

· 舆情研究 ·

基于混合神经网络的突发公共卫生事件微博谣言识别研究^{*}

冯兰萍^{1 3} 董陈超² 徐绪堪^{1 3}

(1. 河海大学商学院 常州 213022;

2. 东南大学 经济管理学院 南京 211189;

3. 常州工业大数据挖掘与知识管理重点实验室 常州 213022)

摘要 [研究目的] 提高突发公共卫生事件中微博谣言识别效果, 分析微博特征对谣言识别的影响, 为公众识别谣言、政府及相关部门干预治理提供参考。[研究方法] 在用户基本特征和内容特征、传播特征的基础上引入用户历史特征, 构建基于混合神经网络的多特征融合谣言识别模型, 该模型采用 BiLSTM+CNN 抽取代表性深度语义特征, 与其他特征融合后, 使用 DNN 网络进行谣言识别。[研究结论] 实验结果表明: 该模型的准确率、查准率、查全率、F1 值均高于 94%, 准确率、查全率、F1 值优于其他谣言识别模型; 语义特征是谣言识别的关键特征, 在此基础上融合单一特征能够大幅提高谣言识别效果, 提高程度由高到低, 依次为统计特征、传播特征、用户基本特征、用户历史特征和情感特征。递进融合特征越多, 模型平均性能、泛化能力越强, 谣言识别效果越好。

关键词: 公共卫生事件; 微博谣言; 谣言识别; 多特征融合; 混合神经网络

中图分类号: TP393

文献标识码: A

文章编号: 1002-1965(2022)12-0081-08

引用格式: 冯兰萍, 董陈超, 徐绪堪. 基于混合神经网络的突发公共卫生事件微博谣言识别研究[J]. 情报杂志, 2022, 41(12): 81-88.

DOI: 10.3969/j.issn.1002-1965.2022.12.013

Research on Weibo Rumor Identification in Public Health Emergencies Based on Hybrid Neural Network

Feng Lanping^{1 3} Dong Chenchao² Xu Xukan^{1 3}

(1. Business School, Hohai University, Changzhou 213022;

2. School of Economics and Management, Southeast University, Nanjing 211189;

3. Changzhou Key Laboratory of Industrial Big Data Mining and Knowledge Management, Changzhou 213022)

Abstract [Research purpose] To improve the identification effectiveness of Weibo rumor in public health emergencies, and to analyze the influence of Weibo features on the identification of rumor, so as to provide support to the public to identify rumors and the government and relevant departments to intervene in the governance of rumors. [Research method] Based on the user basic features, the content features, and the spreading features, user historical features are introduced, a multi-feature fusion rumor identification model is constructed based on hybrid neural network. This model use BiLSTM+CNN to extract the representative depth semantic features and then merge with other features, use DNN network for rumor identification. [Research conclusion] The experimental results show that the accuracy, precision, recall and F1 value of the model are all higher than 94%, the accuracy, recall and F1 value are obviously superior to other rumor identification models. The fusion of semantic features and other features can greatly improve the effectiveness of rumor identification. From high to low in the improvement degree of rumor identification, they are statistical features, spreading features, user basic features, user

收稿日期: 2022-07-30

修回日期: 2022-09-22

基金项目: 国家社会科学重大基金项目“总体国家安全观下重大突发事件的智能决策情报体系研究”(编号: 20&ZD125) 子课题“总体国家安全观下重大突发事件智能决策情报体系的基本理论与方法研究”研究成果之一。

作者简介: 冯兰萍, 女, 1970 年生, 博士, 副教授, 研究生导师, 研究方向: 信息管理与数据挖掘; 董陈超, 男, 1999 年生, 硕士研究生, 研究方向: 数据挖掘; 徐绪堪, 男, 1976 年生, 博士, 教授, 博士生导师, 研究方向: 知识组织、情报分析。

historical features, and emotional features in turn. The more incremental features, the better the average performance and generalization ability of the model, and the better the rumor identification.

Key words: public health emergencies; Weibo rumor; rumor identification; multi-feature fusion; hybrid neural network

社交媒体的谣言识别现已得到学术界广泛深入的研究,主要采用传统的机器学习方法和深度学习方法,在微博内容特征的基础上融入更多的辅助特征提升实时谣言识别效果^[1]。微博谣言特征主要包括用户特征、内容特征和传播特征。突发公共卫生事件微博谣言内容通常具有一些固定的句式^[2],同时存在较多与疫情相关的语词,且分布较为分散;用户历史微博情感对谣言识别有辅助作用^[1,3],且用户特征和传播特征并不随着领域的不同而产生差别^[4],因而用户历史特征对谣言识别亦有一定的辅助作用,而多特征融合中引入用户历史特征的研究相对不多。因此如何充分抽取谣言的关键特征,构建多特征融合的谣言识别模型,提高谣言识别效果;探讨各特征对谣言识别的影响,为政府及相关部门谣言治理、用户谣言识别提供参考,是本文研究的问题。

相对于传统的机器学习,深度学习自动构建特征可进一步提升谣言检测的准确性与稳定性^[4]。因此,本文以微博数据为研究对象,提出一种基于混合神经网络的多特征融合谣言识别模型。该模型引入用户历史特征,充分利用 BiLSTM 和 CNN 的优势,采用 BiLSTM+CNN 抽取微博代表性深度语义特征,与其它多特征融合后,使用全连接 DNN 网络进行谣言识别,在此基础上分析各个特征对谣言识别的影响,为谣言自动识别研究提供新的思路。

1 国内外相关研究

社交媒体谣言识别主要包括模型和方法选取、特征抽取及其基础上的谣言判定^[5]。

a.在模型和方法方面。传统的机器学习方法有支持向量机^[5-6]、随机森林^[7]、决策树^[8]等,由于分类算法对谣言的识别效果有很大影响,因此主要集中于分类算法的选择和改进、特征的有效识别^[9]。深度学习因其强大的学习能力,不需要人工构建特征在谣言识别领域得到了广泛的关注。Ma 等^[10]通过构建循环神经网络(RNN)模型首次对微博谣言事件进行检测,提高了谣言事件自动检测的效率。尹鹏博等^[3]采用卷积神经网络(CNN)和 LSTM 分别提取用户特征和微博文本特征,提出基于 C-LSTM 的检测模型进行早期谣言检测。赵月华等^[11]针对小样本领域提出基于 BERT 虚假信息识别模型。石锴文等^[2]构建 BiLSTM+DNN 模型对多特征向量进行谣言判别。刘勘等^[4]将深度迁移网络应用于跨领域谣言识别。Chen 等^[12]提出注意

力机制与 RNN 相结合的谣言检测模型,能更准确地检测早期谣言。潘德宇等^[13]在 CNN 网络引入了注意力机制,获得较好地识别效果。现有研究多采用单一的 CNN、RNN 及衍生网络进行谣言识别,或者利用注意力机制对已提取特征进一步提取关键信息,以提高谣言识别的精度。但是未能充分利用 RNN 和 CNN 网络提取文本语义的优势,因而谣言识别效果仍存在一定的提升空间。

b.在特征抽取挖掘方面。通常在社交媒体谣言检测中用到的特征主要有用户特征、内容特征和传播特征^[9]。例如,Castillo 等^[8]从字符串长度、词语个数、符号(问号、感叹号、微笑和皱眉等)、标签、链接等维度对文本特征细分。贺刚等^[14]认为浅层的文本特征难以反映微博文本内容的可信度,在 Yang 等^[15]提出的微博文本内容、用户属性信息和传播深度 3 个特征的基础上,引入了文本的符号、链接、关键词分布以及时间差四类新特征。首欢容等^[16]构建基于情感分析的领域谣言自动识别模型,并运用到“食品养生”“医学健康”两个领域实现领域谣言自动识别。石锴文等^[2]运用 TF-IDF 增强 Word2Vec 词向量语义特征,利用 LDA 模型对文本进行主题分析,从多个维度融合文本特征。黄学坚等^[1]在谣言识别模型中除了用户基本特征和内容统计特征,从用户历史微博和微博评论中挖掘用户深层特征。现有研究根据不同的需求,从不同的角度对微博特征进行抽取挖掘,主要集中于文本内容特征与其融合其他特征研究,在谣言识别中融合用户历史特征的研究不多。

2 多特征融合的微博谣言识别模型

2.1 模型框架

谣言内容、发布用户以及传播是构成识别谣言和非谣言的关键要素^[9],不同特征之间的融合将对谣言识别起到较大作用。突发公共卫生事件具有公众参与度高、负面倾向性强、民众恐慌加剧和极端用语多等独有的特征^[2],且用户特征和传播特征对谣言识别的影响与突发事件所属领域无关,用户历史行为对谣言的识别具有一定的辅助作用。由于 BiLSTM 的双向机制保证了每一个词在充分考虑上下文的条件下获得语义,实现微博文本特征的深度抽取^[17]。CNN 可提取语义表示和捕获平面结构中的显著特征^[18],具有捕获局部上下文的优势^[19]。因此本文采用 BiLSTM 获取微博内容的上

下文语义, CNN 对提取特征使用不同的卷积核捕获多粒度的局部上下文语义信息,从而获得谣言内容语词间上下文语义结构,以最大池化方式抽取代表性语义特征。BiLSTM+CNN 输出语义特征向量为自动构建的高维向量,解释性不强,全连接 DNN 的多隐藏层可以抽象地表示特征,具有传统的算法所不具备的灵活性,对于大部分非线性函数的拟合能力都较好,只要通过增大全连接神经网络的深度,就可以对函数进行精确地拟合^[20]。因而在语义特征与其他特征拼接融合后,可采用全连接型的 DNN 网络进行谣言识别。综上,结合相关研究成果,在用户基本特征和内容特征、传播特征的基础上引入用户历史特征,本文提出一种突发公共卫生事件下的多特征融合谣言识别模型框架(如图 1 所示),以实现微博谣言自动有效识别。

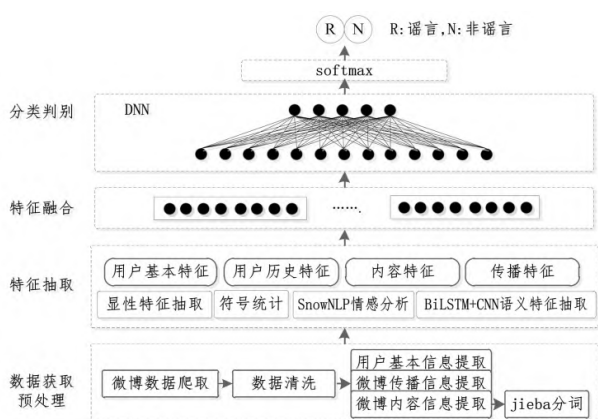


图 1 突发公共卫生事件下的多特征融合微博谣言识别模型框架

该模型包括数据获取与预处理、特征抽取、特征融合、分类判别等。具体为首先采用相关工具或编程爬取微博相关数据,并对其进行清洗,根据微博数据标签提取用户基本信息、传播信息、微博内容信息,对微博内容文本进行符号统计、Jieba 库分词和 SnowNLP 情感分析^[21]、BiLSTM+CNN 模型抽取微博内容特征,在此基础上进一步处理获得用户历史特征。接着将所提取特征拼接后,输入 DNN 全连接层判别谣言的类别。

2.2 特征构建及抽取

2.2.1 特征构建

根据上文分析,本文谣言微博特征采用用户基本特征、用户历史特征、内容特征和传播特征 4 个维度表征。内容特征现有的研究主要包括:统计特征(T1)、情感特征(T2)、语义特征(T3),能够很好地区分谣言和非谣言,因此本文将从这 3 个维度对内容特征进一步细分。谣言传播者和普通用户的基本特征上存在较为明显的差异,例如谣言用户具有地域差异,为了增加发布谣言的曝光度会在较短的时间内关注大量用户,以获得较多粉丝,话题型用户名具有更高的可信度

等^[9]。谣言用户发博时间通常选择普通用户刷微博的高峰期,以获得较高的关注度,因此传播特征除了评论数、转发数、点赞数外,发博时间亦是重要的传播特征。另外虽然用户历史发博的领域不定,但用户特征和传播特征与领域无关,因此用户历史特征可用用户历史传播特征(H1)、历史发博特征(H2)、历史情感特征(H3)表征。结合清博指数 BCI 和相关研究,用户基本特征(U)、内容特征(T)、传播特征(S)、用户历史特征(H)具体如表 1—表 3 所示。

表 1 微博内容特征、传播特征

特征类别	特征属性	特征值
统计特征(T1)	URL 链接	URL 链接个数
	@	@ 用户个数
	表情	表情个数
	?	? 个数
	!	! 个数
情感特征(T2)	话题	话题个数
	符号总数	除?、! 外的符号个数
	多媒体	0: 不包含图片或视频, 1: 包含图片或视频
语义特征(T3)	情感特征	文本内容情感得分, 区间为[0, 1], 得分越接近 1 代表情感越积极, 得分越接近 0 代表越消极。
	语义特征	语义特征抽取模型抽取的语义向量
传播特征(S)	转发数	转发总数
	评论数	评论总数
	点赞数	点赞总数
	微博发布时间	发布时间

表 2 用户基本特征

特征属性	特征值
用户性别	1: 女性 2: 男性 3: 空值
用户所在地	用户人工选择所在地信息, 非 IP 属地 1: 海外 2: 北京 3: 上海 4: 广东……
关注数	用户关注的账号数
粉丝数	用户被关注的账号数
微博总数	用户发布的微博总数
用户昵称类型	1: 原始用户名 2: 网络用户名 3: 话题型用户名 4: 职业型用户名 5: 官方用户名

表 3 用户历史特征

特征类别	特征属性	特征值
历史传播特征(H1)	历史微博点赞数	用户历史微博平均获得的点赞数
	历史微博评论数	用户历史微博平均被评论次数
	历史微博转发数	用户历史微博平均被转发次数
	历史微博发布时间	用户历史微博发布的平均时间
历史发博特征(H2)	历史微博来源	0: 来源缺失, 1: 手机端 2: 客户端
	历史微博原创性	0: 非原创, 1: 原创
历史情感特征(H3)	历史微博情感	用户历史微博平均情感得分

2.2.2 特征抽取

用户基本特征、传播特征根据微博数据标签提取,统计特征采用统计学方法、python 编程或 EXCEL 统计文本中包含的符号数完成,情感特征值直接采用 SnowNLP 库计算内容情感得分。符号统计和 SnowNLP 情感计算比较成熟,微博文本语义特征反映微博用

户观点、意见、情感的词语表达,具有不确定性,是谣言识别的关键部分,本文使用 BiLSTM+CNN 抽取语义特征。其结构图如图 2 所示。

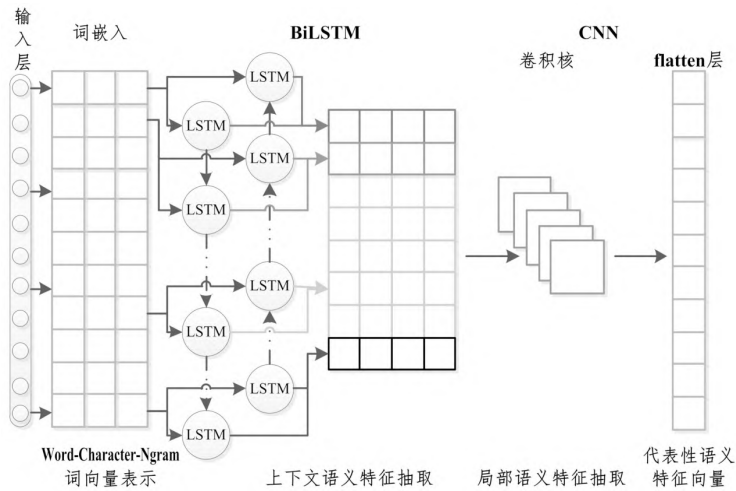


图 2 BiLSTM+CNN 语义特征抽取模型

该模型包括三层:词嵌入、BiLSTM 层、CNN 层。

a.词嵌入。将微博内容文本分词后的词转化为词向量以作为 BiLSTM 层的输入。北师大和人大开源的已经训练好的词向量词典 Word + Character + Ngram^[22] 相比其他全领域的词向量词典更适合微博谣言识别的任务^[1]。因此本文选取 Word + Character + Ngram (300d) 将文本每个词表示为 300 维的词向量。

b.BiLSTM 层。使用 BiLSTM 编码微博单词的语义及其前后信息。通过顶部和底部网络分别接收正向和反向的词向量,学习正向和反向两个维度的上下文语义,将最后隐藏向量融合后输出上下文语义特征向量。设 t 时刻词向量的正向 LSTM 输出最后隐藏特征向量为 \vec{h}_t ,反向 LSTM 输出最后隐藏特征向量为 \overleftarrow{h}_t ,BiLSTM 层输出上下文语义特征向量 h_t 如公式 (1) 所示:

$$h_t = [\vec{h}_t \parallel \overleftarrow{h}_t] \quad (1)$$

所有词向量的语义特征向量构成文本特征矩阵 A 。

c.CNN 层。针对不同语词在微博内容中分布不同,对文本特征矩阵 A 进行上下文结构局部特征提取,CNN 从上下文语义中捕获结构信息,提取代表性特征,去除冗余信息、消除歧义。本文通过设置不同宽度的卷积核对文本特征矩阵 A 进行卷积运算、最大池化,选取宽度 16 对微博中句子进行编码,最大池化后再选取宽度 4 对句子中四元组作进一步的语义编码,做最大池化。最后将池化后特征向量 flatten 成一维语义特征向量,形成代表性语义特征向量 $T3$ 输出。

2.3 特征融合

特征融合是将不同角度的特征结合,为分类提供

更多可依据的信息^[4]。将用户基本特征、用户历史特征、传播特征、统计特征、情感特征等相关数据归一化后,与 BiLSTM+CNN 模型提取的 256 维代表性语义特征向量 $T3$ 拼接成一个多元特征向量 TU ,

作为 DNN 网络的输入。特征融合过程如公式 (2) — (4) 所示。

$$T = T1 \oplus T2 \oplus T3 \quad (2)$$

$$H = H1 \oplus H2 \oplus H3 \quad (3)$$

$$TU = U \oplus T \oplus S \oplus H \quad (4)$$

式中 \oplus 代表级联操作,即对各特征向量进行拼接。

2.4 谣言识别

谣言识别采用多层 DNN 网络进行分类训练,DNN 的输入为多元特征向量 TU 。分类模型训练过程中,先对类别标签进行 One - Hot 编码,在隐藏层使用

tanh 作为激活函数,并使用 categorical_crossentropy 作为模型的损失函数对模型参数进行优化。最后通过 softmax 层进行分类,输出微博文本为谣言和非谣言的概率,二者概率之和为 1,其具体计算过程如公式 (5)、(6) 所示。

$$\mathcal{X}_k(x_i) = W_k^T \cdot x_i \quad (5)$$

$$P(y_i = k | x_i) = \frac{\exp(\mathcal{X}_k(x_i))}{\sum_{i=1}^K \exp(\mathcal{X}_k(x_i))} \quad (6)$$

式中 W_k 是第 k 个类别的权重系数矩阵, x_i 表示第 i 个文本的多元特征向量, y_i 表示第 i 个文本的类别(谣言或非谣言), $\mathcal{X}_k(x_i)$ 为第 i 个文本在第 k 个类别上的计算结果值, K 为类别总数。

3 实验分析

为了验证本文模型的有效性,首先对本文模型中语义特征采用不同的方法处理,对比实验结果分析本文模型语义特征在谣言识别中的作用及 BiLSTM + CNN 谣言语义特征学习能力;接着与现有谣言模型实验结果对比分析本文模型谣言识别效果;最后在语义特征的基础上递进融合特征,分析各特征对谣言识别效果的影响。实验数据集训练集比例 70%,测试集比例 30%。使用准确率 (Accuracy)、查准率 (Precision)、查全率 (Recall) 和 F1 值作为模型效果评价指标。

3.1 数据来源

本文实验数据为 2020 年 1 月 15 日—2022 年 2 月 7 日间新冠疫情相关微博数据,谣言数据为微博社区管理中心平台已确认的新冠疫情谣言,非谣言数据根据微博话题#新冠疫情#,采用 Python 编程爬取,用户历史数据通过循环滚动的方式,获取用户最近 50 条微

博历史特征。最后获取数据为谣言用户基本信息 713 个,非谣言用户基本信息 1 177 个,谣言微博数 770 条,非谣言微博数 1 200 条(随机选取),谣言用户历史微博数 27 939 条,非谣言用户历史微博数 77 736 条。

3.2 数据清洗与预处理

首先对爬取到的页面进行初步清洗,根据标签获取本文所需数据,获得用户基本特征、传播特征和微博内容信息,保存为 CSV 文件。接着主要作如下处理:

a.缺失值处理。用户自己未填写相关特征信息(如性别、所在地等)或者由于微博发布时间过早、用户主动删除了部分已发布的微博,导致部分信息缺失时,本文将缺失值全部填充为-1,以区分缺失值和获取到的真实值。例如某条微博的点赞数为-1表示点赞数缺失。

b.文本内容特征处理。首先对文本中符号进行统计,获得统计特征数据,采用 python 编程通过正则化表达式去除文本中的表情、标点符号、多媒体等信息,Jieba 库对微博文本进行分词,SnowNLP 库计算微博文本的情感得分。

c.归一化处理。采用 min-max 对统计特征、传播特征等相关数据进行归一化处理,将值转化到区间[0,1]。

d.微博谣言时间分布统计分析。以 1 小时为时间段统计每个时间段内谣言发布量,每个时间段内谣言发布量如图 3 所示。

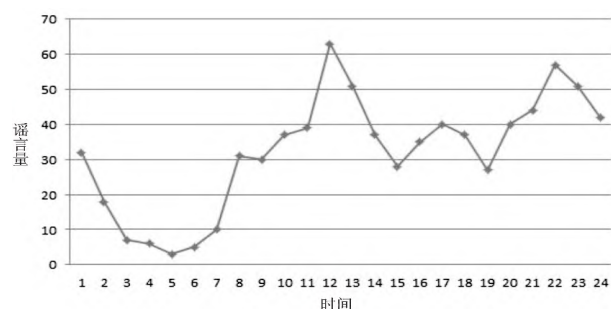


图 3 微博谣言时间分布统计

由图 3 可知,24 小时内谣言量呈波动状态,谣言量最少的时间段是 4-5 点,5 条;最高峰是 11-12 点,63 条;每小时平均约 32 条。谣言量较多的时间段为 10-13 点、20-24 点,2-7 点极少。

3.3 主要参数设置

本文实验环境为 Python 3.8、tensorflow 2.1.0、keras 2.3.1 等。为了保证公平性和合理性,本文采用网络规模和尺寸保持一致,主要参数如表 4 所示。

3.4 不同语义特征抽取模型实验结果分析

为研究本文模型的可行性和有效性,在模型结构与参数不变情况下,与以下 3 种类别模型对比谣言识别效果,分析本文模型中语义特征的作用及 BiLSTM +

CNN 语义特征抽取的性能。具体类别为:a.类别 1,不考虑语义特征,其他特征不变;b.类别 2,仅考虑语义特征,采用 BiLSTM+CNN 抽取语义特征;c.类别 3,采用常用深度学习模型及组合抽取语义特征,其他特征不变;谣言识别效果评价各项指标值是按照训练集和测试集的比例进行加权后计算得到的,实验结果见表 5。

表 4 主要参数

类别	参数名称	参数值
词嵌入	最大分词数	150
	词嵌入维度	300
	词嵌入词典最大包含词向量个数	400000
	LSTM 输入序列数	150
BiLSTM	LSTM 单元神经元个数	128
	Dropout	0.1
	Recurrent_dropout	0.1
	return_sequences	True
	输出层的尺寸大小	2
	输出层激活函数	softmax
CNN 模型	CNN 输入层大小	[150, 256, 1]
	第 1 层卷积层卷积核个数、尺寸、步长	8, 16 * 16, 1 * 1
	第 2 层卷积层卷积核个数、尺寸、步长	16, 4 * 4, 1 * 1
	第 1、2 层卷积层激活函数	tanh
	第 1、2 层池化层	MaxPooling (最大池化)
	第 1 层池化层尺寸	16 * 16
	第 2 层池化层尺寸	4 * 4
	Dropout	0.1
	全连接层尺寸	256
	全连接层激活函数	tanh
训练参数	输出层的尺寸大小	2
	输出层激活函数	softmax
	损失函数	categorical_crossentropy
	优化器	adam
	epochs	10
	batch_size	32

表 5 不同语义特征抽取模型实验结果对比分析(单位: %)

模型类别	语义特征抽取模型	准确率	查准率	查全率	F1
类别 1 模型	未抽取语义特征	83.56	81.75	77.46	79.53
类别 2 模型	BiLSTM+CNN	91.68	92.20	87.21	89.81
	TextCNN	86.70	72.28	92.05	80.96
	TextRNN	92.33	95.15	86.55	90.63
类别 3 模型	BiLSTM	94.77	94.81	92.03	93.40
	LSTM+CNN	95.33	94.01	94.00	94.01
	BiLSTM-CNN	92.19	96.49	85.39	90.59
本文模型	BiLSTM+CNN	96.46	94.43	96.40	95.40

3.4.1 语义特征的作用分析

由表 5 可知,语义特征是微博谣言识别的关键要素,其他特征能够显著地提高谣言识别效果。类别 2 模型与类别 1 模型相比各评价指标大幅提高。类别 1 模型准确率为 83.56%、查准率 81.75%,查全率和 F1 值不到 80%;虽然融合了微博其他特征,但未考虑谣言文本的语义特征,对谣言识别能力相对较弱,因而微博语义特征是识别谣言的重要特征。本文模型与类别

2 模型相比,本文模型中语义特征融合了其他特征能够显著地提高谣言识别效果。

3.4.2 BiLSTM+CNN 语义特征抽取性能分析

由表 5 可知,BiLSTM+CNN 谣言语义特征识别能力明显高于其他常用深度学习模型及组合。与类别 3 模型相比,本文模型的查准率略低于 TextRNN、BiLSTM、BiLSTM-CNN 模型,但准确率、查全率、F1 值都远高于类别 3 模型,类别 3 模型(除 TextCNN 准确率外)与模型类型 1 相比各指标有显著提高,具有较好地谣言识别能力。

a. BiLSTM 与 TextCNN、TextRNN 相比,谣言深度语义特征学习能力更强。BiLSTM 模型准确率(94.81%)略低于 TextRNN(95.15%),但其准确率、查全率、F 值远高于两者。TextRNN 能够抽取深层语义信息,获得较高的准确率,略高于 BiLSTM,但明显降低了查全率。TextCNN 抽取局部信息,获得代表性语义特征,但明显降低了其查准率,反而低于模型类型 1;虽然获得较高的查全率,具有较强的泛化能力,相关指标均低于 BiLSTM。

b. 本文模型查准率略低于 BiLSTM 模型,但查全率、F1 值指标上均有较明显的提升。表明 CNN 网络对 BiLSTM 网络抽取的语义特征向量进行二次学习,学习到质量更高的具有代表性语义特征,具有较强的泛化能力。

c. BiLSTM-CNN 模型采用 BiLSTM、CNN 分别抽取微博语义特征并拼接,语义特征向量维度增加提高了语义特征的专指度,查准率高于本文模型,但准确率、查全率和 F1 值明显低于本文模型。

d. 本文模型相关指标均高于 LSTM+CNN。说明 BiLSTM 谣言深层语义特征能力学习优于 LSTM。

综上所述,本文模型采用 BiLSTM 抽取文本内容上下文深度语义特征,CNN 进一步捕获代表性上下文语义结构,查全率和综合性能更好,具有更强地泛化能力。

3.5 不同谣言识别模型实验结果分析

为了进一步验证文本模型的有效性,在本文数据集上对现有谣言识别模型 S-SVM^[6]、TextCNN+Attention^[14]、C-LSTM^[3]和 BiGRU+Attention^[1]进行复现,与本文模型谣言识别效果对比分析。相关评估指标值的计算与 3.4 一致,实验结果见表 6。

a. S-SVM 模型较为简单,相关指标均低于本文模型,这是由于 S-SVM 无法充分学习到谣言的文本特征,将不能确定为谣言的文本统一分类为非谣言文本,导致谣言文本的查准率较低,获得较好的查全率。

b. 除 S-SVM 模型外,本文模型查准率(94.43%)

低于现有谣言模型,查全率(96.40%)、准确率(96.46%)、F1 值(95.40%)均高于现有谣言模型,说明现有谣言模型提取语义特征相对较深,但本文模型的泛化能力较强、综合性能较高。BiGRU+Attention、TextCNN+Attention 模型对谣言文本的查准率较高,结合表 5 可知,这是由于 Attention 机制进一步识别了谣言的关键信息,提高了谣言语义特征深度,但明显地降低了查全率,远低于本文模型。

表 6 本文模型与其他模型效果对比分析(单位: %)

模型类别	模型名称	准确率	查准率	查全率	F1
现有模型	S-SVM	87.56	72.11	94.90	81.95
	TextCNN+Attention	94.77	95.19	91.72	93.42
	C-LSTM	95.99	97.21	92.76	94.92
	BiGRU+Attention	96.14	98.90	91.77	95.20
本文模型	BiLSTM+CNN	96.46	94.43	96.40	95.40

查准率和查全率是一对矛盾的度量。当查全率较高时,虽然降低了将非谣言识别为谣言的可能性,但降低了查准率,提高非谣言识别为谣言的可能性。突发公共卫生事件谣言滋生后,由于信息不对称、认知、资源、经济等的有限性,公众极易产生恐慌、焦虑情绪^[23],严重时甚至对社会安定和民生保障构成巨大威胁。公众、政府及相关机构期望应用平台尽可能识别网络中潜在的谣言,帮助自身快速地识别谣言,从而降低公众自身的损失(例如经济损失、恐慌、焦虑情绪等),政府及相关机构及时采取相应的措施以遏制谣言的传播和消除谣言的影响。而如果误判的话,政府、相关机构及应用平台可能会失去公众的信任等,因此在此背景下不能仅考虑查准率或查全率作为评价模型好坏的指标。F1 值表示模型的综合性能,兼容查准率和查全率;准确率兼顾谣言和非谣言的查准率,本文将此作为评价模型的指标。本文模型准确率、查全率、F1 值均明显高于现有模型,因此从查全率、F1 值和准确率来看,本文模型对谣言的识别效果更佳。

3.6 不同特征融合谣言识别实验结果分析

参考黄学坚等^[1]、石锴文等^[2]的研究,本文选取不同的特征组合进行特征消融实验,分析各个特征对谣言识别的影响。结合 3.4,语义特征是影响谣言识别的关键因素,在保持网络结构和参数不变、特征不同的情况下,在语义特征(T3)的基础上,对比分析递进融合特征后模型对谣言识别效果的影响。设极差表示各模型训练集和测试集准确率之差,反映出模型的平均性能和泛化能力。准确率差值、F1 值差值为仅语义特征模型与其他模型的准确率与 F1 值的差值。具体实验结果如表 7 所示。

表 7 不同特征融合的模型实验结果对比(单位: %)

融合特征	数据集	准确率	查准率	查全率	F1	准确率 差值	F1 差值	极差
T3	训练集	92.25	92.36	88.66	90.74	-	-	1.91
	测试集	90.34	91.82	83.82	87.64	-	-	
U+T3	训练集	94.13	91.09	94.00	92.52	1.88	1.78	0.91
	测试集	93.22	92.27	89.82	91.03	2.88	3.39	
T1+T3	训练集	94.49	92.18	93.89	93.03	2.24	2.29	1.10
	测试集	93.39	93.18	89.52	91.31	3.05	3.67	
T2+T3	训练集	94.20	91.64	93.68	92.65	1.95	1.91	1.32
	测试集	92.88	91.36	89.73	90.54	2.54	2.90	
S+T3	训练集	94.13	90.18	94.84	92.45	1.88	1.71	0.74
	测试集	93.39	90.91	91.32	91.12	3.05	3.48	
H+T3	训练集	94.06	90.73	94.15	92.41	1.79	1.67	1.01
	测试集	93.05	90.91	90.50	90.70	2.71	3.06	
T1+T2+T3	训练集	94.86	93.82	93.31	93.56	2.61	2.82	0.79
	测试集	94.07	92.27	91.86	92.06	3.73	4.42	
T1+T2+S+T3	训练集	95.51	92.55	96.04	94.26	3.26	3.52	0.59
	测试集	94.92	92.27	93.98	93.12	4.58	5.48	
T1+T2+S+U+T3	训练集	96.23	94.36	96.11	95.23	3.98	4.49	0.13
	测试集	96.10	95.00	94.57	94.78	5.76	7.14	
T1+T2+S+U+H+T3	训练集	96.52	94.18	97.00	95.57	4.27	4.83	0.25
	测试集	96.27	95.00	95.00	95.00	5.93	7.36	

3.6.1 不同特征融合对谣言识别效果的影响分析

a. 单个特征的融入模型能够不同程度地提高谣言识别效果。由表 7 可知, 根据准确率和 F 值, 单个特征影响程度由高到低, 依次为统计特征、传播特征、用户基本特征、用户历史特征和情感特征。其中统计特征和用户基本特征全面提升了谣言预测查准率和查全率, 传播特征、用户历史特征和情感特征提升了谣言预测的查全率, 而查准率略有降低。查准率贡献最大的是统计特征, 用户历史特征贡献其次, 最后是传播特征和情感特征。查全率贡献最大的是传播特征, 其次为用户基本特征, 最后是传播特征和用户历史特征。单个特征分别融合后, F1 值都明显的提升, 提高了模型的综合性能。

b. 内容特征仍然是谣言识别的主要特征。由表 7 可知, T3 融入 T1、T2 后, 即内容特征(T1+T2+T3) 明显提高谣言识别预测的查全率(92.27%)、查准率(91.86%)、F1 值(92.06%)。

c. 基于内容特征, 随着融入特征增加, 模型谣言识别能力越强。由表 7 可知, 融入 S, 模型查准率持平, 查全率持续提升; 递进融入 U 后, 模型查准率、查全率显著提高; 递进融入 H 后, 模型查准率持平, 查全率提升, 所有指标达到最高, 准确率和 F1 值分别达到了 96.27% 和 95%。

因此, 公众或政府部门相关机构及应用平台等可

根据微博特征对谣言识别的影响程度采取相关的措施, 例如关注和识别微博文本特征, 尤其是在谣言高发阶段 10—13 点和 20—24 点(见图 3), 重点关注微博语义特征(例如语词表达)、情感表达(符号、情感词的使用等), 及其引起的传播效果; 同时建立用户信息数据库保存用户基础信息和用户历史信息, 以辅助谣言识别。在谣言爆发初期, 其他信息较少的情况下, 可充分利用用户的历史信息对谣言进行识别。

3.6.2 特征融合对谣言识别模型性能影响分析

随着融合特征递进增加, 模型平均性能、泛化能力越强, 预测效果更好。由表 7 差值可知, 仅使用 T3 模型在训练集和测试集上的准确率之差最大, 达到 1.91%。而融入了任一单特征后, 模型极差明显下降, 表明模型的泛化能力得到明显提升。而递进融入 2 个以上特征后, 极差大幅下降, 泛化能力大幅提升, 预测能力更强。

4 结 语

在公共卫生事件发展过程中准确、高效的谣言自动识别模型能够辅助公众、政府及相关部门更早地发现谣言并阻断谣言传播。本文在现有研究的基础上, 从用户基本特征、用户历史特征、内容特征、传播特征 4 个维度, 构建基于混合神经网络的多特征融合的公共突发事件谣言识别模型。该模型采用 BiLSTM + CNN 网络抽取代表性的语义深度特征、DNN 网络对

谣言进行分类。实验结果表明: a. 本文模型取得了较好地识别效果, 综合性能优于常用的深度学习模型和现有的谣言识别模型; b. 文本语义特征是影响谣言识别的关键要素, 统计特征、传播特征、用户基本特征和历史特征都能够不同程度地提升谣言识别效果。

本文研究还存在一些局限性, 例如未考虑谣言评论内容特征的影响、本文提出模型仅使用新冠疫情数据验证等, 下一步将在此基础上做进一步探讨, 以更好为相关部门的决策参考提供依据。

参考文献

- [1] 黄学坚, 王根生, 罗远胜, 等. 融合多元用户特征和内容特征的微博谣言实时检测模型[J/OL]. 小型微型计算机系统: 1-12[2022-07-26]. <http://kns.cnki.net/kcms/detail/21.1106.TP.20211021.1044.002.html>.
- [2] 石锴文, 刘 勤. 突发公共卫生事件中微博谣言的识别[J]. 图书情报工作, 2021, 65(13): 87-95.
- [3] 尹鹏博, 潘伟民, 彭 成, 等. 基于用户特征分析的微博谣言早期检测研究[J]. 情报杂志, 2020, 39(7): 81-86.
- [4] 刘 勤, 杜好宸. 基于深度迁移网络的 Twitter 谣言检测研究[J]. 数据分析与知识发现, 2019, 3(10): 47-55.
- [5] 张仰森, 彭媛媛, 段宇翔, 等. 基于评论异常度的新浪微博谣言识别方法[J]. 自动化学报, 2020, 46(8): 1689-1702.
- [6] 马 鸣, 刘 云, 刘地军, 等. 基于主题和预防模型的微博谣言检测[J]. 北京理工大学学报, 2020, 40(3): 310-315.
- [7] 曾子明, 王 婧. 基于 LDA 和随机森林的微博谣言识别研究——以 2016 年雾霾谣言为例[J]. 情报学报, 2019, 38(1): 89-96.
- [8] Castillo C, Mendoza M, Poblete B. Information credibility on twitter[C]//Proceedings of International Conference on World Wide Web, Hyderabad, 2011: 675-684.
- [9] 刘雅辉, 靳小龙, 沈华伟, 等. 社交媒体中的谣言识别研究综述[J]. 计算机学报, 2018, 41(7): 1536-1558.
- [10] Ma J, Gao W, Mitra P, et al. Detecting rumors from microblogs with recurrent neural networks [C]//Proceedings of IJCAI, 2016.
- [11] 赵月华, 朱思成, 苏苏宁. 面向网络虚假医疗信息的识别模型构建研究——一种基于预训练的 BERT 模型[J]. 情报科学, 2021, 39(12): 165-173.
- [12] Chen T, Li X, Ying H Z, et al. Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection [C]// Proceedings of the 2018 Pacific Asia Conference on Knowledge Discovery and Data Mining 2018: 40-52.
- [13] 潘德宇, 宋玉蓉, 宋 波. 一种新的考虑注意力机制的微博谣言检测模型[J]. 小型微型计算机系统, 2021, 42(2): 348-353.
- [14] 贺 刚, 吕学强, 李 卓, 等. 微博谣言识别研究[J]. 图书情报工作, 2013, 57(23): 114-120.
- [15] Yang F, Liu Y, Yu X, et al. Automatic detection of rumor on sina weibo [C]//Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics, Beijing: ACM, 2012: 1-7.
- [16] 首欢容, 邓淑卿, 徐 健. 基于情感分析的网络谣言识别方法[J]. 数据分析与知识发现, 2017, 1(7): 44-51.
- [17] Chen T, Xu R, He Y, et al. Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN [J]. Expert Systems with Applications, 2017, 72: 221-230.
- [18] Liu Y, Wei F R, Li S J, et al. A dependency-based neural network for relation classification[J]. ArXiv: 1507.04646, 2015.
- [19] 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, Beijing, 2015: 365-371.
- [20] 于雅洁, 刘贤达, 蒋启梅, 等. 融合 LSTM-DNN 的工业安全态势预测模型[J/OL]. 小型微型计算机系统: 1-8 [2022-09-22]. <https://kns.cnki.net/kcms/detail/21.1106.TP.20211213.1115.002.html>.
- [21] 冯兰萍, 严 雪, 程铁军. 重大突发事件微博舆情演化分析——以“天津大爆炸事故”为例[J]. 竞争情报, 2021, 17(3): 21-29.
- [22] Li S, Zhao Z, Hu R, et al. Analogical reasoning on chinese morphological and semantic relations [C]//56th Annual Meeting of the Association for Computational Linguistics, ACL, 2018: 138-143.
- [23] 范 敏, 周建新. 信息畸变与权力博弈: 重大疫情下网络谣言的生成与传播机制[J]. 新闻与传播评论, 2020, 73(4): 64-72.

(责编/校对: 王育英)

(上接第 37 页)

fornia Press, 1980.

- [24] Gamson W A, Modigliani A. Media discourse and public opinion on nuclear power: A constructionist approach [J]. American Journal of Sociology, 1989, 95(1): 1-37.
- [25] Entman R M. Framing: Toward clarification of a fractured paradigm [J]. Journal of Communication, 1993(43): 51-58.
- [26] 庞 琴, 赵滢竹. 回归以来英国智库对香港问题的分析与解读[J]. 当代港澳研究, 2019(2): 90-119.
- [27] 张虹倩. “民主国家峰会”与拜登政府的对华战略——基于布鲁金斯政策报告话语的框架分析[J]. 社会科学, 2021(7): 22-34.
- [28] 赵恒煜. 美国智库的“中国观”研究——以布鲁金斯学会中国问题研报的综合分析为例[J]. 情报杂志, 2020, 39(2): 35-43.
- [29] 赵恒煜. 欧洲智库的“中国观”研究——基于三大欧洲智库研究报告的综合比较分析[J]. 情报杂志, 2021, 40(1): 111-119.
- [30] Kuckartz U. 质性文本分析: 方法、实践与软件指南[M]. 朱志勇, 范晓慧, 译. 重庆: 重庆大学出版社, 2017.

(责编/校对: 王平军)