

基于篇章信息和 Bi-GRU 的中文事件检测



朱培培 王中卿 李寿山 王红玲

苏州大学计算机科学与技术学院 江苏 苏州 215006

(ppzhu@stu.suda.edu.cn)

摘要 事件抽取是信息抽取中一个重要的研究方向,其中事件检测是事件抽取的关键。目前,中文神经网络事件检测方法均是基于句子的方法,这种方法获得的局部上下文的信息不足以解决事件触发词的歧义性。针对这个问题,文中探索了篇章信息的作用。首先,以双向门控循环单元网络(Bidirectional Gated Recurrent Units, Bi-GRU)模型为基线,定义3个窗口来学习句子特征;然后,将句子表示进行拼接,利用双向门控循环单元网络学习句子的上下文特征;最后,将句子表示和上下文表示进行融合,以丰富句子的语义信息,并减少候选触发词语义模糊现象,通过 Softmax 函数进行事件触发词的分类。在 ACE2005 数据集上的实验结果表明,句子的上下文特征能够有效提升中文事件检测方法的性能,该中文事件检测方法的 F1 值比当前最好的模型高 1.5%。

关键词: 事件抽取;事件检测;篇章信息;双向门控循环单元网络;ACE2005

中图法分类号 TP391

Chinese Event Detection Based on Document Information and Bi-GRU

ZHU Pei-pei, WANG Zhong-qing, LI Shou-shan and WANG Hong-ling

School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China

Abstract Event extraction is an important research task in information extraction and event detection is the key to event extraction. Existing Chinese neural network event detection methods are sentence-based and the local context information obtained by this method is not enough to resolve the event triggers semantic ambiguity. In order to solve this problem, this paper studies document information effects. Firstly, based on the bidirectional gated recurrent units network (Bi-GRU), this paper defines three windows to learn sentence features. Then, the sentence-level representation is concatenated and the document features are learned by using the bidirectional gated recurrent units network. Finally, to enrich the semantic information of sentences and reduce the event-trigger semantic event triggers ambiguity, it merges the sentence-level representation and the document-level representation and then classifies event triggers through the Softmax function. Experimental results on the ACE2005 dataset show that the sentences-context representation can improve the Chinese event detection performance and this event detection method outperforms state-of-the-art results by 1.5% on F1.

Keywords Event extraction, Event detection, Document information, Bidirectional gated recurrent units, ACE2005

1 引言

事件抽取作为信息抽取的一个子任务,被定义为从无结构化的文本中识别事件的发生并抽取结构化的事件元素,如触发词(trigger,标识事件的谓词)及其类型、论元(argument,事件的参与元素)及其角色。根据 ACE(Automatic Content Extraction)评测,事件抽取一般分为4个子任务:触发词识别、事件类型分类、论元识别及论元角色分类。其中,触发词识别和事件类型分类可以合并为事件检测任务,本文的工作中集中于事件检测任务,不考虑论元分类任务。

在中文事件检测任务中,一些触发词在特定的句子中存在严重的语义模糊问题(触发词包含多种语义,可以作为多种事件类型),且现有的中文事件检测方法都是基于句子信息的,一个孤立的句子所包含的语义信息不足以准确判断一个模糊词是否是触发词及其对应的事件类型。如表1所列,句子“大家要求他下台。”中触发词“下台”可以作为“移动”类型,也可以作为“离职”类型,仅根据单个句子很难准确地判断触发词“下台”的事件类型。为了解决这一问题,本文研究了篇章信息的作用。由于词语构成句子,句子构成篇章,篇章内部各层次语言单元之间的信息既相互联系又相互影响,因此,

到稿日期:2019-11-05 返修日期:2019-12-30 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:国家自然科学基金青年科学基金(61806137,61702518);江苏省高校自然科学基金(18KJB520043)

This work was supported by the Young Scientists Fund of the National Natural Science Foundation of China(61806137,61702518) and Natural Science Foundation of the Higher Education Institutions of Jiangsu Province, China (18KJB520043).

通信作者:王中卿(wangzq@suda.edu.cn)

对于事件检测任务,不仅词语的上下文信息对有效完成任务有辅助作用,句子的上下文信息对有效完成事件检测任务也有很大的帮助。句子的上下文信息能为事件检测提供更为丰富的语义信息,从而提升事件检测任务的性能。如表1所列,对于句子“大家要求他下台。”,如果考虑上下文,通过“表演”和“舞台”等信息,“下台”就可以很容易地被标记为“移动”类型。因此,本文提出基于篇章信息的考虑句子上下文信息的神经网络模型。

表1 触发词在特定句子中语义模糊示例

Table 1 Example of trigger semantic ambiguity in particular

| sentence | |
|----------|--------------------------------------|
| 句子 | 大家要求他下台。 |
| 上下文 | 艾斯特的表演太糟糕了。大家要求他下台。 他们觉得舞台不应该属于他。 |

图1给出了一个完整的利用篇章信息的ACE2005中文事件结构示例。在第一句话中,“病逝”是事件的触发词,触发的事件类型为死亡类型。“这个月25号”“苏爸爸”和“医院”分别对应死亡事件模板的3个论元角色标签,即“时间”“目标”和“地点”。在第二句话中,“离开”是事件的触发词,通过第一句话中提供的信息可知“离开”所触发的事件类型为死亡类型,“他”即“苏爸爸”对应死亡事件模板中的论元角色标签“目标”。

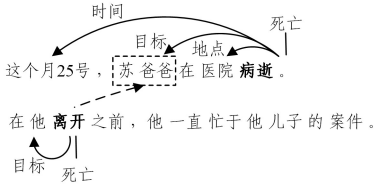


图1 ACE2005 中文事件示例

Fig. 1 ACE2005 Chinese event example

为了能够有效地利用篇章信息,本文利用句子和句子上下文的融合信息来推测未知事件信息。具体来说,本文首先采用 Word2vec 训练的词向量作为词语的初始向量表示,并以 Bi-GRU 为基线,定义 3 个窗口对句子信息进行表示,然后拼接句子表示,通过 Bi-GRU 学习句子的上下文信息;最后将句子的上下文信息与句子信息进行融合,用于弥补句子级语义信息的不足,以解决事件检测中触发词模糊问题。在 ACE2005 中文事件数据集上的实验结果表明,本文提出的基于篇章信息和 Bi-GRU 的神经网络模型的性能在中文事件检测任务中优于基于句子信息的神经网络模型,验证了句子的上下文信息对完成事件检测任务的有效性。

2 相关工作

近年来,事件检测方法主要研究各种不同层次的信息对事件检测性能的影响。根据层次结构,事件检测方法可以分为句子级事件检测方法和篇章级事件检测方法。

2.1 句子级事件检测方法

传统句子级事件检测方法^[1-6]均采用人工构建事件触发词的特征表示,然后借助分类模型判断其事件类型。Chen 等^[1]探讨了触发词和论元中词法、语法和语义特征以及实体

信息的作用,训练了一个最大熵马尔可夫模型(Maximum Entropy Markov Model, MEMM),比较了基于词和基于字的触发词标注的效果。Li 等^[2]用词语的结构语法、形态结构和原义推测未知触发词,提高了系统的召回率。Wang 等^[3]把句子位置、长度和标题词等特征组合起来,结合语义角色标注(Semantic Role Labling, SRL)、命名实体识别(Name Entity Recognition, NER)技术和基于规则的方法抽取 5W1H(who, what, whom, when, where, how)元素并填充到事件本体中,用于事件抽取任务。此外,研究者们还利用同义词、中英平行语料和跨语言信息^[4]等来扩展中文触发词,以解决未知事件触发词(未在训练集中出现的触发词)和数据稀疏等问题。

神经网络事件检测方法^[7-15]使用神经网络模型自动学习和组合上层特征,最后通过 Softmax 模型进行事件的分类。在英文事件检测任务中,Nguyen 等^[7]利用当前词与候选触发词的位置信息以及实体类型来辅助事件识别任务,并通过卷积神经网络(Convolutional Neural Network, CNN)模型自动挖掘预训练词向量的隐含特征,有效地解决了传统方法中的错误传播问题并解除了复杂的特征工程限制。Chen 等^[8]提出动态多池卷积神经网络(Dynamic Multi-Pooling Convolutional Neural Network, DMCNN)模型来抽取更有效的特征以提高事件抽取的性能。除了使用卷积神经网络学习短语级别的特征外,研究者们还通过循环神经网络(Recurrent Neural Network, RNN)对文本序列进行建模,以获得更丰富的上下文特征。Nguyen 等^[9]提出基于循环神经网络的联合学习方法,该方法采用 C-CBOW(Concatenation-based Variant of Continues Bag-Of-Words)模型训练词向量,结合实体类型向量和二元依存关系向量获得序列特征,通过上下文窗口获得局部特征,并利用全局特征挖掘触发词和论元角色之间的依存关系,有效地提高了系统的性能。Feng 等^[10]采用基于卷积神经网络和循环神经网络的混合模型来获取短语和序列信息。此外,研究者们还探索了论元信息^[11]和外部语义知识^[12]等对事件检测任务的影响。在中文事件检测任务中,针对中文数据集中存在的词与触发词之间不匹配的问题,Lin 等^[13]提出基于字的触发词块提议模型,该模型首先采用 DM-CNN 模型分别学习基于字和基于词的语义和语法信息,然后在字词混合表示的基础上,利用中文事件触发词的字符组成结构,提出基于字的触发词块生成模型,并分类触发词块。

2.2 篇章级事件检测方法

篇章级事件检测方法的核心思想是根据实体类型一致性、事件类型一致性和论元一致性等,利用已知的事件信息推理同类型或类似事件中的未知事件信息。在英文事件检测任务中, Ji 等^[16]将事件抽取的范围从单个文档扩展到与主题相关的文档集,并采用基于规则的方法在句子和文档间传播一致的触发词和论元,将全局信息与局部信息结合,在事件检测任务中取得了较好的结果。Liao 等^[17]提出的文档集跨事件推理方法,不局限于同一类型的事件,利用其他类型的事件信息和事件一致性等特征来预测给定事件的类型和解决事件的歧义性。Hong 等^[18]提出跨实体推理的方法,利用实体类型的一致性进一步提高事件抽取性能。Zhao 等^[19]基于 RNN 的层次和监督注意力机制学习事件检测的篇章级嵌入,以优

化另外一个 RNN 事件检测模型。在中文事件检测任务中, Li 等^[20]针对论元缺失问题,利用篇章一致性预测相同词语触发的事件类型。Chen 等^[21]提出了联合模型,利用事件类型一致性和篇章的主题来预测触发词,并把字符、语义角色标注、触发词概率、零指代、论元一致性等特征用于事件检测任务中。

目前,中文篇章级事件检测方法均为传统事件检测方法,本文利用神经网络高效地学习特征,避免了人工构建特征。传统中文事件检测方法利用篇章一致性和触发词类型一致性等进行推理。例如,如果一个词是触发词,那么在同一篇章内的其他相同的该词也是触发词;如果一篇文档内大多数触发词为死亡类型,则定义该篇文档的主题为死亡事件,该文档内的所有触发词有很大的概率为死亡事件。这两种情况也存在一定的局限性,前者不能判断一篇文档内同义词的事件类型,如“出任”和“转任”;后者忽略了一种事件导致另外一种事件发生的情况,如攻击事件导致死亡事件。

与上文所述利用触发词、实体和篇章主题来推理候选触发词不同,本文不是利用规则来推理触发词,而是根据数据集的特点,利用小规模的上下来丰富句子的语义信息,从而增强事件触发词的表现能力。由实验结果可以看出,本文方法简单有效。

3 基于篇章信息的神经网络模型

本文提出的基于篇章信息的中文事件检测模型不仅考虑了词语的上下文信息,还考虑了句子的上下文信息,因此本文选择 Bi-GRU 对序列信息进行建模,以获得更丰富的特征。该模型如图 2 所示,主要由以下 4 部分组成。

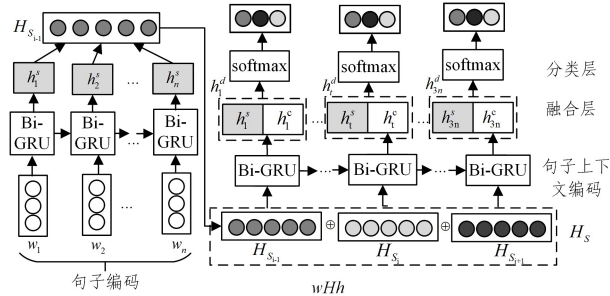


图 2 基于篇章信息的神经网络模型

Fig. 2 Neural network model using document information

(1) 句子编码:采用预训练的词向量作为词语的初始向量表示,并通过 Bi-GRU 模型学习句子级特征,得到句子表示 $H_{S_i} = [h_{i1}^s, h_{i2}^s, \dots, h_{in}^s]$ 。

(2) 句子上下文编码:将句子表示及其上下邻近的句子表示拼接起来,通过 Bi-GRU 模型学习句子的上下文序列特征,得到句子上下文表示 $H_C = [h_{i1}^c, \dots, h_{in}^c, h_{3n}^c]$ 。

(3) 融合层:融合句子表示和句子上下文表示中每个对应词语的隐层表示,得到句子的最终隐层表示 $H = [h_{i1}^d, \dots, h_{i3n}^d, \dots, h_{3n}^d]$ 。

(4) 分类层:用 Softmax 模型判断句子中的每个词语是否是触发词,及其对应的事件类型。

3.1 句子表示

对于给定的一个包含 n 个词语的句子 $S_i = \{\omega_1, \omega_2, \dots,$

$\omega_n\}$, $\omega_t \in V, t \in [1, n]$ (全局词表 V 为预训练词向量中所包含的所有词语集合),首先通过查询预训练向量矩阵得到句子内所有词语的初始化向量表示 $X_i = [x_1, x_2, \dots, x_n]$,然后把向量 X_i 输入 Bi-GRU 模型学习从前往后和从后往前两个方向的句子级序列特征,从而得到隐层向量表示 H_{S_i} 。

本文将由句子 S_i 得到正向隐层表示序列 $\overrightarrow{H_{S_i}}$ 的过程定义为:

$$\overrightarrow{GRU}(x_1, x_2, \dots, x_n) = [\overrightarrow{h_1}, \overrightarrow{h_2}, \dots, \overrightarrow{h_n}] \quad (1)$$

将由句子 S_i 得到反向隐层表示序列 $\overleftarrow{H_{S_i}}$ 的过程定义为:

$$\overleftarrow{GRU}(x_n, x_{n-1}, \dots, x_1) = [\overleftarrow{h_n}, \overleftarrow{h_{n-1}}, \dots, \overleftarrow{h_1}] \quad (2)$$

其中,对于第 t 步的隐层表示 $\overrightarrow{h_t} = \overrightarrow{GRU}(\overrightarrow{h_{t-1}}, x_t)$ 可由当前输入向量 x_t 和上一时刻的隐层表示向量 $\overrightarrow{h_{t-1}}$ 计算得到,计算公式如下:

$$\begin{aligned} z_t &= \sigma(W_z \cdot [\overrightarrow{h_{t-1}}, x_t]) \\ r_t &= \sigma(W_r \cdot [\overrightarrow{h_{t-1}}, x_t]) \end{aligned} \quad (3)$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * \overrightarrow{h_{t-1}}, x_t])$$

$$\overrightarrow{h_t} = (1 - z_t) * \overrightarrow{h_{t-1}} + z_t * \tilde{h}_t$$

句子的隐层表示 H_{S_i} 由双向的隐层表示 $\overrightarrow{H_{S_i}}$ 和 $\overleftarrow{H_{S_i}}$ 中对应词语的隐层表示拼接而成,即词语 ω_t 对应的隐层表示 h_t^i 由双向的隐层表示 $\overrightarrow{h_t}$ 和 $\overleftarrow{h_t}$ 拼接而成,公式如下:

$$h_t^i = [\overrightarrow{h_t}; \overleftarrow{h_t}] \quad (4)$$

3.2 句子的上下文表示

本文依次读取一篇文档内连续的 3 条句子 S_{i-1}, S_i 和 S_{i+1} ,其中 $i \in (1, n)$, n 为该篇文档内句子的总数,形如 $S_1, S_2, S_3, S_2, S_3, S_1, \dots, S_{n-2}, S_{n-1}, S_n$ 。文档的首句和末句均读取 1 次,第二句和倒数第二句均读取 2 次,其他句无特殊情况均读取 3 次。若 n 小于 3,则以文档最后一条句子补齐 3 条句子,将 3 条句子分别输入 3.1 节中的模型得到句子的隐层表示 $H_{S_{i-1}}, H_{S_i}$ 和 $H_{S_{i+1}}$,然后通过拼接得到句子的表示 H_S :

$$H_S = [H_{S_{i-1}}; H_{S_i}; H_{S_{i+1}}] \quad (5)$$

最后将 H_S 输入 Bi-GRU 模型学习句子的上下文序列特征,得到句子的上下文隐层表示 H_C 。具体过程定义为:

$$\overrightarrow{GRU}(H_S) = [\overrightarrow{h_{i1}^c}, \dots, \overrightarrow{h_{i3n}^c}, \dots, \overrightarrow{h_{3n}^c}] \quad (6)$$

$$\overleftarrow{GRU}(H_S) = [\overleftarrow{h_{3n}^c}, \dots, \overleftarrow{h_{i1}^c}, \dots, \overleftarrow{h_{i1}^c}]$$

句子上下文的隐层表示 H_C 由双向的隐层表示中对应词语的隐层表示拼接而成,即词语 ω_t 对应的隐层表示 h_t^c 由双向的隐层表示 $\overrightarrow{h_t^c}$ 和 $\overleftarrow{h_t^c}$ 拼接而成,具体表达式如下:

$$h_t^c = [\overrightarrow{h_t^c}; \overleftarrow{h_t^c}] \quad (7)$$

3.3 融合表示

对于句子中的某个词语 ω_t ,本文通过句子的隐层表示 h_t^i 和句子上下文的隐层表示 h_t^c 得到最终的融合表示 h_t^d 。本文通过以下两种方式计算 h_t^d 并进行实验比较。

$$\text{拼接: } h_t^d = [h_t^i; h_t^c] \quad (8)$$

$$\text{相加: } h_t^d = [h_t^i + h_t^c] \quad (9)$$

通过实验验证,本文最终采用相加的方式进行句子和句子上下文信息的融合。

3.4 事件类型分类

对于句子中的词语 w_i , 隐层表示为 h_i^d , 通过一个全连接层得到每个类别的打分, 公式如下:

$$\mathbf{O} = \mathbf{W} \times \mathbf{h}_i^d + \mathbf{b} \quad (10)$$

其中, \mathbf{W} 是权重矩阵, \mathbf{b} 是偏移量。

计算得到打分之后, 通过 Softmax 层来归一化打分, 得到词语的事件类型概率。词语 w_i 属于事件类别 i 的概率为:

$$P(i|w_i, \theta) = \frac{e^{O_i}}{\sum_{j=1}^K e^{O_j}} \quad (11)$$

其中, θ 是模型参数, O_i 是词语 w_i 属于类别 i 的打分, K 为总类型。

本文依次读取一篇文档内连续的 3 条句子, 因此本文最终会得到句子中词语 w_i 在 3 个不同的上下文语境中属于事件类别 i 的概率, 为了保持同一触发词事件类型的一致性, 本文对词语 w_i 属于事件类别 i 的 3 种概率进行求和操作, 得到 $P_e(i|w_i, \theta)$ 。

3.5 训练与优化策略

假设句子 $S = \{w_1, w_2, \dots, w_n\}$ 对应的事件类型序列为 $Y = \{y_1, y_2, \dots, y_n\}$, 本文使用交叉熵计算损失函数, 计算式如下:

$$J(\theta) = - \sum_{i=1}^n \log P_e(y_i | w_i, \theta) \quad (12)$$

本文采用自适应矩估计 (Adaptive Moment Estimation, Adam) 优化算法对参数进行更新, 并在输入层之后使用 Dropout 机制随机断开一定比例的神经元, 以解决过拟合问题。

4 实验

4.1 实验数据与评价方法

本文选取 ACE2005 中的 633 篇中文事件抽取文档作为实验数据集。该数据集由新闻、广播、对话、博客和论坛等组成, 定义了 8 种事件类型和 33 种子类型。本文提出的事件检测方法只考虑了 33 种事件子类型, 把识别子类型作为本文的任务, 不考虑层次结构。本文通过联合学习触发词识别和事件类型分类任务, 将事件检测任务视为一个基于词语的多分类任务, 具体操作为给定一个句子, 将句子中的每个词语分为 34 类 (33 种子类型 + Null) 中的 1 类。与文献 [5] 相同, 本文分别选择 567 篇和 66 篇文档作为训练集和测试集, 并保留训练集中的 33 篇文档作为验证集。与以往的工作相同, 本文采用准确率 (Precision)、召回率 (Recall) 和 F1 值作为评价指标, 通过如下评价标准来判断触发词的识别和触发词的事件类型分类是否正确。

(1) 触发词的识别: 如果一个触发词的位置与标注的文档内该触发词的位置一致, 则该触发词识别正确。

(2) 触发词的事件类型分类: 如果一个触发词的位置和事件类型与标注的文档内该触发词的位置和事件类型一致, 则该触发词的事件类型分类正确。

4.2 实验参数设置

本文采用大规模中文混合语料训练的词向量¹⁾, 该向量

表示基于 skip-gram 模型, 采用 Word2vec 进行训练。实验的详细参数设置如表 2 所列。

表 2 实验详细参数设置

Table 2 Parameter setting of experiment

| 参数 | 取值 |
|---------|--------|
| 词向量维度 | 300 |
| 隐层表示维度 | 100 |
| dropout | 0.5 |
| 批处理大小 | 20 |
| 学习率 | 0.0001 |

4.3 实验结果及分析

4.3.1 与基准系统的比较

本节将提出的基于篇章信息和 Bi-GRU 的中文事件检测方法与已有的中文事件检测方法进行对比。本文选择以下基准系统进行对比。

(1) Rich-C: Chen 等^[21] 提出的传统事件检测方法。该方法联合学习触发词识别和事件类型分类任务, 利用事件类型一致性和篇章的主题预测触发词, 并运用了丰富的语言学特征。

(2) HNN-Errata: Zeng 等^[14] 提出的神经网络事件检测方法。该方法将事件检测任务视为序列标注问题, 同时采用 Bi-LSTM 和 CNN 模型获取短语和序列信息。除此之外, 该模型还加入了触发词勘误表以修正事件类型。

(3) NPNs: Lin 等^[13] 提出的神经网络事件检测方法。该方法采用 DMCNN 模型分别学习基于字符和基于词语的语义和语法信息, 然后将词表示和字表示融合, 提出基于字的触发词块生成模型并分类触发词。

(4) CSDC: Li 等^[20] 提出的传统事件检测方法。该方法运用成分语义, 针对论元缺失问题, 提出利用篇章一致性预测相同词语触发的事件类型。

(5) TFJM: Li 等^[5] 提出的传统事件检测联合模型。该模型引入触发词过滤模式过滤掉尽可能多的假触发词实例。

(6) Bi-GRU-doc: 本文提出的基于篇章信息的神经网络模型。本文用 Bi-GRU 学习句子和句子的上下文表示并进行融合, 利用句子的上下文信息来丰富句子的语义信息。

在基准系统中, Rich-C 系统采用十折交叉验证法, 其他系统均选择了不同的文档的 10% 作为测试集。为了进行公平的对比, 本文选择了与 CSDC 及 TFJM 完全相同的测试集, 同样采用中文分词工具 ICTCLAS 对句子进行分词, 保留了句子中的实体和触发词并将其作为一个词语。

事件检测的实验结果如表 3 所列。表 3 中, HNN-Errata 和 NPNs 的结果优于 Rich-C, Bi-GRU-doc 的结果优于 CSDC 及 TFJM, 这证明了利用神经网络学习序列特征在效果上优于传统方法的人工构建特征。相比当前结果最好的模型 TFJM, 本文提出的模型 Bi-GRU-doc 在准确率和 F1 值上均有提升, 在触发词识别任务上 F1 值提升了 3.3%, 在事件类型分类任务上 F1 值提升了 1.5%。虽然 TFJM 采用了联合学习的方法, 考虑了实体和触发词之间的联系, 但是这种方法也是基于句子的, 而 Bi-GRU-doc 方法证明了篇章信息对事

¹⁾ <https://github.com/Embedding/Chinese-Word-Vectors>

件检测任务的有效性。与篇章级事件检测方法 CSDC 相比, CSDC 利用篇章一致性,预测一篇文档内相同触发词的事件类型。但是,这种方法并没有解决不同的事件触发词或有歧义的事件触发词所触发的事件类型判别问题。Bi-GRU-doc 的结果优于 CSDC,表明句子的上下文信息能够丰富句子的语义信息,缓解事件触发词的歧义性,从而提高事件检测的性能。

表 3 事件检测的实验结果

Table 3 Experimental results of event detection

(单位:%)

| | Trigger Identification | | | Trigger Classification | | |
|------------|------------------------|-------------|-------------|------------------------|-------------|-------------|
| | P | R | F1 | P | R | F1 |
| Rich-C | 62.2 | 71.9 | 66.7 | 58.9 | 68.1 | 63.2 |
| HNN-Errata | 76.0 | 63.8 | 69.3 | 69.8 | 59.9 | 64.5 |
| NPNs | 64.8 | 73.8 | 69.0 | 60.9 | 69.3 | 64.8 |
| CSDC | 79.3 | 63.5 | 70.5 | 75.2 | 60.2 | 66.9 |
| TFJM | 74.4 | 71.9 | 73.1 | 71.4 | 68.9 | 70.2 |
| Bi-GRU-doc | 81.3 | 72.1 | 76.4 | 76.3 | 67.7 | 71.7 |

4.3.2 篇章信息不同的融合方法比较

(1)Baseline:针对 3.1 节得到的句子表示,通过 Softmax 分类层对句子中的每个词语进行分类,得到其事件类型。

(2)Context:与 3.2 节内容相同,依次读取一篇文档内连续的 3 条句子 S_{i-1} , S_i 和 S_{i+1} , 将其分别输入 3.1 节中的模型,得到句子的隐层表示 $H_{S_{i-1}}$, H_{S_i} 和 $H_{S_{i+1}}$, 然后将 3 个模型的损失函数的和作为整体网络的损失函数,考虑句子上下文信息对事件检测任务的影响。

(3)Concatenation:采用 3.3 节提出的拼接方法进行句子和句子上下文信息的融合。

(4)Add (Proposed):采用 3.3 节提出的相加方法进行句子和句子上下文信息的融合。

不同融合方法的实验结果对比如表 4 所列。首先,Context 模型的 F1 值比 Baseline 高出 0.7%,这说明句子的上下文信息对事件检测任务有一定影响。其次,将句子表示和句子上下文表示进行融合后,准确率、召回率和 F1 值均有提升,这证明句子的上下文表示能够丰富句子的语义信息,一定程度地解决了触发词语义模糊问题,从而提升事件检测任务的性能。由表 4 可知,通过相加操作得到的融合信息取得了最好的结果。

表 4 不同融合方法的结果

Table 4 Results of different fusion methods

(单位:%)

| | Trigger Classification | | |
|---------------|------------------------|------|------|
| | P | R | F1 |
| Baseline | 71.2 | 66.8 | 68.9 |
| Context | 74.1 | 65.6 | 69.6 |
| Concatenation | 75.3 | 67.6 | 71.3 |
| Add(Proposed) | 76.3 | 67.7 | 71.7 |

4.3.3 不同窗口设置的方法比较

ACE2005 中文数据集包含 633 篇文档,大多数文档包含的句子数为 4~10,且该数据集由新闻、广播、对话等组成,具有简洁明了的特点。如表 5 所列,该篇文档是一则新闻,从中可以看出中国爱乐乐团演出事宜和中国爱乐乐团的背景资料

这两方面内容均只用了短短 3 句话进行介绍,这充分体现了该数据集简洁明了的特点。因此,为了获取同一主题同一方面的信息,本文选择小窗口进行句子的特征学习,利用句子的上下文信息来补充句子的语义信息。

表 5 文档示例

Table 5 Document sample

中央台消息:新世纪来临之际,受到社会普遍关注的中国爱乐乐团今天将在北京宝利剧院首次登台亮相。
向首都观众献上一台融中、外音乐于一体的精彩音乐会。
中央人民广播电台文艺调频 90.0 兆赫将在今晚 19:45 现场直播这场演出。
中国爱乐乐团是在中国广播交响乐团的基础上组建,直属于国家广播电影电视总局的国家级乐团。
也是我国第一个以爱乐命名的国家级交响乐团。
乐团从今年 5 月 25 号宣布成立后,在短短半年多时间里,经过严格考核完成了对乐手的招聘并立即投入封闭式训练。

本文分别选择窗口 2-窗口 5 进行对比, windows-2, windows-3(Proposed), windows-4, windows-5 分别表示按照 3.2 节中的方法依次读取 2,3,4,5 句输入到 Bi-GRU 模型。不同窗口设置方法的实验结果如图 3 所示。由图 3 可知,选择 3 个窗口的 Bi-GRU 模型的效果最好。针对该数据集文档短和语言简洁明了的特点,本文选择 3 个窗口最多可以获得一个句子的前面 2 个句子和后面 2 个句子的信息,有利于确保获得同一主题的不同方面的句子信息。

本文运用上下文信息虽然丰富了句子的语义,但因为数据集简明的特点,上下文能提供的信息具有局限性,同一主题不同方面的内容产生了一定的干扰,从而影响事件检测任务的性能。例如,表 5 中的文档以中国爱乐乐团为主题,前 3 句和后 3 句分别介绍了演出方面和背景资料方面;一场事故,前后分别介绍事故的起因和伤亡方面等。如图 3 所示, windows-5 选择 5 个窗口的 Bi-GRU 模型,我们可以获得一个句子前面 4 个句子和后面 4 个句子的信息,由于 ACE2005 中多数文档包含的句子数为 4~10,因此 windows-5 方法增大了同一主题不同方面信息的干扰,从而影响了实验的效果。

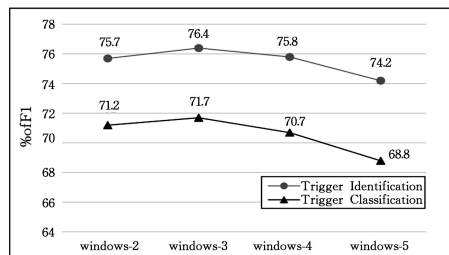


图 3 不同窗口设置方法的结果

Fig. 3 Results of different window set

结束语 本文采用双向门控循环单元网络模型自动学习序列特征表示,避免了复杂的人工构建特征工程。针对触发词语义模糊的问题,本文进一步研究了基于篇章信息的神经网络模型,利用句子的上下文信息丰富了句子的语义信息,进而提升了事件检测任务的性能。在 ACE2005 中文数据集上的实验结果显示,本文提出的基于篇章信息和 Bi-GRU 的中文事件检测方法比最好的基准系统在事件类型分类的 F1 值上高出 1.5%。

在中文事件检测任务中,除了词语的语义模糊问题,还有较为严重的未知事件触发词和事件类型稀疏问题。未知事件

触发词指在训练集中未出现的触发词,事件类型稀疏指某类事件的触发词在训练集中出现次数较少。这些问题阻碍了事件检测任务性能的提升,因此本文的下一步研究方向是探索如何通过基于篇章信息和字符信息融合,来解决事件触发词和事件类型稀疏问题。

参 考 文 献

- [1] CHEN Z,JI H. Language specific issue and feature exploration in chinese event extraction[C]//Proceedings of 47th ACL. Boulder, Colorado; Association for Computational Linguistics, 2009: 209-212.
- [2] LI P F,ZHOU G D. Employing morphological structures and sememes for chinese event extration[C]//Proceedings of COLING 2012. Mumbai, India, 2012:1619-1634.
- [3] WANG W. Chinese news event 5w1h semantic elements extraction for event ontology population [C] // Proceedings of the WWW 2012. Lyon, France, 2012:197-202.
- [4] JI H. Cross-lingual Predicate Cluster Acquisition to improve Bilingual Event Extraction by Inductive Learning[C]//Proceedings of the Workshop on UMSLLS 2009. Boulder, Colorado; Association for Computational Linguistics, 2009:27-35.
- [5] LI P F,ZHU Q M,DIAO H J,et al. Joint modeling of trigger identification and event type determination in chinese event extraction[C] // Proceedings of COLING 2012. Mumbai, India, 2012:1635-1651.
- [6] LI Q,JI H,HUANG L. Joint event extraction via structured prediction with global features [C] // Proceedings of the 51st ACL. Sofia, Bulgaria; Association for Computational Linguistics, 2013:789-797.
- [7] NGUYEN T H,GRISHMAN R. Event detection and domain adaptation withconvolutional neural networks[C]//Proceedings of 53rd ACL. Beijing, China; Association for Computational Linguistics, 2015:365-371.
- [8] CHEN Y B,XU L H,LIU K,et al. Event extraction via dynamic multi-poolingconvolutional neural networks[C]//Proceedings of 53rd ACL. Beijing, China; Association for Computational Linguistics, 2015:167-176.
- [9] NGUYEN T H,CHO K,GRISHMAN R. Joint event extraction via recurrent neural networks [C] // Proceedings of NAACL-HLT 2016. San Diego, California, USA; Association for Computational Linguistics, 2016:300-309.
- [10] FENG X C,HUANG L F,TANG D Y,et al. A language independent neural network for event detection[C]//Proceedings of 54th ACL. Berlin, Germany; Association for Computational Linguistics, 2016:66-71.
- [11] LIU S L,CHEN Y B,LIUK,et al. Exploiting argument information to improve event detection via supervised attention mechanisms[C]//Proceedings of 55th ACL. Vancouver, Canada; Association for Computational Linguistics, 2017:1789-1798.
- [12] LIU S L,CHEN Y B,HE S Z,et al. Leveraging framenet to im-
- prove automatic event detection[C]//Proceedings of 54th ACL. Berlin, Germany; Association for Computational Linguistics, 2016:2134-2143.
- [13] LIN H Y,LU Y J,HAN X P,et al. Nugget proposal networks forchinese event detection[C]//Proceedings of 56th ACL. Melbourne, Australia; Association for Computational Linguistics, 2018:1565-1574.
- [14] ZENG Y,YANG H H,FENG Y S,et al. A convolution bilstm neural network model for chinese event extraction[C] // Proceedings of NLPCC-ICCPOL. 2016:275-287.
- [15] ZENG Y,FENG Y S,MA R,et al. Scale up event extraction learning via automatic training data generation [C] // Proceedings of 32nd AAAI Conference on Artificial Intelligence. New Orleans, Louisiana, USA; AAAI Press, 2018:6045-6052.
- [16] JI H,GRISHMAN R. Refining event extraction through cross-document inference[C]//Proceedings of 46th ACL. Columbus, Ohio, USA; Association for Computational Linguistics, 2008: 254-262.
- [17] LIAO S S,GRISHMAN R. Using document level cross-event inference to improve event extraction[C]// Proceedings of 48th ACL. Uppsala, Sweden; Association for Computational Linguistics, 2010:789-797.
- [18] HONG Y,ZHANG J F,MA B,et al. Using cross-entity inference to improve event extraction [C] // Proceedings of 49th ACL. Portland, Oregon, USA; Association for Computational Linguistics, 2011:1127-1136.
- [19] ZHAO Y,JIN X L,WANG Y Z,et al. Document embedding enhanced event detection with hierarchical and supervised attention[C]//Proceedings of 56th ACL. Melbourne, Australia; Association for Computational Linguistics, 2018:414-419.
- [20] LI P F,ZHOU G D,ZHU Q M,et al. Employing compositional semantics and discourse consistency in chinese event extraction [C]//Proceedings of EMNLP 2012. Jeju Island, Korea; Association for Computational Linguistics, 2012:1006-1016.
- [21] CHEN C,NG V. Joint modeling ofchinese event extraction with rich linguistic features [C] // Proceedings of COLING 2012. Mumbai, India, 2012:529-544.



ZHU Pei-pei, born in 1995, postgraduate, is a member of China Computer Federation. Her main research interests include natural language processing and so on.



WANG Zhong-qing, born in 1987, Ph.D. His main research interests include natural language processing and so on.