

融合触发词特征的事件抽取^①



王立才, 李兴宇, 黄杨琛, 罗琪彬

(华北计算技术研究所 大数据研发中心, 北京 100083)

通信作者: 罗琪彬, E-mail: lqb1139134758@163.com

摘要: 事件抽取是信息抽取领域的重点研究方向. 为了提升事件抽取效果, 解决通用事件抽取方法无法充分利用文本特征信息的问题, 提出了融合触发词特征的事件抽取方法. 通过构建远程触发词库, 为事件类型分类模型提供额外特征信息, 增强事件触发词的发掘能力, 再融合事件类型与触发词距离特征, 提升事件要素抽取模型的表示学习能力, 最后, 将事件类型分类模型与事件要素抽取模型串联, 提升事件抽取效果. 在 DuEE 数据集上进行实验, 与其他模型相比, 本模型提升了准确率、召回率、 $F1$ 值, 证明了本模型的有效性.

关键词: 事件抽取; 预训练模型; 特征工程; 深度学习; 自然语言处理

引用格式: 王立才,李兴宇,黄杨琛,罗琪彬.融合触发词特征的事件抽取.计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/9097.html>

Fusion of Trigger Word Features for Event Extraction

WANG Li-Cai, LI Xing-Yu, HUANG Yang-Chen, LUO Qi-Bin

(Big Data R&D Department, North China Institute of Computing Technology, Beijing 100083, China)

Abstract: Event extraction is a key research area in information extraction. To improve the effect of event extraction and solve the problem that general event extraction methods cannot make full use of text feature information, an event extraction method fused with trigger word features is proposed. A remote trigger word database is constructed to provide additional feature information for the event classification model and enhance the discovery ability of event trigger words. Then, the event type and the distance features of trigger words are integrated to improve the representation and learning ability of the event element extraction model. Finally, the event classification model and the event element extraction model are connected in series to improve the event extraction effect. Experiments on the DuEE dataset demonstrate that compared with other models, this model improves the accuracy, recall, and $F1$ value, which proves the effectiveness of this model.

Key words: event extraction; pre-trained model; feature engineering; deep learning; natural language processing (NLP)

事件抽取作为信息抽取领域的重点研究方向之一, 近年来在自然语言处理领域备受瞩目, 广泛应用于信息搜索、智能问答、文本摘要等领域. 事件抽取任务中的事件由 ACE 会议定义, 包括事件触发词和事件元素. 事件触发词指能够触发特定事件的关键词, 是决定事件类型的最重要因素, 事件元素指事件涉及的时间、地点、实体、属性值等, 是事件的细粒度刻画, 两

者相结合, 完整描述事件本身. 事件抽取的核心是从非结构化文本中抽取事件类型、发生地点、时间、参与角色以及具体内容, 并将这些内容展示为结构化信息.

事件抽取通常分为事件类型分类和事件元素抽取两个子任务, 事件类型分类指从文本信息中识别能够代表事情发生的触发词, 并根据触发词判断事件类型, 事件元素抽取指从文本中抽取出事件的时间、地点、

① 收稿时间: 2022-09-28; 修改时间: 2022-10-27, 2022-11-08; 采用时间: 2022-12-23; csa 在线出版时间: 2023-05-19

实体等关键要素,并匹配其在事件类型中扮演的角色。类似于关系抽取,事件抽取中的事件类型及其对应的事件角色也是预先定义的,如表1会见事件类型中,包含的事件角色有时间、地点、会见主体、会见对象。

表1 会见事件类型抽取示例

文本	事件角色	事件元素
国务委员兼外交部长王毅 2019年8月1日在泰国曼谷 会见美国国务卿蓬佩奥。	时间	2019年8月1日
	地点	泰国曼谷
	会见主体	国务委员兼外交部长王毅
	会见对象	美国国务卿蓬佩奥

事件抽取任务目前也有着许多难题与缺点,一是一个句子中可能同时出现多个事件触发词触发不同类型的事件,例如,句子1:沈跃跃会见俄罗斯妇女联盟主席拉霍娃并出席纪念中俄建交70周年妇女交流图片展开幕式。句子1同时出现了会见和开幕两种事件类型。二是将事件类型分类和事件元素抽取视为两个子任务时,事件类型只确定对应事件模板,并不参与事件元素抽取,忽视了事件类型与事件元素之间存在的关系。三是在进行事件元素抽取时,不考虑事件元素与事件触发词位置之间的关系,可能导致事件元素的抽取不准确。

针对上述问题,本文提出了一种融合触发词特征的事件抽取方法,分为两个部分:一是融合远程触发词位置特征的事件类型分类模型。具体来说,我们设计规则构建了一个远程触发词库,并通过将远程触发词表示与文本表示拼接到一起作为文本特征进行事件类型分类。通过这种方式,模型增强了对于触发词的特征表示能力与挖掘能力。二是融合事件触发词距离特征与事件类型特征的事件元素抽取模型。具体来说,针对事件触发词与事件类型不参与事件元素抽取的问题,我们将事件类型与文本拼接到一起作为预训练语言模型的输入文本,并将触发词与各个字符的距离作为特征与文本表示拼接起来增强事件元素与事件触发词、事件类型的关联。

我们的贡献总结如下。

- 我们提出了一种构建远程触发词特征的方法。在事件类型分类模型中,将这种特征与通过预训练语言模型的文本向量表示融合后,可以捕获更多事件触发词、更准确的事件类型。

- 我们提出了一种在事件元素抽取过程中融合事

件触发词距离特征与事件类型特征的方法。增强了事件触发词、事件类型与事件元素的关联关系。

- 实验表明在包含65种事件类型的DuEE数据集上,相较于之前的模型,本文方法提升了2.1%~2.6% F1值。

1 相关研究

从事件抽取任务出现至今,其主要解决方法分为3种^[1]:基于模式匹配的事件抽取、基于机器学习的事件抽取、基于深度学习的事件抽取。

首先出现的是基于模式匹配的事件抽取,模型通过学习标注语料的相应模式,对待抽取文档进行模式匹配。Riloff^[2]提出了Auto slog系统,通过对标注语料的学习来生成相应领域的触发词知识库与匹配规则。Kim等^[3]通过将WordNet获取的语义信息融入到模式结构中进行特定领域的事件抽取。使用这类方法的重点在于模式的构建,需要领域专家指导且费时费力,模式的迁移性差,因此适合在某一特定领域使用。

之后,基于机器学习的事件抽取把事件类别分类与事件元素抽取视为分类问题,将事件触发词分类为对应事件类型,将事件元素分类为相应事件角色。Chieu等^[4]提出了一种基于最大熵的分类模型,结合其文本对应的领域特征完成事件抽取。Ahn^[5]使用Timbl和MegaM模型,结合候选词的语义、词法、句法等特征进行事件抽取。Llorens等^[6]采用CRF提升了事件元素抽取的效果。事件抽取中的机器学习方法虽然避免了模式的构建,但是需要大规模的标注语料和人工设计特征,依然耗时耗力,且其不可自动学习新特征的特点导致其泛化性较差。

近年来,由于深度学习方法表现出的强大特征表示能力,其迅速被运用于AI领域的各个任务。Nguyen等^[7]首先使用卷积神经网络(CNN)进行事件检测任务,将事件检测问题视作分类问题,通过引入实体信息,增强CNN模型的领域适应性。Chen等^[8]提出动态多池化卷积神经网络模型(DMCNN),结合skip-gram模型提取文本句子特征。接着,Nguyen等^[9]又提出使用双向循环神经网络来学习句子的更丰富的表示,并捕捉事件元素与触发词标签的依赖关系,提升事件抽取效果。Chen等^[10]提出了一种双向动态的长短期记忆网络(BDLSTM-TNNs),并设计了一个特殊的张量层来自动探索候选参数间的相互关系并联合预测事件角色。随着深度学习方法的深入应用,为了能够更好表示词

语以及句子的深层次语义,词向量技术也成为突破方向. Peters 等^[11]提出了 ELMo 模型,结构为 CNN-BiLSTM-CNN,能够根据当前上下文对词向量表示进行动态调整,其在大规模语料集上的预训练模型提升了自然语言处理领域各个任务的效果. Radford 等^[12]将 LSTM 替换为单向 Transformer,提出了预训练语言模型 GPT,接着, Devlin 等^[13]又提出了带有双向 Transformer 的 BERT 模型,只需要在其后进行下游任务的微调就可取得非常好的效果. 事件抽取任务也受其影响,提升了相当的效果.

2 融合触发词特征的事件抽取

融合触发词特征的事件抽取方法如图 1 所示,由事件类型分类模型与事件元素抽取模型组成. 对于事件类型分类模型,我们首先构建一个远程触发词库,包含了训练集中具有代表性和可靠性的触发词. Search

location module 通过在搜寻在文本中出现的远程触发词,并进行定位得到远程触发词位置. 进而将通过分词器得到的文本标记与远程触发词位置输入到事件类型分类模型中,得到触发词 trigger 与事件类型 type. 对于事件元素抽取模型,将事件类型分类模型得到的事件类型与文本拼接并通过分词器得到新的文本标记. 再将文本通过距离计算模块计算文本各个字符与事件类型分类模型得到的触发词的距离,得到 trigger distance. 将新的文本标记与 trigger distance 作为事件元素抽取模型的输入,得到事件元素 argument 及其对应角色 role. 最后通过事件类型对事件角色进行约束并提取该事件角色对应的事件元素,统一整合后作为事件抽取结果.

2.1 事件类型分类模型

事件类型分类是在句子中抽取出现的触发词,并根据触发词将事件分为对应的事件类型. 事件类型分类模型如图 2 所示.

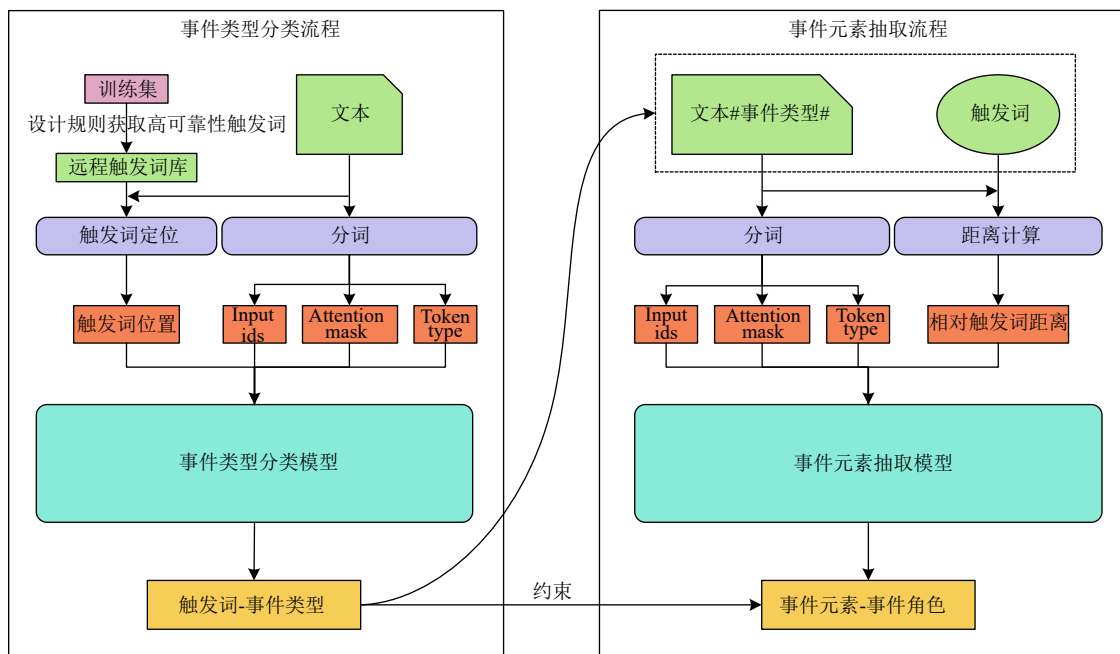


图 1 事件抽取总体流程图

借鉴远程监督的思想,利用训练集中常出现且具有代表性的触发词构建远程触发词库,为事件类型分类过程提供额外特征. 通过人工设计规则将这些触发词统一构建到远程触发词库中,如具有代表性的触发词一般是包含 2-3 个中文字符的词语 (“裁判”“解散”“夺冠”等),使构建的远程触发词库具有较高的可靠性

与代表性. 当远程触发词出现在输入句子中时,将其出现位置进行标注,得到远程触发词起始位置 s , 结束位置 e , 构造相应 *Trigger location* 向量表示.

$$Trigger\ location = \{t_1, t_2, \dots, t_i\} \tag{1}$$

$$t_i = \begin{cases} 1, & s \leq i \leq e \\ 0, & \text{else} \end{cases} \tag{2}$$

通过 Embedding 层将 trigger loc 转化为相应的向量表示 T , 并将其与预训练语言模型得到的句子向量表示 E 拼接到一起作为全部向量表示 I .

$$I = \text{Concat}(T, E) \quad (3)$$

将 I 输入到全连接层得到新的输出 X :

$$X = W \times X + B \quad (4)$$

再使用 Dropout 层以概率 p 保留每个神经元的输出, 丢弃剩下的输出. 使用 Dropout 层随机地丢弃神经元, 能够让现有的模型不断改变, 实现为多个模型的效果和, 提升了模型的鲁棒性与稳定性.

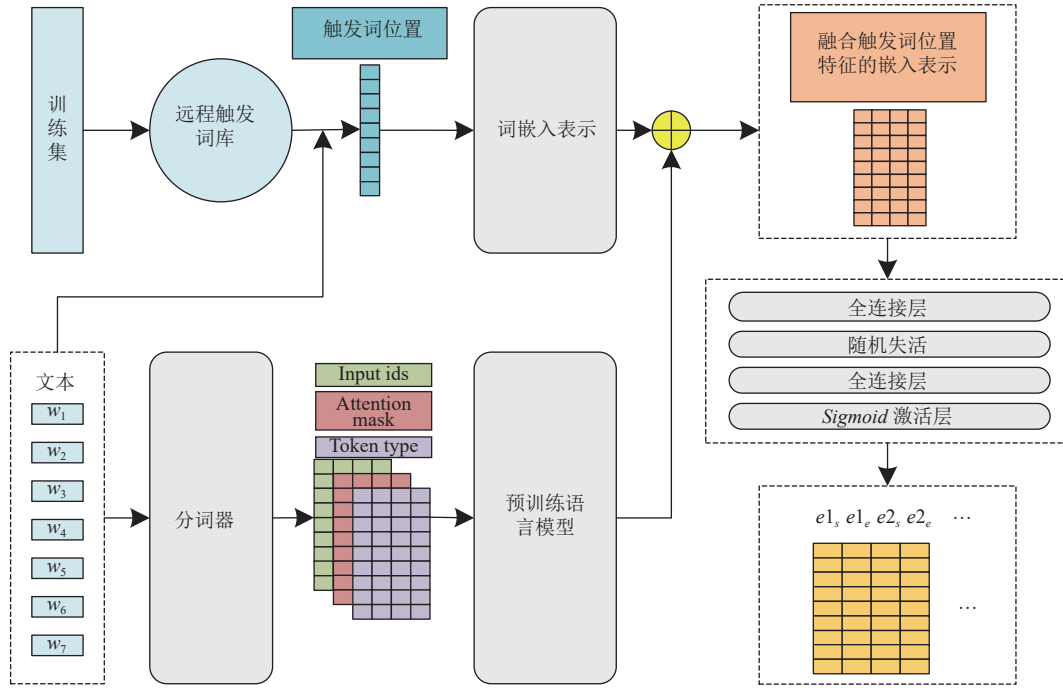


图2 事件类型分类模型

进行触发词位置预测时, 采用指针解码方法, 首先经过全连接层改变 X 最后一维大小得到 D :

$$D = W * X + B, D \in R^{n \times e} \quad (5)$$

其中, n 为句子最大长度, 由于需要预测每个事件类型的起终点位置, 因此 e 为事件类型个数的两倍.

再通过 Sigmoid 激活函数得到字符作为每种事件类型的起终点位置概率矩阵 S .

$$S = \text{Sigmoid}(D) \quad (6)$$

采用二分类交叉熵作为损失函数:

$$\text{Loss} = \frac{1}{N} \sum_{n=1}^N l_n \quad (7)$$

$$l_n = -W[y_n \cdot \log s_n + (1 - y_n) \cdot \log(1 - s_n)] \quad (8)$$

其中, W 为超参数, y_n 为正确标签, s_n 为预测标签, N 为 batch size.

使用 AdamW 作为优化器最小化损失函数. 由于每个字符有作为每种事件类型的起始概率和终点概率,

因此可以通过设置合适的阈值对抽取效果进行调整, 得到事件触发词及其对应的事件类型. 若无事件触发词概率达到阈值, 则选择远程触发词库中出现在文本中概率最高的触发词, 若无远程触发词出现在句子中, 则此文本不抽取事件.

2.2 事件元素抽取模型

事件元素抽取是在句子中抽取与出现事件相关的元素, 并将它们分别与事件角色对应.

本文的事件元素抽取方法借鉴关系抽取的思想, 将 (事件触发词, 事件角色, 事件元素) 视作三元组进行事件元素抽取. 其具体流程与事件类型分类模型相似, 但在处理特征时有所不同.

首先将事件类型分类模型得到的事件类型与输入文本进行拼接得到 text#event type#, 将其通过预训练语言模型得到文本向量表示 I . 再使用事件类型分类模型得到的事件触发词起始位置 t_s 与终点位置 t_e 构造距离特征 C , 增强句子文本的特征向量表示:

$$C = \{c_1, c_2, \dots, c_i\} \quad (9)$$

$$c_i = \begin{cases} t_s - i, & i < t_s \\ 0, & t_s \leq i \leq t_e \\ i - t_e, & i > t_e \end{cases} \quad (10)$$

将距离向量表示 C 与文本向量表示 I 拼接到一起作为更丰富的向量表示 M :

$$M = \text{Concat}(C, I) \quad (11)$$

再将 M 通过全连接层-Dropout 层-全连接层转化为 N :

$$N = W * M + B, N \in R^{n \times a} \quad (12)$$

其中, a 为事件角色个数的两倍。

经过 *Sigmoid* 激活函数得到字符作为每个事件角色起终点概率。损失函数、优化器与事件类型分类模型相同,在此不再赘述。

由于得到的事件角色起终点概率包含所有的事件角色,因此需要使用事件类型分类模型得到的事件类型对其事件角色进行约束,只提取该事件类型下事件角色对应的事件元素作为事件元素抽取模型的结果。

3 实验结果与分析

3.1 数据集

实验数据集采用百度的中文事件抽取数据集 DuEE^[14]。DuEE 定义了 9 种事件大类,每种事件大类包含相应的子事件类型,一共有 65 种事件类型,1.7 万个包含事件信息的句子。其中 1.2 万条作为训练集,0.15 万条作为验证集,0.35 万条作为测试集。事件类型及元素示例如表 2 所示。

表 2 事件类型及元素示例

事件类型	事件角色
组织关系-裁员	时间、裁员方、裁员人数
灾害/意外-爆炸	时间、地点、死亡人数、受伤人数
司法行为-罚款	时间、罚款对象、执法机构、罚款金额

3.2 评价指标

将事件类型预测正确且事件角色与事件元素对应正确的事件元素视为预测正确。

P =预测论元正确个数/所有预测论元的数量。

R =预测论元正确个数/所有人工标注论元的数量。

$F1$ 综合得分: $F1 = (2 \times P \times R) / (P + R)$ 。

3.3 实验结果与分析

模型构建使用 PyTorch 框架。模型最大输入长度

设为 256, batch_size 设为 48, 学习率设为 2×10^{-4} , 共训练 30 个 epoch。将本文的模型与 BiLSTM-CRF, BERT-CRF, RoBERTa-CRF 等模型进行比较,其结果如表 3。

表 3 实验结果 (%)

模型	P	R	$F1$
BiLSTM-CRF	61.9	57.5	59.6
BERT-CRF	73.1	71.8	72.4
RoBERTa-CRF	76.0	75.2	75.6
Our-BERT	74.2	74.8	74.5
Our-RoBERTa	77.1	79.4	78.2

BiLSTM-CRF 模型^[15]是经典的序列标注模型,通过获得文本对应的静态词嵌入,经过 CRF 分类得到最佳序列标注,与预训练语言模型相比,缺少了词嵌入的动态表示,不能准确表示文本特征。BERT-CRF^[16]与 RoBERTa-CRF^[16]都是基于预训练语言模型的序列标注方法,都能获得动态词向量表示,区别在于 RoBERTa 模型采用了动态 mask 机制、去掉了 BERT 模型中的 NSP 预训练任务并且使用了字节级别的编码方式,能够学习到更有效的特征表示,因此得到了更好的效果。将本文的触发词特征融入到模型后,使得模型能够包含更多的特征信息,且本方法能够通过调节模型阈值进行多分类输出,解决了多事件类型以及多事件元素的问题,显著提升了事件抽取效果。

4 总结

本文提出了一种融合触发词特征的事件抽取方法,通过构建远程触发词库为事件类型分类提供额外的触发词特征,再将事件类型分类模型得到的事件触发词构建为触发词距离特征,并将事件类型与文本拼接增强事件类型特征表示。在事件抽取数据集 DuEE 上与其他模型比较,提升了事件抽取效果,验证了融合触发词特征的事件抽取方法的有效性。在后续的工作中,将在特定领域,结合其数据特点,继续进行事件抽取研究,为构建领域知识图谱提供数据支撑。

参考文献

- 1 高强,游宏梁.事件抽取技术研究综述.情报理论与实践,2013,36(4):114-117,128. [doi: 10.16353/j.cnki.1000-7490.2013.04.011]
- 2 Riloff E. Automatically constructing a dictionary for information extraction tasks. Proceedings of the 11th National Conference on Artificial Intelligence. Washington:

- AAAI Press. 1993, 811–816.
- 3 Kim JT, Moldovan DI. Acquisition of linguistic patterns for knowledge-based information extraction. *IEEE Transactions on Knowledge and Data Engineering*, 1995, 7(5): 713–724. [doi: [10.1109/69.469825](https://doi.org/10.1109/69.469825)]
 - 4 Chieu HL, Ng HT. A maximum entropy approach to information extraction from semi-structured and free text. *Proceedings of the 18th National Conference on Artificial Intelligence*. Edmonton: ACM, 2002. 786–791.
 - 5 Ahn D. The stages of event extraction. *Proceedings of the 2006 Workshop on Annotating and Reasoning about Time and Events*. Sydney: ACL, 2006. 1–8.
 - 6 Llorens H, Saquete E, Navarro-Colorado B. TimeML events recognition and classification: Learning CRF models with semantic roles. *Proceedings of the 23rd International Conference on Computational Linguistics*. Beijing: Coling 2010 Organizing Committee, 2010. 725–733.
 - 7 Nguyen TH, Grishman R. Event detection and domain adaptation 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: Association for Computational Linguistics, 2015. 365–371.
 - 8 Chen YB, Xu LH, Liu K, *et al.* Event extraction via dynamic multi-pooling 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: Association for Computational Linguistics, 2015. 167–176.
 - 9 Nguyen TH, Cho K, Grishman R. Joint event extraction via recurrent neural networks. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. San Diego: Association for Computational Linguistics, 2016. 300–309.
 - 10 Chen YB, Liu SL, He SZ, *et al.* Event extraction via bidirectional long short-term memory tensor neural networks. *Proceedings of the 15th China National Conference on Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*. Yantai: Springer, 2016. 190–203.
 - 11 Peters ME, Neumann M, Iyyer M, *et al.* Deep contextualized word representations. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. New Orleans: Association for Computational Linguistics, 2018. 2227–2237.
 - 12 Radford A, Narasimhan K, Salimans T, *et al.* Improving language understanding by generative pre-training. http://openai-assets.s3.amazonaws.com/research-covers/language-unsupervised/language_understanding_paper.pdf. 2018.
 - 13 Devlin J, Chang MW, Lee K, *et al.* BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Minneapolis: Association for Computational Linguistics, 2019. 4171–4186.
 - 14 LI XY, LI FY, PAN L, *et al.* DuEE: A large-scale dataset for Chinese event extraction in real-world scenarios. *Proceedings of the 9th CCF International Conference on Natural Language Processing and Chinese Computing*. Zhengzhou: Springer, 2020. 534–545.
 - 15 余辉, 徐畅, 刘雅茹, 等. 基于 BiLSTM-CRF 的中文临床指南治疗事件抽取. *中华医学图书情报杂志*, 2020, 29(2): 9–14. [doi: [10.3969/j.issn.1671-3982.2020.02.002](https://doi.org/10.3969/j.issn.1671-3982.2020.02.002)]
 - 16 喻雪寒, 何琳, 徐健. 基于 RoBERTa-CRF 的古文历史事件抽取方法研究. *数据分析与知识发现*, 2021, 5(7): 26–35.

(校对责编: 孙君艳)