

基于密集连接卷积神经网络的远程监督关系抽取

钱小梅¹ 刘嘉勇¹ 程芃森^{1,2}

1 四川大学网络空间安全学院 成都 610000

2 中国科学院信息工程研究所中国科学院网络测评技术重点实验室 北京 100093

(qxiaomay@126.com)

摘要 密集连接卷积神经网络(DenseNet)是一种新型深度卷积神经网络架构,通过建立不同层间的连接关系,来确保网络层与层间最大程度的信息传输。在文本远程监督关系抽取任务中,针对现有神经网络方法使用浅层网络提取特征的局限,设计了一种基于密集连接方式的深度卷积神经网络模型。该模型采用五层卷积神经网络构成的密集连接模块和最大池化层作为句子编码器,通过合并不同层次的词法、句法和语义特征,来帮助网络学习特征,从而获取输入语句更丰富的语义信息,同时减轻深度神经网络的梯度消失现象,使得网络对自然语言的表征能力更强。模型在 NYT-Freebase 数据集上的平均准确率达到了 82.5%,PR 曲线面积达到了 0.43。实验结果表明,该模型能够有效利用特征,并提高远程监督关系抽取的准确率。

关键词: 深度学习;关系抽取;远程监督;卷积神经网络;密集连接

中图法分类号 TP391

Distant Supervised Relation Extraction Based on Densely Connected Convolutional Networks

QIAN Xiao-mei¹, LIU Jia-yong¹ and CHENG Peng-sen^{1,2}

1 School of Cybersecurity, Sichuan University, Chengdu 610000, China

2 Key Laboratory of Network Assessment Technology, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China

Abstract Densely connected convolutional networks (DenseNet) is a new architecture of deep convolutional neural network. By using identity mapping for shortcut connections between different layers, it can ensure the maximum information transmission of neural network. In the distant supervised relation extraction task, precious models use shallow convolution neural networks to extract features of a sentence which can only represent partial semantic information. To enhance the representation power of network, a deep convolutional neural network model based on dense connectivity was designed to encode sentences. The proposed model consists of five layers of densely connected convolutional neural networks. It can capture more semantic information by combining different levels of lexical, syntactic, and semantic features. At the same time, it can alleviate the phenomenon of gradient disappearance of deep neural network, which makes the network more capable of characterizing natural language. The experimental results on NYT-Freebase datasets show that the mean accuracy of the proposed model achieves 82.5%, and the PR curve area achieves 0.43. Experimental results show that the proposed model can effectively utilize features and improve the accuracy of distant supervised relation extraction.

Keywords Deep learning, Relation extraction, Distant supervision, Convolutional neural network, Dense connectivity

1 引言

关系抽取是文本信息抽取的任务之一,其主要研究内容是从句子中提取实体间的语义关系,即给定一段包含实体对 e_1 和 e_2 的文本,关系抽取的目的是识别 e_1 和 e_2 的关系类别 r 。关系抽取是构建知识图谱的关键组成部分,对于自然语言处理应用(如语义搜索、智能问答和自动摘要)至关重要^[1]。

经典的关系抽取算法(如基于特征的方法和基于核函数的方法)^[2-3]和近年来广泛使用的神经网络模型(如卷积神经网络(Convolutional Neural Networks, CNN)、循环神经网络(Recurrent Neural Network, RNN)和长短期记忆神经网络(Long Short-Term Memory, LSTM)^[4-6]) 在训练过程中均需要大量标记数据,而数据的标记非常昂贵且费时,因此, Mintz 等提出了远程监督^[7],将已有知识库对应到丰富的非结构化

到稿日期:2019-01-23 返修日期:2019-09-12 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:中国科学院网络测评技术重点实验室开放课题基金(NST-18-001)

This work was supported by Open Research Fund of the Key Laboratory of Network Assessment Technology of Chinese Academy of Sciences (NST-18-001).

通信作者:程芃森(cps11@163.com)

数据中,其一般过程如图 1 所示。

图 1 中,远程关系抽取首先从知识库(KB)中抽取存在关系的实体对。然后从非结构化文本中抽取含有实体对的句子作为训练样例,并假设该样例句子中一定存在实体描述关系,例如“Bill Gates”和“Microsoft”在 KB 中的关系是“Founder”,那么假设类似于“Bill Gates, founder of Microsoft.”的句子都存在“founder”关系。如此,利用 KB 对海量文本数据进行自动标注,得到标注好的数据,随后训练一个分类器,每一个类别表示一种关系,由此实现关系抽取。该方法能带来大量训练数据,但同时也引入了噪声问题,由于假设过于肯定,容易引入错误标签,并且在数据标记过程依赖于自然语言处理(Natural Language Processing, NLP)工具, NLP 工具(如命名实体识别)的误差会在整个数据构造过程中传播和积累,从而影响后续关系抽取的精度。

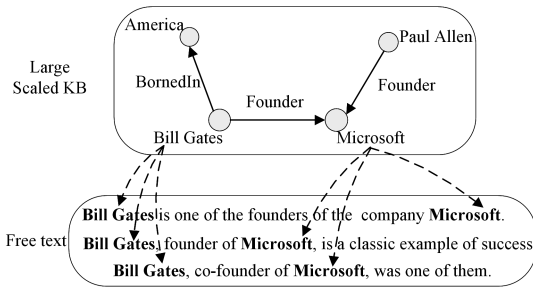


图 1 远程监督关系抽取

Fig. 1 Distant supervision relationship extraction

在文本远程监督关系抽取任务中,现有研究倾向利用深度学习替代特征工程,用词嵌入表示句子中的实体和词语,再采用 CNN、RNN 或相关变体对句子进行建模,最后进行关系分类^[8-10]。如 Zeng 等基于 CNN 提出了一种分段卷积神经网络(Piecewise Convolutional Neural Network, PCNN)模型^[8],在 CNN 池化层通过两个实体位置将特征分为 3 段进行池化,以便于更好地捕获两个实体间的结构化信息;Lin 等在 PCNN 的基础上加入了注意力机制(Attention, ATT),对每个词带实例进行权重学习^[9],并赋予标签正确的实例高权重,反之则赋予低权重,从而提高了分类准确率;Zhou 等对 RNN 做了改进,采用双向循环神经网络(Bi-directional Long Short-Term Memory, BLSTM)作为句子编码器,并使用词级别注意力机制进行权重学习^[10]。以上研究工作表明,深度学习方法能够解决特征表示问题,但以上神经网络方法均为较浅层模型,通常只涉及一个神经网络层和最大池化层作为句子编码器,而浅层网络学习到的特征只能表征自然语言的部分语义信息,在远程监督关系分类时更容易受到噪声数据的影响。

随着复杂神经网络的发展,深层神经网络模型逐渐被应用于关系抽取领域,如文献[11]提出了一种 9 层的深度残差网络用于文本远程监督关系抽取任务,而如此深的网络并非每一层都有效,因此相较于浅层网络性能的提升并不大。为了探究深层神经网络在文本远程监督关系抽取任务上的表现,本文提出了一种基于密集连接卷积神经网络的关系抽取模型,简称 DenseCNN 模型。密集连接卷积神经网络的核心思想是在网络不同层间建立连接关系,使模型能更有效地利用网络中所有层次的特征,其在 ImageNet 图像分类任务中取得了优异表现^[12]。DenseCNN 模型采用 5 层卷积神经网络

构成的密集连接模块和最大池化层作为句子编码器,并通过不同层次特征的合并和复用,使得网络能够在噪声数据中学习较好的特征,从而提升关系抽取性能。

2 关系抽取模型

本文提出的 DenseCNN 关系抽取模型如图 2 所示,该模型由文本向量化层、密集连接卷积层、池化层和分类层构成。该模型首先在文本向量化层采用词向量模型 word2vec 将文本表示成向量形式^[13],与位置向量(各词与两个实体的相对位置)进行拼接作为输入;其次通过密集连接卷积层得到文本的特征;然后通过最大池化层对特征进行下采样,保留显著特征;最后通过全连接的 softmax 分类层进行关系分类^[14]。

密集连接卷积层是该模型与普通卷积神经网络模型的不同之处,这里 5 层卷积网络构成的深度学习模型可以将浅层网络学到的词法或句法特征抽象为更为丰富的语义特征,同时密集连接模块通过合并不同层次的特征,加强了特征的复用和传递,减轻了神经网络的梯度消失现象,使得网络对自然语言的表征能力更强。

句子向量化如图 2 中“Vector representation”部分所示,对于一个长度为 n 的句子 s ,可以将其表示为一个向量列表 $X(s) = (x_1, x_2, \dots, x_n)$,其中每一个向量都包含两个部分:词向量和位置向量。

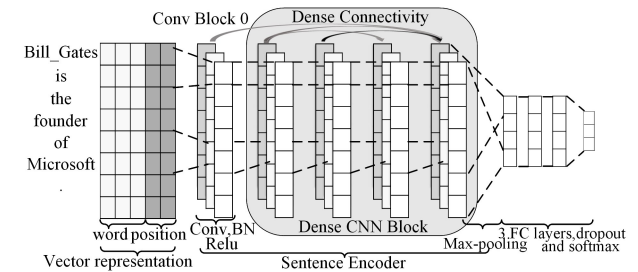


图 2 DenseCNN 模型结构

Fig. 2 Structure of DenseCNN

2.1 向量化层

(1) 词向量

为了捕捉句子的词法和语义信息,需要将句子 $s = \{\omega_1, \omega_2, \dots, \omega_n\}$ 中的每一个词 ω_i 映射为低维实值向量: $x_i^w = W^{\text{word}} \omega_i^v$ 。其中, $W^{\text{word}} \in R^{d_w \times |V|}$ 是用 word2vec 训练得到的向量矩阵, d_w 表示词向量的维度, $|V|$ 表示词典的大小, v^i 是词的独立编码表示。

(2) 位置向量

在关系抽取任务中,靠近实体对的单词对抽取实体对关系有显著影响,将每个单词到两个实体的相对距离映射为低维位置向量并拼接到该单词的词向量后,能更加准确地描述句子含义^[8]。在句子 $s = \{\omega_1, \omega_2, \dots, \omega_n\}$ 中,将词 ω_i 到实体 e_1 和 e_2 的相对距离映射到低维向量空间得到位置向量 $x_{p_1}^i$ 和 $x_{p_2}^i$,其中 $x_p \in R^{d_p \times |P|}$, d_p 表示位置向量维度, $|P|$ 表示距离集大小。

因此,在句子 $s = \{\omega_1, \omega_2, \dots, \omega_n\}$ 中,词 ω_i 的向量化表示为 $x_i = \{x_i^w, x_{p_1}^i, x_{p_2}^i\}$,其中 $x_i \in R^d (d = d_w + d_p \times 2)$ 。

2.2 密集连接卷积层

2.2.1 卷积

模型采用的卷积模块如图2中“Conv Block 0”部分所示。假定句子 s 中第 i 个词到第 $i+j$ 个词通过2.1节向量化后得到的词向量 $\mathbf{x}_i, \mathbf{x}_{i+1}, \mathbf{x}_{i+2}, \dots, \mathbf{x}_{i+j}$ 构成的向量矩阵为 $\mathbf{x}_{i:i+j}$,卷积窗口大小为 h ,词向量维度为 d ,利用一个过滤器 $\mathbf{w}=\mathbf{R}^{h \times d}$ 就能够捕捉连续 h 个词的特征。因此,第 i 个连续 h 个词的向量矩阵 $\mathbf{x}_{i:i+h-1}$ 经过卷积得到的特征 c_{x_i} 如式(1)所示:

$$c_{x_i} = f(\mathbf{w} \cdot \mathbf{x}_{i:i+h-1} + b_0) \quad (1)$$

其中, $b \in \mathbf{R}$ 表示偏置项; f 表示一个非线性函数,包括归一化处理函数BN(Batch Normalization)^[15]和ReLU激活函数^[16]。那么,给定长度为 n 的句子 $s = \{w_1, w_2, \dots, w_n\}$ 的词向量矩阵包括 $[\mathbf{x}_{1:h}, \mathbf{x}_{2:h+1}, \dots, \mathbf{x}_{n-h+1:n}]$,卷积后该句的特征向量表示为 $\tilde{\mathbf{c}} = [\tilde{c}_{x_1}, \tilde{c}_{x_2}, \dots, \tilde{c}_{x_{n-h+1}}]$,其中 $\tilde{\mathbf{c}} \in \mathbf{R}^s$ ($s = n - h + 1$)。

2.2.2 密集连接网络

在标准CNN网络中,假设 c_l^s 是第 l 层的输出,那么 c_l^s 表示前一层输出 c_{l-1}^s 通过一个非线性变换 H_l^s 得到的结果,如式(2)所示:

$$c_l^s = H_l^s(c_{l-1}^s) \quad (2)$$

其中, H 包括了卷积、ReLU线性激活函数和dropout约束函数^[17]。

为了简化深层网络训练,文献[18]提出了残差网络ResNet,引入了一个残差模块,将多个卷积层级联的输出与输入相加,第 l 层的输出结果 c_l^R 为:

$$c_l^R = H_l^R(c_{l-1}^R) + c_{l-1}^R \quad (3)$$

其中, H^R 定义为由多个卷积、ReLU激活函数和归一化处理函数BN级联的模块。残差模块的引入使得特征能够被重用,解决了深层网络梯度消失问题。

文献[12]在残差网络的基础上进行进一步的扩展,设计了一种更复杂的连接方式,即密集连接网络DenseNet,其特点对于网络的任意一层,该层前面所有层的特征都是该层的输入,该层的特征是后面所有层的输入。因此,第 l 层的输出为:

$$c_l = H_l([c_{l-1}, c_{l-2}, \dots, c_0]) \quad (4)$$

其中, $[\cdot]$ 表示将每层的输出特征进行通道合并,是一个非线性函数,包括卷积、ReLU激活函数和归一化处理函数BN。这种密集连接方式使得特征得到了最大程度的复用,特征和梯度的传递更有效,网络更容易训练。

在文本远程监督关系抽取任务中,为了更好地利用句子语义信息捕捉更为丰富的语义特征,本文利用密集连接网络来帮助网络进行特征学习,其网络示意图如图2中“Dense CNN Block”部分所示,该部分由密集连接的5层卷积神经网络构成,具体的网络结构如图3所示。

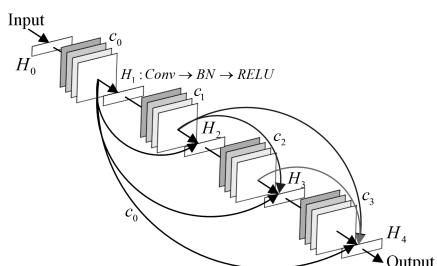


图3 密集连接模块结构图

Fig. 3 Structure of dense CNN block

图3中,密集连接模块的输入是句子 s 的向量化表示 $\mathbf{X}(s) = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$,通过式(1)卷积得到第一层的特征表示为 c_0 ,其中:

$$c_{0i} = f(\mathbf{w}_0 \cdot \mathbf{x}_{i:i+h-1} + b_0) \quad (5)$$

接着通过密集连接模块的特征映射,将 c_0 作为后面所有层的输入,如图3中左下方黑色曲线箭头所示。则第二层卷积输入为 c_0 ,输出特征 c_1 为:

$$c_1 = H_1([c_0]) \quad (6)$$

$$c_{1i} = f(\mathbf{w}_1 \cdot c_{0(i+h-1)} + b_1) \quad (7)$$

最终,网络的输出可以表示为:

$$\text{output} = H_4([c_0, c_1, c_2, c_3]) \quad (8)$$

$$\text{output}_i = f(\mathbf{w}_4 \cdot [c_0, c_1, c_2, c_3]_{i+h-1} + b_4) \quad (9)$$

其中, output 为句子 s 的特征编码表示 c_{out} 。在句子特征表示中,如果浅层到深层网络的特征分别表示词法、句法和语义特征,那么密集连接网络通过特征映射将词法、句法和语义特征结合起来有利于网络进行特征学习,从而能够更准确地表达句子含义。

2.3 池化层

在密集连接卷积神经网络提取特征之后,对输出特征进行池化,如图2中“Max-pooling”部分所示。池化的目的是在保留主要特征的同时减少参数和计算量,常用的池化方法有平均池化和最大池化两种,本文使用最大池化对句子特征 c_{out} 进行降维:

$$q = \max\{c_{\text{out}}\} \quad (10)$$

假定过滤器数量为 m ,那么通过最大池化输出的特征向量为 $\mathbf{z} = [q_1, q_2, \dots, q_m]$ 。

2.4 分类层

经过池化后的向量 \mathbf{z} 作为句子的特征输入到一个全连接的softmax分类器,输出关系上的概率分布为:

$$\mathbf{O} = \mathbf{W}^{\text{rel}} \mathbf{z} \quad (11)$$

其中, $\mathbf{W}^{\text{rel}} \in \mathbf{R}^{r \times m}$ 是转移矩阵; r 为关系类别的数目;输出 \mathbf{O} 是一个 r 维的向量,第 i 维表示第 i 类关系的概率。

2.5 模型训练和优化

本文将交叉熵代价函数^[19]作为目标函数,其定义如下:

$$J(\theta) = \sum_{i=1}^T \log p(\mathbf{O}^i | \mathbf{X}^i, \theta) \quad (12)$$

其中, θ 表示模型中可训练的参数; T 为训练样本的数量; $p(\mathbf{O}^i | \mathbf{X}^i, \theta)$ 是极大似然估计函数,定义为给定参数和句子 \mathbf{X} 的情况下关系类别 i 的概率,如式(13)所示:

$$p(\mathbf{O}^i | \mathbf{X}^i, \theta) = \frac{e^{o_i}}{\sum_{k=1}^r e^{o_k}} \quad (13)$$

利用梯度下降法求解模型的参数,在求导过程中使用反向传播算法。

$$\theta = \theta + \eta \frac{\partial}{\partial \theta} \log J(\theta) \quad (14)$$

其中, η 表示学习率。

3 实验及分析

3.1 实验环境

本文的实验环境如下:操作系统为ubuntu16.04,64位;

内存大小为 8 GB; 处理器为 Intel(R) Xeon(R); 编程平台为 Pycharm, python3.6 版本。实验程序基于开源软件库 Tensorflow 1.10 完成, 在单个 NVIDIA Corporation GPU 上执行。

3.2 数据集和测评标准

为了评估密集连接卷积神经网络关系抽取模型, 本文采用远程监督关系抽取任务中广泛使用的 NYT-Freebase 数据集^[20], 该数据集于 2010 年由 Riedel 等提出, 通过将 Freebase 知识库中的实体与纽约时报(NYT)文本库对齐而生成。该数据集样例如表 1 所列。

表 1 NYT-Freebase 数据集样例

Table 1 Example of NYT-Freebase dataset

Entity Aid/Bid	Entity A/B	Relation	Sentence
m.0kfbz/ m.068963	akihito/ prince_akishino	/children	the princess is the wife of prince_akishino, the second son of emperor akihito.
m.0j95/ m.0nlh7	alberta/ edmonton	/capital	the land is near calgary, while that is one of alberta's largest cities, the capital is edmonton.

该数据集分为训练数据和测试数据, 其中训练数据有 522611 个句子、281270 个实体对和 18252 个关系事实, 测试数据包括 172448 个句子、96678 个实体对和 1950 个关系事实。其中有 39528 个数据来自 Freebase 的实体和 53 类关系, 包括 52 类通用关系和一类特殊关系 NA(表示两个实体没有特定关系)。

本文采用与文献[8-11]一致的评估方法——留出法(hold-out), 该方法通过比较从测试集句子中发现的关系事实与测试集标记的关系事实来评估模型, 避免了人工评估的昂贵代价。实验结果采用准确率-召回率(PR)曲线、平均准确率($P@N$)作为评价指标, 通过对比前 N 项准确率以及 PR 曲线面积来评估模型的性能。准确率、召回率和平均准确率的计算公式如下:

$$precision = \frac{a}{b} \quad (15)$$

$$recall = \frac{a}{c} \quad (16)$$

$$P@N = \frac{a_N}{N} \quad (17)$$

其中, a 表示被正确分类的关系实例个数; b 表示被判别为某类的关系实例数目; c 表示测试集的关系实例总数; $P@N$ 表示将测试关系实例按照预测概率进行排序, 得到概率最大的前 N 项实例关系分类的准确率; a_N 表示前 N 项实例中被正确分类的数目。

3.3 参数设置

实验使用了对 NYT-Freebase 语料库中训练得到的词向量, 词向量维度大小为 50, 输入文本填充到固定大小 100, 在模型训练过程中采用 dropout 策略防止过拟合, 引入 L2 正则化约束网络参数。根据以往卷积神经网络上关系抽取的研究^[9-11]以及调参过程, 实验中所采取的超参数设置如表 2 所列。

表 2 超参数设置

Table 2 Hyperparameter settings

Hyperparameter	value
Word dimension d_w	50
Position dimension d_p	5
Convolution kernel size h	3
Number filters m	64
Initial Learning Rate η	0.001
Batch size B	64
Dropout probability D	0.5
L2 regularization λ_{l2}	0.0001

3.4 实验验证

为了验证提出的 DenseCNN 关系抽取模型的性能, 本文进行了以下两组对比实验。

(1) 本文以标准 CNN 网络为基线模型, 将 5 层 CNN 网络构成的模型(CNN-5)与 5 层密集连接网络构成的模型(DenseCNN-5)做对比, 实验中的参数均保持一致。对比实验的 PR 曲线如图 4 所示。

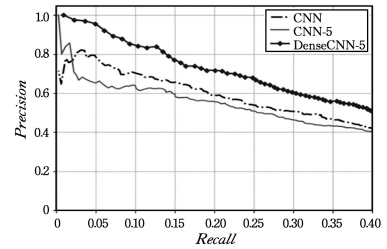


图 4 DenseCNN 模型与 CNN 基线模型的 PR 曲线对比

Fig. 4 Comparison of PR curves of DenseCNN and other CNN baselines

从图 4 可以看出, 首先, 对比 CNN 基线模型, 增加模型深度(如 CNN-5)并没有达到很好的效果, 反而比基线模型差, 这可能是因为深度网络发生了过拟合, 导致模型在测试数据上的表现较差。其次, 使用密集连接模块的 DenseCNN-5 模型在性能上优于 CNN 和 CNN-5, 因为密集连接模块通过特征融合将浅层特征传递至深层网络, 综合利用了浅层的低级特征和深层的复杂特征, 有助于模型具备更好的泛化性能, 防止过拟合。

(2) 本文以密集连接网络为基线模型, 在参数一致的情况下, 对比不同深度的密集连接网络模型(DenseCNN-gxg, 其中 x 为 CNN 的层数)对实验结果的影响。实验的 PR 曲线如图 5 所示。

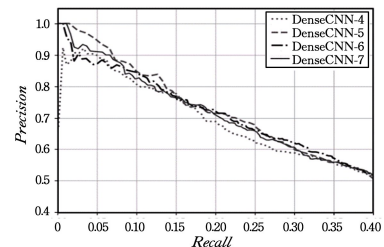


图 5 不同深度 DenseCNN 模型的 PR 曲线对比

Fig. 5 Comparison of PR curves of DenseCNN with different layers

从图 5 可以看出, 随着密集连接网络深度的增加, 关系分类的准确率有较小提升, 当网络深度为 6 层和 7 层时, PR 曲线面积均在 0.42~0.43 之间, 网络性能逐渐达到饱和。同

时,表3列出了不同深度密集连接网络模型的识别效率,其中包括模型训练效率即模型单次遍历完训练集数据的平均时长(Train time per epoch)、测试效率即模型完成测试数据关系抽取的时长(Test time)、模型可训练参数数目(Num params)和参数占用显存大小(Params memory)。

表3 不同深度 DenseCNN 模型识别的效率对比

Table 3 Efficiency comparison of DenseCNN with different layers

Models	Train time per epoch/min	Test time/min	Num params	Params memory/MB
DenseCNN-4	2.54	1.80	416107	1.59
DenseCNN-5	4.47	2.25	1610859	6.14
DenseCNN-6	9.88	3.09	6359659	24.26
DenseCNN-7	29.76	7.31	25294443	96.49

基于以上结果可以得出结论,深层密集连接网络比浅层有效,但随着深度的增加,特征数目也会线性增长,导致模型计算过于复杂,并且训练时需耗费大量时间和 GPU 显存。因此,本文选择 5 层密集连接网络模型(DenseCNN-5)作为后续研究的基准模型,其在效率和复杂度适中的情况下能达到较好的性能。

3.5 实验分析结果

为了评估本文所提模型 DenseCNN 在远程监督关系抽取任务上的性能,将 DenseCNN 模型与近年的几种基线模型进行比较。

CNN+ATT(由 Lin 等于 2016 年提出)利用 CNN 对句子进行建模,利用句子层面的注意力机制进行权重学习,解决了远程监督错误标签问题^[9]。

PCNN+ATT(由 Lin 等于 2016 年提出)在 CNN 的池化层通过两个实体位置将特征分为 3 段进行池化,并采用句子层面的注意力机制进行权重学习^[9]。

BLSTM+ATT(由 Zhou 等于 2016 年提出)使用双向循环神经网络(BLSTM)作为句子特征提取器,并采用词级别的注意力机制^[10]。

ResCNN-9(由 Huang 等于 2017 年提出)使用浅层残差网络作为句子编码器,没有使用注意力机制^[11]。

各个模型的精度-召回率曲线如图 6 所示,在整个召回率范围内,DenseCNN 模型的精确率都优于其他基线模型,并且本模型的 PR 曲线面积为 0.43,超过了 PCNN+ATT 模型的 0.36、BLSTM+ATT 模型的 0.367 和 ResCNN-9 模型的 0.38。

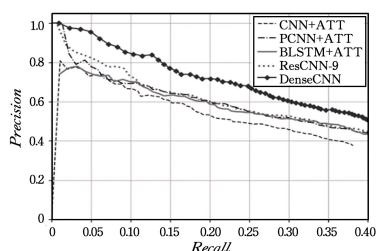


图6 DenseCNN 模型与基线模型的 PR 曲线对比

Fig. 6 Comparison of DenseCNN with other baselines

DenseCNN 模型与其他基线模型的平均准确率如表 4 所列。可以发现,即使没有使用注意力机制,DenseCNN 模型仍

然优于其他模型并取得了更优性能,在平均准确率指标上相对于 PCNN+ATT 和残差网络 ResCNN-9 提高了 7%~8%。同时,相比于 9 层的残差网络模型 ResCNN-9,DenseCNN 使用 5 层的密集连接网络就能达到一个较好的效果,在参数和计算成本更少的情形下实现了比残差网络更优的性能。

表4 DenseCNN 与基线模型的 P@N 对比

Table 4 Comparison of P@N of DenSeCNN with other baselines

P@N	(单位:%)			
	100	200	300	Mean
CNN+ATT	73.2	65.1	64.3	67.2
PCNN+ATT	80.0	74.2	66.8	74.0
BLSTM+ATT	76.0	72.5	67.7	72.1
ResCNN-9	82.0	73.0	70.7	75.2
DenseCNN	86.0	82.5	79.0	82.5

基于以上结果,可得出结论:

(1)密集连接卷积神经网络模型优于一般的卷积神经网络模型,这是因为密集连接网络通过特征通道合并实现特征重用,加强了特征传递,使网络学习出更为丰富的特征;同时能够解决深度神经网络梯度消失现象,网络对自然语言的表征能力更强。

(2)密集连接卷积神经网络不需要注意力机制或多实例学习就能获得较好的抽取效果。近年来,学者们针对远程监督数据噪声问题在神经网络方法上融合了多示例学习或注意力机制来过滤噪声,使得关系抽取的效果得到了提升。本文提出的关系抽取模型可以在句子编码后结合注意力机制或多示例学习来减弱噪声的影响,进一步提升关系抽取的性能。

结束语 本文提出了一种用于远程监督关系抽取的密集连接卷积神经网络模型——DenseCNN。由于密集连接的方式,该模型提升了梯度的反向传播,使得网络更容易训练,同时,模型通过合并不同层的特征,实现了特征复用,加强了特征传递。实验表明,本文提出的模型适用于远程监督关系抽取任务,有助于在有噪声的远程监督关系抽取任务中提取有用特征,从而提升远程监督关系抽取任务的性能。尽管密集连接网络能够在有噪声的远程监督关系抽取任务中提升性能,但是噪声问题并没有得到完全解决。下一步的工作可以考虑结合多示例学习或注意力机制进行关系抽取。

参考文献

- [1] KUMAR S. A Survey of Deep Learning Methods for Relation Extraction[J]. arXiv:1705.03645.
- [2] RINK B, HARABAGIU S. Utd:Classifying semantic relations by combining lexical and semantic resources[C]//Proceedings of the 5th International Workshop on Semantic Evaluation. Association for Computational Linguistics, ACL Anthology, Stroudsburg, PA, 2010:256-259.
- [3] BUNESCU R C, MOONEY R J. A shortest path dependency kernel for relation extraction[C]//Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing. Association for Computational Linguistics(ACL), Stroudsburg, PA, 2005:724-731.
- [4] ZENG D, LIU K, LAI S, et al. Relation classification via convo-

- lutional deep neural network[C] // Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics; Technical Papers. Association for Computational Linguistics, ACL Anthology, Stroudsburg, PA, 2014; 2335-2344.
- [5] EBRAHIMI J, DOU D. Chain based RNN for relation classification[C] // Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies. Association for Computational Linguistics (ACL), Stroudsburg, PA, 2015; 1244-1249.
- [6] XU Y, MOU L, LI G, et al. Classifying relations via long short term memory networks along shortest dependency paths[C] // Proceedings of the 2015 conference on empirical methods in natural language processing. Association for Computational Linguistics (ACL), Stroudsburg, PA, 2015; 1785-1794.
- [7] MINTZ M, BILLS S, SNOW R, et al. Distant supervision for relation extraction without labeled data[C] // Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP. Association for Computational Linguistics (ACL), Stroudsburg, PA, 2009; 1003-1011.
- [8] ZENG D, LIU K, CHEN Y, et al. Distant supervision for relation extraction via piecewise convolutional neural networks [C] // Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics (ACL), Stroudsburg, PA, 2015; 1753-1762.
- [9] LIN Y, SHEN S, LIU Z, et al. Neural relation extraction with selective attention over instances[C] // Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics (ACL), Stroudsburg, PA, 2016; 2124-2133.
- [10] ZHOU P, SHI W, TIAN J, et al. Attention-based bidirectional long short-term memory networks for relation classification [C] // Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics (ACL), Stroudsburg, PA, 2016; 207-212.
- [11] HUANG Y Y, WANG W Y. Deep Residual Learning for Weakly-Supervised Relation Extraction[J]. arXiv; 1707.08866.
- [12] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, 2017; 4700-4708.
- [13] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [C] // Advances in Neural Information Processing Systems. Neural Information Processing Systems Foundation, 2013; 3111-3119.
- [14] NASRABADI N M. Pattern recognition and machine learning [J]. Journal of Electronic Imaging, 2007, 16(4); 049901.
- [15] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. arXiv; 1502.03167.
- [16] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks[C] // Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. Microtome Publishing, Menlo Park, CA, 2011; 315-323.
- [17] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout; a simple way to prevent neural networks from overfitting [J]. The Journal of Machine Learning Research, 2014, 15(1); 1929-1958.
- [18] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, Los Alamitos, CA, 2016; 770-778.
- [19] DE BOER P T, KROESE D P, MANNOR S, et al. A tutorial on the cross-entropy method[J]. Annals of Operations Research, 2005, 134(1); 19-67.
- [20] RIEDEL S, YAO L, MCCALLUM A. Modeling relations and their mentions without labeled text[C] // Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Berlin; Springer, 2010; 148-163.



QIAN Xiao-mei, born in 1995, postgraduate. Her main research interests include information content security and so on.



CHENG Peng-sen, born in 1988, Ph.D candidate. His main research interests include information content security and so on.