

文章编号: 1003-0077(2019)09-0024-07

基于膨胀卷积神经网络模型的中文分词方法

王星, 李超, 陈吉

(辽宁工程技术大学 电子与信息工程学院, 辽宁 葫芦岛 125105)

摘要: 目前, 许多深度神经网络模型以双向长短时记忆网络结构处理中文分词任务, 存在输入特征不够丰富、语义理解不全、计算速度慢的问题。针对以上问题, 该文提出一种基于膨胀卷积神经网络模型的中文分词方法。通过加入汉字字根信息并用卷积神经网络提取特征来丰富输入特征; 使用膨胀卷积神经网络模型并加入残差结构进行训练, 能够更好理解语义信息并提高计算速度。基于 Bakeoff 2005 语料库的 4 个数据集设计实验, 与双向长短时记忆网络模型的中文分词方法做对比, 实验表明该文提出的模型取得了更好的分词效果, 并具有更快的计算速度。

关键词: 中文分词; 膨胀卷积; 深度学习; 自然语言处理

中图分类号: TP391 **文献标识码:** A

Dilated Convolution Neural Networks for Chinese Word Segmentation

WANG Xing, LI Chao, CHEN Ji

(School of Electronic and Information Engineering, Liaoning Technical University, Huludao, Liaoning 125105, China)

Abstract: At present, many deep neural network models deal with Chinese word segmentation tasks with bidirectional long short term memory neural network structure. Issues remain in the aspects that the input features are not rich enough, the semantic understanding is not complete, and the calculation speed is slow. In this paper, a dilated convolution neural networks for Chinese Word Segmentation is proposed. Chinese radical information is integrated to enrich the input and the convolution neural network is applied to extract the feature. The dilated convolution neural networks with residual structure can better understand semantic information and improve training efficiency. Experimented on the four datasets in Bakeoff 2005, the proposed method achieves better performance in terms of accuracy and efficiency compared with the bidirectional long short term memory neural network method.

Keywords: Chinese word segmentation; dilated convolution; deep learning; NLP

0 引言

不同于英语中每个单词之间有分隔符, 汉语中词语没有明确的分隔符。因此中文分词 (Chinese word segmentation, CWS) 是中文处理的基础任务, 分词的好坏直接影响后面的自然语言处理任务^[1]。自从 Xue 等^[2]用最大熵模型四词位标注集进行中文分词之后, 许多方法将中文分词作为序列标注问题进行处理, 这样可以使用监督学习^[3-4]或半监督学习^[5-6]的方法处理, 然而这些方法受到特征选择和特征提取的限制^[7]。

近年来深度神经网络模型已经成功用于多项自然语言处理任务中。Collobert 等^[8]提出一种神经网络模型来完成序列标注任务。此后, 许多方法将神经网络模型应用到中文分词任务中。Zheng 等^[9]首先将带窗口的前馈神经网络模型应用到中文分词任务中。Pei 等^[10]在 Zheng 等^[9]的基础上提出最大张量边界模型 MMTNN, 利用张量和标签间的关系表示标签与上下文之间的关系。Chen 等^[11]提出了一种带有重置门和更新门结构的递归神经网络模型 (GRNN) 来保留上下文信息, 并且提出了分层训练防止梯度扩散现象。Chen 等^[12]提出用单向的 LSTM 神经网络来获取前文信息的方法。Yao

收稿日期: 2019-03-11 定稿日期: 2019-04-10

基金项目: 国家自然科学基金(61402212); 辽宁省高等学校杰出青年学者成长计划项目(LJQ2015045); 中国博士后基金(2016M591452); 辽宁省自然科学基金(2015020098)

等^[13]舍弃了窗口的方法,针对单向 LSTM 的不足进行了改进,提出了基于双向的 LSTM 网络的中文分词方法,有效提高了模型的效果。Xu 等^[14]同样在 GRNN 模型上提出了一种改进方案,利用 LSTM 和递归神经网络更好地结合局部信息。Cai 等^[15]提出一种新颖的神经网络模型,并不是采用上下文的窗口,而是构建句子完整的切分历史,直接在结果上进行建模,但存在训练和测试速度慢的问题。Cai 等^[16]在这基础上做了一定的改进,将模型里的 Beam Search 算法改进为贪婪算法,大幅减少了模型训练和测试的时间。Zhang 等^[17]提出了基于词的转移模型,对句子从左往右进行增量解码,能有效利用字和词的融合特征来进行分词。

以上分词模型取得了不错的效果,但还存在以下不足:(1)输入以汉字为主,没有使用汉字的字根特征、五笔特征等更深层的信息;(2)模型以循环神经网络为主。采用循环神经网络特征提取时,根据输入序列长度进行顺序计算,计算效率受到限制。针对第一个不足,本文在输入特征时加入字根特征,字根特征能够有效提高分词效果^[18],使用卷积神经网络对字根进行特征提取,并将字根特征与字符特征融合作为整体输入特征,能够更好地理解汉字语义信息。针对第二个不足,本文模型采用卷积神经网络,卷积神经网络能够有效提高分词效果^[19],卷积神经网络能够独立于序列长度运算,减少训练和评估模型所花费的时间。在整个序列中,卷积神经网络不是递增表示每个字符,而是将卷积核作用于整个序列,其计算成本随着卷积层数的增加而增加,而不再是由输入序列的长度所决定。为了减少卷积神经网络的层数,提高计算效率,本文采用膨胀卷积作为主要模型。与典型的卷积神经网络层一样,膨胀卷积也是使用滑动窗口作用于文本序列上,不同的是膨胀卷积窗口可以跳过宽度为 δ 的输入,通过控制膨胀卷积窗口跳过的字符长度,能够使用更少的卷积层来提取整个序列的特征。此外,本文在卷积层中加入残差结构来解决神经网络因层数增加而出现的退化问题。实验表明,字根特征的引入,并结合加入残差结构的膨胀卷积神经网络,有效改善了中文分词效果,并提高了中文分词的速度。

1 膨胀卷积神经网络模型

中文分词任务通常被认为是序列标注问题,通过所属标注来判断每个字符在词语中的位置,常用

的标注集是 {B, M, E, S}, 利用这四种标注获取词语的边界信息,其中 B、M、E 表示词语的开头、中间、结尾, S 表示单个字。

基于神经网络的序列标注任务通常由三部分组成:(1)浅层特征层,即向量化表示;(2)神经网络层;(3)标签推理层。

1.1 浅层特征层

在双向长短时记忆网络中文分词模型中,浅层特征层使用汉字的嵌入式表示,通常用查找表来找到字符的向量表示。而本文加入字根特征,如图 1 所示。字根特征用卷积神经网络进行特征提出,并与字符特征融合作为模型的输入特征。用 c 表示字符, r 表示字根,则字符 c 的字根列表 $x_r = [r_1, r_2, \dots, r_n]$, 输入特征如式(1)所示。

$$x = \text{CNN}(x_r) \oplus x_c \quad (1)$$

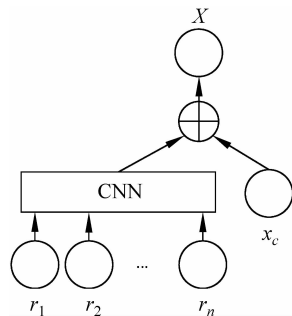


图 1 字符—字根融合特征

1.2 神经网络层

1.2.1 膨胀卷积

卷积操作是卷积神经网络的重要特征,卷积层以特征映射为组织方式,其中的每个单位与前一层的局部感受野相连接,利用共享的卷积核与局部感受野做卷积运算,再经过激活函数做非线性运算来得到特征值。通常神经网络中卷积运算以二维形式出现,但由于自然语言处理任务的特殊性,卷积在自然语言处理任务中以一维卷积的形式出现。对于字符序列 x , 使用一维卷积核 W 卷积,结果 c 可由式(2)得出。

$$c_t = W_c \bigoplus_{k=0}^r x_{t+k} \quad (2)$$

膨胀卷积核与一般卷积具有相同的操作,不同的是可以一次跳过 δ 个输入宽度,在更长的有效输入上定义卷积操作。具体定义如式(3)所示,其中 δ 是膨胀宽度。当 $\delta=1$ 时,膨胀卷积就成为一般卷积操作,当 $\delta>1$ 时,膨胀卷积能够扩展字符上下文

宽度。

$$c_t = W_c \bigoplus_{k=0}^r x_{t \pm k\delta} \tag{3}$$

堆叠的膨胀卷积神经网络能够更好地提取整个文本序列的全局信息^[20],进而更好地理解文本语义。如图 2 所示,当最大膨胀卷积宽度为 4,卷积核滑动步长为 3,堆叠膨胀卷积层为 3 时,堆叠膨胀卷积神经网络能够覆盖到 15 个字符序列长度,当层数为 4 时能覆盖到 31 个字符序列长度,当层数为 n 时,则能覆盖到 $2^{(n+1)} - 1$ 个字符序列长度。覆盖字符序列长度随着层数增加成指数级增长。

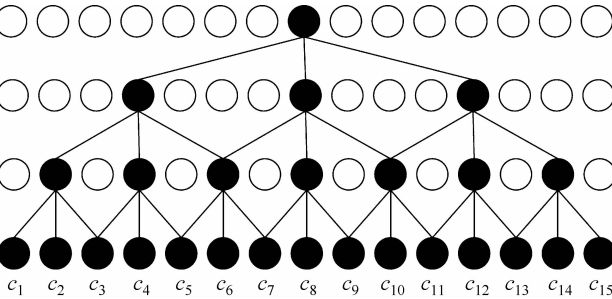


图 2 堆叠膨胀卷积

随着卷积神经网络层数的加深,会出现过拟合现象。在膨胀卷积神经网络中使用批量归一化方法并加入残差结构,使用模块化方法来训练,这种模块化方法将前一层输出特征作为下一层的输入特征,同时能够共享网络参数,能够有效缓解过拟合情况。

1.2.2 膨胀卷积神经网络

定义第 j 层神经网络,膨胀宽度为 δ 的膨胀卷积操作为 $D_\delta^{(j)}$,第一层使用 $\delta=1$ 的膨胀卷积核 $D_1^{(0)}$ 对输入特征表示得到 i_t ,如式(4)所示。

$$i_t = D_1^{(0)} x_t \tag{4}$$

接着使用 n 层膨胀卷积对 i_t 进行卷积操作,开始时 $c_t^{(0)} = i_t$,然后定义堆叠的膨胀卷积如式(5)所示。其中, $r(\cdot)$ 表示 ReLU 激活函数, $\text{bn}(\cdot)$ 表示批量归一化操作。加入残差结构后,定义如式(6)所示。图 3 表示了残差结构。

$$c_t^{(j)} = r(\text{bn}(D_{2^{j-1}}^{(j-1)} c_t^{(j-1)})) \tag{5}$$

$$c_t^{(j)} = r(\text{bn}(D_{2^{j-1}}^{(j-1)} c_t^{(j-1)})) + c_t^{(j-1)} \tag{6}$$

经过 n 层膨胀卷积神经网络操作后,输出时使用 $\delta=1$ 的膨胀卷积网络 $D_1^{(n)}$,如式(7)所示。

$$c_t^{(n+1)} = r(\text{bn}(D_1^{(n)} c_t^{(n)})) \tag{7}$$

将上面的堆叠膨胀卷积作为一个整体模块 $M(\cdot)$,并重复使用 L 次,如式(8)所示,开始时 $m_t^{(1)} = M(i_t)$ 。

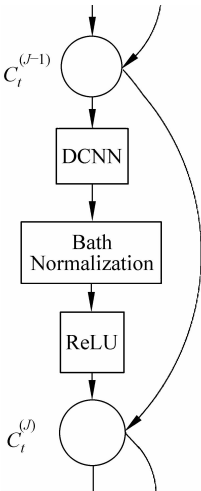


图 3 残差结构

$$m_t^{(k)} = M(m_t^{(k-1)}) \tag{8}$$

最后使用线性变换得到每个字符 x_t 属于每个类别的得分,如式(9)所示。

$$h_t^{(L)} = W_o m_t^{(L)} \tag{9}$$

1.3 标签推理

对于输入句子: $x = (x_1, x_2, \dots, x_n)$, 标签集合 $L = \{B, M, E, S\}$, 在给定条件模型 $p(y | x)$ 下,预测每个字符最有可能的输出标签 $y^* = (y_1^*, y_2^*, \dots, y_n^*)$, 如式(10)所示。

$$y^* = \arg_{y \in L^n} \max p(y | x) \tag{10}$$

本文使用两种方法处理输出标签的预测。首先使用式(11)。这种方法将每个字符的标签认为是条件独立的,使用贪婪算法得到输出标签序列。

$$p(y | x) = \prod_{i=1}^n p(y_i | F(x)) \tag{11}$$

然而输出标签彼此都要相互影响,因此考虑到线性条件随机场模型作为预测模型,如式(12)所示。

$$p(y | x) = \frac{\Psi(y | x)}{\sum_{y' \in L^n} \Psi(y' | x)} \tag{12}$$

其中, $\Psi(y | x)$ 是特征函数。如果具体只考虑连续的两个标签,如式(13)、式(14)所示。

$$\Psi(y | x) = \prod_{i=1}^{n-1} \phi(x, i, y_i, y_{i+1}) \tag{13}$$

$$\phi(x, i, y_i, y_{i+1}) = \exp(s(x, y)) \tag{14}$$

其中, $s(x, y)$ 是每个标签的评分函数,具体如式(15)所示。

$$s(x, y) = \sum_{i=1}^n E_{i, y_i} + \sum_{i=1}^{n-1} T_{y_i, y_{i+1}} \tag{15}$$

线性条件随机场明确表明了相邻标签间的相互作用,虽然需要在输出空间中全局搜索,但能够满足

某些输出约束,具有更好的特征复杂性,对于标签之间的相互关系提供了更多的先验信息。

2 训练

将膨胀卷积神经网络作为一种编码器,由式(9)来得到每个字符的概率。如果将标签看作是相互独立的,可以直接用最大似然方法训练解码。如果将标签看作相互作用的,用条件随机场方法训练解码。

本文采用文献[20]中提出的训练方法。在每个模块之后定义模型的预测,假设 $h_i^{(k)}$ 是由式(9)得出的第 k 个模块的输出结果,式(16)是第 k 个模块后的平均损失函数,训练时最小化模块后的平均损失。

$$\frac{1}{L} \sum_{k=1}^L \frac{1}{T} \sum_{t=1}^T \log P(y_t | h_t^{(k)}) \tag{16}$$

在每个模块之后,通过增强正确的预测值来学习模型,后面的模块则不断优化初始预测。这种处理损失的方法能够缓解深层结构的梯度消失问题,广泛用于计算机视觉领域^[21]。

本文在浅层特征层 x 和每个模块的输出 m_i 之后使用了 dropout 方法来帮助缓解过拟合现象,在膨胀卷积网络中使用批量归一化(bath normalization,BN)方法并加入残差结构,有效规避复杂参数对网络训练产生的影响,加速训练收敛的同时提高了网络的泛化能力。

3 实验设计与分析

3.1 实验环境、数据集和评测标准

本文实验配置为: Inter(R)Core i7-6800K@3.40GHZ 处理器,操作系统 Ubuntu16.04 LTS(64bit),运行内存 16GB(RAM),显卡: GIGA-BYTE GeForce GTX1080Ti,使用 TensorFlow 1.12 构建神经网络模型进行训练和测试。

本文实验数据采用由 Bakeoff2005 提供的 PKU,MSR,CITYU,AS 四种数据集,四种数据集的分词标准各有不同。PKU 是由北京大学计算语言研究所提供的语料库,其分词特点是姓名中姓和名要分开,组织机构等在语法词典中的直接标记,大多数短语性的词语先切分再组合。MSR 是微软亚洲研究院所提供的语料库,其分词特点是由大量的命名实体构成的长单词。AS 是由台湾中央研究院提供的语料库,分词规范与北大制定的分词规范

类似,同时也与台湾地区的语言使用习惯相关。CITYU 是由香港城市大学提供的语料库,分词规范受香港地区的使用习惯影响。数据集规模如表 1 所示。其中随机选取训练数据的 90%作为训练集,10%作为开发集。所有的数据在输入前需要经过预处理,将英文字母替换成 X,数字替换成 0。

表 1 Bakeoff2005 数据集

数据集	训练集词次数	测试集词次数
PKU	1.1M	104K
MSR	2.37M	107K
CITYU	1.46M	41K
AS	5.45M	122K

所有实验均采用标准 Bakeoff 评分程序计算常用的评测指标: P (准确率), R (召回率), F_1 (召回率和准确率的调和平均值),以 F_1 值为主要评测指标。计算公式如式(17)~式(19)所示。

$$P = \frac{\text{输入正确分词的结果个数}}{\text{输入全部结果个数}} \times 100\% \tag{17}$$

$$R = \frac{\text{输入正确分词的结果个数}}{\text{测试集中正确的结果个数}} \times 100\% \tag{18}$$

$$F_1 = \frac{2 \times P \times R}{P + R} \times 100\% \tag{19}$$

3.2 实验设计

本文与中文分词中常用的双向长短时记忆网络模型做对比,解码器分别使用贪婪算法(BiLSTM)和条件随机场(BiLSTM+CRF)。对于膨胀卷积神经网络做了多种对比实验:(1)使用基本的膨胀卷积神经网络模型(DCNN)进行训练;(2)膨胀卷积神经网络模型使用条件随机场解码(DCNN+CRF);(3)膨胀卷积神经网络模型加入残差结构(DCNN+Res);(4)膨胀卷积神经网络模型加入残差结构并使用条件随机场解码(DCNN+Res+CRF)。以上 BiLSTM 模型和 DCNN 模型都使用字符和字根的融合特征作为浅层特征层的输入特征。实验结果如表 2 所示。

表 2 分词模型评价指标对比

模型	指标	数据			
		PKU	MSR	CITYU	AS
BiLSTM	P	94.89	95.49	95.27	96.71
	R	94.71	95.12	95.13	96.49
	F_1	94.80	95.31	95.20	96.60

续表					
模型	指标	数据			
		PKU	MSR	CITYU	AS
BiLSTM+CRF	P	95.77	96.55	95.92	96.90
	R	95.61	95.94	95.94	96.75
	F_1	95.69	96.24	95.93	96.84
DCNN	P	96.10	96.23	96.05	96.90
	R	96.02	96.20	96.07	96.48
	F_1	96.06	96.22	96.06	96.69
DCNN+CRF	P	96.30	96.36	96.19	96.93
	R	96.02	96.23	96.13	96.90
	F_1	96.16	96.30	96.16	96.91
DCNN+Res	P	96.20	96.43	96.09	96.96
	R	96.01	96.16	96.16	96.81
	F_1	96.11	96.30	96.12	96.89
DCNN+Res+CRF	P	96.27	96.48	96.04	97.36
	R	96.13	96.28	96.41	97.14
	F_1	96.20	96.34	96.22	97.25

本文比较了各个模型的处理速度。在每个数据集中,以其中速度最慢的为 1.0,其他模型的速度是与速度最慢的比值。实验结果如表 3 所示。

表 3 分词模型在 Bakeoff2005 数据上处理速度

模型	PKU	MSR	CITYU	AS
BiLSTM	1.15	1.21	1.17	1.16
BiLSTM+CRF	1.00	1.00	1.00	1.00
DCNN	2.95	3.01	2.87	2.99
DCNN+CRF	1.98	2.04	1.85	2.02
DCNN+Res	3.00	3.03	2.83	3.01
DCNN+Res+CRF	2.01	2.01	1.90	2.04

本文将 DCNN 模型与近年来中文分词神经网络模型做对比。实验结果如表 4 所示。

表 4 与其他模型对比

模型	PKU	MSR	CITYU	AS
Chen(2015) ^[12]	94.8	95.6	—	—
Xu(2016) ^[14]	96.1	96.3	—	—
Zhang(2016) ^[17]	95.7	97.7	—	—
Cai(2017) ^[16]	95.8	97.1	95.6	95.3
Chen(2017) ^[22]	94.3	96.0	95.6	94.8
Yang(2018) ^[23]	95.8	97.8	—	—
Gong(2018) ^[24]	96.1	97.7	96.1	95.2

续表				
模型	PKU	MSR	CITYU	AS
Zhang(2019) ^[25]	95.8	97.1	95.2	94.7
DCNN+Res+CRF	96.2	96.3	96.2	97.3

主要参数设置如表 5 所示。字根向量和字符向量采用随机初始化的方式,向量维度都设置为 100。对于浅层特征输入和神经网络中每个模块输出都使用 dropout,且设为 0.15。优化算法使用 Adam 对模型进行优化,初始学习率设为 0.001,并以 1.0 的速度进行衰减。卷积层数为 4。梯度裁剪参数为 5。设置每个 batch 为 128,迭代 100 轮次。

表 5 参数设置

参数名	参数值
字根向量维度	100
字符向量维度	100
初始学习率	0.001
学习衰减率	1.0
Dropout 比率	0.15
CNN 层数	4
Clip_norm	5
Batch_size	128
Max_epochs	100
参数更新方法	Adam

3.3 实验分析

3.3.1 F_1 值结果分析

为了验证条件随机场(CRF)的影响,本文在每个模型解码的时候都分别使用了贪婪算法和条件随机场两种解码方式。以表 2 中 PKU 数据集的实验结果为例,BiLSTM+CRF 比 BiLSTM 提高了 0.89%,DCNN+CRF 比 DCNN 提高了 0.10%,DCNN+Res+CRF 比 DCNN+Res 提高了 0.08%。通过对比可以发现,条件随机场解码确实能够提高模型分词结果,而且对于 BiLSTM 模型的影响较大,对于 DCNN 模型提高的较少。MSR、CITYU、AS 三个数据集也出现同样的现象。表明条件随机场考虑到前后字符之间的相互关系进行全局搜索时,与 BiLSTM 这种与时间序列模型相关性更大。

为了验证 DCNN 模型的效果。将 DCNN 模型

与 BiLSTM 模型的实验结果做对比。以表 2 中 PKU 数据集的实验结果为例,DCNN 模型比 BiLSTM 模型提高了 1.26%,DCNN+CRF 模型比 BiLSTM+CRF 模型提高了 0.47%,DCNN 模型比 BiLSTM+CRF 提高了 0.37%。在 MSR、CITYU、AS 三个数据集中,DCNN(+CRF)模型的 F_1 值也要高于 BiLSTM(+CRF)模型。而且 DCNN 模型在不使用 CRF 情况下,其 F_1 值已经能够达到与 BiLSTM+CRF 模型相同的效果,甚至要更好些。说明 DCNN 模型在处理中文分词中有很好的优势。

为了验证残差结构的效果。将 DCNN+Res 模型与 BiLSTM 模型和 DCNN 模型的实验结果做对比。以表 2 中 PKU 数据集的实验结果为例,DCNN+Res 模型比 BiLSTM 模型提高了 1.36%,DCNN+Res+CRF 模型比 BiLSTM+CRF 模型提高了 0.47%。DCNN+Res 模型比 DCNN 模型提高了 0.05%,DCNN+Res+CRF 模型比 DCNN+CRF 模型提高了 0.04%。加入残差结构后 DCNN 模型的实验结果有一定的提高。

3.3.2 模型速度分析

以表 3 中 PKU 数据集的实验速度为例,DCNN 模型的速度约是 BiLSTM 模型的 2.5 倍,DCNN+CRF 模型的速度是 BiLSTM+CRF 模型的两倍,DCNN+Res(+CRF)和 DCNN(+CRF)速度相差不大。DCNN 模型整体要比 BiLSTM 模型速度要快。每个模型在使用 CRF 解码时,速度变慢许多。在 MSR、CITYU、AS 三个数据集中,实验速度的结果分析和 PKU 数据集中的实验结果相同。说明 DCNN 在处理速度上具有很大优势。

综合表 2 和表 3 分析。DCNN 模型 F_1 值比 BiLSTM 模型好,处理速度更快。DCNN 模型不使用 CRF 情况下与 BiLSTM+CRF 模型的 F_1 值相差不大,处理速度却约是 BiLSTM+CRF 模型的 3 倍。而加入残差结构后,DCNN+Res 模型的 F_1 值和速度都要比 BiLSTM+CRF 模型好。通过分析比较,DCNN 模型在 F_1 值和速度上都取得更好的效果。

3.3.3 与其他模型对比分析

表 4 列出与其他模型的对比。比较对象为 Chen 扩展的 LSTM 模型,Xu 的 GRNN 改进模型,Zhang 的基于词的转移模型模型,Chen 的对抗神经网络模型,Cai 的句子切分模型,Yang 的 Lattice LSTM 模型,Gong 的 LSTM 转换模型,Zhang 的 LSTM 联合学习模型。

由表 4 可知,DCNN 模型与近年来深度学习中中文分词模型相比中取得不错的结果。PKU 数据集和 CITYU 数据集略好与其他模型结果,比 Gong 提高了 0.1%。MSR 数据集上结果略差些。AS 数据集的实验结果得到了很好的提高,比 Gong 提高了 2.1%。

4 结论

本文工作主要有以下几点:(1)针对自然语言处理中的中文分词任务,使用膨胀卷积神经网络结构进行分词;(2)使用字根特征来丰富输入特征,能够更好地理解文本语义信息;(3)将残差结构加入神经网络中,有效缓解了神经网络中梯度消失的问题;(4)在训练时使用模块化方法,并优化模块中平均损失。实验结果表明本文使用的 DCNN 神经网络的方法相比于传统的 BiLSTM 模型能够提高分词的结果,并且能够大幅度提高分词的速度。

本文使用的膨胀卷积神经网络虽然能够对中文分词结果进行有效的改善,但还有进一步改进的空间。在实验中发现,CRF 解码方式严重影响中文分词的速度,后续工作将寻找更高效的解码方式来优化。

参考文献

[1] 黄昌宁,赵海.中文分词十年回顾[J].中文信息学报,2007,21(3):8-19.

[2] XUE N. Chinese word segmentation as character tagging [J]. International Journal of Computational Linguistics and Chinese Language Processing, 2003, 8 (1):29-47.

[3] 宋彦,蔡东风,张桂平.一种基于字词联合解码的中文分词方法[J].软件学报,2009,20(9):2366-2375.

[4] 黄德根,焦世斗,周惠巍.基于子词的双层 CRFs 中文分词[J].计算机研究与发展,2010,47(5):962-968.

[5] 邓丽萍,罗智勇.基于半监督 CRF 的跨领域中文分词[J].中文信息学报,2017,31(4):9-19.

[6] 韩冬煦,常宝宝.中文分词模型的领域适应性方法[J].计算机学报,2015,38(2):272-281.

[7] Bengio Y, Ducharme R, Vincent P, et al. A neural probabilistic language model [M]. Innovations in Machine Learning. Springer Berlin Heidelberg, 2006: 137-186.

[8] Collobert R, Jason W, Michael K, et al. Natural language processing (almost) from scratch [J]. Journal of Machine Learning Research, 2011, 12(1): 2493-2537.

- [9] Zheng X, Chen H, Xu T. Deep learning for Chinese word segmentation and POS tagging [C]//Proceedings of Conference on Empirical Methods in Natural Language Processing, 2013: 647-657.
- [10] Pei W, Ge T, Chang B. Max-margin tensor neural network for Chinese word segmentation [C]//Proceedings of Meeting of the Association for Computational Linguistics, 2014: 293-303.
- [11] Chen X, Qiu X, Zhu C, et al. Gated recursive neural network for Chinese word segmentation [C]//Proceedings of Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015: 1744-1753.
- [12] Chen X, Qiu X, Zhu C, et al. Long short-term memory neural networks for Chinese word segmentation [C]//Proceedings of Conference on Empirical Methods in Natural Language Processing, 2015: 1197-1206.
- [13] Yao Y, Hhang Z. Bi-directional LSTM recurrent neural network for Chinese word segmentation [C]//Proceedings of International Conference on Neural Information Processing. Springer, Cham, 2016: 345-353.
- [14] Xu J, Sun X. Dependency-based gated recursive neural network for Chinese word segmentation [C]//Proceedings of Meeting of the Association for Computational Linguistics, 2016: 567-572.
- [15] Cai D, Zhao H. Neural word segmentation learning for Chinese [J]. Computing Research Repository, 2016: 409-420.
- [16] Cai D, Zhao H, Zhang Z, et al. Fast and accurate neural word segmentation for Chinese [J]. Association for Computational Linguistics, 2017: 608-615.
- [17] Zhang M, Zhang Y, Fu G. Transition-based neural word segmentation [C]//Proceedings of Meeting of the Association for Computational Linguistics, 2016: 421-431.
- [18] Dong C, Zhang J, Zong C, et al. Character-based LSTM-CRF with radical-level features for Chinese named entity recognition [C]//Proceedings of International Conference on Computer Processing of Oriental Languages. Springer International Publishing, 2016: 239-250.
- [19] Wang C, Xu B. Convolutional neural network with word embeddings for chinese word segmentation [C]//Proceedings of the 8th International Joint Conference on Natural Language Processing, 2017: 163-172.
- [20] Emma S, 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, 2017: 2670-2680.
- [21] Christian S, Liu W, Jja Y, et al. Going deeper with convolutions [J]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 1-9.
- [22] Chen X, Shi Z, Qiu X, et al. Adversarial Multi-Criteria Learning for Chinese Word Segmentation [C]//Proceedings of ACC, 2017: 1193-1203.
- [23] Yang J, Zhang Y, Liang S. Subword encoding in lattice LSTM for chinese word segmentation [C]//Proceedings of Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2019: 2720-2725.
- [24] Gong J, Chen X, Gui T, et al. Switch-LSTMs for multi-criteria chinese word segmentation [C]//Proceedings of the 33th AAAI Conference on Artificial Intelligence, 2019: 6457-6464.
- [25] 章登义, 胡思, 徐爱萍. 一种基于双向 LSTM 的联合学习的中文分词方法 [J/OL]. 2019, 36 (10). [2018-07-09]. <http://www.arocmag.com/article/02-2019-10-060.html>.



王星(1983—),通信作者,博士,副教授,主要研究领域为智能数据与知识工程。

E-mail: xingwang_lntu@163.com



李超(1991—),硕士研究生,主要研究领域为自然语言处理、知识图谱。

E-mail: ChaoLeed@163.com



陈吉(1981—),硕士,工程师,主要研究领域为知识工程、图像处理。

E-mail: jichen_lntu@163.com