第49卷第1期
2021年1月

华南理工大学学报(自然科学版)
Journal of South China University of Technology
(Natural Science Edition)

Vol. 49 No. 1
January 2021

doi: 10.12141/j.issn.1000-565X.200489

重大突发疫情事件中的谣言识别

刘勘[1] 黄哲英[2]

(1. 中南财经政法大学 信息与安全工程学院，湖北 武汉 430073；2. 南开大学 商学院，天津 300071)

摘 要: 新冠疫情爆发以来，相关谣言时有传播，但传统的谣言识别模型却难以有效判别疫情谣言，因为相较于大量历史谣言数据，疫情谣言的数量还不足以训练出良好的分类器。因此，建立一个以少量谣言数据为基础的疫情谣言识别模型紧迫且重要。针对训练数据量不足的问题，为了提高疫情谣言鉴别效果，文中提出了一种基于文本增强和生成对抗网络(GAN)的疫情谣言识别方法。首先，分析疫情谣言的文本特征，提取能表征疫情谣言的特征词；然后，基于GAN构建疫情谣言生成模型，将不含疫情谣言特征的历史谣言，利用疫情谣言特征词库进行文本增强，并生成大量含有疫情谣言特征的新谣言数据；最后，在疫情谣言中补充新生成的谣言数据，从而训练出更准确的疫情谣言分类模型。实验结果表明，使用GAN扩充训练集后，识别效果提高了3个百分点，明显优于传统机器学习和深度学习算法，为重大突发疫情事件中谣言的识别提供了新的途径。

关键词: 新冠疫情；谣言识别；生成模型；文本增强

中图分类号: TP391 **文章编号:** 1000-565X(2021)01-0018-11

2020年1月中下旬开始，新型冠状病毒肺炎疫情逐渐严峻，该重大突发事件来临时，民众本就处于一个相对紧张的状态，此时一些造谣者却在网上大肆发布疫情相关的谣言信息，不断加剧民众焦虑，对疫情的有效治理、国家政策实施及社会稳定带来极为不利的影响。相关辟谣平台虽然会及时更新辟谣信息，但辟谣过程需要向相关专家及政府部门求证，辟谣结果虽准确，但耗时较长，难以满足突发的重大疫情背景下对谣言更快速地进行判别的要求。因此，建立一个及时准确的疫情谣言鉴别模型就显得尤为紧迫和重要。

针对现阶段疫情形势和网络谣言研究现状，本文以提供可靠的疫情谣言鉴别模型为最终目的，以实现对重大突发疫情事件中相关信息的真伪鉴别。其中的核心问题在于，与历史谣言数据相比，疫情谣言的数据量相对较少，且疫情谣言有其独特特征，难以利用现有的判别模型进行鉴别。对此，本文从文本增强和生成对抗两个方面构建疫情谣言判别模型。首先基于疫情谣言的文本特征构建疫情谣言词库，以便对历史谣言进行符合疫情谣言特征的文本增强；再建立基于生成对抗网络(GAN)的疫情谣言生成模型，以扩充疫情谣言的数据量，实现更加精准的疫情谣言鉴别效果。

1 相关研究

早在2002年，网络谣言就引起学者们的关注，

收稿日期: 2020-08-14

基金项目: 国家自然科学基金面上项目(71573196)；中南财经政法大学中央高校基本科研业务费专项资金资助项目(2722020JX007)

Foundation item: Supported by the General Program of the National Natural Science Foundation of China (71573196)

作者简介: 刘勘(1970-)，男，博士，教授，主要从事机器学习和数据挖掘、社交网络与舆情分析研究。E-mail: liukan@zuel.edu.cn

但多属于新闻与传媒领域范畴，聚焦于谣言的传播及监管举措。随着互联网与数据挖掘技术的发展进步，越来越多的计算机技术应用于网络谣言研究，并呈现出较为明显的几个发展阶段。

以特征提取为中心是研究早期的整体趋势。造谣者往往会使用一些情感丰富、表述新颖的语言来吸引大众眼球，因此谣言信息在语言表述、词语使用、标点符号等方面都有一定的相似性。如 Takahashi 等[1]选择信息转发率、爆发点以及词汇分布为主要特征，研究 Twitter 上关于“日本地震引发海啸”事件的数据，发现词汇分布这一特征在谣言与非谣言之间有着显著的差别。Castillo 等[2]同样以 Twitter 数据为研究样本，通过对情感词汇的数目、类别以及发布者的年龄、关注与被关注的数量等特征信息进行分析，人工抽取得到一组可以较好预测信息可信度的特征组合。但这些特征往往是研究者精心设计所得的，时间消耗过多，而且某一数据集上表现良好的特征组合在另外的数据集上效果并不理想，应用局限性较强。因此，近年来以深度学习为核心的谣言鉴别方法应运而生。

深度神经网络在语义挖掘方面成果显著，可巧妙避开早期特征提取的繁琐过程。深度神经网络(DNN)、卷积神经网络（CNN)、循环神经网络(RNN)、长短期记忆（LSTM）等神经网络模型，在研究文本的潜在含义方面具有较高的价值。如 Ma[3-4]等利用 RNN 模型分析某一话题相关帖子的上下文联系，有效减弱了无价值帖子对谣言鉴别结果的影响。刘政等[5]提出了使用 CNN 挖掘文本深层特征的方法，与传统算法相比优势更明显。针对 RNN 模型需要分析所有上下文的不足，Chen 等[6]提出了使用 LSTM 获取文本语义特征的方法，进一步提高了模型检测准确率。Li 等[7]提出了将注意力机制与多任务学习相结合，使用 LSTM 为基础算法的谣言检测方法，其检测效果明显优于主流方法。针对早期谣言检测过程中评论短缺的问题，Qian 等[8]提出了带有评论生成器的卷积神经网络模型，将产生的评论与原文结合后进行谣言检测。Bian 等[9]对谣言的散布结构加以考虑，通过自上而下以及自下而上得到双向信息，其检测效果超过多种前沿方法。

由于缺少对领域特征的提取，以上模型主要针对普通的谣言识别，对于突发事件中网络谣言的研究，仍然停留在传播学、心理学和法学的角度。陈堂发[10]从法律角度出发，提出了追责新思路；韩海峰[11]从医疗角度出发，探讨了公共卫生突发事件中的救治模式；文献［12-13］从政府角度出发寻找了舆论引导的新方法；王雅琪[14]以合肥市 2016 年重大突发事件为例，分析了次生舆情的影响；吴明等[15]研究新媒体在舆情引导中的重要作用。相比之下，针对突发事件谣言鉴别方面的文献较少。张鹏等[16]针对突发事件，结合遗传算法和 BP 神经网络构建出最终的突发事件预警模型。王芳等[17]以新冠疫情谣言为研究对象，建立了疫情相关谣言的真实度评价体系，并构建了信息真实度计算模型，进一步挖掘了谣言真实度与辟谣信息之间的相关关系。

深度神经网络技术是谣言鉴别问题的有效手段，其高效率、高准确率的特点使其得到广泛的应用，但这类模型只能实现基于大量历史数据的谣言预测。新冠疫情从爆发到现在只有几个月时间，虽然期间产生了不少谣言，但相对于数以万计的历史数据而言，疫情谣言数据量仍显不足，如何在此基础上训练出好的判别模型正是本文研究的重点。

2　疫情谣言特征

2.1　数据来源

本文收集新型冠状病毒肺炎疫情爆发以来，从 2020 年 1 月 1 日到 4 月 30 日 161 天内的疫情谣言数据共计 730 条，全部来自于微博社区管理中心且都已经被证实为谣言。作为对比，本文还采集了部分历史谣言数据，共计 39570 条，为微博社区管理中心不实信息板块 2012—2019 年所记载的谣言信息，作为疫情谣言生成模型中的历史谣言数据。同时，选取新华网、人民网、央视网新闻和凤凰资讯 4 大新闻官方网站获取历史非谣言数据。为保证数据涵盖社会生活的各个方面，采集时政、国际、社会、财经、生活、文化、体育、科普、健康、养生、医药、明星、人物、公益、救助共 15 个板块的内容，共计 90337 条非谣言数据，从中随机抽取 4 万条作为历史非谣言数据集。

数据采集使用 Selenium 模拟浏览器获取网页信息，通过 BeautifulSoup 解析网页内容，最后用正则表达式与 Find 函数匹配所需字段。爬取的数据分为疫情谣言、历史谣言和历史非谣言 3 部分。由于一些数据存在数据缺失、数据异常等现象，将此类数据去除后，选用 Jieba 分词器与哈尔滨工业大

学的停用词表完成分词、去除停用词的步骤。使用Word2Vec对文本信息进行向量化转化，将每个词语转化为一个固定长度的向量，对应于固定维度的特征空间。

2.2 疫情谣言的主题特征

通过对疫情谣言的分析，最终将其划分为5个主题，分别为疾病防治、民生保障、政策解读、人物聚焦和灾害救助，其分布比例分别为33.8%、29.2%、12.1%、19.0%和5.9%。

疾病防治类的谣言信息占到总数的三分之一，其次是民生保障类。在每一类主题下，使用TF-IDF找出疫情相关的5个主题词，可真实反映出疫情谣言的主题特征（见表1）。

表1 疫情谣言主题词汇表

Table 1 Glossary of epidemic rumor topics

主题	词1	词2	词3	词4	词5
疾病防治	设施	抢救	医疗	病毒	聚集
灾害救助	药物	捐赠	支援	调用	无偿
政策解读	违法	公布	复工	调任	解封
人物聚焦	英雄	演讲	医生	行程	重症
民生保障	交通	小区	快递	拦截	消毒

2.3 疫情谣言文本特征

根据刘知远等[18]的研究，谣言在语义方面有一定的特点，不同类型的谣言间又有一定的差异。由于历史谣言数据中不包含疫情谣言特征，因此本文通过对疫情谣言情感、词频和权重等方面的分析，获得疫情谣言的文本特征。

(1) 疫情谣言情感词

使用知网情感分析用词语集（beta版）[19]对谣言语句的情感进行分析，得到疫情谣言情感词云图，如图1所示。可以发现疫情谣言的情感具有以下几方面的特点：①情绪化现象严重，“可怕”“恐怖”“不幸”等词语在谣言文本中多次出现，带有较深程度的恐惧、愤怒、同情等情感倾向；②用词消极，夸大现象明显，“允悲”“严峻”“野蛮”等词语出现的次数极多，对国家政策实施及社会稳定带来了不良的影响。

图1 疫情谣言情感词云图

Fig. 1 Word cloud diagram of epidemic rumors emotion

(2) 疫情谣言高频词

疫情谣言文本在词语使用方面也呈现一定的规律，表现在：①有关“极”“非常”的词语（如“细思极恐”“非常多”等）、有关“重”的词语（如包含“重大新闻”“重要”等）以及“一定”“千万”等程度词语出现的频率较高，这类词语会吸引大众眼球，且使得谣言看起来更加可信；②有关“防治”的词语（如“病毒防治”“治疗”“预防”等）、有关医学的词语（如“医院”“专家”“药”“酒精”“消毒”等专业词语）频繁出现，这类词语含有疫情的独有特征，试图让民众将其视为正常的医学救治、防治内容来混淆视听。各词语在730条微博疫情谣言中的出现频次如表2所示。

表2 疫情谣言词语的频数

Table 2 Frequency of epidemic rumor words

词语	次数	词语	次数
不要	48	医院	60
重…	44	消毒	34
极	28	酒精	27
千万	26	药	19
一定	16	防治	17
非常	11	专家	10
…	…	…	…

(3) 疫情谣言关键词

疫情谣言在其主题上具有高度一致性。使用TF-IDF进行词汇分析后，选取在所有文本上结果之和最大的500个词汇，得到疫情谣言关键词表，如表3所示。其中关于中国、美国、武汉、“封城”、医疗、医院、疫情、肺炎等内容的谣言最为突出。

表3 疫情谣言关键词表

Table 3 Keyword list of epidemic rumors

词语	TF-IDF权重和	词语	TF-ID权重和
美国	18.22	肺炎	8.80
中国	13.85	医院	8.66
医疗	13.43	日本	8.21
疫情	10.45	俄罗斯	8.16
允悲	10.16	“封城”	7.90
开学	9.71	消毒	7.82
武汉	9.44	…	…

2.4 谣言文本增强

在计算机视觉领域，对图像数据的增强可提高

图像处理的效果[20]。在自然语言处理领域，也尝试了一些文本增强的方法[21-22]，主要使用词库或向量进行词汇替换、反向翻译、噪声注入等，如 Zhang 等[23]使用基于词典替换的文本增强技术来提高文本分类的效果；Wei 等[24]通过对词语进行删除、插入、替换等操作，在数据量较少时，也能实现分类效果的提升。

由于历史谣言数据在文本内容、情感倾向等方面，与疫情谣言有较大的差别，故本文利用疫情谣言的文本特征对历史谣言进行文本增强。利用 2.3 节的疫情谣言文本的词频、关键词及倒排文档方法构建基于 TF-IDF 的疫情谣言词库，包含程度词（53 条）、情绪词（35 条）、领域词（48 条）和主题词（25 条）4 个部分，对历史谣言文本内容按照全词匹配进行词汇转换，这些词语总的转换次数分别为：程度词（3 304 次）、情绪词（2 986 次）、领域词（2 603 次）、主题词（4 663 次），部分转换关系如表 4 所示。从表 4 可知：

（1）程度词转换　由于疫情谣言中夸大现象明显，在相应事件的表述中，程度副词使用较多，且多为深度副词；通过程度词表，可将历史谣言中程度表述较浅的词语转化为深度程度副词。

（2）情绪词转换　疫情谣言中情绪消极与极端化现象明显，通过构建情绪词表，可将历史谣言数据中情绪不太明显以及情绪较轻的词语，转化为情感极性相同但程度较重的词语。

（3）领域词转换　疫情谣言包含较多的医疗救助领域词语，因此构建疫情领域词表，可将历史谣言中带有明确疾病疾控方面的词，转化为相近的疫情词汇。

表 4　疫情谣言词库表

Table 4　Thesaurus of epidemic rumors

程度词表		情绪词表		领域词表		主题词表	
原词	替换	原词	替换	原词	替换	原词	替换
很	极	坏	阴险	疾病 病情	疫情	设备 装置	设施
较为	极为	粗鲁	野蛮	治疗 预防	防治	急救 急诊	抢救
愈加	更为	生气	暴怒	非典 SARS	新冠	药品 药	药物
要	一定要	危险	惊险	帮助 救助	支援	高铁 公交	交通
少	极少	可怕	恐怖	筹款 借款	捐赠	勇士 伟人	英雄
多	极多	放纵	嚣张	上班 开学	复工	犯法 作案	违法
…	…	…	…	…	…	…	…

3　基于 GAN 的疫情谣言识别模型

3.1　生成对抗学习

近十年来，国内没有爆发过大规模的疫情，历史数据中与疫情相关的信息较少，而此次新冠疫情爆发以来，相关数据量也不足以训练出较好的鉴别模型。因此，本研究基于 GAN 的思想，使用 GAN 对此次新冠疫情的少量谣言进行特征提取，建立疫情谣言生成模型，将普通谣言转化为具有疫情谣言独有特征的生成谣言，从而补充疫情谣言判别模型的训练数据集，其主要流程图如图 2 所示。

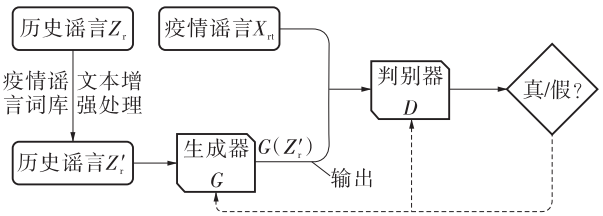

图 2　基于 GAN 的谣言识别框架

Fig. 2　Framework of rumor identification based on GAN

GAN 是 Goodfellow[25] 提出的一种具有对抗机制的生成模型。GAN 由生成器 G 与判别器 D 组成。生成器 G 接收随机数据作为输入，输出与真实数据相似的伪样本；判别器 D 则预测输入样本的真伪，并据此产生预测误差，根据误差依次对生成器 G 与判别器 D 进行更新。

Goodfellow[25] 已经证明，GAN 在训练生成器的过程中，通过改变随机数据的分布，可生成趋向于真实数据的新数据。因此，如果将 GAN 的这种特性应用于疫情谣言的生成，将经过文本增强处理的历史谣言数据作为原始模型中的随机数据，再通过对抗网络改变历史谣言的数据分布，使其转化为具有疫情谣言特征的生成谣言，从而获得大量与此次新冠疫情相关联的谣言数据，就可以弥补数据集中疫情谣言数据量过少的不足。

3.2　模型框架

根据生成对抗的思想，本文从两个方面展开研究：①建立疫情谣言生成模型，将历史谣言转换成一定数量的生成谣言，扩充疫情谣言数据；②构建疫情谣言鉴别模型，将原有的历史谣言、非谣言数据与生成的疫情相关谣言合并后展开训练，扩充数据集，实现鉴别谣言的功能。整体流程如图 3 所示。

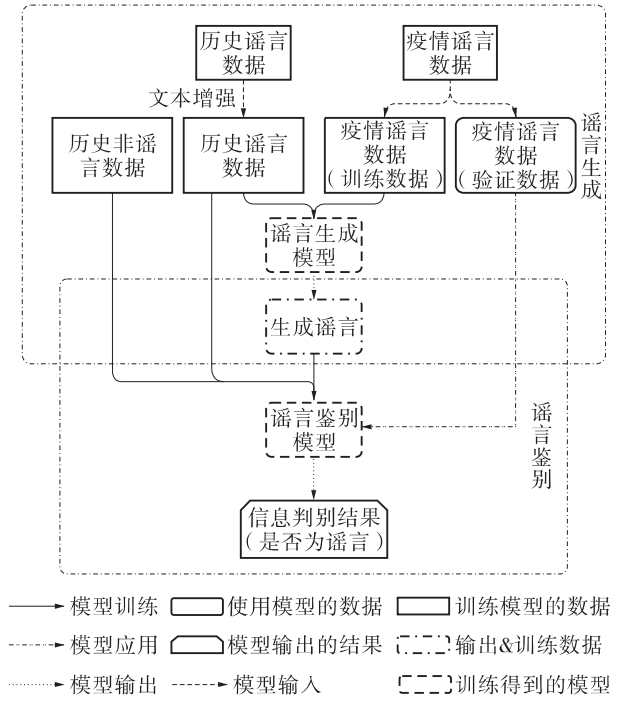

图3 基于GAN的疫情谣言识别流程图

Fig. 3 Identification flowchart of epidemic rumors based on GAN

这里的对抗是历史谣言与疫情谣言的对抗，谣言的数据分布在两者间有一定的差异，因此生成模型的任务就是将历史谣言迁移到疫情谣言的空间，产生新的生成谣言来扩充标注数据集。在生成更多的疫情谣言样本后，疫情谣言鉴别模型利用已有数据与生成数据，进行随后的鉴别过程，得到谣言与非谣言的鉴别结果。

3.3 疫情谣言生成模型

根据图2的生成对抗原理构建图4所示的疫情谣言生成模型，生成器将历史谣言 Z_r' 转变成带有疫情谣言特征的生成谣言 $G(Z_r')$，判别器区分真实的疫情谣言和生成器生成的谣言。在模型迭代足够的次数后，判别器的准确率趋近于50%，疫情谣言和生成谣言具有较高的相似性，判别器无法辨别数据来源。此时生成谣言 $G(Z_r')$ 便可以用来扩充训练数据集，作为标注数据展开模型训练。

生成器 G 对历史谣言 Z_r' 进行重编码，通过与判别器 D 的博弈，生成器生成与真实样本分布相似的假样本，使历史谣言的数据分布不断趋向于疫情谣言 X_{rt}。

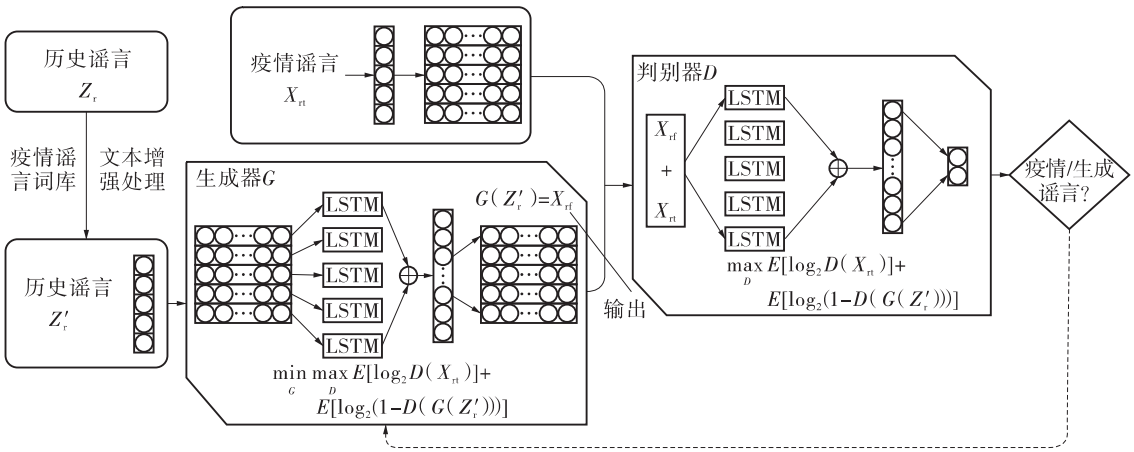

图4 疫情谣言生成模型

Fig. 4 Generation model of epidemic rumors

根据时间将获取的谣言数据划分为历史谣言 Z_r 与疫情谣言 X_{rt}，文本预处理后，利用疫情谣言词库表，实现历史谣言的文本增强操作，使用Word2Vec实现文本向量化。之后将 Z_r' 输入生成器，采用LSTM结构，一方面利用神经网络进行空间迁移，另一方面尽可能保留上下文之间的相关关系。

设历史谣言 $Q=\{[Z_r^{1'},S^1],[Z_r^{2'},S^2],\cdots,[Z_r^{i'},S^i],\cdots,[Z_r^{n'},S^n]\}$，疫情谣言 $R=\{[X_{rt}^1,S^1],[X_{rt}^2,S^2],\cdots,[X_{rt}^j,S^j],\cdots,[X_{rt}^m,S^m]\}$，其中 $Z_r^{i'}$ 表示第 i 个历史谣言，X_{rt}^j 表示第 j 个疫情谣言，S^i 表示其来源（$S^i=1$ 表示来自疫情谣言，$S^i=0$ 表示来自历史谣言）。记生成器生成的第 v 个伪样本为 X_{rf}^v，则 $X_{rf}^v=G(Z_r^{v'},\theta^G)$，其中 θ^G 是生成器的全部参数，根据判别器的样本真假误差不断更新。

判别器以准确区分样本来源（生成谣言/疫情谣言）为目标，依次训练进行对抗博弈，直到判别器无法辨别，至此完成历史谣言的特征迁移工作。判别器接收生成谣言 X_{rf} 与疫情谣言 X_{rt} 作为输入，经过BiLSTM层、全连接层及输出层的对应处

理，判别样本来源。设判别器的输入为 $X_{rf}=\{X_{rf}^1, X_{rf}^2,\cdots,X_{rf}^i,\cdots,X_{rf}^h\}$ 与 $X_{rt}=\{X_{rt}^1,X_{rt}^2,\cdots,X_{rt}^i,\cdots,X_{rt}^m\}$，即 $X=\{X_{rf},X_{rt}\}$，$|X|=h+m$，将 X 输入到 BiLSTM 层得到输出流 $f=\{f_1,f_2,\cdots,f_i,\cdots,f_{h+m}\}$。定义全连接层权值 w_1 和偏置 b_1，输出 $H=\{H_1,H_2,\cdots,H_i,\cdots,H_{h+m}\}$；输出层用于最终判别，其权值为 w_s，偏置为 b_s，输出 $O_s=\{O_{s,1},O_{s,2},\cdots,O_{s,i},\cdots,O_{s,h+m}\}$，则有

$$H_i=\sigma(w_1 f_i+b_1) \tag{1}$$

$$O_{s,i}=\sigma(w_s H_i+b_s) \tag{2}$$

疫情谣言生成模型的全局损失函数由判别器判别样本来源的损失与生成器的损失两部分构成。模型首先训练判别器的参数。设 X_{rt} 的标签为 1，表示疫情谣言；$G(Z_r')$ 的标签为 0，表示生成谣言，判别器以尽可能识别疫情谣言和生成谣言为目标，更新模型参数。判别器更新完成后，固定判别器的参数，更新生成器参数。设历史谣言 Z_r' 的标签为 1（其实 Z_r' 并非疫情谣言），希望判别器将生成数据判别为 1，生成器的参数更新后，可以更好地生成迷惑判别器的数据集。损失函数如下:

$$l=\min_G\max_D\{E[\log_2 D(X_{rt})]+E[\log_2(1-D(G(Z_r')))]\} \tag{3}$$

式（3）中首先最大化右侧的两项内容，此时生成器参数固定，对于疫情谣言 X_{rt}，$D(X_{rt})$ 趋向于 1，第一项趋向于最大值；对于生成谣言 $G(Z_r')$，判别器准确率越高，$D(G(Z_r'))$ 越趋向于 0，第二项也就越大。两项相加，得到的即为最大值，判别器参数得到更新。

然后最小化右侧的两项内容，此时固定判别器参数，$D(X_{rt})$ 即为固定值，第二项中 $G(Z_r')$ 跟随生成器的变化而变化。为更好地迷惑判别器，$D(G(Z_r'))$ 越趋于 1 越好，即第二项内容越小越好，两项相加后，得到的即为最小值，此时生成器参数得到更新。

两步更新完成后，判别器和生成器的参数都得到更新，反复迭代直至判别器无法辨别数据真假为止，至此谣言生成模型构建结束，生成的数据即可在谣言鉴别模型中加以应用。

3.4　疫情谣言鉴别模型

生成大量疫情谣言之后，就可以构建分类器实现疫情谣言鉴别。本文使用基于 BiLSTM 网络的分类器模型，其具体结构如图 5 所示。

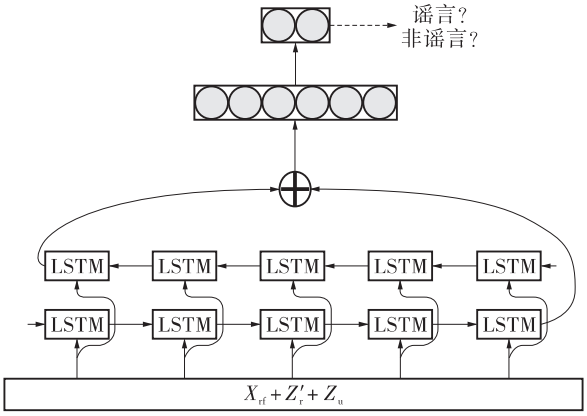

图 5　疫情谣言鉴别模型的结构

Fig. 5　Structure of epidemic rumors identification model

分类器由 BiLSTM 网络、全连接层及输出层组成。模型接收历史谣言、非谣言数据（Z_r' 与 Z_u）及生成谣言 X_{rf} 作为输入，经过 BiLSTM 层进行特征抽取，将正反两个方向的序列化输出相加，再送入全连接层与输出层，对样本是否为谣言进行预测。

分类器输入为 $X_{rf}=\{X_{rf}^1,X_{rf}^2,\cdots,X_{rf}^i,\cdots,X_{rf}^h\}$，$Z_r'=\{Z_r^{1'},Z_r^{2'},\cdots,Z_r^{i'},\cdots,Z_r^{n'}\}$，$Z_u=\{Z_u^1,Z_u^2,\cdots,Z_u^i,\cdots,Z_u^k\}$，$X_{rt}=\{X_{rt}^1,X_{rt}^2,\cdots,X_{rt}^i,\cdots,X_{rt}^m\}$，即 $X'=\{X_{rf},Z_r',Z_u\}$，$|X'|=h+n+k$，将 X' 输入 BiLSTM 层得到正向输出与反向输出，分别为 $L_f=\{L_{f,1},L_{f,2},\cdots,L_{f,i},\cdots,L_{f,h+n+k}\}$ 与 $R_f=\{R_{f,1},R_{f,2},\cdots,R_{f,i},\cdots,R_{f,h+n+k}\}$。定义全连接层权值 w_1'、偏置 b_1' 和输出 $H'=\{H_1',H_2',\cdots,H_i',\cdots,H_{h+n+k}'\}$；输出层用于判断是否为谣言，其权值为 w_c，偏置为 b_c，输出 $O_c=\{O_{c,1},O_{c,2},\cdots,O_{c,i},\cdots,O_{c,h+n+k}\}$，则有

$$H_i'=\sigma[w_1'(L_{f,i}+R_{f,i})+b_1'] \tag{4}$$

$$O_{c,i}=\sigma(w_c H_i'+b_c) \tag{5}$$

4　实验与结果分析

4.1　实验设计

根据本研究的最终目标，将疫情谣言鉴别实验设计为以下 4 种:

实验 1（与传统的机器学习和深度学习方法的对比实验）: 分别使用传统机器学习算法（NB、SVM、DT、集成学习算法 XGBoost、BP 神经网络算法）、深度学习算法（CNN、RNN、LSTM 和 BiLSTM）与本文算法（GAN + BiLSTM/LSTM + BiLSTM）进行比较，其中: GAN + BiLSTM 表示使用 BiLSTM 作为分类器; LSTM + BiLSTM 表示 GAN

模型中，生成器使用 LSTM，判别器使用 BiLSTM 算法。

实验 2（文本增强前后效果对比实验）：对文本进行增强处理后，使用实验 1 中的算法进行实验，检验文本增强处理在此研究中的应用效果。

实验 3（GAN 的生成器与判别器使用不同算法进行对比实验）：分别使用 GAN、VAE 与 LSTM、BiLSTM 算法作为生成器与判别器的核心算法进行对比，选择最优的生成模型。

实验 4（GAN 训练完成后，连接不同分类器的对比分析实验）：分别使用 GAN + CNN、RNN、LSTM、BiLSTM 进行对比，选择最优的鉴别模型。

在实验过程中，采用五折交叉验证评估各模型性能。此外，为确保模型验证的可靠性，对疫情谣言数据按照 8∶2 进行划分，80% 的数据用于疫情谣言生成模型的训练，剩下 20% 的数据不参与整个训练过程，而是用于最终的模型效果检验。

4.2 结果分析

4.2.1 实验 1

首先，使用历史谣言数据集 Z_r 与历史非谣言数据集 Z_u，采用 NB、SVM、DT、XGBoost、BP、CNN、RNN、LSTM 和 BiLSTM 进行实验。然后，将 Z_r 作为生成器的输入信息，从 730 条疫情谣言数据中抽取 584 条谣言，组成 X_{rt} 数据集，训练谣言生成模型。多次迭代后，判别器的准确率在 50% 上下波动，利用此时的生成模型，得到 1 万条生成数据。最后，将生成数据与 Z_u、Z_r 输入谣言鉴别模型，得到本文算法的实验结果（设生成数据与 Z_r 的标签为 1，表示谣言数据；Z_u 为 0，表示非谣言数据），如表 5 所示。其中 A_{train} 表示在训练集上的预测准确率，A_{test} 表示使用训练数据集的 20% 作为测试集的预测准确率，A_n 表示模型在疫情谣言验证集上的预测准确率。从表 5 可以发现：

(1) 在传统的机器学习算法中，NB、SVM、BP 神经网络算法在历史数据的测试集上的预测准确率达到 95% 以上，XGBoost 的为 91.36%，而 DT 算法在测试集上的预测准确率只有 76.44%，分析其原因，可能是由于此研究过程中涉及到的数据量较大，且各数据之间没有明确的特征区分，使得 DT 算法无法进行节点划分，导致最终的预测准确率较低。在疫情谣言验证数据集上，这 5 种算法的预测准确率都在 25%~55% 之间，疫情谣言鉴别结果不理想，一方面是由于训练数据集中没有疫情相关数据，另一方面这类算法较为简单，无法充分学习谣言与非谣言之间的区别，使得算法对此类信息的鉴别能力较弱。

(2) 4 种深度学习算法在历史数据的训练集上都有 95% 以上的预测准确率，CNN 的预测准确率甚至达到 98.78%；算法在测试集上的预测准确率也稳定在 97% 左右，模型训练效果较优；在疫情谣言验证数据集上的预测准确率在 72%~82% 之间，与前面 5 种传统的机器学习算法相比，模型识别效果明显提升，说明在大量的历史数据条件下，深度学习算法更有利于模型的学习。

表 5 实验 1 的结果对比

Table 5 Result comparison of experiment 1

模型	A_{train}/%	l	A_{test}/%	A_n/%
NB			95.57	30.82
SVM			96.77	52.05
DT			76.44	25.62
XGBoost			91.36	45.34
BP			95.65	50.14
CNN	98.78	0.0347	96.58	74.66
RNN	96.76	0.0848	96.68	72.60
LSTM	98.01	0.0532	97.93	80.82
BiLSTM	98.30	0.0432	98.07	81.51
GAN + BiLSTM (LSTM + BiLSTM)	97.43	0.0704	95.20	82.88

(3) GAN 与 BiLSTM 结合的模型在训练集与测试集上的预测准确率与其他几个模型相比并不理想，但在疫情谣言验证集上的实验结果表明，在历史数据的基础上，加上生成数据后，本文模型在疫情谣言上得到的预测效果比未使用生成数据时提高了 1.3 个百分点。可见，生成模型经过训练，得到的生成数据被赋予疫情谣言的特征，可以在疫情谣言的鉴别中定向提高模型的预测准确率。

部分 GAN 生成的数据如图 6 所示，可以发现，生成的数据与疫情、疾病、公益、卫生等内容密切相关，抽象的、情感丰富的词语也较多，符合谣言的特点。

- 环保法好评，交房严查清洁。
- 医院挽救一例北极熊，棒！
- 全民星沦落炮灰，寄托公益挣扎。
- 药品监督管理局公开审理，医科大夫心寒。

图 6 利用 GAN 模型生成的谣言数据

Fig. 6 Rumor data generated by GAN model

4.2.2　实验 2

为了验证本文提出的文本增强方法，对历史谣言进行文本增强处理后，得到历史谣言 Z_r'，再使用实验 1 中的 10 种算法进行实验，结果如表 6 所示。从表 6 可知:

(1) 文本增强处理后，10 种算法在测试集上的预测准确率都有所提高。

(2) 文本增强处理后，SVM 与 XGBoost 算法在疫情谣言验证集上的预测效果较文本增强前有所下降，其余 8 种算法的预测准确率都有所提高，说明本文提出的文本增强方法，对于疫情谣言的鉴别有很好的辅助作用。

表 6　实验 2 的结果对比

Table 6　Result comparison of experiment 2　　%

模型	A_{test}		A_n	
	增强前	增强后	增强前	增强后
NB	95.57	96.22	30.82	30.96
SVM	96.77	97.05	52.05	46.71
DT	76.44	80.48	25.62	27.53
XGBoost	91.36	91.70	45.34	36.99
BP	95.65	96.25	50.14	50.68
CNN	96.58	96.87	74.66	78.08
RNN	96.68	97.45	72.60	80.82
LSTM	97.93	98.09	80.82	81.51
BiLSTM	98.07	98.16	81.51	83.56
GAN + BiLSTM (LSTM + BiLSTM)	95.20	97.77	82.88	84.93

4.2.3　实验 3

为了使疫情谣言鉴别模型拥有最优的性能，对生成器、判别器的算法进行调整，展开对比实验，结果如表 7 所示。对比 1 与 2、3 与 4 号实验（这两组实验是在判别器固定的情况下改变生成器的算法）可以发现，在疫情谣言验证数据集的实验结果上，生成器算法的选择并不能独立影响模型的效果。对比 1 与 3、2 与 4 号实验（这两组实验是在生成器固定的情况下改变判别器的算法）可以发现，判别器算法的选择也不能独立影响模型的效果。对比 1 - 4 号实验可以发现，当生成器使用 LSTM、判别器使用 BiLSTM 时，模型的预测准确率最高，达到 84.93%。对比 5 - 8 号实验可以发现，使用 VAE 作为生成器时，疫情谣言的鉴别准确率维持在 76%~85%，整体效果与 GAN 差不多，但最优准确率为 84.25%，与 GAN 下的最优情况 84.93% 相比，降低了近 0.7 个百分点。

表 7　实验 3 的结果对比

Table 7　Result comparison of experiment 3

序号	模型	A_{train}/%	l	A_{test}/%	A_n/%
1	GAN + BiLSTM (BiLSTM + LSTM)	97.72	0.0606	97.67	83.56
2	GAN + BiLSTM (LSTM + LSTM)	97.77	0.0600	97.63	79.45
3	GAN + BiLSTM (BiLSTM + BiLSTM)	97.89	0.0575	96.17	60.27
4	GAN + BiLSTM (LSTM + BiLSTM)	97.77	0.0604	97.66	84.93
5	VAE + BiLSTM (BiLSTM + LSTM)	97.44	0.0689	97.58	82.88
6	VAE + BiLSTM (LSTM + LSTM)	97.46	0.0705	96.66	76.71
7	VAE + BiLSTM (BiLSTM + BiLST)	98.13	0.0520	97.60	84.25
8	VAE + BiLSTM (LSTM + BiLSTM)	97.76	0.0600	97.76	83.56

4.2.4　实验 4

完成 GAN 的训练后，利用分类器进行谣言判别，不同分类器得到的结果如表 8 所示，从表 8 可以发现:

(1) 对比 1 - 4 号实验（这 4 组实验的生成器均为 LSTM，判别器均为 BiLSTM，采用不同的分类模型）结果，在训练集和测试集上，LSTM 与 BiLSTM 的鉴别效果优于 CNN 与 RNN；在疫情谣言验证数据集上，BiLSTM 也为最优模型，鉴别准确率达 84.93%，而 CNN、RNN 与 LSTM 的分别只有 71.23%、70.55% 与 80.82%，BiLSTM 的优势明显。

(2) 对比 5 - 8 号实验（这 4 组实验的生成器使用 BiLSTM，判别器使用 LSTM，改变谣言鉴别模型的算法）结果，CNN、RNN、LSTM 与 BiLSTM 在训练集与测试集上的鉴别准确率均在 94% 以上，是一个比较理想的结果；而在疫情谣言验证数据集上，BiLSTM 的鉴别准确率为 82.19%，而 CNN、RNN 与 LSTM 的只达到 77.40%、79.45% 与 76.03%，可见，CNN 只能前向传播、RNN 无法解决长时依赖、LSTM 无法获取逆向信息的问题，使

得在同样的数据集上与 BiLSTM 产生较大的差异。

(3) 对比 4、8 号实验，在使用 LSTM、BiLSTM 为生成器与判别器算法，以 BiLSTM 为鉴别模型算法的情况下，模型预测效果是所有实验中最优的，达到 84.93%，为本文研究所得的最终谣言鉴别模型，与文本增强处理前的最佳准确率 81.51% 相比，提高了 3 个百分点。

表 8 实验 4 的结果对比

Table 8 Result comparison of experiment 4

序号	模型	A_{train}/%	l	A_{test}/%	A_n/%
1	GAN + CNN (LSTM + BiLSTM)	97.18	0.0761	97.08	71.23
2	GAN + RNN (LSTM + BiLSTM)	94.92	0.1329	95.08	70.55
3	GAN + LSTM (LSTM + BiLSTM)	97.83	0.0588	97.66	80.82
4	GAN + BiLSTM (LSTM + BiLSTM)	97.77	0.0604	97.66	84.93
5	GAN + CNN (BiLSTM + LSTM)	97.19	0.0767	96.60	77.40
6	GAN + RNN (BiLSTM + LSTM)	95.03	0.1309	95.90	79.45
7	GAN + LSTM (BiLSTM + LSTM)	97.65	0.0636	94.33	76.03
8	GAN + BiLSTM (BiLSTM + LSTM)	97.72	0.0606	97.67	82.19

5 结语

新冠疫情当前，为了使谣言对社会的危害降到最低，就要争取及早准确地鉴别疫情相关谣言。为了弥补疫情数据的不足，实现更加精准的疫情谣言鉴别，本文首先构建疫情谣言词库，利用其进行文本增强；然后使用 GAN 对此次疫情的相关谣言进行特征提取，将历史谣言转化为具有疫情谣言特征的生成谣言，获得大量与此次疫情相关联的谣言数据；最后使用增强后的训练数据集进行判别模型训练，以提高谣言判别的准确率。实验结果表明，使用 GAN 扩充训练集后，识别效果提高了 3 个百分点，明显优于传统的机器学习和深度学习算法。随着时间的推移，获得证实的疫情谣言数量越来越多，此时可以对疫情谣言数据集进行补充更新，使谣言生成模型更加准确地学习疫情谣言的特征，不断提高生成数据与真实疫情谣言之间的相似度，获得更高质量的生成谣言，进而达到更加精准的谣言鉴别效果。

本文的研究方法是针对新冠疫情数据提出的，不论是此次疫情还是未来可能会出现的疫情，本文构建的模型都可以对重大突发疫情的谣言治理起到辅助作用，为普通网民提供相应的判别依据。但疫情毕竟是极少发生的，对疫情谣言及传播者的特征提炼仍有难度，如果加入更多的特征，如谣言传播主体的行为特征，有可能继续提高谣言判别的效果。此外，现实中谣言与非谣言的数据不平衡现象明显，如何在研究中加入对现实情况的考量，也需要展开进一步的研究。

参考文献:

[1] TAKAHASHI T, IGATA N. Rumor detection on twitter [C]// Proceedings of the 6th International Conference on Soft Computing and Intelligent Systems, and the 13th International Symposium on Advanced Intelligence Systems. Kobe: IEEE, 2012: 452-457.

[2] CASTILLO C, MENDOZA M, POBLETE B. Information credibility on twitter [C]// Proceedings of the 20th International Conference on World Wide Web. Hyderabad: ACM, 2011: 675-684.

[3] MA J, GAO W, WEI Z, et al. Detect rumors using time series of social context information on microblogging websites [C]// Proceedings of the 24th ACM International Conference on Information and Knowledge Management. Melbourne: ACM, 2015: 1751-1754.

[4] MA J, GAO W, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks [C]// Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York: AAAI Press, 2016: 3818-3824.

[5] 刘政，卫志华，张韧弦. 基于卷积神经网络的谣言检测 [J]. 计算机应用，2017，37(11)：3053-3056，3100.
LIU Zheng, WEI Zhihua, ZHANG Renxian. Rumor detection based on convolutional neural network [J]. Journal of Computer Applications, 2017, 37(11): 3053-3056, 3100.

[6] CHEN T, LI X, YIN H, et al. Call attention to rumors: deep attention based recurrent neural networks for

early rumor detection［C］//Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining. Melbourne: Springer, 2018: 40-52.

［7］ LI Q, ZHANG Q, SI L. Rumor detection by exploiting user credibility information, attention and multi-task learning［C］//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: ACL, 2019: 1173-1179.

［8］ QIAN F, GONG C, SHARMA K, et al. Neural user response generator: fake news detection with collective user intelligence［C］//Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm: IJCAI, 2018: 3834-3840.

［9］ BIAN T, XIAO X, XU T, et al. Rumor detection on social media with bi-directional graph convolutional networks［C］// Proceedings of the AAAI Conference on Artificial Intelligence. New York: AAAI Press, 2020: 549-556.

［10］ 陈堂发. 突发危机事件中谣言追责的理性问题: 基于区块链技术支撑的讨论［J］. 人民论坛·学术前沿, 2020(5): 15-21.
CHEN Tangfa. The rationality issue of holding rumor spreaders accountable in emergencies: discussion based on the blockchain technology［J］. Frontiers, 2020(5): 15-21.

［11］ 韩海峰. 关于公共卫生突然事件的应急救治模式探讨［J］. 智库时代, 2019, 3(2): 24, 26.
HAN Haifeng. Discussion on emergency treatment mode of public health emergencies［J］. Think Tank Era, 2019, 3(2): 24, 26.

［12］ 吕蒙蒙. 突发事件中政府信息公开和公信力建设问题研究［J］. 新闻研究导刊, 2017, 8(16): 106.
LÜ Mengmeng. Research on government information disclosure and credibility construction in emergencies［J］. Journal of News Research, 2017, 8(16): 106.

［13］ 史少春. 加强信息公开与新闻宣传做好重大突发事件舆论引导［J］. 中国行政管理, 2020, 36(2): 27-28.
SHI Shaochun. Strengthen information disclosure and news propaganda to guide public opinion in major emergencies［J］. Chinese Public Administration, 2020, 36(2): 27-28.

［14］ 王雅琪. 公共危机中“次生舆情”的考察: 以合肥市 2016 年重大突发事件为例［J］. 视听, 2017, 12(9): 127-128.
WANG Yaqi. Investigation of “secondary public opinions” in public crisis: a case study of major emergencies in Hefei in 2016［J］. Radio & TV Journal, 2017, 12(9): 127-128.

［15］ 吴明, 甄贞, 王小玲. 在突发事件中政务新媒体的舆论引领研究［J］. 中国管理信息化, 2017, 20(17): 204-206.
WU Ming, ZHEN Zhen, WANG Xiaoling. Research on public opinion guidance of new media of government affairs in emergencies［J］. China Management Informationization, 2017, 20(17): 204-206.

［16］ 张鹏, 兰月新, 李昊青, 等. 突发事件网络谣言危机预警及模拟仿真研究［J］. 现代情报, 2019, 39(12): 101-108, 137.
ZHANG Peng, LAN Yuexin, LI Haoqing, et al. Study on the simulation and crisis early-warning of Internet rumors about emergencies［J］. Journal of Modern Information, 2019, 39(12): 101-108, 137.

［17］ 王芳, 连芷萱. 公共危机中谣言真实度计算及其与正面信息的交锋研究［J］. 图书与情报, 2020, 41(1): 34-50.
WANG Fang, LIAN Zhixuan. Authenticity grade calculation of rumor in public crisis and its confrontation with positive information［J］. Library & Information, 2020, 41(1): 34-50.

［18］ 刘知远, 张乐, 涂存超, 等. 中文社交媒体谣言统计语义分析［J］. 中国科学(信息科学), 2015, 45(12): 1536-1546.
LIU Zhiyuan, ZHANG Le, TU Cunchao, et al. Statistical and semantic analysis of rumors in Chinese social media［J］. Scientia Sinica (Informationis), 2015, 45(12): 1536-1546.

［19］ 董振东. 知网情感分析用词语集 (beta 版)［EB/OL］. (2007-12-22)［2020-05-31］. http://www.keenage.com/html/c_bulletin_2007.htm.

［20］ ZHANG H, CISSE M, DAUPHIN Y N, et al. Mixup: beyond empirical risk minimization［EB/OL］. (2018-04-27)［2020-05-31］. https://arxiv.org/pdf/1710.09412.pdf.

［21］ KOBAYASHI S. Contextual augmentation: data augmentation by words with paradigmatic relations［EB/OL］. (2018-05-16)［2020-05-31］. https://arxiv.org/pdf/1805.06201.pdf.

[22] ANABY-TAVOR A, CARMELI B, GOLDBRAICH E, et al. Not enough data? Deep learning to the rescue! [EB/OL]. (2019-11-27) [2020-05-31]. https://arxiv.org/pdf/1911.03118.pdf.

[23] ZHANG X, ZHAO J, LECUN Y. Character-level convolutional networks for text classification [C]// Proceedings of Advances in Neural Information Processing Systems. New York: Curran Associates, 2015: 649-657.

[24] WEI J, ZOU K. EDA: easy data augmentation techniques for boosting performance on text classification tasks [EB/OL]. (2019-08-25) [2020-05-31]. https://arxiv.org/pdf/1901.11196.pdf.

[25] GOODFELLOW I. NIPS 2016 tutorial: generative adversarial networks [EB/OL]. (2017-04-03) [2020-05-31]. https://arxiv.org/pdf/1701.00160.pdf.

Rumor Identification in Major Sudden Epidemic Situation

LIU Kan[1] *HUANG Zheying*[2]

(1. School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan 430073, Hubei, China;
2. School of Business, Nankai University, Tianjin 300071, China)

Abstract: Since the outbreak of the covid-19 epidemic, related rumors have spread rampantly. Traditional rumor identification models have difficulties in epidemic rumor identification because the size of epidemic rumors is not large enough to train a good classification and identification model. Therefore, it is an urgent task to build a rumor identification model based on a small amount of epidemic rumor data. To deal with the problem of insufficient training data, text enhancement and generative adversarial networks (GAN) methods were used to generate a large amount of information similar to epidemic rumors and to improve the identification effect of epidemic rumors. First, the textual characteristics was analyzed to extract keyword of epidemic rumors. Second, epidemic rumor generation model was constructed based on the idea of GAN, and historical rumors which do not contain epidemic rumor features were textually enhanced by the epidemic rumor feature thesaurus, and a large amount of new rumor data containing epidemic rumor features were generated. Finally, the newly generated rumor data are combined with the epidemic rumor data to train a more accurate classification model of the epidemic rumor. Experiment results show that the rumor identification effect is improved by 3% after using the GAN extended training set. The new model is evidently much better than the traditional machine learning and deep learning algorithms, and it provides a new way for the identification of rumors in public health emergency.

Key words: covid-19 epidemic; rumor identification; generation model; text enhancement