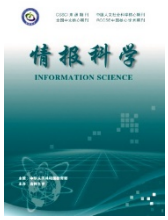

情报科学
Information Science
ISSN 1007-7634,CN 22-1264/G2

《情报科学》网络首发论文

题目：基于信息可信度评估的突发公共卫生事件谣言识别研究
作者：钱旦敏，郑建明，王文敬，马野青
网络首发日期：2023-09-19
引用格式：钱旦敏，郑建明，王文敬，马野青．基于信息可信度评估的突发公共卫生事件谣言识别研究[J/OL]．情报科学.
https://link.cnki.net/urlid/22.1264.g2.20230914.2050.006

网络首发时间：2023-09-19 11:37:00
网络首发地址：https://link.cnki.net/urlid/22.1264.g2.20230914.2050.006

基于信息可信度评估的突发公共卫生事件谣言识别研究*

钱旦敏[1,2,3] 郑建明[3] 王文敬[2] 马野青[1]
(1.南京大学 商学院，江苏 南京 210008；2.南通大学 医学院，江苏 南通 226000；
3.南京大学 信息管理学院，江苏 南京 210023)

摘　要：【目的/意义】随着突发公共卫生事件不断演变，人们对其认识有一个从模糊到精确的过程。笔者对突发公共卫生事件网络信息可信度进行量化分级，为更细致化的谣言识别提供数据支持。【方法/过程】分析并选取信息文本关键词、情感、评论、信源和媒体五大静态特征，融合时间和当日新增确诊数两大动态特征，结合熵值法将其量化得到谣言指数RI，基于此引入“宽容区间”，并借助朴素贝叶斯分类器确定界限，将谣言识别结果的可信度分为低、中、高三类。【结果/结论】该模型在训练集和验证集上表现良好，正确率分别为95%和90.20%，与决策树和SVM两个基线模型相比，模型各项性能指标均有显著提升。【创新/局限】本研究建立了一种基于信息可信度评估的谣言识别模型，通过建立RI指数，并创新纳入有关疫情状况的动态指标，提高谣言识别的精准度；引入对谣言进行可信度分级，突破传统的谣言识别二分类检测方法的局限性。

关键词：突发公共卫生事件；谣言识别；网络舆情；可信度评估；谣言指数；熵值法
中图分类号：G203

0 引言

2023年3月，中国互联网络信息中心CNNIC发布的第51次《中国互联网络发展状况统计报告》显示：2022年，我国网民规模为10.67亿，互联网普及率达75.6%[1]。互联网平台的发展，降低了信息传输成本，加快了信息传播的速度、广度和深度，推动了新型传播模式的形成。然而，由于缺乏有效的信息监控手段，谣言频发引起了社会恐慌，形成了错误的舆论，破坏了社会秩序[2]。谣言既与公众心理契合，又与公众切身利益关联，可快速唤起公众焦虑、愤懑等共有情绪，轻易将子虚乌有的事件升级扩大，实现快速传播，造成恶劣影响[3]。

突发公共卫生事件的发生通常毫无征兆，事件影响范围广，波及人数多，且涉及人民切身的健康安全，阻碍了经济发展和社会稳定，迫使人们对其给予高度重视[4]。同时，卢建平教授提到重大疫情从初露端倪、渐成规模到集中暴发总是一个渐进的过程，人们的认识也需要由模糊到精确[5]。目前，已有学者通过不同方法对舆情热度[6]和信息真实性[7]进行研究，但对于模糊的、难以定性的言论依然无法有效辨别。

基于上述考虑，笔者在谣言识别过程中融入信息可信度分级的思想，选取关键词、情感、评论、信源、媒体、时间、当日新增确诊数等七大特征，结合熵值法将其量化得到谣言指数RI（Rumor Index，简称RI），并引入“宽容区间”将其分为低R、中G、高T三个等级，从

*基金项目：教育部人文社会科学研究基金项目“新市民公共卫生信息精准化服务模式研究”(项目编号：17YJCZH140)；2022年南通大学大学生创业训练计划项目“突发公共卫生事件中用户集成式健康信息精准服务模型创新研究”(项目编号：CY2022014)。
作者简介：钱旦敏（1981-），女，江苏启东人，博士，副教授，主要从事健康管理、数据挖掘研究，通讯作者补充邮箱 qdm11@163.com；郑建明（1960-），男，江苏南京人，博士，教授，博导，主要从事信息资源管理、公共文化服务研究；王文敬（2001-）女，江苏如东人，本科生，主要从事医学信息学研究；马野青（1966-），男，江苏海安人，博士，教授，博导，主要从事数字经济、国际贸易研究。

而提高突发公共事件中谣言识别准确率，有效抑制其传播，维护健康的网络生态环境。

1 相关文献研究

1.1 信息可信度研究

信息可信度极大程度上影响受众的直观感受和观点倾向，从而改变舆情方向和结果[8]。目前国内外诸多学者从不同角度进行了相关研究：①理论研究方面，如宋士杰等梳理了信息可信度的相关理论研究，并在 iField 视域下对其未来研究方向进行了前沿展望[9]。徐静等利用时效性、权威性、影响力和关注度四大特征提出了一个多维度加权结合的可信度计算方法，并利用 Web 信息验证方法的有效性[10]。Castillo 等基于信息内容、话题、受众、传播方式等四大特征七十四个分项建立多级社交网络信息可信度评价指标体系，为信息内容可信度评估提供参考依据[11]。ODonovan 等分析了 Twitter 中各功能的效用，发现关注数、转发数、内容长度、URL 等可作为信息可信度评估的最佳指标[12]。②可信度在突发公共卫生事件方面的研究，如谢娟等以新冠疫情为例，提出用户信息可信度的判断依据[13]。曹树金等基于微信数据，研究突发公共卫生事件中信息可信度影响的若干因素[14]。曾祥敏、何鸿雁等利用媒介信息的可信度评价，探讨新冠肺炎疫情中用户媒介信息接触及传播意愿[15-16]。Bora 等从内容、转发数、浏览次数、来源等方面对寨卡疫情期间 YouTube 发布的信息可信度进行评估[17]。Thomson 等引入传播者与事件暴发点的地理位置对信息可信度进行评价，发现与现场距离越近越能增加信息的可信度[18]。③可信度在谣言识别方面的研究，如滕婕等构建了基于 Multi-Agent 的信息主体信任识别模型，寻找出网络谣言传播事件中可靠的信息主体[19]。曾子明等利用用户可信度、微博影响力、文本-主题三大特征变量，采用随机森林算法进行谣言识别[20]。Wu 基于用户特征、内容特征、时间特征和评论特征等对新浪微博信息进行可信度分析，采用逻辑回归分类算法将信息分为谣言信息和非谣言信息[21]。Liu 等基于信息来源的可信度、平台、发布者等特征采用实时性算法对谣言信息进行实时监测[22]。

1.2 突发公共卫生事件谣言识别研究

在网络大环境下，突发公共卫生事件期间谣言频发，由此对疫情防控、经济发展、社会稳定、人民心理等各个方面都会产生巨大的负面影响。为了有效抑制因谣言传播而造成的各类伤害，国内外学者纷纷投入到突发公共卫生事件的谣言识别研究工作中。曾子明等提出基于 Transformer 和 BiLSTM 模型，并纳入情感识别，解决突发公共卫生事件情景下谣言识别任务数据不平衡且带标签数据量少的问题[20]。孙冉等构建融合文本语义特征的多特征谣言识别模型，并计算不同特征在谣言识别中的重要程度[23]。冯兰萍等利用用户基本特征、内容特征、传播特征、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[24]。石锴文等使用 BiLSTM+DNN 模型对谣言文本的四大主要特征向量进行分类判别[25]。Guo 采用朴素贝叶斯算法结合拉普拉斯平滑对社交网络文本中的谣言进行识别，实现对谣言的特征以及网络消息文本的自动识别与分类[26]。Suthanthira 等提出准确性检测神经网络算法，有效提高推文中谣言识别的准确率[27]。Srinivasan 等为识别具有数据稀疏性的谣言，提出了一种具有新激活函数的双重卷积神经网络方法[28]。Yuan 等提出了一种基于扩张卷积的模型 DBCN，实现以更少的信息损失自动提取输入的特征，较其他线性模型有效提高了 F1 平均度量[29]。

1.3 相关研究述评

综观国内外现有相关研究发现：①虽然信息可信度分别在谣言识别和突发公共卫生事件两个领域都有涉及，但国内研究较少从信息文本的特征出发进行可信度评估：可信度在谣言识别领域的研究偏向于对用户粉丝数、关注数和已发微博数等用户主体信息和信息推荐者及其评价值进行综合计算得到信息主体的信任度；可信度在突发公共卫生事件方面的研究则主要凭借公众对媒介的依赖程度和对媒介的主观评价来测算可信度。②疫情的走势对谣言产生

有较强影响[30]，但有关突发公共卫生事件谣言识别模型的研究中尚未引入不同时间段有关疫情状况的动态指标。针对上述现象，笔者基于信息文本的关键词、情感、评论、信源和媒体等五大静态特征，并融合时间、当日新增确诊数等两大动态特征，构建突发公共卫生事件谣言识别可信度分级模型。

2 基于信息可信度评估的谣言识别模型构建

笔者通过对信息的关键词、情感、评论、信源、媒体、时间、每日新增确诊数等七大特征进行分析，采用熵值法优化特征权重，使用加权计算和 sigmoid 函数获得谣言指数 RI，将信息属于舆情的可信度量化，引入“宽容区间”划分可信度等级。

2.1 构建特征指标

（1）关键词特征：随着时间的推移与事件的进展，所涉及的关键词往往受政策背景与舆论导向等影响而改变，其特征与谣言的演化密切相关。关键词作为谣言重要文本特征之一，其频次分布差异特征可有效区分谣言与非谣言[31]。因此，笔者拟对训练集中谣言和非谣言样本进行 jieba 分词与去停用词后分别统计词频，得到谣言与非谣言的词频数据集 A 和 B。将待识别样本分别与两个数据集取交集，得到集合 C 和 D。结合贺刚等[31]的关键词特征公式，笔者综合考虑交集中关键词的数量及其在样本中的的词频及在谣言和非谣言数据集中的相应的词频占比，建立公式（1）和（2）用于获得关键词谣言指数 kw_r（rumorkeywords）和关键词事实指数 kw_t（truthkeywords），并用 kw_r - kw_t 得到样本的关键词特征值。

$$kw_r = N_C * \sum_{i=1}^{n} M_i * \frac{n_i}{n_C} \quad （1）$$

N_c 表示集合 C 的关键词数量，M_i 表示集合 C 中关键词 i 在集合 A 中的词频，n_i 表示关键词 i 在集合 C 中的词频，n_C 表示集合 C 中所有关键词的总频次。

$$kw_t = N_D * \sum_{j=1}^{n} M_j * \frac{n_j}{n_D} \quad （2）$$

N_D 表示集合 D 的关键词数量，M_j 表示集合 D 中关键词 j 在集合 B 中的词频，n_j 表示关键词 j 在集合 D 中的词频，n_D 表示集合 D 中所有关键词的总频次。

（2）情感特征：情感特征能在谣言识别中发挥重要作用，且在人人自危的突发公共卫生事件中，公众恐慌、焦虑、敏感多疑、紧张等各种负面情绪愈加凸显，谣言情感特征被放大[32]。因此，笔者拟使用 python 软件包 snownlp 对训练集中的样本进行情感分析，得到谣言和非谣言情感因子得分分布，情感得分在 0 到 1 之间，接近 0 表示正面情绪，接近 1 则为负面情绪。

（3）评论特征：评论中包含许多验证消息真实性的潜在证据，通常是公众对原文的真实反映和观点[33]。当人们无法对信息真伪进行评价时，常产生负面情绪，并使用质疑和批判性质的词语[34]。因此，笔者拟将评论中含有“谣言”“辟谣”“假的”等表示质疑的词汇（如图 1 所示）的样本编码为 1，否则编码为 0；对无评论的样本，则视为无质疑或反对等意见，编码为 0。

comment 1	不是合肥，大县城小机场没有摆渡车下穿通道，此外，合肥之前接收的4批航班都是大半夜，只有16日是上午9点40，而且靠停廊桥
comment 2	官方已辟谣，压根就不是合肥的！
comment 3	辟谣了已经
comment 4	这是造谣
comment 5	额，虽然但是，这是乌龙，那个人不是红码
comment 6	朋友 你在搞什么啊 很多年前的事情了？？？？你这样别人不知道以为是现在的事情
comment 7	2年前的事情，你现在发出来造谣？？？
comment 8	我不看就是没有，我不找就是不存在，太专业你也看不懂不是吗？
comment 9	都假的还不删？
comment 10	这个事发生在昆明，不是合肥。

图 1　谣言的评论示例

Figure 1 Sample comments for rumors

（4）信源特征：用户信息一定程度上隐含了用户权威程度和可信度，经过验证的用户可能更信任源于新闻媒体的信息，非认证用户散布谣言的可能性大于认证用户[35-36]。因此，笔者拟根据以下标准对信息来源进行赋值：中央新闻门户赋分 1，经认证的官方组织赋分 2，有认证的个人用户赋分 3，无认证用户赋分 4。

（5）媒体特征：随着多媒体技术的发展，谣言形式呈现出多样化。多数学者认为谣言中加入图片、视频等来传递更多的信息，是为了增加谣言的迷惑性[23]。因此，笔者拟根据样本中是否含有图片、音频或视频数据进行分类，若带有上述媒体特征，则编码为 1，否则编码为 0，从而构建谣言识别的媒体特征。

（6）时间特征：信息传播受到信息发布时间的影响，发布时间的差异会导致接收到信息的用户数量不同[23]。因此，笔者先考虑纳入 3 个指标，包括：信息发布的时间段、星期和是否在节假日发布，用以刻画用户发帖的时间特征。其中，将信息发布时间段划分为三个阶段：上午（6:01-12:00）、下午（12:01-18:00）和夜间（18:01-24:00，00:01-6:00）。在同一谣言事件中，一般先发布的信息容易引起较大的关注，而且在辟谣的过程中，大众对谣言的关注度会逐渐减弱。因此，本文又引入了时间差特征 T，表示在一个谣言事件中，信息与最先发布谣言的信息之间的时间差。如公式（3）所示：

$$T=T_{pq}-T\_f_p \tag{3}$$

其中，T_{pq} 表示第 p 个主题第 q 条信息发布的时间，T_f_p 表示第 p 个主题第一条信息发布的时间。

（7）当日新增确诊数：新增确诊数量是疫情发展态势的一大特征，对谣言的产生具有较强的影响[30]。因此，笔者以中国疾病预防控制中心发布的新型冠状病毒感染疫情最新情况（https://www.chinacdc.cn/jkzt/crb/zl/szkb_11803/jszl_11809/）为数据源，采集全国每日新增确诊数量作为该指标数值。

2.2 生成谣言指数 RI

2.2.1 权重优化

由于各指标对谣言识别的影响大小存在差异，笔者针对上述构建的多维度特征指标体系，借鉴林平等评价政府开放数据服务水平的方法，采用熵值法作为指标权重的优化方法，确定七大特征在该指标体系中各自所占权重[37]：首先，利用训练集构造一个具有 m 条文本、n 个评价指标的初始数据矩阵 X，$X=\left(X_{ij}\right)_{m\times n}$，$X_{ij}\left(i=1,2,\cdots,m,\ j=1,2,\cdots,n\right)$ 表示第 i 条文本的第 j 项评价指标值。然后，分别对正、负向指标数据进行归一化处理后得到 X'_{ij}（$X'_{ij}\in[0,1]$）。最后，通过计算指标的比重、信息熵值、信息熵冗余度，得到各特征指标权重 w_j（$j=1,2,\cdots,n$）。

2.2.2 特征融合

为将七个特征融合得到谣言指数 RI，笔者利用上述熵值法得到的权重进行加权计算，并将其结果使用 sigmoid 函数映射到[0,1]的范围之内，得到谣言指数 RI，如公式（4）所示，将识别结果量化。

$$RI = sigmoid(W \bullet X^{t}) \quad (4)$$

RI 代表每个样本为谣言的可能性，$sigmoid(x)=\frac{1}{1+e^{-x}}$，$W$ 代表权重矩阵，X 代表特征矩阵，$x_1,...,x_7$ 分别代表关键词特征、情感特征、评论特征、信源特征、媒体特征、时间特征、当日新增确诊数七个特征值，$\omega_1,...,\omega_7$ 分别代表各特征值对应的权重值，记训练集中谣言数据的谣言指数为 RI_r，记非谣言数据的谣言指数为 RI_t。

2.3 实现网络舆情可信度分级

由于突发公共卫生事件中社交平台的言论结构复杂，难以给出“非黑即白”的定义，使用二分类方法对事件的真实性和可信度进行判断不够严谨，这样的“一刀切”往往存在较为明显的假阳性或者假阴性问题。因此，笔者基于信息的可信度及真实性评分对网络舆情进行分级，利用 RI 引入“宽容区间”，将结果的可信度分为三级：低（R）、中（G）、高（T）。其中，宽容区间为言论性质模糊的灰色 RI 区间，其值越大，信息越容易被划分为 G，反之，越难被划分为 G。宽容区间的确定方法如下：

输入：训练集中谣言数据特征集 $F_r=\{X_1, X_2,...X_n\}$ 与非谣言数据特征集 $F_t=\{X'_1, X'_2,...X'_m\}$。谣言数据谣言指数 $D_{RIr}=\{RI_1,RI_2,...RI_n\}$，非谣言数据谣言指数集 $D_{RIt}=\{RI'_1,RI'_2,...RI'_m\}$，原分类标签 Y_o。其中 $X_i=\{x_{1i},x_{2i},x_{3i},x_{4i},x_{5i},x_{6i},x_{7i}\}$，$Y_o \in \{0,1\}$，n 为谣言数据数量，m 为非谣言数据数量，i 为样本数量。宽容区间的确定方式如下：

Step1:分别求出 D_{RIr} 和 D_{RIt} 的概率密度函数 $PDF_r(x)$与 $PDF_t(x)$及其期望 $E_r(x)$与 $E_t(x)$，由 $PDF(RI=b|Y_o=1)=PDF(RI=b|Y_o=0)$，$(E_t(x)<x<E_r(x))$，得到决策边界 b。

Step2:记宽容区间长度的一半为 w_h，朴素贝叶斯分类模型为 f(x)，以$[b-w_h,b+w_h]$为宽容区间按如下规则将样本划分 R、G、T 三级：数据集中谣言指数小于宽容区间下界者划分为 T，大于宽容区间上界者划分为 R，其余为 G，得到 $Y=\{y_1,y_2,...,y_k\}$，其中 k 为样本总数，$y_i \in \{R,G,T\}$，i 为第 i 个样本。将 w_h 初值赋为 0，步长为 0.001。

Step3:判断 wh 是否在区间[0.001,0.5]内，若在则执行 Step4，反之则退出。

Step4:在训练集上利用训练 f(x)得到$\hat{Y}$。

Step5:若在训练集中 $P(\hat{Y}=Y)>0.95$，则计算 f（x）在验证集上的正确率并输出 w_h，反之扩展宽容区间，$w_h:=w_h+0.001$，返回 Step3。

输出：宽容区间界限 $b-w_h$ 与 $b+w_h$。

3 实证分析

3.1 数据采集

3.1.1 数据来源

根据国家卫生健康委发布的最新公告，自 2023 年 1 月 8 日起，我国新冠疫情已基本结束，进入“乙类乙管”常态化疫情防控阶段。故笔者收集 2019 年 12 月 1 日至 2023 年 1 月 8 日在微博、知乎、“科普中国”和“科学辟谣”等知名度较高的官方微信公众号中发表的关于新

冠疫情、猴痘、出血热等突发公共卫生事件的信息，随机选取 7 类话题共 10 352 条，其中谣言 5 865 条，非谣言 4 487 条。由于微信公众号平台在信源特征、媒体特征两方面较弱，故用该平台的文本信息来构建训练集的语料库。微博和知乎具有本文涉及的所有特征，可作为本文的训练集、测试集和验证集。随机分别抽取谣言与非谣言的 80%作为训练集，将剩余的 20%作为验证集。

多数学者对谣言识别的研究止于模型的构建与检验，利用的是已知真实性的数据集，鲜有采集未知真实性的数据进行实证研究。杨建林等指出我国图书情报学领域的舆情研究成果多偏于学术化与理论化，难以真正落地实施[38]。因此，笔者以新冠疫情为例，用“新冠疫情”关键词在微博中检索，随机选取疫情发展各阶段共 470 条微博作为测试集。由于突发公共卫生事件具备突发性的属性，故无法构建出适用于所有突发公卫事件的语料库，为缓解该问题，笔者在测试集中构建具有针对性的词频数据集，但训练集中的词频数据集无须更改，以减轻后期构建突发公卫事件模型时需重建训练集的工作量。

3.1.2 数据预处理

首先分别对训练集、验证集和测试集进行数据集成、数据清洗，去除纯记叙类文本和无意义图片，然后在 python3.10 编程环境下采用 jieba 分词器与哈尔滨工业大学的停用词表，分别对三个数据集的文本数据进行分词和去除停用词，以便对关键词词频进行统计。

3.2 基于信息可信度的谣言分级

3.2.1 生成谣言指数 RI

首先，笔者基于优化的突发公共卫生事件谣言识别指标体系，获取训练集中各特征指标值。其中，关键词特征值是基于对训练集中文本的关键词词频分布统计结果进行的计算，如图 2 所示；情感特征由 python 软件包 snownlp 处理训练集中的数据分析得到，其范围为[0,1]，表示该文本情感为正面、积极的概率，如图 3 所示；评论特征、信源特征、媒体特征、时间特征（如图 4）、当日新增确诊数（如图 5）由前述规则赋值或计算得到。然后，利用熵值法优化各指标权重，其中，关键词特征、情感特征、评论特征、信源特征、媒体特征、时间特征、当日新增确诊数特征所占权重分别为[0.006，0.08，0.371，0.178，0.168，0.058，0.139]（保留三位小数）。其中，关键词特征的权重占比相对较低，其原因可能与突发公共卫生事件中群众关注较为集中导致关键词具有一定的相似性有关；评论特征的权重明显较高，这可能由于此类事件中群众的高参与度以及对谣言的高讨论度。最后，笔者将加权计算的结果使用 sigmoid 函数映射到[0,1]的范围之内，得到训练集中谣言与非谣言样本的谣言指数 RI，其分布情况如图 6 所示。

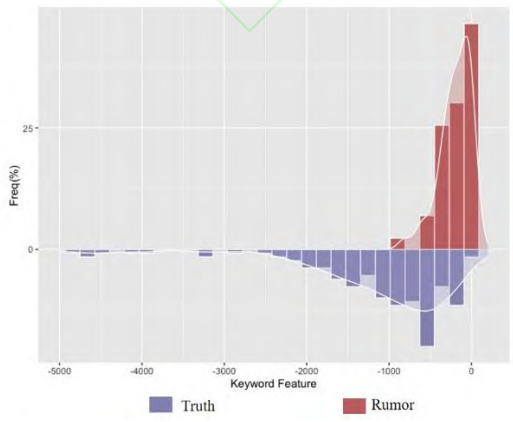

图 2 关键词特征分布图

Figure 2 Keywords Feature distribution map

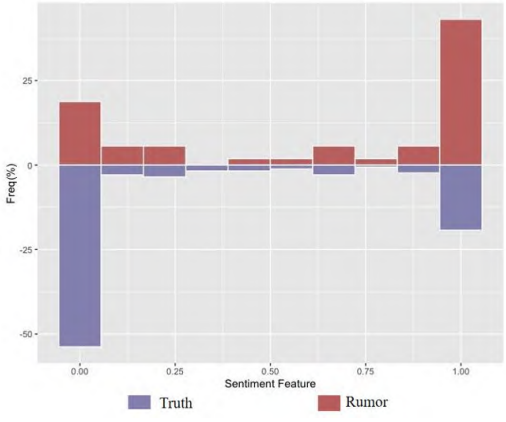

图 3 情感特征分布图

Figure 3 Emotional feature distribution map

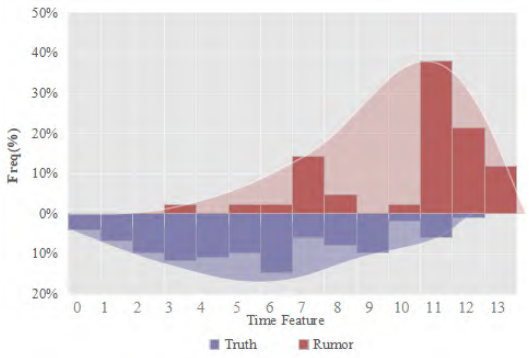

图 4 时间特征分布图

Figure 4 Time Feature distribution map

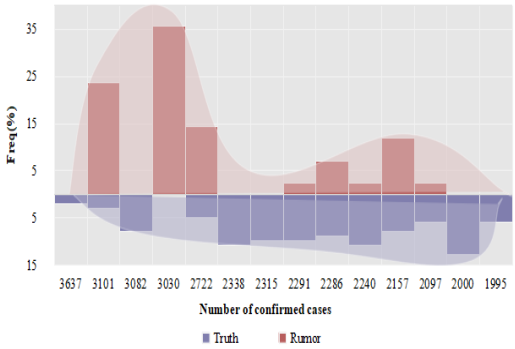

图 5 当日新增确诊数特征分布截图

Figure 5 Number of newly confirmed cases distribution map

3.2.2 划分界限的确定

根据上述宽容区间的确定方法执行 Step1~Step5，得到训练集上三分类结果的正确率 P 与 w_h 的关系（如图 6 所示）。当 w_h 达到 0.170 时，正确率 P 达 0.962，随着宽容区间进一步扩大，正确率发生波动，此时的宽容区间无法很好地区分出 G，可能是由于宽容区间过宽导致非 G 类样本过少而导致的假性上升。综上所述，将宽容区间设定在 0.170 最为合适，既有效区分出 G 类，又不会因为宽容区间过宽造成样本数量不均。训练集中谣言与非谣言样本的谣言指数 RI 分布情况及 R、G、T 的区间如图 7 所示，对应的 R、G、T 划分界限如表 1 所示。使用最终得到的贝叶斯模型在验证集上验证，正确率达 90.20%，可见其没有出现过拟合的情况。

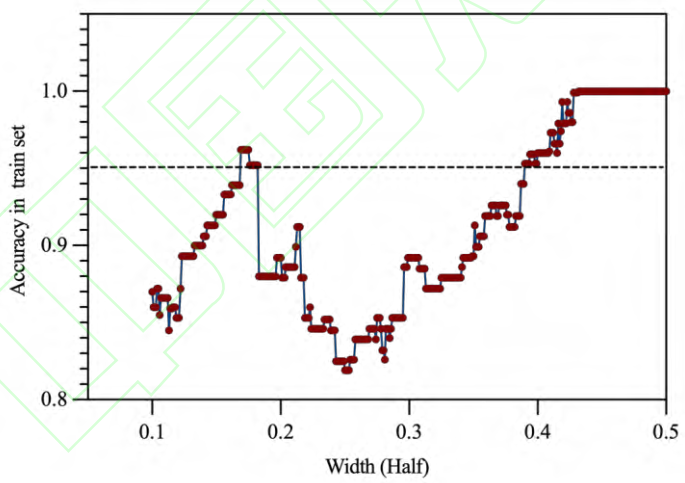

图 6 宽容区间训练过程图

Figure6 Tolerance interval training process diagram

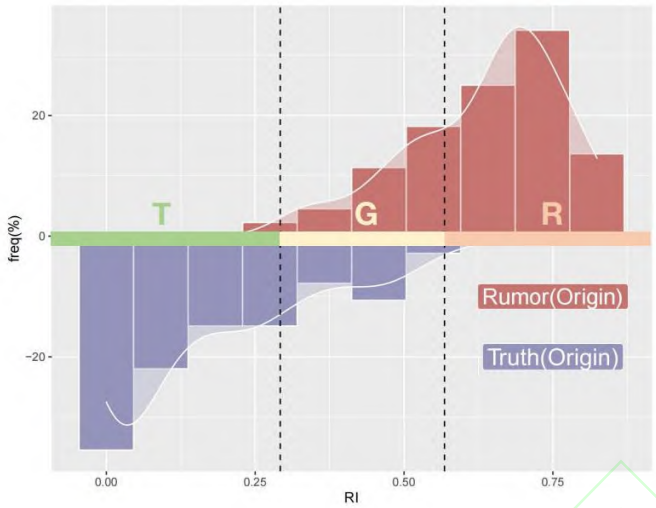

图 7 谣言指数 RI 分布图

Figure 7 Map of rumor index distribution

表 1 谣言多分类区间

Table1 Multiple classification intervals for rumors

RI	可信度分级	字母表示
<0.398	高	T
0.398-0.699	中	G
>0.699	低	R

3.3 以“新冠疫情”为例的三分类结果及分析

笔者利用优化的网络舆情可信度模型对测试集中 470 个样本进行分类，共计 362 个样本被判定为 T，84 个样本被判定为 G，24 个样本被判定为 R。符合实际生活中，正常信息量通常远大于谣言信息量的特征[39]。

其中，T 中的样本主要为官方对新冠疫情相关事件的报道、通知及科普等，部分为个人用户或经认证的网站或媒体等对新冠的客观见解或科普，经人工核验均为真实信息。

R 中的样本主要与疫情发展现状以及新冠病毒危害性有关，包括其症状、死亡率和后遗症等，大多以群众的偏激猜测和伪专家的不合理讲解等形式呈现，如“新冠死亡率低于流感！”“得了新冠到底会怎样？会马上好”“病毒专家：血氧饱和度是重要自测指标”等，此类言论与科学常识相悖，甚至被官方辟谣。

G 中的样本构成较为复杂，笔者参考国家互联网辟谣平台和国际权威研究成果进行人工核实将其分为以下三类，并选取典型样本，如表 2 所示：（1）对事实的模糊、绝对化或篡改：如样本①，虽然华泰证券曾发出该报告，但原报告的研究对象为新变异毒株 BA.5，范围限制在部分亚太地区，且病死率低于 0.1%的情况仅出现在其中的部分地区，显然不具有代表性，而样本①的文本模糊了大量细节，有误导大众相信新冠死亡率普遍低于流感之嫌；样本②中官方给出是关于“long covid”的统计结果，其定义与样本提到的“后遗症”并不一致；样本③提到的相关报道是新冠疫情带来的失业，而非“大量劳动力因病丧失劳动能力”。（2）基于个人经历的主观臆测或不合理建议：如样本④⑤⑥根据主观猜测误导公众将新冠当作感冒，若大量传播可能降低群众对新冠病毒的防范意识，而广东省新冠肺炎中医药防治专家组组长张忠德指出，奥密克戎只是症状与流感相似，本质上不能等同于流感或普通感冒。且有研究表明在政策放开后的几个月里，COVID-19 可能会导致大约 100 万人死亡，此外，新冠

病毒可能带来急性和长期神经功能障碍，疲劳、呼吸短促、头痛和嗅/味觉障碍等症状可能持续数月，给正常生活带来诸多不便[40-41]。（3）负面情绪发泄：主要体现在对各阶段防疫政策的不满，如样本⑦⑧，在新冠病毒流行毒株危害性较大，需要进行严格管控时，表示对封控的不满；样本⑨在流行毒株危害性降低，可以适当放开，解决医疗资源挤兑等问题时，出于对病毒的恐惧，对政策进行讽刺。这三类言论虽不及谣言那样颠倒是非、无中生有，但所表达的内容及方式可能引起群众的恐慌、愤怒和过度松懈等，影响国家的宏观调控，存在一定的误导性与危害性。

表 2 G 中样本摘选

Table 2 A representative sample of G

序号	文本	RI
①	华泰证券：新冠海外新变种死亡率或已低于流感！好事，希望以后就等于感冒一样。	0.552
②	美国 CDC（疾病控制与预防中心）做了 long covid 的统计，新冠 1/5 的人有后遗症	0.611
③	日经新闻 9 月 27 日报道，新冠疫情下，日本的大量劳动力因病丧失劳动能力，在此情况下，以餐饮和住宿服务等非制造业为中心，企业的人手短缺现象加剧	0.425
④	几年后得新冠就是感冒了，复阳就是又感冒了，不小心死掉就是抵抗力太差，小老百姓命不好而已总比因为居家隔离小病熬成大病，最后活活拖死强	0.675
⑤	国家都放开了，个人还在坚持什么呢，与其提心吊不如直接阳了，心理放下对新冠的恐惧才是疫情的结束，早点阳早点还自己自由	0.565
⑥	事实证明！新冠病毒就是感冒产生的一种新型细菌，和感冒的症状一模一样，吃吃感冒药就行了！	0.440
⑦	如此“封城”三个月有多大意义，瞎折腾	0.422
⑧	有本事别让老师出校门啊，怎么老师身体免疫新冠？学生出校门就感染呗	0.538
⑨	现在已经躺成这样，还能怎么躺？哦，入境改成 0+0 吧～～放 600 多种毒株进来，大乱炖一下，也许就能彻底免疫了呢	0.496

3.4 实验结果

为了验证本文提出的模型的有效性，我们使用常见的评价指标——准确率（acc）、精确率（pre）、召回率（rec）和 F1 值对其进行了评估，并将实验结果与决策树和支持向量机两个基线模型在同一数据集上的结果进行了对比，结果如表 3 所示：

表 3 模型性能验证

Table 3 Model performance verification

评价指标	决策树	支持向量机	本模型
准确率(Accuracy)	85.4%	82.9%	90.2%
精确率(Precision)	78.69%	94.60%	94.44%
召回率(Recall)	79.86%	88.77%	98.77%
F1	79.26%	91.36%	96.34%

从表 3 可以看出，笔者提出的模型在所有评价指标上的表现都优于决策树和支持向量机两个基线模型。具体来说，相较于决策树和支持向量机，本模型在准确率、精确率和召回率上表现更好，精确率略低于支持向量机。此外，综合评价指标 F1 值也显示本模型的表现明显超过决策树和支持向量机。因此，本文提出的模型在测试集上表现出了良好的性能。

4 结语

笔者基于多维度指标特征，通过计算“谣言指数 RI”将样本的可信度量化，实现将识别结果分为三级可信度，即高 T、中 G、低 R。并以新冠疫情为例，从真实性未知的各项性能指标均优于基线模型。

本研究创新性主要表现在：将信息可信度评估方法引入网络谣言文本识别研究，生成 RI 指数；基于 RI 指数，结合贝叶斯分类方法，确定宽容区间，实现对谣言识别结果的可信度分级；基于大量文献调研，笔者创新引入“当日新增确诊数”指标提高模型性能指标。

疫情实时动态情况对谣言的产生与传播有非常重要的影响，目前笔者仅纳入“当日新增确诊数”这一指标，后续将考虑增加如死亡率、康复率等其他有关疫情发展的动态指标。另外，随着人工智能技术的发展、GPT 技术应用的逐步普及，不可避免地会增加信息迷雾，从而加剧谣言的产生，所以笔者将进一步思考细化信息可信度评价指标体系粒度，从而进一步提高模型的科学性、准确性和适用性。

参考文献

1 中国互联网信息中心.CNNIC 发布第 51 次《中国互联网络发展状况统计报告》[EB/OL].[2023-03-02].https://www.cnnic.net.cn/n4/2023/0302/c199-10755.html.

2 强子珊,顾益军.基于多模态异质图的社交媒体谣言检测模型[J/OL].数据分析与知识发现:1-16[2023-04-02].http://kns.cnki.net/kcms/detail/10.1478.G2.20230207.0905.001.html.

3 张桂蓉,董志香,夏霆.突发公共卫生事件网络谣言网格化预警模型研究[J/OL].中国管理科学:1-13.[2023-04-06].DOI:10.16381/j.cnki.issn1003-207x.2021.1501.

4 侯晓艳,邓朝华,周利琴.重大突发公共卫生事件中的网络虚假信息识别与治理[J].中华医学图书情报杂志,2021,30(8):9-14.

5 卢建平.疫情灾难下谣言的传播与治理[EB/OL].[2022-09-28].https://www.spp.gov.cn/spp/llyj/202002/t20200205_453778.shtml.

6 牟冬梅,靳春妍,邵琦.基于情感分析的突发公共卫生事件网络舆情热度预测模型仿真[J].现代情报,2021,41(10):59-66.

7 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):6.

8 李保珍,王亚.社交媒体环境下网络信息可信度评估研究综述[J].情报学报,2015,34(12):1314-1321.

9 宋士杰,赵宇翔,朱庆华.iField 视域下的信息可信度研究:概念溯源、主题演化与未来展望[J].中国图书馆学报,2022,48(1):107-126.

10 徐静,杨小平,柳增.基于内容信任的 Web 信息可信度验证方法研究[J].北京理工大学学报,2014,34(7):710-715.

11 Castillo C, Mendoza M, Poblete B. Information credibility on Twitter[C]//Proceedings of the 20th international conference on World wide web.ACM,2011:675-684.

12 ODonovan J, Kang B, Meyer G, et al. Credibility in Context: An Analysis of Feature Distributions in Twitter[J]. International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing,2012,545(3):293-301.

13 谢娟,李文文,沈鸿权,等.信息爆炸和信息不确定语境下的可信度判据研究——以 COVID-19 疫情为例[J].情报学报,2021,40(7):714-724.

14 曹树金,常倞玮.社交媒体中的突发公共卫生事件信息可信度影响因素研究——以微信为例[J].现代情报,2020,40(9):3-14.
15 曾祥敏,张子璇.场域重构与主流再塑:疫情中的用户媒介信息接触、认知与传播[J].现代传播(中国传媒大学学报),2020,42(5):65-74,83.
16 何鸿雁,韩鸿.突发公共卫生事件中的微信谣言传播与第三人效果影响研究[J].传媒观察,2021(4):83-92.
17 Bora K, Das D, Barman B, et al. Are internet videos useful sources of information during global public health emergencies? A case study of YouTube videos during the 2015-16 Zika virus pandemic[J]. Pathog Glob Health,2018,112(6):320-328.
18 Thomson R, Ito N, Suda H, et al. Trusting tweets: The Fukushima disater and information source credibility on Twitter[C]//Proceedings of the 9th International ISCRAM Conference,2012:1-10.
19 滕婕,夏志杰,罗梦莹,等.基于 Multi-Agent 的网络谣言传播事件中信息主体信任识别研究[J].情报杂志,2020,39(3):105-114.
20 曾子明,张瑜.基于数据增强和多任务学习的突发公共卫生事件谣言识别研究[J/OL].数据分析与知识发现:1-16[2023-04-22].http://kns.cnki.net/kcms/detail/10.1478.g2.20230208.1132.001.html.
21 Wu S, Liu Q, Liu Y, et al. Information credibility evaluation on social media[C]//Proceedings of the 13th AAAI Conference on Artificial Intelligence,2016, DOI:10.48550/arXiv.1609.09226.
22 Liu X, Nourbakhsh A, Li Q, et al. Real-time Rumor Debunking on Twitter[JC]//International Conference on Information and Knowledge Management, 2015：1867-1870.
23 孙冉,安璐.突发公共卫生事件中谣言识别研究[J].情报资料工作,2021,42(5):42-49.
24 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.
25 石锴文,刘勘.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.
26 Guo L. Social Network Rumor Recognition Based on Enhanced Naive Bayes[J]. Journal of New Media,2021(3):99-107.
27 Suthanthira D P, Karthika S. Rumor Identification and Verification for Text in Social Media Content[J]. The Computer Journal,2021,65(2):436-455.
28 Srinivasan S, Dhinesh B. A Parallel Neural Network Approach for Faster Rumor Identification in Online Social Networks[J]. International Journal on Semantic Web and Information Systems (IJSWIS),2019,15(4):69-89.
29 Yuan Y, Wang Y L, Liu K. Perceiving more truth: A dilated-block-based convolutional network for rumor identification[J]. Information Sciences,2021,569(8):746-765.
30 姚艾昕,马捷,林英,等.重大突发公共卫生事件谣言演化与治理策略研究[J].情报科学,2020,38(7):22-29.
31 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.
32 首欢容,邓淑卿,徐健.基于情感分析的网络谣言识别方法[J].数据分析与知识现,2017,1(7):44-51.
33 王莉.网络虚假信息检测技术研究与展望[J].太原理工大学学报,2022,53(3):397-404.
34 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.
35 Sun S Y, Liu H Y, He J, et al. Detecting event rumors on Sina Weibo automatically

[C].AP Web,2013,7808:120-131.
36 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].小型微型计算机系统,2022,43(12):2518-2527.
37 林平,何思奇,段尧清.数据与用户视角下政府开放数据服务水平评价研究[J].图书情报工作,2020,64(2):23-29.
38 黄茜茜,杨建林.近十年我国图书情报学领域网络舆情研究方法应用分析[J].现代情报,2022,42(7):167-177.
39 王繁,郭军军,余正涛.融合评论的多任务联合谣言检测方法[J].计算机工程与科学,2022,44(9):1702-1710.
40 Etter M M, Martins T A, Kulsvehagen L, et al. Severe Neuro-COVID is associated with peripheral immune signatures, autoimmunity and neurodegeneration: a prospective cross-sectional study[J]. Nat Commun, 2022,13(1):1-21.
41 Ishii S, Sugiyama A, Ito N, et al. The role of discrimination in the relation between COVID-19 sequelae, psychological distress, and work impairment in COVID-19 survivors[J]. Sci Rep,2022,DOI：10.1038/s41598-022-26332-6.

Rumor Identification of Public Health Emergency Based on Information Credibility Assessment

QIAN Danmin[1,2,3], ZHENG Jianming[3], WANG Wenjing[2], MA Yeqing[1]

（1. Business School, Nanjing University, Nanjing 210008, China;
2. School of Medicine, Nantong University, Nantong 226000, China;
3.School of Information Management, Nanjing University, Nanjing 210023, China）

Abstract: 【Purpose/significance】With the continuous evolution of public health emergencies, people's understanding of them has a process from vagueness to accuracy. The author quantified and graded the credibility of online information of public health emergencies to provide data support for more detailed rumor identification. 【Method/process】Five static features of the information text, namely keywords, emotion, comment, information source and media, and two dynamic features, namely time and the number of newly confirmed cases on the same day, were selected, the rumor index RI was quantified with the entropy method. Based on this, the "tolerance interval" was introduced, and the boundary was determined by the naive Bayes classifier. The reliability of rumor identification results was divided into three categories: low, medium and high. 【Result/conclusion】The model performed well in the training set and the verification set, with the accuracy of 95% and 90.20%, respectively. Compared with the decision tree and SVM baseline models, the model's performance indexes were significantly improved. 【Innovation/limitation】In this study, a rumor identification model based on information credibility evaluation was established. RI index was established and dynamic indicators related to the epidemic situation were innovatively incorporated to improve the accuracy of rumor identification. The credibility classification of rumor is introduced to break through the limitation of the traditional binary classification detection method of rumor identification.

Keywords: public health emergency; rumor recognition; online public opinion; credibility evaluation; rumor index; entropy evaluation method