计算机科学
COMPUTER SCIENCE

基于联合模型的端到端事件可信度识别

曹金娟, 钱忠, 李培峰

引用本文

曹金娟, 钱忠, 李培峰.基于联合模型的端到端事件可信度识别[J]. 计算机科学, 2023, 50(2): 292-299.

CAO Jinjuan, QIAN Zhong, LI Peifeng. End-to-End Event Factuality Identification with Joint Model[J]. Computer Science, 2023, 50(2): 292-299.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.211200108

基于联合模型的端到端事件可信度识别

曹金娟　钱　忠　李培峰
苏州大学计算机科学与技术学院　江苏 苏州 215006
(20194227035@stu.suda.edu.cn)

摘　要　事件可信度是对文本中事件真实情况的一种描述,是自然语言处理领域许多相关应用的基本任务。目前,大多数关于事件可信度的相关研究都是使用标注的事件进行事件可信度识别,不方便实际应用,并且忽略了不同事件源对事件可信度的影响。针对现有问题,提出了一个端到端的事件可信度识别的联合模型JESF。该模型可以同时进行事件识别、事件源识别、事件可信度识别3个任务;使用BERT(Bidirectional Encoder Representations from Transformers)和语言学特征加强单词的语义表示;使用注意力机制(Attention)和依存句法树构建图卷积神经网络(Graph Convolutional Network,GCN),以有效地提取语义和句法特征。特别地,该模型也可以应用于只考虑默认源(文本作者)的事件可信度任务。在FactBank,Meantime,UW,UDS-IH2等语料上的实验结果显示,所提模型优于基准模型。

关键词:事件可信度识别;端到端;联合模型;图卷积神经网络;BERT
中图法分类号　TP391

End-to-End Event Factuality Identification with Joint Model

CAO Jinjuan,QIAN Zhong and LI Peifeng
School of Computer Sciences and Technology,Soochow University,Suzhou,Jiangsu 215006,China

Abstract　Event factuality is the description of the real situation of events in text. It is the basic task of many related applications in the field of natural language processing. At present,most researches on event factuality use labeled events to identify event factuality,which is inconvenient for practical application,and ignores the impact of different event sources on event factuality. Aiming at the existing problems,an end-to-end joint model JESF for event factuality identification is proposed. The model can carry out three tasks at the same time:event identification,event source identification and event factuality identification. Using bidirectional encoder representations from transformers(BERT) and linguistic features to strengthen the semantic representation of words. Graph convolutional network(GCN) is constructed by using attention mechanism and dependency syntax tree to effectively extract semantic and syntactic features. In particular,the model can also be applied to the task of event factuality considering only the default source(text author). Experimental results on FactBank,Meantime,UW and UDS-IH2 show that the proposed model is better than the benchmark model.

Keywords　Event factuality identification,End-to-End,Joint model,Graph convolutional network,Bidirectional encoder representations from transformers

1　引言

当人们在描述和谈论某一事件时,往往会表达自身对当前事件发生与否的态度和看法,即事件可信度。作为自然语言理解领域的一种深层语义信息,事件可信度在自然语言处理的研究和实践领域有着重要作用。在问答系统中,事件可信度可以作为分析依据为系统筛选更准确、有效的答案;在谣言检测中,事件可信度识别是谣言检测的基础任务;在知识库的创建中,其可以作为选择知识的判断依据。

之前的研究大多是基于标注的事件,并且在只考虑默认源(文本作者、AUTHOR)的情况下进行的。如例1所示,之前的研究是给定事件“rose”和“rise”,判断事件“rose”和“rise”的事件可信度,其中事件“rose”不受任何不确定和否定信息的影响,其可信度值为“一定可信”,事件“rise”在不确定线索词“may”的作用下可信度值为“可能可信”。而在真实情况中,事件并未给出,并且同一事件对于不同事件源的事件可信度不同。对此,本文提出了一个联合模型来识别事件、事件源以及对应的事件可信度。本文将事件可信度值分为一定可信(CT+)、一定不可信(CT-)、可能可信(PS+)、可能不可信(PS-)和不确定(Uu)5类。一个句子中有可能含有多个事件,每个事件对应不同的可信度值,而且同一事件对于不同事件源,其可信度也不同,这提高了事件可信度识别的难度。

到稿日期:2021-12-09　返修日期:2022-03-16
基金项目:国家自然科学基金(61772354,61836007,61773276);江苏高校优势学科建设工程资助项目
This work was supported by the National Natural Science Foundation of China(61772354,61836007,61773276) and Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.
通信作者:钱忠(qianzhong@suda.edu.cn)

如例 2 所示，句子中有两个事件，即“claimed”和“caught”，对于文本作者来说，“claimed”是一个客观事实，即可信度为 CT＋，但对“caught”的可信度不持立场，可信度为 Uu。同时，对于事件源“police”来说，事件“caught”的可信度为 CT＋。

例 1　Stocks rose sharply today and may rise again tomorrow.（股票今天大幅上涨，明天可能再次上涨。）

例 2　The police claimed that they had caught the fugitive.（警察宣称他们已经抓住了逃犯。）

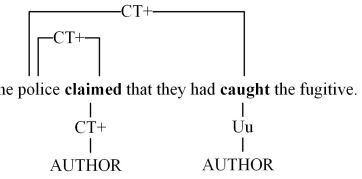


图 1　例 2 事件源和事件可信度对应图

Fig. 1　Example 2 correspondence diagram of event source and event factuality

近年来的相关研究大多数是基于神经网络的模型，Qian 等[1]提出了具有辅助分类的生成式对抗网络事件可信度识别模型，使用不确定标签和线索词标签来判断事件的可信度，可以有效提升数据较少类别的性能，但是对于多个线索词一同作用的事件可信度无法识别，如 PS－，被否定线索词和不确定线索词共同作用。Veyseh 等[2]为了捕获线索词到事件的信息，使用 BERT[3]对词向量进行编码，提出了基于 GCNs[4]的模型捕获句法和语义信息，从而提升了识别事件可信度的性能，但是仅考虑事件在 ATUTHOR 下的事件可信度，没有考虑事件在其他事件源下的事件可信度，并且只使用了事件对应节点捕获的特征识别事件可信度，忽略了句子的整体信息。Cao 等[5]提出了一个基于 GCN 和 BiLSTM 的混合模型，用于事件可信度识别，需要先对事件、事件源、线索词等信息进行抽取，再进行事件可信度识别，级联误差导致模型效果不佳。Jiang 等[6]探究了 BERT 在事件可信度识别任务上的性能，实验结果表明 BERT 没有获得最优的性能。

基于以上分析，本文提出了一个端到端的联合模型 JESF，可以识别未标注事件的事件可信度，并且在只考虑文本作者和考虑所有源的事件可信度识别任务上都有效果。具体来说，JESF 模型使用 1 个共享编码器和 3 个解码器进行事件、事件源和事件可信度识别。由于事件源是由特殊的事件引入的，因此在事件源识别时使用了事件识别的特征，事件可信度与事件和事件源有关，因此在可信度识别时融入了事件和事件源识别信息，体现了 3 个子任务之间的依赖关系，从而提升了模型的性能。本文的主要贡献如下：

（1）提出了一个联合模型 JESF，可以识别句子中的事件、事件源以及事件在事件源下的事件可信度。本文模型的实验结果比目前的端到端事件可信度识别模型的效果好，特别是对于只考虑默认源下的事件可信度识别也是有效的。

（2）使用 BERT 词向量和语言学特征（词性、依存关系、词原形）作为单词表示，并利用双向 LSTM 通过考虑单词的上下文来增强单词在句子中的语义表示，从而解决事件的歧义问题。使用注意力机制和依存句法树构建 GCNs，进行事件信度识别，它可以更有效地整合语义和句法信息，从而提升模型性能。

（3）探究了串联模型和联合模型的特点。基于本文提出的联合模型 JESF，使用了相同设置下的串联模型 PESF 进行对比。在只考虑默认源的事件可信度识别中，本文也使用 PEF，JEF 来探究串联模型和联合模型的特点。

2　相关工作

目前大多数的研究都是基于标注事件进行可信度识别，并且只考虑对于文本作者的事件可信度。传统的事件可信度识别算法结合了特征工程和机器学习对事件可信度进行识别，Saurí[7]使用相关特征信息（主要有上层谓词、否定词、情态词等），提出了一种自上而下遍历依存句法树分析相关信息和事件关系，从而计算事件可信度的算法。Lotan 等[8]分析了谓词、线索词、特殊句子结构（如状语从句）等信息对事件可信度的影响，提出了一种基于依存句法树的规则模型，用于计算事件可信度。Marneffe 等[9]使用上层谓词类型及词原形、事件主语、情态词、否定词、句法结构等特征，提出了一种基于最大熵的方法。Saurí 等[10]提出了一种基于 SVM 分类的方法。Qian 等[11]提出了一种分两步的机器学习和规则相结合的方法，用于识别事件可信度。随着神经网络的兴起，研究者开始尝试使用神经网络模型去识别事件可信度。Veyseh 等[2]提出了一种基于 GCNs 的模型，能够有效地融合语法信息和句法信息进行事件可信度识别。Zhang 等[12]提出了基于门控卷积网络的篇章级事件可信度识别方法。Cao 等[13]提出了一个不确定的局部到全局的神经网络（Uncertain Local-to-Global Network），用于篇章级事件可信度识别。

对于未标注事件的事件可信度识别，先前的研究没有考虑仅在默认源下的事件可信度。Qian 等[1]提出了一种基于 ACGAN 模型的方法，将事件、事件源、源引入谓词以及线索词等信息抽取出来，再通过 ACGAN 进行事件可信度识别。Cao 等[5]提出了一种混合神经网络，通过规则获取事件、事件源等相关信息，再通过混合神经网络进行事件可信度识别。这些方法都是基于串联的方法，会产生级联错误。目前事件可信度识别并没有使用联合学习的方法。

Saurí 等基于 TimeBank 创建了面向事件可信度的语料库 FactBank[14]，标注了事件、事件源等信息，并从情态和极性两个部分对事件进行事件可信度值的标注，主要包括 CT＋，CT－，PS＋，PS-，Uu 等。Minard 等[15]构建了一个多语言事件和时间语料库 MEANTIME，事件可信度标签根据 certainty 和 polarity 两个属性获得。Lee 等[16]构建了更细粒度的语料库 UW，将事件可信度的值映射到[－3，3]上，可将分类任务转化为回归任务，从而量化给定事件发生的可能性，其中－3 表示一定不发生，3 表示一定发生。White 等[17]构建了 UDS-IH1 数据集，标注了超过 6 920 个事件。为了扩展事件可信度数据集，Rudinger 等[18]基于 UDS-IH1 构建了数据集 UDS-IH2，从 UNDERSTANDABLE，PREDICAT，HAPPENED，CONFIDENCE 这 4 个角度进行数据标注。

3　任务定义

事件可信度识别任务是识别事件在文本中的可信度，端到端事件可信度则是给定句子，识别句子中的事件、事件源，

并识别事件对事件源的事件可信度。

3.1 相关定义

(1)事件。在事件可信度识别中事件的表示形式为触发词(如例1中的"rose"),事件是一种能够判断其是否发生的情况,具有一定的可信度。

(2)事件源。事件源是事件的参与者或评价者,对事件的可信度持有一定的看法,不同的事件源可能对同一事件有不同的看法,即事件的可信度不同。之前的研究大多仅考虑了文本作者作为事件的唯一事件源,本文考虑了句子内所有的事件源。特别地,本文还考虑了泛化源(Generic Source,GEN)和虚拟源(Dummy Source,DUMMY)两种特殊的事件源。如果句子中出现了非人称结构(如"it seems/appears/is clear/…")或者被动语态结构(如"it is expected/confirmed/reported/…"),在类似的表达中,可以用"每个人"或"某人"来表达,对此,本文使用泛化源来表示。如果事件所在的句子作为直接引语出现在引号中,即没有给出相应的事件源,则使用虚拟源来表示。

3.2 事件可信度值的统一表示

相关研究[7,10,14,19]均表明事件可信度的取值与情态(也称为确定性)和极性有关。其中极性表达事件是否发生,情态表明事件发生可能性的大小。尽管目前现有的英文语料库标注方式不一样,但是均按照情态和极性进行标注。因此,本文采用FactBank的标注方式,按照情态和极性对FactBnak,Meantime,UW,UDS-IH2语料库进行标签转换,将事件可信度分为CT+,CT-,PS+,PS-以及Uu这5类。

表1列出了标签的转换方式。其中,对于FactBank语料库,本文将其中的PR(Probable)转换为PS(Possible),主要原因在于PR和PS的区分困难,并且区分意义不大,以及PR和PS类别数量较少。对于Meantime,本文根据Certainty(确定性)和Polarity(极性)对标签进行转换。UW语料库是将事件可信度值映射在[-3,+3]范围内,本文则根据正负(极性)以及可信度值(确定性)四舍五入将标签进行转换。对于UDS-IH2,本文则是根据HAPPENED(极性)和CONFIDENCE(确定性)进行标签转换。

表1 标签的转换

Table 1 Label conversion

FactBank	Meantime		UW	UDS-IH2		转换标签
可信度标签	Certainty	Polarity	可信度值	HAPPENED	CONFIDENCE	
CT+/CT-	CERTAIN	POS/NEG	[2.5,3]/[-3,-2.5]	True/False	4	CT+/CT-
PS+(PR+)	POSSIBLE(PROBABLE)	POS	(0.5,2.5)	True	3,2,1	PS+
PS-(PR-)	POSSIBLE(PROBABLE)	NEG	(-2.5,-0.5)	False	3,2,1	PS-
Uu	UNDERSPECIFIED	POS/NEG	[-0.5,0.5]	True/False	0	Uu

4 模型结构

本节主要介绍本文提出的事件可信度识别的联合模型,旨在先识别句子中的事件和事件源,然后计算事件的可信度,模型的框架如图2所示。该模型主要由输入层、事件识别层、事件源识别层和可信度识别层4个模块组成。

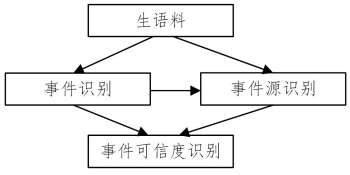


图2 JESF模型框架

Fig. 2 Frame of JESF model

4.1 输入层

为了获得丰富的语义信息,输入向量由单词向量、词性向量、依存向量和词原形向量组成。对于一个长度为 n 的句子 $\boldsymbol{X}=\{x_1,x_2,\cdots,x_n\}$,其中 x_i 表示句子中的第 i 个单词。单词向量则是句子通过BERT得到的,同时本文使用随机初始化词向量分别获得词性、依存关系、词原形的词向量表示,将向量拼接获得最终句子的词向量表示 $\boldsymbol{E}$。

(1)单词向量。本文使用"BERT-Base,Uncased,12 layers"预训练好的词向量对句子进行编码,以获得单词的词向量表示 $(\boldsymbol{e}_1^b,\boldsymbol{e}_2^b,\cdots,\boldsymbol{e}_n^b)$。

(2)词性向量。对于一个长度为 n 的句子 $\boldsymbol{X}$,通过Stanfordnlp获得对应单词词性的序列 $(p_1,p_2,\cdots,p_n)$,其中 p_i 为单词 x_i 的词性。本文使用随机初始化方法获得词性向量的初始表示 $(\boldsymbol{e}_1^p,\boldsymbol{e}_2^p,\cdots,\boldsymbol{e}_n^p)$。

(3)依存关系向量。依存关系指每个单词在依存句法树上的与父节点的依存关系,本文使用Stanford Parser获得单词对应的依存关系,使用随机初始化的词向量对应的向量表示 $(\boldsymbol{e}_1^d,\boldsymbol{e}_2^d,\cdots,\boldsymbol{e}_n^d)$。

(4)词原形向量。本文使用spacy获得单词的词原形,之后使用GloVe[20]初始化的词向量获取词原形的向量表示 $(\boldsymbol{e}_1^w,\boldsymbol{e}_2^w,\cdots,\boldsymbol{e}_n^w)$。

最终,将向量拼接起来获得句子的词向量表示 $\boldsymbol{E}=\{\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n\}$,其中 $\boldsymbol{e}_i=\boldsymbol{e}_i^b\oplus\boldsymbol{e}_i^p\oplus\boldsymbol{e}_i^d\oplus\boldsymbol{e}_i^w$。对于获得的词向量表示,本文使用双向LSTM(Bidirectional LSTM,BiLSTM)学习文本特征。

$$\overrightarrow{\boldsymbol{h}_i}=LSTM_f(\boldsymbol{e}_i,\overrightarrow{\boldsymbol{h}_{i-1}}) \tag{1}$$

$$\overleftarrow{\boldsymbol{h}_i}=LSTM_b(\boldsymbol{e}_i,\overleftarrow{\boldsymbol{h}_{i-1}}) \tag{2}$$

BiLSTM通过将正向LSTM和反向LSTM获得的特征进行拼接,来获得具有前向和后向的文本信息 $\boldsymbol{h}_i=[\overrightarrow{\boldsymbol{h}_i},\overleftarrow{\boldsymbol{h}_i}]$,$\boldsymbol{H}=\{\boldsymbol{h}_1,\boldsymbol{h}_2,\cdots,\boldsymbol{h}_n\}$。

4.2 事件识别层

事件大多为名词和动词,且同一单词的不同形式大多有类似的语义。针对事件的这一特点,本文在输入时加入了词性向量和词原形向量,以有效地解决事件识别中的歧义问题。

本文使用一个多头注意力机制[21]来捕获事件识别重要信息 $\boldsymbol{T}=\{\boldsymbol{t}_1,\boldsymbol{t}_2,\cdots,\boldsymbol{t}_n\}$,使用一个全连接层(Fully Connected Layer,FC),通过sigmoid进行事件识别。

$$\boldsymbol{T}=MultiHeadAttenion(\boldsymbol{H},\boldsymbol{H},\boldsymbol{H}) \tag{3}$$

$$\boldsymbol{y}_i^e = \text{sigmoid}(\boldsymbol{W}_t \boldsymbol{t}_i + \boldsymbol{b}_t) \tag{4}$$

4.3 事件源识别层

为了识别句子中所有的事件源，对于不存在于句子中的AUTHOR，GEN 和 DUMMY 这 3 种事件源，本文在句子前面加入这 3 种源，以便于事件源的识别，如图 3 所示。事件源识别层由一个多头注意力层和一个全连接层构成。事件源识别层的输入 $\boldsymbol{F}=\{\boldsymbol{f}_1,\boldsymbol{f}_2,\cdots,\boldsymbol{f}_n\}$ 由事件识别输出向量 $\boldsymbol{T}$ 和 BiLSTM 输出向量 $\boldsymbol{H}$ 拼接构成。本文使用多头注意力机制进一步微调事件源的向量表示，以捕获事件源的相关特征 $\boldsymbol{S}=\{\boldsymbol{s}_1,\boldsymbol{s}_2,\cdots,\boldsymbol{s}_n\}$，然后使用全连接层识别事件源。具体表达式如下：

$$\boldsymbol{f}_i = [\boldsymbol{h}_i, \boldsymbol{t}_i] \tag{5}$$

$$\boldsymbol{S} = MultiHeadAttenion(\boldsymbol{F}, \boldsymbol{H}, \boldsymbol{H}) \tag{6}$$

$$\boldsymbol{y}_i^s = \text{sigmoid}(\boldsymbol{W}_s \boldsymbol{s}_i + \boldsymbol{b}_s) \tag{7}$$

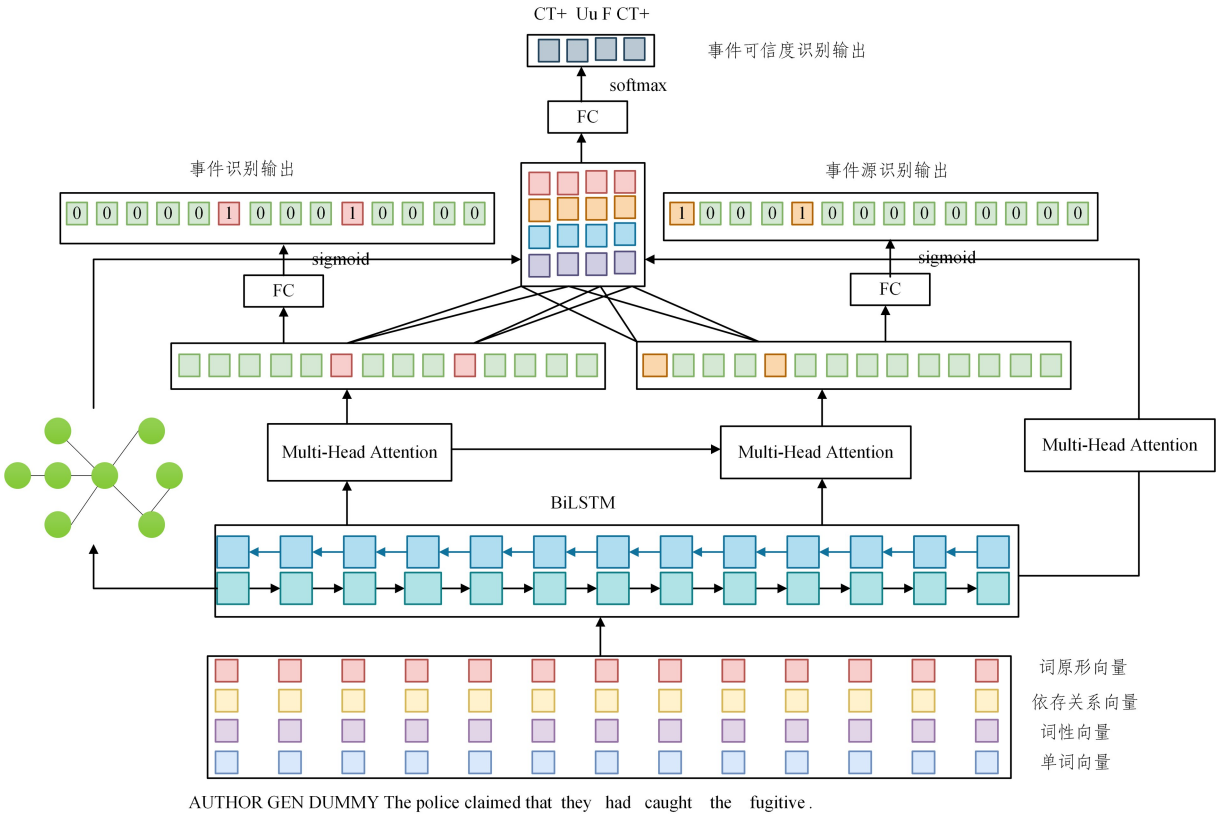


图 3 JESF 模型结构图

Fig. 3 Structure of JESF model

4.4 事件可信度识别层

事件可信度识别层主要包括一个多头注意力机制、一个 GCN[2] 和一个前馈神经网络。输入包含输入层获得的语义特征表示 $\boldsymbol{H}$、事件识别向量 $\boldsymbol{T}$ 和事件源识别向量 $\boldsymbol{S}$。

对于输入向量 $\boldsymbol{H}$，本文通过多头注意力机制和全连接层来捕获句子的语义特征 $\boldsymbol{m}$。同时，本文使用 GCNs 捕获句子的语义和句法信息。Veyseh 等[2] 通过函数将 x_i 和 x_j 的词向量拼接计算，并使用 sigmoid 激活函数得到语义矩阵 $\boldsymbol{A}_{\text{sem}}$，这种计算方式得到的矩阵没有结合上下文，只是单词向量间的相似度计算。为了关注全局的语义信息，本文使用来自 BiLSTM 的向量根据 Attention 中的方式获得语义矩阵 $\boldsymbol{A}_{\text{sem}}$，通过依存句法树获得句法矩阵 $\boldsymbol{A}_{\text{syn}}$，并通过一个可学习参数 p 结合 $\boldsymbol{A}_{\text{sem}}$ 和 $\boldsymbol{A}_{\text{syn}}$ 得到权重矩阵 $\boldsymbol{A}_s$。最后，将特征向量 $\boldsymbol{H}$ 和权重矩阵 $\boldsymbol{A}_s$ 输入到两层 GCNs 中，得到了语法和语义特征 $\boldsymbol{G}$。具体表达式如下：

$$\boldsymbol{A}_{ij}^{\text{sem}} = softmax(\boldsymbol{W}_q \boldsymbol{H} * (\boldsymbol{W}_k \boldsymbol{H})^{\mathrm{T}}) \tag{8}$$

$$\boldsymbol{A}_{ij}^{\text{syn}} = (\widetilde{\boldsymbol{D}}^{-\frac{1}{2}} \widetilde{\boldsymbol{A}} \widetilde{\boldsymbol{D}}^{-\frac{1}{2}}) \tag{9}$$

$$\boldsymbol{A}_s = p\boldsymbol{A}_{\text{sem}} + (1-p)\boldsymbol{A}_{\text{syn}} \tag{10}$$

$$\boldsymbol{g}_i^1 = \text{relu}(\boldsymbol{A}_s \boldsymbol{h}_i \boldsymbol{W}_{g1}) \tag{11}$$

$$\boldsymbol{g}_i = \text{relu}\, \boldsymbol{A}_s \boldsymbol{g}_i^1 \boldsymbol{W}_{g2}) \tag{12}$$

其中，$\boldsymbol{A}$ 是依存句法树对应的邻接矩阵，$\widetilde{\boldsymbol{A}}=\boldsymbol{A}+\mathbf{I}$（单位矩阵），$\widetilde{\boldsymbol{D}}$ 是 $\widetilde{\boldsymbol{A}}$ 对应的度矩阵。

本文将通过事件识别结果获得对应的事件候选词 $\boldsymbol{T}_i$（$\boldsymbol{x}_i$ 被识别为事件），通过事件源识别结果获得事件源候选词 $\boldsymbol{S}_j$（$\boldsymbol{x}_j$ 被识别为事件源），对每个事件和事件源进行可信度识别。最终对每个事件和事件源进行可信度识别的表达式如下：

$$\boldsymbol{O}_t = \text{relu}(\boldsymbol{W}_c[\boldsymbol{t}_i; \boldsymbol{s}_j; \boldsymbol{g}_i; \boldsymbol{m}] + \boldsymbol{b}_c) \tag{13}$$

$$y_t = \text{softmax}(\boldsymbol{W}_o \boldsymbol{O}_t + \boldsymbol{b}_o) \tag{14}$$

本文模型的目的是学习 3 个任务之间的联系，从而提升模型的性能。本文的损失函数为：

$$L(\theta) = -\frac{1}{n}\sum_{i=0}^{n-1}(\log P(y_i^e|\theta) + \log P(y_i^s|\theta)) - \frac{1}{m*k}\sum_{t=0}^{m*k-1}\log P(y_t|\theta) \tag{15}$$

其中，θ 为模型中可学习的参数，n 为输入序列长度，m 和 k 为被识别为事件的数目和被识别为事件源的数目。

5 实验

5.1 实验及语料设置

由于目前只有 FactBank 语料库标注了文本作者以外的事件源的事件可信度，因此本文仅在 FactBank 上探究事件在不同事件源下的可信度，而在现有英文数据集 FactBank，Meantime，UW，UDS-IH2 语料库上探究在默认源（即文本作者）下的事件可信度识别性能。由于不同数据集的标注方式不一样，本文将标签进行了转换，转换方式如表 1 所列，表 2 列出了数据集的数据分布情况，表 3 列出了数据集的划分（其是 FactBank，Meantime，UW 和 UDS-IH2 上标准的 train/dev/test 划分）。

表 2 数据集的统计

Tabel 2 Statistics of datasets

语料库	事件源	CT−	PS−	Uu	PS+	CT+
FactBank	所有源	446	79	4658	578	8152
	默认源	206	19	3674	198	5664
	其他源	240	60	984	380	2488
MeanTime	默认源	26	1	42	67	1259
UW	默认源	159	430	416	3957	8682
UDS-IH2	默认源	1711	7992	876	10517	6193

表 3 默认源下的数据统计

Tabel 3 Data statistics under default source

语料库	Train		Dev		Test	
	事件	句子	事件	句子	事件	句子
FactBank	6636	2162	2462	795	663	222
Meantime	1012	347	195	87	188	77
UW	9422	2963	3358	1074	864	274
UDS-IH2	22108	12543	2642	2002	2539	2077

本文中的 BERT 词向量维度为 768，词性向量、依存向量以及词原形向量维度均为 300，dropout 比例为 0.3，batch 大小为 8，BiLSTM 的输出维度为 300，GCN 的输出维度也为 300。另外，本文采用学习率为 0.0003 的 Adam 优化器优化任务中的损失函数。实验中，本文采用 F 值作为衡量事件识别和事件源识别性能的评价指标。与 Qian 等[1]和 Cao 等[5]一样，本文使用 Micro-F1 和 Macro-F1 作为衡量事件可信度识别性能的评价指标。

5.2 基准模型

5.2.1 基于不同事件源的事件可信度识别模型

(1)Cao[5]。该模型是由 Cao 等提出的一个端到端的事件可信度识别模型，其先抽取事件、事件源、线索词等相关信息，再通过一个基于 BiLSTM 和 GCN 的混合模型进行事件可信度识别。

(2)PESF。该模型是先进行事件识别，接着进行事件源识别，再通过事件识别和事件源的识别进行可信度识别。其是一个串联模型，用于比较串联和联合方法的差异，是本文模型的串联形式。

5.2.2 基于默认源的事件可信度识别模型

由于目前没有端到端的仅考虑默认源的事件可信度的相关研究，为了进行公平的比较，对于使用标注事件进行可信度识别的方法，本文均使用基于 BiLSTM 加多头注意力机制的模型（本文的事件识别模块）进行事件识别，之后使用基于标注事件的事件可信度模型进行识别。

(1)L-BiLSTM[18]。使用栈式双向线性 LSTM 对句子进行编码，之后通过事件的单词编码进行事件可信度识别。

(2)H-BiLSTM[18]。使用栈式双向线性 LSTM 和栈式双向树结构 LSTM（Stacked Bidirectional Tree LSTM，T-biLSTM）混合模型对句子进行编码，之后通过事件的单词编码进行事件可信度识别。

(3)Veyseh[2]。使用 BERT 对句子进行编码，之后使用基于 GCNs 的模型对事件可信度进行识别，能有效地结合语义和句法信息。

(4)Jiang[6]。使用 BERT 进行可信度识别，模型结构与 L-BiLSTM 类似。

(5)PEF。先进行事件识别，再进行可信度识别。该模型是本文模型只使用事件识别和事件可信度识别的串联模型。

(6)JEF。同时进行事件识别和事件可信度识别，是本文模型只使用事件识别和事件可信度识别的联合模型，即除去模型中相关源识别部分。

5.3 实验分析

5.3.1 基于不同事件源下可信度的结果分析

从表 4 可以看出，与 Cao 等提出的模型相比，本文模型能够有效提升各个类别的实验性能，尤其宏平均 F1 值（Macro-F1）有大幅度提升，这是因为使用 BERT 和语言学特征词向量能表示更丰富的语义信息，并且基于注意力机制和依存句法树构建的 GCNs 模型能有效捕捉线索词、事件、事件源以及事件可信度之间的关系，因此该模型能够有效提升 PS+，CT− 等被线索词所作用事件的性能。特别地，从表中可以发现，PS− 的结果均为 0，主要原因在于 PS− 样本不足，仅占训练集的 0.63%，而测试集仅有 4 个 PS− 样本，并且 PS− 受文本中不确定信息和否定信息共同影响，模型识别困难。

表 4 考虑不同事件源的事件可信度识别结果（FactBank）

Tabel 4 Results of event factuality identification considering different event sources on FactBank

模型	事件源	CT+	Uu	CT−	PS+	PS−	Mcro-F1	Macro-F1
Cao	All	63.63	59.01	33.42	26.63	0.00	60.00	36.54
	AUTHOR	70.97	64.98	38.59	31.09	0.00	67.37	41.13
	Other Source	43.62	23.63	28.03	21.52	0.00	37.04	23.36
PESF	All	65.26	62.98	54.55	38.81	0.00	63.03	44.32
	AUTHOR	76.02	69.19	51.61	44.44	0.00	72.35	48.25
	Other Source	41.14	37.25	**57.14**	35.00	0.00	40.75	34.11
JESF	All	**66.83**	**66.79**	**54.55**	**40.63**	0.00	**65.34**	**45.76**
	AUTHOR	**76.83**	**71.96**	**60.00**	**48.00**	0.00	**74.22**	**51.36**
	Other Source	**45.84**	**42.86**	50.00	**35.90**	0.00	**44.60**	**34.92**

表 5 仅考虑默认源的事件可信度识别结果

Tabel 5 Result of event factuality identification under default source

模型	FactBank		Meantime		UW		UDS-IH2	
	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro -F1
L-biLSTM	63.03	38.32	75.57	24.81	63.28	34.33	49.67	33.39
H-biLSTM	62.89	38.08	73.21	22.87	62.73	31.31	47.78	37.15
Veyseh	68.77	41.98	77.98	16.47	67.77	37.06	54.46	35.39
Jiang	70.14	32.95	78.38	16.62	65.78	37.79	53.17	34.55
PEF	72.74	47.23	78.92	28.60	68.20	40.53	58.93	38.03
JEF	73.21	47.36	78.87	29.85	69.32	40.97	60.00	40.99
JESF	**74.14**	**50.48**	**80.42**	**30.93**	**69.93**	**42.60**	**61.30**	**43.32**

PESF 和 JESF 相比，可以体现出随着任务的深入，串联模型会产生级联误差，即上一任务结果对下一任务的结果产生影响，串联模型 PESF 在经过事件识别、事件源识别后，模型会产生较大的级联误差，使用联合模型可以有效地避免级联误差。JESF 模型效果的提升表明，本文使用的联合模型能够有效地融合事件识别、事件源识别和事件可信度识别任务。

5.3.2 基于默认源下可信度的结果分析

与 L-BiLSTM，H-BiLSTM 相比，PEF 的微平均（Micro-F1）和宏平均（Macro-F1）显著提升，主要原因在于本文使用了基于 BERT 的词向量，并且使用 GCNs 捕获句法信息。对比 Veyseh 等、Jiang 等提出的模型，PEF 的效果更好，这表明本文提出的基于注意力机制和依存句法树构建的 GCNs 能够有效捕捉文本的句法和语义信息，并且能够有效地捕捉线索词（“not”“maybe”“possible”等）特征。同时注意力机制通过对重要特征进行捕捉，从而减少了冗余信息的干扰。

JEF 与 PEF 模型的对比，则展示了串联模型和并联模型的不同。在不考虑其他源的情况下，串联模型的结果仅依赖于事件识别结果，因此级联误差会较小，联合模型和串联模型的效果相差不大。JEF 模型的效果较好，表明了事件识别有助于事件可信度识别。

特别地，在只考虑文本作者这个默认源下，依然使用事件源识别模块的 JESF 相比 JEF 模型的效果有所提升，本文认为这是因为加入了事件源，模型在训练过程中无形地增加了可信度识别的数据量（即事件和事件源匹配数量比事件数量多），并且加入了事件源信息，有助于模型更好地识别事件可信度，从而提升模型性能。

5.3.3 事件、事件源的结果分析

对于事件识别，本文尝试使用 CNN 和 BiLSTM 等常见模型进行实验，根据表 6 的实验结果，PEF 明显优于 CNN 和 BiLSTM。这是因为 CNN 提取的是文本局部信息，而 BiLSTM 提取的是上下文信息，在提取层面能够更好地捕获事件信息。加入多头注意力机制后，文本能更好地捕获重要信息，以对事件特征进行捕获。因此，本文选择使用基于 BiLSTM 加多头注意力机制的模型进行可信度识别。

表 6 仅考虑默认源的事件识别结果

Tabel 6 Result of event identification under default source

	Factbank			Meantime			UW			UDS-IH2		
	P	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*
CNN	83.98	78.28	81.03	83.94	86.17	85.04	80.52	86.57	83.44	92.77	95.08	93.91
LSTM	86.37	79.34	82.70	86.59	82.45	84.47	85.40	85.30	85.35	96.70	95.83	96.26
PEF	**87.70**	80.69	84.05	88.05	**91.28**	89.64	**85.76**	87.15	**86.45**	**97.18**	**97.56**	**97.37**
JEF	84.29	84.16	84.23	**93.97**	85.78	89.69	85.29	85.88	85.58	95.64	96.73	96.18
JESF	82.90	**86.27**	**84.55**	93.30	89.45	**91.33**	84.98	**87.73**	86.33	96.05	96.77	96.41

从表 6 中的联合模型 PEF 和串联模型 JEF 的性能可以看出，随着数据量的增大，JEF 事件识别的效果比 PEF 的效果好，但比 PEF 的效果差。这是由于 PEF 是针对事件识别任务进行的，随着数据量的增加，模型捕获到更深层次的特性，模型效果更好，此时 JEF 由于其他信息的干扰，其效果反而不如 PEF。

如表 7 所列，本文事件识别 F 值比 Cao 等提出的模型提高了 2.38%，事件源识别 F 值则提高了 11.29%。主要原因在于，针对同一单词不同词性的意思不同、同一语义单词表现形式不同的问题，本文引入词性和词原形向量来解决。如例 3－例 5，例 3 中的事件“watch”是动词，意思是“监视”，而例 4 中的“watch”不是事件，其词性是名词，意思是“手表”。例 5 中，“watching”是事件，词原形是“watch”，意思为“监视”，与例 3 一致。

表 7 考虑不同事件源下事件和事件源的识别结果

Tabel 7 Results of event identification and event source identification considering different event sources

	事件识别			事件源识别 P		
	P	*R*	*F*	*P*	*R*	*F*
Cao	81.95	82.20	82.08	83.33	73.88	78.32
PESF	**87.70**	80.69	84.05	**92.25**	84.18	88.03
JESF	83.28	**85.67**	**84.46**	90.69	**88.55**	**89.61**

与联合模型 JESF 相比，串联模型 PESF 对事件、事件源识别的准确率较高，但是召回率较低，这是由于模型针对事件、事件源识别进行训练，对于已经出现的事件、事件源能够进行较好地识别。联合模型则是受到事件源、事件可信度信息的影响，模型能够捕获到更多特征，从而增大模型识别出未出现实例的概率，但是与此同时，联合模型受到其他信息的干扰，对于事件、事件源识别的准确率反而有所降低。

例 3 Police kept watch outside and about 25 undercover police were scattered throughout the crowd.(警方在外面进行监视,大约 25 名卧底警察分散在人群中。)

例 4 This could cause significant harm to watch producers in the United States and the Virgin Islands.(这可能对美国和维尔京群岛的手表生产商造成重大损害。)

例 5 Police have been watching the suspect.(警方一直在监视嫌疑犯。)

5.4 消融实验

为了评估不同模块对模型的贡献,本文进行了消融实验。表 8 列出了 FactBank 上消融实验的结果。在单词表示中加入词性、依存关系和词原形等语言学特征,是为了加强单词的语义表示。单词在句子中的词形表明了句子的时态等信息,蕴含上下文信息,词性则体现了单词的语法特性,依存关系则体现了单词在句子中的结构,词原形则体现单词原本的语义信息。融入这些语言学特征,则是综合考虑了词的上下文、语法、结构、语义等信息。

实验结果表明,词性、依存关系和词原形等语言学特征能够有效提高事件识别、事件源识别以及事件可信度识别的效果。

表 8 基于 FactBank 的消融实验

Tabel 8 Ablation experiment on FactBank

模型	事件识别			事件源识别			事件可信度识别	
	P	*R*	*F*	*P*	*R*	*F*	Micro-F1	Macro-F1
JESF	83.28	**85.67**	**84.46**	**90.69**	**88.55**	**89.61**	**65.34**	**45.76**
-词性向量	81.42	84.62	82.99	86.75	88.22	87.48	62.96	43.04
-依存关系向量	84.25	83.11	83.69	88.38	84.51	86.40	62.46	41.06
-词原形向量	84.40	77.53	80.82	87.93	85.86	86.88	60.03	45.53
-BERT	83.86	76.77	80.16	88.77	85.19	86.94	58.82	37.61
-GCNs	**85.53**	81.15	83.28	87.07	86.20	86.63	62.50	43.91
-BiLSTM	84.77	81.45	83.08	86.81	84.18	85.47	60.70	42.78

由于 BERT 模型的优越性,本文使用 BERT 对句子进行编码,结果显示使用 BERT 对句子编码可以有效提升模型性能。

由于 BERT 的基本架构是基于自注意力的 Transformer[21],其编码过程弱化了位置信息和方向信息,在事件可信度识别中,位置信息和方向信息是比较重要的。例如,否定词"not"在事件的前后,对于事件的影响也是不同的。因此,本文加入 BiLSTM 捕获句子的语义特性,实验结果证明了 BiLSTM 在学习序列化特征方面的有效性。

例 6 Bush denied that he had been arrested.(布什否认他曾经被捕。)

对于部分句子的结构,例如倒装、从句等,线索词与事件的相对位置较远,模型无法捕捉到线索词和事件的关系,但是在句法树中线索词和事件距离较近。如例 6,事件"arrested"在否定词"denied"的作用下,事件可信度为 CT−。在文本中"denied"与事件"arrested"有一定距离。在依存句法树中(见图 4),否定词"denied"与事件"arrested"直接相连,因此使用依存矩阵可以有效地捕捉句法特征,有助于模型捕捉这种远距离依赖的关系。

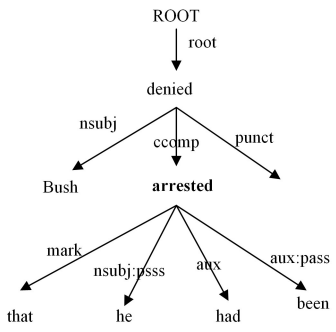


图 4 例 6 的依存句法树

Fig. 4 Dependency syntax tree of example 6

图 5 是例 6 对应的权重矩阵图,是融入了句法矩阵和语义矩阵得到的,单独使用句法矩阵可以学习到依存句法树结构的信息,即事件源"Bush"和事件"arrested"之间的信息是通过连接点"denied"间接学习的。融入语义矩阵后,可以发现"denied"和"arrested"以及"Bush"两两之间都有较大的权重,进一步加强了事件、事件源、线索词之间的联系。特别地,可以看到权重矩阵是非对称矩阵,这是因为融入了语义矩阵,语义矩阵经过 softmax 标准化,获得了不同单词之间的关注度得分,即"Bush"对于"arrested"和"arrested"对于"Bush"的关注度得分不同,这样更能体现句子中单词之间的关系。因此使用句法矩阵和语义矩阵作为权重矩阵的 GCNs 能够学习到更多的特征,从而进一步提升模型效果。

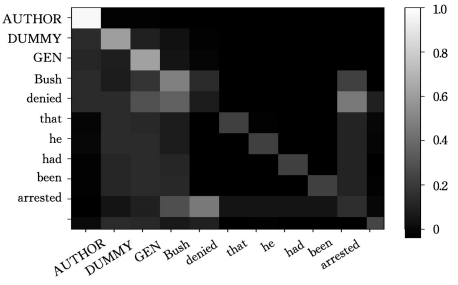


图 5 例 6 对应的权重矩阵

Fig. 5 Corresponding weight matrix of example 6

5.5 错误分析

本节针对 FactBank 上不同事件源下的事件可信度进行错误分析。

对于事件识别,79.43%的错误来自存在多个事件。其主要原因在于,存在多个事件的句子往往有更复杂的结构,并且存在单词相同语义不同的情况(如例 3、例 4),这使得模型识别事件更加困难。

对于事件源识别,事件源识别错误主要在于模型难以判断"源"是否对当前事件具有一定看法和态度,容易错将普通"源"识别为对事件具有看法和态度的事件源。如例 7,

“Kopp”对“leave”没有表达看法，即“Kopp”不是事件源。

例 7 Kopp may leave Paris，said Freeh.（弗里说，科普可能离开巴黎。）

对于事件可信度识别，事件识别错误或事件源识别错误导致事件可信度识别错误，是事件识别错误的主要原因。对于事件识别类别识别错误，62.63%的错误来自于 CT+和 Uu 的混淆识别，识别错误的原因主要是不同事件源对同一事件持不同态度，如例 2，对于事件源“police”和文本作者，事件“caught”的可信度为 CT+和 Uu。在例 8 中，对于事件源“watchmaker”和文本作者，事件“made”的可信度均为“CT+”，如何判断一句话是在陈述一个客观事实还是在描述某事件源所认为的事实，这是界定 CT+和 Uu 的难点。

例 8 The watchmaker tried to repair the pocket watch that was made in the early 1870s.（钟表匠试图修理 19 世纪 70 年代早期制造的怀表。）

结束语 本文提出了一种通用的端到端事件可信度识别的联合模型 JESF，该模型可以同时适用于仅考虑默认源和考虑所有源的事件可信度识别任务。具体而言，JESF 使用一个 BERT 对句子进行编码，并加入词性向量、依存关系向量、词原形向量，从而丰富句子的语义表示，并使用 BiLSTM 捕获句子的序列化语义特征；其次，使用多头注意力机制学习事件特征，进行事件识别；然后，用多头注意力机制特征辅以事件特征进行事件源识别；最后，使用 GCNs 捕获事件的句法和语义特征，使用多头注意力机制捕获句子的语义特征，并融入事件和事件源特征进行事件可信度识别。实验结果表明，本文提出的 JESF 模型在端到端事件可信度识别任务上的性能优于其他基准。

可以很明显地看出，数据的不平衡导致模型对少量类别的数据识别效果不佳，且目前可信度识别方面的语料太少，因此如何从不平衡的少量数据中捕获最有效的特征，将是未来工作的主要方向。

参 考 文 献

[1] QIAN Z，LI P F，ZHANG Y，et al. Event factuality identification via generative adversarial networkswith auxiliary classification [C]// Proceedings of the Joint Conference on Artificial Intelligence. 2018：4293-4300.

[2] VEYSEH A P B，NGUYEN T H，DOU D. Graph based neural networks for event factuality prediction using syntactic and semantic structures[C]// Proceedings of ACL. 2019：4393-4399.

[3] DEVLIN J，CHANG M W，LEE K，et al. Bert：pre-training of deep bidirectional transformers for language understanding [C]// Proceedings of NAACL-HLT. 2019：4171-4186.

[4] KIPF T N，WELLING M. Semi-supervised classification with graph convolutional networks [C]// Proceedings of ICLR. 2017：1-14.

[5] CAO J J，QIAN Z，LI P F，et al. End-to-end event factuality identification via hybrid neural networks[C]// Proceedings of CCKS. 2020：196-208.

[6] JIANG J J，DE MARNEFFE M C. He thinks he knows better than the doctors：bert for event factuality fails on pragmatics [J]. Transactions of the Association for Computational Linguistics，2021，9：1081-1097.

[7] SAURÍ R. A factuality profiler for eventualities in text[D]. Massachusetts：Brandeis University，2008.

[8] LOTAN A，STERN A，DAGAN I. Truthteller：annotating predicate truth[C]// Proceedings of NAACL-HLT. 2013：752-757.

[9] DE MARNEFFE M C，MANNING C D，POTTS C. Did it happen? the pragmatic complexity of veridicality assessment [J]. Computational Linguistics，2012，38(2)：301-333.

[10] SAURÍ R，PUSTEJOVSKY J. Are you sure that this happened? assessing the factuality degree of events in text[J]. Computational Linguistics，2012，38(2)：261-299.

[11] QIAN Z，LI P F，ZHU Q M. A two-step approach for event factuality identification[C]// Proceedings of IALP. 2015：103-106.

[12] ZHANG Y，LI P F，ZHU Q M. Document-level event factuality identification method with gated convolution networks[J]. Computer Science，2020，47(3)：206-210.

[13] CAO P F，CHEN Y B，YANG Y Q，et al. Uncertain local-to-global networks for document-level event factuality identification[C]// Proceedings of EMNLP. 2021：2636-2645.

[14] SAURÍ R，PUSTEJOVSKY J. Factbank：a corpus annotated with event factuality [J]. Language Resourcesand Evaluation，2009，43(3)：227-268.

[15] MINARD A L，SPERANZA M，URIZAR R，et al. Meantime，the newsreader multilingual event and time corpus[C]// Proceedings of LREC. 2016：4417-4422.

[16] LEE K，ARTZI Y，CHOI Y，et al. Event detection and factuality assessment with non-expert supervision [C]// Proceedings of EMNLP. 2015：1643-1648.

[17] WHITE A S，REISINGER D，SAKAGUCHI K，et al. Universal decompositional semantics on universal dependencies[C]// Proceedings of EMNLP. 2016：1713-1723.

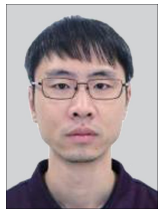
[18] RUDINGER R，WHITE A S，DURME B V. Neural models of factuality[C]// Proceedings of NAACL-HLT. 2018：731-744.

[19] HORN L R. A natural history of negation[M]. Chicago：University of Chicago Press，1989.

[20] PENNINGTON J，SOCHER R，MANNING C. Glove：global vectors for word representation[C]// Proceedings of EMNLP. 2014：1532-1543.

[21] VASWANI A，SHAZEER N，PARMAR N，et al. Attention is all you need[C]// Proceedings of NIPS. 2017：5998-6008.

CAO Jinjuan，born in 1997，postgraduate. Her main research interests include natural language processing and so on.



QIAN Zhong，born in 1989，Ph. D，lecturer. His main research interests include natural language processing and so on.

（责任编辑：喻藜）