

基于字词联合表示的中文事件检测方法

吴凡 朱培培 王中卿 李培峰 朱巧明

苏州大学计算机科学与技术学院 江苏 苏州 215006

(944721805@qq.com)



摘要 事件检测作为事件抽取的一个子任务,是当前信息抽取的研究热点之一。它在构建知识图谱、问答系统的意图识别和阅读理解等应用中有着重要的作用。与英文字母不同,中文中的字在很多场合作为单字词具有特定的语义信息,且中文词语内部也存在特定的结构形式。根据中文的这一特点,文中提出了一种基于字词联合表示的图卷积模型 JRCW-GCN(Joint Representation of Characters and Words by Graph Convolution Neural Network),用于中文事件检测。JRCW-GCN 首先通过最新的 BERT 预训练语言模型以及 Transformer 模型分别编码字和词的语义信息,然后利用词和字之间的关系构建对应的边,最后使用图卷积模型同时融合字词级别的语义信息进行事件句中触发词的检测。在 ACE2005 中文语料库上的实验结果表明,JRCW-GCN 的性能明显优于目前性能最好的基准模型。

关键词: 中文事件检测;联合表示;图卷积

中图法分类号 TP391

Chinese Event Detection with Joint Representation of Characters and Words

WU Fan,ZHU Pei-pe,ZHANG Zhong-qing,LI Pei-feng and ZHU Qiao-ming

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

Abstract As a sub-task of event extraction,event detection is a research hotspot in information extraction. It plays an important role in many NLP applications,such as knowledge graph,question answering and reading comprehension. Different with English character,a Chinese character can be regarded as single-character word and has its certain meaning. Moreover,there are specific structures in Chinese words. Therefore,this paper proposes a Chinese event detection model based on graph convolution neural network,called JRCW-GCN(Joint Representation of Characters and Words by Graph Convolution Neural Network),which integrates Chinese character and word representation. It uses the latest BERT and Transformer to encode the semantic information of words and characters respectively,and then uses the relationship between words and characters to construct the corresponding edges. Finally,it uses the graph convolution model to detect Chinese events by integrating the semantic information in Chinese character and word level. The experimental results on the ACE2005 Chinese corpus show that the performance of our model outperforms the state-of-the-art models.

Keywords Chinese event detection,Joint representation,Graph convolution

1 引言

事件作为信息的一种表现形式,定义为特定的人、物在特定时间和特定地点相互作用的客观事实。事件的主要内容包括:1)触发词,是句子中能够表示事件发生的核心词,多为动词或名词;2)事件类型,指事件所属的类别;3)论元,是事件的参与元素和;4)论元角色,指论元在事件中扮演的角色。本文将关注事件触发词检测任务,暂不考虑识别论元及其角色。传统的事件检测方法采用人工构建的特征来表示每个候选触发词的词法、句法、语义信息,并借助基于统计的分类模型判

断其事件类型。然而,人工特征的构建比较耗时耗力,且依赖于专家知识。

近年来,基于神经网络的特征学习方法在自然语言处理任务中展现出了巨大的优势。不同于传统的离散型特征,神经网络特征是连续型向量特征,不仅能够建模语义信息,还能自动组合构建更上层的特征。目前,神经网络模型已经广泛应用于事件检测任务,但是绝大多数应用针对的是英文,在中文事件检测任务上只有极少工作初步探索了神经网络方法的应用。

一般地,一个事件句中相对应的事件必然有与之对应的触发词。除了触发词本身的语义信息,触发词的内部结构和

到稿日期:2020-03-25 返修日期:2020-09-20 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:国家自然科学基金项目(61836007,61806137,61772354);江苏省高等学校自然科学研究面上项目(18KJB520043)

This work was supported by the National Natural Science Foundation of China(61836007,61806137,61772354) and Natural Science Foundation of Jiangsu Province(18KJB520043).

通信作者:朱巧明(qmzhu@suda.edu.cn)

构成触发词的单字的语义信息都有助于事件检测。下面以 S1 和 S2 为例进行说明。

- S1:拉莫斯接替伊瓦塞纳少将出任海军司令。
- S2:由特哈达将军接任。
- S3:他认为任何文明存在的期限,从进入宇宙起,都不超过 600 年。

事件句 S1 中的触发词是“出任”,事件句 S2 中的触发词是“接任”,虽然两个触发词不完全一致,但是在结构上具有相似性,其核心语义来自于单字“任”,都指向了现实世界中同一个“start-position”类型的事件。但是,由于 S3 中的触发词“任何”在结构上不同,核心词是“任何”,很容易将这个“任”字与前两个“任”字区别开来。另外,对于传统的词级别神经网络模型,“接任”可能在测试集中是一个未登录词,模型并不能很好地学习到“出任”和“接任”之间的相似度。已有的一些工作说明,中文的字信息也具有很丰富的语义信息(单字词),传统的词级别神经网络模型缺乏对字的语义信息的挖掘。同时,由于中文词存在很严重的未登录词问题,单纯使用词层面的语义给模型带来了很大的挑战,因此在事件检测中不仅要依靠词语的完整语义信息,还要挖掘出词语内部结构和字层面的语义信息。

为了解决上述问题,本文提出了一个基于字词联合表示的事件检测图卷积模型 JRCW-GCN。该模型将字和词看作图中的节点,其中词语节点为分类节点,字节点为非分类节点。字节点可看作词语节点的子节点,通过图卷积神经网络 GCN^[1] (Graph Convolution Neural Network)可将字节点的信息融合到词语节点中。具体地,该模型首先使用最新的预训练语言模型 BERT 去编码字级别的语义信息,使用 Transformer 模型编码词级别的语义信息,然后用编码后的输出初始化字词节点特征,最后根据之前构建好的边创建邻接矩阵,将字词节点输入到 GCN 层中融合字词之间的语义特征,并根据输出结果识别分类节点是否为特定类型的事件。实验结果证明,本文提出的模型在 ACE2005 中文语料上的性能优于目前最好的模型。

2 相关工作

作为一种有监督多元分类任务,事件检测方法包括两个步骤:特征选择和分类模型。传统事件检测方法的分类模型大多是最大熵模型和支持向量机分类模型,神经事件检测方法的分类模型是 softmax 分类模型,不同之处在于特征选择和表示方法不同。传统事件检测方法的重点在于研究人工构建特征来检测触发词。英文事件检测方法采用的具体特征包括:1)词法特征;2)句法特征;3)实体信息;4)篇章级特征。除了词法、句法、实体信息外,中文事件检测方法^[2-4]针对未登录词和分词错误问题,探索了字符信息、语义角色标注、触发词概率、零指代、触发词一致性、同义词扩展、跨语言信息等信

息的作用。

神经网络模型的特征学习能力在英文事件检测任务中得到了验证。英文神经事件检测方法^[5-8]将预训练的词向量作为每个单词的初始向量表示,建模了单词的语义和语法信息。除了词向量表示外,其还通过向量化实体类型特征和位置特

征融入了更丰富的特征。至于上层特征学习,文献[9]使用卷积神经网络(Convolutional Neural Network,CNN)学习短语级别特征,并通过循环神经网络(Recurrent Neural Network,RNN)^[10-11]学习句子级别的特征。此外,研究者们还探索并验证了论元信息^[12]、外部语义知识^[13]、跨语言信息^[14]、依存信息^[15]等对于事件检测的效果。

现有的中文神经网络事件抽取方法初步探索了词级别的信息和字级别的信息对事件检测的作用^[16-17]。由于字序列相对较长,受限算力模型和模型架构,早期的大部分研究仍着重于研究基于词语表示进行事件检测,但是中文字和词的微妙关系决定了单纯使用字或单纯使用词层面的语义都不能有效地完成中文事件检测任务。

随着算力和模型的日益改善,研究者开始探索词语表示和字表示的区别。Zeng 等^[16]通过两种方式来解决中文事件检测任务,即基于词的分类模型和基于字的序列化标注模型,并比较了两种模型的效果。两种模型的特征表示分别用基于词的混合表示模型和基于字的混合表示模型来解决。还有研究尝试将字词信息融合来完成中文事件检测任务。Lin 等^[17]提出基于字的事件块模型来解决词和触发词的不匹配问题。在基于字的事件块模型中,他们采用 DMCNN 的框架分别学习基于词的表示和基于字的表示,并将词表示融合到字表示中。Ding 等^[18]提出通过对预定义的语义词汇表融入外部知识,同时使用一种基于触发词感知的层次化结构网络来同时融合字词信息,以进行触发词识别和事件类型分类。

3 中文事件检测模型

传统的中文事件抽取一般只考虑了词语级别的语义信息。近年来,一些方法研究了字级别模型,其由于编码能力偏弱,仍存在许多问题。其中最重要的是,传统的字编码不足以解决中文的歧义问题。最新的 BERT 中文字模型很好地弥补了该不足。基于此,本文提出了全新的字词融合模型,在不破坏原有 BERT 模型的基础上构建了一个字词模型。

事件检测需要检测出一个特定的词语作为触发词,单纯使用字级别的编码在实验中存在触发词边界问题。一方面,当把事件检测看作一个序列标注问题时,由于中文触发词的边界模糊不清,因此会降低模型的准确率。并且,由于事件检测任务不同于 NER 的序列标注任务,在实际的标注语料中存在很多无类别词语,这导致了标签的稀疏性,因此在实际应用中其效果并不是很理想。

另一方面,一些词语单纯从字符层面很难推断其语义信息,如“落入法网”,从单个字很难推断其正确的语义信息,但是当我们把它作为一个成语时,就能很容易地推测出它属于事件类型“逮捕入狱”。因此,在中文事件检测模型上更倾向于将字的语义信息和词语级别的语义信息相融合,然后对词语进行分类,其效果往往更好。

本文提出了一种基于字词联合表示的神经网络模型 JRCW-GCN,使用 BERT 预训练模型编码字级别的语义信息,使用 Transformer 编码层来捕获词语级别的语义信息,最终用图卷积融合字词的语义信息进行分类。JRCW-GCN 模型的架构如图 1 所示。

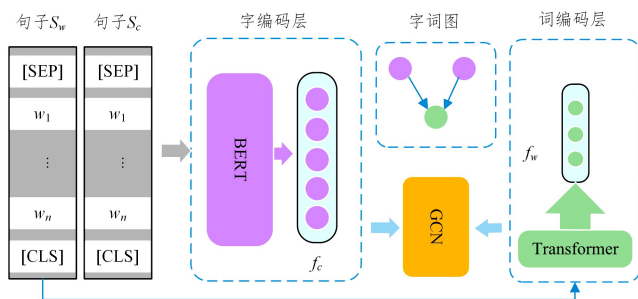


图1 JRCW-GCN模型

Fig.1 JRCW-GCN Model

首先给定一个原始句子 S , 我们使用分词工具将其切分成 n 个词语序列 $S_w = (w_1, w_2, w_3, \dots, w_n)$ 。同样地由词语序列能得到每个词的字序列 $w_1 = (c_{11}, c_{12}, \dots, c_{1m})$, 由此展开句子的字序列 $S_c = (c_{11}, c_{12}, \dots, c_{nm})$ 。分别将 S_c 和 S_w 输入到对应的编码层中进行编码, 输出分别为 f_c 和 f_w 。由于维度不同, 需要经过线性层进行维度一致化, 将词语表示和字表示投射到同一维度空间。最后, 根据分词结果构建相应的图, 使用字词向量初始化图中的节点, 并将图节点以及邻接矩阵输入到对应的图卷积层, 经过两层图卷积层获得最终的词语节点向量表示并进行分类。

3.1 词编码

Transformer^[19] 是最新的 encode-decode 框架, 本文只使用 Transformer 的 encode 部分进行编码。Transformer 中使用的注意力模型, 即文献[19]所提到的多头注意力机制。

词编码层的架构如图2所示, 首先将分好词的句子序列输入到词编码层, 使用预训练好的词向量 0 进行初始化(初始化的词向量具有更多的语义信息)。同时, 使用自注意力机制可以捕获不同单词间的语义关联信息, 进而可以更好地建模相应的词语表示, 最终模型的输出为词语向量 f_w , 该向量作为后续输入到字词图中的词语节点特征。

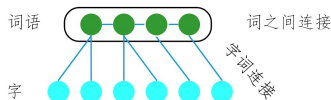


图2 字词图(电子版为彩色)

Fig.2 Word and character graph

3.2 字编码

BERT 模型^[20] 是当下最热门的语言模型。BERT 创新性地提出了一种新的预训练目标, 即遮蔽式语言模型, 用于消除传统的语言模型单向性的局限性。遮蔽式语言模型是指在预训练时随机地遮蔽一部分单词, 然后让模型预测被遮蔽的单词。最新的实验结果证明, BERT 预训练模型能够很好地捕获语法、语义层面的信息。中文字蕴含的语义往往是很复杂的, 而 BERT 强大的编码能力正好能够补充这一点。

如图1所示, 本文不改变 BERT 的模型架构, 直接用 BERT 作为一个巨大的编码器; 同时, 在使用谷歌预训练好的参数模型的基础上, 原始句子经过微处理后被对应地输入到 BERT 编码器中。由于 BERT 在捕获序列信息时的能力不如 BiGRU, 因此本文在 BERT 的输出层后面加了一层 BiGRU, 用于捕获序列信息。如图1所示, BERT 的输出结果为 f_c , 作

为后续图中的字节节点特征向量。

3.3 字词融合微调 BERT

字和词的语义表示分别为 f_c 和 f_w , 为了将字词表示投射到同一维度空间, 先经过一层线性变化层将其维度变化到相同维度得到 f'_c 和 f'_w 。如图2所示, 可以将 f'_c 和 f'_w 代入节点作为节点特征, 形成一张图。其中, 词语节点(蓝色节点)作为分类节点, 其为最小的分类单位, 而由于字节节点的颗粒度过小, 对其进行分类会影响整体的性能, 因此字节节点(绿色节点)作为非分类节点。同时, 前后词语之间存在相应的边用于捕获上下文的语义信息。

词语节点从其对应的字节节点中捕获局部特征帮助分类, 如“打死”中“打”为一个代表性的“攻击”事件触发词(单字词), 传统的事件检测模型更倾向于从“打”字中捕获语义信息。同时使用边来控制词语之间语义信息的流动, 以捕获上下文的语义信息(传统 RNN 序列信息流动的作用)。

最后, 利用多次迭代训练边的权重使模型能够捕获不同字对词语的语义贡献程度, 更大程度地解决了词语的分歧问题。另外, JRCW-GCN 模型通过对 BERT 的微调使其更好地适应图卷积的模型框架。

3.4 图卷积层

本文根据文献[21]中的图卷积方法, 对节点特征信息进行卷积操作。本文使用多层的图卷积层, 其计算公式如下:

$$h_i^{(l)} = \sigma \left(\sum_{j=1}^n A_{ij} W^{(l)} h_j^{(l-1)} + b^{(l)} \right) \quad (1)$$

其中, A 为邻接矩阵, $A_{ij} = 1$ 代表节点 i 到节点 j 之间有一条单向边。 W 为权重矩阵, b 为偏置矩阵。 $h_j^{(l-1)}$ 为前一层隐层的状态, $h_i^{(l)}$ 为当前层隐层的状态。其中, 第一层使用字编码层和词编码层的输出作为初始化的隐藏状态。通过多层的图卷积动态地将字节节点信息融入到其对应的词语节点中。

图中所有的节点特征经过多层的图卷积层, 为了防止过拟合, 多层图卷积层中额外使用了 Felu 激活函数和 Dropout, 最后经过多层图卷积输出的节点特征为 O 。

3.5 输出分类层

本文模型将图卷积网络层的输出 O 作为最后的分类向量, 使用交叉熵损失函数计算损失进行反向传播。假设句子输出 $O = (o_1, o_2, \dots, o_m)$, 对应的事件类型序列为 $Y = (y_1, y_2, \dots, y_m)$, 则目标函数如下:

$$J(\theta) = - \sum_{i=0}^n \log p(y_n | o_n, \theta) \quad (2)$$

该模型使用了 Dropout 策略和权重衰减等策略来防止过拟合。为计算模型参数, 本文通过 Adam 算法最小化目标函数, 并且在对 BERT 进行微调时对不同层进行学习率的衰减, 上层的衰减相对较慢。

4 实验

4.1 实验设置

本文使用了 ACE2005 中文数据集进行实验。ACE2005 语料集包括 633 篇来自广播新闻、微博、新闻电讯的文档。本文和文献[6]相同, 选择 66 篇文档作为测试集, 并从剩下的 567 篇文档中随机选择 33 篇作为开发集, 将剩余的 534 篇作为训练集。

本文的词向量使用文献[22]所提到的开源的预训练词向量,其他相关的超参数如表 1 所列。

表 1 实验超参数

Table 1 Hyper parameter

超参数	数值
学习率	5×10^{-5}
Batch size	32
Transformer 层数	6
Dropout	0.1
词向量维度	300
BERT 字向量维度	768
节点特征维度	768
训练轮数	40

与文献[2]相同,本文采用如下评价标准来判断触发词和论元识别分类正确与否。

(1)触发词识别正确的前提是其位置匹配到一个标准触发词的位置。

(2)事件类型分类正确的前提是触发词的位置和事件类型匹配到一个标准触发词的位置和类型。

本文采用准确率(Precision, P)、召回率(Recall, R)和 F1(F)值作为评价指标。

4.2 实验结果

本文选择如下模型作为基准模型。

(1)Rich-C,是文献[4]中的包含丰富语言学特征的传统事件检测模型,特征包括语义角色标注、触发词概率、零指代、触发词一致性等。

(2)HNN,是文献[23]中的神经事件检测模型,该混合模型同时采用双向 LSTM 和 CNN 模型学习特征。

(3)HNN-errata,是文献[16]中的神经事件检测模型,该模型是基于双向 LSTM 和 CNN 的混合模型。不同于 HNN,该模型加入了触发词勘误表以修正事件类型。触发词勘误表记录了训练集中常见的触发词及其事件类别,若勘误表中触发词出现在测试集中,则直接将该词的事件类别设定为勘误表中对应的事件类别。

(4)NPN,是文献[17]提出的基于混合表示的神经网络模型。该文用 RNN 学习句子和词语中的序列信息,并使用词语的混合表示作为特征。其使用词语生成器尝试解决因为分词错误导致无法识别触发词的问题。

(5)TLNN,是文献[18]提出的可以同时解决触发词不匹配和多语义触发问题的模型,其借助分层表示学习和触发感知特征提取器,有效地利用了多颗粒信息并学习了深层的语义特征。

将本文提出的中文事件检测模型和以上模型进行对比,具体实验结果如表 2 所列。由表 2 可知,与基准实验结果相比,JRCW-GCN 模型的性能有了较大的提升。相比 TLNN 模型,JRCW-GCN 模型在触发词识别上的效果更优,由于 TLNN 融入了外部的 HowNet 知识,因此其在分类任务上效果显著,在事件类型分类上的差距不是很大。对比一些先前的模型,JRCW-GCN 的效果明显优于它们。同时,我们基于 BERT 模型搭建了一个简单的基准模型。与单纯使用 BERT 模型时的实验效果相比,JRCW-GCN 性能仍有所提升。由于 BERT 缺乏词边界信息,存在对边界识别错误的问题,因此在

最终结果上不如 JRCW-GCN,这进一步证明了字词融合模型优于单纯使用字序列模型进行序列化标注任务。

表 2 本文模型与基准模型的对比

Table 2 Comparison of proposed model with baseline

模型	触发词识别			事件类型分类		
	P	R	F	P	R	F
Rich-C	62.2	71.9	66.7	58.9	68.1	63.2
HNN	74.2	63.1	68.2	77.1	53.1	63.0
HNN-errata	76.0	63.8	69.3	69.8	59.9	64.5
NPN	64.8	73.8	69.0	60.9	69.3	64.8
TLNN	67.3	74.7	70.8	64.5	71.5	67.8
BERT	76.1	66.6	71.0	71.7	62.7	66.9
JRCW-GCN	80.8	65.7	72.5	76.0	61.8	68.2

表 2 中,JRCW-GCN 具有较高的 P 值,对比其他模型,它挖掘出更多的中文词语内部字的信息,既融入了 BERT 的字编码的先验语言知识,又融入了词语信息。因此,与先前的模型相比,JRCW-GCN 更容易识别一些低频词和相似词,这是由中文的特点决定的:语义相近的词语具有更接近的内部结构,如“出任”和“接任”的词语相似度明显很高,但是由于 OOV(Out of Vocabulary)问题,词语模型往往不能很好地识别这类词,而 JRCW-GCN 可以根据词语的子节点(字节节点)的特征分析其词语的相似度。

4.3 实验分析

JRCW-GCN 的几个简化模型及其对比实验如表 3 所列。其中,Char_only 为仅保留字编码信息(即图 1 中的 BERT 编码部分)的字表示模型,Word_only 为只保留词编码信息(即图 1 中的 Transformer 编码部分)的词表示模型,JRCW 为去掉字词图卷积(即图 1 中的字词图和 GCN 部分)而只保留字编码信息和词编码信息的字词模型。

表 3 对比实验

Table 3 Comparative experiment

模型	触发词识别			事件类型分类		
	P	R	F	P	R	F
JRCW-GCN	80.8	65.7	72.5	76.0	61.8	68.2
Word_only	70.3	54.0	61.1	65.6	50.4	57.0
Char_only	77.9	61.8	68.9	74.4	59.1	65.8
JRCW	72.0	70.0	71.0	68.1	66.6	67.2

单独使用词表示的模型 Char_only 与单独使用字表示的模型 Word_only 相比较,字表示模型的性能更好,有大幅提升。相比字模型,词语表示存在较大的 OOV 问题,测试集中约有 5%的触发词是未登录词或者稀有词,单纯的词语表示大大影响了对这些词语的探测,而使用字表示模型基本不存在这样的问题,而且借助强大的预训练模型,单纯使用字表示的模型也取得了相对较好的实验结果。

单纯将字词融合而没有考虑字词关系的模型 JRCW,相比词表示模型 Char_only 和字表示模型 Word_only 在性能上都有所提升。一方面,单纯使用字表示模型时,由于缺乏词语的语义信息,对一些单纯使用字表示的词很难进行判断,如“进口”,单独的“进”和“口”都有其不同的语义,但是当两个字和在一起时,就具有更深层次的语义信息;同样地,单纯使用词表示模型时,由于无法捕获词语内部字特征的信息,因此很难泛化其相似的词语。

本文模型 JRCW-GCN 比 JRCW 在触发词识别和事件类

型分类任务的性能上分别有1个点的提升,这证明了深度融合字词信息的重要性,也验证了本文方法的有效性。JRCW-GCN创新性地使用图的结构表示字词结构,其作为词语内部的字具有对词语解释补充的作用,因此字词图表示能够很好地保留字词的内部结构信息,同时也一定程度地保留了词的各种特性,能更好适应类似事件检测的词组级分类任务。

结束语 本文提出了一种基于字词联合表示的模型来进行中文事件检测任务,在充分考虑中文语言学知识的情况下,同时捕获了词级别的语义信息以及字级别的语义信息。巧妙地将这两者信息进行融合,一定程度上有助于解决一些上下文语义相似度过高的触发词分类,也能一定程度地解决未登录词的问题。事实证明,相比中文事件检测最好的模型,JRCW-GCN模型性能在事件触发词识别和触发词分类上都有了一定的提升。

参考文献

- [1] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[J]. arXiv:1609.02907, 2016.
- [2] LI P, ZHU Q, DIAO H, et al. Joint modeling of trigger identification and event type determination in Chinese event extraction [C]// Proceedings of COLING 2012. 2012:1635-1652.
- [3] LI P, ZHOU G, ZHU Q, et al. Employing compositional semantics and discourse consistency in Chinese event extraction[C]// Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. 2012:1006-1016.
- [4] CHEN C, NG V. Joint modeling for chinese event extraction with rich linguistic features[C]// Proceedings of COLING 2012. 2012:529-544.
- [5] GRISHMAN R, WESTBROOK D, MEYERS A. Nyu's english ace 2005 system description[J]. ACE, 2005, 5.
- [6] JI H, GRISHMAN R. Refining event extraction through cross-document inference[C]// Proceedings of ACL-08: HLT. 2008: 254-262.
- [7] 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. Association for Computational Linguistics, 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. Association for Computational Linguistics, 2011: 1127-1136.
- [9] COLLOBERT R, WESTON J, BOTTOU L, et al. Natural language processing(almost) from scratch[J]. Journal of Machine Learning Research, 2011, 12(Aug): 2493-2537.
- [10] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [11] CHO K, VAN MERRIËNBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. arXiv:1406.1078, 2014.
- [12] LIU S, CHEN Y, LIU K, 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. 2017:1789-1798.
- [13] LIU S, CHEN Y, HE S, et al. Leveraging framenet to improve automatic event detection[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. 2016: 2134-2143.
- [14] LIU J, CHEN Y, LIU K, et al. Event detection via gated multi-lingual attention mechanism[C]// Thirty-Second AAAI Conference on Artificial Intelligence. 2018.
- [15] NGUYEN T H, GRISHMAN R. Graph convolutional networks with argument-aware pooling for event detection[C]// Thirty-Second AAAI Conference on Artificial Intelligence. 2018.
- [16] ZENG Y, YANG H, FENG Y, et al. A convolution BiLSTM neural network model for Chinese event extraction[M]// Natural Language Understanding and Intelligent Applications. Springer, Cham, 2016: 275-287.
- [17] LIN H, LU Y, HAN X, et al. Nugget Proposal Networks for Chinese Event Detection[J]. arXiv:1805.00249, 2018.
- [18] DING N, LI Z, LIU Z, et al. Event Detection with Trigger-Aware Lattice Neural Network[C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019:347-356.
- [19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is All you Need[C]// Neural Information Processing Systems. 2017:5998-6008.
- [20] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv:1810.04805, 2018.
- [21] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[J]. arXiv:1609.02907, 2016.
- [22] LI S, ZHAO Z, HU R, et al. Analogical reasoning on chinese morphological and semantic relations[J]. arXiv:1805.06504, 2018.
- [23] FENG X, QIN B, LIU T. A language-independent neural network for event detection[J]. Science China Information Sciences, 2018, 61(9): 092106.



WU Fan, born in 1996, postgraduate. His main research interests include natural language processing and so on.



ZHU Qiao-ming, born in 1963, Ph. D, professor, Ph.D supervisor, is a member of China Computer Federation. His main research interests include natural language processing and so on.