

基于分层次多粒度语义融合的中文事件检测

丁玲 向阳

同济大学电子与信息工程学院 上海 201804

(dling@tongji.edu.cn)



摘要 事件检测是信息抽取领域中一个重要的研究方向,其主要研究如何从非结构化自然语言文本中提取出事件的触发词,并识别出事件的类型。现有的基于神经网络的方法通常将事件检测看作单词的分类问题,但是这会引入中文事件检测触发词与文本中词语不匹配的问题。此外,由于中文词语的一词多义性,在不同的语境下,相同的词语可能会存在歧义性问题。针对中文事件检测中的这两个问题,提出了一个分层次多粒度语义融合的中文事件检测模型。首先,该模型利用基于字符序列标注的方法解决了触发词不匹配的问题,同时设计了字符-词语融合门机制,以获取多种分词结果中词语的语义信息;然后,通过设计字符-句子融合门机制,考虑整个句子的语义信息,学习序列的字-词-句混合表示,消除词语的歧义性;最后,为了平衡“O”标签与其他标签之间的数量差异,采用了带有偏差的损失函数对模型进行训练。在广泛使用的 ACE2005 数据集上进行了大量实验,实验结果表明,所提模型在精确率(Precision,P)、召回率(Recall,R)和 F1 值这 3 个指标上比现有的中文事件检测模型至少高出 3.9%,1.4%和 2.9%,证明了所提方法的有效性。

关键词: 信息抽取;中文事件检测;多粒度语义融合;预训练语言模型;卷积神经网络;双向长短期记忆模型

中图法分类号 TP182

Chinese Event Detection with Hierarchical and Multi-granularity Semantic Fusion

DING Ling and XIANG Yang

School of Electronics and Information Engineering, Tongji University, Shanghai 201804, China

Abstract Event detection is an important task in information extraction field, which aims to identify trigger words in raw text and then classify them into correct event types. Neural network based methods usually regard event detection as a word-wise classification task, which suffers from the mismatch problem between words and triggers when applied to Chinese. Besides, due to the multiple word senses of a trigger word, the same trigger word in different sentences causes the ambiguity problem. To address the two problems in Chinese event detection, we propose a Chinese event detection model with hierarchical and multi-granularity semantic fusion. First, we adopt a character-based sequence labelling method to solve the mismatch problem, in which we devise a Character-Word Fusion Gate to capture the semantic information of words in different segmentation ways. Then we devise a Character-Sentence Fusion Gate to learn a character-word-sentence hybrid representation of sequence, which takes the semantic information of the entire sentence into condition and solves the ambiguity problem. Finally, in order to balance the influence the label “O” and the other labels, a loss function with bias is applied to train our model. The experimental results on the widely used ACE2005 dataset show that our approach outperforms at least 3.9%, 1.4% and 2.9% than other Chinese event detection models under the metrics of accuracy (Precision, P), recall (Recall, R) and F1.

Keywords Information extraction, Chinese event detection, Multi-granularity semantic fusion, Pre-trained language model, Convolutional neural network, Bidirectional long short-term memory model

1 引言

事件抽取是信息抽取领域中一项基本且重要的研究方向,在智能问答^[1]、信息检索^[1]和股票预测^[2]等领域有着广泛的应用;事件检测是事件抽取任务的第一步,也是关键的一步。

根据 ACE(Automatic Content Extraction)会议的定义,事件主要由触发词(Trigger)和事件元素(Argument)组成。其中,触发词指句子中最能体现事件发生的词语;事件元素描述了事件的结构信息,包括事件参与者、时间、地点等。事件检测主要研究如何从非结构化自然语言文本中提取出事件的触发词,并识别出事件的类型。例如,给定文本“他出生于上

收稿日期:2020-08-10 返修日期:2020-09-15 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:国家重点基础研究发展计划(2019YFB1704402)

This work was supported by the National Basic Research Program of China(2019YFB1704402).

通信作者:向阳(tjdxxyangyang@gmail.com)

海,但是却是在山东长大”,我们需要抽取出触发词“出生”,并判断出所触发的事件类型为“Be-Born”。

现有的事件检测方法主要分为基于特征抽取的方法和基于神经网络的方法。基于特征抽取的方法^[3-6]从词法、句法和语义的角度出发,利用现有的自然语言处理工具设计提取了很多有意义的特征用于事件检测。这些方法的准确率较高,但是高度依赖人工设计的特征,可移植性较差。基于神经网络的方法^[7-12]将文本用高维度的词向量(Word Embedding)进行表示,然后通过各种不同结构的神经网络从文本中自动抽取高质量的、可用于事件检测的特征。与基于特征抽取的方法相比,这种方法不需要太多的人工干预,也避免了使用其他自然语言工具造成的误差传递,因此目前的事件检测研究大多采用基于神经网络的方法。

但是现有的基于神经网络的方法大多基于英文语料,将事件检测看作单词的分类问题,忽略了中文自然语言处理的特性,因此无法很好地处理触发词和文本中的词语完全不匹配的问题^[13]。

具体来说,因为英文中最小的语义单位是单词,所以可以首先根据单词之间的空格进行分词,然后判断每个单词是否是触发词以及触发的事件类型;然而,中文中最小的语义单位是词语,且词语间没有空格进行分割,因此通常需要先进行分词处理,再对词语进行分类。但是,分词可能会造成触发词与文本中的词语不匹配的问题,即触发词是词语的一部分(part-of-word trigger)或者几个词语合起来组成一个触发词(multi-word trigger)。图1(a)给出了1个词语包含2个触发词的例子,即“听证会”包含了2个触发词“听证”和“会”,分别触发了“Trial-Hearing”和“Meet”事件;图1(b)给出了3个词语组成1个触发词的例子,“写”“的”和“信”一起触发了“Phone-Write”事件。因此,如果将中文事件检测直接看作词语的分类问题(即如果判断“听证会”“写”“的”和“信”触发的事件类型),那么就无法正确对以上词语进行分类。

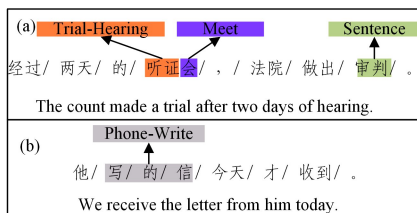


图1 触发词与文本词语不匹配的示例

Fig. 1 Examples of trigger-word mismatch

Chen等^[14]、Qin等^[15]和Li等^[16]分别提出了基于特征抽取的方法来解决这个问题,但是这些方法严重依赖人工对触发词的进一步处理。Lin等^[13]针对这个问题,设计了基于神经网络的NPN(Nugget Proposal Networks)模型,该模型以字符作为输入序列的基本单位,同时将词语的信息加入到字符中,从而得到输入序列的字词混合表示,然而该方法存在触发词重叠的问题。

此外,由于中文词语的一词多义性,相同的词语在不同的语境下会有完全不同的意思。如图2所示,语句“离开多年的工作岗位,他内心不舍”中“离开”一词触发的是“End-Posi-

tion”事件类型,而语句“离开超市后,他直接去了机场”中“离开”一词触发的是“Transport”事件。

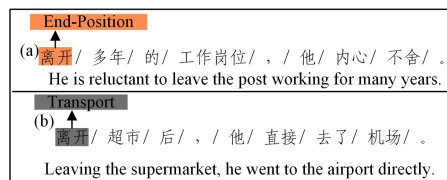


图2 一词多义引起的歧义性问题

Fig. 2 Ambiguity problem caused by a word with multiple senses

本文提出了一种多层次多粒度语义融合的事件检测模型来解决上述问题。为了避免由于分词错误造成的触发词和文本中词语不匹配的问题,我们将字符作为输入序列的基本单位,同时充分考虑了序列存在的多种分词结果,将多个词语的语义信息灵活地融入与之相关的字符表示中;此外,为了解决中文词语的一词多义性问题,我们将整个语句的信息也融入到序列的表示中,最终得到序列的字-词-句混合表示。

本文的主要贡献如下:1)针对中文事件检测中的文本-触发词不匹配问题和一次多义性问题,本文提出了一种多层次多粒度语义融合的事件检测模型;2)充分考虑了多种不同的分词结果以及整体语句的信息,设计了字符-词语和字符-语句两种不同层次的融合门机制;3)在ACE中文语料上进行大量实验,实验结果表明本文方法优于现有的中文事件检测方法。

2 相关工作

事件检测是信息抽取领域一项重要而富有挑战性的任务,现有的事件检测方法主要分为两种:基于特征抽取的方法^[3-6]和基于神经网络的方法^[7-12]。

基于特征抽取的方法利用自然语言处理工具提取出候选词的词法、句法和语义特征,并用 one-hot 向量进行表示,然后利用最大熵、支持向量机等模型对候选触发词进行分类,完成事件检测。Grishman等^[17]、Huang等^[6]和Li等^[18]从不同的角度分析了词法特征在事件检测中的重要作用,设计了多种不同的词法特征;Li等^[19]、Lu等^[20]和Yang等^[21]从句子和篇章的全局角度出发,提取了很多有意义的全局语义特征。这些方法的准确率较高,在ACE英文语料上均取得了不错的效果;但是由于他们严重依赖人工设计的特征和现有的自然语言处理工具,其方法的迁移性差且存在不可避免的误差传递问题。

基于神经网络的方法用词向量来表示输入序列,并通过设计不同的神经网络结构自动地学习输入序列中的关键特征,节省了大量的人力且避免了现有的自然语言处理工具的误差传递问题,因此目前的事件检测研究大多采用基于神经网络的方法。Chen等^[7]和Nguyen等^[22]首次在事件检测中提出了利用动态多池化卷积神经网络模型和领域自适应卷积神经网络模型,在ACE英文语料上取得了当时最好的实验结果;Zhao等^[23]和Chen等^[8]提出利用不同层次的注意力机制来获取句子级和文档级的语义信息;Lin等^[24]通过对损失函数进行改进来解决标签迷惑性问题。这些方法虽然在ACE

的英文语料上都取得了很好的效果,但是都没有考虑中文自然语言处理的特性,即触发词和文本中的词语不能完全匹配的问题。

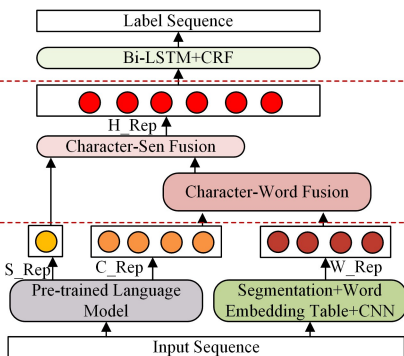
为了解决这一问题,Chen 等^[14]提出了基于字符的候选事件表示模型,并设计了全局触发词勘误表,但是该方法没有充分考虑词语的语义信息且严重依赖人工设计的特征;Qin 等^[15]提出自动扩展候选触发词集的方法来解决这一问题,但是该方法产生的对照表不够灵活和全面。此外,Xia 等^[25]、Zeng 等^[26]和 Lin 等^[13]以字符为基本单位,同时融合了词语的语义信息,利用神经网络学习输入序列的混合表示,但是他们没有考虑到序列的多种分词结果对混合表示的影响以及句子级语义信息对事件检测的作用。

鉴于以上方法的不足,本文提出了一种基于多层次多粒度语义融合的中文事件检测模型,首先通过两个不同层次的融合门机制,动态融合序列的字符级、词语级和句子级等不同粒度的表示,然后利用字符级序列标注模型来完成事件检测。

3 中文事件检测

3.1 整体模型框架

本文提出了一种基于多层次多粒度语义融合的序列标注模型来解决中文自然语言处理中触发词不匹配以及一词多义性问题,模型的整体结构如图 3 所示。该模型主要包含 3 个部分:1)多层次文本表示,首先将输入序列分别用字符和多种分词结果进行表示,然后通过预训练语言模型和词向量模型最终得到序列的字符级表示、词语级表示和句子级表示;2)基于分层融合门的多粒度语义混合表示,通过不同层次的融合门机制将序列的字符级表示、词语级表示和句子级表示进行融合,最终得到序列的字-词-句多粒度混合表示;3)基于序列标注模型的触发词预测,通过双向长短期记忆(Bidirectional Long Short-term Memory, BiLSTM)模型和条件随机场模型(Conditional Random Field, CRF)对序列的混合表示进行标注,识别出文本中的触发词以及触发的类型。



注:S_Rep、C_Rep、W_Rep 和 R_Rep 分别表示序列的句子级表示、字符级表示、词语级表示和字-词-句混合表示

图 3 模型的整体结构

Fig. 3 Architecture of model

3.2 多层次文本表示

给定输入序列 $S = \{c_1, c_2, \dots, c_N\}$, 其中 c_i 表示序列中的第 i 个字符。按照预训练语言模型 BERT^[27] 的要求,首先在序列 S 的开始位置添加字符 [CLS] 来表示整个序列,然后将

序列输入到 BERT 中,得到每个字符 c_i 的向量表示 x_i^s 和整个序列 [CLS] 的表示 x^s :

$$x^s, x_i^s = \text{BERT}(S) \quad (1)$$

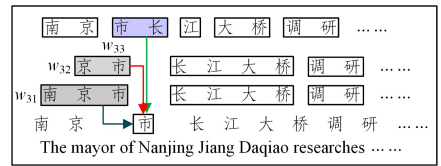
此外,利用分词工具,序列 S 还可以被表示为 $S = \{\omega_1, \omega_2, \dots, \omega_M\}$, 其中 ω_i 表示序列中的第 i 个词语,然后通过查询预训练的词向量表 e ,可以得到序列中每个词语 ω_i 的表示 x_i^w :

$$x_i^w = e(\omega_i) \quad (2)$$

值得注意的是,序列 S 可能存在多种分词方式,如果只考虑一种容易造成误差传递问题。如图 4 所示,序列正确的语义为“南京/市长/江大桥/调研……”,如果只考虑一种分词,将“南京市”“长江大桥”的语义信息错误地加入字符表示中,会影响模型对真正句意的理解。因此,本文首先利用结巴分词工具识别出序列中存在的所有词语,然后针对字符 c_i ,保留与之相关的 3 个词语 ω_{i1}, ω_{i2} 和 ω_{i3} (若多于 3 个则舍弃,若少于 3 个则用 0 向量填充),得到字符 c_i 对应的多个词语表示 $x_{c_i}^w$:

$$x_{c_i}^w = \{x_{i1}^w, x_{i2}^w, x_{i3}^w\} \quad (3)$$

如图 4 所示,通过结巴分词工具得到序列中存在的所有词语,然后保留与“市”相关的 3 个词语(即“南京市”“京市”与“市长”),作为“市”的多个词语表示,然后进行下一步处理。



注:以“市”为例说明如何将多种分词得到的词语“南京市”“京市”和“市长”融入到混合表示中

图 4 序列的多种分词结果实例

Fig. 4 Examples of multiple segmentation results for sequence

3.3 基于分层融合门的多粒度语义混合表示

本文设计了字符-词语、字符-句子两种不同层次的融合门机制,将字符、词语和句子中不同粒度的语义信息融合在一起,学习序列的字-词-句混合表示。

3.3.1 字符-词语融合门机制

为了避免触发词和文本中词语不匹配的问题,可以利用中文词语中丰富的语义信息,将词语的信息融入到字符的表示中。最常见的思路是将词语的表示与字符的表示直接拼接在一起,并认为它们具有相同的重要程度。但是,在中文事件检测中,不同的语境字符和词语的重要程度是不同的。例如,在识别句子“经过两天的听证会”的候选触发词“会”时,模型应该更加关注字符“会”的语义信息而不是词语“听证会”的信息,这是因为词语“听证会”含有“听证”和“会”两个触发词,如果模型过多地关注“听证会”的语义含义,就容易错误地将“会”识别成“Trail-Hearing”事件的触发词;在识别句子“这部电影杀青了”中的触发词时,相比字符“杀”的语义,模型应该更加关注词语“杀青”的语义,这是因为字符的语义和词语整体的语义有较大的偏差,过多地关注字符的语义容易误导模型对语句整体语义的理解。

为了让模型可以根据具体的语境,自主地确定字符和词语信息的重要程度,本文设计了字符-词语语义融合门机制。

具体来讲,由于卷积神经网络可以很好地提取局部特征,我们首先将字符 c_i 的多个词语表示 $\mathbf{x}_{c_i}^w$ 输入到卷积神经网络中,抽取多个词语中最重要的语义信息表示 $\mathbf{x}_{c_i}^{w'}$,以避免由于只利用一种分词信息造成的噪声干扰问题:

$$\mathbf{x}_{c_i}^{w'} = \tanh(\mathbf{W}_{om} \cdot \mathbf{x}_{i1:i3}^w + \mathbf{b}_{om}) \quad (4)$$

其中, \mathbf{W}_{om} 是卷积神经网络的过滤器, \mathbf{b}_{om} 是偏置项, $\mathbf{x}_{i1:i3}^w$ 是 ω_{i1} 到 ω_{i3} 的词向量表示。

然后,对于字符 c_i 的字符级表示 \mathbf{x}_i^c 和词语级表示 $\mathbf{x}_{c_i}^{w'}$,我们通过融合门 \mathbf{G}_i 来动态地计算其字-词混合表示 \mathbf{r}_i ,具体表达式如下:

$$\mathbf{G}_i = \sigma(\mathbf{W}_g(\tanh(\mathbf{W}_{gc} \cdot \mathbf{x}_i^c + \mathbf{W}_{gw} \cdot \mathbf{x}_{c_i}^{w'})) + \mathbf{b}_g) \quad (5)$$

$$\mathbf{r}_i = \mathbf{G}_i \odot \mathbf{x}_i^c + (1 - \mathbf{G}_i) \odot \mathbf{x}_{c_i}^{w'} \quad (6)$$

其中, \mathbf{W}_{gc} , \mathbf{W}_{gw} 和 \mathbf{W}_g 分别是字符表示、词语表示和门机制的权重, \mathbf{b}_g 是门机制的偏置项, σ 是sigmoid激活函数, \odot 是按元素相乘的运算符, \mathbf{G}_i 是学习到的融合门,旨在对字符信息和词语信息的重要程度进行动态的建模, \mathbf{r}_i 是学习到的字符 c_i 的字词混合表示。

我们利用字符-词语融合门机制,得到序列 S 的字-词混合表示 $\mathbf{R} = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N\}$,其包含了序列的字符组成结构和丰富的词语语义。

3.3.2 字符-句子融合门机制

为了解决由中文的一词多义引起的歧义性问题,本文设计了字符-句子融合门机制,在识别触发词时考虑了整个句子的语义,以消除词语的歧义。具体来讲,对于字符 c_i 的字符-词语混合表示 \mathbf{r}_i 和语句的整体表示 \mathbf{x}^s ,我们计算其字-词-句混合表示 \mathbf{r}_i' 如下:

$$\mathbf{r}_i' = \text{ReLU}(\mathbf{W}_{gs}[\mathbf{r}_i; \mathbf{x}^s; |\mathbf{r}_i - \mathbf{x}^s|; \mathbf{r}_i \odot \mathbf{x}^s]) \quad (7)$$

其中, $;$ 表示向量的拼接, $|\mathbf{r}_i - \mathbf{x}^s|$ 表示向量差的绝对值。我们首先通过绝对值运算和点乘运算来衡量 \mathbf{r}_i 和 \mathbf{x}^s 之间的差异性和相似性,然后通过一个全连接层对两个向量自身的表示、向量之间的差异性和向量之间的相似性进行加权。

经过两个不同层次的融合门机制,我们将序列 S 的字符级、词语级和句子级等不同粒度的表示融合在一起,得到序列的字-词-句混合表示 $\mathbf{R}' = \{\mathbf{r}_1', \mathbf{r}_2', \dots, \mathbf{r}_N'\}$ 。

3.4 基于序列标注模型的事件触发词预测

为了有效地捕捉到序列的上下文信息,同时考虑连续标签之间的依赖性,本文采用BiLSTM模型和CRF模型识别序列中的触发词的位置和类型。此外,本文采用“BIO”序列标注模式,标签“O”表示字符不属于任何触发词,标签“B-事件类型”和“I-事件类型”表示相应事件类型触发词的开始位置(Begin)和其他位置(Inside)。

首先对于每个字符表示 \mathbf{r}_i' ,前向LSTM和后向LSTM分别基于不同的上下文计算隐藏状态:

$$\vec{\mathbf{h}}_i = \overrightarrow{\text{LSTM}}(\vec{\mathbf{h}}_{i-1}, \mathbf{r}_i') \quad (8)$$

$$\overleftarrow{\mathbf{h}}_i = \overleftarrow{\text{LSTM}}(\overleftarrow{\mathbf{h}}_{i+1}, \mathbf{r}_i') \quad (9)$$

然后,通过拼接 $\vec{\mathbf{h}}_i$ 和 $\overleftarrow{\mathbf{h}}_i$ 得到字符 c_i 完整的隐藏状态表示 $\mathbf{h}_i = [\vec{\mathbf{h}}_i; \overleftarrow{\mathbf{h}}_i]$ 。最后,将序列的表示 $\mathbf{H} = \{\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_N\}$ 输入到CRF模型中来预测触发词标签序列 $Y = \{y_1, y_2, \dots, y_N\}$ 的概率,即:

$$O(y_i) = \mathbf{W}_{CRF}^y \mathbf{h}_i + \mathbf{b}_{CRF}^y \quad (10)$$

$$p(Y|O) = \frac{\exp(\sum_i (O(y_i) + T(y_{i-1}, y_i)))}{\sum_{Y'} \exp(\sum_i (O(y_i') + T(y_{i-1}', y_i')))} \quad (11)$$

其中, O 是计算 \mathbf{h}_i 的预测标签为 y_i 的得分函数, \mathbf{W}_{CRF}^y 和 \mathbf{b}_{CRF}^y 是计算分数的权重和偏置项, \mathbf{T} 是状态转移矩阵, Y' 是所有可能的标签序列中的任意一个。本文利用维特比算法作为解码器得到分数最高的触发词标签序列,最终得到句子中的候选触发词。

3.5 损失函数

本文采用交叉熵损失函数作为模型的损失函数 $J(\theta)$,同时为了平衡标签“O”过多的问题,利用偏差权重 α 加强非“O”触发词标签对模型的影响,即:

$$J(\theta) = \sum_{j=1}^{N_s} \sum_{i=1}^{N_j} (\log p(Y_i | O_i) \cdot I + \alpha \log p(Y_i | O_i) \cdot (1 - I)) \quad (12)$$

$$I = \begin{cases} 1, & \text{if tag} = \text{“O”} \\ 0, & \text{if tag} \neq \text{“O”} \end{cases} \quad (13)$$

其中, N_s 是一个训练批次中句子的数量, N_j 是句子 S_j 含有的字符数, I 是指示函数,如果字符 c_i 的标签为“O”,那么 I 的值为1,否则为0。

4 实验及结果分析

4.1 数据集和实验设置

4.1.1 数据集的来源和评价标准

本文将ACE2005中文事件抽取语料作为实验数据,该数据集共包含633篇中文文档,且标注了事件的触发词、类别、实体、元素角色等信息。在实验过程中,为了与前人的结果进行公平的对比,本文采用与之前的工作^[14, 26]相同的划分方法,将569/64篇文档分别作为训练集/测试集,并从训练集中随机选择33篇文档作为验证集。此外,我们采用经典的 P 、 R 、 $F1$ 值的评价方法对模型在中文事件检测中的表现做出全面评价。

4.1.2 实验超参数设置

本文在验证集上通过网格搜索法调整模型的参数。其中,预训练语言模型BERT的参数与文献^[27]保持一致;词语的词向量通过查找预训练的中文词向量表得到;BiLSTM和CNN模型的网络参数设置为300维;式(12)中的偏差参数 α 设置为2.0;模型采用Adam优化器,并将其学习率设为0.001;此外,模型的dropout设置为0.5,且在模型训练过程中微调所有的向量表示。

4.2 整体实验结果及分析

本文将提出的多层次多粒度语义融合模型与前人提出的事件检测模型进行比较,具体模型包括:

(1)LSFE(Language Specific Feature Exploration)模型^[14],是一种基于特征抽取的方法,该模型利用人工设计的各种特征与触发词词典表进行中文事件检测。

(2)Rich-C(Rich Component)模型^[28],是一种基于特征抽取的方法,该模型针对中文事件检测的特性,人工设计了不同的特征来联合抽取事件触发词和事件元素角色。

(3)DMCNN(Dynamic Multi-pooling Convolutional Neu-

ral Networks)模型^[7],是一种基于神经网络的方法,该模型设计了一种动态多池化卷积神经网络结构来自动地提取出候选句子中的特征以进行事件检测。

(4)C-BiLSTM(Convolution BiLSTM Neural Network)模型^[26],是一种基于神经网络的方法,该模型将卷积神经网络和双向 LSTM 模型相结合,从候选句子中提取相关特征。

(5)HBTNGMA(Hierarchical and Bias Tagging Networks with Gated Multi-level Attention)模型^[8],是一种基于神经网络的方法,该模型设计了基于多层次注意力机制的标注网络结构,从整篇文档和句子中抽取有意义的特征。

(6)NPN(Nugget Proposal Networks)模型^[13],是一种基于神经网络的方法,该模型根据中文事件检测的特征,利用字符级语义信息和词语级语义信息的混合表示来解决触发词和文本中词语不匹配的问题。

各个模型的实验结果如表 1 所列,其中 DMCNN,C-BiLSTM 和 HBTNGMA 这 3 个模型分别基于字符和词语来实现。

表 1 模型的总体结果

Table 1 Overall results of models

Models	Trigger Identification			Trigger Classification		
	P	R	F	P	R	F
LSFE ^[14]	82.4	50.6	62.7	78.8	48.3	59.9
Rich-C ^[28]	62.2	71.9	66.7	58.9	68.1	63.2
DMCNN(c)	60.1	61.6	60.9	57.1	58.5	57.8
CBiLSTM(c) ^[26]	65.6	66.7	66.1	60.0	60.9	60.4
HBTNGMA(c)	41.6	59.3	48.9	38.7	55.1	45.5
DMCNN(w)	66.6	63.6	65.1	61.6	58.8	60.2
CBiLSTM(w) ^[26]	75.8	59.0	66.4	69.8	54.2	61.0
HBTNGMA(w)	54.3	62.8	58.3	49.9	57.6	53.5
NPN ^[13]	64.8	73.8	69.0	60.9	69.3	64.8
Ours	67.2	73.1	70.0	64.8	70.7	67.7

注:(c)表示模式基于字符表示实现,(w)表示模型基于词语表示实现

由表 1 可知:1)与其他方法相比,本文方法在触发词识别和触发词分类两个任务中均取得了最好的结果,F1 值至少分别提高了 1%和 2.9%;2)与 LSFE 模型和 Rich-C 模型相比,本文模型在两个任务中均取得了明显的提高,且准确率和召回率更加平衡,这说明与基于特征的方法相比,基于神经网络的方法可以自动地抽取出文本中更丰富的语义特征;3)与 NPN 模型相比,本文模型在两个任务中的 F1 值取得了明显的提高,这说明除了动态地利用字符信息和词语信息,句子中蕴含的整体语义信息也有助于解决触发词的歧义性问题;4)与 DMCNN(Char)模型相比,DMCNN(Word)模型在两个任务中的 F1 值分别提高了 4.2%,2.4%,与 C-BiLSTM(Word)和 HBTNGMA(Word)相比,C-BiLSTM(Char)和 HBTNGMA(Char)也取得了更好的实验结果,这说明在中文事件检测中,基于词粒度的模型比基于字符粒度的模型的效果更好,侧面反映了中文是词语驱动的语言;5)没有考虑中文自然语言处理特性的模型性能下降严重,例如,HBTNGMA 模型在英文语料触发词分类任务中得到的 P 值为 77.9,R 值为 69.1,F1 值为 73.3(数据来源于原始论文),而 HBTNGMA(Word)在中文语料中得到的 P,R,F1 这 3 个指标与之相比分别下降了 28%,11.5%和 19.8%,这说明中文自然

语言处理的特性对模型的性能有较大影响,在英文事件检测语料上表现很好的模型不一定适用于中文事件检测。

4.3 对比实验结果及分析

为了验证本文提出的分层次多粒度语义融合机制以及考虑多种分词结果在中文事件检测中的作用,我们进行了一系列的对比实验。具体对比模型的设置如表 2 所列,实验结果如表 3 所列。

表 2 对比模型的设置

Table 2 Setting of ablation study

Models	Description
Char-Only	没有采用融合门机制,仅利用输入序列的字符表示
Char-Word	只采用字符-词语融合门机制得到输入序列的字-词混合表示,其中序列的词语表示利用一种分词结果
Char-S	只采用字符-句子融合门机制得到输入序列的字-句混合表示
Char-* Word	只采用字符-词语融合门机制得到输入序列的字-词混合表示,其中序列的词语表示利用多种分词结果
Char-Word-S	采用字符-词语、字符-句子两种融合门机制得到输入序列的字-词-句混合表示,其中序列的词语表示利用一种分词结果
Char-* Word-S	采用字符-词语、字符-句子两种融合门机制得到输入序列的字-词-句混合表示,其中序列的词语表示利用多种分词结果

表 3 对比实验结果

Table 3 Results of ablation study

Models	Trigger Identification			Trigger Classification		
	P	R	F	P	R	F
Char-Only	59.6	68.5	63.7	58.0	66.2	61.9
Char-Word	66.5	69.7	68.1	63.8	67.1	65.3
Char-S	64.3	68.4	65.9	60.8	65.7	63.1
Char-* Word	67.1	70.6	68.9	64.5	68.3	66.2
Char-Word-S	66.9	71.8	69.3	64.1	69.0	66.4
Char-* Word-S	67.2	73.1	70.0	64.8	70.7	67.7

由表 3 可知:1)与 Char-Only 模型相比,Char-Word 模型在两个任务上的 F1 值分别提升了 4.4%和 3.4%,Char-Sen 在两个任务上的 F1 值分别提升了 2.2%和 1.2%,这说明序列的词语表示和语句表示中含有的语义信息对事件检测均有明显的提升作用,且词语的语义信息对效果提升得更明显;2)Char-Word-S 模型在两个任务上的 F1 值比 Char-Word 模型分别提升了 1.2%和 1.1%,且 Char-* Word-S 模型在两个任务上的结果相比 Char-* Word 模型也有所提升,这说明分层次语义融合机制在事件检测中的作用以及语句中含有的语义信息可以对字-词混合表示进行有效补充;3)与 Char-Word 相比,Char-* Word 在两个任务上的 F1 值分别提升了 0.8%和 0.9%,Char-* Word-S 的表现也好于 Char-Word-S,这说明考虑多种分词可以更加灵活地考虑序列中的词语的语义,在一定程度上解决了由序列分词错误导致的误差传递问题。

4.4 案例分析

本节利用两个示例进一步说明模型在解决触发词与文本中词语不匹配问题以及一词多义性问题等方面的作用。

针对示例 1“经过/两天/的/听证会/,/法院/做出/审判”,C-BiLSTM(Word)模型、NPN 模型和本文模型对其的标注如表 4 所列。由表 4 可知,当文本中词语“审判”与触发词匹配时,基于词语的模型 C-BiLSTM(Word)可以正确地识别

触发词并分类;但是当文本中词语“听证会”包含两个触发词“听证”与“会”时,C-BiLSTM(Word)模型只能对整个词语“听证会”进行分类,导致两个触发词(“听证”“会”)识别错误;此外由于NPN模型采用滑动窗口的方式,没有充分考虑在不同情况下词语与字符之间的权重问题,因此导致遗漏触发词“会”。但本文模型采用字符序列标注的方式,不受分词的影响;此外,模型考虑多种分词方式并利用字符-词语融合门机制,可以根据具体的语境对字符和词语分配不同的权重,在对“听”“证”两个字符进行标注时,充分考虑词语“听证”的语义信息,增大了词语的权重;在对“会”进行标注时,充分考虑字符“会”的语义信息,增大了字符的权重。该模型对整个示例1中语句具体的标注结果如图5所示。

表4 示例1的标注结果

Table 4 Results of labeling for example 1

Models	Trigger Identification/Classification		
	Sentence 1	Sentence 2	
CBiLSTM	听证会, T-H	审判, Se	/
NPN	听证, T-H	审判, Se	/
Ours	听证, T-H	会, Meet	审判, Se
Answer	听证, T-H	会, Meet	审判, Se

注: T-H表示 Trail-Hearing事件, Se表示 Sentence事件

表5 示例2的标注结果

Table 5 Results of labeling for example 2

Models	Trigger Identification /Classification	
	Sentence 1	Sentence 2
CBiLSTM	送, Transport	送走, Transport
NPN	送, Transport	送, Transport
Ours	送, Transfer-Money	送, Transport
Answer	送, Transfer-Money	送, Transport

OOOOO B-TH I-TH B-Mt OOOOO B-Se I-Se O
经过两天的 听 证 会 , 法院做出 审 判 。

The count made a trial after two days of hearing.

注: T-H表示 Trail-Hearing, Se表示 Sentence, Mt表示 Meet

图5 示例1的完整标注

Fig. 5 Labeling for whole sequence of example 1

针对示例2中“送/他/一笔/赴/欧洲/旅行/的/旅费”和“然后/把/他/送走/了”, C-BiLSTM(Word)模型、NPN模型和本文模型对其的标注如表5所列。两个例句的触发词都为“送”,但是却有不同的含义:第一个是“送钱”的意思,第二个是“送别”的意思。如果不考虑整个句子的语义,模型很难区分这两个触发词语义中的差别,因此C-BiLSTM(Word)模型与NPN模型对第一个语句的触发词分类均出现了错误。但是,本文模型通过字符-句子融合门机制考虑了整个句子的语义,并学习序列的字-词-句混合表示,因此可以结合词语“旅费”以及句子的整体语义判断出语句1中“送”的含义为“送钱”,从而对语句1的触发词进行正确识别并分类。

结束语 本文提出了一种基于多层次多粒度语义融合的事件检测模型来解决中文事件检测中的触发词和文本中的词语不匹配问题以及一词多义性问题。该模型首先利用 Bert, Skip-gram 等方法得到序列的3种不同粒度的表示,然后通过两种不同层次的融合门机制将字符、词语和语句中含有的语义信息进行融合,得到序列的字-词-句混合表示,最后通过基于字符的序列标注模型识别出候选句子中的触发词和事件类

型。此外,为了平衡“O”标签与其他标签之间的数量差异,本文采用了带有偏差的损失函数对模型进行训练。实验结果表明了本文模型在中文事件检测中的有效性。

中文事件检测还存在着一些其他问题,例如如何根据上下文的语境抽取候选句子中的触发词、如何通过无监督的方式抽取用户感兴趣的事件类型等,这些都将成为我们下一步的研究重点。

参 考 文 献

- [1] AHN D. The stages of event extraction[C]//Proceedings of the Workshop on Annotating and Reasoning about Time and Events. 2006:1-8.
- [2] YANG H, CHEN Y, LIU K, et al. DCFEE: A document-level Chinese financial event extraction system based on automatically labeled training data[C]//Proceedings of ACL 2018, System Demonstrations. 2018:50-55.
- [3] PATWARDHAN S, RILOFF E. A unified model of phrasal and sentential evidence for information extraction[C]//Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. 2009:151-160.
- [4] LIAO S, GRISHMAN R. Using document level cross-event inference to improve event extraction[C]//Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. 2010:789-797.
- [5] HONG Y, ZHANG J, MA B, et al. Using cross-entity inference to improve event extraction[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. 2011:1127-1136.
- [6] HUANG R, RILOFF E. Modeling textual cohesion for event extraction[C]//Twenty-Sixth AAAI Conference on Artificial Intelligence. 2012:1-7.
- [7] CHEN Y, XU L, LIU K, et al. Event extraction via dynamic multi-pooling convolutional neural networks[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. 2015:167-176.
- [8] CHEN Y, YANG H, LIU K, et al. Collective event detection via a hierarchical and bias tagging networks with gated multi-level attention mechanisms[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018:1267-1276.
- [9] LIU X, LUO Z, HUANG H. Jointly multiple events extraction via attention-based graph information aggregation[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018:1247-1256.
- [10] NGUYEN T H, CHO K, GRISHMAN R. Joint event extraction via recurrent neural networks[C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016:300-309.
- [11] NGUYEN T H, GRISHMAN R. Modeling skip-grams for event detection with convolutional neural networks[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016:886-891.

- [12] SHA L, QIAN F, CHANG B, et al. Jointly extracting event triggers and arguments by dependency-bridge rnn and tensor-based argument interaction[C]//Thirty-Second AAAI Conference on Artificial Intelligence. 2018;5916-5923.
- [13] LIN H, LU Y, HAN X, et al. Nugget proposal networks for Chinese event detection[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (volume 1; Long papers). 2018;1565-1574.
- [14] CHEN Z, JI H. Language specific issue and feature exploration in Chinese event extraction[C]//Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics. 2009;209-212.
- [15] QIN B, ZHAO Y, DING X, et al. Event type recognition based on trigger expansion[J]. Tsinghua Science and Technology, 2010, 15(3):251-258.
- [16] LI P, ZHOU G. Employing morphological structures and sememes for Chinese event extraction[C]//Proceedings of COLING 2012. 2012;1619-1634.
- [17] GRISHMAN R, WESTBROOK D, MEYERS A. Nyu's english ace 2005 system description[C]//Proceedings of the ACE 2005 Evaluation Workshop. 2005.
- [18] LI Y, BONTCHEVA K, CUNNINGHAM H. Using uneven margins SVM and perceptron for information extraction[C]//Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005). 2005;72-79.
- [19] LI Q, JI H, HUANG L. Joint event extraction via structured prediction with global features[C]//Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics. 2013;73-82.
- [20] LU W, ROTH D. Automatic event extraction with structured preference modeling[C]//Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics. 2012;835-844.
- [21] YANG B, MITCHELL T. Joint extraction of events and entities within a document context [C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016;289-299.
- [22] NGUYEN T H, GRISHMAN R. Event detection and domain adaptation with convolutional neural networks[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2; Short Papers). 2015;365-371.
- [23] ZHAO Y, JIN X, WANG Y, et al. Document embedding enhanced event detection with hierarchical and supervised attention[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. 2018;414-419.
- [24] LIN H, LU Y, HAN X, et al. Cost-sensitive regularization for label confusion-aware event detection[J]. arXiv:1906.06003, 2019.
- [25] XIA Y, LIU Y. Chinese Event Extraction Using Deep Neural Network with Word Embedding[J]. arXiv:1610.00842, 2016.
- [26] ZENG Y, YANG H, FENG Y, et al. A convolution BiLSTM neural network model for Chinese event extraction[M]. Cham: Springer, 2016;275-287.
- [27] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv:1810.04805, 2018.
- [28] CHEN C, NG V. Joint modeling for chinese event extraction with rich linguistic features[C]//Proceedings of COLING 2012. 2012;529-544.



DING Ling, born in 1995, doctoral student, is a member of China Computer Federation. Her main research interests include natural language processing, information extraction and event extraction.



XIANG Yang, born in 1962, Ph.D, professor, is a member of China Computer Federation. His main research interests include machine learning, data mining and natural language processing.