# CCML 2021 第十八届中国机器学习会议 The 18th China Conference on Machine Learning

## 2021年8月6-8日 中国·长沙 AUG 6-8, 2021 CHANGSHA, CHINA

# 基于空洞卷积块架构的命名实体识别模型

袁钺<sup>1</sup>, 王艳丽<sup>2</sup>, 刘勘<sup>2</sup>

#### (1.北京大学信息管理系,北京 100871; 2.中南财经政法大学信息与安全工程学院,湖北 武汉 430073)

(yuanyue@stu.pku.edu.cn wangyanli@stu.zuel.edu.cn liukan@zuel.edu.cn)



#### ! " # \$ % & ' () \* + ), - . / 0 1 2 3 4 5 6 7 8 9 : ;

传统卷积神经网络往往会叠加多个卷积层和池化层来扩大网络的感受野来获取上下 文文本的高阶特征。但是池化层的添加往往会导致空间信息的丢失。



#### < = > 9 : ? @ ! " # \$ % & ' ( < A B C D E F G H I 8 J K L M N 0 ;</pre> 传统卷积神经网络从文本中获取包含更多的上下文信息的高阶文本特征变得困难,进 而限制了其在命名实体识别任务中的准确率。;

#### **OP#\$%&'(Q&R#\$%&'(STUVW-+),0;**

与标准的卷积相比,空洞卷积的卷积核覆盖了比标准卷积的卷积核更大的区域,从而 覆盖更大的范围的输入数据。给定一个一维输入X,假设空洞卷积具有大小为S的卷 积核k,那么一维输入X经过空洞卷积后的输出Y[i]可以定义为:

 $Y[i] = \sum_{s=1}^{S} [i + r * s - 1]k[s], i \in [0, L)$ 

式中,r为空洞卷积核的空洞率( $r \ge 1, r \in N^*$ )。相比于标准卷积,空洞卷积网络的感 受野能更有效率地扩增,即选择更大的卷积核尺寸S或增加空洞率r,如下图1所示。



图 1 增加卷积核尺寸 S 或空洞率感 r 受野变化示例

此外,扩大空洞卷积网络感受野还可以通过叠加空洞卷积层并随着层数的增加按照指 数的方式(eg, 2<sup>°</sup>)增加空洞率r来实现。图 2 给出了一个通过叠加空洞卷积层改变感 受野大小的例子。



图 4 (a)基于空洞卷积块架构的命名实体识别模型结构及(b)细节示例

#### $Z [ 0 0 # $ ^ 8 A B C D E F \setminus ] (DCL+Bi-LSTM)$

为了验证空洞卷积块架构的性能,本文去除了空洞卷积块架构中的残差卷积层、批归 一化层、激活层,并以此构建了相应的销蚀模型,即仅基于空洞卷积层来设计命名实 体识别模型(Dilated Convolutional Layer Based Named Entity Recognition)。图 5(a)展 示了该销蚀模型的结构,包含空洞卷积层、Bi-LSTM 层和 CRF 层。如图 8(b)中展示 的是该销蚀模型的细节示例。



图 2 堆叠卷积层时标准卷积和空洞卷积感受野变化示例



0 P # \$ X Y Z (Dilated Convolutional Block Architecture, DCBA)



#### 图 3 空洞卷积块架构与列式空洞卷积

为了保证空洞卷积网络的稳定性,本文设计了空洞卷积块架构。 图 3 所示的空洞卷积 **块架构由三个部分组成:空洞卷积块、残差卷积层、串联层** 

- 残差卷积层和串联层可以融合原始输入数据本身的特征,提高堆叠空洞卷积 块后架构的整体稳定性,从而增加空洞卷积块架构的鲁棒性。
- 空洞卷积块中批归一化层和激活层来稳定空洞卷积层在堆叠时的性能,空洞 卷积层对输入数据进行列式空洞卷积,



## a 人民日报 1998 中文标注语料库。

#### b c A B C D E F C d ;

		Location			Person			Or		Total				
		Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1	
without CRF	Bi-LSTM+Softmax	0.7502	0.715	0.7322	0.7645	0.7707	0.7676	0.5433	0.6221	0.58	0.7044	0.7129	0.7086	
	IDCNN+Softmax	0.7699	0.7709	0.7704	0.7978	0.7838	0.7907	0.6318	0.6769	0.6536	0.7471	0.7548	0.751	
	DCL+Bi- LSTM+Softmax	0.7929	0.7584	0.7753	0.8169	0.8165	0.8167	0.6338	0.6529	0.6432	0.765	0.7544	0.7596	
	DCBA+Bi- LSTM+Softmax	0.7976	0.8248	0.811	0.823	0.7969	0.8097	0.6811	0.7333	0.7062	0.779	0.7962	0.7875	
with CRF	Bi-LSTM+CRF	0.8699	0.8085	0.838	0.846	0.817	0.8313	0.7556	0.7318	0.7435	0.837	0.7947	0.8153	
	IDCNN+CRF	0.9009	0.8248	0.8612	0.8497	0.815	0.832	0.7749	0.7708	0.7729	0.8558	0.8101	0.8323	
	DCL+Bi-LSTM+CRF	0.9061	0.8617	0.8833	0.8974	0.8377	0.8665	0.8354	0.7776	0.8054	0.8883	0.8359	0.8613	
	DCBA+Bi-LSTM+CRF	0.9187	0.8794	0.8986	0.8891	0.8241	0.8553	0.8353	0.8152	0.8251	0.891	0.8479	0.8689	
(a)				(b)	COM				(e)			(f)		
				(d)	- Shiterates in a state at the intervales discussion			( <i>g</i> )				(h)		
				177 KAN	Hatrafra Herrelea Herrelea Herrelea								Alternal Barrel Barrel Charteral	

## DCBA+Bi-LSTM+CRF 模型和 DCL+Bi-LSTM+CRF 模型的总体 Precision、

提出了列式空洞卷积(Column-wise Dilated Convolution)。与一般空洞卷积不 同,对于二维文本嵌入,列式空洞卷积能输出一个二维的特征映射以供后续空 洞卷积层堆叠处理。

图 3 中 $V_{\mu} \in \mathbb{R}^{5 \times 8}$  是一个包含G个词嵌入文本的二维(2-D)嵌入(输入矩阵), 卷积核k的 大小为 $W \times l$ , W为卷积核宽度, l为二维文本嵌入 $V_{\pm}$ 的宽度(亦即词嵌入的维度)。可 以看作是l个宽为W,长度为1的空洞卷积核 $k' \in \mathbb{R}^{((s)\times)}$ 与l个长度为G的一维(1-D)输 入嵌入之间一维卷积的组合。与标准卷积不同,这里的卷积按列操作,每个输出元素。 的计算公式如下:

 $f[x][y] = \langle V_i[x: x + W_{(s,r)} - 1, y], k[:, y] \rangle_F$ 

下标 F 表示 Frobenius 内积 (即空洞卷积核 k和二维文本嵌入  $V_{\pm}$ 对应元素的乘积之和, 输出的二维矩阵中每一个元素的f[x][y]是由部分输入的二维文本嵌入与部分空洞卷 积核进行内积运算而得到的。批归一化层对每个输入的数据特征进行规范化和线性变 化。批归一化层的输出将输入到激活层,激活层使用线性整流函数(ReLU)。

### $[ \ 0 \ P \ # \ X \ Y \ Z \ 8 \ A \ B \ C \ D \ E \ F \ ] \ \hat{OCBA+Bi-LSTM}$

图 4(a)展示了基于空洞卷积块架构的命名实体识别模型的结构 ,包含空洞卷积块架构、 Bi-LSTM 层和 CRF 层。空洞卷积块架构的输入是由一维字符嵌入拼接而成的二维矩 阵。空洞卷积块架构从输入的二维矩阵中提取包含更多上下文信息的高阶文本特征并 与各个字符嵌入拼接,然后输出到 Bi-LSTM+CRF 模型中进行命名实体识别。图 4(b) 展示了基于空洞卷积块架构的命名实体识别模型的细节示例。





- Recall 和 F1值优于其他所有模型。其中, DCBA+Bi-LSTM+CRF 模型最高。
- DCBA+Bi-LSTM+CRF 比 DCL+Bi-LSTM+CRF 在命名实体识别任务中收敛 更快且性能更稳定。

#### + ) . C d



- 空洞卷积块架构的感受野与命名实体识别的准确度并不是正相关。空洞率过 大时,卷积核非零值相距较远,模型丢失了过多输入文本的邻近特征,造成模 型识别能力降低。
- 感受野增大并不一定带来识别效果的提升,必须适当调整空洞率以获得大小 较为合适的感受野,降低网格效应干扰以得到更好识别效果。

本文相关代码请参见:

https://github.com/Acstream/Dilated-Convolutional-**Block-Architecture-Based-Named-Entity-Recognition** 



# 投稿人 姓名:袁钺1 王艳丽2 刘勘2 单位: 1.北京大学信息管理系 2.中南财经政法大学信息与安全工程学院