

# 基于单句表示的篇章可信度识别方法

张刘敏, 张赟, 李培峰

(苏州大学计算机科学与技术学院 苏州 215006)

**摘要:** 事件可信度表示文本中事件的真实状况, 描述了事件是一个事实, 或者是一种可能性, 又或是一种不可能, 是自然语言处理中一个重要的语义任务。目前, 大多数关于事件可信度分析的方法都集中在句子级, 很少涉及篇章级的事件可信度分析。本文基于卷积神经网络, 结合篇章中的句子级特征 (包括句子的语义、语法以及线索词特征表示), 使用对抗训练来识别篇章可信度。在中英文数据集上的结果显示, 本文的方法与 Qian 的相比, 微平均 F1 值分别提高了 3.51% 和 6.02%, 宏平均 F1 值分别提升了 4.63% 和 9.97%。同时, 该方法在训练速度上也提高了 4 倍。

**关键词:** 事件可信度识别; 句子级特征; 线索词

## An Approach of Document-Level Event Factuality Identification Based on Sentence-Level Representations

ZHANG Liumin, ZHANG Yun, LI Peifeng

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

**Abstract:** Event factuality represents the factual nature of events in texts, and it describes whether an event is a fact, a possibility, or an impossible situation. Event factuality identification is an important semantic task in natural language processing. However, most of the current work focus on sentences level, and only a few aims at document-level. Based on the convolutional neural network, this paper introduces the sentence-level features in the text, including the semantic, grammar and clues of the sentence, and the confrontation training to identify the document-level factuality. Experimental results on both the Chinese and English corpus show that, the micro-average F1 is increased by 3.51% and 6.02%, respectively, and the macro-average F1 is increased by 4.63% and 9.97%, respectively compared with the baseline. The training speed of this method is also increased by four times at the same time.

**Keywords:** event factuality identification; sentence-level representation; clue word

### 1 引言

事件可信度表达了相关来源对事件事实本质的承诺, 表达了事件被描述为事实、可能或不可能的情况。篇章事件可信度就是一篇文章中事件的事实性状态, 是自然语言处理的重要基础, 可用于问答系统<sup>[1]</sup>、信息抽取<sup>[2]</sup>和篇章语义分析<sup>[3]</sup>等。事件可信度识别主要分为句子级以及篇章级事件可信度识别。其中, 句子级可信度识别是识别句子中事件的可信度; 篇章级事件可信度识别则是识别一篇文章中主要事件的可信度。在例 1 所示的文档中多次谈到“双号”这一事件。

在各个事件句 (包含事件的句子) 中, 该事件的可信度值就是其句子级可信度。如例 1 的首句“爱尔兰乐透开奖现一球双号 (CT-) 回应: 光的反射”中, 事件“双号”的回应是对于事件的否定, 因此, 该句中事件“双号”句子级可信度值为“一定不发生/CT-”。在其他涉及该事件的句子中, 其句子级可信度值有“一定发生/CT+”或“可能发生/PS+”。但从全文角度综合考虑, 该事件的篇章级可信度值为“一定不发生/CT-”。一个事件在文章中各个句子中的可信度值可以不同, 但是它的篇章级可信度值唯一。

例 1 爱尔兰乐透开奖现球双号 (CT-)

**项目基金:** 61836007, 61772354, 61773276

**作者简介:** 张刘敏 (1995-), 女, 硕士研究生, 主要研究领域为自然语言处理; 张赟 (1993-), 男, 硕士研究生, 主要研究领域为自然语言处理; 李培峰 (1971-), 男, 教授, 博士生导师, 主要研究领域为自然语言处理和机器学习。

回应：光的反射。

爱尔兰乐透奖开奖直播惊现一球双号 (CT+) 引争议。

(环球网综合报道) 据英国《每日邮报》10月8日报道, 当地时间10月7日晚, 爱尔兰乐透加一在电视上进行开奖直播, 公布价值450万英镑(约合人民币3906万元)乐透大奖的中奖号码。

然而在开奖时, 38号红球上却惊现数字“33”, 这一现象让观众很困惑, 他们质疑镜头画面是否被改动过, 但组织方却回应称, 这只是“光的反射”问题。

这一现象出现后, 热心的彩民们纷纷在社交媒体上发帖, 质疑爱尔兰乐透开奖中一球双号 (PS+) 的现象。

对此, 爱尔兰国家彩票局的一位发言人称, 这只是一种“幻象”, 是由“光的瞬时反射”所致。

她表示, 一球双号 (CT-) 的情况不可能存在, 因为所有乐透球的重量、大小和数字在开奖前都经过了严格的检查。

此外, 乐透开奖的所有流程都受毕马威会计事务所的审计员监督。

除了38号争议球外, 此次开奖被抽中的数字还有1、4、29、45、46以及奖金球26号。

句子级可信度识别只需要考虑该事件在句内的可信度, 可信度值可以是一定发生、一定不发生、可能发生等; 而对于篇章级可信度识别来说, 需要综合考虑全文, 理解全文的内容, 才能判断该事件在全文的可信度值。目前, 由于缺乏篇章级可信度相关的语料库, 篇章级可信度的研究还处于起步阶段。仅有的一些关于中文的工作是 Qian<sup>[4]</sup> 构建了一个篇章级事件可信度识别语料库, 提出了一种基于对抗网络的篇章级事件可信度识别模型。但是, Qian<sup>[4]</sup> 模型中输入采用的是篇章级输入, 将一篇文章中的所有的事件句或者依存路径连接在一起编码, 使用 BILSTM 和 Attention 进行篇章特征捕获, 但是 BILSTM 神经网络存储单元有限, 这种篇

章级的输入不仅会让神经网络遗忘之前的特征信息, 而且不利于神经网络学习句子级特征。同时 BILSTM 的神经元依赖之前神经元的输入, 不能实行并行操作, 无法有效地利用计算资源, 需要更长时间来进行模型训练。

为了解决 Qian<sup>[4]</sup> 存在的问题, 本文将直接使用单层卷积神经网络, 通过结合篇章中句子级表示进行篇章级别的事件可信度识别。本文将从篇章中的句子出发, 先通过卷积神经网络<sup>[5]</sup> 学习篇章中句子级特征, 包括句子级的语义、语法和否定、不确定线索词信息, 比如说例1中的其中一个句子级语义特征“她表示, 一球双号的情况不可能存在, 因为所有乐透球的重量、大小和数字在开奖前都经过了严格的检查。”之后拼接“不可能”这个否定线索词; 然后进行向量拼接, 通过全连接网络来学习篇章级特征, 判断文章的篇章级可信度。

拼接不确定、否定线索词的句子级输入与篇章级输入相比优势在于:

- 1) 句子级特征表示比较短, 增加的不确定、否定线索词信息, 可以使得神经网络学习句子中重点内容。同时对一篇文章中的所有事件句分别提取特征, 不仅学习到了句子的语义特征, 还学习到了篇章中所有句子之间的结构特征;
- 2) 采用单层卷积神经网络可以大大缩短模型的训练时间, 减少模型的复杂性。本文的模型识别篇章可信度的准确性比 Qian<sup>[4]</sup> 的在中英文数据集上微平均 F1 值分别提高了 3.51% 和 6.02%, 宏平均 F1 值分别提升了 4.63% 和 9.97%。该方法在训练速度上也比 Qian<sup>[4]</sup> 提高了 4 倍。

本文的结构如下: 首先简要介绍了该领域目前的相关工作; 然后主要介绍本文的模型架构以及实现的细节; 接着介绍实验的相关设置和结果分析; 最后是对全文工作的总结以及对于之后工作的展望。

## 2 相关工作

本节将从事件可信度的语料库以及相关识别方法两个方面来介绍事件可信度识别领域的相关工作。

首先介绍事件可信度语料库的相关工作。目前,可信度识别这个领域的语料库主要都是关于句子级可信度的语料库。Rubin<sup>[6]</sup>提出了一种识别文本中的确定性信息的理论框架,并且标注了新闻文本中显示的确定性信息;Sauri<sup>[7]</sup>构建了基于 TimeBank 的英文事件可信度语料库 FactBank,目前大多数英文的句子级可信度研究工作在此语料库上进行;曹媛<sup>[8]</sup>在 ACE2005 中文事件语料库的基础上,标注了各个事件的可信度及其相关属性。而篇章级事件可信度语料库的相关工作比较少,Qian<sup>[9]</sup>从中国日报网和新浪双语新闻上选取了 1730 篇英文文档和 4650 篇中文文档,标注了一个篇章级事件可信度语料库。标注的事件可信度分为“不清楚/Uu”、“一定不发生/CT-”、“可能不发生/PS-”、“可能发生/PS+”、“一定发生/CT+”,语料库中同时标注了句子级事件可信度以及篇章可信度。

由于句子级事件可信度语料比较多,因此主要研究工作也集中在句子级事件可信度识别中,分为三类:基于规则方法、基于统计方法以及基于神经网络。基于规则的方法中,Sauri<sup>[10]</sup>和 Sauri<sup>[11]</sup>提出了计算事件可信度的算法;基于统计方法中,Sauri<sup>[11]</sup>和 Werner<sup>[12]</sup>均采取了基于支持向量机 SVM 和特征工程的方法,Qian<sup>[13]</sup>提出了一个统计和规则相结合的模型。基于神经网络的方法中,Qian<sup>[14]</sup>提出了一个基于生成式对抗网络的方法,将句子级可信度识别分为两步,第一步从生语料中抽取特征,然后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。在篇章级事件可信度识别中,Qian<sup>[4]</sup>在自主标注的语料库上提出了一种基于对抗网络的篇章级事件可信度识别模型。该模型通过使用 BILSTM 对包含事件的句子信

息和句法路径进行编码,通过 Attention 提取句子语义和句法信息中主要特征;将编码后的信息融合在一起,得到综合全文的事件信息的特征表示;然后将全文特征输入到全连接网络中进行特征学习;最后,将特征输入到 softmax 层对事件进行分类识别。本文的工作是在 Qian<sup>[4]</sup>的基础上进行。

## 3 模型结构

在篇章级事件可信度识别上,Qian<sup>[4]</sup>提出了基于对抗网络的篇章级事件可信度识别,模型采用 BILSTM + Attention。其中句法信息方面 Qian<sup>[4]</sup>认为事件句的前后邻句对判断事件可信度有重要作用。根据当前句子的邻句中是否含有不确定或否定线索词,将句法路径定义为当前事件到达当前句子根节点的依存句法路径,以及邻句线索词到达其所在根节点的依存句法路径。然后将两个根节点合并起来。Qian<sup>[4]</sup>采用的是篇章级输入,即将一篇文章中所有的事件句或者句法路径连接在一起编码输入到 BILSTM 和 Attention 之中。虽然 BILSTM 可以学习远距离的信息,但是存储单元毕竟有限,篇章级输入的长文本会增加神经网络学习序列的难度,无法准确捕获到句子中的特征。并且由于 BILSTM 依赖上一层的计算结果,无法实现并行运算,导致模型训练的时间比较长。

由于 Qian<sup>[4]</sup>模型存在的文本信息遗失以及训练速度慢的问题,本文采用单层卷积神经网络模型。不同于 Qian<sup>[4]</sup>的篇章级序列特征,本文的输入编码是将文章中事件所在句的句子级序列表示进行编码。除了句子信息以及依存路径信息,本文还考虑将句子中的否定以及不确定线索词信息加入到句子的语义表示之中。本文模型结构如图 1 所示,包含输入层、卷积层、池化层以及可信度识别层。

### 3.1 模型输入

(1) 事件所在的句子。首先通过对语

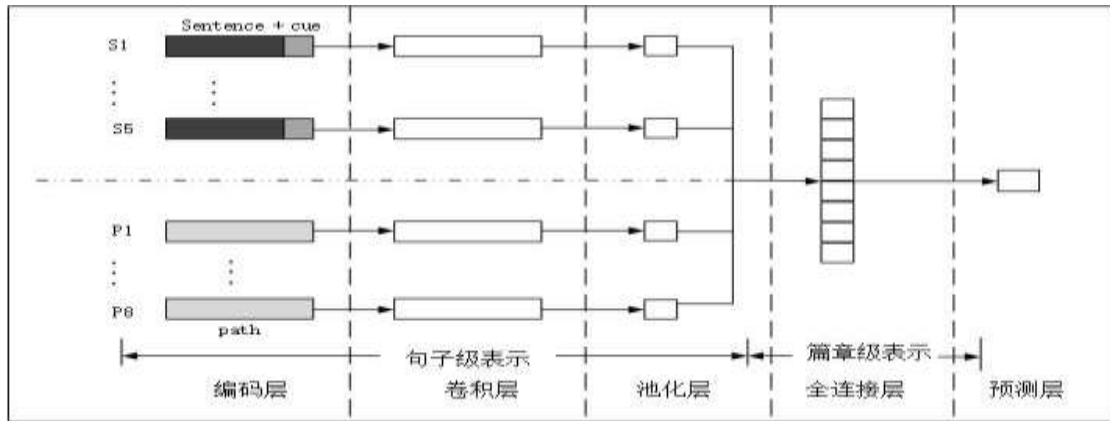


图 1 模型结构图

料库的统计以及训练，可以得到词语的低维分布式表示，接着联合 word2vec 将事件句进行预训练，得到模型所需要的句子语义输入，表示为  $S_1$ 、 $S_2$ 、……、 $S_n$ 。其中， $S_1$  表示所有文章中第一句事件句的矩阵，以此类推， $S_2$  表示所有文章中第二个事件句的矩阵。

(2) 事件所在句中包含的否定线索词或者不确定线索词。本文将采用字典匹配的方法，根据不确定以及否定线索词字典识别出事件句中的不确定和否定线索词

(3) 不确定和否定线索词到达事件的依存句法路径。如果邻句中存在不确定或者否定线索词，则其到达当前句子中事件的依存路径为：该线索词到达邻句根节点的依存路径，加上当前句子根节点到达事件的依存路径，两句的根节点合并为一个根节点。该特征只是简单考虑了句子中否定或者不确定特征对于邻句事件的影响，依存路径中只包含到达线索词的路径信息，未考虑线索词自身的语义特征对于篇章可信度识别的影响，表示为  $P_1$ 、 $P_2$ 、……、 $P_n$ 。与 (1) 中的表示类似，其中  $P_1$  表示所有文章中第 1 句的句法路径的矩阵， $P_2$  表示所有文章中第 2 句句法路径的矩阵。

(2)、(3) 中包含的否定和不确定线索词的来源如下：英文线索词来自 BioScope 语料库，中文线索词来自 CNeUn 语料库。将得到的否定、不确定线索词向量 (2) 拼接到句子的语义向量 (1) 后面，丰富句子

的语义表示。所得到的向量都通过 word2vec 工具进行预训练得到需要的序列矩阵。

单句编码的句子级输入序列均使用矩阵表示，形状大小为  $m \times n \times d_0$ ，其中， $m$  表示表示所有文章的数量， $d_0$  词嵌入向量的维度， $n$  表示单句序列长度，固定大小为 120。

与 Qian<sup>[4]</sup> 相同的篇章级输入的矩阵大小为  $m \times a \times d_0$  或是  $m \times b \times d_0$ ，全文信息直接记作 *Full*。其中  $m$  表示所有文章的数量， $a$  表示每篇文章中所有句子的长度大小，固定大小为 120， $b$  表示每篇文章中所有句法路径的长度，固定大小为 70， $d_0$  表示词嵌入向量的维度。

篇章级输入与句子级输入相同的是  $m$  以及  $d_0$ ，不同的是矩阵的第二维。其中，句子级输入矩阵的第二维表示每篇文章中一个句子或者句法路径的长度，而篇章级输入矩阵表示的是每篇文章中所有句子或者句法路径拼接之后的长度。

### 3.2 特征提取

本文训练模型使用单层卷积神经网络、全局最大化和全连接网络。

采用单层的卷积神经网络的原因如下：一是因为本文采用的是句子级特征表示输入，输入的序列长度都不是特别长，卷积神经网络可以有效的学习到句子的特征；二是因为卷积神经网络可以进行并行运算，大大加快了模型的运行速度，更具有实用性。

本文通过卷积神经网络, 获得序列的表示  $R_{si}$  或是  $R_{pi}$  (其中  $s$  与  $p$  分别代表句子和依存路径,  $i$  表示篇章中第几个句子或依存路径)。作为卷积神经网络的输出, 序列的表示  $R$  为:

$$R_{si} = S_i * W_1 + b_1 \quad (1)$$

$$R_{pi} = P_i * W_2 + b_2 \quad (2)$$

$$R = R_{si} \oplus R_{pi} \quad (3)$$

其中,  $\oplus$  表示将句子和依存路径通过连接的方式拼接在一起,  $W_1$ 、 $W_2$  是参数矩阵。

由于句子级特征表示经过全局最大化, 提取了每句中最活跃的信息, 篇章文本长度缩短, 然后通过全连接网络学习整篇文章的语义信息以及句法特征更加清晰、全面、迅速。

### 3.3 对抗训练

为了提高模型的鲁棒性, 类似之前研究者的工作, 本文对模型增加了小幅的摄动。本文采用 Fast Gradient Sign Method (FGSM) 来计算出摄动, 添加到词向量上, 利用修改后的词向量, 进行模型训练, 得到扰动后的交叉熵损失, 按照系数的不同与原有的交叉熵损失相加得到新的 Loss, 通过优化这个 Loss 来得到鲁棒性更好的模型。

## 4 实验

### 4.1 数据集以及设置

本文实验所用的数据集与 Qian<sup>[4]</sup> 一致。英文数据集上, 篇章级值为 CT- 中句子级和篇章级可信度不同的占篇章级值为 CT- 的总文档的 41.94%; 类似的, 篇章级为 PS+ 中句子级和篇章级可信度不同的占篇章级为 PS+ 的 66.42%, 而篇章级值为 CT+ 的文档仅占 12.52%。中文数据集里面, 篇章级值为 CT- 中句子级和篇章级可信度不同的文章占篇章级值为 CT- 的总文档的 63.79%; 类似的, 篇章级为 PS+ 中句子级和篇章级可信度不同的占篇章级为 PS+ 的 62.38%; CT+ 的文

档仅占 14.23%。以上数据可以看出句子级可信度与篇章级可信度冲突比较多的是 CT- 以及 PS+, 识别难度比较大, CT+ 的识别难度较小。

实验在中英文语料库上进行了 10 折交叉验证。在每折中, 按照 8:1:1 来划分训练集、验证集以及测试集。使用准确率 P (Precision)、召回率 R (Recall) 以及 F1 值 (F1-measure) 来描述其性能, 同时通过微平均 (Micro-average) 和宏平均 (Macro-average) 来描述对各个分类的综合性能。所有的向量都通过 word2vec 进行初始化, 本文的句子长度固定长度为 120, 句法路径的长度固定为 70。本文卷积神经网络使用的窗口大小为 2, 卷积核的数量为 1024, 提取卷积神经网络的特征使用 globalmaxpooling1d 函数。

### 4.2 实验模型

由于无论是中文还是英文数据集中关于 PS- 以及 Uu 的语料太少了, 分别占据文档的 1.22% 和 1.56%。因此, 本文将和 Qian<sup>[4]</sup> 一样, 主要考虑实验中 CT+, PS+ 以及 CT- 的性能。为了验证本文句子级特征的性能, 本文把 Qian<sup>[4]</sup> 以及输入与 Qian<sup>[4]</sup> 相同的卷积神经网络作为基准系统。在中英文数据集上均做了实验, 实验结果表明本文所选取的语义特征以及句子间结构的有效性。

为了验证本文模型结构的有效性, 本文考虑了以下两个基准系统:

**Qian<sup>[4]</sup>:** 该基准系统与本文模型的区别在于, 采用篇章内序列相连的篇章级序列矩阵。

**Full:** 输入为多句连接的篇章级序列矩阵, 与 Qian<sup>[4]</sup> 的输入相同。

本文对比实验中的模型和输入如下:

**Full\_Neg:** 输入为篇章中多句连接的篇章级序列矩阵, 每个句子向量拼接否定线索词向量, 每个句子向量的大小仍然设为 120。

表 1 本文系统 Sen\_Neg 和 Qian 等基准系统的性能对比

语言	模型	CT+(P/R/F1)	CT-(P/R/F1)	PS+(P/R/F1)	微平均	宏平均
中文	Full	86.85/90.97/88.84	78.14/80.97/79.35	76.90/65.77/70.44	83.52	80.19
	Qian	85.36/89.86/87.52	83.68/83.17/83.35	79.05/69.74/74.06	84.03	81.64
	Sen_Neg	89.98/90.01/90.01	86.11/87.00/86.46	81.58/82.89/82.13	87.54	86.27
英文	Qian	88.76/91.04/89.84	77.21/77.78/76.87	66.89/58.46/62.14	83.56	76.28
	Full	90.70/92.98/91.81	82.18/81.98/81.94	75.72/73.21/74.03	86.92	82.77
	Sen_Neg	93.25/92.68/92.95	85.85/88.52/87.10	75.56/81.91/78.36	89.58	86.25

表 2 中文数据集上不同输入的实验性能对比

	模型	CT+(P/R/F1)	CT-(P/R/F1)	PS+(P/R/F1)	微平均	宏平均
中文	Full	86.85/90.97/88.84	78.14/80.97/79.35	76.90/65.76/70.44	83.51	80.19
	Full_Ps	87.13/88.67/87.85	82.58/82.14/82.23	76.69/77.61/77.07	84.24	82.51
	Full_Neg	87.79/89.18/88.43	86.59/84.97/85.72	79.04/80.54/79.67	85.26	83.42
	Sen_Neg_Ps	89.66/89.51/89.52	83.95/85.74/84.73	79.52/81.02/80.20	86.38	84.89
	Sen	89.07/90.15/89.04	86.59/84.97/85.72	79.04/80.54/79.67	86.87	85.17
	Sen_Ps	90.18/90.04/90.08	86.22/86.06/86.05	78.78/83.17/80.87	87.21	85.73
	Sen_Neg	89.98/90.01/90.01	86.11/87.00/86.46	81.58/82.89/82.13	87.54	86.27

表 3 英文数据集上不同输入的实验性能对比

	模型	CT+(P/R/F1)	CT-(P/R/F1)	PS+(P/R/F1)	微平均	宏平均
英文	Full	90.67/92.98/91.81	82.18/91.99/91.94	75.73/73.21/74.03	86.92	82.77
	Full_Ps	90.05/93.79/91.82	82.32/84.59/83.24	74.03/68.19/70.46	87.09	82.11
	Full_Neg	91.06/93.42/92.17	83.14/87.39/84.82	77.69/72.49/76.59	88.35	84.01
	Sen_Neg_Ps	92.16/94.33/93.18	82.92/86.22/84.36	78.31/73.67/75.41	89.06	84.54
	Sen	91.78/94.09/92.88	83.80/84.37/83.81	78.03/76.09/76.59	88.89	84.63
	Sen_Ps	92.36/92.85/92.58	84.89/85.04/84.75	75.42/79.23/77.06	88.82	84.93
	Sen_Neg	93.25/92.68/92.95	85.85/88.52/87.10	75.56/81.91/78.36	89.58	86.25

**Full\_Ps:** 输入为篇章中多句连接的篇章级序列矩阵, 每个句子向量拼接不确定线索词向量, 每个句子向量的大小仍然设为 120。

**Sen:** 基于卷积神经网络, 输入为篇章中单句编码的句子级序列矩阵。

**Sen\_Ps:** 输入为单句编码的句子级序列矩阵, 每个句子向量拼接不确定线索词向量, 每个句子向量的大小仍然设为 120。

**Sen\_Neg:** 输入为单句编码的句子级序列矩阵, 每个句子向量拼接否定线索词向量, 每个句子向量的大小仍然设为 120。

**Sen\_Neg\_Ps:** 输入为单句编码的句子级序列矩阵, 每个句子向量按照线索词在事件句中出现的顺序拼接否定和不确定线索词

向量, 每个句子向量的大小仍然设为 120。

表 1 显示了本文模型与两个基准系统性能对比。其中, 相比 Qian<sup>[4]</sup>的性能, 本文模型在中英文数据集上微平均 F1 值分别提高了 3.51%和 6.02%, 宏平均 F1 值分别提升了 4.63%和 9.97%; 相比与 Full 的性能, 本文模型在中英文数据集上微平均 F1 值分别提高 5%和 2.6%, 宏平均 F1 值分别提升 2.6%和 3.5%。

### 4.3 实验分析

表 2 和表 3 给出了在中文、英文数据集上不同输入的实验性能对比。从表 2 和表 3 可以发现:

- 1) 从 *Full* 以及 *Sen* 的性能可以看出来, 当改变了输入编码方式, 先进行句子级特征提取再进行篇章级特征提取要比一开始就进行篇章级特征提取的性能优越。*Sen* 在中英文数据集上微观 F1 值比 *Full* 模型的性能分别提高了 3.3% 和 1.9%, 宏观 F1 值分别提高了 5% 以及 1.9%。分析得出, 单句编码的句子级输入相比于多句连接的篇章级编码输入有以下几个方面的优势: 一是句子级的编码输入实际上隐藏考虑了事件在文章中的相对位置。新闻文本有比较明显的结构特征, 一般来说, 新闻文本的首句或者结尾句都是一篇文章的总结句。考虑句子相对位置的句子级编码输入效果比篇章级编码输入明显提升很多; 二是本文在经过卷积神经网络提取特征之后, *Full* 与 *Sen* 都经过了一个 *global-maxpool1d*, 提取第二维度上最活跃的特征, 但是 *Full* 中的篇章级输入导致序列长度太长, 在获得第二维最活跃特征的时候就会丢失更多的信息; 三是句子级文本长度较短, 卷积神经网络已经能够很好的捕获文本特征。
- 2) 从 *Full*、*Full\_Neg*、*Full\_Ps*、*Sen\_Neg\_Ps*、*Sen*、*Sen\_Ps* 以及 *Sen\_Neg* 模型中可见, 在篇章级编码输入中无论是增加不确定线索词还是否定线索词, 都可以增加篇章可信度识别的准确性。即说明不确定、否定线索词对于神经网络识别篇章可信度的有效性。但是, 从表 2 和表 3 中可以明显看出来拼接不确定线索词的性能效果没有拼接否定线索词的好。其中, 在中英文数据集上, *Full\_Neg* 的微观 F1 值相比于 *Full* 分别提高了 1.7% 和 1.2%, 宏观 F1 值分别提高了 3.3% 和 1.3%。分析数据集可以发现, 如例 2 中有“从未 (*neg\_cue*)”、“考虑 (*ps\_cue*)”、“未 (*neg\_cue*)”、“否认 (*neg\_cue*)”四个不同的线索词, 根据实际的句意, 本句中的“辞职”实际上已经被“从未 (*neg\_cue*)”这

个否定线索词将句子可信度值变成了一定不发生/CT-, 当我们只增加不确定线索词“考虑”的时候, 扰乱了文本信息。因此可以看出, 增加线索词符合人类的判断可信度标准, 是比较有效的语义特征, 有助于神经网络理解文章的语义特征, 不过需要考虑线索词的优先级别。但是, 当同时增加不确定以及否定线索词的时候, 模型的效果相对于只增加否定或者不确定线索词下降。从表 2 和表 3 可以看出来下降比较明显的类别是在 CT-以及 PS+上的性能。分析数据集以及模型得出以下结论: 一是句子级编码输入文本长度比较短, 卷积神经网络已经能够学习到线索词的信息; 二是有些句子中的不确定或者否定线索词比较多, 神经网络无法准确提取特征。比如说例 2 中一共有 4 个线索词, 包括不确定以及否定线索词, 两种线索词混合连接, 神经网络无法准确判断哪一个线索词最有效。

例 2 美国务卿称从未考虑辞职未否认曾称特朗普是白痴。

- 3) 本文能够提高篇章可信度识别准确性的主要原因在于 CT-以及 PS+两个类别识别准确性的提升。分析数据集以及模型可以看出来, 主要在于增加的否定、不确定线索词信息使得神经网络能够更加容易地学习到线索词的信息。例如, 一般想要判断一个事件一定不发生, 必然需要有确定的否定线索词。原来将 CT-判断为 PS+, 由于增加了否定线索词, 能够准确判断 CT-。一个类别性能的提升必然会导致另外一个类别性能的提升。因此不仅是 CT-类别的准确性上升, PS+的性能也相应上升。
- 4) 本文的模型方法在英文数据集上的效果明显优于中文数据集上。对比中英文各个类别的性能, 在英文数据集上 CT+的性能较好, 中文数据集上 PS+性能较好。分析数据集以及模型得出, 一是英文表达比较直接, 否定、不确定都有明

确的单词表示,有利于卷积神经网络获取特征;二是英文数据集比较小,卷积神经网络更容易获取特征;三是由于数据集的不均衡造成,文章中句子可信度与篇章可信度不相同的文档中,占比比较低的篇章可信度识别难度会相应下降。

- 5) 由于卷积神经网络相对于 BILSTM 能够并行运算,基于本文方法的模型训练速度是之前工作的 1/4。

## 5 总结与展望

本文提出的基于单句表示的篇章可信度识别,使用卷积神经网络捕获句子级的空间以及语言学特征,即句子在文章中的位置以及否定、不确定线索词特征;然后通过全连接网络学习篇章的序列特征;最后判断该事件在篇章级别中的可信度值。实验结果表明本文的方法无论是宏平均 F1 值和微平均 F1 值,还是在各个类别的准确性上都明显优于 Qian<sup>[4]</sup>。在接下来的工作中,一是考虑通过工具提取句子的主干信息,再拼接一些线索词信息,使得神经网络更加容易的理解学习篇章中的信息,提升篇章可信度识别的准确性;二是考虑将全文信息编码成图神经网络,增加篇章中句子与句子之间的关系信息,从而提升篇章可信度识别的准确性。

### 参考文献.

- [1] Yu A W, Dohan D, Luong M T, et al. Qanet: Combining Local Convolution with Global Self-Attention for Reading Comprehension[J]. arXiv preprint arXiv, 2018:1804.09541.
- [2] Yan Z, Tang D, Duan N, et al. Assertion-Based QA with Question-Aware Open Information Extraction[C]//Proc. of Thirty-Second AAAI Conference on Artificial Intelligence. arXiv preprint arXiv, 2018:1801.07414.
- [3] 褚晓敏, 朱巧明, 周国栋. 自然语言处理中的篇章主次关系研究[J]. 计算机学报, 2017, 40(4): 842-860.

- [4] Qian Z, Li P, Zhu Q, et al. Document-Level Event Factuality Identification via Adversarial Neural Network[C]//Proc. of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019:2799-2809.
- [5] Kim Y. Convolutional Neural Networks for Sentence Classification[J]. 2014:1746-1751.
- [6] Rubin V L, Liddy E D, Kando N. Certainty Identification in Texts: Categorization Model and Manual Tagging Results[J]. Information Retrieval, 2005, 20: 61-76.
- [7] Saur í R, Pustejovsky J. FactBank: A Corpus Annotated with Event Factuality[J]. Language Resources and Evaluation, May 2009, 43(3): 227-268.
- [8] 曹媛, 朱巧明, 李培峰. 中文事件事实性信息语料库的构建方法[J]. 中文信息学报, 2012, 27(6): 38-44.
- [9] 钱忠. 事件可信度识别方法研究[D]. 苏州大学, 2018.
- [10] Saur íR. A Factuality Profiler for Eventualities in Text[D]. PhD Thesis, Waltham, MA, USA: Brandeis University, 2008.
- [11] Klenner M, Clematide S. How Factuality Determines Sentiment Inferences[C]//Proc. of the Fifth Joint Conference on Lexical and Computational Semantics, 2016:75-84.
- [12] Werner G J, Prabhakaran V, Diab M, Rambow O. Committed Belief Tagging on the Factbank and Lu Corpora: A Comparative Study[C]// Proc. of the Second Workshop on Extra-Propositional Aspects of Meaning in Computational Semantics, 2015:32-40.
- [13] Qian Z, Li P F, Zhu Q M. A Two-step Approach for Event Factuality Identification[C]//Proc. of 2015 International Conference on Asian Language Processing, 2015:103-106.
- [14] Qian Z, Li P F, Zhang Y, et al. Event Factuality Identification via Generative Adversarial Networks with Auxiliary Classification[C]//Proc. of the Joint Conference on Artificial Intelligence, 2018: 4293-430.



---

联系方式

本文负责人：张刘敏；电话：18351037696，

E-mail:20185227086@stu.suda.edu.cn

通讯作者：李培峰；电话：13606216291，E-mail:

pfli@suda.edu.cn