

# 融合多特征 BERT 模型的中文实体关系抽取<sup>①</sup>



谢 腾, 杨俊安, 刘 辉

(国防科技大学 电子对抗学院, 合肥 230031)

通讯作者: 谢 腾, E-mail: 2956957138@qq.com

**摘 要:** 关系抽取是构建知识图谱的一项核心技术. 由于中文具有复杂的语法和句式, 同时现有的神经网络模型提取特征有限以及语义表征能力较差, 从而影响中文实体关系抽取的性能. 文章提出了一种融合多特征的 BERT 预训练模型的实体关系抽取算法. 首先对语料进行预处理, 提取关键词、实体对信息和实体类型特征并进行融合, 以此来强化 BERT 模型的语义学习能力, 极大地减少了语义信息特征的丢失, 最后通过 Softmax 分类器进行关系分类. 实验结果表明, 文章模型优于现有的神经网络模型. 在人工标注的中文数据集上本文模型取得了 97.50% 的  $F1$  值.

**关键词:** 实体关系抽取; 双向编码表征模型; 词向量; 外部特征; 特征融合

引用格式: 谢腾, 杨俊安, 刘辉. 融合多特征 BERT 模型的中文实体关系抽取. 计算机系统应用, 2021, 30(5): 253-261. <http://www.c-s-a.org.cn/1003-3254/7899.html>

## Chinese Entity Relation Extraction Based on Multi-Feature BERT Model

XIE Teng, YANG Jun-An, LIU Hui

(College of Electronic Engineering, National University of Defense Technology, Hefei 230031, China)

**Abstract:** Relation extraction is a core technology to construct a knowledge graph. The complexity of Chinese grammar and sentence structure as well as the limited feature extraction and poor semantic representation of the existing neural network model restrict the relation extraction of Chinese entities. A relation extraction algorithm based on a BERT pretraining model is proposed in this study. It preprocesses the corpus by extracting keywords, entity pairs and entity type and integrating them to strengthen the semantic learning ability of the BERT model, greatly reducing the loss of semantic features. Results are obtained by a Softmax classifier, which show that this model is better than the existing neural network model. In particular, the model reaches a  $F1$ -score of 97.50% on the Chinese data set.

**Key words:** relation extraction; BERT model; word vector; external features; feature fusion

随着互联网的快速发展以及网络的广泛应用, 产生的信息呈爆炸式地增长, 让人们无法在海量数据中快速有效地找到准确的知识. 如何从海量信息中准确高效地找到用户所需要的信息成为目前的研究热点, 因此, 信息抽取技术受到学术界与工业界的广泛关注. 信息抽取技术作为自然语言处理的研究重点与热点. 它的研究热点主要包括实体抽取、关系抽取和事件抽取. 其中的关系抽取则是作为其核心任务. 关系

抽取主要解决的问题是从非结构化的文本语句中自动抽取实体间的语义关系, 并形成统一格式的结构化信息. 具体而言, 就是预先给定语句和候选实体, 关系抽取模型则需要依靠自身的特征提取能力来预测实体之间存在可能的语义关系. 关系抽取处理海量数据有很强的优势, 因此其在机器翻译、知识图谱的构建、问答系统以及智能搜索等应用中具有很大的研究价值.

<sup>①</sup> 基金项目: 安徽省自然科学基金 (1908085MF202)

Foundation item: Natural Science Foundation of Anhui Province (1908085MF202)

收稿时间: 2020-09-10; 修改时间: 2020-10-09; 采用时间: 2020-10-13; csa 在线出版时间: 2021-04-28

## 1 相关工作

实体关系抽取是自然语言处理的重要课题之一。随着知识图谱的发展, 实体关系抽取已经逐渐深入到其中, 并发挥相当重要的作用。就目前而言, 实体关系抽取算法主要分为3类: 一是基于模板特征匹配的关系抽取算法; 二是基于特征的关系抽取算法; 三是基于神经网络的关系抽取算法。基于模板的关系抽取是通过预先设定规则和语法或语义模式, 将其与文本进行匹配, 从而实现关系实例的抽取。该算法在小规模数据集上容易实现同时构建较简单, 并且能够达到较高的准确率, 但是它所耗费的人工成本相当高, 可移植性相当差, 而且这种算法很难定义文本中所有的关系模板, 因此也进而导致相对较低的召回率。

基于特征的关系抽取, 它主要依赖于上下文的各种词汇、句法以及语义信息等。具体而言, 给出相应的训练样本, 然后通过词汇、句法和语义分析等对训练语料进行特征提取, 再将提取的特征结果输入到分类器当中进行关系分类, 从而完成关系抽取。该算法的实质是将关系抽取看成是一个多分类问题, 通过构建一个分类器类进行实体关系抽取。该算法同样是严重依赖人工选择的特征, 如果这些特征能够为关系抽取提供有用的信息, 那么算法性能就会有大的提升; 但是若这些特征无法给出相应的有用信息, 它们就会严重影响关系抽取的性能, 尤其当训练样本的增多时, 这些特征反而会带来严重的噪声影响, 从而使得关系抽取的性能变得极其之差。

传统的关系抽取算法需要消耗大量的人力去设计特征, 对于大规模的抽取任务而言, 人工设计的特征作用并不大, 最重要的是语句中的一些隐性特征很难进行定义, 因而传统算法难以胜任大规模的关系抽取任务。而基于神经网络的关系抽取算法, 它不再依赖于手工设计的特征, 而是能够自动学习出有效的特征。这种方法可以在不使用自然语言处理工具的情况下, 在多种自然语言处理任务能够取得更优良的性能。近几年来, 神经网络模型已经成为了关系抽取任务中的重要方法。

Socher 等<sup>[1]</sup>采用递归神经网络模型, 使用矩阵向量表征分析树的各个节点, 从而提取句子中重要的语义信息, 但是同时因为借助了自然语言处理工具, 容易引起错误传播问题。为了更加准确地捕捉句子的语义, Zhang 和 Wang<sup>[2]</sup>在2015年采用双向循环网络(BRNN)

来对句子进行建模, 学习其中的语义信息。双向的RNN可以从前向和后向获取句子的信息, 但是由于RNN存在梯度消失和梯度爆炸问题, 胡新辰等<sup>[3]</sup>采用双向长短期记忆网络(BiLSTM)来实现对句子特征的提取, 同时又借助传统的词汇与句子特征来进行关系抽取。罗计根等<sup>[4]</sup>提出一种融合梯度提升树的BiLSTM模型, 在中医领域上实现了关系抽取, 并取得相当不错的效果。近期注意力机制<sup>[5]</sup>在各大领域上有了比较成功的应用。Zhou 等<sup>[6]</sup>在2016年就在双向长短期记忆网络(BiLSTM)的基础上加入注意力机制, 并且在关系抽取任务上取得了更优良的性能。周文焯等<sup>[7]</sup>则是融合了多层注意力机制与双向长短期记忆网络进行语义关系抽取, 取得一定的效果。李卫疆等<sup>[8]</sup>在ATT-BiLSTM模型基础上引入多种外部特征来进行关系分类。朱姗姗和唐慧丰<sup>[9]</sup>则是把ATT-BiLSTM模型应用到军事领域上, 同时结合词性特征, 在军事类实体关系上取得较好的F1值。

卷积神经网络在关系抽取上也有长足的应用。Zeng<sup>[10]</sup>采用卷积神经网络模型, 再结合词向量以及位置特征, 此算法无需提取复杂的特征就取得超越传统方法的性能。随后, Zeng 等<sup>[11]</sup>又基于实体位置信息, 提出了分段池化的思想, 根据实体位置将池化的向量分为3段, 然后各取这3段的最大值进行拼接。dos Santos 等<sup>[12]</sup>提出分段排序结构来取代Softmax分类层器, 它有效地提高了不同实体类型的区分度, 并且取得了相当好的效果。Mou 等<sup>[13]</sup>和刘伟等<sup>[14]</sup>则利用树形结构的神经网络进行关系抽取并且取得很好的效果。Lin 等<sup>[15]</sup>改进卷积神经网络模型, 在此基础上加入了注意力机制, 解决了实体相对应的句子噪音问题, 该模型可以使用有效的句子学习, 因而使得关系抽取性能进一步提高。Bing 等<sup>[16]</sup>通过深度卷积神经网络模型提取句子的低级和高级特征, 在NYT数据集上取得较高的F1值。Liu 等<sup>[17]</sup>利用循环神经网络对依存句法树进行建模, 然后又利用卷积网络对语句的最短依存路径进行建模, 统一将两者的语法分析结果添加到深度学习框架中。张晓斌等<sup>[18]</sup>则是利用卷积网络和循环网络两者各自的优点, 弥补相应的不足, 在此基础上提出了一种结合CNN和LSTM的实体关系抽取算法, 该算法取得了一定的效果。李孟颖等<sup>[19]</sup>利用CNN-BiLSTM联合模型在微生物生长环境领域上进行关系抽取, 并取得了相当不错的效果。Li 等<sup>[20]</sup>提出一种预训练CNN模型来探索语义关系抽

取并实现了较高的性能。

综上所述,前两者方法严重依赖人工特征,而构建这些特征耗时耗力;对于现有的神经网络模型而言,根据 Tang 等<sup>[21]</sup>的研究表明,相较于 Transformer,神经网络的综合特征提取能力有限以及语义表征能力较差,因而导致整个网络模型无法获取丰富的语义信息特征,因此在实体关系抽取上已经很难再有所突破性进展。而近期的 BERT (Bidirectional Encoder Representations from Transformers)<sup>[22]</sup>模型在 NLP 领域取得重大突破,凭借其独特的网络结构在多个研究方向得到成功应用,同时在关系抽取上也存在相应的应用。Wu 等<sup>[23]</sup>和 Giorgio 等<sup>[24]</sup>利用 BERT 模型实现过英文关系抽取,并取得相当不错的效果,甚至优于当前的深度学习网络模型。因此本文就引入 BERT 模型来进行中文实体关系抽取。本文的主要贡献如下:

(1) 在中文实体关系抽取任务上,大多数算法主要集中在循环神经网络模型和卷积神经网络模型等。而本文所采用的模型不同于之前的神经网络模型,本文则是使用了 BERT 预训练模型来实现中文关系抽取任务;

(2) 在进行中文实体关系抽取时添加了多种外部特征:关键词、实体类型和实体对信息。这些特征能够在词汇和句法上体现出整个语句一定的语义信息,从而利用这种优势进一步来加强 BERT 模型的语义特征提取能力,因此可以进一步提高中文实体关系抽取性能。

## 2 融合多特征 BERT 模型的关系抽取

本文是基于 BERT 预训练模型来实现中文关系抽取任务的。其整个训练过程如图 1 所示,主要工作如下:

① 语料预处理,构造语料特征(本文采用的外部特征有关键词、实体类型和实体对信息);

② 建立相关的网络模型(本文采用的是 BERT 模型),其输入是①中所处理的数据,并获取其相关的分布式词向量特征;

③ 将得到的各种外部特征向量进行拼接融合,通过全连接层得到最终的向量表征;

④ 采用 Softmax 分类器,对最后的实验结果进行关系分类。

本文的模型最大优势就在于不需要提前进行额外的字向量或是词向量训练,只要把语句直接输入到 BERT 预训练模型中即可,该模型会自动将训练出带有较丰富语义的特征向量,然后直接利用该特征向量进行关系抽取。

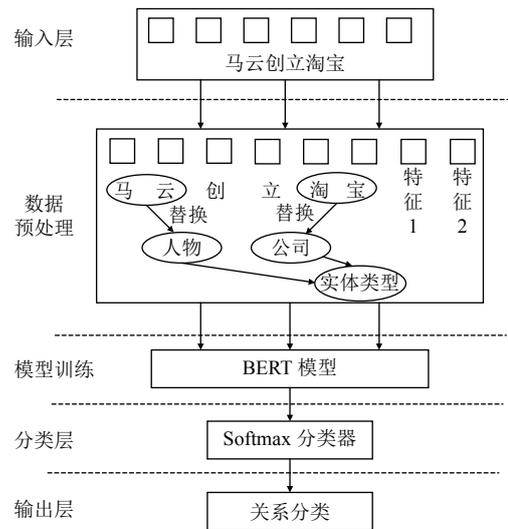


图 1 基于 BERT 模型的中文实体关系抽取流程

### 2.1 外部特征的提取融合

#### 2.1.1 数据预处理

数据预处理的主要目的是获取高质量的训练数据,其过程主要包括两个方面:数据的整理标注以及外部特征的提取融合。数据的整理标注主要是处理成模型训练所需要的语料格式,而外部特征的提取融合则是整个数据预处理的重点所在,下面依次对外部特征的提取和融合进行细致分析。

#### 2.1.2 外部特征的提取

在中文实体关系抽取当中,不同于英文实体关系抽取,中文实体关系抽取具有不同的语法结构,尤其是在特定的领域中,中文更是具有独特的语法结构,因此为进一步改善中文实体关系抽取的性能,本文将选取关键词、实体类型和实体对信息作为外部特征。

##### (1) 关键词

在进行关系抽取这个分类任务时,引入关键词对提高抽取性能具有一定辅助作用。提取出来的关键词大都是有很高的区分度,甚至可以增强语义信息,因此关键词特征与抽取的关系相关性越高,那么对关系抽取的性能提高就越大。关键词作为一个显式特征,它可以从另一个方面能够弥补 BERT 模型训练得到的特征的不足之处。例如,“天安门是北京市的一处旅游景点”,通过提取关键词可以得到“旅游景点”,由此可以预测出关系可能是“地理位置关系”。因此提取一些类别关键词更有利于关系抽取。对于任一个语句  $X$  而言,提取其中的关键词后输入到 BERT 模型中进行向量表征。如果  $k_i$  表示为第  $i$  个关键词的特征向量,那么所有关键

词的特征向量则为  $K = [k_1, \dots, k_n]$ .

(2) 实体类型

实体类型是作为一种浅层的语义特征,它主要为实体标注一些特定的语义角色.实体类型通常在于浅层语义的理解,而对于语句的深层语义特征不进行分析.在实际应用过程中,浅层语义特征往往也显得非常重要,甚至能学习到更多的语义信息.对于关系两边的实体类型,经常是实体对的约束条件.例如,指定一个关系“出生地”,则可以知道主体的实体类型属于“人物”,而客体的实体类型是“地点”.在此模型中,本文把语料中的候选实体替换为相应的实体类型,然后再将整个语句输入到BERT模型中进行学习,从而提取语义特征.

(3) 实体对信息

对于限定关系抽取而言,实体对是作为一种约束信息.例如给定“马云”和“淘宝”,通过模型训练后,可以大概率将“出生地”或是“所在城市”等一些与之相关性不高的关系排除掉.因此它在一定程度上是可以得到相应的语义信息,从而能够帮助判断实体关系类型.将实体对信息输入到模型训练中,可以得到其特征向量:  $T = [t_1, t_2]$ ,其中  $t$  表示实体向量.

2.1.3 特征融合

在实现对外部特征的提取后,需要将所提取的特征进一步地拼接融合,使得外部特征能够显示地表达出来,从而利用其所表现出来的语义特征来强化模型训练,使模型输出特征向量能够尽可能地表达出更多的语义信息,进而提升中文实体关系抽取效果.具体特征融合过程如图2所示.

对于任意语句  $X = [x_1, x_2, \dots, x_v]$ ,在数据处理阶段将其中的候选实体用相应的实体类型进行替换后,得到一个新的语句表示:

$$X' = [x'_1, x'_2, \dots, x'_v] \quad (1)$$

在  $X'$  当中,其可以理解为把实体类型添加到语句当中,通过替换实则是将不同实体之间的关系转为两种实体类型之间的关系,同时实体类型本身就带有一定的语义信息,这两者的共同作用降低了噪声的干扰.随后将  $X'$  输入至模型训练,提取语句的语义特征,经过训练后其语句的向量表征形式如下公式:

$$E'_1 = BERT(\{X'\}) \quad (2)$$

其次通过自然语言处理工具对整个语句  $X$  的关键

词进行提取.关键词是对语句内容的一个整体性概括,关键词排序越靠前的对语句的影响力就越大.关键词最主要的作用就是提取语句中语义较丰富的特征词,进而降低无关的信息对关系抽取的影响,因此把提取出来的关键词按序进行排列.关键词排序靠前的优先作为外部特征的输入.同时为了尽可能地丰富句子的语义,本文选取排序前20%的关键词,而后续的关键词已经很难体现句子的语义信息,甚至可能会带来额外的噪声影响,从而使得关系抽取性能变差.在此用  $g_i$  表示提取的第  $i$  个关键词,经过训练后的关键词的向量表征形式如下:

$$K = BERT(g_1, g_2, \dots) \quad (3)$$

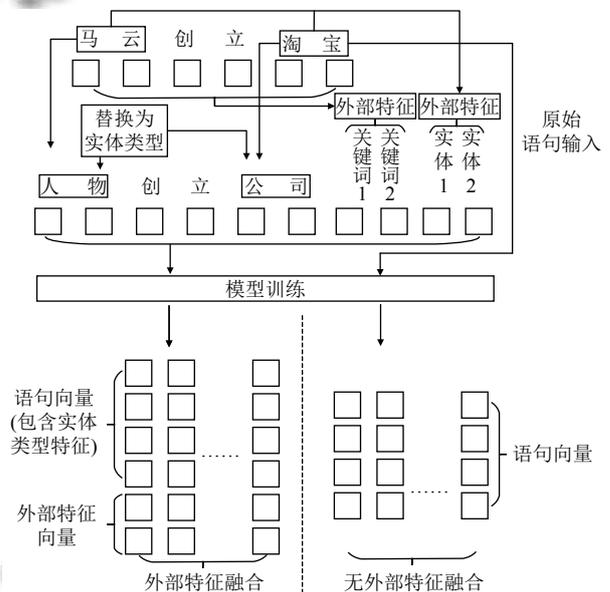


图2 特征融合

对于实体对信息而言,实体对是作为一种对关系的约束信息,同时能够间接反映出实体所在的位置信息,对提高关系抽取性能很有帮助.而对于关系抽取而言,能够利用各种信息加强对语义的理解都是可以作为一种有效的外部特征进行输入.在数据的处理过程中,实体对分为头实体和尾实体,在此分别用  $s$  和  $o$  来表示,那么训练后的实体特征向量表示为如下:

$$T = BERT(s, o) \quad (4)$$

三者训练过后得到了相应的特征向量,紧接着就是将三者的特征融合在一起.特征融合主要有两种方式:特征拼接和加权求和.两种融合方式具体如图3所示.

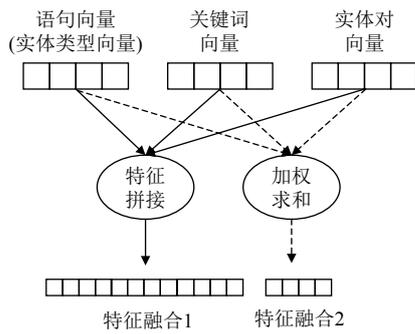


图3 特征融合方式对比

从图3可以看出,两种特征融合方式都是可行的,但两种方式各有所长.对于加权求和的特征融合方式,其优点在将三者向量的一些关键信息融入到一起,降低了向量的维度,便于计算,但是其缺点是在加重的过程中容易造成某些关键的语义信息的缺失.而对于特征拼接而言,它把所有的向量都拼接在同一个向量中,虽然使得向量维度提高了,但它保留了最原始的语义信息特征,尤其对于关系抽取任务而言,它更加注重对整个语句的理解程度,因此从理论上来说,尽可能地保留更多的语义信息是有助于提高关系抽取的性能.因此本文就采用特征拼接的融合方式,而非所有特征加权求和,所以外部特征融合的向量表示如下:

$$H = E_1' \oplus K \oplus T \tag{5}$$

图2中的右边展示的是无任何外部特征的加入,其训练出来的语句向量表示为如下:

$$E_1 = BERT(X) \tag{6}$$

从上述可以看出特征向量  $H$  主要包括两种类型的向量:语句向量  $E_1'$  和外部特征向量,前者又含有实体类型特征,后者则是包含关键词和实体对信息的特征向量  $K$ 、 $T$ . 整体而言,本文将所选外部特征融合到一个向量表征中.而图中右边表示的是把原始语料输入到模型之中进行训练,得到只是一个单纯的语句向量  $E_1$ . 因此两者对比分析可以得到,本文输出向量融合了外部特征向量,弥补了自动学习特征的不足,因而能够表征出更多的语义信息.

### 2.2 基于 BERT 模型的词向量表征

在处理自然语言过程中,词向量是一种相当重要的表征形式.目前而言,先进的词向量是一种分布式的表征形式,它是通过用一个低维的实数向量来表征一个词语.这样的词向量包含丰富的句法和语义信息特征.词向量是通过对大量的无标注文档训练而得到的.

在本文中,采用 BERT 模型来训练词向量,不同于以往的静态词向量,前者能够在不同的语境下表征不同的语义,因此这样的词向量更加体现出强大的语义表征能力.其训练过程主要分两步进行的:一是对输入语句的预处理,二是通过基于双向 Transformer 的网络结构进行训练.

本文词向量的具体训练工作如下:

给定一个句子  $X = [x_1, x_2, \dots, x_v]$ , 则可以表示为一个向量矩阵  $E = R^{V \times d_m}$ , 其中  $V$  是句子中词语的个数,  $d_m$  是词向量维度.

该句子  $X$  经过数据预处理,将所得到的特征联合拼接得到语句  $X'$ ,  $X'$  就作为 BERT 模型的输入进行预训练,其中句向量和外部特征向量的表征过程如图4所示.

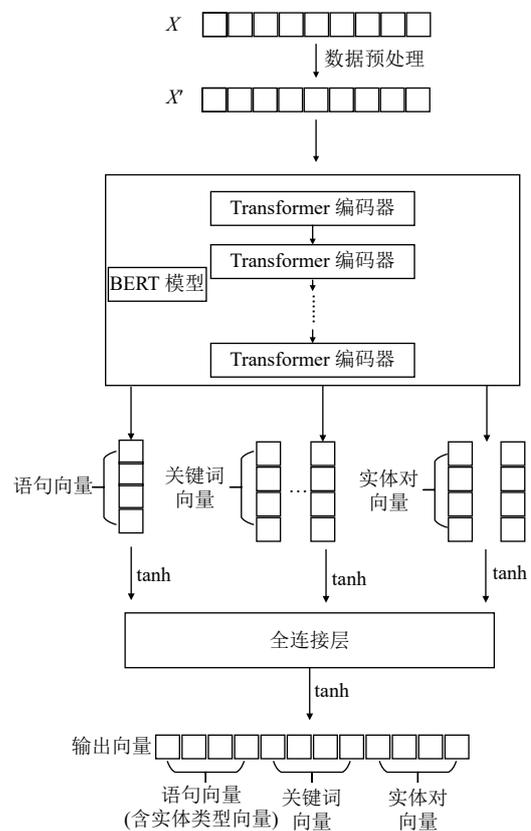


图4 向量表征

从图4来看, BERT 模型的整体框架是由 12 层 Transformer 搭建而成,但是在具体的训练过程需要按照实际的需求来确定 Transformer 层数,以此来达到最佳的训练效果.从理论上分析,数据量较大时,Transformer 层数可以选取较大的值,因此在训练过程中可以获取更加深层次的语义信息;数据量较小时,Transformer 层

数可以选取较小的值,因为数据量小,即使设置选取较大的值也很难获取其中深层次的语义信息,反而容易导致“过学习”的状态,使得训练无法收敛,从而导致关系抽取的性能下降.从本文所选取的数据以及后期的实验结果来看,Transformer层数应该设置2或3是最为合适的.

在经过BERT模型的预训练后,得到句子向量 $E'$ 、关键词向量 $K$ 以及实体对向量 $T$ ,随后通过激活函数,其具体表示如下:

$$E'_1 = W_1 \tanh(E') + b_1 \quad (7)$$

$$K' = W_2 \tanh(K) + b_2 \quad (8)$$

$$T' = W_3 \tanh(T) + b_3 \quad (9)$$

紧接着通过全连接层将所有的向量拼接在一起,其表示如下:

$$H' = \text{concat}(E'_1, K', T') \quad (10)$$

最后经由激活函数激活得到最终的输出向量:

$$H = W_4 \tanh(H') + b_4 \quad (11)$$

其中, $W$ 是权重矩阵, $b$ 是偏置向量.实验过程中融合了多种外部特征,使得训练出来的语句表征向量含有实体类型特征、关键词特征以及实体对信息特征,而这些外部特征能够更加便于整个模型理解句子的语义,从而更有利于解决关系抽取问题.

### 2.3 关系分类

在得到BERT预训练模型输出的组合特征向量后,将其输入到全连接层,获得一个含有丰富语义的向量 $H$ ,随后开始进行关系分类.本文采用的是Softmax分类器.该分类器的主要作用是对最后的结果进行归一化处理,从而可以得到一组概率值,通过判断概率的大小来进行关系分类.向量 $H$ 通过Softmax分类器第 $i$ 种关系的条件概率如下:

$$P(r|X) = \frac{\exp(h_i)}{\sum_j \exp(h_j)} \quad (12)$$

在训练过程中目标函数采用的是交叉熵损失函数.交叉熵是用来衡量多种分布之间的相似性,因此也可以判断网络结构实际的输出与期望的输出的接近程度.其目标函数 $loss$ 计算如下:

$$loss = -\frac{1}{n} \sum_M [y \ln y' + (1-y) \ln(1-y')] \quad (13)$$

其中, $y, y'$ 分别表示为真实值和预测值, $M$ 为样本, $n$ 为

样本数量.模型优化过程中采用随机梯度下降法,模型参数的更新则采用反向传播的方法进行调整,计算过程如下:

$$\theta \leftarrow \theta + \lambda \frac{\partial J(\theta)}{\partial \theta} \quad (14)$$

其中, $\lambda$ 表示学习速率.

## 3 实验

本文主要采用人工标注的中文数据集进行了相应的实验,实验过程主要分为以下两组:

(1) 在BERT模型上依次添加外部特征,验证外部特征是否能够有效地强化BERT模型的学习能力;

(2) 对比基于ATT-BiLSTM和CNN模型的经典方法,验证BERT模型是否能够达到预期效果.

### 3.1 数据集与衡量指标

本文采用的数据集是来源浙大知识引擎实验室人工标注的中文数据集.该数据集主要分为训练集和测试集,其中训练数据集总共包含4000条语句实例,测试集总共包含1000条语句实例.该数据集的关系类型主要限定于以下10种,具体如表1所示,部分语料样本如图5所示.

表1 候选实体关系类型

编号	关系名称	编号	关系名称
1	国籍	6	所在城市
2	祖籍	7	所属专辑
3	导演	8	连载网站
4	出生地	9	出品公司
5	主持人	10	毕业院校

舒歌平,男,中国神华煤制油有限公司主任工程师,国籍,舒歌平,人物,0,中国,国家,6  
《血战异世》是所调节写的网络小说连载于17k小说网,连载网站,血战异世,网络小说,1,17k小说网,网站,19  
丘顺卿是泉州府人,为元朝驸马,官武略大将军,出生地,丘顺卿,人物,0,泉州府,地点,4  
明朝末年抗清英雄黄得功,本姓王,安徽合肥人后改姓黄,出生地,黄得功,人物,8,安徽合肥,地点,16  
于玮,毕业于四川音乐学院,魏晨的大学同学,中国四川人,毕业院校,于玮,人物,0,四川音乐学院,学校,6

图5 部分语料展示

对于实验结果评测,本文采用官方数据集中的macro-averaged  $F1$  进行评价模型性能. macro-averaged  $F1$  的计算具体过程如下所示:

$$P_i = \frac{\text{正确分类的关系总数}}{\text{抽取第}i\text{类的所有关系总数}} \quad (15)$$

$$R_i = \frac{\text{正确分类的关系总数}}{\text{标准结果第}i\text{类的所有关系总数}} \quad (16)$$

$$F1_i = \frac{2P_i R_i}{P_i + R_i} \quad (17)$$

因此得到 macro-averaged  $F1$  值的公式如下所示:

$$\text{macro-averaged } F1 = \frac{1}{N} \sum_{n=1}^N F1_n \quad (18)$$

### 3.2 实验参数设置

实验超参数设置主要是为了使模型达到最佳性能. 实验过程中为防止模型出现过拟合现象, 训练过程中引入了 dropout 策略. 此外, 模型采用 Adam 优化器进行优化. 本文实验在参数设置过程中发现 Learn\_rate 与 Num\_hidden\_layers 这两个参数对实验影响是最大的. 为取得实验最佳效果, 多次对这两个参数进行微调, 最终确定以下结果. 具体超参数设置如表 2 所示.

表 2 超参数设置

实验参数	参数值
Learn_rate	1e-4
Decay_rate	0.5
Batch_size	64
Dropout	0.3
Num_hidden_layers	2

### 3.3 实验效果与分析

本文为验证所选特征的有效性以及对比较神经网络模型方法, 主要做了以下两组实验, 其中 KW、EI、ET 分别代表关键词、实体对信息和实体类型. 衡量指标采用 macro-average  $F1$  进行评价 (以下的  $F1$  均表示为 macro-average  $F1$  值).

#### (1) 实验一. 验证特征的有效性

本文模型为提取更多的语义特征, 引入了关键词 (KW)、实体对信息 (EI) 和实体类型 (ET) 三种特征. 本文在中文数据集上做了相应的实验来验证这 3 种特征的有效性. 实验结果如图 6 和表 3 所示.

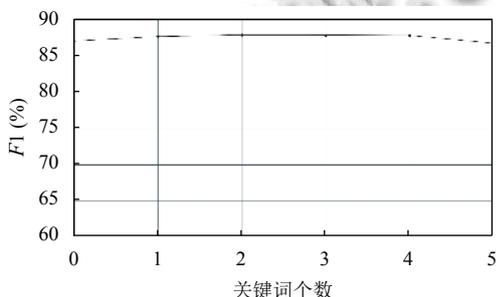


图 6 关键词个数的影响

图 6 展示了关键词个数对模型性能的影响. 从图中可以看出, 关键词个数是对模型性能有影响的. 起初, 关键词个数递增,  $F1$  也随之上升, 但是到达一定程度

后, 伴随着关键词个数的增加,  $F1$  反而逐渐下降. 在整个实验过程中, 当关键词个数为 2 至 4 个, 实验效果表现最佳. 从这可以看出, 选取适量的关键词, 尤其是能够体现出句子语义信息的关键词, 对模型性能是有提升的; 但是当选取的关键词增多时, 由于其中开始出现一些不相关的词汇, 而它们将会给关系抽取任务带来额外的噪声影响, 从而使得模型性能下降.

表 3 特征对 BERT 模型的影响 (%)

特征	$F1$
None	87.05
+KW	87.87
+KW、EI	93.22
+KW、EI、ET	97.50

本文选取无任何外部特征加入的 BERT 模型作为基准模型. 然后在该模型基础上依次加入外部特征. 从上述表 3 数据结果可以明显地看出, 所选取的这三种外部特征是能够帮助关系分类的. 其中, 加入关键词后,  $F1$  提升了 0.82%; 再加入实体对信息, 由于实体对信息对所属关系具有一定的约束性, 同时其能够间接体现出实体的位置特征, 因而使得模型性能提高了 5.35%; 最后加入实体类型, 其是属于浅层语法信息, 能够提取较为丰富的语义信息, 因此模型性能得到了 4.28% 的提升. 从上述分析可以得到, 所选的外部特征能够强化 BERT 模型的学习能力. 实体对信息和实体类型特征对性能提升效果明显, 而关键词相较次之.

#### (2) 实验二. 与其他模型的对比实验

为了对比不同方法在关系抽取任务上的实验效果, 而 ATT-BiLSTM 和 CNN 模型常用于在关系抽取中, 词向量特征 (WV) 与位置特征 (PF) 是作为神经网络模型的固定特征. 因此本文主要利用了这两种经典的神经网络模型和本文模型在同一数据上进行测试比较. 测试结果如表 4 和图 7 所示.

表 4 不同模型的实验对比 (%)

模型+特征	$F1$
ATT-BiLSTM	74.65
CNN	85.35
<b>BERT</b>	<b>87.05</b>
ATT-BiLSTM+WV、PF	84.83
CNN+WV、PF	92.13
<b>本文</b>	<b>97.50</b>

通过表 4 可分析:

(1) ATT-BiLSTM 模型是一种最经典的关系抽取

模型,其优势主要在于其双向结构,能够获取上下文信息,获取整个语句的全局序列信息,因此在处理长距离依赖问题具有其他神经网络无可比拟的优势,同时结合注意力机制,又融合词向量与位置特征,在该数据集上取得 84.83% 的  $F1$  值.然而该模型对于局部特征的重点语义信息很难捕捉到,且提取特征能力相对较差(相较于 BERT 的 Transformer 结构),因此在中文关系抽取上该模型性能并不占优.

(2) CNN 模型相较于 ATT-BiLSTM 模型,它更善于提取句子的局部语义特征,同时结合词向量与位置特征在本文中文数据集上达到了 92.13% 的  $F1$  值,实验结果表明,相对于 ATT-BiLSTM 模型而言,CNN 模型在关系抽取任务效果更好.但是 CNN 过于注重局部信息而忽视了整体全局特征信息的重要性,因此 CNN 模型在全局特征的应用和表征是有所欠缺的.

(3) BERT 预训练模型不同于 BiLSTM 和 CNN 神经网络,BERT 模型采用一种全新的 Transformer 结构,它是基于注意力机制构建的网络,该网络把关注的重点放在每个句子中任意两个词之间的关联程度,这样的处理既可以获取整个全局信息,同时又重点考虑了局部特征,可以说该模型在很大程度上综合这两种神经网络模型的优势.为了对比这 3 种模型的性能,在不加任何外部特征的情况下对比其实验效果.正如表 4 所示,BERT 模型取得 87.05% 的  $F1$  值,远高于 ATT-BiLSTM 模型和 CNN 模型.因此也证实了 BERT 模型在特征提取能力强于这两者经典神经网络模型.

对于关系抽取而言,该问题的解决在很大程度上在于模型对整个文本的理解程度.因此为了使得模型更好地学习到数据的语义特征,本文在 BERT 模型的基础上融合了多种外部特征,减少了一些关键信息特征的丢失,从而让模型更利于关系抽取.从实验结果来看,加入外部特征是有效可行的,这也表明了融合外部特征是可以进一步加强 BERT 模型的语义学习能力.与其他模型相比较,本文提出的模型具有相当明显的优势.

图 7 展示的是上述模型在前 20 轮训练的  $F1$  更新状况.从图中可以看出,在训练初期时,两种神经网络就处于一个比较低的水平,只有在多次迭代训练后,性能才不断往上升;而对于 BERT 模型,一开始就保持相当高的水平,并且能够稳定保持,同时也能看到本文模型的  $F1$  值始终处于 BiLSTM 和 CNN 神经网络模型

之上,这就意味着 BERT 模型的性能稳定性优于 ATT-BiLSTM 模型和 CNN 模型.

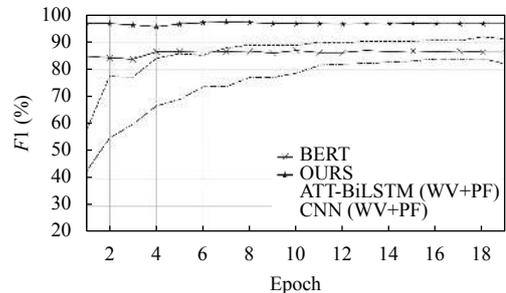


图 7  $F1$  的更新情况

## 4 结束语

本文主要探讨了 BERT 预训练模型在中文实体关系抽取的应用. BERT 模型通过它特有的结构能够更好地学习到有效的语义特征,并在此基础上融合了关键词、实体信息以及实体类型三大外部特征,弥补了自动提取特征的不足,又最大限度地降低了关键信息的损失,因此使模型学习到更多额外的语义特征,从而提升了关系抽取的性能.本文对数据稀疏和不均衡问题没有做过多的研究和分析,因此下一步需研究数据稀疏和不均衡问题对该模型的影响.

## 参考文献

- 1 Socher R, Huval B, Manning CD, et al. Semantic compositionality through recursive matrix-vector spaces. 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.
- 2 Zhang DX, Wang D. Relation classification via recurrent neural network. arXiv preprint arXiv: 1508.01006, 2015.
- 3 胡新辰. 基于 LSTM 的语义关系分类研究 [硕士学位论文]. 哈尔滨: 哈尔滨工业大学, 2015.
- 4 罗计根, 杜建强, 聂斌, 等. 基于双向 LSTM 和 GBDT 的中医文本关系抽取模型. 计算机应用研究, 2019, 36(12): 3744-3747.
- 5 Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. arXiv preprint arXiv: 1706.03762, 2017.
- 6 Zhou P, Shi W, Tian J, et al. Attention-based bidirectional long short-term memory networks for relation classification. Proceeding of the 54th Annual Meeting of the Association

- for Computational Linguistics. Berlin, Germany. 2016. 207–212.
- 7 周文辉, 刘亮亮, 张再跃. 融合多层注意力机制与双向 LSTM 的语义关系抽取. 软件导刊, 2019, 18(7): 10–14, 18.
- 8 李卫疆, 李涛, 漆芳. 基于多特征自注意力 BLSTM 的中文实体关系抽取. 中文信息学报, 2019, 33(10): 47–56, 72.
- 9 朱珊珊, 唐慧丰. 基于 BiLSTM\_Att 的军事领域实体关系抽取研究. 智能计算机与应用, 2019, 9(4): 96–99. [doi: 10.3969/j.issn.2095-2163.2019.04.021]
- 10 Zeng DJ, Liu K, Lai SW, *et al.* Relation classification via convolutional deep neural network. Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. Dublin, Ireland. 2014. 2335–2344.
- 11 Zeng DJ, Liu K, Chen YB, *et al.* Distant supervision for relation extraction via piecewise convolutional neural networks. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal. 2015. 1753–1762.
- 12 dos Santos CN, Xiang B, Zhou BW. Classifying relations by ranking with convolutional neural networks. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Beijing, China. 2015. 626–634.
- 13 Mou LL, Li G, Zhang L, *et al.* Convolutional neural networks over tree structures for programming language processing. Processing of the 30th AAAI Conference on Artificial Intelligence. Phoenix, Arizona, USA. 2016. 1287–1293.
- 14 刘伟, 陈鸿昶, 黄瑞阳. 基于 Tree-based CNN 的关系抽取. 中文信息学报, 2018, 32(11): 34–40. [doi: 10.3969/j.issn.1003-0077.2018.11.005]
- 15 Lin YK, Shen SQ, Liu ZY, *et al.* Neural relation extraction with selective attention over instances. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany. 2016. 2124–2133.
- 16 Bing XY, Shen LD, Zheng LY. A moderately deep convolutional neural network for relation extraction. Proceedings of the 2019 11th International Conference on Machine Learning and Computing. Zhuhai, China. 2019. 173–177. [doi: 10.1145/3318299.3318326]
- 17 Liu Y, Wei FR, Li SJ, *et al.* A dependency-based neural network for relation classification. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Beijing, China. 2015. 285–290.
- 18 张晓斌, 陈福才, 黄瑞阳. 基于 CNN 和双向 LSTM 融合的实体关系抽取. 网络与信息安全学报, 2018, 4(9): 44–51.
- 19 李孟颖, 王健, 王琰, 等. 基于融合式神经网络的微生物生长环境关系抽取. 模式识别与人工智能, 2019, 32(2): 177–183.
- 20 Li Q, Li LL, Wang WN, *et al.* A comprehensive exploration of semantic relation extraction via pre-trained CNNs. Knowledge-Based Systems, 2020, 194: 105488. [doi: 10.1016/j.knosys.2020.105488]
- 21 Tang GB, Müller M, Rios A, *et al.* Why self-attention? A targeted evaluation of neural machine translation architectures. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium. 2018. 4263–4272.
- 22 Devlin J, Chang MW, Lee K, *et al.* BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv: 1810.04805, 2018.
- 23 Wu SC, He YF. Enriching pre-trained language model with entity information for relation classification. Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing, China. 2019. 2361–2364.
- 24 Giorgio J, Wang XD, Sahar N, *et al.* End-to-end named entity recognition and relation extraction using pre-trained language models. arXiv preprint arXiv: 1912.13415, 2019.