DOI:10.13718/j.cnki.xsxb.2021.11.008

基于最短依存路径和 BERT 的关系抽取算法研究①

陈 珂, 陈振彬

广东石油化工学院 计算机学院,广东 茂名 525000

摘要:深度学习模型依靠文本单一的词特征、位置特征在文本关系抽取任务中取得了不错的效果.但以往研究未能充分理解句子语义,数据稀疏和噪声传播问题对分类模型的影响依旧存在.随着注意力机制和预训练语言模型的研究不断深入,BERT(bidirectional encoder representations from transformers)预训练模型为自然语言处理任务提供了更好的词句表示.因此,该文提出结合 BERT 预训练语言模型获得更具语义表现力的特征表示,同时使用依存句法分析提取出最短依存路径作为额外信息输入分类模型,降低了噪声词汇对分类模型的影响.该算法在中文人物关系抽取数据集和 SemEval2010 Task 8 语料集上进行对比实验,最终实验效果 F 值可达到 0.865.

关 键 词:关系抽取;依存句法分析;最短依存路径;BERT

中图分类号:TP391.1 **文献标志码**:A **文章编号**:1000-5471(2021)11-0056-11

Entity Relation Extraction Based on Shortest Dependency Path and BERT

CHEN Ke, CHEN Zhenbin

College of Computer, Guangdong University of Petrochemical Technology, Maoming Guangdong 525000, China

Abstract: The deep learning model relies on the single word feature and position feature of the text to achieve good results in the task of text relation extraction. However, in the existing research results, sentence semantics are not fully understood, and the impact of data sparsity and noise propagation on the classification model is still serious. With the development of attention mechanism and pretraining language model, Bert (bidirectional encoder representations from transformers) pretraining model provides a better way to express words and sentences for natural language processing tasks. On the other hand, dependency parsing is used to extract the shortest dependency path as additional information input to the classification model to reduce the impact of noisy words on the classification model. The method of this paper is to conduct a comparative experiment on Chinese character relationship extraction dataset and semeval2010 task 8 corpus, and the final experimental effect F value can reach 0.865.

Key words: relation extraction; dependency parsing; shortest dependency path; BERT

近年来人工智能技术发展迅速,智能算法已广泛应用于各领域,其中智能技术的实现多依赖于大规模、高质量和宽领域的结构化知识库[1-2].传统的知识库构建主要依赖手工,通过该方式构建的知识库扩

① 收稿日期:2020-08-04

基金项目:国家自然科学基金项目(61172145);广东省自然科学基金项目(2018A030307032);广东省普通高校重点科研平台和项目(2020ZDZX3038).

作者简介:陈 珂,教授,硕士,主要从事自然语言处理研究.

展性较差、规模较小且具有局限性，因此，如何自动化构建知识库成为近年的研究热点. 从大量非结构化数据中抽取出结构化数据，成为构建大型知识库的关键技术之一. 从自然文本中自动提取出多个实体并判别其关系类型是关系抽取任务的目的. 目前，已有关系抽取的方法大致可归纳为：基于模式匹配的关系抽取方法、基于词典的关系抽取方法、基于文本语法和语义的关系抽取方法、基于机器学习的关系抽取方法及混合抽取方法. 目前，基于机器学习的关系抽取方法的核心思想是使用表示学习等方法，组织和衍生特征向量，通过提取自然语言文本中的特征，组织成可被深度神经网络或者其他学习网络接受的张量形式进行分类器的训练. 特征的组织和模型的优化是关系抽取方法的关键步骤，也是影响分类准确率的重要因素.

在以往研究中，普遍采用基于 Skip-gram 模型和 CBOW 模型的单一字向量和词向量作为文本特征，再结合具体任务训练语料的特点，构建特定任务的概率模型. 这种方法虽然效果不错，但噪声传播问题仍是关系抽取任务要解决的难点. 通过构建特定任务的模型能在一定程度上解决该问题，但局限性成为进一步提升抽取效果的瓶颈. 另一方面，传统的预训练语言模型，虽然能在一定程度上反映文本字词语义，但其表示能力受滑动窗口的限制，并不能充分表示上下文语义，存在一词多义现象. 在特定语境下，传统的字词特征仍有改进空间.

依存句法分析是自然语言处理领域的分析方法之一. 基于转移和基于图的依存分析方法是依存句法分析的两个主要思路. 基于转移的依存分析方法是构建一条从初始转移状态到终结状态的转移动作序列并逐步生成依存树；基于图的依存分析方法则将文本序列转换为有向完全图，在图中求解最大生成树问题. 通过依存句法分析能简洁反映文本实体词之间的直接或间接的关联关系，以降低噪声词对训练分类模型的影响，更好地解决噪声传播问题；同时，依存句法分析能衍生更丰富的实体和语法特征，能更好地挖掘文本语义.

目前注意力机制和语言模型的不断发展为自然语言处理任务提供了更好语义表示方案. 基于 Self-Attention 机制的 Transfomers 被提出后，以其为基础的 BERT(bidirectional encoder representations from Transformers)预训练模型也应运而生. BERT 能更好地综合考虑文本的上下语境，增强了模型的泛化能力，充分描述了字符级、词级、句子级甚至句间关系的特征，在自然语言处理领域引起了重大的反响，在 GLUE(多种英语语言理解任务的集合，包括文本蕴涵、情感分析和语法判断等)任务中取得不错的分数，充分证明 BERT 强大的语义表示能力. 本研究提出了一种基于 BERT 模型并使用最短依存路径特征的文本实体关系抽取模型. 最短依存路径特征从句法结构的角度筛选出对于句中实体有较大意义的信息；BERT 模型能更充分考虑文本上下文语境，在句子语义表示方面更加优秀. 本文主要研究内容如下：

(1) 使用句法依存分析获得依存句法树，对依存句法树进行剪枝等处理并获得实体间的最短路径. 对路径上特定词性的特征词赋予较高权重值，使文本特征更具区分性，并且降低噪声词的影响.

(2) 利用 BERT 模型对文本进行特征提取和特征表示作为下游 NLP 任务输入的部分输入特征，经过微调后的 BERT 模型将更适用于当前语料并具备更好的语义信息.

(3) 在下游任务的分类模型 LSTM 上加入注意力机制，使得模型训练过程中能更好地注意到重要的特征词，更好地提升关系抽取模型的分类能力.

1　相关工作

在已有的显示关系抽取研究中，基于特征向量的关系抽取方法占绝大多数. 但该方法非常依赖文本特征的提取，所提取的特征质量将直接影响最终的抽取效果，因此基于特征向量的关系抽取的关键在于文本提取和组织有用的语义信息.

丰富的语义和语法特征可以更好地提升关系抽取任务的分类效果. Jiang 等[3]利用统一特征空间对不同特征及其对关系抽取效果的影响进行研究，实践证明通过组合基本特征能有效提高关系抽取的效果；奚斌等[4]通过对词法、语法和语义等特征进行多种组合，同样也证明了组合特征能有效提高关系抽取性能；Zhang 等[5]则通过利用实体之间的多种位置关系并将其特征融合到特征集中，在 ACE2005 公开数据集中进行实验，证明该特征更好地改善了关系抽取的效果.

但基于特征向量的关系抽取方法中大多仅仅考虑文本词特征，句子的句法和语义特征并没有被更好地挖掘. 已有的研究表明，动词对于文本语义的理解有较大帮助，提升了模型效果. 甘丽新等[6]围绕句中动

词提出“最近动词依赖特征”，并使用依存句法分析进行关系抽取任务；李明耀等[7]在依存分析基础上把动词分为3类：动词作谓语、动词短语作谓语、复杂动词作谓语，再分别对这3种情况进行处理和计算，抽取出实体关系. 这些方法均取得了较好的效果.

注意力机制在自然语言处理领域和图像处理领域已有诸多应用. 例如，Kambhatla等[8]已经在关系抽取任务中使用注意力机制，提取文本词特征、位置特征和词性特征后，使用卷积神经网络，并加入基于注意力机制的上下文选择器和MLP层，最终取得不错的效果.

针对机器翻译任务中的难点，文献[9]基于注意力机制提出一种被称为“Transformer”的网络结构. Transformer不同于在自然语言处理任务中广泛使用的循环神经网络和Encoder-Decoder结构，它放弃了递归结构而使用注意力机制去刻画输入与输出之间的关系. Transformer并没有使用递归结构，使其能够并行计算，训练速度方面将更优于循环神经网络，另外，Transformer结构中包含多个Multi-Head Attention(多头注意机制)层，能更好地考虑单词的上下文及其语境.

在文献[9]基础上，文献[10]提出了一种双向的Transfomer结构(BERT). 目前，在关系抽取任务中应用BERT模型的研究相对较少，但远程监督的关系抽取中使用该模型居多. 例如，ALT C等[11]人使用BERT模型在NYT数据集上进行远程监督任务，并与较为流行的PCNN+ATT关系抽取分类模型进行对比实验，最终取得不错的效果. 另外，在非远程监督的关系抽取中，Shi等[12]使用简单的BERT模型为下游的关系抽取任务提取文本词特征，最终实验分类精确率达到73.3%.

2 基于BERT与句法依存分析的实体关系抽取

基于机器学习的关系抽取任务中，特征的提取和组织是其关键性步骤，并最终将影响模型的抽取效果. 本研究从文本句法和语义特征的角度出发，以改善和优化对关系抽取影响较大的噪声传播问题. 本研究方法使用依存句法分析以突出文本重要语义表示，理解句子实体之间的关键信息. BERT语言模型则充分考虑单词或单句的上下文，为下游分类模型提供更优质的文本语义的向量表示.

2.1 任务定义

对于给定的句子 $s=\{w_0, w_1, w_2, \cdots, w_n\}$，用 ε 和 $\Re$ 分别表示实体集和关系集，通过命名实体识别抽取出该句中的两个实体 w_i, w_j，其中 $w_i, w_j \in \varepsilon$，通过对句子中的词序列和实体对进行特征抽取和衍生，并以此训练学习模型，将句子中所描述的两实体 w_i, w_j 的关系映射到实体集中的某个 r，其中 r 即为所求的文本中所描述的实体关系.

2.2 依存句法分析

关系抽取中引起噪声传播的因素之一是当文本较长，并且文本中对于关系描述的词语相对隐晦，无用信息过多，导致句子中的有效信息没有被关系分类模型更好学习. 本研究使用依存句法分析[13]对文本中的重要信息进行抽取，称其为“预抽取”.

随着深度学习的发展，目前的依存句法分析主要依靠概率统计学习模型进行实现. 依存关系，即句子成分之间的支配与被支配关系，在依存句法分析中可由弧进行表示，并且依存句法分析中认为支配语句的是其核心动词. 因此，句子的句法结构可以表示为一个由单词作为结点，关系作为边的结构图，以句子“在那个时候，华生是福尔摩斯的朋友和全方位的助手.”为例，其依存句法结构见图1.

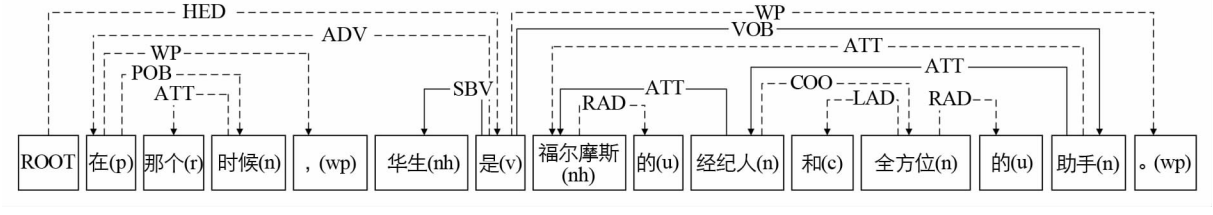

图1 依存句法结构图

图1中的ROOT指向句子中的核心动词“是”，其余各个句子成分之间ADV,ATT,SBV表示支配成分与被支配成分的关系. 表1是依存句法分析中常用到的几种关系.

表 1　依存句法关系标注集

符号	依存关系	例
SBV	主谓关系	我做实验(我←做)
VOB	动宾关系	我做实验(做→实验)
FOB	前置宾语	他什么文献都看(文献←看)
IOB	间宾关系	我送她本书(送→她)
ATT	定中关系	漂亮的景色(漂亮←景色)
ADV	状中关系	非常优秀(非常←优秀)
CMP	动补结构	做完了实验(做→完)
COO	并列关系	小明和小红(小明→小红)
POB	介宾关系	在盒子里(在→里)
DBL	兼语	我请我吃饭(请→我)
LAD	左附加关系	大山和大海(和←大海)
RAD	右附加关系	兄弟们(兄弟→们)
HER	核心关系	我送她一本书(→送)
WP	标点	.

为了使句子成分间关系更为直观，图 1 的依存句法结构可转换为如图 2 的依存句法树. 依存句法树中，支配与被支配关系表示为树的“双亲－孩子”关系，另外树的根节点为句子的核心动词，因此，可以给出依存句法树的定义：依存树 T 可记为 $T=(R,V,A)$，其中 R 为依存树的根节点，即核心动词，V 为依存树中的节点，即句中分词，A 为依存句法树中的弧，即词间的支配与被支配关系. 其中依存句法树 T 需满足：

① 根节点的入度为 0；

② 根节点到其他任一结点皆有路径；

③ 除根节点外，各结点有且仅有一个双亲结点.

使用依存句法树结构表示依存句法分析，通过计算，可以获得句子的最短依存路径(SDP)[14-15]. 最短依存路径是基于依存句法分析计算获得的特征序列. 最短依存路径已经被证实可以有效表示实体之间的语义关系结构，路径中包含的词汇信息是根结点到头尾两个实体结点之间的路径上的信息，其中包含的词汇信息足够表征文本的主要信息，并且能减少冗余的噪声信息. 图 2为句子“在那个时候，华生是福尔摩斯的朋友和全方位的助手. ”的依存句法分析树，其中加粗的是实体华生和实体福尔摩斯之间的最短依存路径，很明显可以看出该最短路径下涵盖了实体之间关系的重要信息. 本研究所使用的最短依存路径计算方法如下：

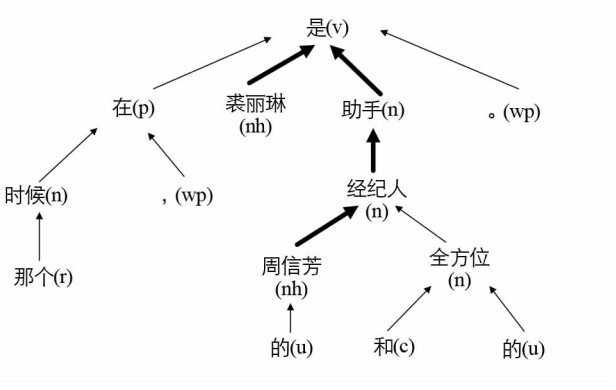

图 2　依存句法树

(1) 语料预处理及依存句法分析

经过文本预处理后，对句子进行依存句法分析，并将结果组织成树结构得到依存句法树 T.

(2) 调整与剪枝

本研究的剪枝与决策树的剪枝操作不同，决策树的后剪枝是从底层向上计算是否剪枝，如需剪枝，则把子树删除，本文对于依存句法树的剪枝，其目的是进一步排除句子中的噪声，降低无关词语对分类结果的影响，但也需考虑句子信息完整性，防止剪枝后丢失重要的语义信息. 因此，本研究将对依存句法树去掉根节点后所得的多棵子树进行调整和剪枝. 若实体词存在，则保留完整子树；若实体词不存在，只保留该子树中的动词和名词.

若舍弃的结点为叶子结点，则直接舍弃；

若舍弃的结点为非叶子节点，则选择孩子节点中的动词结点作为新的双亲结点；当存在多个动词，则按照如下优先级进行选择：实义动词(如，教育、写作等)＞趋向动词＞系动词＞助动词(“＞”表示“优先于”).

(3) 获得最短依存序列

将已剪枝的依存句法树视为特殊的图结构 G_t，以两实体词结点分别作为起始节点 V_{e1} 和终点结点 V_{e2}，使用 Djkstra 最短路径算法求出两个实体节点之间的最短路径，定义其为最短依存路径 P_t，其表达式为：

$$P_t = \text{Djkstra}(G_t, V_{e1}, V_{e2}) \tag{1}$$

其中，在最短依存路径上的词语，组成了该文本的最短依存序列 $P_w = \{w_i, w_{i+1}, \cdots, w_j\}$，$i, j \leqslant n$.

2.3 基于 BERT 的表示学习

如下图 3 所示，本研究使用的 BERT 预训练模型其实质是以 Transformer 模型的编码器作为基模型的一种多头注意力机制构建的模型. 在以往的研究中，BERT 和 RNNs 等模型不同，在并行性方面、语义理解方面的表现更加突出.

BERT 模型只是用了 Transformer 模型的编码器，而其编码器由 6 个相同的层组成，每一层又由两个子层组成，接下来对编码器子层的主要部分进行介绍，其结构如图 4 所示.

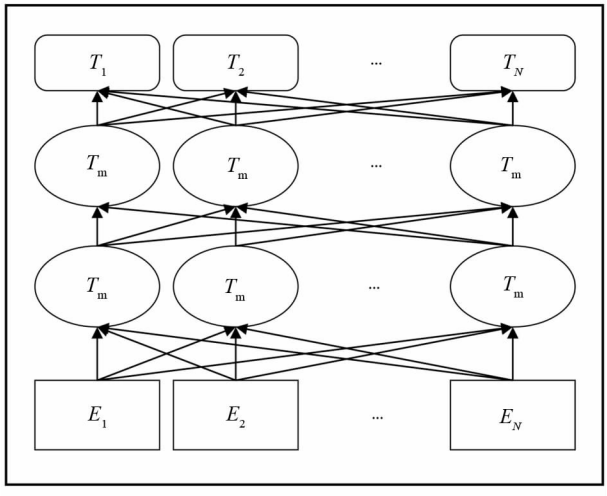

图 3 BERT 模型结构图

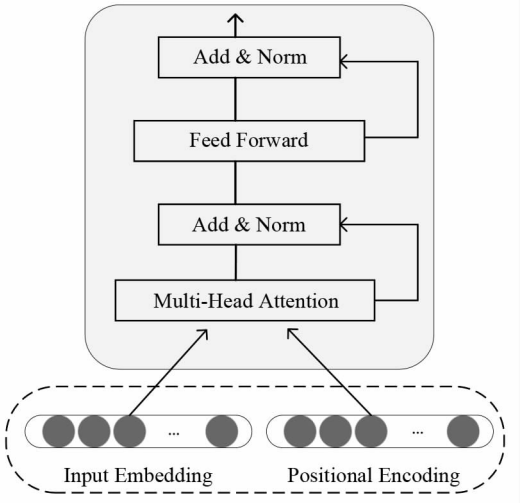

图 4 Transformer 编码器结构图

本研究将经处理后的原序列中的单词映射成多维词向量 $\boldsymbol{e}_i \in \mathbb{R}^d$，$d$ 为词向量的维度. 然后得到句子 s 的词向量集合 $X = \{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n\}$，其中 $X \in \mathbb{R}^{n \times d}$. 因此，多头注意力机制层大致可以表示为：

$$Z = \text{MultiHead}(Q, K, V) = HW^o \tag{2}$$

其中 $W^o \in \mathbb{R}^{hn \times k}$ 为多头注意力的权重矩阵. 模型中的多头自注意力是指首先对 Q，K 和 V 进行不同的线性变换，再计算相似度，这个过程重复做 h 次，然后将 h 次的结果拼接起来再进行线性变换作为多头自注意力机制的结果. 其计算方法为：

$$Q = XW^Q \tag{3}$$

$$K = XW^K \tag{4}$$

$$V = XW^V \tag{5}$$

其中，$W^Q \in \mathbb{R}^{k \times n}$；$W^K \in \mathbb{R}^{k \times n}$；$W^V \in \mathbb{R}^{k \times n}$ 分别为 Q, K, V 的权重矩阵. 然后重复 h 次之后，最终多头注意力的输出就是将各头输出进行拼接，其表达式为：

$$H = \text{head}_1 \oplus \text{head}_2 \oplus \cdots \oplus \text{head}_h \tag{6}$$

其中：$H \in \mathbb{R}^{n \times hn}$，$\oplus$ 为拼接操作. 综上所述，head_i 的表达式为：

$$\text{head}_i = \text{soft max}\left(\frac{(XW_i^Q)(XW_i^K)^{\mathrm{T}}}{\sqrt{k}}\right)(XW_i^V) \tag{7}$$

其中：$W_i^Q \in \mathbb{R}^{k \times n}$，$W_i^K \in \mathbb{R}^{k \times n}$，$W_i^V \in \mathbb{R}^{k \times n}$.

多头注意力机制层的结果，经过残差和归一化处理后，进入前馈神经网络层，该层通过简单的线性激活的运算得到文本语义的向量表示，其过程为：

$$C = \max[0, ZW_1 + b_1]W_2 + b_2 \tag{8}$$

其中：W_1，W_2 为前馈网络的权重矩阵；b_1，b_2 为前馈网络的偏置.

2.4　基于依存句法的关系抽取模型

本研究在依存句法分析工作的基础上，提出基于依存句法的 BiLSTM 模型，即称为 DS-BiLSTM，模型的结构图见图 5.

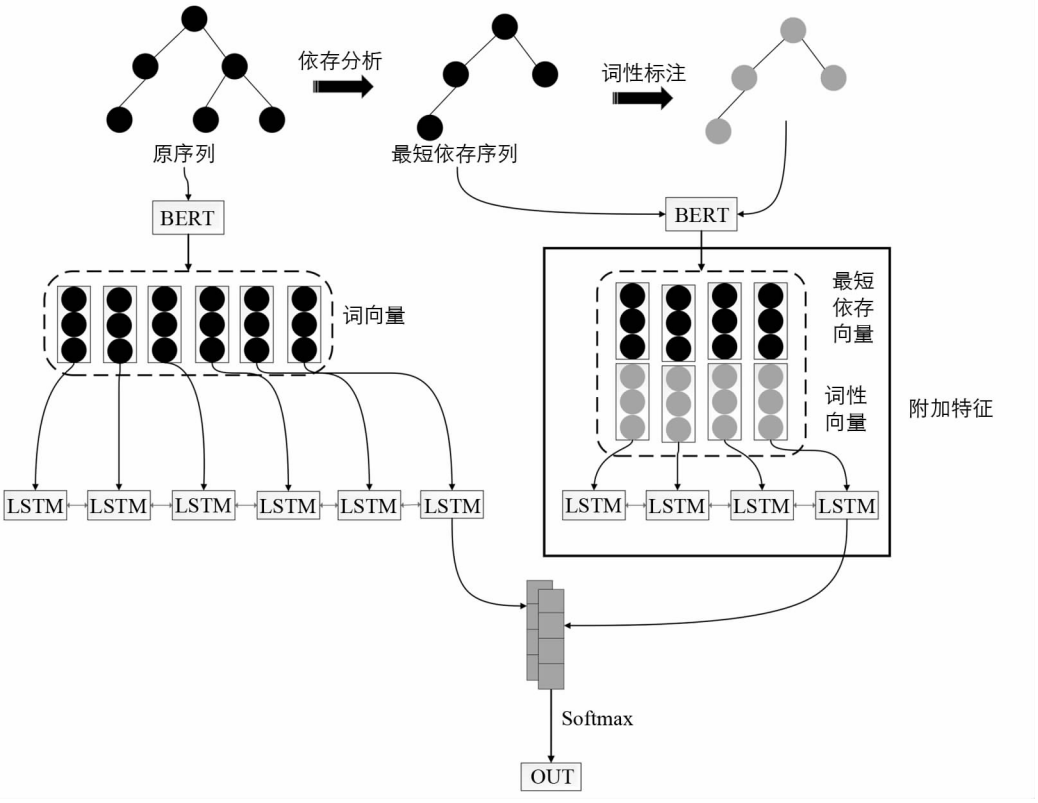

图 5　DS-BiLSTM 模型结构图

在 2.2 节中已经得到了依存句法处理后序列的语义表示 C，为了能更好地注意关键性信息，本文将最短依存序列单独作为 BERT 的一个输入，并且对最短依存序列进行词性标注，将其词性映射成相应的语义向量 $pos_i \in \mathbb{R}^d$，其中 d 为前文提到的词向量维度，pos_i 为第 i 个词的词性，可求最短依存序列的词性特征为：$pos_{0:n} = \{pos_0, pos_1, \cdots, pos_m\}$. 因此，对于给定的句子 $s = \{w_0, w_1, w_2, \cdots, w_n\}$，通过句法分析可得其最短依存序列并映射成相应的词向量序列，可得 $s = \{e_{p0}, e_{p1}, e_{p2}, \cdots, e_{pm}\}$，其中 $m \leqslant n$，将最短依存序列特征及其词性特征进行拼接得到：

$$X_p = s_p \odot pos_{0:m} \tag{9}$$

为更好利用特征并进行编码，通过(2) ～ (5) 式分别计算出其 Self-attention 的查询向量(Query vector)，键向量(Key vector) 和值向量(Value vector)，代入自注意力机制的公式后获得 E_p，再将 E_p 进行简单的线性激活后得到最短依存序列特征 P.

$$E_p = \text{softmax}\left(\frac{Q_p K_p^T}{\sqrt{k}}\right) V_p \tag{10}$$

$$P = \tanh(W_p \cdot E_p + b_p) \tag{11}$$

然后，分别将 P 和 C 分别传入传统的双向 LSTM 神经网络进行分类，其过程大致为：

$$f_t = \sigma(w_f \cdot [h_{t-1}, e_t] + b_f) \tag{12}$$

$$i_t = \sigma(w_i \cdot [h_{t-1}, e_t] + b_i) \tag{13}$$

$$\widetilde{C}_t = \tanh(w_c \cdot [h_{t-1}, e_t] + b_c) \tag{14}$$

$$\overrightarrow{C_t} = f_t \cdot C_{t-1} + i_t \cdot \widetilde{C}_t \tag{15}$$

$$O_t = \sigma(w_o \cdot [\overrightarrow{h_{t-1}}, e_t] + b_o) \tag{16}$$

$$\overrightarrow{h_t} = o_t \cdot \tanh(\overrightarrow{C_t}) \tag{17}$$

$$h_t = \overrightarrow{h_t} + \overleftarrow{h_t} \tag{18}$$

$$\text{output} = \text{softmax}(h_{t1} \oplus h_{t2}) \tag{19}$$

其中，$\overrightarrow{h_{t-1}}$ 是上一时间步的隐含状态，e_t 为当前时间步输入，W 和 b 分别为LSTM各个门的权重矩阵和偏置矩阵. (12) 和(13) 式将上一时间步 h_{t-1} 传来的隐含状态和当前时间步 e_t 的输入通过 sigmod 函数将其映射到一个[0，1]区间，确定遗忘权重和记忆权重. (14) 和(15) 式选择性地将当前时间步特征更新到细胞状态. (16) 和(17) 式则通过前一时间步的隐含状态$\overrightarrow{h_{t-1}}$和当前的细胞状态$\overrightarrow{C_t}$，经过 sigmod 层得出权重 o_t，最后再通过线性激活得到当前的隐含状态$\overrightarrow{h_t}$. (18) 式综合考虑当前时间步的双向信息，即文本上下文，将两个方向的隐含状态综合计算得到当前时间步最后的输出 h_t. 最后，将两个 LSTM 模型的输出进行拼接，并用 softmax 得到最后的模型输出结果，因此采用的代价函数为交叉熵代价函数(cross-entropy).

$$\text{Loss} = -\frac{1}{n}\sum^{D}[y - \ln R + (1-y)\ln(1-R)] \tag{20}$$

其中，D 为数据集的大小，y 为标签，R 为模型的输出.

3 实　　验

本实验将分别在中文人物关系抽取数据集和 Semeval-2010 Task 8 数据集中进行. 数据集统计如表 2 和表 3 所示. 其中人物关系抽取数据集共 10 万条文本，其中包含 11+1 种关系，由于该数据集通过远程监督获得，数据稀疏问题较为严重，经筛选整合后成为本文实验语料，但需要自行划分训练集和测试集；Semeval语料集中训练集含 8 000 个样例，测试集包含 2 700 多样例，涵盖 9+1 种关系. 其中中文语料集将进行随机采样并选取其中 70%作为训练数据，20%作为测试数据，10%用于模型评估.

表 2　人物关系抽取语料集数据统计

序号	关系类型	频次	占比/%
0	未知	0	0
1	父母	12 630	12.63
2	夫妻	15 312	15.31
3	师生	9 450	9.45
4	兄弟姐妹	9 360	9.36
5	合作	8 265	8.27
6	情侣	10 923	10.92
7	祖孙	6 439	6.44
8	好友	8 975	9.00
9	亲戚	9 388	9.39
10	同门	5 671	5.67
11	上下级	3 317	3.32

表 3　Semeval-2010 语料集数据统计

关　系	训练集		测试集	
	频次	占比/%	频次	占比/%
Cause-Effect	1 003	12.5	328	12.1%
Instrument-Agency	504	6.3	156	5.7
Product-Producer	717	9.0	231	8.5
Content-Container	540	6.8	192	7.1
Entity-Origin	716	9.0	258	9.5
Entity-Destination	845	10.6	292	10.8
Component-Whole	941	11.8	312	11.5
Member-Collection	690	8.6	233	8.6
Message-Topic	634	7.9	261	9.6
Other	1 410	17.6	454	16.7

3.1　数据预处理

对于中英文数据集首先进行数据清洗，去除数据中存在的乱码、标点符号、数字替换等. 中文分词采用可自定义分词词典的 Jieba 分词工具，优势在于可以自定义分词词典，对于命名实体识别出来的特定词语和目标实体相对较为友好，英文分词方法相对较为简单则不再赘述. 依存句法分析和分词标注使用 Pyltp 自然语言处理工具包来构建依存句法树和求最短依存路径.

3.2　超参数

Transformer 模型的编码器部分是本文所使用 BERT 模型的主要组成成分，通过双向、多层的 Transfomer 连接而成，表 4 是本研究使用模型的参数设置；对于分类模型 BiLSTM，表 5 为其超参数设置.

表 4　Transformer 编码器参数设置

参数	参数描述	取值
d_model	Dimensionality of the word vectors	400
head_num	Number of heads in multi-head self-attention	8
encoder_num	Number of encoder components	1
hidden_dim	Hidden dimension of feed forward layer	400

表 5　LSTM 分类模型超参数设置

参数	参数描述	取值
Epoch	The times of training on the whole dataset	300
BatchSize	Number of samples selected ineach training	128
Maxlength	The length of the sentence matrix	100
learning_rate	Convergence rate of objective function	0.001
EMBEDDING_DIM	Dimensionality of the word vectors	400

3.3　实验设置

为了验证本文方法的有效性，利用实验抽取的语料设置了对比实验. 首先，本研究方法采用控制变量原则，分为以下 4 组：①word2vec＋BiLSTM；②BERT＋LSTM；③BERT＋BiLSTM；④BERT＋DS-BiLSTM(本文模型).

通过本研究方法验证本文模型在关系抽取任务中效果的提升，同时，可以比较 BERT 和传统 Word2vec 模型的效果，并且基于单向的 LSTM 和双向 LSTM 的模型效果差异也体现出来. 另一方面，本实验与部分现有的关系抽取模型进行对比，实验的文本编码均使用 BERT 模型架构.

(1) APT。Culotta 等[16]提出的基于增强句法树(Augmented Parse Trees)模型.

(2) SVM。Zhao 等[17]使用的 SVM 核方法.

(3) CNN。Zeng 等[18]使用的卷积神经网络.

(4) CNN-ATT。在 CNN 中加入普通的注意力机制.

(5) LSTM。使用 simple-LSTM 进行关系抽取任务.

(6) LSTM-ATT。在(5)中的 simple-LSTM 输出层中加入普通的注意力机制.

(7) BiLSTM-ATT。Zhou[19]提出的在双向 LSTM 上加入注意力机制模型.

(8) Multi-BiLSTM。Xu 等[20]人提出的多通道 LSTM(multi-channel)引入额外信息.

(9) 本研究提出的 DS-BiLSTM 模型.

3.4　实验结果与分析

根据 3.3 中的实验设置进行实验，首先本研究方法进行了对比实验，结果见表 6 和表 7. 其中表 6 表示在人物关系抽取数据集中的模型表现，表 7 则是在 Semeval-2010 Task 8 关系抽取数据集中的表现：

表 6　中文人物关系抽取模型效果

模型	准确率	召回率	F 值
Word2vec＋BiLSTM	0.793	0.823	0.807
BERT＋LSTM	0.812	0.835	0.823
BERT＋BiLSTM	0.846	0.852	0.837
BERT＋DS-BiLSTM	0.857	0.849	0.852

表7 英文Semv2010关系抽取模型效果

模型	准确率	召回率	F值
Word2vec+BiLSTM	0.811	0.828	0.819
BERT+LSTM	0.843	0.850	0.846
BERT+BiLSTM	0.867	0.856	0.861
BERT+DS-BiLSTM	0.871	0.860	0.865

从表6可以比较明显地看出，在中文的人物关系抽取实验中，在预训练模型方面，在特征表示上BERT的效果比Word2vec预训练模型会相对较好，即使在使用同样的分类模型BiLSTM的情况下，准确率、召回率和F值3个指标均提升了5.3%、2.9%和3.0%. 另一方面，普通LSTM和双向LSTM模型单从F值综合性指标进行分析，也有一定提升，但不明显. 而本文方法DS-BiLSTM结合依存句法分析并且将最短依存路径作为额外信息后，实验的3个评价指标都有一定的提高，最终的F值可以到达0.852. 对于英文数据集Semeval-2010的实验效果和中文语料实验效果基本一致，比较全面地说明了本文方法中的语言表示层BERT模型和依存句法分析对于关系抽取任务的效果有一定的提升，噪声传播等问题有一定的改善.

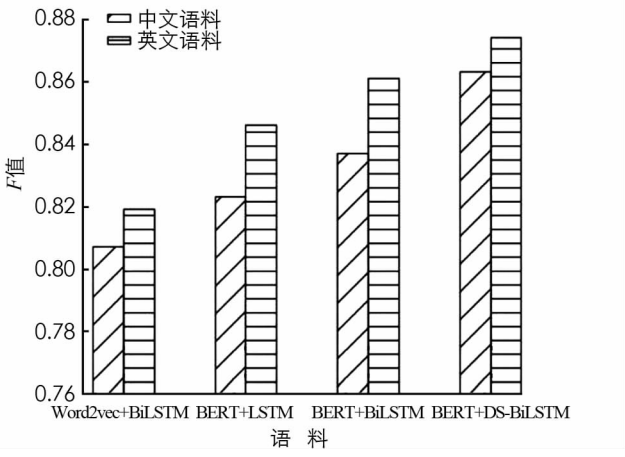

图6 中英文语料效果对比图

如图6所示两个语料集使用相同方法和模型进行实验对比. 从整体的效果来看，Semv2010语料集的F值均较高，原因与数据质量有关. 人物关系语料集通过远程监督获取并人工进行优化，其质量依旧不如Semv2010语料；另一方面，在句法分析上，英语文本中词间关系的句法歧义性较低，在中文语料中发现部分句子由于分词方式的不同导致句子产生歧义. 该问题可通过优化领域分词词典和命名实体识别算法得到解决.

从以上实验结果可见本文方法的有效性，以下的对比实验将与已有的模型进行比较. 其比较结果见表8.

表8 不同模型的关系抽取F值

模型	中文语料集	英文语料集
APT	0.588	0.597
SVM	0.693	0.706
CNN	0.800	0.825
CNN-ATT	0.825	0.813
LSTM	0.827	0.852
LSTM-ATT	0.841	0.843
BiLSTM-ATT	0.843	0.840
Multi-BiLSTM	0.835	0.859
DS-BiLSTM(Our)	0.852	0.865

从表8可以看出，在较早时期Aron等人提出的基于增强句法树(APT)用于关系抽取任务，并未使用BERT等语言模型时，其关系分类效果只有0.588和0.597；而Zhou等人的方法采用word2vec对文本进行编码学习，基于SVM核方法进行分类，效果的F值已经可以提升到0.70左右，但该方法没有考虑句子的噪声问题导致F值相对还是较低；而Zeng使用的模型是当时相对流行的CNN外，还关注到了实体词和其他词之间的距离，也获得了不错的效果. 本文实验在该模型基础上进行尝试，一方面，在输出层加入注意力机制过滤部分噪声，最终的CNN-ATT模型提升0.9%；另一方面，目前较为主流的循环神经网络，Zhou使用的带注意力机制的双向LSTM在目前的关系抽取效果上较为突出，可以达到0.840和0.843的F值，但经过实验，在数据质量较差时，其效果降低也较为明显. 人物关系语料集未经人工处理前，使用该

模型进行实验，最终效果只有 0.68，其对噪声的隔离能力较差；而 Xu 等提出的多通道双向 LSTM 的效果则较为优秀，能达到 0.859 的 F 值.

此外，CNNs 和 RNNs 的分类模型比较，RNNs 的模型方法相对来说表现较好，也体现 RNN 对于序列数据的优势. 本研究方法通过调参后可达到 0.87 的 F 值，相对于以上已有的模型有一定的提升. 但在召回率上，本研究方法还不太平稳. 因此，本研究将中文数据集切割成不同的 size，然后观测其精确率、召回率和 F 值的变化情况(图 7).

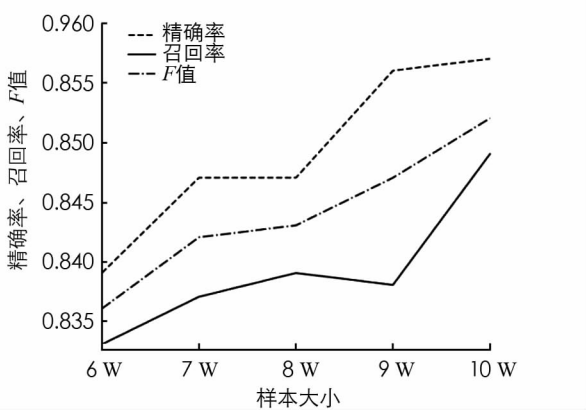

图 7　DS-BiLSTM 模型不同样本效果图

从图 7 可见，随着训练语料规模的不断增加，准确率和 F 值都在呈稳步上升的趋势，但召回率在 8 W 和 9 W 数据的地方反而降低. 该现象仅在本研究的中文语料的关系抽取任务中出现，因此其原因在于部分数据仍存在较多噪声并且质量不佳，而本研究模型虽然仍会受到噪声的影响，但是总体相对稳定，并且取得不错的分类效果.

4　结束语

通过结合依存句法分析，提取句子中的最短依存序列作为额外信息，另外对于原序列根据一定规则进行过滤，尽可能排除噪声，并且使用 BERT 模型挖掘文本深层语言特征，进行更有效的表示学习. 实验证明，本研究方法对于关系抽取中噪声传播的问题有一定的改善.

本研究发现，按本文方法依旧存在弊端，最短依存序列中的语义信息不一定包含实体间关系的语义特征词. 一方面，如何优化和调整依存句法树，使其有效信息能更好地被抽取出来时下一步研究工作的方向；另一方面，如何更好地利用和衍生依存句法分析得到的最短依存路径特征是解决噪声传播问题的关键，也是目前关系抽取领域的重要课题.

依存句法分析单纯从句法层面对文本进行理解和分析，而语义角色标注能更好地识别文本中实体名词之间复杂的语义联系，未来的优化方向将结合语义角色标注进行展开.

参考文献：

[1] 王　伟，吴　芳. 基于注意机制和循环卷积神经网络的细粒度图像分类算法 [J]. 西南师范大学学报(自然科学版)，2020，45(1)：48-56.

[2] 张敏军，华庆一，贾　伟，等. 基于深度神经网络的个性化推荐系统研究 [J]. 西南大学学报(自然科学版)，2019，41(11)：104-109.

[3] JIANG J，ZHAI C X. A systematic exploration of the feature space for relationextraction [C]//Proceedings of Human Language Technologies：the Conferenceof the North American Chapter of the Association for ComTonal Linguistics. 2007：113-120.

[4] 奚　斌，钱龙华，周国栋，等. 语言学组合特征在语义关系抽取中的应用 [J]. 中文信息学报，2008，22(3)：44-49，63.

[5] ZHANG P，LI W J，WEI F R，et al. Exploiting the Role of Position Feature in Chinese Relation Extraction [C]//Proceeding of International Conference on LanguageResources and Evaluation，2008：1-5.

[6] 甘丽新，万常选，刘德喜，等. 基于句法语义特征的中文实体关系抽取 [J]. 计算机研究与发展，2016，53(2)：284-302.

[7] 李明耀，杨　静. 基于依存分析的开放式中文实体关系抽取方法 [J]. 计算机工程，2016，42(6)：201-207.

[8] KAMBHATLA N. Combining Lexical，Syntactic，and Semantic Features with Maximum Entropy Models for Information Extraction[C]//The Association for Computational Linguistics，2004：178-181.

[9] VASWANI A，SHAZEER N，PARMAR N，et al. Attention Is All You Need[C]//Advances in Neural Information Processing Systems 30：Annual Conference on Neural Information Processing Systems(NIPS)，2017：5998-6008.

[10] DEVLIN J，CHANG W M，LEE K，et al. BERT：Pre-training of Deep Bidirectional Transformers for Language Under-

standing[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics(NAACL),2019:4171-4186.

[11] 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. Florence, Italy. Stroudsburg, PA, USA: Association for Computational Linguistics, 2019: 1388-1398.

[12] SHI P, LIN J. Simple BERT Models for Relation Extraction and Semantic Role Labeling [EB/OL]. 2019.

[13] ŞAHiN G G, EMEKLiGiL E, ARSLAN S, et al. Relation Extraction via One-Shot Dependency Parsing on Intersentential, Higher-Order, and Nested Relations [J]. Turkish Journal of Electrical Engineering and Computer Sciences, 2018, 26(2): 830-843.

[14] NINGTHOUJAM D, YADAV S, BHATTACHARYYA P, et al. RelationExtraction between the Clinical Entities Based on the Shortest Dependency Path Based LSTM [EB/OL]. 2019: arXiv: 1903. 09941[cs. CL]. https://arxiv.org/abs/1903. 09941

[15] 温 政, 段利国, 李爱萍. 基于最短依存路径与神经网络的关系抽取 [J]. 计算机工程与设计, 2019, 40(9): 2672-2676, 2696.

[16] CULOTTA A, SORENSEN J. Dependency Tree Kernels for Relation Extraction [C]//Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics-ACL'04. July 21-26, 2004. Barcelona, Spain. Morristown, NJ, USA: Association for Computational Linguistics, 2004: 423-429.

[17] ZHAO S B,GRISHMAN R. Extracting Relations with Integrated Information Using Kernel Methods [C]//The Association for Computer Linguistics(ACL),2005:419-426.

[18] ZENG D J,LIU K,LAI S W,et al. Relation Classification via Convolutional Deep Neural Network [C]//The Association for Computer Linguistics(ACL),2014:2335-2344.

[19] 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). Berlin, Germany. Stroudsburg, PA, USA: Association for Computational Linguistics, 2016: 207-212.

[20] XU Y, MOU L 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. Lisbon, Portugal. Stroudsburg, PA, USA: Association for Computational Linguistics, 2015: 1785-1794.

责任编辑 潘春燕