



计算机工程  
Computer Engineering  
ISSN 1000-3428, CN 31-1289/TP

## 《计算机工程》网络首发论文

题目: 基于 CNN-BiGRU 模型的事件触发词抽取方法  
作者: 苗佳, 段跃兴, 张月琴, 张泽华  
DOI: 10.19678/j.issn.1000-3428.0058756  
网络首发日期: 2020-10-15  
引用格式: 苗佳, 段跃兴, 张月琴, 张泽华. 基于 CNN-BiGRU 模型的事件触发词抽取方法. 计算机工程. <https://doi.org/10.19678/j.issn.1000-3428.0058756>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。



# 基于 CNN-BiGRU 模型的事件触发词抽取方法

苗佳<sup>1</sup>, 段跃兴<sup>1</sup>, 张月琴<sup>1</sup>, 张泽华<sup>1</sup>

(1. 太原理工大学, 信息与计算机学院, 山西晋中 030600)

**摘要:** 在事件触发词抽取任务中, 传统方法在特征提取过程中对自然语言处理工具产生过度依赖, 耗费大量人力、容易出现错误传播和数据稀疏性等问题。针对存在的这些问题, 提出采用 CNN-BiGRU 模型来进行事件触发词抽取。该模型将词向量和位置向量拼接起来作为输入, 既能提取词级别特征也能提取句子全局特征, 可使得触发词的抽取效果达到更好。提出的模型通过 CNN 来提取词汇级别特征, 利用 BiGRU 来获取文本上下文语义信息。在 ACE2005 英文语料库和中文突发事件语料库 CEC 上的实验结果显示, 该模型事件触发词识别 F 值分别达到 74.9% 和 79.29%, 表明提出的模型有利于提升事件触发词的抽取性能。

**关键词:** 事件抽取; 触发词检测; 事件类型识别; 卷积神经网络; 循环神经网络; 双向门控循环单元; 特征提取

开放科学(资源服务)标志码(OSID):



## Event Trigger Word Extraction Method Based on CNN-BiGRU Model

Miao Jia<sup>1</sup>, Duan Yuexing<sup>1</sup>, Zhang Yueqin<sup>1</sup>, Zhang Zehua<sup>1</sup>

(1. College of Information and Computer, Taiyuan University of Technology, Jinzhong, 030600, China)

**[Abstract]** In the task of event trigger word extraction, traditional method relies too much on natural language processing tools in the feature extraction process, which consumes a lot of manpower, and is prone to problems such as error propagation and data sparseness. In response to these problems, the CNN-BiGRU model is proposed to extract event trigger word. The model combines word vectors and position vectors as input, and can extract both lexical-level features and sentence global features, which can make the extraction effect of trigger word better. The proposed model uses CNN to extract lexical-level features and BiGRU to obtain contextual semantic information. The experimental results on the ACE2005 English corpus and the Chinese emergency corpus CEC show that the F value of the model event trigger word recognition reaches 74.9% and 79.29%, respectively, indicating that the proposed model is beneficial to improve the performance of event trigger word extraction.

**[Key words]** event extraction; trigger word detection; event type recognition; convolutional neural network; recurrent neural network; bidirectional gated recurrent unit; feature extraction

DOI:10.19678/j.issn.1000-3428.0058756

## 0 概述

随着科技的飞速发展, 网络成为人们获取信息的最快途径。然而, 面对这些海量的网络信息, 人们从中挑选出真正感兴趣的信息变得异常困难。作为信息抽取<sup>[1]</sup>的重要组成部分, 事件抽取的研究具有巨大的现实意义。事件抽取<sup>[2]</sup>技术的主要作用是为用户抽取结构化信息, 从而提高用户的工作效率。事件抽取在自动文摘<sup>[3]</sup>、信息检索<sup>[4]</sup>、自动问答<sup>[4]</sup>、情报收集<sup>[5]</sup>等领域有着广泛的应用。在事件抽取任务中触发词抽取的研究很有必要, 因为触发词直接引起事件发生, 事件类型主要由触发词决定。触发词抽取的步骤就是先判断出哪个是触发词, 然后划分出触发词属于的事件类别。

近年来, 事件抽取得到了许多研究机构和研究者越来越多的关注。作为事件抽取领域最具有影响力的评测会议, 自动内容抽取(Automatic Content Extraction, ACE<sup>①</sup>)评测会议对事件<sup>[6]</sup>的定义如下: 事件是指发生了的事情, 包含直接引起事件发生的触发词和事件的参与者。本文以 S1 为例来阐述触发词抽取任务。在 S1 中, 包含了 Life\_Die 和 Conflict\_Attack 两种事件, 两种事件对应的触发词分别为 died 和 fired。

S1: In Baghdad, a cameraman died when an American tank fired on the Palestine Hotel. (在巴格达, 当一架美国坦克向巴勒斯坦酒店开火时, 一名摄影

<sup>①</sup> <http://projects.ldc.upenn.edu/ace/>

**基金项目:** 国家自然科学基金资助项目 (61503273)

**作者简介:** 苗佳(1995—), 女, 硕士研究生, 研究方向为知识图谱的构建与应用, 事件触发词抽取; 段跃兴(1964—), 男, 硕士, 副教授, 研究方向为知识图谱的构建与应用; 张月琴(1963—), 女, 硕士, 教授, 研究方向为智能信息处理; 张泽华(1981—), 男, 博士, 讲师, 研究方向为社区发现与挖掘。E-mail: 1007193557@qq.com

师死亡。)

作为信息抽取领域的重要研究方向,事件抽取近年来也得到了许多关注,国内外的研究人员也纷纷展开了对事件抽取的研究。AHN 等<sup>[4]</sup>人提出的 MaxEnt 方法采用最大熵模型进行事件抽取,将事件抽取任务分为若干串行的子任务;Liao 等<sup>[7]</sup>人提出基于跨事件推理的事件抽取模型;Hong 等<sup>[8]</sup>人提出的方法是采用跨实体推理的方法进行事件触发词的抽取;Li 等<sup>[9]</sup>人提出在基于结构化感知机的基础上,把触发词识别和元素识别结合起来,利用多种局部特征和全局特征,从而更有效地提高了抽取的效果。以上这些方法在特征提取过程中过分依赖自然语言处理工具,容易出现错误传播和数据稀疏等问题。

随着词向量在自然语言处理领域的成功应用,词向量技术吸引了大量事件抽取研究者的眼球,他们纷纷在事件抽取任务中运用了词向量技术。NGUYEN 等<sup>[10]</sup>人使用 CNN 自动挖掘隐含特征,并取得了良好的效果;CHEN 等<sup>[11]</sup>人提出了一种 DMCNN 模型,该模型增加了一个动态多池化层,使得模型能够抽取到多个事件,也达到了不错的抽取效果。除了 CNN 的特征学习模型外,还有研究者采用循环神经网络(RNN)对文本中的序列信息进行建模。NGUYEN 等<sup>[12]</sup>实现了一种双向循环神经网络的联合框架,实现了触发词与元素的联合抽取模型,并探究了多种不同的记忆特征;陈斌等<sup>[13]</sup>人提出一种事件触发词抽取方法,该方法将卷积神经网络和长短期记忆网络相结合进行事件触发词抽取。LIU 等<sup>[14]</sup>人分别利用多种有监督和无监督机制挖掘实体信息,在测试集元素未知和元素已知的条件下,研究对触发词抽取产生的作用。

本文将 CNN-BiGRU 模型运用于事件触发词抽取任务中,使得抽取到的特征更加充分,同时也达到了非常不错的抽取效果。将 CNN 和 BiGRU 模型融合起来在很多关联领域或其他领域都有运用到,郑巧夺等<sup>[6]</sup>人提出用双层 CNN-BiGRU-CRF 模型进行事件因果关系的抽取;王丽亚等<sup>[15]</sup>人提出在 CNN-BiGRU 模型中引入注意力机制从而来进行文本情感分析,他们也都取得了不错的效果。因此,将 CNN 和 BiGRU 模型进行融合是完全具有必要性和可行性的。

本文提出的模型相比传统模型特征提取更加充分,该模型通过 CNN 提取词级别特征,利用 BiGRU 来获取文本上下文语义信息得到句子级别特征,拼接词级别特征和句子级别特征,最后通过 softmax

层对提取出的事件相关特征进行分类,识别触发词并预测触发词所属事件类别。

## 1 CNN-BiGRU 模型

在本节中,将介绍一种新的事件触发词抽取模型,该模型将 CNN 和 BiGRU 融合起来进行事件触发词抽取,模型流程结构图如图 1 所示。

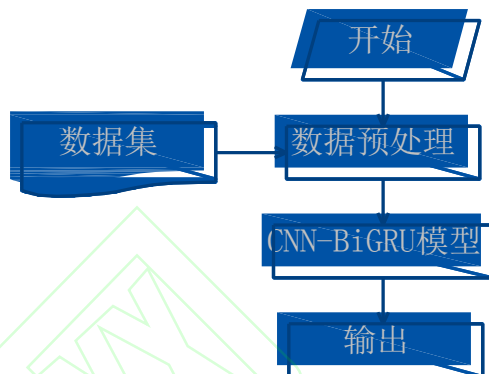


图 1 模型整体流程结构

Fig1 Model overall process structure

### 1.1 输入映射层

输入映射层是将句子序列中的每个词转换为向量表达,把文本数据表示为计算机能够理解的方式。对于一个句子序列:

$$S = \{W_1, W_2, \dots, W_n\} \quad (1)$$

其中,  $n$  表示句子的长度,即  $S$  中单词的个数,  $W_i$  表示句子中第  $i$  个单词,本文将每个单词  $W_i$  转化为实值向量  $X_i$ ,  $X_i$  是由词向量和位置向量拼接而成。

对于一个长度为  $n$  的句子序列,可表示为:

$$X = \{X_1, X_2, \dots, X_n\} \quad (2)$$

其中,  $X_i$  表示  $X$  中每个单词的特征向量。

本文将句子的最大长度设置为 50,因为在 CNN 里一般会将输入向量设定为固定的长度。如果句子长度大于 50,则截取前 50 个单词作为该句子;如果句子长度小于 50,则采用补零的方式将句子补齐。

#### 1.1.1 词向量

词向量是表示词语特征的最常用方式,具有良好的语义特征,是把深度学习相关模型引入到自然语言处理领域的一个关键技术。本文的实验采用谷歌预训练的 Word2vec 词向量表进行初始化。Word2vec 技术包括 CBOW 和 Skip-gram 两种模型。CBOW 模型是已知当前词的上下文,对当前词进行预测;Skip-gram 模型恰恰相反,是已知当前词的情况下,对其上下文进行预测。本文的训练方式采用 Skip-gram 模型,其模型结构如图 2 所示。本文采用

300 维的词向量来表示词语特征, 单词  $W_i$  的词向量表示为  $V(W_i)$ 。

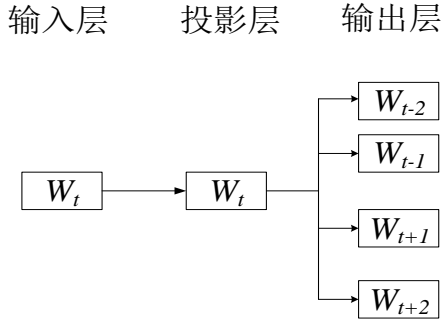


图 2 skip-gram 模型  
Fig2 skip-gram model

### 1.1.2 位置向量

位置向量<sup>[16]</sup>形成的方式是通过随机初始化, 它代表上下文单词在窗口中的相对位置。本文采用 50 维的位置向量来表示位置特征, 单词  $W_i$  的位置向量表示为  $V(pe_i)$ 。

综上所述, 单词  $W_i$  的特征向量  $X_i$  可表示为:

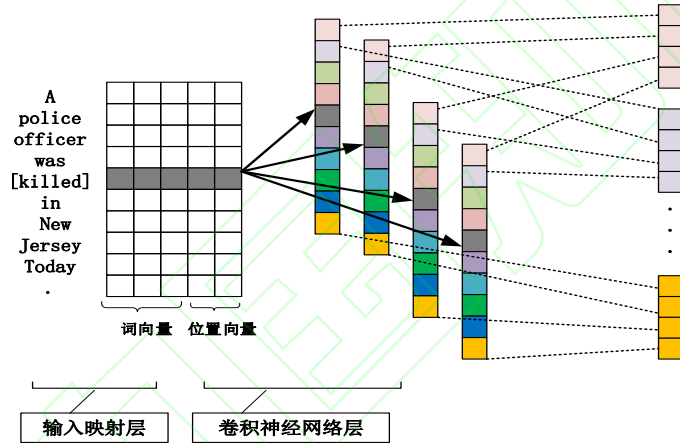


图 3 基于卷积的词级别特征提取

Fig3 lexical-level feature extraction based on convolution

设计一个  $u \in \mathbb{R}^{k \times d}$  的卷积核来抽取单词语义和相关上下文特征, 卷积核的宽度和我们的输入的宽度一样, 因为在文本处理中, 句子矩阵的每一行代表一个词的词向量, 卷积核通常覆盖上下几行的词。单词  $W_i$  通过卷积核  $u_i$  进行卷积操作, 可得到特征向量  $c_i$ , 如式 4 所示。

$$c_i = f(u \cdot X_{i-k/2:i+k/2} + b) \quad (4)$$

在式 4 中,  $X_{i-k/2:i+k/2}$  表示单词  $W_i$  周围  $k$  个单词的特征,  $f$  表示一个非线性激活函数,  $b$  表示一个偏置量,  $c_i$  代表单词  $W_i$  通过卷积操作得到的局部特征。

$$X_i = V(W_i) \oplus V(pe_i) \quad (3)$$

其中,  $\oplus$  表示将词向量和位置向量拼接在一起,  $X_i \in \mathbb{R}^d$ 。

## 1.2 特征提取

本文通过 CNN 和 BiGRU 将词级别特征与句子级别特征相结合来提高触发词的抽取效果。

### 1.2.1 CNN 提取词级别特征

CNN<sup>[17]</sup> 是一种前馈神经网络, 本文主要利用卷积神经网络的特性, 对句子中的每个词提取局部特征, 最后以特征图的形式表现出来。触发词抽取是一个由序列到标签的分类问题, 卷积就相当于采用卷积核在输入矩阵上滑动进行乘积求和的一个过程。

卷积过程如图 3 所示, 给定句子的向量表示  $X = \{X_1, X_2, \dots, X_n\}$ ,  $X_i \in \mathbb{R}^d$ 。

将卷积核扫描句子中的每个单词, 最终可得到整个句子的特征图:

$$g = [c_1, c_2, \dots, c_n] \quad (5)$$

在式 5 中,  $g \in \mathbb{R}^n$ ,  $n$  表示句子的长度。

我们设计一个卷积集合  $u = \{u_1, u_2, \dots, u_m\}$  来获取不同的特征, 本文选择卷积核大小  $k=5$ , 卷积核数量  $m=400$ 。重复以上的卷积过程, 需要根据窗口的大小, 采用补零的方式将句子边界补齐。整个句子的特征图可以表示为:

$$C = [g_1, g_2, \dots, g_m] \quad (6)$$

在式 6 中,  $C \in \mathbb{R}^{n \times m}$ ,  $C$  中每行表示当前单词通过 CNN 所提到的词汇级别特征。



### 1.2.2 BiGRU 提取句子级别特征

循环神经网络容易产生梯度消失和长期依赖等问题, 因为随着输入的逐渐增多, RNN 较难学习到连接之间的关系。为了解决这个问题, 先后提出了 LSTM 和 GRU 这两个扩展模型。本文选择 GRU<sup>[7]</sup>, GRU 在各种任务中表现出的性能与 LSTM 类似, 但 GRU 结构较为简单, 相较于 LSTM 少了一个门, 参数也比较少, 因此收敛速度比 LSTM 更快, 更加适合触发词抽取的任务。

GRU 是由 Cho 等人提出的, GRU 不使用单独的存储器单元, 而是使用门控机制来跟踪序列的状态。GRU 有重置门  $r$  和更新门  $z$  两种类型的门, 它们共同负责更新信息的方式。GRU 的单元结构如图 4 所示。

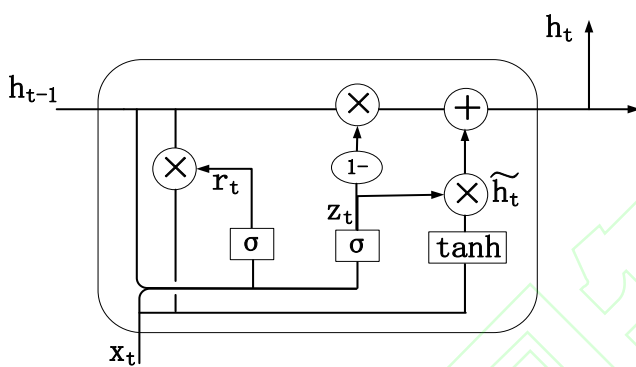


图 4 GRU 单元结构  
Fig4 GRU unit structure

在图 4 中,  $z_t$  为更新门, 用于控制保留的过去的信息量, 以及被添加的新信息量;  $r_t$  为重置门, 用于控制前一时刻状态有多少信息被写入到当前候选集  $\tilde{h}_t$  上, 如果  $r_t$  的值为 0, 则表示忘记之前所有的状态。

在  $t$  时刻, GRU 计算新状态如公式 (7) ~ (10) 所示:

$$z_t = \sigma(w_z \cdot [h_{t-1}, X_t]) \quad (7)$$

$$r_t = \sigma(w_r \cdot [h_{t-1}, X_t]) \quad (8)$$

$$\tilde{h}_t = \tanh(w \cdot [r_t * h_{t-1}, X_t]) \quad (9)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (10)$$

$w_z$ 、 $w_r$ 、 $w$  为权重参数;  $X_t$  为  $t$  时刻 GRU 的输入;  $\tilde{h}_t$  表示一个 GRU 单元在  $t$  时刻的激活值;  $h_t$  表示当前 GRU 单元的激活值;  $\sigma$  为 sigmoid 函数。

在 GRU 对文本序列建模时, 不考虑反向上下文, 不能将上下文信息全部利用起来, 因此增加反向输入的 GRU。双向 GRU 利用了两个并行通道, 使得模型既能获得正向的累积依赖信息, 又能获得

反向的未来的累积依赖信息。通过这种方式, 使得提取的特征信息更加丰富。在训练过程中, 假设 BiGRU 将隐藏层的数量设定为  $s$ , 那么 BiGRU 的输出结果就如式 11 表示:

$$H = [h_1, h_2, \dots, h_n] \quad (11)$$

其中,  $H \in \mathbb{R}^{n \times (2 \times s)}$ ,  $H$  中每一行代表一个单词通过 BiGRU 提取得到的句子级别特征。

### 1.3 输出层

由上文可知,  $C \in \mathbb{R}^{n \times m}$  为 CNN 提取到的词特征,  $H \in \mathbb{R}^{n \times (2 \times s)}$  为 BiGRU 提取到的句子全局特征, 将这两个特征进行拼接, 如式 12。这样可以将特征提取的更加充分。输出层的目的是将 CNN-BiGRU 模型提取得到的特征放入 softmax 层进行触发词预测, 如式 13。

$$y = C \oplus H \quad (12)$$

在式 12 中,  $\oplus$  表示向量的拼接,  $y \in \mathbb{R}^{n \times (m + 2 \times s)}$ 。

$$O = \text{softmax}(W_l y + b_l) \quad (13)$$

在上式中,  $W_l$  表示连接权重矩阵,  $b_l$  表示偏置。

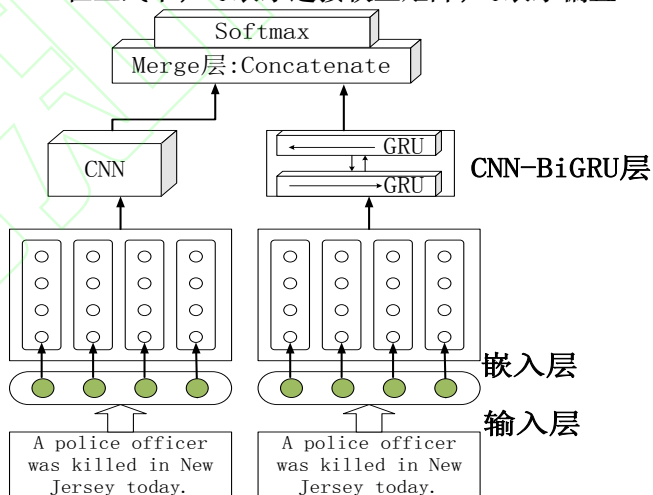


图 5 CNN-BiGRU 模型  
Fig5 CNN-BiGRU model

## 2 实验与分析

### 2.1 实验数据集

本文使用的数据集是 ACE2005<sup>②</sup>英文语料库和中文突发事件语料库 CEC, 它们是事件触发词抽取任务中最常使用的数据集。

ACE2005 英文语料库是事件抽取任务一直被广泛采用的数据集, 该语料库包含 599 篇文档, 内容涉及新闻、广播、对话、博客、论坛和电话录音等。ACE 评测会议的主要目标是识别出新闻语料中的实

<sup>②</sup> <https://catalog.ldc.upenn.edu/LDC2006T06>

体、关系和事件。在 ACE2005 英文语料库中,事件被分为 8 个大类,33 个子类,本文把识别子类作为主要任务。本文随机选取校验集 30 篇,测试集 40 篇,再将训练集选取为剩下的 529 篇。

中文突发事件语料库 CEC<sup>[18]</sup>是上海大学开源的数据集,该数据集总共包括地震、食物中毒、恐怖袭击、火灾以及交通事故 5 类新闻报道。CEC 事件语料库的标注格式是 XML 语言,该语料库对事件、事件元素以及环境等全面标注,总共包含 11908 个事件。本文实验设定训练集、校验集和测试集的比例为 7:1:2。

## 2.2 实验评价标准

事件触发词识别/分类:触发词被正确识别 $\Leftrightarrow$ 触发词在文本的位置和类型与标准标注文档中的候选事件触发词的位置和类型完全匹配。

本文与大多数事件抽取研究的评价方法一致,将准确率 (precision)、召回率 (recall) 和  $F_1$  值 ( $F_1$ -measure) 作为衡量事件抽取模型好坏的指标。precision 表示系统正确标注结果的总数与系统标注结果总数的比值。recall 表示系统正确标注结果的总数与语料中标注结果的总数的比值。 $F_1$  是 precision 值和 recall 值的一个综合度量。具体计算公式如下:

$$F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (14)$$

事件抽取模型的性能通常通过  $F_1$  值来综合判断, $F_1$  值越大,模型性能越好。

## 2.3 实验参数设置

在训练过程中,采用 softmax 函数作为分类器,进而预测出触发词类型的概率矩阵,触发词类型的判断是通过 argmax 函数,损失函数为交叉熵代价 (cross-entropy error) 函数,将 Adam 优化算法作为梯度下降算法。本文模型实验参数的设置如表 1 所示。

表 1 超参数设置值  
Table1 Hyperparameter setting value

参数名	值
词向量的维度	300
位置向量的维度	50
卷积核大小	5
卷积核数量	400
BiGRU 隐藏层数量	200

## 2.4 实验结果分析

为了验证本文模型对于实验的有效性,采用多种模型与本文模型进行对比。表 2、表 3 分别展示了在 ACE2005 英文语料库上不同模型触发词抽取性能和触发词分类性能的对比;表 4 展示了在中文

突发事件语料库 CEC 上不同模型触发词抽取性能的对比。

表 2 ACE2005 英文语料库上触发词抽取性能对比  
Table2 Comparison of trigger word extraction performance on ACE2005 English corpus

方法模型	准确率	召回率	F 值
MaxENT <sup>[4]</sup>	76.2	60.5	67.4
Li's structure <sup>[9]</sup>	76.9	65.0	70.4
DMCNN <sup>[11]</sup>	<b>80.4</b>	67.7	73.5
S-CNNs <sup>[19]</sup>	78.1	<b>71.8</b>	74.8
JRNN <sup>[12]</sup>	68.5	65.7	71.9
Our model	80.1	70.4	<b>74.9</b>

分析上表可知,传统的事件触发词抽取方法表现出来的效果普遍不如基于神经网络的事件触发词抽取方法效果要好。在触发词抽取性能中,本文模型准确率虽然较 DMCNN 低 0.3%,但召回率却提升 2.7%,F 值提升 1.4%;S-CNNs 模型召回率比本文模型高 1.4%,准确率却比本文模型低 2%,F 值比本文模型低 0.6%。总体来说,在触发词抽取性能对比中,本文提出的模型表现最好。

表 3 ACE2005 英文语料库上触发词分类性能对比  
Table3 Comparison of trigger word classification performance on ACE2005 English corpus

方法模型	准确率	召回率	F 值
MaxENT <sup>[4]</sup>	74.5	59.1	65.9
Li's structure <sup>[9]</sup>	73.7	62.3	67.5
DMCNN <sup>[11]</sup>	75.6	63.6	69.1
S-CNNs <sup>[19]</sup>	74.1	64.8	69.1
JRNN <sup>[12]</sup>	66.0	<b>73.0</b>	69.3
Our model	<b>74.9</b>	65.2	<b>69.7</b>

由表 3 分析可知,在触发词分类性能对比中,JRNN 的事件抽取方式,虽然召回率的提高非常显著,但是它的准确率却较本文模型低 8.9%,F 值较本文低 0.4%。综上所述,在触发词抽取这一任务中,本文提出的模型是目前所有模型中效果最好的。

表 4 中文突发事件语料库 CEC 上触发词抽取性能对比  
Table4 Comparison of trigger word extraction performance on Chinese emergencies corpus CEC

方法模型	准确率	召回率	F 值
SVM <sup>[20]</sup>	76.45	71.04	73.65
SVM+Embedding <sup>[21]</sup>	77.85	73.62	75.68
CNN <sup>[22]</sup>	<b>79.21</b>	76.53	77.85
Our model	79.16	<b>79.42</b>	<b>79.29</b>

分析表 4 可知,本文提出的模型达到的抽取效果是最好的,相比起现有表现最好的模型 F 值提高 1.44%。虽然本文模型准确率比 CNN<sup>[22]</sup>模型低 0.05%,但召回率却比 CNN<sup>[22]</sup>模型高出 2.89%,整体表现出的抽取效果也是最好的。

CNN 进行词级别特征提取时,本文对 CNN 重要参数调整对模型影响进行了探究,实验结果如图 6 所示。

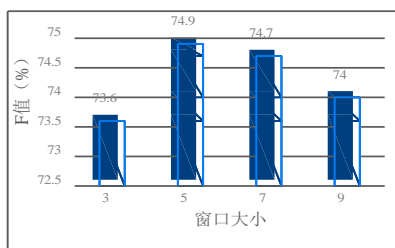


图 6 窗口大小对模型的影响

Fig6 The effect of window size on the model

由图 6 可知,当把其他参数设为定值时,卷积窗口的大小设定为 5 时,可使得模型的 F 值达到最大,效果达到最好。分析可知,当 CNN 卷积窗口过小时,不能充分利用上下文语义进行特征提取,对应的 F 值较小;当卷积窗口过大时,也不能达到最好的特征抽取效果,因为窗口过大时,不利于提取句子的局部特征。因此参数的选取对于实验也是非常重要的。

### 2.5 模型收敛特性

本文提出的 CNN-BiGRU 模型具有很好的事件触发词抽取效果,因为该模型特征提取相当充分。同时本文模型 also 具有很好的收敛特性,收敛情况对比如图 7 所示。

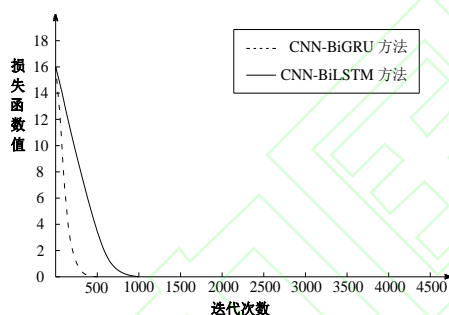


图 7 收敛情况对比

Fig7 Convergence comparison

由图 7 可知,本文提出的模型迭代次数达到 500 次左右就可收敛,CNN-BiLSTM 模型迭代次数达到 1000 次左右才可收敛。分析可知,BiGRU 在事件触发词抽取任务上的表现比 BiLSTM 效果更好,能更快能达到收敛。本质是因为 BiGRU 网络结构相对 BiLSTM 更加简单,所需参数少,能有效缩短模型训练时间。

## 3 结束语

本文提出基于 CNN-BiGRU 模型进行事件触发词抽取的方法。在模型结构上,分解为两个步骤,通过卷积神经网络提取词级别特征,再通过 BiGRU 提取句子级别特征,使得特征提取足够充分。同时 BiGRU 所需参数少,可以有效缩短模型的训练时间。

经过实验对比,相比现有其他事件触发词抽取方法,本文提出模型的抽取效果有了进一步的提升。在接下来的工作中本文将考虑尝试引入更多的有效特征作为输入来提高模型的抽取性能;另外,由于主题事件抽取将是未来的发展方向,考虑将模型扩展到主题事件抽取任务中。

### 参考文献

- [1] Zhu Muyijie, Bao Bingkun, Xu Changsheng. Research progress on the development and construction of knowledge Graph[J].Journal of Nanjing University of Information Science&Technology(Natural Science Edition),2017,9(06):575-582.  
朱木易洁,鲍秉坤,徐常胜.知识图谱发展与构建的研究进展[J].南京信息工程大学学报(自然科学版),2017,9(06):575-582.
- [2] Gao Lizheng, Zhou Gang, Luo Junyong, Lan Mingjing.Review of Research on Meta-Event Extraction[J].Computer Science,2019,46(8):9-15.  
高李政,周刚,罗军勇,兰明敬.元事件抽取研究综述[J].计算机科学,2019,46(8):9-15.
- [3] Sun Peipei, Liao Tao, Liu Zongtian. Automatic abstract extraction based on event arguments[J].Computer and Digital Engineering,2015,43(10):1829-1833.  
孙佩佩,廖涛,刘宗田.基于事件要素的自动文摘抽取[J].计算机与数字工程,2015,43(10):1829-1833.
- [4] AHN D. The stages of event extraction [C]//Proceedings of the Workshop on Annotating and Reasoning about Time and Events. Cambridge,USA:MIT Press,2006:1-8.
- [5] Qi Guilin, Gao Huan, Wu Tianxing. Research Progress of Knowledge Graph[J].Journal of Information Engineering,2017,3(1):004-025.  
漆桂林,高桓,吴天星.知识图谱研究进展[J].情报工程,2017,3(1):004-025.
- [6] Zheng Qiaoduo, Wu Zhendong, Zou Junying. Event causality extraction method based on two-layer CNN-BiGRU-CRF model [J/OL]. Computer Engineering: 1-9[2020-09-07].<https://doi.org/10.19678/j.issn.1000-3428.0057361>.  
郑巧夺,吴贞东,邹俊颖.基于双层 CNN-BiGRU-CRF 模型的事件因果关系抽取方法 [J/OL]. 计算机工程:1-9[2020-09-07].<https://doi.org/10.19678/j.issn.1000-3428.0057361>.



- [7] Liao S, Grishman R. Using document level cross-event inference to improve event extraction[C]//Proceedings of the 48th Annual Meetings of the Association for Computational Linguistics. Cambridge,USA:MIT Press,2010:789-797.
- [8] 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.Cambridge,USA:MIT Press,2011:1127-1136.
- [9] Li Q, JI H, HUANG L. Joint event extraction via structured prediction with global features[C]//Proceedings of Meeting of the Association for Computational Linguistics.Cambridge,USA:MIT Press,2013:73-82.
- [10] NGUYEN TH,GRISHMAN R. Event detection and domain adaptation with convolutional neural network[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing.Stroudsburg,PA:ACL,2015:365-371.
- [11] CHEN Y,XU L,LIU K, et al. Event extraction via dynamic multi-pooling convolutional neural networks [C]//Proceedings of Meeting of the Association for Computational Linguistics.Cambridge,USA:MIT Press,2015:167-176.
- [12] NGUYEN T H,CHO K,GRISHMAN R.Joint event extraction via recurrent neural networks [C]//Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Cambridge,USA:MIT Press,2016:300-309.
- [13] Chen Bin, Zhou Yong, Liu Bing. Event trigger extraction based on convolutional bidirectional long short term memory network[J].Computer Engineering,2019,45(1):153-158.  
陈斌,周勇,刘兵.基于卷积长短期记忆网络的事件触发词抽取[J].计算机工程,2019,45(1):153-158.
- [14] LIU Shulin, CHEN Yubo, LIU Kang, et al. Exploiting argument information to improve event detection via supervised attention mechanisms [C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics.Stroudsburg,USA:ACL,2017:1789-1798.
- [15] Wang Liya, Liu Changhui, Cai Dunbo, Lu Tao. Sentiment analysis of Chinese text with attention mechanism introduced in CNN-BiGRU network[J].Computer Applications,2019,39(10):2841-2846.  
王丽亚,刘昌辉,蔡敦波,卢涛.CNN-BiGRU 网络中引入注意力机制的中文文本情感分析[J].计算机应用,2019,39(10):2841-2846.
- [16] Liu Yiting, Li Peifeng. Event trigger word extraction based on local entity features[J].Computer Engineering,2019,45(11):213-217+224.  
柳亦婷,李培峰.基于局部实体特征的事件触发词抽取[J].计算机工程,2019,45(11):213-217+224.
- [17] Wu Wentao, Li Peifeng, Zhu Qiaoming. A joint extraction method of entities and events based on hybrid neural network[J].Journal of Chinese Information Processing,2019,33(8):77-83.  
吴文涛,李培峰,朱巧明.基于混合神经网络的实体和事件联合抽取方法[J].中文信息学报,2019,33(8):77-83.
- [18] Yi Shixiang, Yin Hongpeng, Zheng Hengyi. Public safety event trigger word recognition based on BiLSTM[J].Journal of Engineering Science,2019,41(9):1201-1207.  
易士翔,尹宏鹏,郑恒毅.基于BiLSTM的公共安全事件触发词识别[J].工程科学学报,2019,41(9):1201-1207.
- [19] ZHANG Z,XU W,CHEN Q. Joint event extraction based on skip-window convolutional neural networks[C]//Proceedings of International Conference on Computer Processing of Oriental Languages.Berlin,Germany:Springer,2016:324-334.
- [20] Sampo Pyysalo, Tomoko Ohta, Makoto Miwa, Han-Cheol Cho, Jun&apos; ichi Tsujii, Sophia Ananiadou. Event extraction across multiple levels of biological organization[J]. Bioinformatics,2012,28(18):i575.
- [21] Zhou Deyu, Zhong Dayou,He Yulan. Event trigger identification for biomedical events extraction using domain knowledge.[J]. Bioinformatics (Oxford, England),2014,30(11):1587.
- [22] Jian Wang, Honglei Li, Yuan An, Hongfei Lin, Zhihao Yang. Biomedical event trigger detection based on convolutional neural network[J]. Int. J. of Data Mining and Bioinformatics,2016,15(3):195.