文融合效果差的问题, 提出使用基于位置的词向量在嵌入层融合实体的上下文信息, 进一步提升抽取效果。

3) 针对序列标注层语块与标签之间关联计算的缺陷, 提出使用注意力机制, 计算实体特征与其上下文之间的相关性, 进一步捕捉实体语块在序列中的潜在语义特征。

未来工作中, 将研究在网络模型中添加高质量的词典, 并利用迁移学习对少量标注样本进行训练。

参考文献

- [1] Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging [EB/OL]. (2015-08-09) [2021-03-20]. <https://arxiv.org/abs/1508.01991>
- [2] Ma X, Hovy E. End-to-end sequence labeling via bidirectional lstm-cnns-crf [EB/OL]. (2016-01-19) [2021-03-20]. <https://arxiv.org/abs/1603.01354>
- [3] Zhang Y, Yang J. Chinese NER using lattice LSTM [EB/OL]. (2018-11-29) [2021-03-20]. <https://arxiv.org/abs/1805.02023>
- [4] Shang J, Liu L, Gu X, et al. Learning named entity tagger using domain-specific dictionary // Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, 2018: 2054-2064
- [5] Li X, Yan H, Qiu X, et al. FLAT: Chinese NER using flat-lattice transformer [C/OL] // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020 [2021-03-20]. <https://aclanthology.org/2020.acl-main.611.pdf>
- [6] Ma R, Peng M, Zhang Q, et al. Simplify the usage of lexicon in Chinese NER [C/OL] // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020 [2021-03-20]. <https://aclanthology.org/2020.acl-main.528.pdf>
- [7] Li X, Feng J, Meng Y, et al. A unified mrc framework for named entity recognition [C/OL] // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020 [2021-03-20]. <https://aclanthology.org/2020.acl-main.519.pdf>
- [8] Kruengkrai C, Nguyen T H, Aljunied S M, et al. Improving low-resource named entity recognition using joint sentence and token labeling [C/OL] // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020 [2021-03-20]. <https://aclanthology.org/2020.acl-main.523.pdf>
- [9] Chaudhary A, Xie J, Sheikh Z, et al. A little annotation does a lot of good: a study in bootstrapping low-resource named entity recognizers // Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, 2019: 5167-5177
- [10] Rijhwani S, Zhou S, Neubig G, et al. Soft gazetteers for low-resource named entity recognition [C/OL] // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020 [2021-03-20]. <https://aclanthology.org/2020.acl-main.722.pdf>
- [11] Lison P, Barnes J, Hubin A, et al. Named entity recognition without labelled data: a weak supervision approach [C/OL] // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020 [2021-03-20]. <https://aclanthology.org/2020.acl-main.139.pdf>
- [12] Safranchik E, Luo S, Bach S. Weakly supervised sequence tagging from noisy rules. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(4): 5570-5578
- [13] Luo L, Yang Z, Yang P, et al. An attention-based BiLSTM-CRF approach to document-level chemical named entity recognition. Bioinformatics, 2018, 34(8): 1381-1388
- [14] Xu K, Yang Z, Kang P, et al. Document-level attention-based BiLSTM-CRF incorporating disease dictionary for disease named entity recognition. Computers in Biology and Medicine, 2019, 108: 122-132
- [15] Lample G, Ballesteros M, Subramanian S, et al. Neural architectures for named entity recognition // Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. San Diego, 2016: 260-270
- [16] Zhang Y, Yang J. Chinese NER using lattice LSTM // Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne, 2018: 1554-1564

- [17] Liu W, Xu T, Xu Q, et al. An encoding strategy based word-character LSTM for Chinese NER // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis, MN, 2019: 2379–2389
- [18] Cui L, Zhang Y. Hierarchically-refined label attention network for sequence labeling // Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Minneapolis, MN, 2019: 4106–4119
- [19] Zhu Q, Li X, Conesa A, et al. GRAM-CNN: a deep learning approach with local context for named entity recognition in biomedical text. *Bioinformatics*, 2018, 34(9): 1547–1554
- [20] Strubell E, Verga P, Belanger D, et al. Fast and accurate entity recognition with iterated dilated convolutions // Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Copenhagen, 2017: 2670–2680
- [21] Cho M, Ha J, Park C, et al. Combinatorial feature embedding based on CNN and LSTM for biomedical named entity recognition. *Journal of Biomedical Informatics*, 2020, 103: 103381
- [22] Chiu J P C, Nichols E. Named entity recognition with bidirectional LSTM-CNNs. *Transactions of the Association for Computational Linguistics*, 2016, 4: 357–370
- [23] Gui T, Ma R, Zhang Q, et al. CNN-based Chinese NER with lexicon rethinking // Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. Macao, 2019: 4982–4988
- [24] Wu F, Liu J, Wu C, et al. Neural Chinese named entity recognition via CNN-LSTM-CRF and joint training with word segmentation // The World Wide Web Conference. San Francisco, 2019: 3342–3348
- [25] Zhu Y, Wang G. CAN-NER: Convolutional attention network for Chinese named entity recognition // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis, MN, 2019: 3384–3393
- [26] Chen H, Lin Z, Ding G, et al. GRN: gated relation network to enhance convolutional neural network for named entity recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 33(1): 6236–6243
- [27] Gehring J, Auli M, Grangier D, et al. Convolutional sequence to sequence learning // International Conference on Machine Learning. Sydney, 2017: 1243–1252