



开放科学
(资源服务)
标识码
(OSID)

基于预训练模型的政务领域实体关系抽取

葛世奇¹ 孙新^{1,2} 寇桓锦¹ 袁燕^{1,2}

1. 北京理工大学计算机学院 北京 100081;
2. 北京市海量语言信息处理与云计算应用工程技术研究中心 北京 100081

摘要: [目的/意义] 海量电子政务信息资源极大地方便了人们对信息的获取,同时也给人们有效获取信息和知识提出了挑战。关系抽取是信息抽取的核心任务,政务领域的关系抽取对后续面向政务领域的智能检索、智能问答及政府公文生成等智慧化服务与建设具有深远影响。但是政务领域的文本中存在着大量的长实体,使得传统的关系抽取模型在政务领域数据集中的表现不尽如人意。[方法/过程] 本文提出一种适用于政务领域的中文关系抽取模型 CPRE-BERT, 首先使用预训练 BERT 模型作为编码器,能够在一定程度上解决多义性、边界模糊以及领域关系数据数量不足的问题。其次,采用分类池化的思想分别处理短实体和长实体,最大程度上保留实体信息。最后,在中文领域人物关系公共数据集和政务领域数据集上的对比实验结果验证了模型的有效性。[结果/结论] 实验结果表明,本文提出的方法在政务领域数据集上关系抽取的准确率和 F1 指标较基线模型分别提高了 2.3% 和 2.2%。

关键词: 关系抽取; 深度学习; BERT 模型; 分类池化

中图分类号: G35; TP391

Relation Extraction Based on Pre-trained Model for E-government

GE Shiqi¹ SUN Xin^{1,2} KOU Huanjin¹ YUAN Yan^{1,2}

1. School of Computer Science, Beijing Institute of Technology, Beijing 100081, China;
2. Beijing Engineering Applications Research Center on High Volume Language Information Processing and Cloud Computing, Beijing 100081, China

Abstract: [Objective/Significance] Massive e-government information resources not only bring great convenience, but also pose a challenge to people's effective access to information and knowledge. Relation extraction is the core task of information extraction, and plays an important role in the intelligent services such as intelligent retrieval, intelligent question answering etc. Due to the large number of long entities in the field of e-government, the performance of the classical relationship extraction

基金项目 中国科学技术信息研究所情报工程实验室开放基金项目“面向政府公文的实体关系抽取研究”。

作者简介 葛世奇(1997-), 硕士研究生, 研究方向为自然语言处理、关系抽取; 孙新(1975-), 博士, 副教授, 研究方向为人工智能、自然语言处理; 寇桓锦(1998-), 硕士研究生, 研究方向为自然语言处理; 袁燕(1989-), 硕士, 实验师, 研究方向为自然语言处理, E-mail: yuanyan@bit.edu.cn。

引用格式 葛世奇, 孙新, 寇桓锦, 等. 基于预训练模型的政务领域实体关系抽取 [J]. 情报工程, 2022, 8(4): 3-13.

model in the data set of the government field is not satisfactory. [Methods/Processes] To address this issue, we propose a novel relation extraction model, which employs BERT model to solve the problems of ambiguity, fuzzy boundary, and insufficient annotated data. Secondly, the classification pool method is used to process short entities and long entities respectively which can obtain better entity information. Finally, the experimental results on public datasets and government domain datasets demonstrate the effectiveness of the model. [Results/Conclusions] Compared with the baseline model, the accuracy and F1-score of this model are improved by 2.3% and 2.2% respectively.

Keywords: Relation Extraction; deep learning; BERT Model; classification pooling

引言

近年来,“互联网+政务服务”已逐渐成为政府工作的常态,海量电子政务信息资源给人们有效获取信息和知识提出了挑战,如何利用自然语言处理(Natural Language Processing, NLP)技术从形式复杂、多源分散的公文文本中提取出高质量、结构化知识便成为当前亟待解决的问题。

实体关系抽取任务是信息抽取、信息检索等领域的核心任务和重要环节,目的是发现和识别句子中已标记实体对之间的语义关系。实体关系抽取技术从基于规则的方法到现在基于深度学习的方法,在准确率、召回率等很多方面取得了巨大的进步^[1]。深度学习的方法能够自动地从大规模标注语料中获取语义关系的有效特征,为实体关系抽取任务带来了新的突破,典型方法包括:基于卷积神经网络(Convolutional Neural Network, CNN)的关系抽取^[2]、基于循环神经网络(Recurrent Neural Network, RNN)的关系抽取、基于长短期记忆网络(Long Short-Term Memory, LSTM)的关系抽取^[3]及其改进结构^[4]等。BERT(Bidirectional Encoder Representations from Transformers)模型^[5]能够将句子生成带有上下文语义信息的词向量序列,

在NLP的诸多领域取得了出色的成绩。BERT模型强大的语义学习能力使关系抽取模型的性能取得了进一步的提升^[6-9]。

深度学习的发展使得关系抽取任务在通用领域上取得了良好效果,也有学者开展了中文特定领域的关系抽取研究^[10,11],但是鲜有政务领域的中文关系抽取。政务领域中海量新闻报道以及电子公文里蕴含着大量的实体关系,并且通常较长实体,例如政务关系中的职位“XX集团董事会战略规划与可持续发展委员会委员”、会议名称如“中共中央第十一届三中全会”等,政务领域的实体特点使得经典的关系抽取模型在长实体占比大的政务领域数据集上的表现欠佳。

本文重点研究面向中文政务领域的关系抽取模型,针对政务领域包含较长实体的特点以及中文文本具有复杂的语法和句式的问题,提出了一种基于BERT的关系抽取模型(Classification Pooling Relation Extraction based on BERT, CPRE-BERT),采用分类池化的思想分别处理短实体和长实体。首先,采用BERT模型作为编码器获得包含文本丰富语义信息的表示,在一定程度上解决多义性、边界模糊以及领域关系数据数量不足的问题。其次,根据中文和领域文本特点,在池化层进行改进,最

大程度上保留实体信息。最后在政务领域的对比实验结果验证了 CPRE-BERT 模型的有效性。

1 相关工作

有监督关系抽取方法和远程监督关系抽取方法是实体关系抽取的常用方法^[12]。有监督的关系抽取旨在从大规模标注语料中获取表达实体间语义关系的有效特征，在已标注的数据基础上训练分类器，主要方法有基于规则的、基于特征工程的以及基于深度模型的方法。为了能够自动获取高质量的标注数据，Mintz 等^[13]于 2009 年提出了远程监督的方法，通过数据自动对齐远程知识库来解决开放域中大量无标签数据自动标注的问题，但存在远程监督假设引入噪声数据的问题^[14,15]。

深度学习方法相比传统方法能够获取更为隐含的语义信息，随着深度学习技术的逐渐成熟，基于深度学习的方法开始被广泛采用，包括基于循环神经网络 RNN 的方法^[16]、卷积神经网络 CNN 的方法^[2]，以及基于多实例学习与分段卷积神经网络模型（Piecewise Convolutional Neural Networks, PCNN）的远程监督^[14]的方法等。Lin 等^[17]在前人研究的基础上引入注意力机制，综合利用包中所有语句的信息，进一步提升了关系抽取模型的性能。后来随着 LSTM 的出现，Zhou 等^[3]将 LSTM 应用到了关系抽取中。李志欣等^[4]提出一种以完全依赖树作为直接输入，利用图卷积网络结合两个并行的注意力模块，自主学习如何有选择地关注对关系抽取任务有用的信息。这些方法的主要瓶颈在于特征提取以及语义表征能力有限，导致

实体嵌入之间缺少必要的语义信息关联，从而影响关系抽取效果。

2018 年，Google 提出了 BERT 预训练语言模型^[5]。BERT 模型是一个双向的基于 transformer 的模型，它利用大量的无标注语料进行训练，能够全面并准确地捕捉语句的特征并生成带有这些特征的词向量，进而生成质量远高于 LSTM 等传统神经网络结构的词向量。Wu 等^[6]提出了 R-BERT 的关系抽取模型，使用实体和实体位置在预训练模型中的结合方式，对编码之后的实体向量序列做平均池化之后，与 BERT 模型输出的 [CLS] 向量进行拼接，再通过一层全连接网络映射到关系集合中，最后通过 SoftMax 方法来预测实体之间存在的关系，该模型在英文公开数据集中取得了较好的效果。

本文引入 BERT 模型来进行中文政务领域的实体关系抽取，通过使用预训练模型自动获得具有较丰富语义的特征向量，提高关系抽取的效果。

2 基于 BERT 的分类池化关系抽取模型

本文提出一种政务领域的关系抽取模型 CPRE-BERT，采用 R-BERT^[6]作为本文模型的核心部分，挖掘政务领域实体关系特征信息。针对政务领域中存在大量较长实体的问题，提出了分类池化的思想，对短实体和长实体进行分别处理，避免对所有实体都进行平均池化的方法无法很好表达实体语义信息的问题。此外，模型 CPRE-BERT 直接将 BERT 模型输出的头实体和尾实体拼接出来作为关系分类的中间向

量,提高算法效率。

2.1 BERT模型

Transformer^[18]是Google团队在2017年6月提出的NLP经典之作。Transformer模型有效地避免了CNN和RNN的缺点,引入了残差前馈网络和多头机制,使得Transformer解决了长期依赖问题。BERT模型是Transformer一个最出众的改进版本,该模型主要由多层双向Transformer编码器部分组成,模型结构示意图如图1所示。

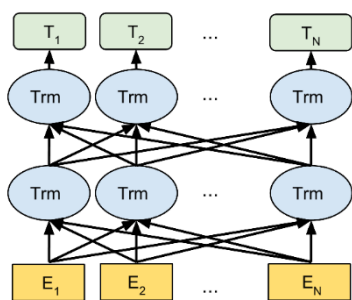


图1 BERT模型结构示意图^[5]

对于句子 $X=\{x_1, x_2, \dots, x_l\}$,其中为序列长度, x_i 表示序列中第 i 个字, $i=1, 2, \dots, l$,BERT模型的字嵌入模块会对该序列中的每一个字进行符号嵌入(记为 w_i)和位置嵌入(记为 p_i)。最终的嵌入表示记为 e_i ,即将每个字向量 w_i 编码 e_i , e_i 表示为二者之和。

$$e_i = w_i + p_i \quad (1)$$

之后,经过BERT用于语言理解的Transformer预训练模型,将嵌入结果送入多层Transformer编码器模块中进行编码,最终得到输入语句 X 的字向量表示 $[t_1, t_2, \dots, t_l]$,如式(2)所示,其中, l 为序列长度。 t_i 为对应 e_i 的输出向量。

$$[t_1, t_2, \dots, t_l] = \text{BERT}([e_1, e_2, \dots, e_l]) \quad (2)$$

2.2 CPRE-BERT模型

CPRE-BERT模型的主要思想是:采用BERT模型进行编码,根据头实体和尾实体的长度的不同,分别采用平均池化和最大池化的方法得到头实体和尾实体对应的向量,之后经过一个线性变换和一个激活函数的激活之后,将两者进行拼接,再经过一层全连接网络完成关系映射。最后,根据映射的结果使用softmax进行筛选,得到概率最大的关系作为最终的关系预测结果。模型使用交叉熵损失函数作为优化目标。

CPRE-BERT模型包括三个部分,分别是编码层、池化层和预测层,如图2所示。编码层的作用是使用BERT将输入的句子变为向量序列;池化层的作用是将头、尾实体所包含的向量序列进行压缩,得到最终实体的向量表示;预测层的作用是使用头、尾实体的表示向量预测两个实体之间的关系。CPRE-BERT模型如图2所示。

2.2.1 编码层

由于BERT在词嵌入层使用的是符号嵌入和位置嵌入加和的方式^[5],因此在将输入句子送入BERT模型之前,首先对句子进行预处理:在头实体和尾实体的两端分别插入分隔符^[6]。这样做的目的是让BERT能够捕捉到头实体和尾实体的位置信息。对于中文BERT模型来说,我们注意到其词典中有[unused]符号,因此为了避免句中其他特殊符号与分隔符发生冲突,在头实体和尾实体头尾插入[unused1]和[unused2],用来使BERT模型捕捉头尾实体的位置信息。

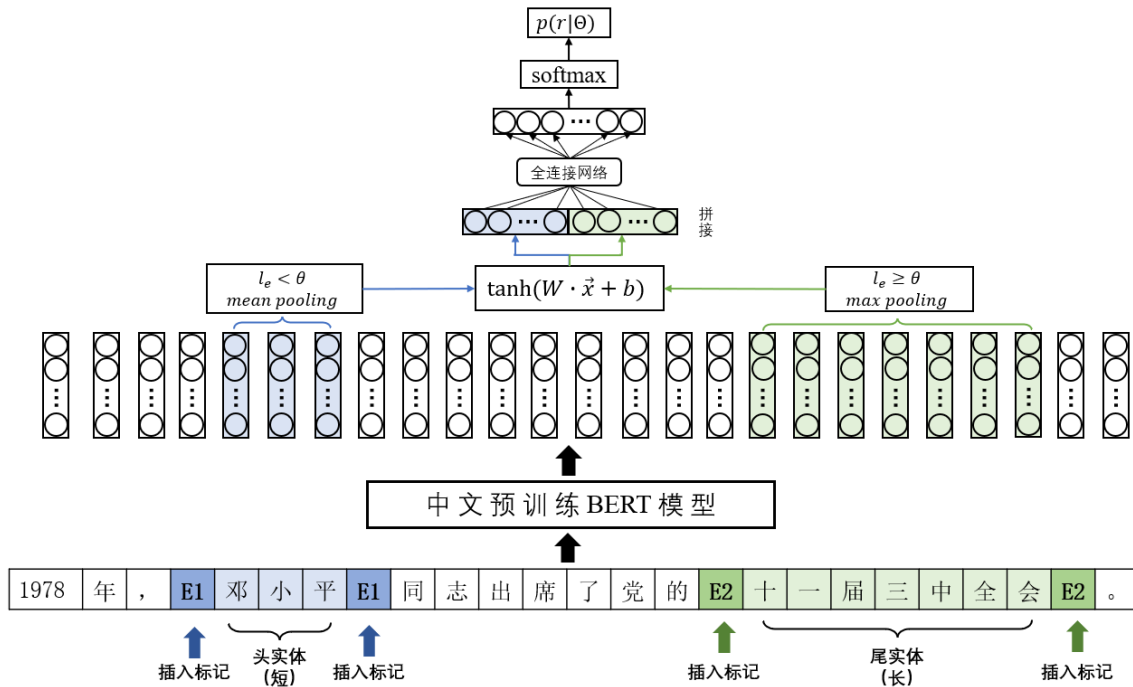


图 2 CPRE-BERT 模型图

因此，对于例句：

“1978年，邓小平同志出席了党的十一届三中全会。”

经过预处理之后，送入 BERT 模型的句子为：

“1978年，[unused1]邓小平[unused1]同志出席了党的[unused2]十一届三中全会[unused2]。”

将预处理之后的句子送入预训练好的 Bert-base-Chinese 模型中，经过 BERT 模型的嵌入处理，得到句子的嵌入序列 $[e_1, e_2, \dots, e_l]$ 。之后通过多层注意力机制网络，得到句子的表示向量，即每一个单位的表示向量组成的向量序列 $[t_1, t_2, \dots, t_l]$ 。

2.2.2 分类池化方法

池化层的目的是将包含多个字向量的头实体和尾实体压缩成为一个向量，送给预

测层进行预测。这里采用分类池化的方法，首先设置阈值 θ ，对于句子 X 中的实体长度 l_e ，当 $l_e < \theta$ 时，对 BERT 输出的实体对应的向量序列使用平均池化进行处理，当 $l_e \geq \theta$ 时，采用最大池化，并将池化过后的向量送入激活函数中进行激活。实体向量表示 h_e 的计算方式如下：

$$h_e = \tanh(c_pooling([t_i, t_{i+1}, \dots, t_j])) \quad (3)$$

i, j 代表的是实体在句子中的开始位置和结束位置， \tanh 为双曲正切激活函数。

$c_pooling$ 即为分类池化操作， $c_pooling$ 的具体实现如下：

$$c_pooling([t_i, t_{i+1}, \dots, t_j]) = \begin{cases} \frac{\sum_{k=i}^j t_k}{j-i+1} & \text{当 } l_e < \theta \\ \maxpool([t_i, t_{i+1}, \dots, t_j]) & \text{当 } l_e \geq \theta \end{cases} \quad (4)$$

其中， i, j 代表的是实体在句子中的开始位

置和结束位置, $l_e=j-i+1$, 代表实体的长度; i 和 j 分别为实体的开始位置和结束位置, $max\text{-pool}$ 为最大池化操作。

采用分类池化的思想, 主要理由如下:

(1) 采用平均池化 (mean-pooling) 的方法体现了“平均”的思想, 将每一个向量赋予相同权重进行相加。这种方式考虑到了每一个向量的值, 并将其都加入到池化结果之中。这种思想对短实体来说, 能够更好地表达整个短实体的语义, 通过这种方式可以获取特征的平均信息。

(2) 最大特征池化 (max-pooling) 的方法选取最显著的特征。对于每一个分量, 取特征中得分最大的值作为池化结果, 其它特征值全部抛弃。对于长实体, 这种思想相比平均池化更能够表达整个长实体的语义。因为对于长实体来说, 如果将所有向量都赋予相同权重进行相加, 会冲淡重要信息的影响, 可能使得整体语义的含义偏离整体的含义。因此, 使用最大池化能够保留长实体的向量序列中重要的信息, 可以使池化结果更好地表达相应的语义, 模型可以获取到文本单词的最显著特征。

2.2.3 预测层

预测层的功能是使用前一步生成的头实体向量 h_e^h 和尾实体向量 h_e^t , 预测出该实体对之间存在每一种关系的概率。首先对于池化层输出的头尾实体向量, 通过一层全连接网络和激活函数, 生成最终的实体向量表示 v_e :

$$v_e = \tanh(W_e h_e + b_e) \quad (5)$$

其中 W_e 代表可训练的权重, h_e 为池化层输出的实体表示向量, 为方便表示, 这里隐去向

量 h_e 和 v_e 上标 h 和 t , b_e 为对应偏置值, \tanh 表示双曲正切激活函数。这里, 头尾实体对应的权重和偏置均相等。

下一步将头、尾实体向量拼接起来, v_e^h 做归一化处理, v_e^t 表示经过全连接网络和激活函数的头实体向量, 表示相应尾实体向量, $concat$ 为向量拼接操作, $norm$ 为归一化操作:

$$h_r = norm(concat(v_e^h, v_e^t)) \quad (6)$$

最后, 对该向量使用 $softmax$ 方法, 得到预测结果:

$$p(r|\Theta) = softmax(h_r) \quad (7)$$

其中 Θ 代表模型中所有的参数, $p(r|\Theta)$ 即为在当前模型参数下, 头尾实体之间包含关系集中每种关系的概率。

模型的损失函数采用常规的交叉熵损失函数, 如公式 (8) 所示:

$$L = \sum \log(p(r|\Theta)) \quad (8)$$

3 实验结果及分析

为了验证 CPRE-BERT 模型进行关系抽取的性能, 本文设计了三组实验, 分别为: (1) 在公共数据集上进行不同池化方法的对比实验; (2) 在政务领域数据集, 针对分类池化思想模型和未分类池化的模型进行对比实验, 验证本文方法在政务领域中的有效性; (3) 针对模型是否带有 [CLS] 向量进行对比实验。

3.1 实验数据及设置

实验使用的数据集分别是中文领域人物关系公共数据集^[19]和自行标注的政务领域数据集, 基本信息如表 1 所示。

表 1 实验用数据集信息

	人物关系公共数据集	政务领域数据集
关系数量	12	16
训练集数量	15000	2859
测试集数量	3000	813

数据集中长、短实体的分布统计情况如表 2 所示。从表 2 中可以看出，政务领域数据集相对于公共数据集，包含了更多的长实体，占比达到 38%。由于这些长实体的存在，本文模型对于长实体和短实体之间的分类处理显得尤为必要。

表 2 数据集实体占比情况

数据集名称	长度<6的实体占比	长度6的实体占比
政务领域数据集	62%	38%
公共数据集	99.4%	0.6%

模型训练采用的参数如表 3 所示。其中，在公共数据集上 Max sentence length 值为 128，而在政务领域 Max sentence length 为 256，这是由政务领域数据集中数据的特点所决定的。

表 3 训练参数

参数	值
Batch size	8
Max sentence length	128/256
Learning rate	1e-5
Epochs	20
Dropout rate	0.3

评价指标采用关系抽取任务常用的评价指标 F1 值来进行评测，F1 值的计算公式如下：

$$Pre = \frac{\text{抽取出的正确关系数量}}{\text{抽取出的所有关系数量}} \quad (10)$$

$$Pec = \frac{\text{抽取出的正确关系数量}}{\text{数据集中实际含有的关系数量}} \quad (11)$$

$$F1 = \frac{2 \cdot Pre \cdot Rec}{Pre + Rec} \quad (12)$$

其中，Pre 为准确率，Rec 为召回率，F1 值为两者的调和平均值。

本文采用的对比模型有：LSTM+ATT 模型^[3]，R-BERT 模型^[6]，以及替换掉原 R-BERT 模型池化方式的 R-BERT_{Max} 模型，该模型将原 R-BERT 模型中的平均池化替换为了最大池化；以及 R-BERT_{LogSumExp} 模型，该模型将原 R-BERT 模型中的平均池化替换为 LogSumExp 池化^[20]。

为了验证分类池化思想对实验结果的影响，基于本文 CPRE-BERT 分类池化的思想，实验部分除了本文提出的 CPRE-BERT 外，还增加了 CPRE-BERT_{LogSumExp}、CPRE-BERT_{LSTM}、CPRE-BERT_{ATT}。三个模型。这三个模型对于短实体的池化方式均保留平均池化方式不变，对于长实体的池化公式如表 4 所示。其中， \bar{h}_e 和 \bar{h}_e 分别为正向 LSTM 和反向 LSTM 编码器的输出值， t_k 为 BERT 模型对第 k 个符号的输出向量。 t_{cls} 为 BERT 输出的句子 [CLS] 向量的值。i 和 j 分别为实体在句中的开始位置和结束位置。

表 4 对比模型池化公式

模型名称	池化公式
CPRE-BERT _{LogSumExp}	$pool([t_i, t_{i+1}, \dots, t_j]) = \log \sum_{k=i}^j \exp(t_k)$
CPRE-BERT _{LSTM}	$\bar{h}_e = \overline{LSTM}([t_i, t_{i+1}, \dots, t_j])$ $\bar{h}_e = \overline{LSTM}([t_i, t_{i+1}, \dots, t_j])$ $pool([t_i, t_{i+1}, \dots, t_j]) = (\bar{h}_e + \bar{h}_e) / 2$
CPRE-BERT _{ATT}	$\alpha = softmax(t_{cls} \cdot [t_i, t_{i+1}, \dots, t_j]^T)$ $pool([t_i, t_{i+1}, \dots, t_j]) = \alpha \cdot [t_i, t_{i+1}, \dots, t_j]$

此外，为了验证 [CLS] 向量对关系抽

取结果的影响,本文也添加了基于CPRE-BERT模型,并禁用分类池化机制的3个变种模型:CPRE-BERT_{Mean-all}、CPRE-BERT_{Max-all}和CPRE-BERT_{LogSumExp-all}。上述3种模型基于本文CPRE-BERT基本框架,将长实体和短实体的池化方式统一为平均池化(CPRE-BERT_{Mean-all}),最大池化(CPRE-BERT_{Max-all})和LogSumExp池化^[20](CPRE-BERT_{LogSumExp-all})。

3.3 实验结果

3.3.1 不同池化方法的实验结果

在公共数据集中,由于长实体只占了所有实体数量的0.6%,实体普遍较短,因此主要实验目的是验证不同的池化方式对关系抽取模型的影响。公共数据集上的实验结果如表5所示。

首先,Bi-LSTM+ATT模型和其他的基于BERT模型的结果相差较大,这说明BERT模型的特征提取能力是要远高于传统的LSTM模型的。

其次,在公共数据集上,最大池化方式(R-BERT_{Max}、CPRE-BERT_{Max-all})与使用平均池化的方式(R-BERT、CPRE-BERT_{Mean-all})相比较差,主要原因是,在短实体占比相当大的情况下,对短实体使用平均池化方式能够考虑实体中每一个字的语义,最大程度保留了实体整体的语义信息。对短实体采用最大池化方式会导致短实体所包含的本就不多的向量序列出现信息损失。而LogSumExp池化是一种平滑的最大池化,介于平均池化和最大池化之间,该池化方式的模型(R-BERT_{LogSumExp}、CPRE-BERT_{LogSumExp-all})取得介于两者之间的结果,说明对于短实体,平均池化是最能够表达

其语义的方式。

表5 公共数据集上各个模型的表现

模型	准确率	召回率	F1值
Bi-LSTM+ATT	0.645	0.641	0.643
R-BERT	0.813	0.811	0.812
R-BERT _{Max}	0.808	0.804	0.806
R-BERT _{LogSumExp}	0.810	0.808	0.809
CPRE-BERT _{Mean-all}	0.812	0.810	0.811
CPRE-BERT _{Max-all}	0.807	0.809	0.808
CPRE-BERT _{LogSumExp-all}	0.806	0.810	0.808

3.3.2 关系抽取的对比实验

第二组实验在本文自行标注的政务领域数据集上进行。在该数据集中,长实体占比达到38%,对长实体进行分类池化对于结果的影响变得明显,因此,在该组实验中增加本文中提出的分类池化思想的系列模型。另外,由于表4中Bi-LSTM+ATT的结果相对于BERT模型的结果相差较多,因此这里不再使用LSTM模型作为对比。各个模型在该数据集中的实验结果如表6所示。

表6 政务数据集上各个模型的表现

模型	准确率	召回率	F1值
R-BERT	0.849	0.850	0.849
R-BERT _{Max}	0.860	0.854	0.857
R-BERT _{LogSumExp}	0.841	0.867	0.854
CPRE-BERT _{Mean-all}	0.870	0.858	0.864
CPRE-BERT _{Max-all}	0.867	0.861	0.864
CPRE-BERT _{LogSumExp-all}	0.860	0.874	0.867
CPRE-BERT _{LSTM}	0.841	0.867	0.854
CPRE-BERT _{ATT}	0.859	0.865	0.862
CPRE-BERT _{LogSumExp}	0.870	0.862	0.866
CPRE-BERT	0.872	0.870	0.871

从表 6 中可以看出,对长实体占比较多的政务数据集,使用最大池化比使用平均池化能够更好地捕捉长实体的语义信息。另外,通过 CPRE-BERT_{Mean-all} 和 CPRE-BERT_{Max-all} 的对比也可以发现,对于长实体占比较大的数据,全部采用最大池化和全部采用平均池化的结果相当,这是因为不管是平均池化和最大池化,针对长实体和短实体都分别有优点和缺点。而 CPRE-BERT_{LogSumExp-all} 使用平滑的最大池化 LogSumExp 池化方式,介于二者之间,取得了比两者都要好的效果。

此外,基于分类池化的思想,对比相关模型(CPRE-BERT_{LSTM}、CPRE-BERT_{ATT}、CPRE-BERT_{LogSumExp}、CPRE-BERT)可以发现,分类池化思想在对短实体使用平均池化的基础上,对长实体使用最大池化能够取得最好的效果。使用 LSTM 的模型(CPRE-BERT_{LSTM})由于缺少实体两侧的上下文信息,因此在 LSTM 中,会错误地记忆与语义无关的词向量信息,导致整体效果较差。使用 [CLS] 向量做注意力机制的模型(CPRE-BERT_{ATT})效果较差的原因是:在长实体中,一些对实体特征贡献较小的词,可能对整体语义的贡献会比较大,因此使用注意力机制,同时使用 [CLS] 向量作为查询向量,会导致权重分配出现偏差。使用 LogSumExp 池化的模型(CPRE-BERT_{LogSumExp})尽管使用最大池化的思想,但是在计算过程中仍然会引入其他向量的成分对最终结果造成影响,从而导致结果仍然不如最大池化。

本文提出的 CPRE-BERT 模型效果最好。主要原因是:通过使用预训练 BERT 模型最大程度获得包含上下文信息的具有丰富语义的特

征向量,对这些向量直接采用最大池化的操作,可以很好地保留了该序列中最具有代表性的特征,而舍弃其他无关特征。模型的对比实验证明了在分类池化的思想下,对长实体采用最大池化能够达到最好的效果。

3.3.3 [CLS]向量影响对比实验

进一步在公开数据集和政务领域数据集上增加对比实验,验证 [CLS] 向量对关系抽取模型的影响,实验结果如表 7 所示。可以看出,在训练数据充足的公开数据集中,三种去掉 [CLS] 的方式(CPRE-BERT_{Mean-all}、CPRE-BERT_{Max-all}、CPRE-BERT_{LogSumExp-all}) 在 F1 值上比带有 [CLS] 的同类方法(R-BERT、R-BERT_{Max}、R-BERT_{LogSumExp}) 结果相差仅 0.001 的 F1 值,且使用最大池化的两个模型(CPRE-BERT_{Max-all}、R-BERT_{Max}) 对比可以发现,没有 [CLS] 向量的模型结果甚至优于带有 [CLS] 向量的同类模型。因此,在训练数据充足的数据集上,在数据能够保证参数拟合的情况下,[CLS] 向量对关系抽取的结果并无影响。

表 7 对比实验结果

模型	公共数据集	政务领域数据
	F1值	集F1值
R-BERT	0.812	0.849
R-BERT _{Max}	0.806	0.857
R-BERT _{LogSumExp}	0.809	0.854
CPRE-BERT _{Mean-all}	0.811	0.864
CPRE-BERT _{Max-all}	0.808	0.864
CPRE-BERT _{LogSumExp-all}	0.808	0.867

从表 7 中可以看出,在训练数据相对较少的政务领域数据集上,带有 [CLS] 向量的模型(R-BERT、R-BERT_{Max}、R-BERT_{LogSumExp}) 结果

明显比去掉 [CLS] 向量的模型 (CPRE-BERT_{Mean-all}、CPRE-BERT_{Max-all}、CPRE-BERT_{LogSumExp-all}) 要差。因此, 在训练数据不充足的数据集上, 由于 [CLS] 向量对实验结果并无帮助, 所以去掉 [CLS] 向量的模型能够更好地拟合数据, 并取得更好的效果。

从结果中可以看出, 无论是在公开数据集还是政务领域数据集, [CLS] 向量对实验结果的影响都可以忽略不计。当训练数据充足 (公开数据集) 时, 带有 [CLS] 向量与否的模型结果基本持平, 而在训练数据较少的时候, 不带有 [CLS] 向量的模型效果则明显优于带有 [CLS] 向量的同类模型。原因是在 BERT 模型作为编码器时, 由于使用了多层的多头注意力机制, 所以生成的字向量本身即带有上下文信息。因此在对实体进行池化操作后, 无需再将 [CLS] 向量加入作为语义信息的补充。去掉 [CLS] 向量, 能够减少模型的参数, 使模型能够使用更简单的结构达成更好的效果

3.3.4 实例分析

为了进一步说明本模型的特点, 选取 3 个代表性示例进行深入分析, 本小节使用了 CPRE-BERT_{Mean-all}、CPRE-BERT_{Max-all} 和 CPRE-BERT 3 个模型作为代表模型验证实验效果。3 个用例以及对应的三种模型的预测结果如表 8 所示。

由表 8 可以看出: (1) 示例 1 中, 对于头尾实体均为长实体的句子, 使用最大池化的模型可以更好地预测其中蕴含的关系类型, 而平均池化对此的表现不佳。(2) 示例 2 中, 对于头尾实体都是短实体的句子, 使用最大池化在某些时候会导致实体语义表达出错, 进而导致预测结果出现差错, 而使用平均池化则能更好

地保留相应信息。(3) 示例 3 中, 对于同时包含长实体和短实体的情况, 如果对头尾实体都使用平均池化或最大池化, 会导致不同实体的信息出现偏差, 从而出现不同表现的错误。而在三种示例中, CPRE-BERT 模型采用分类池化的思想, 能够根据句中实体长度的不同, 分别采用不同的策略, 进而取得更好的预测结果。

表 7 对比实验结果

模型	公共数据集	政务领域数据集
	F1值	F1值
R-BERT	0.812	0.849
R-BERT _{Max}	0.806	0.857
R-BERT _{LogSumExp}	0.809	0.854
CPRE-BERT _{Mean-all}	0.811	0.864
CPRE-BERT _{Max-all}	0.808	0.864
CPRE-BERT _{LogSumExp-all}	0.808	0.867

4 总结

本文针对政务领域的关系抽取, 结合政务领域数据特点, 简化并改进了 R-BERT 模型, 将其应用到了中文领域, 并调整了其训练参数, 进而在政务领域的数据集上相对一系列基线模型取得了最好的效果。提出分类池化思想, 并通过实验证明了对于长实体来说使用最大池化要比使用平均池化更能表达实体的语义信息。同时, 实验证明了 BERT 生成的词向量本身就具有包含上下文语义信息的特点, 因而也无需在头尾实体之间添加 [CLS] 向量来进行语义信息的补充操作。但是本文所提出的模型在性能上仍然存在不足, 后续可以尝试调整 BERT 模型的训练参数, 使之能够根据实体长度的特点生成相应的词向量; 同时改进尝试对头尾实体

向量表示, 更好地实现关系映射。

参 考 文 献

- [1] 鄂海红, 张文静, 肖思琪, 等. 深度学习实体关系抽取研究综述[J]. 软件学报, 2019, 30(6):1793-1818.
- [2] Zeng D, Liu K, Lai S, et al. Relation classification via convolutional deep neural network[C]. Proceedings of the 25th International Conference on Computational Linguistics (COLING 2014). 2014:2335-2344.
- [3] 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 (Volume 2: Short Papers). 2016: 207-212.
- [4] 李志欣, 孙亚茹, 唐素勤, 等. 双路注意力引导图卷积网络的关系抽取[J]. 电子学报, 2021(2):315-323.
- [5] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[C]. Proceedings of Annual Conference of the North American Chapter of the Association for Computational Linguistics. Minnesota: NAACL-HLT, 2019: 4171-4186.
- [6] Wu S C, He Y F. Enriching pre-trained language model with entity information for relation classification[C]. Proceedings of the 28th ACM International Conference on Information and Knowledge Management. 2019: 2361-2364.
- [7] Alt C, Hübner M, Hennig L. Fine-tuning pre-trained transformer language models to distantly supervised relation extraction [C]. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019:1388-1398.
- [8] Wang H, Focke C, Sylvester R, et al. Fine-tune bert for DocRED with two-step process[J]. arXiv preprint arXiv:1909.11898, 2019.
- [9] Zhou W, Huang K, Ma T, et al. Document-level relation extraction with adaptive thresholding and localized context pooling[C]. Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence. 2021:14612-14620.
- [10] 李芊芊, 张克亮. 基于依存分析的军事领域英文实体关系抽取研究[J]. 情报工程, 2019, 5(1):98-112.
- [11] 刘秀磊, 王延飞, 刘思含, 等. 科技情报对象关系抽取的技术选择[J]. 情报工程, 2018, 4(3):39-47.
- [12] 杨穗珠, 刘艳霞, 张凯文, 等. 远程监督关系抽取综述[J]. 计算机学报, 2021(8):1636-1660.
- [13] Mintz M, Bills S, SNOW R, et al. Distant supervision for relation extraction without labeled data[C]. Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics, Singapore: Association for Computational Linguistics. 2009:1003-1011.
- [14] Zeng D J, Liu K, Chen Y B, 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. 2015:1753-1762.
- [15] Shang Y M, Huang H Y, Sun X, et al. Are Noisy Sentences Useless for Distant Supervised Relation Extraction?[C]. Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20). 2020:8799-8806.
- [16] Socher R, Huval B, Manning C D, et al. Semantic Compositionality through Recursive Matrix-Vector Spaces[C]. Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Jeju Island: Republic of Korea. 2012:1201-1211.
- [17] Lin Y K, Shen S Q, Liu Z Y, et al. Neural relation extraction with selective attention over instances[C]. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016, 1: 2124-2133.
- [18] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]. Proceedings of the Annual Conference on Neural Information Processing Systems. California: NeurIPS. 2017:5998-6008.
- [19] 中文人物关系数据 [EB/OL]. [2022-01-03]. <https://github.com/buppt/ChineseNRE/raw/master/data/people-relation/train.txt>.
- [20] Jia R, Wong C, Poon H. Document-level N-ary relation extraction with multiscale representation learning[C]. Proceedings of Annual Conference of the North American Chapter of the Association for Computational Linguistics. Minnesota: NAACL-HLT. 2019: 3693-3704.