

基于多维度行文风格特征的假新闻检测方法研究

李小艳

(贵州大学计算机科学与技术学院, 贵州 贵阳 550000)

✉946730793@qq.com



摘 要: 传统的基于行文风格的假新闻检测方法对文本的行文风格特征考虑不全面, 只考虑符号、句长、句数和特征词, 而忽略情绪词、虚词、实词等行文风格特征, 并且不能很好地提取真假新闻行文风格之间的差异性信息, 因此探索一种基于词、句、篇三个维度的行文风格特征, 用于自动检测社交媒体上低可信度文本的方法。充分研究文本多维度行文风格特征对假新闻检测的作用, 使用多通道卷积神经网络提取行文风格的高阶抽象信息, 利用注意力机制捕获各维度特征对假新闻检测的影响力。实验显示, 提出的Multi_CNNSA模型显著提高了假新闻检测效果, 在weibo数据集上取得86.95%的F1值。

关键词: 假新闻检测; 行文风格特征; 卷积神经网络

中图分类号: TP399 **文献标识码:** A

Research on Fake News Detection Method based on Multi-dimensional Writing Style Features

LI Xiaoyan

(School of Computer Science and Technology, Guizhou University, Guiyang 550000, China)

✉946730793@qq.com

Abstract: Traditional fake news detection methods based on writing style do not fully consider the writing style characteristics of the text. They only consider symbols, sentence length, sentence number and feature words, while ignoring the writing style features such as emotional words, functional words, and substantial words. They are not good enough to extract difference information between the writing styles of true and fake news. In view of these problems, this paper proposes to explore a method for automatic detection of low-credibility texts on social media based on the three-dimensional writing style features of words, sentences, and contexts. After a full study on the effect of text multi-dimensional style features on fake news detection, multi-channel convolutional neural network is used to extract high-level abstract information of style, and attention mechanism is used to capture the influence of each dimension feature on fake news detection. Experiments show that the proposed Multi_CNNSA model significantly improves the effect of fake news detection, achieving an F1 value of 86.95% on the weibo dataset.

Keywords: fake news detection; writing style features; convolutional neural network

1 引言(Introduction)

近年来, 互联网快速发展, 在为信息传递带来便利的同时, 也成为假新闻传播的摇篮^[1], 如2016年美国总统大选期间, 网站上充斥着各种耸人听闻的总统大选假新闻^[2]。更严重的是, 由于此类假新闻持续、广泛地传播, 最终间接影响了总统选举的结果, 因此假新闻检测开始为人们所重视。

研究表明, 社交网络上的信息快速传播, 导致网络平台

上的信息存在爆炸性、持续性增长的特点^[3]。在这样的情形下, 人工识别和检测假新闻的方法存在效率低、具有时滞性等问题, 往往难以满足假新闻检测的时效性及大规模可应用性等需求, 更遑论筛选和阻止假新闻的传播。因此, 如何有效的识别和检测网络平台上大量的假新闻成为一个急需解决的难题。如果能研究出一个有效的假新闻检测方法, 不仅能减少相关监管部门的工作量, 而且能减少网络舆论带来的负

面影响,还能增强网络媒体的社会公信力,从而构建一个更加和谐稳定的网络生态环境。

目前,对网络平台上的假新闻自动检测研究仍处于起步阶段,大部分的研究工作都将这一问题作为分类任务来处理,可分为基于传统文本分类的假新闻检测和基于深度学习文本分类的假新闻检测两大类方法。

(1)基于传统文本分类的假新闻检测方法,主要是将整个文本分类问题拆分成特征工程和分类器两部分。特征工程可以分为文本预处理、特征提取、文本表示三个部分,最终的目的是要把文本转换成计算机可理解的格式并封装。特征工程构造的特征主要包括浅层的统计特征,如假新闻的内容^[4]、用户属性^[5]、传播方式^[6]等,以及深层的文本内容特征,如假新闻事件情感倾向性^[7]、事件主题^[8]、事件关键词^[9]等。在得到特征工程构造的特征后,分类器基于构造的特征来进行文本分类,从而实现假新闻检测的功能。

(2)基于深度学习文本分类的假新闻检测方法,主要是通过深层的神经网络去探索真假新闻之间的差异,从而实现假新闻检测。典型的方法或模型如O'BRIEN^[10]等人利用卷积神经网络(CNN)从新闻主题的角度检测假新闻,表明CNN可以捕获虚假和真实新闻语料在语法修辞上的差异性,从而实现假新闻的识别工作;UMER等人^[11]通过使用CNN-LSTM混合神经网络,在文本特征向量传递到分类器之前对其进行降维操作,以此提高假新闻检测的准确率。目前主流的XLNet^[12]、BERT^[13]等模型也被用于假新闻检测。

对假新闻进行检测主要有四个途径^[14]。(1)基于知识库的方法:该方法利用已知的信息做一个知识库,判断另一个信息的真假。PAN等人^[15]使用已有数据集分别构建真假知识图谱,从而对新闻进行检测。(2)基于行文风格的方法:该方法通过寻找真假新闻可区分的行文风格特征,判断新闻的真假。行文风格特征有普通特征(字频、词频、句法和篇章)和聚合特征(普通特征的融合)。PRZYBYLA等人^[16]基于行文风格提出Stylometric Classifier模型用于检测网页源假新闻。(3)基于信息来源的方法:真假新闻的发布者有着不同的立场,基于用户对内容的操作(评论、点赞、举报等)构建矩阵或者图模型,从而判断其真假。WANG等人^[17]利用信息来源的用户画像作为一个特征,用于假新闻检测。(4)基于传播途径的方法:类似PageRank的行为传递。假新闻比真新闻传播速度快,而且传播的路径图也有所差异。因而基于传播途径的假新闻检测常转化为图的相似性比较。BIAN等人^[18]使用双向图卷积神经网络对社交媒体谣言进行检测。

研究发现,网络平台上的假新闻常为了快速传播或吸引网民的目光,在行文风格上往往使用不规则字符、煽动性言语等^[19],如图1所示。因此,针对以上问题,本文聚焦新闻行文风格^[20-21]的分析研究,提出了一种多维度行文风格特征的假新闻检测方法。该方法主要聚焦行文风格的词维度、句维度及篇维度来提取文本的特征因子,用于分析真假新闻文本之间存在的潜在关系,从而更加有效地对假新闻进行检测。

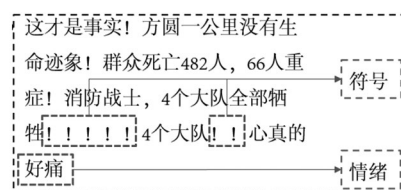


图1 假新闻行文风格实例

Fig.1 Examples of fake news style

2 相关工作(Related work)

2.1 假新闻定义

假新闻检测是一个新兴的问题,目前对假新闻的定义和识别没有明确的方案和较为严谨的说法,常见的定义如表1所示。因此,在研究之前我们必须明确假新闻的定义。在本次研究中,假新闻定义为:故意捏造并被证实为假的信息。

表1 假新闻定义

Tab.1 Definition of fake news

概念	真实性	意图	是否为新闻
虚假新闻	非事实	误导	是
假新闻	非事实	未定义	是
讽刺新闻	未统一	娱乐	是
故意的假情报	非事实	娱乐	未定义
误报	非事实	未定义	未定义
筛选	常见事实	误导	未定义
标题党	未定义	误导	未定义
谣言	未定义	未定义	未定义

2.2 行文风格特征分析

本文从多角度分析行文风格对假新闻检测的影响力,并根据词、句、篇的属性,将这些特征详细划分为三个维度,即词维度、句维度和篇维度,并结合文本的语义信息对假新闻检测进行研究。多维度行文风格特征的具体分析如下:

(1)词维度:实词、情绪词

词典是文本分类中经常使用的工具。POTTHAST等人^[22]在2017年使用带有词性标签的词典用于“超党派”新闻识别,该工具的缺点在于词典的大小有限。而在2020年,PRZYBYLA等人^[16]使用Word2vec表示,用相似词对实词词典进行扩展最终得到GI词典用于网页源假新闻检测,并取得一定成果。因此,本文将带有词性标签的实词划分为词维度行文风格特征。为了使GI词典更加丰富,本文将英文单词翻译为中文,并且为了防止原有词义的丢失,将单词的全部翻译结果作为词典的扩展,得到最终的GI词典。此外,假新闻常使用夸张的情绪词^[23]来吸引人们的目光,因此情绪词也被划分为词维度的行文风格特征。目前已公开的情绪词语集有很多,文本选用知网提供的情感分析用词语集,因为它有着丰富的情感词类别,分别为:主张、正面情感、负面情感、正面评价、负面评价和程度词语。

(2)句维度:句长、平均句长^[23]、句数

通过对数据集的研究,发现真假新闻文本的句长和平均句长也存在差异。因此,本文将句长、平均句长和句数作为

句维度的行文风格特征。

(3)篇维度：符号^[24]、虚词

对于整个篇章，本文关注文本的整体符号使用情况。除文本的语义信息，文章的叹词、助词、介词等虚词的使用也是文本的行文风格之一，而且据图2统计可以看出，在新闻的真假文本中，虚词的使用有很大的差异性，因此本文将虚词作为篇维度的行文风格特征。

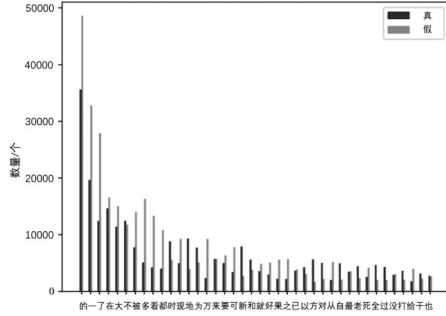


图2 常用虚词使用频率统计

Fig.2 Frequency statistics of common function words

2.3 语义特征及多维度行文风格特征提取方法

语义特征是文本分类的基础，做好内容理解，准确提取特征是文本分类的基石。卷积神经网络的核心思想是捕捉局部特征，对于文本来说，局部特征就是由若干单词组成的滑动窗口，类似于N-gram。卷积神经网络的优势在于能够自动地对N-gram特征进行组合和筛选，获得不同抽象层次的语义信息。因此本文使用Text_CNN提取文本的语义信息，Text_CNN模型如图3所示。

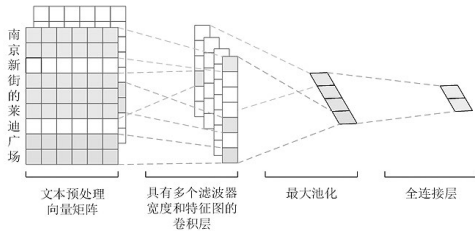


图3 Text_CNN模型

Fig.3 Text_CNN model

对于直观的行文风格特征，如GI词典、情绪词、句长、平均句长和句数，本文使用统计学习方法对其进行表示。对于较为复杂的复合特征如虚词和符号，本文使用词频-逆文本频率指数(TF-IDF)和互信息相结合的推荐算法筛选出特征的最优组合，详细介绍如下。

(1)TF-IDF

结合TF-IDF的思想，将其类比于符号(虚词)-新闻真假关系上，确定核心符号(虚词)，从而去掉不必要的噪声。具体公式如下：

$$S(s) = \left(\sum_{d[i] \in d_{set}} \frac{\text{count}(s \in d[i])}{\text{doc}[i]} \right) \times \log \left(\frac{\text{len}[d]}{\sum_{d[j] \in d} \text{count}(s \in d[j])} \right) \quad (1)$$

式(1)中， s 表示一种符号(虚词)， $d[i]$ 表示第 i 篇文本， d_{set} 表示所有文本形成的集合， $\text{doc}[i]$ 表示一篇文本包含的总符号(虚词)数， $\text{len}[d]$ 表示文本总量， $\sum_{d[j] \in d} \text{count}(s \in d[j])$ 表示包含某符号(虚词)的文本数。

(2)互信息

本文在式(1)的基础上，进一步挖掘一些深层知识，计算其与新闻真假的互信息，以得到最优的符号(虚词)组合特征。互信息的基本定义如下：

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (2)$$

应用到符号(虚词)特征选择：

$$I(U; C) = \sum_{e_t \in \{1,0\}} \sum_{e_c \in \{1,0\}} P(U = e_t, C = e_c) \log_2 \frac{P(U = e_t, C = e_c)}{P(U = e_t)P(C = e_c)} \quad (3)$$

其中， U 、 C 都是二值随机变量，当文档包含特征 t 时， U 的取值为 $e_t=1$ ，否则 $e_t=0$ ；当文档属于类别 c 时， C 的取值为 $e_c=1$ ，否则 $e_c=0$ 。用最大似然估计时，上面的概率值都是通过统计文档中特征和类别的数目计算的，因而实际计算公式如下：

$$I(U; C) = \frac{N_{11}}{N} \log_2 \frac{N N_{11}}{N_1 N_{.1}} + \frac{N_{01}}{N} \log_2 \frac{N N_{01}}{N_0 N_{.1}} + \frac{N_{10}}{N} \log_2 \frac{N N_{10}}{N_1 N_{.0}} + \frac{N_{00}}{N} \log_2 \frac{N N_{00}}{N_0 N_{.0}} \quad (4)$$

其中， N_{xy} 表示 $x=e_t$ 和 $y=e_c$ 情况下对应的文档数目，比如 N_{10} 表示包含特征 t (此时 $e_t=1$)但不属于类别 c (此时 $e_c=0$)； $N_{1.}=N_{10}+N_{11}$ 表示所有包含特征 t 的文档数目， $N_{.1}=N_{11}+N_{01}$ 表示所有属于类 c 的文档数目， $N=N_{00}+N_{01}+N_{10}+N_{11}$ 表示所有文档的数目。

3 模型构建(Model building)

本文基于行文风格特征的方法，从多维度提取行文风格特征，并结合文本的语义信息对假新闻进行检测。其中提取的行文风格特征来自不同的维度，若将其放入一个神经网络进行训练，会模糊特征之间的潜在关系，因此本文使用CNN模型分别捕获各特征的高阶抽象信息。此外每个维度的特征因为其性质不同，对假新闻检测的影响力度也不同，而Attention机制可以看成一种自动加权机制，它可以把多个不同模块通过加权的方式进行关联，在NLP任务中有着广泛的应用。因此，本文使用Attention机制突出各维度特征对假新闻检测的重要程度，从而构建了一个Multi_CNNSA模型，该模型结构图如图4所示。

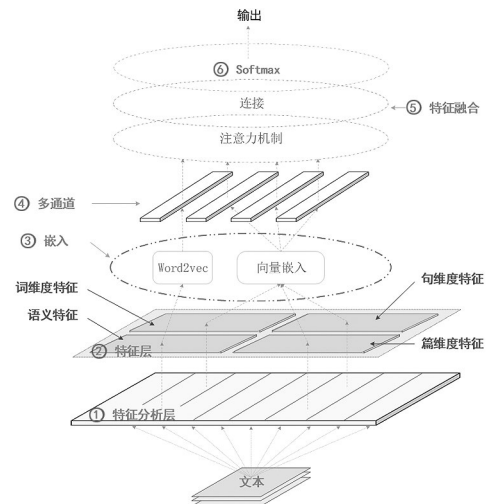


图4 Multi_CNNSA模型

Fig.4 Multi_CNNSA model

3.1 输入层

令 T 为新闻文本，对文本进行预处理，得到 $T_1=t_1, t_2, \dots$ ，其中 t_i 表示一个词。令预处理好的Word2vec词向量查找表为 $R^{V \times D}$ ， V 是查找表的大小， D 是向量矩阵 $X=[x_1, x_2, x_L]$ ，其中 $x_i \in R^{V \times D}$ ， L 为最大文本长度，长度不足的文本补零。根据前文的行文风格特征处理方法得到词维度特征向量矩阵 W 、句维度特征向量矩阵 S 、篇维度特征向量矩阵 A 。

3.2 Multi_CNNSA层

本文的Multi_CNNSA层一共有三个CNN通道和一个Text_CNN通道，分别用于计算每个特征的高阶抽象信息，且每个通道后接一个Attention层，用于计算每个特征对假新闻检测的影响力度。其中Text_CNN通道接收输入层的语义信息数据 X ，输出结果记为 X' ，另外三个CNN通道分别接收来自输入层的三个维度的行文风格特征数据 W 、 S 、 A ，输出结果记为 W' 、 S' 、 A' 。然后将每个结果作为Attention层的输入，得到语义特征结果 X'' 、词维度特征结果 W'' 、句维度特征结果 S'' 、篇维度特征结果 A'' 。最后通过特征拼接得到本层的输出结果 H ，即：

$$H = X'' \oplus W'' \oplus S'' \oplus A'' \quad (5)$$

本文中Attention的计算公式如下：

$$a = \tanh(U) \quad (6)$$

$$\alpha = \text{softmax}(W^T a) \quad (7)$$

$$V^* = U \alpha^T \quad (8)$$

$$V = \tanh(V^*) \quad (9)$$

其中， $U = u_1, u_2, \dots, u_L$ 是多通道CNN的输出结果， L 是通道数， W 是权值矩阵， T 为转置， V 是Attention层的输出结果。

3.3 Softmax层

本层的目的是通过Softmax函数将上层结果进行归一化，最终实现对新闻的二分类，即将上层的输出结果 H 作为本层的输入，通过Softmax函数进行归一化操作。具体公式如下：

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{c=1}^C e^{z_i}} \quad (10)$$

4 实验与分析(Experiment and analysis)

4.1 数据集

本次实验聚焦社交媒体的假新闻检测，使用的是2018年KDD论文发布的weibo假新闻数据集^[23]，该数据集来源于微博，并由人工进行标注，因此其数据的真实性有效。本文将其整理为短文本假新闻二分类数据集。其中假新闻的标签为“1”，真新闻的标签为“0”，详细信息如表2所示。

表2 weibo假新闻数据集详情

Tab.2 Dataset details of weibo fake news

数据集	真新闻	假新闻	合计
训练集	3,823	3,706	7,529
验证集	478	463	941
测试集	478	463	941

4.2 评价指标

在假新闻检测的任务中，模型需要尽可能多地检测出假新闻，避免将真新闻误判为假新闻，所以为保证实验的公平性和可对比性，将综合考虑检测结果的精确率 P 、召回率 R 以及 $F1$ 值作为模型的评估指标。为使每个类别都得到公平对待，使用宏平均计算最终结果。其混淆矩阵如表3所示，其中TP表示真阳性，即实际为假预测为假；TN表示真阴性，即实际为真预测为真；FP表示假阳性，即实际为真预测为假；FN表示假阴性，即实际为假预测为真。

表3 假新闻检测混淆矩阵

Tab.3 The confusion matrix of fake news detects

真实结果	预测结果	
	假新闻	真新闻
假新闻	TP	FN
真新闻	FP	TN

指标计算公式如下：

$$P = \frac{TP}{TP + FP} \quad (11)$$

$$R = \frac{TP}{TP + FN} \quad (12)$$

$$F1 = 2P \frac{R}{P + R} \quad (13)$$

4.3 参数设置

数据预处理时，使用Word2vec方法进行词向量处理，向量维度为100。其最大文本长度设置为500，长度不足时，在文本后面补零表示。初始学习率为1，使用Adam函数进行优化。批次数为50，迭代数为100。使用的数据集随机按8:1:1比例分为训练集、验证集和测试集。

4.4 实验结果与分析

在本部分中，本文通过四个对比实验来验证我们提出的多维度行文风格假新闻检测方法的可行性和高效性。实验结果如表4所示。

表4 各模型性能指标结果

Tab.4 Performance index results of each model

模型	$P/\%$	$R/\%$	$F1/\%$
Naïve Bayes	56.57	86.79	68.48
Text_CNN	76.34	82.76	79.42
EANN	84.76	81.25	82.93
SCNN	72.41	73.87	73.13
Multi_CNNSA	87.08	86.78	86.95

其中，朴素贝叶斯(Naïve Bayes)方法是使用最为广泛的分类模型之一。该方法在贝叶斯算法的基础上进行简化，即假定给定目标值时属性之间相互条件独立。也就是说没有哪个属性变量对于决策结果来说占有较大的比重，也没有哪个属性变量对于决策结果占有较小的比重。虽然这个简化方式在一定程度上降低了贝叶斯分类算法的分类效果，但是在实际的应用场景中，极大地简化了贝叶斯方法的复杂性。因此本

文将其作为模型的对比实验之一。

CNN最早用于图像训练,如今其变体Text_CNN已应用到文本挖掘任务。除预训练文本之外,Text_CNN可以通过不同大小的卷积核并行提取文本的高阶抽象信息,同时通过最大池化突出关键词信息,最终实现分类任务。而且其训练过程参数数目少,计算量少,训练速度快。因此本文将其作为对比实验之一。

EANN模型是WANG等人^[25]在Text_CNN提取文本语义信息的基础上,加入文本的图片信息,从而构建多模态假新闻检测模型。它证明了假新闻检测任务中,除文本语义信息外,其他辅助特征也可以加入模型,而且可以显著提高假新闻检测的效果。作者使用的数据集与本文与我们使用的是同一个weibo数据集,因此我们直接引用其结果,而没有复现该方法。

为进一步证明本文提出的方法对假新闻检测的有效性,本文提出一个SCNN模型,该模型只使用多维度行文风格特征对假新闻进行检测,即去掉本文提出模型的文本语义特征提取部分。为保证实验的公平性,该模型的参数设置、特征提取和表示方法不变。

从表4可以看出,本文提出的Multi_CNNSA模型优于另外四个模型,其F1值达到了86.95%,比另外四个模型都高,这说明本文提出的方法和模型可以更好地表达真假新闻文本行文风格之间的差异性。其中Text_CNN模型优于Naïve Bayes模型,其精确率和F1值都比对方高。EANN是基于Text_CNN模型加入辅助特征提出的模型,其结果比单独的Text_CNN效果好。这证明了假新闻检测任务中,除文本的语义特征外,其他辅助特征的重要性。而本文提出的SCNN模型的表现虽没有Text_CNN表现好,但是其结果却高于Naïve Bayes,这从侧面证明了辅助特征的重要性。

5 结论(Conclusion)

本文研究真假新闻行文风格之间的差异性,并将这些差异性特征根据词、句和篇这三个维度对其进行归类。针对真假新闻的行文风格特征提出一种基于多维度行文风格的假新闻检测方法,然后使用多通道卷积神经网络对各维度的行文风格特征进行融合。实验结果表明,该方法能够有效提取真假新闻行文风格特征之间的差异性,并利用这些潜在的差异性对假新闻进行检测。

在文本的句维度行文风格特征中,句子的结构信息在真假新闻文本中也存在一定的差异性,那么如何利用句子的结构信息对假新闻进行检测将是在未来的工作中会着重思考的问题。

参考文献(References)

- [1] 吴广智,郭斌,丁亚三,等.假消息认知机理研究综述[J].计算机科学,2021,48(6):306-314.
- [2] 秦宇.探访马其顿假新闻工厂[N].南方都市报,2017-02-26(RB13).
- [3] 李晨璐.微博假新闻现象的探析与应对[J].传播与版权,2018(06):121-122,125.
- [4] LIU X, NOURBAKSH A, LI Q, et al. Real-time rumor

- debunking on twitter[C]// BAILEY J. Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. New York, USA: Association for Computing Machinery, 2015:1867-1870.
- [5] LIANG G, HE W, XU C, et al. Rumor identification in microblogging systems based on users' behavior[J]. IEEE Transactions on Computational Social Systems, 2015, 2(3): 99-108.
- [6] MA J, GAO W, WONG K F. Detect rumors in microblog posts using propagation structure via kernel learning[C]// NG H T. Research Collection School of Computing and Information System. Vancouver, Canada: Association for Computational Linguistics, 2017:708-717.
- [7] 毛二松,陈刚,刘欣,等.基于深层特征和集成分类器的微博谣言检测研究[J].计算机应用研究,2016,33(11):3369-3373.
- [8] 杨文太,梁刚,谢凯,等.基于突发话题和领域专家的微博谣言检测方法[J].计算机应用,2017,37(10):2799-2805.
- [9] 武庆圆,何凌南.基于多标签双词主题模型的短文本谣言分析研究[J].情报杂志,2017,36(3):92-97.
- [10] O'BRIEN N, LATESSA S, EVANGELOPOULOS G, et al. The language of fake news: Opening the black-box of deep learning based detectors[J/OL]. (2018-11-01) [2021-07-12]. <http://34.201.211.163/handle/1721.1/120056>.
- [11] UMER M, IMTIAZ Z, ULLAH S, et al. Fake news stance detection using deep learning architecture(cnn-lstm)[J]. IEEE Access, 2020, 8:156695-156706.
- [12] YANG Z, DAI Z, YANG Y, et al. Xlnet: Generalized autoregressive pretraining for language understanding[J/OL]. (2019-06-19) [2021-07-13]. <https://arxiv.org/abs/1906.08237>.
- [13] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[DB/OL]. (2018-10-11) [2021-07-15]. <https://arxiv.org/abs/1810.04805>.
- [14] PIERRI F, CERI S. False news on social media: A data-driven survey[J]. ACM Sigmod Record, 2019, 48(2):18-27.
- [15] PAN J Z, PAVLOVA S, LI C, et al. Content based fake news detection using knowledge graphs[C]// Center for Biomedical Informatics Research Stanford University. International Semantic Web Conference. Cham: Springer, 2018:669-683.
- [16] PRZYBYLA P. Capturing the style of fake news[C]// CONITZER V. Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2020, 34(01):490-497.
- [17] WANG Y, YANG W, MA F, et al. Weak supervision for fake news detection via reinforcement learning[C]// CONITZER V. Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2020, 34(01):516-523.

(下转第16页)