

文章编号: 1003-0077(2021)01-0072-09

一种基于门控空洞卷积的高效中文命名实体识别方法

王笑月, 李茹, 段菲

(山西大学 计算机科学与技术学院, 山西 太原 030006)

摘要: 近年来, 基于 RNN 的模型架构在命名实体识别任务中被广泛采用, 但其循环特性导致 GPU 的并行计算能力无法被充分利用。普通一维卷积虽可以并行处理输入文本, 显著缩短模型训练时长, 但处理长文本时往往需要堆叠多个卷积层, 进而增加梯度消失的风险。针对以上问题, 该文采用可通过参数调节感受野范围的空洞卷积, 并引入了带有残差连接的门控机制, 以强化有效信息, 降低无效信息的影响, 同时改善梯度消失问题; 针对字向量表示能力有限的问题, 该文还将字向量与其所属词的位置信息融合, 以丰富文本特征。为验证所提出方法的有效性, 在 MSRA 数据集以及 Sina Resume 数据集上进行了实验, F_1 值分别达到了 92.97% 与 94.98%。与传统基于 Bi-LSTM-CRF 的命名实体识别模型相比, 模型训练速度提升 5~6 倍, 且表现优于一般的 RNN 架构。

关键词: 空洞卷积; 门控机制; 中文命名实体识别

中图分类号: TP391

文献标识码: A

An Efficient Chinese Named Entity Recognition Method Based on Gated-Dilated Convolution

WANG Xiaoyue, LI Ru, DUAN Fei

(School of Computer Science and Technology, Shanxi University, Taiyuan, Shanxi 030006, China)

Abstract: To further improve gated convolution neural network for Chinese named entity recognition (NER) and resolve the gradient vanishing occurred in the stacking multiple convolution layers, adopt recurrent architecture, in particular long short-term memory networks (LSTM). Because of the recurrent nature of those methods, the parallel computing capability of GPU cannot be utilized at their full potential. Although vanilla 1-D convolution operation can be adopted to process texts in parallel, stacking multiple layers are often necessary to obtain satisfactory receptive fields so as to better model long-range dependencies in texts. we propose to replace vanilla convolution operations with recently proposed dilated convolution operations whose receptive fields can be controlled via a dilated factor. To further strengthen effective information and reduce the negative impacts induced by invalid information, we propose a gating mechanism with residual connections. To enrich textual features, we also fuse the character embedding with the word position information. Validated on MSRA dataset and Sina Resume dataset, the results show that, compared with conventional Bi-LSTM-CRF models, our proposed method shows very promising performance, as well as 5x~6x speedup in the training phase compared to RNN architecture.

Keywords: dilated convolution; gated mechanism; Chinese named entity recognition

0 引言

作为自然语言处理中的一项基础任务, 命名实体识别旨在从一段非结构化文本中抽取与任务要求相关的实体, 例如人名、地名、机构名等。它在机

器翻译、信息抽取、问答系统等需要关键内容抽取技术的领域具有广泛的应用价值。

传统命名实体识别主要基于特征工程, 即由线性统计模型(如隐马尔可夫(HMM)^[1]、支持向量机(SVM)^[2]或条件随机场(CRF)^[3]等)与人工定义的特征相结合构造模型, 但由于自然语言语法结构的

收稿日期: 2019-09-09 定稿日期: 2019-10-13

基金项目: 国家自然科学基金(61772324); 国家重点研发计划(2018YFB1005100)

多样性,基于特征工程的方法需要针对不同的数据、结合领域知识人工总结特征和定义规则,其缺乏灵活性且无法保证泛化能力。高性能通用并行计算设备(如 GPGPU、TPU)的出现与进步使一系列神经网络架构被成功应用于序列标注任务中,Ling 等^[4]、Ma 等^[5]、Lample 等^[6]提出的基于 RNN 的神经网络模型虽然取得了优异的结果,但由于其采用循环处理序列的方式,故未能充分利用 GPU 的并行计算能力。

不同于 RNN 循环处理输入序列的方式,卷积运算可一次性地并行处理整个输入序列。虽然在并行处理方面卷积神经网络具有先天的优势,但通常用于文本处理的一维卷积感受野受限输入窗口大小 w 。在第 l 层,卷积对于输入文本的感受野大小为 $r=l(w-1)+1$,当输入序列长度增加时,需要堆叠的层数也随之增加。为扩大感受野,同时减少堆叠层数,可以采用池化操作,但池化会导致信息损失,因而不适用于序列标注任务。为在同样的窗口大小下扩大卷积的感受野,Yu 等^[7]提出了空洞卷积,其在相同的过滤器上通过更改空洞率 D 的大小,在不同范围的文本上操作,使其有效感受野随网络深度以指数级增长,且不产生信息损失,也不会增加模型参数;Strubell 等^[8]首次将空洞卷积用于命名实体识别任务上,但只是简单地将 Bi-LSTM 替换为空洞卷积,并未进一步对信息流动的方式进行考虑和处理;Dauphin 等^[9]在普通卷积的基础上引入了门控机制,进一步强化了数据中的有效信息,降低了无效信息的影响,实验结果表明其在语言建模任务上是有效的;Wang 等^[10]将这种门控机制运用到命名实体识别任务中,首次使用门控卷积结构进行实体识别。

Wang 等^[10]所提出的模型易出现梯度消失的问题,本文针对该缺陷进行改进,采用可通过参数调节感受野范围的空洞卷积,并引入了带有残差连接^[11]的门控机制(详见 2.2 节)。

除模型架构外,自然语言处理任务中的另一关键点在于如何更好地表示输入文本。对于输入文本常采用基于字的划分粒度或基于词划分粒度两种方式。在基于字的模型中,每个字通常只有一种表示,无法兼顾一字多义的情况,例如,“京”在“北京”和“京东”中有着不同的意义;而基于词的模型常因为分词错误和误差传递导致实体识别错误,而且在真实应用中构建领域相关的大规模词典代价昂贵,在匹配特征时时间和空间复杂度较高,从而导致模型

十分低效;此外,大规模词典中带来的噪声影响难以消除。本文使用 Chen 等^[12]所提出的基于位置的字向量,针对每一个字在所属词中的位置,对于每个字给予了四种预训练的向量,从而在字级别的基础上针对每一个字加入了所属词的位置特征,丰富了文本信息表示,同时也在一定程度上改善了一字多义的问题。Chen 等^[12]以词为划分粒度结合基于位置的字向量得到文本表示,继而输入模型中,本文考虑到以词为划分粒度易产生 OOV(out-of-vocabulary)问题,故以字为划分粒度。

本文的主要贡献可概括为以下两方面:

(1) 首次在中文数据集上使用空洞卷积进行命名实体识别,显著提升了模型训练速度;

(2) 为避免由于网络层数加深而产生的梯度消失现象,并使信息可以在多通道内传输,引入了带有残差连接的门控机制。

1 相关工作

传统模型大多基于手工定义的特征和线性链条件随机场(He 等^[13],Mao 等^[14])。基于 CNN 或者 RNN 等的神经网络模型效果超过传统的机器学习方法而无需手工定义特征。由于 LSTM 具有长期记忆的特点,LSTM-CRF 架构被广泛运用到命名实体识别的任务中。Huang 等^[15]使用 Bi-LSTM 抽取词级别的文本信息;Lample 等^[6]在其基础上结合了字级别的文本特征;Peng 等^[16]基于如何更丰富地表示文本信息问题,使用 Chen 等^[12]的字词嵌入融合方法,提出了一个“字+位置”嵌入和 NER 任务联合训练的模型;Zhang 等^[17]构造了一个词格模型,通过使用外部词典,匹配文本中所有可能的分词结果,进而避免了分词错误;Cao 等^[18]考虑到中文分词任务和中文命名实体识别任务存在共同特性,即实体边界识别,使用对抗学习来联合训练命名实体识别任务和中文分词任务,抽取共享的词边界信息;Tan 等^[19]通过自注意力机制捕捉输入文本的全局信息,结果表明自注意力机制在命名实体识别任务中的有效性;最近,随着 ELMo^[20]和 BERT^[21]等大型预训练语言模型的出现,在自然语言处理的下游任务中,许多工作使用了这类模型并获得了显著的性能提升,但由于其参数量巨大,仅微调需要的时间也较长,本文出于降低资源消耗的考虑,未采用这两种方式。

2 模型架构

本文将实体识别视为序列标注的任务,为文本中的每个字预测对应的标签,并对文本采用 BIO^[22]的标注方式,即字位于实体开头标注为 B,位于实体内部标注为 I,非实体的字标注为 O。

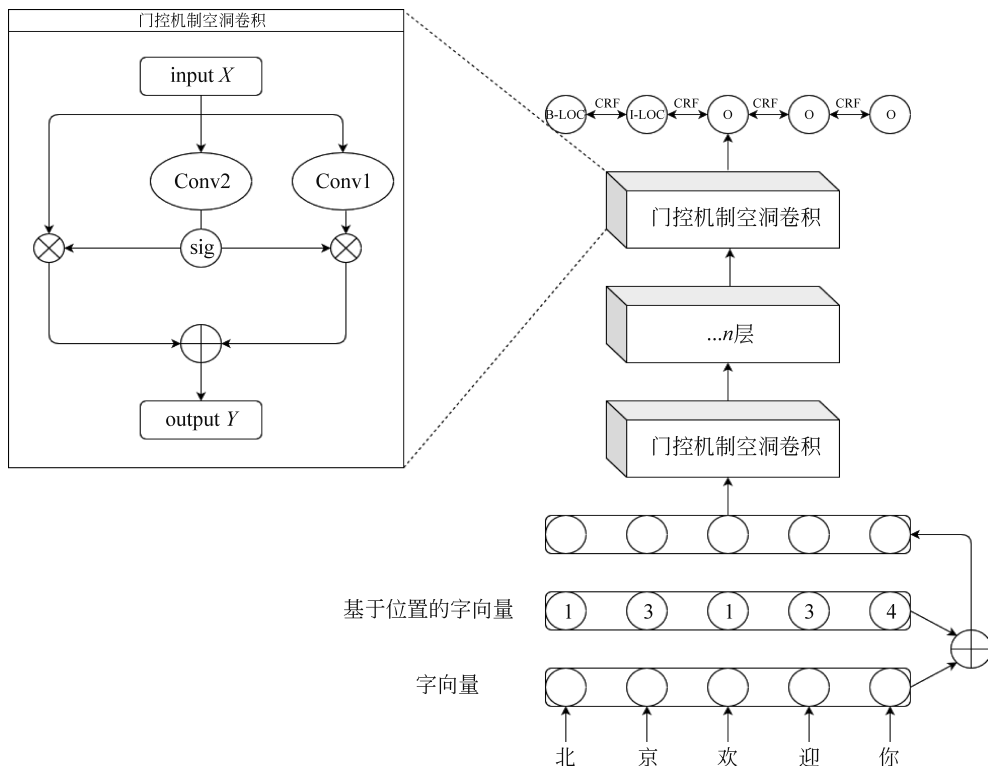


图1 模型架构

2.1 空洞卷积

自然语言处理任务中所使用的卷积类型通常为一维卷积。卷积层相当于仿射变换,对每个字,卷积操作如式(1)所示,其中 W_c 表示窗口大小为 r 的滤波器, c_t 代表经过卷积变换后的输出, \oplus 代表向量拼接操作。

$$c_t = W_c \bigoplus_{k=0}^r x_{t \pm k} \quad (1)$$

空洞卷积(Yu 等^[7])是卷积的一种变体,与普通卷积不同,其跨越与空洞率 d 相同的文本片段对输入文本间隔采样,通过堆叠以指数级别增长的空洞率的空洞卷积(例如,本文所采用的空洞率为 1, 2, 4, 1, 2, 4),可以实现以较少的层数覆盖到大多数句子的长度,例如第 L 层的感受野可达 $2^{L+1} - 1$ 。如果设置卷积核窗口大小为 3,堆叠 4 层空洞卷积,则其有效感受野的宽度为 31,而普通的卷积感受野

模型整体架构如图 1 所示,其中“门控机制空洞卷积”单元为带有门控机制的卷积层,细节将在 2.2 节中阐述,为张量拼接操作;将输入文本的字向量与所对应的基于位置的词向量拼接,所得到的文本表示输入到若干个堆叠的卷积层中,之后通过 CRF 解码得到实体识别结果。

仅为 9。

空洞卷积的计算方式如式(2)所示。

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

其中 δ 为空洞率, \oplus 为向量拼接操作, W_c 为窗口大小为 r 的滤波器, c_t 为输出。

2.2 带有残差连接的门控机制

为强化数据中的有效信息,并增加流通渠道的多样性,Dauphin 等^[9]在卷积的基础上加入了门控机制,提升了语言模型实体识别的效果。本文在其基础上进一步拓展,引入了残差机制,以改善深层网络带来的梯度消失问题,同时使信息可在多通道内传输。本文采用的门控卷积架构如图 2 所示。

每一个卷积门控单元内的操作可用式(3)表示:

$$Y = X \otimes (1 - \sigma(\text{Conv2}(X))) + \sigma(\text{Conv2}(X)) \otimes (\text{Conv1}(X)) \quad (3)$$

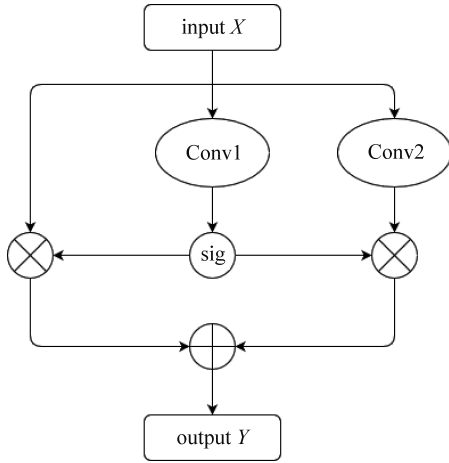


图2 带有残差连接的门控卷积架构

其中 X 表示输入, Conv1 定义为卷积操作 1, Conv2 定义为卷积操作 2, Conv1 与 Conv2 都为空洞卷积, 二者所设定的滤波器个数、窗口大小一致, 但权值不共享, σ 代表 sigmoid 激活函数, \otimes 代表张量的 Hardamard 积, $+$ 代表张量相加。

2.3 基于词中位置的字嵌入

采用字向量作为输入, 可避免分词错误带来的影响, 同时不存在 OOV 的问题, 但其所包含的文本信息较少, 而词向量又无法充分利用词中的字间关系, 因此本文提出一种融合字向量与词特征的方案, 具体介绍如下。

考虑到每一个字在词中的位置有四种情况: ①位于词的开始; ②位于词的中间; ③位于词的末尾; ④单个字表示词。故构建特征矩阵 $W \in R^{4 \times s}$, 其中 4 表示四种位置情况, s 为词特征矩阵的维度。

给定输入序列 $X = \{x_1, x_2, \dots, x_n\}$, 其中 x_i 代表输入序列中的字, 在经过分词之后, 得到输入词序列 $C = \{c_1, c_2, \dots, c_m\}$, 其中 c_i 表示序列中索引为 i 的词。按照划分好的词信息, 对每个字赋予位置标签, 0 表示所属词的开始, 1 表示所属词的中间, 2 表示所属词的末尾, 3 表示单个字为词。基于位置的词向量表示如图 3 所示。

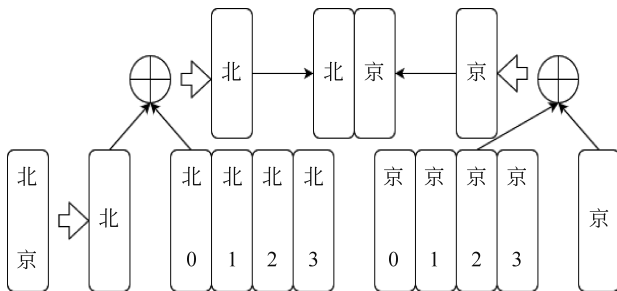


图3 基于位置的词向量表示

根据所赋予的位置信息标签, 对文本序列中的每一个字在特征矩阵 W 中查找对应的词特征向量, 继而得到与输入文本相对应的词特征向量矩阵 $E_2 \in R^{n \times s}$ 。

将最终得到的字向量矩阵 $E_1 \in R^{n \times c}$ (其中 n 表示输入序列长度, c 表示字向量矩阵维度) 与词特征向量矩阵 $E_2 \in R^{n \times s}$ 进行拼接, 得到输入向量矩阵 $E \in R^{n \times (c+s)}$ 。

2.4 解码与训练

通过构造转移矩阵, 条件随机场可考虑到相邻标签之间标注的合理性, 并输出一个概率值最大的标注序列。本文在模型输出层采用了条件随机场。

给定一个文本 $X = \{x_1, x_2, \dots, x_n\}$, 通过 CNN 层后可得到对应的分数矩阵 $H \in R^{n \times L}$, 其中 $H_{i,j}$ 表示输入序列中第 i 个单词对应第 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, 其中 L 为标签数量。令其可能的标注序列表示为 $Y = \{y_1, y_2, \dots, y_n\}$, 定义其得分如式(4)所示。

$$S(X, Y) = \sum_{i=1}^n H_{i, y_i} + \sum_{i=0}^n T_{y_i, y_{i+1}} \quad (4)$$

其标注序列概率计算方式为式(5):

$$P(Y | X) = \frac{e^{S(X, Y)}}{\sum_{Y' \in U} e^{S(X, Y')}} \quad (5)$$

其中 U 为所有可能的标注序列构成的集合, 采用负对数似然作为损失函数, 如式(6)所示。

$$L(X, Y^*) = -\log P(Y^* | X) \quad (6)$$

其中 Y^* 为正确的标注序列。在训练阶段, 通过反向传播最小化该损失函数; 在测试阶段, 使用维特比算法将条件概率最大的标注序列作为最终的输出。

3 实验

3.1 数据集

本文在已经公开划分好的 Sina Resume 数据集^[17]和 SIGHAN Bakeoff 2006^[23]的 MSRA 数据集上进行实验, 两个数据集的规模如表 1 所示。

Resume 数据集包含八种实体, 分别是国家 (CONT)、教育组织 (EDU)、地点 (LOC)、人名 (PER)、机构 (ORG)、职业 (PRO)、种族 (RACE)、职业名 (TITLE)。

表 1 实验所采用数据集介绍

数据集	种类	训练集(K)	测试集(K)	验证集(K)
MSRA	句子	46.4	4.4	—
	字符	2 169.9	172.6	—
	实体	74.8	6.2	—
Sina Resume	句子	3.8	0.48	0.46
	字符	124.1	15.1	13.9
	实体	1.34	0.15	0.16

MSRA 数据集包含三种实体,分别是人名(PER)、机构名(ORG)、地名(LOC)。

3.2 评价指标

实验采用 F_1 值来评估命名实体识别效果,其中 F_1 值由 P (准确率)和 R (召回率)来决定。计算公式如式(7)~式(9)所示。

$$P = \frac{\text{正确识别的实体个数}}{\text{识别的实体个数}} \times 100\% \quad (7)$$

$$R = \frac{\text{正确识别的实体个数}}{\text{语料中所有的实体个数}} \times 100\% \quad (8)$$

$$F_1 = \frac{2 \times P \times R}{P + R} \times 100\% \quad (9)$$

3.3 实验设置

模型参数设置如表 2 所示。

模型中字向量(char-embedding)维度设置为 100,词特征向量维度(seg-embedding)设置为 20,词特征矩阵采用了随机初始化的方式,字向量使用 Word2Vec^[24] 工具训练,选用了 skip-gram^[24] 模型,上下文扫描窗口设置为 5,负采样值设置为 8,迭代次数设置为 8。采用了 BIO 标注方式进行文本标注,空洞卷积的窗口大小设置为 3,滤波器个数设置为 120,采用优化函数为 Adam,初始学习率大小为 0.001,学习率缩减步长 lr_decay 设置为 0.05,dropout 率设置为 0.5,batch_size 设置为 32,epoch 设置为 100。

表 2 模型参数设置

模型层	参数	参数值
Word2Vec	model	skip-gram
	negative sampling	8
	window size	5
	iteration	8

续表

模型层	参数	参数值
空洞卷积	kernel size	3
	filter number	120
	dilated rate	1,1,2,4,1,2,4
	learning_rate	0.001
	lr_decay	0.05
	dropout	0.5
	optimizer	Adam
	batch_size	32
	epoch	100
	char-embedding	100
	seg-embedding	20

3.4 实验结果与分析

本实验采用准确率、召回率和 F_1 值作为评估指标。

3.4.1 MSRA 数据集命名实体识别结果与分析

表 3 展示了在 MSRA 数据集上的实验结果,将本文提出的模型同先前的实验结果做了对比。其中 baseline 为不加门控机制与词特征的普通空洞卷积;Chen 等^[25]、Zhang 等^[26]、Zhou 等^[27] 运用了丰富的人工定义特征结合 CRF 进行命名实体识别;Dong 等^[28] 采用 Bi-LSTM-CRF 的架构进行中文命名实体识别,并且基于汉字的组合结构构建了 radical-level 特征,将其与字向量融合;Zhou 等^[29]

表 3 MSRA 数据集命名实体识别实验结果(%)

模型	P	R	F_1
Chen 等(2006) ^[25]	91.22	81.71	86.20
Zhang 等(2006) ^[26]	92.20	90.18	91.18
Zhou 等(2013) ^[27]	91.86	88.75	90.28
Dong 等(2016) ^[28]	91.28	90.62	90.95
Wang 等(2017) ^[10]	92.34	90.15	91.23
Cao 等(2018) ^[18]	91.30	89.58	90.64
Zhou 等(2018) ^[29]	92.04	91.31	91.67
Zhang 等(2018) ^[17]	93.57	92.79	93.18
baseline	91.64	90.15	90.89
dgcnn	93.51	92.44	92.97

使用了基于字符级别的 CNN-Bi-LSTM-CRF 的架构,通过 Bi-LSTM 来捕捉文本全局信息,CNN 捕捉文本局部信息;Wang 等^[10]使用完全基于卷积的架构,并在卷积基础上引入了门控机制;Cao 等^[18]使用了联合学习的方法,提取分词和命名实体识别中的共同词边界信息;Zhang 等^[17]提出了一种词格结构,将词典信息加入到神经网络中,避免了分词错误。虽然其模型性能最优,但运用了外部数据,且结果依赖于词典的质量。本文所提出的模型在不引入外部资源的情况下,效果达到了最优。

3.4.2 Resume 数据集命名实体识别结果与分析

Resume 数据集实体识别结果如表 4 所示,(1)表示 Zhang 等^[17]所采用的基于字的 LSTM 模型,(2)为 Zhang 等^[17]所采用的基于词的 LSTM 模型,(3)为 Zhang 等^[17]所提出的词格模型。Zhang 等^[17]使用了词格模型,并且加入了额外的词典信息,可以看到本文提出的模型性能达到了 F_1 值 94.98%,显示出本文所提出模型的优越性。

表 4 Resume 数据集命名实体识别实验结果(%)

模型	P	R	F_1
Zhang 等 ^[17] (1)	94.53	94.29	94.41
Zhang 等 ^[17] (2)	94.07	94.42	94.24
Zhang 等 ^[17] (3)	94.81	94.11	94.46
baseline	93.72	93.70	93.71
dgcnn	95.05	94.91	94.98

通过对比实验结果,发现在 MSRA 数据集上,本文所提出的模型结果略低于 Zhang 等^[17],而在 Resume 数据集上,本文所提出的模型结果高于 Zhang 等^[17];笔者分析导致该现象的原因之一为,对于 MSRA 数据集,数据涉及的领域范围较广,在不同文本场景下,同一实体边界粒度不一致,模型无法从数据集中获取足够的信息,准确地区分不同场景下的实体边界,较依赖于外部资源;但对于 Resume 数据集,数据涉及的范围领域较单一,实体边界划分较准。故在 MSRA 数据集上本文所提出模型效果低于 Zhang 等^[17],而在 Resume 数据集上模型效果略高于 Zhang 等^[17]。

3.4.3 对比实验

为了验证所提出改进方案的有效性,分别在 MSRA 和 Resume 数据集上进行了一系列的对比实验,结果分别如表 5 和表 6 所示。为增强结果的直观性,将表 5 结果可视化于图 4 中,而将表 6 结果

可视化于图 5 中。其中 baseline 设置为普通空洞卷积,baseline+gate 为空洞卷积加上门控机制,baseline+seg 为空洞卷积加上词特征,dgcnn 为本文所提出的模型,dgcnn-residual 为本文所提出的模型去除残差结构部分。

表 5 MSRA 数据集命名实体识别对比实验结果(%)

模型	P	R	F_1
baseline	91.64	90.15	90.89
baseline+gate	92.54	90.40	91.46
baseline+seg	92.87	91.22	92.04
dgcnn	93.51	92.44	92.97
dgcnn-residual	93.42	91.85	92.63

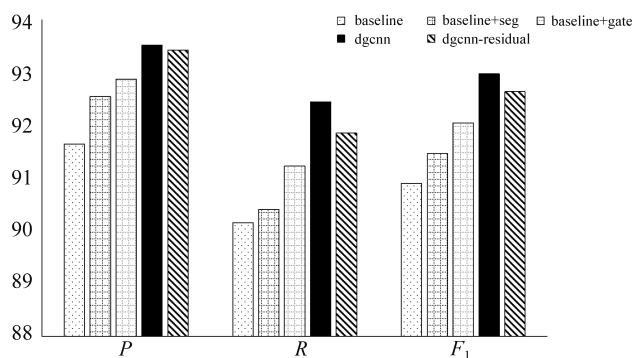


图 4 MSRA 数据集命名实体识别对比实验结果

由表 5 和图 4 可看出,引入带有残差连接门控机制和词特征的模型(dgcnn),与基于普通空洞卷积的模型(baseline)相比,其准确率和召回率都有所提升,表示出在无效信息的过滤以及文本表示上具有更优的性能。在 MSRA 数据集上 F_1 值提升 2.08%;相比于未引入残差连接门控机制的空洞卷积模型(dgcnn-residual), F_1 值提升 0.34%;相比于未引入词特征的空洞卷积与带有残差连接的门控机制相结合模型(baseline+gate),结合词特征后具有更加丰富的文本信息,改善了实体边界划分不准的现象, F_1 值提升 1.51%,召回率提升了 2.04%;相比于未引入带有残差连接的门控机制模型(baseline+seg), F_1 值提升 0.93%。以上结果显示本文所提出模型架构的有效性。

表 6 Resume 数据集命名实体识别对比实验结果(%)

模型	P	R	F_1
baseline	93.72	93.70	93.71
baseline+gate	94.41	94.27	94.34

续表

模型	P	R	F_1
baseline+seg	94.36	94.26	94.31
dgcn	95.05	94.91	94.98
dgcn-residual	94.97	94.71	94.84

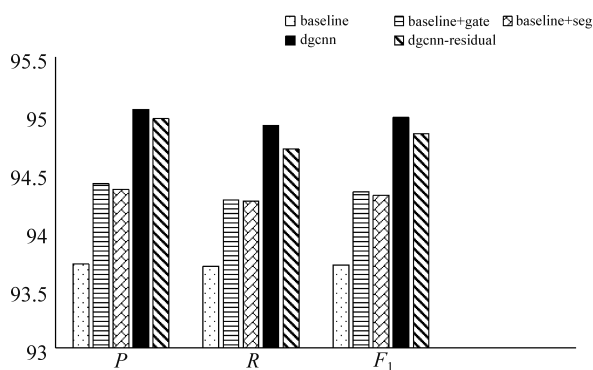


图5 Resume 数据集命名实体识别对比实验结果

从表6和图5(图5为表6的可视化结果)中可看出,门控机制与词特征是模型性能提升的主要因素,相比于基于普通空洞卷积的模型架构(baseline),引入词特征的模型(baseline+seg)与引入残差连接的门控机制模型(baseline+gate),在Resume数据集上,分别提升 F_1 值0.6%与0.63%;将词特征与带有残差连接的门控机制相结合模型(dgcn)学习到了更多的上下文特征,与baseline相比 F_1 值提升1.27%,召回率 R 与精确率 P 均得到明显的提升。其中在门控特征部分,本文在Dauphin^[9]所提出的方案的基础上加入了残差连接。为了验证残差连接对模型性能的影响,本文还加入了对比实验(dgcn-residual),与之相比引入残差连接的模型(dgcn)不仅改善了梯度消失的现象,同时也使得信息可以在多通道内流动, F_1 值提升0.14%。此结果表明该改进可以进一步提升命名实体识别效果。

3.4.4 MSRA 数据集模型训练时间对比及分析

由于Resume数据集规模较小,各模型训练时间未表现出明显差异,故本文选择在规模较大的MSRA数据集上进行训练时长对比,结果示于表7。采用的对比模型基准为Bi-LSTM-CRF架构,并将普通卷积与空洞卷积速度进行对比,由表7可看出,当本文所提出模型“带有门控机制空洞卷积”单元层数设置为4层时,其在训练阶段速度为基于Bi-

LSTM-CRF模型的5倍,并且在拥有与普通卷积相同感受野时,其性能优于普通卷积架构;当感受野范围为15时(空洞卷积堆叠层数为3,普通卷积堆叠层数为7),模型训练速度约提升了3倍,并且 F_1 值也同时提升了1.4%;当感受野范围为31时(空洞卷积堆叠层数为4,普通卷积堆叠层数为15),模型训练速度提升了约5倍, F_1 值也同时提升了0.09%。以上结果表明了本文所提出模型在速度和性能上的优越性。

表7 MSRA 数据集上时间对比

模型	模型层数	时间/h	F_1 值
Bi-LSTM	2	20	88.36
Bi-LSTM	4	60	90.18
CNN+gate	7	12	89.64
CNN+gate	15	21	91.23
dgcn	3	4	91.04
dgcn	4	4	91.32
dgcn	7	5	92.97

3.4.5 错误案例分析

本文进一步对所提出模型在MSRA数据集上各实体类别的识别效果进行分析,如表8所示。

表8 MSRA 数据集各个类别实体识别效果(%)

模型(dgcn)	P	R	F_1
LOC(地名)	94.64	91.55	93.07
ORG(机构名)	92.17	90.02	91.08
PER(人名)	94.19	94.04	94.12

由表8可以看出,PER(人名)的识别效果准确率与召回率较为一致,而LOC(地名)和ORG(机构名)的实体识别效果准确率高出召回率。

出于研究模型在MSRA数据集中,实体识别结果准确率高出召回率的原因,本文选取了测试集中50条错误数据进行了分析,从表9中列举出的错误案例可以看出,产生误差的主要原因在于机构名同地名容易混淆,由实体边界识别不清所致。例如“特立尼达和多巴哥-中国友好协会”属于机构名,但模型识别结果将“特立尼达”与“多巴哥”判断为地名,“中国友好协会”判断为“机构名”,因此导致准确率较高,召回率偏低。

表 9 典型错误案例

文本	dgcnn 预测实体	Zhang 等 ^[17] 预测实体	真实实体
企业界人士有：比利时通用银行总行长费尔迪南·沙德尔。	比利时通用银行(ORG) 费尔迪南·沙德尔(PER)	比利时(LOC) 通用银行总行(ORG) 费尔迪南·沙德尔(PER)	比利时通用银行总行(ORG) 费尔迪南·沙德尔(PER)
在香港特区设立终审法院是落实“一国两制”、“港人治港”、高度自治方案政策的重要体现	香港特区设立终审法院(ORG) 港(LOC) 港(LOC)	香港特区(LOC) 终审法院(ORG) 港(LOC) 港(LOC)	香港特区(LOC) 终审法院(ORG) 港(LOC) 港(LOC)
特立尼达和多巴哥—中国友好协会会长克劳德·阿兴	特立尼达(LOC) 多巴哥(LOC) 中国友好协会(ORG) 克劳德·阿兴(PER)	特立尼达和多巴哥—中国友好协会(ORG) 克劳德·阿兴(PER)	特立尼达和多巴哥—中国友好协会(ORG) 克劳德·阿兴(PER)

Zhang 等^[17] 由于使用了大型外部词典,大大减少了实体边界识别错误的可能性,故表中第三条文本中实体“特立尼达和多巴哥-中国友好协会”可正确识别为机构;对表中第二条文本数据识别出谓词“设立”,该谓词将实体“香港特区”与“终审法院”分隔开,正确划分出两实体边界,从而降低了这种易混淆的噪声数据影响;然而 dgcnn 于本例中无法正确判断出谓词“设立”,导致实体边界划分错误,将“香港特区设立终审法院”错误地判断为一个实体。故推测通过外部词典匹配,进而大幅度提高实体边界识别准确率,是 Zhang 等^[17] 所提出模型效果在 MSRA 数据集上优于本文所提出模型效果的主要原因。

4 结论

本文提出了一个基于空洞卷积的模型架构,并在此基础上加入了残差门控机制,既保证了信息能够在多通道流通,同时也缓解了梯度消失现象。基于位置的词特征与字向量融合,改善了字向量表示信息有限的问题,丰富了上下文的特征表示。实验结果表明,本文所提出的模型在速度和性能上具有明显优势。

在未来的工作中,我们将继续研究如何在降低模型计算量的同时保持模型的性能,并将该架构拓展到自然语言处理的其他任务上。

参考文献

- [1] Eddy S R. Hidden Markov models[J]. Current Opinion in Structural Biology, 1996, 6(3): 361-365.
- [2] Tong S, Koller D. Support vector machine active learning with applications to text classification [J].

- Journal of Machine Learning Research, 2002, 2(1): 999-1006.
- [3] Luo G, Huang X, Lin C Y, et al. Joint entity disambiguation[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2014: 879-888.
- [4] Wang Ling, Tiago Luís, Luís Marujo, et al. Finding function in form: Compositional character models for open vocabulary word representation[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2015: 324-331.
- [5] Xuezhe Ma, Yingkai Gao, Zhiting Hu, et al. Dropout with expectation-linear regularization [C]//Proceedings of the International Conference on Learning Representations, 2017: 467-475.
- [6] Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, et al. Neural architectures for named entity recognition[C]//Proceedings of the North American Chapter of the Association for Computational Linguistics, 2016: 387-396.
- [7] Fisher Yu, Vladlen Koltun. Multi-scale context aggregation by dilated convolutions[C]//Proceedings of the International Conference on Learning Representations, 2016: 674-685.
- [8] Emma Strubell, Patrick Verga, David Belanger, et al. Fast and accurate entity recognition with iterated dilated convolutions[C]//Proceedings of the Association for Computational Linguistics, 2017: 465-478.
- [9] Dauphin Y N Fan, et al. Language modeling with gated convolutional networks [C]//Proceedings of the Association for Computational Linguistics, 2016.
- [10] Chunqi Wang, Wei Chen, et al. Named entity recognition with gated convolutional neural networks[C]//Proceedings of the Conference on Computational Linguistics, 2017: 134-145.
- [11] Kaiming He, Xiangyu Zhang, Shaoqin Ren, et al. Deep residual learning for image recognition [C]//

- Proceedings of the Association for Computational Linguistics, 2015:376-389.
- [12] Xinxiong Chen, Lei Xu, Zhiyuan Liu, et al. Joint learning of character and word embeddings[C]//Proceedings of the International Joint Conference on Artificial Intelligence, 2015: 207-216.
- [13] Jingzhou He, Houfeng Wang. Chinese named entity recognition and word segmentation based on character [C]//Proceedings of the 6th SIGHAN Workshop on Chinese Language Processing, 2008.
- [14] Xinnian Mao, Yuan Dong, Saike He, et al. Chinese word segmentation and named entity recognition based on conditional random fields[C]//Proceedings of the 6th SIGHAN Workshop on Chinese Language Processing, 2008.
- [15] Zhiheng Huang, Wei Xu, Kai Yu. Bidirectional LSTM-CRF models for sequence tagging[J]. arXiv preprint arXiv: 1508.01991, 2015.
- [16] Nanyun Peng, Mark Dredze. Named entity recognition for Chinese social media with jointly trained embeddings[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2015:548-554.
- [17] Y Zhang, J Yang. Chinese NER using lattice LSTM [C]//Proceedings of the Association for Computational Linguistics, 2018:671-682.
- [18] Pengfei Cao, Yubo Chen, Kang Liu, et al. Adversarial transfer learning for Chinese named entity recognition with self-attention mechanism[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2018:1461-1473.
- [19] Zhixing Tan, Mingxuan Wang, Jun Xie, et al. Deep semantic role labeling with self-attention[C]//Proceedings of the Association for the Advance of Artificial Intelligence, 2018:543-554.
- [20] Matthew Peters, Mark Neumann, Mohit Iyyer, et al. Deep contextualized word representations[C]//Proceedings of the Association for Computational Linguistics, 2018: 2227-2237.
- [21] Jacob Devlin, Mingwei Chang, Kenton Lee, et al. Bert: Pre-training of deep bidirectional transformers for language understanding [C]//Proceedings of North American Chapter of the Association for Computational Linguistics, 2019:1048-1056.
- [22] Jie Yang, Shuailong Liang, Yue Zhang. Design challenges and misconceptions in neural sequence labeling [C]//Proceedings of CoNLL, 2018: 3879-3889.
- [23] Gina-Anne Levow. The third international Chinese language processing bakeoff: Word segmentation and named entity recognition[C]//Proceedings of the 5th SIGHAN Workshop on Chinese Language Processing, 2006: 548-554.
- [24] Yoav Goldberg, Omer Levy. Word2Vec explained: Deriving Mikolov et al.'s negative sampling word-embedding method[J]. arXiv preprint arXiv:1402.3722, 2014.
- [25] Aitao Chen, Fuchun Peng, Roy Shan, et al. Chinese named entity recognition with conditional probabilistic models[C]//Proceedings of the 5th SIGHAN Workshop on Chinese Language Processing, 2006:324-331.
- [26] Suxiang Zhang, Juan Wen, Xiaojie Wang. Word segmentation and named entity recognition for SIGHAN bakeoff3 [C]//Proceedings of the 5th SIGHAN Workshop on Chinese Language Processing, 2006: 158-161.
- [27] Junsheng Zhou, Weiguang Qu, Fen Zhang. Chinese named entity recognition via joint identification and categorization[J]//Chinese Journal of Electronics, 2013,22(2): 225-230.
- [28] Chuanhai Dong, Jiajun Zhang, Chengqing Zong, et al. Character based LSTM-CRF with radical-level features for Chinese named entity recognition[C]//Natural Language Understanding and Intelligent Applications, 2016: 239-250.
- [29] Fan Yang, Jianhu Zhang, Gongshen Liu, et al. Five-stroke based CNN-BiRNN-CRF network for Chinese named entity recognition[C]// Proceedings of the International Conference on Natural Language Processing and Chinese Computing, 2018:532-541.



王笑月(1998—),硕士研究生,主要研究领域为自然语言处理。

E-mail: wangxy0808@163.com



段菲(1979—),博士,讲师,主要研究领域为机器学习、计算机视觉、多模态人机交互。

E-mail: duanf07@gmail.com



李茹(1963—),通信作者,博士,教授,主要研究领域为自然语言处理。

E-mail: liru@sxu.edu.cn