

文章编号: 1003-0077(2019)06-0057-07

基于序列到序列模型的事件识别

张俊青, 孔 芳

(苏州大学 计算机科学与技术学院, 江苏 苏州 215006)

摘 要: 事件识别是以事件为单位进行信息抽取的起点, 对后续各个子任务都意义重大。针对事件识别任务, 该文提出了一种融入文档信息的序列到序列方法, 一方面借助神经网络减少了特征工程产生的人工依赖, 另一方面借助注意力机制将局部的词、实体与全局的文档中事件的共现等信息统一建模。在 LDC2017E02 语料上实验结果表明, 该方法能有效提高事件识别的性能。

关键词: Seq2Seq; 事件识别; 注意力机制

中图分类号: TP391 **文献标识码:** A

Event Detection Based on Sequence-to-Sequence Model

ZHANG Junqing, KONG Fang

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

Abstract: Event detection is the start point and impacts much on subsequent subtasks of information extraction. Towards better event detection, this paper proposes a sequence-to-sequence approach considering additional document-level information. It reduces the dependency on manual feature engineering by means of neural network, constructing an unified model combining local words, entities and global events co-occurrence in one document by the attention mechanism. Experimental results on LDC2017E02 corpus show that this method can effectively improve the performance of event recognition.

Keywords: Seq2Seq; event detection; attention mechanism

0 引言

事件抽取是当前自然语言处理(natural language processing, NLP)领域的研究热点之一, 具有重要的理论及应用价值, 不仅推动数据挖掘和机器学习等学科理论的发展, 而且其研究成果被广泛应用于信息检索等领域。事件抽取首先进行事件识别, 即判别给定句子是否表示事件描述, 如果是, 则对事件描述进行类别的判定; 识别出事件后再寻找事件的各个参与对象, 并为他们分配相应的语义角色。显然, 事件识别是事件抽取的基础, 它的性能将直接影响后续事件参与对象的语义角色的识别。

具体而言, 事件识别又包含两个子任务: 事件触发词识别和事件类型确定。例如, 给定句子,

“Portillo acknowledged he had killed two of his former students.”事件触发词识别将识别出触发词“killed”触发了一个事件, 而事件类型确定过程则将该事件归为“Life. Die”类别。基于事件识别的结果, 后续事件论元的识别将进一步识别出该“Die”事件的攻击者是“Portillo”, 受害者是“two of his former students”。

目前, 事件识别所用的语料主要有 Doddington 等^[1]使用的 ACE 数据, Kim 等^[2]使用的 BioNLP 分享的生物医疗数据和 Dubbin 等^[3]使用的 TACKBP 数据。事件抽取相关的研究方法主要分为基于特征的方法和基于词表示的方法两类。其中, 基于特征抽取的研究工作, 大多数学者是基于句子级别信息进行事件抽取, 代表性的工作有 Ahn 等^[4]、Grishman 等^[5]、Li 等^[6]。然而, 对于有些事件, 只考虑句

收稿日期: 2018-07-26 定稿日期: 2018-09-25

基金项目: 国家自然科学基金(61836007, 61876118)

子级别信息很难判断事件所属的类型,于是部分学者把目光放转向跨句子的事件推理,如 Gupta 等^[7],Hong 等^[8],Liao 等^[9],以及跨文档的事件推理,如 Ji 等^[10]。近年来,深度学习在自然语言处理任务中表现优异,众多学者也开始专注基于词表示方法的事件识别研究,代表性工作包括:Nguyen 等^[11]利用卷积神经网络来自动学习一些特征完成触发词的识别;Chen 等^[12]在管道框架上,利用动态多池化卷积神经网络进行事件抽取;Nguyen 等^[13]构建双向循环神经网络,利用记忆单元捕获论元和触发词之间的依赖关系,同时完成论元和触发词的抽取工作;Duan 等^[14]通过循环神经网络利用词向量和文档向量的方法提高事件识别的性能。

基于句子级别的事件识别^[15-16],虽然有其可取之处,但忽略了文档中的其他信息,限制了事件识别的性能。例如,给定句子,“Elop is the one that needs to go^①”。基于句子级别的事件识别模型在识别该句子时,由于信息不足,不能充分理解 Elop 是从公司离职的语意,难以把触发词“go”识别为 End-Position 事件。但如果充分考虑事件的上下文信息,比如该文档主旨是 Stephen Elop 从 Microsoft 离职,那么触发词“go”触发的事件类型就比较容易判別了。

此外,语料统计分析发现,相似的事件经常出现在相同的文档中,比如 Attack, Injure, Die 事件经常出现在同一个文档,但是很少和 Transfer Money 事件同时出现。换言之,一个文档中发生的事件具有一种很强的联系。

综上所述,为了减少因抽取大量特征而消耗的时间与精力,并把文档信息应用于事件抽取中,本文针对事件识别任务展开,提出了一个序列到序列(Seq2Seq)的双向循环神经网络模型,并在模型中引入注意力机制,对局部的词、实体和全局的文档信息进行统一向量化。在 KBP2016 测试集上的实验证明了本文方法的有效性,实验结果表明,触发词识别的 F_1 值达到了 58.56%,比实验基准 1 和基准 2 分别提高了 3.76% 和 2.60%,而事件类型判别的 F_1 值达到了 51.82%,比实验基准 1 和基准 2 分别提高了 4.83% 和 1.19%。

1 任务描述

根据 TAC-KBP2016 的定义,事件是发生的事情或导致某种状态的改变,共有 8 大类型和 18 个子

类型,具体如表 1 所示。

表 1 TACKBP2016 数据集上定义的事件类型

Type	SubType	Type	SubType
Conflict	Attack	Movement	Transport, Artifact
	Demonstrate		Transport, Person
Contact	Meet	Justice	Arrest-Jail
	Correspondence	Manufacture	Artifact
	Broadcast	Life	Injure
	Contact		Die
Transaction	Transfer Money	Personnel	Start Position
	Transaction		End Position
	Transfer Ownership		Elect

下面引入 KBP 测评中与事件相关的术语。

实体: 现实世界中一个对象或者一组对象。

触发词: 可以清晰表达出事件发生的主要词汇。

事件类型: 事件所属的类型。

真伪性: 辨别事件所属的 REALIS 类别(Actual, Generic, Other)。

本文着重于事件识别任务,具体包括事件触发词识别和事件类型确定两个子任务。例如,针对句子“Portillo acknowledged he had killed two of his former students”,事件识别系统需要识别出触发词“killed”和其触发的事件类型“Life, Die”。根据先前的工作^[8],我们把类型分成 18 子类型和 None 类型。

2 模型

本节将详细地描述 Seq2Seq 模型。假设 $W = w_1 w_2 \cdots w_n$ 代表句子, n 代表句子的长度, w_i 表示句子的第 i 个词。 $E = e_1 e_2 \cdots e_k$ 表示在该句子中的实体, k 表示句子中实体的数量。本文将事件识别任务看作对触发词及其他词的序列标注问题,对于 w_i 词,对其分配子类型标签(18 类之一)或 None 类型标签。

2.1 句子编码

句向量是 Seq2Seq 模型的基础,本节将详解句向量的构建过程。将句子 W 编码成向量 \mathbf{X} , \mathbf{X} 中包含如下四个方面信息:

① 从“ENG_DF_001471_20140717_G00A0FRCS”文档中选取

(1) W 中 w_i 的词向量, 词表示可以自动学习词之间隐藏的丰富特征, 将从预训练好的词向量中查找^①。

(2) E 中 e_j 的实体向量, 使用 Stanford 工具包^②对句子进行实体识别, 向量随机初始化。

(3) W 中 w_i 词性向量, 词性是用 Sanford 工具包进行抽取的, 并采取随机初始化向量的方法构建词性向量表。

(4) W 的文档向量, 文档还有句子之外其他的信息, 并且一个文档中所发生的事件总是相关的, 比如 Attack、Die 和 Injure 事件, 很有可能发生在同一个文档中。本文通过 PV-DM 模型^[17]训练文档向量, 由使式(1)最大化所得:

$$\frac{1}{n} \sum_{t=m}^{n-m} \log P(w_t | w_{t-m}, \dots, w_{t+m}, \text{doc}) \quad (1)$$

其中, w_1, w_2, \dots, w_n 是文档中包含的词, w_{t-m}, \dots, w_{t+m} 是 w_t 的上下文, m 的窗口大小为 7, doc 是包含训练数据的向量, 与词向量维度相同的随机初始化向量。

预测的工作主要通过多元分类器完成, 例如 softmax 算法, 如式(2)所示。

$$p(w_t | w_{t-m}, \dots, w_{t+m}, \text{doc}) = \frac{e^{y_{w_t}}}{\sum_i e^{y_{w_i}}} \quad (2)$$

式(2)中, y_{w_i} 是词 w_i 非正则化的统一概率, 计算方法如式(3)所示。

$$y = b + Uh(w_t | w_{t-m}, \dots, w_{t+m}, \text{doc}) \quad (3)$$

式(3)中, b, U 是 softmax 参数, h 是将 $w_t | w_{t-m}, \dots,$

w_{t+m} , doc 级联或求平均得到。由于 doc 是共享的, 每次训练中输入都包含该向量, 因此能表达文档的主旨信息。

综上可知, 最终的输入向量是由上述向量拼接组合而成, 把句子 W 映射成向量 X , 作为 Seq2Seq 模型的输入。

2.2 Seq2Seq 模型

输入的向量 $X = (x_1, x_2, \dots, x_n)$, 输出的标签 $Y = (y_1, y_2, \dots, y_n)$ 。对于编码阶段, 把向量 X 输入到双向循环神经网络中, 总体模型如图 1 所示。在前馈网络中, 得到隐节点 $\vec{h}_i = f(\vec{h}_{i-1}, x_i)$, \vec{h}_i 表示当前隐节点, \vec{h}_{i-1} 表示前一隐节点, x_i 表示当前输入的向量。后馈网络中, 隐节点 $\overleftarrow{h}_j = f(\overleftarrow{h}_{j-1}, x_j)$, \overleftarrow{h}_j 表示前一个隐节点, \overleftarrow{h}_{j-1} 表示前一隐节点, x_j 表示当前输入的向量。由于是双向循环神经网络, 故编码阶段最终的隐节点为 $h_i = [\vec{h}_i; \overleftarrow{h}_i]$ 。

由于我们使用了 BahdanauAttention^[18] 注意机制的 Seq2Seq 网络模型, 输出序列为式(4)。即隐藏状态 s_i 、语义向量 c_i 和时间 $i-1$ 的输出 y^{i-1} 经过非线性变换得到时间 i 的输出 y_i 。

$$P(y_i | y_1, \dots, y_{i-1}, X) = g(y_{i-1}, s_i, c_i) \quad (4)$$

对于隐藏状态 s_i , 经隐藏状态 s_{i-1} 、语义向量 c 和时间 $i-1$ 的输出 y_{i-1} 非线性变换得到, 如式(5)所示。

$$s_i = f(s_{i-1}, y_{i-1}, c_i) \quad (5)$$

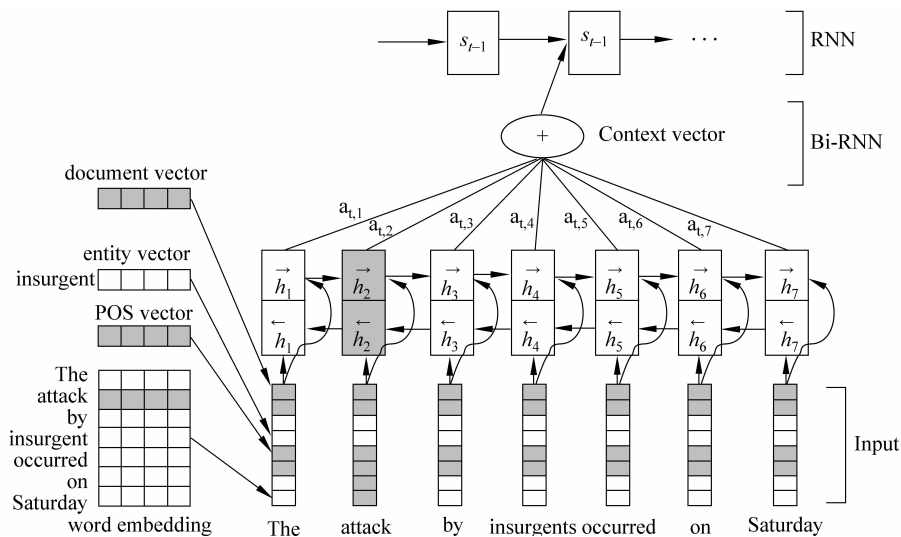


图 1 输入句子“The attack by insurgents occurred on Saturday”后, 识别事件 Conflict, Attack 示意图

① <https://github.com/stanfordnlp/GloVe>

② <https://stanfordnlp.github.io/CoreNLP>

隐藏状态 c_i 等于 T_x 个输入向量与其权重相乘求和,具体如式(6)所示。

$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j \quad (6)$$

式(6)中权重向量 α_{ij} 由式(7)得到,式(7)中 m_{ij} 由式(8)得到,式(8)中函数 a 通常为非线性函数。

$$\alpha_{ij} = \frac{\exp(m_{ij})}{\sum_{k=1}^{T_x} \exp(m_{ik})} \quad (7)$$

$$m_{ij} = a(s_{i-1}, h_j) \quad (8)$$

简而言之,相互独立的四种向量拼接成输入向量 \mathbf{X} ,向量 \mathbf{X} 进入 Seq2Seq 模型中,注意力机制通过给四种向量不同的权重以表达句子不同成分对事件识别的影响,再经过解码阶段得到相应的序列类别,根据交叉熵得到预测与标准类别的损失值,同步更新权重。

3 实验

3.1 基准平台

参加 KBP2016 测评队伍中,触发词识别 F_1 值最高的为 Mihaylov 和 Frank^[19] 的模型,触发词分类 F_1 值最高的为 Lu 和 Ng^[20] 的模型;Yang^[21] 等将事件识别当作序列化标注任务,在 KBP2016 测试集结果表明,其方法达到了先进水平。因此,本文选择了两类基准平台。

基准 1 Mihaylov 和 Frank 把事件触发词识别当作序列化标注任务,使用词、词性与依存关系等构建词向量,作为双向 LSTM 模型的输入向量,通过 softmax 进行标签分类,在 KBP2016 语料上触发词的识别 F_1 值为 54.80%。Lu 和 Ng 的 UTD1 模型先对训练集测试集进行词性还原,把测试集中词的词性还原和训练集中词的词性还原相同的词作为候选触发词;其次利用 Stanford CoreNLP 抽取候选触发词的语义依存关系,对于动词性候选触发词,提取该词的主客体头部信息;对于名词性候选触发词,提取该词的施事与受事头部信息;最后把提取的信息作为 k 近邻模型输入特征,对候选触发词进行识别与分类。在 KBP2016 语料上对触发词分类结果的 F_1 值为 46.99%。

基准 2 Yang 等首先使用最大池化的卷积神经网络获得字表示,其次利用双向循环神经网络获得词表示,最后通过 CRF 进行推理获得最终的标注

标签。他们的方法在 KBP2016 语料上进行事件识别评估,获得的触发词识别与分类的 F_1 值分别为 55.96% 和 50.63%。

3.2 数据集

本文使用的语料为官方提供的 LDC2017E02 语料集,包含 TACKBP2014 训练集与测试集, TACKBP2015 训练集与测试集和 TACKBP2016 测试集。采用 Mihaylov 和 Frank^[14] 的训练集、验证集和测试集的划分方法,把 2014 英文训练集与测试集、2015 英文训练集当作本文的训练集,2015 英文测试集为验证集,2016 英文测试集为测试集,实验具体的文档和事件分布情况如表 2 所示。

表 2 训练集、验证集与测试集所包含的文档与 18 类事件

	文档数	事件数
训练集	509	11 684
验证集	202	4 995
测试集	169	4 155

对于训练集中的事件句,我们采用 BIO 标签,B 表示事件的开始,I 表表示事件的内部,O 表示非事件,可以把事件触发词识别和事件的识别联合抽取出来。比如输入句子为 $W = \{\text{The, attack, by, insurgents, occurred, on, Saturday}\}$,对应的标签为 $Y = \{\text{O, B-Conflict, Attack, O, O, O, O, O}\}$

3.3 评价指标

对于触发词和事件类型,本文采用 P (准确率)、 R (召回率)和 F_1 值的评价指标,并采用 TAC-KBP2016 组织者提供的官方评价脚本^①进行测评。值得注意的是,事件类型评价得分是基于触发词识别的错误传播的,若触发词触发了两种事件类型,评价指标中应含有这两种事件类型。

例如,句子“the murder[Conflict, Attack, Life, Die]of John on Tuesday and Bill on Wednesday”,触发词“murder”触发了“Conflict, Attack”和“Life, Die”两种事件。因此,如果一个文档中含有 100 个触发词,其中 6 个触发词触发了两种事件,在评价指标中应有 106 个触发词和其触发的相应事件类型。

① <http://cairo.lti.cs.cmu.edu/kbp/2016/event/Event-Mention-Detection-scoring-2016-v29> Official Event Nugget Detection and Coreference Scoring for TAC KBP 2016.

3.4 结果及分析

我们首先关注本文提出方法在事件抽取任务上的有效性。

表 3 给出了本文方法取得的事件识别的性能。从表 3 的结果可以看到：

(1) 在词向量的基础上依次加入词性向量、实体向量和文档向量,最终识别的 F_1 值逐渐提高。

(2) 从 C 组实验和 D 组实验对比结果推理可知,实体向量与文档向量蕴含的信息相当,在触发词识别和事件类型确定两个子任务上得到的性能相当。

(3) E 组实验结果表明,事件触发词识别的 F_1 值为 58.56%,事件类型的 F_1 值为 51.82%。对比 C 和 D 组实验与 E 组实验的结果可以看到,相比 C 和 D 组式样,E 组实验在两个子任务上都获得了更好的准确率、召回率和 F_1 值,说明实体向量和文档向量具有一定的互补性,实体向量包含事件局部的参与者信息,而文档向量包含事件之间的关系,将两者结合能更好地进行事件的识别与分类。

总体而言,本文提出的 Seq2Seq 模型在 KBP2016 语料上触发词识别和事件类型分类 F_1 值分别为 58.56%和 51.82%,特别地,引入的实体向量和文档向量能有效提升系统性能。

表 3 用 KBP2016 官方测评各组实验结果(%)

	触发词识别结果			事件类型识别结果		
	P	R	F_1	P	R	F_1
词向量(A)	61.90	48.83	54.59	51.88	43.97	47.60
词向量+词性向量(B)	64.49	49.24	55.84	53.07	44.50	48.40
词向量+词性向量+实体向量(C)	64.77	50.70	56.87	53.94	44.75	48.91
词向量+词性向量+文档向量(D)	60.99	53.02	56.73	51.01	47.32	49.09
词向量+词性向量+实体向量+文档向量(E)	63.64	54.23	58.56	54.08	49.75	51.82

在本文提出的 Seq2Seq 模型中我们引入了注意力机制,接着我们关注各种注意力机制对事件识别任务性能的影响。

表 4 给出了不同注意力机制下本文方法在事件抽取上的性能。从表 4 可知：

(1) 对比加入了 BahadanauAttention 机制的 Seq2Seq 模型和没有注意力的模型,事件触发词识

别的 F_1 值提高了 4.45%,事件类型识别的 F_1 值提高了 6.71%,说明注意力机制的引入能较大提升系统的 F_1 值；

(2) 不同注意力的表现略有差别。使用 BahadanauAttention 机制与使用 LuongAttention^[22] 机制,我们的 Seq2Seq 模型在触发词识别上的 F_1 值性能提高了约 0.64%,事件类型识别的 F_1 值提高了约 0.56%。

表 4 各种 attention 实验对比结果(%)

	触发词识别结果			事件类型识别结果		
	P	R	F_1	P	R	F_1
without attention	52.49	58.62	54.11	44.94	45.29	45.11
BahadanauAttention	63.64	54.23	58.56	54.08	49.75	51.82
LuongAttention	64.06	52.86	57.92	54.91	48.06	51.26

与基准系统相比,本文给出的方法不论在触发词识别还是事件类型的识别上都取得了最佳的 F_1 值。

表 5 给出了 E 组实验结果与实验基准结果的详细对比。从表 5 可以看到,本文触发词识别最终的结果比实验基准 1 和基准 2 分别提高了 3.765%和 2.60%,在事件类型分类上分别比实验基准 1 和基准 2 提高了约 4.8%和 1.2%,证明了该方法的有

效性。UID1 的 k 近邻模型在提取候选触发词时,只提取了和训练集中触发词词性还原一致的词,这导致忽略了其他词性还原不一致的词;而本文的模型是利用 Seq2Seq 模型自动挖掘词之间的联系,能抽取和训练集中触发词词义相近的词,提高了触发词的识别 F_1 值。Mihaylov 和 Frank 模型只局限于句子信息,忽略了实体信息和文档信息;实体含有部分事件类型的信息,文档含有句子之外的其他信息,

如相似类型的事件经常出现在同一个文档中,本文利用词性、词、实体和文档的信息进行事件识别,并且加入了注意力机制,从而提高了事件类型的识别 F_1 值。句子中不同的成分对事件识别的重要性不同,

Yang 等的模型并没有对词的权重加以区分,而本文引入注意力机制,使得模型能区别不同词在事件识别中的作用,赋以不同权值,从而提高了事件识别的性能。

表 5 E 组实验与基准对比结果(%)

		触发词识别结果			事件类型识别结果		
		P	R	F_1	P	R	F_1
基准 1	UTD1	55.36	53.85	54.59	47.66	46.35	46.99
	Mihaylov 和 Frank	58.41	51.60	54.80	48.45	42.81	45.45
基准 2	Yang 等	58.43	53.69	55.96	52.31	49.06	50.63
Seq2Seq(本文实验 E)		63.64	54.23	58.56	54.08	49.75	51.82

最后我们进一步分析了事件类型识别在各个不同类型上的识别性能。表 6 给出了 E 组实验在 18 类不同事件上的识别性能,可以看到部分事件类型识别的 F_1 值仍然偏低,主要原因为:

- (1) 训练集规模不大,共有 11 684 个事件实例。
- (2) 词组形式触发词的训练数据不足,训练集

中,词组形式的触发词仅有 251 个,导致词组的触发词识别正确的仅有两个。

(3) 事件类型分布不平衡,有些事件类型实例比较少,比如 Transaction_Transaction 类型的训练实例仅有 42 个,所占比例为 0.36%,从表 6 可知,Transaction_Transaction 类型识别的 F_1 值 11%,远远低于事件类型识别的 F_1 值 51.82%。

表 6 E 组实验 18 类事件识别结果的 F_1 值

事件类型	F_1 值	事件类型	F_1 值
Conflict_Attack	0.60	Manufacture_Artifact	0.57
Conflict_Demonstrate	0.71	Movement_Transport-Artifact	0.04
Contact_Broadcast	0.37	Movement_Transport-Person	0.50
Contact_Contact	0.36	Personnel_Elect	0.70
Contact_Correspondence	0.10	Personnel_End-Position	0.71
Contact_Meet	0.48	Personnel_Start-Position	0.32
Justice_Arrest-Jail	0.82	Transaction_Transaction	0.11
Life_Die	0.73	Transaction_Transfer-Money	0.58
Life_Injure	0.74	Transaction_Transfer-Ownership	0.46

4 总结

为了提高事件识别的性能,本文给出了一种序列到序列的神经网络模型,并将词、词性、实体和文档信息以向量的形式进行表征,并通过注意力机制的引入更好地凸显各个成分的重要性。在 TACK-BP2016 语料上的实验结果证明了方法的有效性。

参考文献

[1] Doddington G R, Mitchell A, Przybocki M A, et al.

The automatic content extraction (ACE) program-tasks, data, and evaluation [C]//Proceedings of LREC. 2004, 2: 1.
[2] Kim J D, Ohta T, Pyysalo S, et al. Overview of BioNLP09 shared task on event extraction[C]//Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing: Shared Task. 2009:1-9.
[3] Dubbin G, Bhatia A, et al. Event nugget detection and argument extraction with DISCERN[C]//Proceedings of the 29th International Florida Artificial Intelligence Research Society Conference. 2016.
[4] Ahn D. The stages of event extraction[C]// Proceed-

- ings of the Workshop on Annotating and Reasoning about Time and Events. Association for Computational Linguistics, 2006;1-8.
- [5] Grishman R, Westbrook D, Meyers A. Nyu's English ace 2005 system description [J]. Journal on Satisfiability, 2005, 51(11):1927 - 1938.
- [6] Li Q, Ji H. Incremental joint extraction of entity mentions and relations[C]// Proceedings of Meeting of the Association for Computational Linguistics. 2014; 402-412.
- [7] Gupta P, Ji H. Predicting unknown time arguments based on cross-event propagation[C]//Proceedings of the ACL-IJCNLP 2009 Conference Short Papers. Association for Computational Linguistics, 2009; 369-372.
- [8] Hong Y, Zhang J, Ma B, et al. Using cross-entity inference to improve event extraction[C]//Proceedings of Meeting of the Association for Computational Linguistics; Human Language Technologies. Association for Computational Linguistics, 2011;1127-1136.
- [9] Liao S, Grishman R. Using document level cross-event inference to improve event extraction[C]// Proceedings of the Meeting of the Association for Computational Linguistics, July 11-16, 2010, Uppsala, Sweden. DBLP, 2010;789-797.
- [10] Ji H, Grishman R. Refining event extraction through cross-document inference[C]// Proceedings of Meeting of the Association for Computational Linguistics, June 15-20, 2008, Columbus, Ohio, Usa. DBLP, 2008;254-262.
- [11] 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 Proceedings. 2015, 2; 365-371.
- [12] 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, 1; 167-176.
- [13] Nguyen T H, Cho K, Grishman R. Joint event extraction via recurrent neural networks[C]// Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016;300-309.
- [14] Duan S, He R, Zhao W. Exploiting document level information to improve event detection via recurrent neural networks[C]//Proceedings of the 8th International Joint Conference on Natural Language Processing. 2017, 12;352-361.
- [15] Li Q, Ji H, Huang L. Joint event extraction via structured prediction with global features[C]// Proceedings of the Meeting of the Association for Computational Linguistics, 2013;73-82.
- [16] Ghaeini R, Fern X, Huang L, et al. Event nugget detection with forward-backward recurrent neural networks[C]// Proceedings of the Meeting of the Association for Computational Linguistics. 2017; 369-373.
- [17] Le Q, Mikolov T. Distributed representations of sentences and documents[C]//Proceedings of the International Conference on Machine Learning. 2014; 1188-1196.
- [18] Bahdanau, D Cho, K Bengio Y. Neural machine translation by jointly learning to align and translate [C]//Proceedings of the ICLR 2015; 1-15.
- [19] Mihaylov T, Frank A. AIPHES-HD system at TAC KBP 2016: Neural event trigger span detection and event type and realis disambiguation with word embeddings[C]// Proceedings of the TAC. 2016.
- [20] Lu J, Ng V. UTDs event nugget detection and coreference system at KBP 2016[C]//Proceedings of the TAC. 2016.
- [21] Yang J, Liang S, Zhang Y. Design challenges and misconceptions in neural sequence labeling[J]. arXiv preprint arXiv:1806.04470, 2018.
- [22] Luong M T, Pham H, Manning C D. Effective approaches to attention-based neural machine translation [J]. arXiv preprint arXiv:1508.04025, 2015.



张俊青(1992—), 硕士研究生, 主要研究领域为自然语言处理、深度学习。
E-mail: 2397405605@qq.com



孔芳(1977—), 通信作者, 博士, 教授, 主要研究领域为自然语言处理、指代消解、篇章处理。
E-mail: kongfang@suda.edu.cn