● 金 燕[1,2]，徐何贤[1]，毕崇武[1,2,3]
（1. 郑州大学信息管理学院，河南 郑州 450001；2. 郑州市数据科学研究中心，河南 郑州 450001；3. 郑州大学政治与公共管理学院，河南 郑州 450001）

多维特征融合的虚假健康信息识别方法研究：基于LightGBM算法*

摘 要：［目的/意义］为解决虚假健康信息自动识别效率低、准确度不高的问题，提出基于多特征融合的虚假健康信息识别方法。［方法/过程］首先，从内容特征、情感特征、发布者特征3个维度构建虚假健康信息特征指标体系；其次，分别采取不同方法进行特征提取，并转换成可处理的结构化数据；再次，基于LightGBM分类模型融合多特征属性，实现虚假健康信息自动识别；最后，以微信公众号上的健康信息为例进行实证验证。［结果/结论］该方法在微信公众号数据集实验的准确率达到92.22%，判别效果优于基于内容、情感、发布者等单维特征的识别方法，能够在一定程度上解决人工识别存在的及时性差、效率低、数量有限等问题，能够更全面、更接近人工识别准确率地实现虚假健康信息自动化识别。

关键词：多维特征；特征融合；虚假健康信息；LightGBM；识别方法；信息治理
DOI：10.16353/j.cnki.1000-7490.2023.08.019
引用格式：金燕，徐何贤，毕崇武．多维特征融合的虚假健康信息识别方法研究：基于LightGBM算法［J］．情报理论与实践，2023，46（8）：156-164.

Research on Health Misinformation Identification Method Based on Multi-dimensional Feature Fusion: Based on LightGBM Algorithm

Abstract: ［Purpose/significance］In order to solve the problem of low efficiency and low accuracy of automatic identification of health misinformation, a health misinformation identification method based on multi-dimensional features fusion is proposed. ［Method/process］Firstly, the health misinformation characteristic indicator system is constructed from three dimensions: content characteristics, emotional characteristics, and publisher characteristics. Secondly, different methods are adopted to extract features and convert features into structured data that can be processed. Then, based on the LightGBM classification model, multi-dimensional feature attributes are fused to realize the automatic identification of health misinformation. Finally, take the health information on the WeChat public account as an example for empirical verification. ［Result/conclusion］The accuracy rate of this method in the WeChat public account dataset experiment reaches 92.22%, and the discrimination effect is better than the identification method based on single-dimensional features such as content, emotion, and publisher, which can solve the problems of poor timeliness, low efficiency and limited number of manual identification to a certain extent, and realize the automatic identification of health misinformation more comprehensively and closer to the accuracy of manual identification.

Keywords: multi-dimensional features; feature fusion; health misinformation; LightGBM; identification method; information governance

0 引言

随着社交网络的迅猛发展，网上健康信息数量激增。网络信息传播迅速、匿名性与真实性并存、虚拟社会关系网络与真实社会关系网络混杂，加上普通公众的健康素养、科学素养有限，以及辟谣、监管的滞后等因素，导致虚假健康信息滋生并以难以想象的速度传播、扩散，给人们的健康信息选择、利用与健康决策带来了不利影响。目前，对健康信息进行治理主要就是对虚假健康信息的治理，即识别虚假健康信息并对其进行整治。虽然当前已经有很多针对虚假健康信息识别的研究，但在实践中，虚假健康信息识别与治理的效果依然不甚理想。虚假信息识别主要有自动识别和人工识别两种方式，但现有自动识别方法大多是基于单维度特征进行的，针对健康类虚假信息很难达到理想的识别效果[1]，准确度、全面性有待提高；人

* 本文为国家社会科学基金一般项目“群体参与视角下在线健康信息质量治理研究”的成果之一，项目编号：21BTQ054。

工识别准确率较高，但需要耗费大量的人力，其及时程度、效率、能够处理的数量都受到限制，无法应对互联网上层出不穷的虚假健康信息。因此，迫切需要效率更高、准确度更强的虚假健康信息自动识别方法。

鉴于此，本文融合“内容—情感—发布者”多维特征，利用 LightGBM 算法，构建虚假健康信息识别模型，并将其应用于微信公众号上发布的健康类信息验证其识别效果。同时对内容、情感、发布者 3 个维度的特征进行不同组合评估模型效果，展示每个维度及其属性特征对虚假健康信息识别的贡献。

1 概念界定及相关研究

1.1 虚假健康信息及其特征

虚假健康信息又称“伪健康信息”“健康谣言”等，通常是错误的、不准确的、不真实的健康信息。目前，关于虚假健康信息尚无完全统一的定义，但定义视角却是相似的，即从“虚假”的角度出发进行概念界定。邓胜利等[2]认为虚假健康信息是存在不准确、有误导性、叙述不当或是完全捏造等问题的虚假信息中关于健康的知识、技术、技能、观念和行为的信息。侯筱蓉等[3]认为虚假健康信息是不真实的健康信息，其有两个特点：一是虚，即信息本身没有事实依据，没有任何证据能证明其为真；二是假，即能被证明该信息确实不真。A. Ghenai[4]认为虚假健康信息是与当前已建立的医学理解相矛盾的错误信息。李月琳等[5]认为伪健康信息是指健康信息中的谣言、迷信、伪科学等各种无用甚至是有害的信息。张帅[6]将虚假健康信息定义为缺乏科学证据和专家意见支持的错误健康信息，该定义可以用来描述社交媒体上不同类型的错误健康信息，如谣言、失真健康信息、伪健康信息及其他变体。

总结前人研究成果发现，虚假健康信息是指已被证实为错误的、不真实的、不准确的健康信息，通常缺乏可信度、准确性、合理性和相关支持[7]；在语言、主题、情感、行为、形式、来源、语义、用户等方面具有一定特征[6,8-9]，如来源上存在信源模糊、假借权威，形式上存在元数据缺失、格式混乱[6]；在内容、语义上存在无中生有、捏造数据、偷换概念、夸大其词、暗度陈仓、狐假虎威、断章取义、术语包装等现象[10]；喜欢用“科学”包装自己，指向不明，情感上存在诱导性、宣传性语言[11]等。这些研究成果采用文献计量和内容分析方法，从文本的形式特征、内容特征、情感倾向、用户行为等方面研究了虚假健康信息特征，为虚假健康信息的识别奠定了理论基础。

1.2 基于单特征维度的虚假信息识别方法

当前，虚假信息的识别主要有人工识别和机器自动识别两大类。人工识别由平台选取的行业专家对领域内相关信息进行真实性鉴定[12]，识别精度高、准确度强，但识别效率低、时间和人工成本高。如“中国互联网联合辟谣平台”“腾讯较真辟谣”“微信辟谣助手”等辟谣平台，通常是接受群众举报，然后由行业专家对其真实性进行人工核实，最后对举报信息进行辟谣或发布相关科普知识[13]以正视听。机器自动识别多是在分析内容特征、情感特征、发布者特征的基础上，应用朴素贝叶斯、支持向量机等传统分类方法，或深度学习的文本分类方法实现。

1）基于内容特征维度识别，即通过分析文本的语义信息对虚假信息进行识别。於张闲等[1]采用基于词向量的深度神经网络模型和基于双向编码的语言表征模型，对健康信息文本进行自动分类，识别虚假健康信息，模型准确率达 88.1%。也有研究者针对信息的内容特征，采用支持向量机、朴素贝叶斯、决策树算法和逻辑回归模型对虚假评论、虚假新闻进行识别，收到较好的识别效果[14-16]。

2）基于情感特征维度识别，即通过分析文本中表现出的情感倾向对虚假信息进行识别。周娅等[17]基于程度副词与情感词之间的依赖关系计算出评论的情感极性，提出一种基于 XGBoost 算法的虚假评论识别方法。陈燕方等[18]提出一种基于评论产品属性情感倾向评估的模型，用于在线商品虚假评论的识别。

3）基于信息发布者特征维度识别，即通过分析信息发布者的社交行为、关系特征、影响力和信誉，识别出不可靠的信息发布者从而实现对虚假信息的识别。张小旭[19]对垃圾评论用户群组、评论用户和产品三者的基础特征和关系特征进行提取，使用迭代方式通过 GroupRank 算法识别出垃圾评论用户。B. Rodrigo 等[20]根据用户信息及其在社交网络中的行为方式，提取出用户的基本属性、社交属性、评论特征和可信度特征，使用逻辑回归模型、决策树模型等进行虚假评论分类。M. A. Baker 等[21]发现信息发布者的可靠性影响其创作内容的可靠性，可以基于此特征进行不可靠信息的识别。

总体说来，上述虚假信息自动识别方法，虽然在某些方面取得了较好的效果，但仍存在以下问题：①大多集中在对虚假信息的识别，针对健康类虚假信息识别的研究较少。②识别维度不够全面，侧重于信息的单一或部分维度特征。鉴于虚假健康信息在形式、内容、来源、情感等特征上具有隐蔽性，识别起来比较困难，加之单一维度的识别具有不全面等问题，本文试图综合虚假健康信息的多维度特征，构建一种多维度融合的自动识别算法，改进虚假健康信息识别的低效率、低准确率等问题。

2 多维特征融合的虚假健康信息识别模型

多维特征融合的虚假健康信息识别模型包括 3 个模

块，见图1。从图1可知，多维特征体系构建模块主要通过指标识别，筛选出符合研究标准的虚假健康信息特征指标；多维属性特征提取模块主要实现内容特征值、情感特征值、发布者特征值的提取、转换和计算；虚假健康信息判定模块主要将计算出的特征值输入LightGBM模型，实现对健康信息虚假与否的判断。

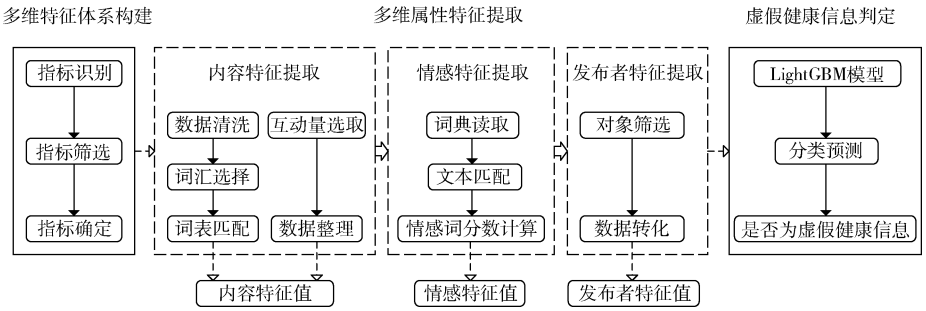

图1 多维特征融合的虚假健康信息识别框架

Fig. 1 Health misinformation identification framework based on multidimensional features fusion

2.1 虚假健康信息多维特征指标体系的构建

如前所述，虚假健康信息在内容、情感、发布者维度上具有内容与发布者不可靠、情感倾向明显等特征。根据相关研究，本文构建了虚假健康信息的多维特征指标体系，见表1。该指标体系基于已有文献并考虑指标是否容易提取等因素建立，可能不够完善，但可以在内容、情感、发布者等维度建立起相对可用的“虚假”特征指标。

表1 虚假健康信息多维特征指标

Tab. 1 Multidimensional characteristic indicators of health misinformation

一级指标	二级指标	具体描述	文献来源
内容特征	缺乏相关支持	没有确切来源的数字和统计；需要却缺少源文档来源；在其他资源中找不到可以印证的信息或知识	文献［7，10，22］
	缺乏准确性	观点片面或太绝对	文献［10］
	缺乏可信性	缺乏确证；捏造数据；无中生有；假借权威；信息老旧	文献［7，10-11，22］
	语义失衡	关键词特征；主题倾向性；受众明确性	文献［23-24］
	内容传播力弱	在看数少；阅读数少；点赞数少	文献［23，25-26］
情感特征	语气不当	信息诱导；夸大其词；术语包装；否定性信息	文献［7，10-11，23-24］
	缺乏客观性	主观性强；情感倾向极端	文献［18，27-28］
发布者特征	可靠性低	发布者信息完整度、披露程度低；粉丝数少	文献［20，29-31］
	社交值低	活跃度低；传播力度低；短时间发帖数量多	文献［20-21，31］
	缺乏权威性	信誉评分低；信任度低；专业水平低	文献［20，29-30，32］

2.2 虚假健康信息多维特征值的提取

特征值的提取就是把健康信息的虚假表征转换为可以计算的数值，作为识别模型的输入。转换方法见表2。

2.2.1 内容特征值提取方法 内容特征表示信息传播力和文本的语义信息，用缺乏相关支持、缺乏准确性、缺乏可信性、语义失衡和内容传播力弱指标进行表示。前4个指标采用关键词表匹配的方法进行其特征值计算，最终转换为关键词覆盖率。具体做法是：①对文本进行数据清洗，去除其中的噪声数据；②分别对虚假和真实的语料进行词频统计并降序排列，将虚假与真实语料中的词汇进行比对；③选择出在虚假语料中存在且次数较多但在真实语料中不存在或次数较少的相关词汇，作为关键词加入关键词表中。考虑到题目和正文的差异性，本文分别建立标题和正文两个关键词表。后一个指标不需要进行计算，可直接提取出信息的互动量。

表2 虚假健康信息特征指标——数据转换表

Tab. 2 Health misinformation characteristic indicators: data transformation tables

特征维度	虚假表征	测度对象
内容特征	缺乏相关支持	标题中关键词覆盖率；正文中关键词覆盖率
	缺乏准确性	
	缺乏可信性	
	语义失衡	
	内容传播力弱	信息互动量
情感特征	语气不当	情感得分
	缺乏客观性	
发布者特征	可靠性低	粉丝数；领域类别
	社交值低	传播能力
	缺乏权威性	认证状态；账号主体类型

2.2.2 情感特征值提取方法 情感特征指健康信息中表现出的情感倾向。本文提取语气不当、缺乏客观性两个指标来进行情感倾向的判定，并将其转换为文本内容的情感得分。

实验采用基于情感词典的方法计算情感极性，并引入了程度副词和否定词词典来提高准确率。选择包含情感词汇相对较多的大连理工大学中文情感词汇本体库[33]作为本文的情感词典。该词典对每个词汇的情感极性都进行了强度值的标注[34]。每个词在每一类情感下都对应了一个极性，为便于计算机作情感计算，用0、1、-1分别代表中性、褒义、贬义；情感词的初始情感强度被设置为1、

3、5、7、9 共 5 个等级。情感词的情感值计算如公式（1）所示[35]。

$$s(w) = v(w)p(w) \tag{1}$$

式中，$s(w)$ 代表词汇的情感值；$v(w)$ 代表词汇的情感强度；$p(w)$ 代表词汇的情感极性。

文本与词典匹配提取出否定词和程度副词，设定情感词初始权重为 W，否定词的权重为 -1，程度副词有不同的权重 W'，代表不同级别的情感倾向。根据句中情感词间是否有否定词或程度副词，以及否定词和程度副词的前后顺序，进行情感词的分值计算，情感词分值计算方法见表 3。情感词分数累加得出信息的整体情感得分，最终得分作为特征值输入分类模型进行健康信息分类。

表 3　情感词分值计算

Tab. 3　Score calculation for emotional words

位置关系	情感词得分
情感词	$s(w)$
否定词 + 情感词	$s(w) \times W \times (-1)$
程度副词 + 情感词	$s(w) \times W \times W'$
否定词 + 程度副词 + 情感词	$s(w) \times [W \times (-1) + W']$
程度副词 + 否定词 + 情感词	$s(w) \times W \times (-1) \times W'$

2.2.3　发布者特征值提取方法　信息发布者特征反映信息发布者的社交特征、影响力和信誉。发布者特征属性包括可靠性、社交值、权威性 3 个指标。可靠性指标转换为发布主体的粉丝数及领域类别；社交值指标转换为发布主体的传播能力；权威性指标转换为发布主体的平台认证状态及账号主体类型。直接获取平台提供的发布者基本信息数据，对于无法直接获取的数据，可通过权威的第三方平台而获取。

2.3　基于 LightGBM 模型的虚假健康信息判定方法

本文实现虚假健康信息判定的基本原理是对健康信息进行真假分类，借助 Boosting 集成分类方法进行。该方法不仅可以将多个单一分类器得到的分类信息进行综合提高分类的精度[36]，还可将弱学习器提升为强学习器[37]，代表模型有 XGBoost[38]、LightGBM 等。LightGBM 算法在 XGBoost 的基础上优化了基于直方图的决策树算法、Leaf 的叶子生长策略、Cache 命中率优化、多线程优化等[39]，速度更快、占用内存更低，可以更加快速地处理海量数据[40]。因此，选择更具优势的 LightGBM 模型进行分类识别虚假健康信息，并选择同类型的 XGBoost 算法进行比较。

LightGBM 模型的基本思想是通过 M 棵弱回归树线性组合为强回归树，如公式（2）所示：

$$F(x) = \sum_{m=1}^{M} f_m(x) \tag{2}$$

式中，$F(x)$ 为最终的输出值；$f_m(x)$ 为第 m 棵弱回归树的输出值。LightGBM 的创新在于其在基于直方图的 GBDT 算法中引入了两个新技术：互斥特征捆绑技术 EFB（Exclusive Feature Bundling）和单边梯度采样技术 GOSS（Gradient-based One-Side Sampling）。EFB 能够在不损失准确性的前提下将一些特征进行融合绑定，以降低特征维度；GOSS 算法可保持信息增益估计的准确，在保证算法精度的同时又减少了样本数量[41]。

3　实证研究

3.1　数据获取与预处理

选取“微信辟谣助手”“腾讯较真辟谣”小程序和“新榜”平台为实验数据来源，人工获取健康类信息文本数据、信息传播数据和发布者数据。虚假健康信息数据来自以下两个数据源：①“腾讯较真辟谣”中标注为“虚假”的健康信息；②“微信辟谣助手”2021 年 6 月 22 日—2022 年 1 月 28 日已辟谣的健康类信息。真实健康信息数据来自“腾讯较真辟谣”中标注为“真实”的健康信息。“新榜”是权威的第三方公众号数据监测平台，拥有微信官方真实的公众号信息数据。本文统一参考“新榜”平台 2022 年 2 月 7 日提供的微信公众号相关数据，保障获取数据的可用性和有效性。将采集到的健康信息数据、信息传播数据和微信公众号数据一起作为实验数据，并从中选取真实、虚假健康信息各 150 条作为数据集。将数据集拆分为 70% 训练集和 30% 测试集，然后通过 LightGBM 分类方法从内容、情感和发布者的角度对标记的文本信息进行训练测试，以达到虚假健康信息自动识别的目的。

数据集的预处理步骤是：首先，运用 Jieba 分词工具进行分词处理，并通过停用词词典去除文本中的停用词以及图片、标点符号等非文本数据；其次，进行数据清洗，即对周新榜指数、信息点赞数和在看数数据中的缺失值进行补 0 处理；最后，进行数据转化，即将非数值型数据转换为数据对象：发布者领域类别为健康和非健康，分别赋值为 1 和 0，发布者的政府认证、媒体认证、企业认证、个人认证和其他认证 5 类账号主体从 5 到 1 依次赋值，将发布者账号中已经认证和未认证的分别赋值为 1 和 0。

3.2　参数设置与评价标准

1）参数设置。机器学习算法模型需要对参数进行调整以达到更好的训练与测试效果，LightGBM 模型和 XGBoost 模型的参数设置情况见表 4。

2）评价标准。实验采用准确率（Accuracy）、精确度（Precision）、召回率（Recall）、$F1$ 值（$F1$-score）作为评价指标。对于健康信息真假识别二分类问题，分类混淆矩阵见表 5。其中，TP 表示健康信息的真实标签为真且模型识别结果也为真的个数；FP 表示健康信息的真实标签为假而模型识别结果为真的个数；FN 表示健康信息的真实

表4 机器学习算法参数设置
Tab. 4 Parameter settings for the machine algorithm

模型	参数设置
XGBoost	learning_rate = 0.1，n_estimators = 181，max_depth = 1，min_child_weight = 1，gamma = 0，subsample = 0.75，colsample_bytree = 0.75，nthread = 4，scale_pos_weight = 1，seed = 27
LightGBM	learning_rate = 0.1，n_estimators = 190，max_depth = 5，num_leaves = 10，max_bin = 25，min_data_in_leaf = 1，bagging_fraction = 0.8，bagging_freq = 10，feature_fraction = 0.6，lambda_l1 = 1e-05，lambda_l2 = 0.1，min_split_gain = 0

表5 分类混淆矩阵
Tab. 5 Classification confusion matrix

预测结果	实际标签	
	真实（Positive）	虚假（Negative）
真实（Positive）	TP	FP
虚假（Negative）	FN	TN

标签为真而模型识别结果为假的个数；TN表示健康信息的真实标签为假且模型识别结果也为假的个数。

准确率A指信息真假识别正确的个数在所有识别个数中的占比，通常来说，准确率越高，分类效果越好，计算方法如公式（3）所示；精确度P表示被识别为真实的信息中实际确实为真实的比例，计算方法如公式（4）所示；召回率R又称为查全率，反映了真实类别是真实的样本被分类后仍被识别为真实的概率大小，计算方法如公式（5）所示；$F1$值是精确率和召回率的调和均值，当P和R都高时，$F1$也会高，计算方法如公式（6）所示。

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$P = \frac{TP}{TP + FP} \tag{4}$$

$$R = \frac{TP}{TP + FN} \tag{5}$$

$$F_1 = \frac{2 \times P \times R}{P + R} \tag{6}$$

3.3 虚假健康信息识别

3.3.1 多维属性特征提取

1）内容特征提取。本文根据前述的关键词选取方法，分别提取出标题和正文的关键词，并进行匹配计算，信息包含的关键词在关键词表中覆盖率的大小可反映信息的虚假程度。提取出的部分关键词详见表6。对于微信公众号发布信息的互动量，可直接通过微信提取阅读数、点赞数、在看数。

2）情感特征提取。根据前述方法，计算出每篇文章信息的情感得分数值，虚假和真实情感得分绝对值见图2。从图2可以看出，虚假信息的情感得分绝对值相对较高，反映出虚假信息内容的情感倾向往往较为明显，真实信息内容的情感倾向不明显，较为客观。

表6 关键词表（部分）
Tab. 6 Keyword list (partial)

标题关键词	医生；建议；癌症；很多；越来越；提醒；每天；天天；告诉；中医；排毒…
正文关键词	身体；作用；医生；食用；非常；每天；造成；血管；能够；癌症；预防…

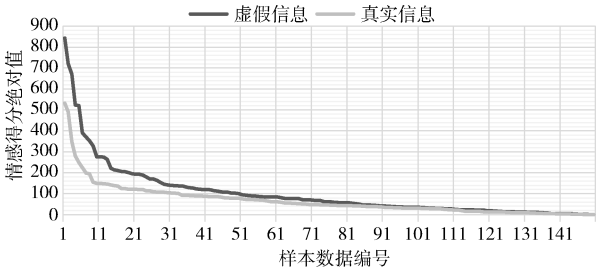

图2 情感得分绝对值分布
Fig. 2 A plot of the distribution of absolute values for sentiment scores

3）发布者特征提取。对于微信公众号，具体特征值提取过程是：①依据“新榜”对公众号进行的类别说明标注类别是否为健康相关。②依据“新榜”计算的公众号周新榜指数，获取公众号一周内的传播能力。③依据微信标注的公众号账号主体，获取其隶属的组织机构。通常而言，政府部门掌握的信息具有全面性、权威性，由政府出面获取一个事件的真实信息通常也是更加准确迅速[42]，而企业和个人偏商业化。基于此，本文对账号主体依据其可靠性从高到低排序依次为政府认证、媒体认证、企业认证、个人认证和其他认证。④参考微信提供的发布者平台认证信息，获取发布者的信用等级。鉴于微信平台认证需要一定的门槛，故本文认为已认证的账号比未认证的账号信用等级更高。

提取整理出的发布者特征值见表7。由表7可以看出，虚假和真实信息发布者的特征值存在较明显的差异。相比较来说，虚假信息发布者的账号主体相对缺乏权威，微信认证情况较差，粉丝数较少，类别通常为非健康类，周新榜指数较低，说明本文选取的发布者特征属性具有合理性。

3.3.2 基于LightGBM模型的虚假健康信息判定 以缺乏相关支持、缺乏准确性、缺乏可信性等10个指标为特征变量，以健康信息的真假标签为预测变量，构建LightGBM分类模型。按照表2所示的参数设置并训练模型，对测试集数据进行虚假健康信息判定的测试结果如下：准确率为92.22%，精确度为91.84%，召回率为93.75%，$F1$值为92.78%。实验得到的ROC曲线（接收者操作特征曲线），见图3。从图3可以看到，ROC曲线在随机分类器

表 7　发布者特征值表（部分）
Tab. 7　Publisher characteristic value table（partial）

公众号名称	账号主体类型	认证状态	粉丝数	领域类别	周新榜指数	信息标签
智慧人生	个人	否	4641	否	397. 1	
葛欣阅读写作	个人	否	4427	否	414. 0	虚假
医学资料全库	个人	否	121144	否	616. 3	
丁香医生	企业	是	1000000 +	是	992. 8	
人民网科普	媒体	是	842152	是	850. 7	真实
科普中国	政府	是	1000000 +	否	964. 8	

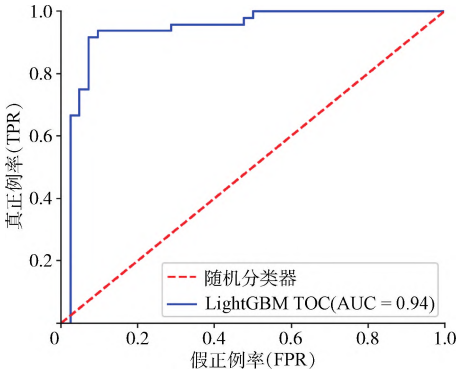

图 3　LightGBM 模型 ROC 曲线图
Fig. 3　ROC graph of the LightGBM model

曲线之上，接近左上角，曲线下的面积（AUC 值）为 0. 94，说明该 LightGBM 模型分类性能好。用 XGBoost 模型对相同的数据集进行同样处理，并将其得到的结果与 LightGBM 模型进行对比，见图 4。由图 4 可以看出，LightGBM 模型的准确率、精确度和 *F*1 值 3 个指标均优于 XGBoost 模型。

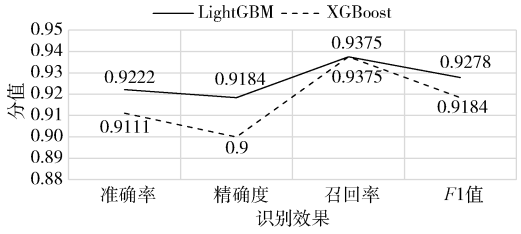

图 4　多特征融合的虚假健康信息识别效果
Fig. 4　Health misinformation identification effect based on multidimensional features fusion

3. 4　虚假健康信息识别结果分析

3. 4. 1　不同特征及其组合对虚假健康信息识别结果的影响　本文选取不同的特征组合进行特征消融实验，分析各个特征对虚假健康信息识别的影响。对比分析单个不同特征及其组合后模型对虚假健康信息识别效果的影响，其中 Fcontent 为内容特征，Femotion 为情感特征，Faccount 为发布者特征。具体实验结果详见表 8。

结果表明，仅使用内容特征时，识别的准确率和精确率分别是 0. 8000 和 0. 7778，表明采用关键词表匹配提取出的内容特征具有一定的有效性。将 2 维特征组合的实验结果与 3 维特征组合的实验结果对比发现，发布者特征对召回率的提高帮助最大，表明发布者特征更有助于提高微信中虚假健康类信息识别的全面性，能够减少将虚假健康信息识别为真实健康信息的概率。

表 8　不同特征及其组合的实验结果对比
Tab. 8　Comparison of experimental results of different features and their combinations

特征维度	准确率（%）	精确度（%）	召回率（%）	*F*1 值（%）
Fcontent	0. 8000	0. 7778	0. 8750	0. 8235
Femotion	0. 6111	0. 6444	0. 6042	0. 6237
Faccount	0. 8667	0. 8913	0. 8542	0. 8723
Femotion + Faccount	0. 8556	0. 8431	0. 8958	0. 8687
Fcontent + Faccount	0. 8889	0. 8654	0. 9375	0. 9000
Fcontent + Femotion	0. 7667	0. 7368	0. 8750	0. 8000
Fcontent + Femotion + Faccount	0. 9222	0. 9184	0. 9375	0. 9278

就单维度而言，发布者特征对模型效果的影响最为显著，说明发布者特征维度对健康信息的真假影响最大。其原因可从以下 3 个层面分析。①信誉层面，发布者的信誉是由其历史行为及其结果积累起来的[32]。发布者信誉低，意味着其存在发布虚假、低质量健康信息的行为，后续发布虚假健康信息的概率也相对较大；发布者信誉高，意味着其前期传播健康信息的质量较高、真实性有保障，后续发布真实健康信息的概率较大。②能力层面，健康信息对发布者的能力即专业性、科学素养、信息素养等的要求高。发布者能力弱，就难以判断健康信息的真假及适用性，容易创建、发布、传播虚假、不可靠的健康信息；发布者能力强，就能够对健康信息的质量、真实性进行把关，有助于创建、传播真实可靠的健康信息。③动机层面，动机是发布者行为产生的驱动力，并能影响发布行为的频率和持续性[43]。若发布者出于获取经济利益、吸引眼球等的外部动机，其发布、传播虚假、低质量健康信息的概率就会增加；若发布者出于利他主义的内部动机，如致力于为大众普及权威、科学的健康知识，则其创建、发布真实健康信息的概率就会增加。

基于全特征分类结果的准确率和精确率高于单一维度特征或两维度特征的组合很多，说明内容、情感和发布者 3 个维度特征中任一维度都不能单独决定健康信息的真假，需要结合多属性特征综合进行识别。如为了提高可信度，有的发布者会刻意加入一些专业词汇，造成“内容可靠”的假象，这会对普通用户产生较强的迷惑作用；信息内容使用情感色彩较浓的词汇并不代表不可靠，真实信息也存在采用吸引读者的方式宣传健康知识的情况；总体质量高的发布者并不能保证发布的每条信息都准确，也可能

存在考察不全、进行功利性宣传的情况，且信息真假具有不稳定性，可能本来真实的信息后来又被推翻。因此，在进行虚假健康信息识别时，对以上3个维度特征进行综合考虑是必要的。

3.4.2 不同特征指标对虚假健康信息识别的重要性分析 每个特征指标与健康信息真假之间的相关关系、关系密切程度不同。不同特征与数据标签即健康信息真假之间有不同的相关性，包括正相关和负相关关系，见图5。由图5可以看出，除了标题中关键词覆盖率、正文中关键词覆盖率和情感得分这3个特征指标与健康信息的真假性呈负相关外，其余指标均与其呈正相关关系。具体而言，阅读数、点赞数、在看数、账号主体类型等指标的值越高，健康信息的真实性就越高；相反，标题中关键词覆盖率、正文中关键词覆盖率和情感得分3个指标的值越高，健康信息的真实性就越低，即越有可能是虚假健康信息。

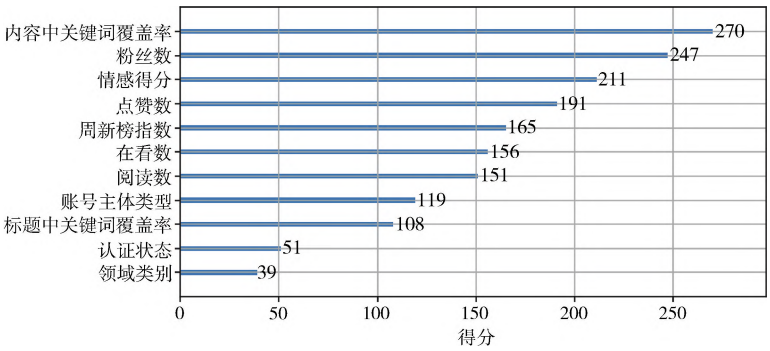

图6 不同特征对虚假健康信息判断的影响程度

Fig. 6 The influence of different characteristics on the judgment of health misinformation

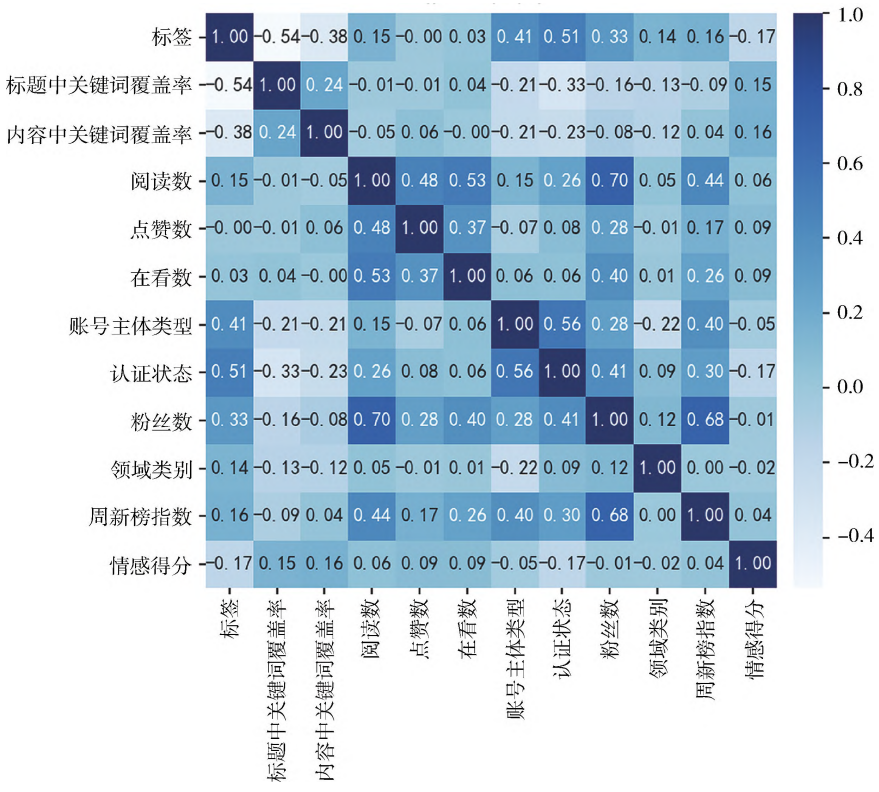

图5 不同特征与健康信息真假的相关性分析

Fig. 5 Correlation analysis of different characteristics and health information authenticity

每个特征指标对虚假健康信息识别的影响程度不同，见图6。由图6可以看出，内容特征维度中的正文覆盖率和点赞数对识别虚假健康信息的影响较大；情感特征维度中的情感得分重要性也很高；发布者特征维度中的粉丝数和周新榜指数最重要。虽然总体而言，信息发布者维度对虚假健康信息识别最重要，但是单个指标中正文中关键词覆盖率、情感得分和点赞数对虚假健康信息的识别贡献相对较大，也应重点关注。

对虚假健康信息识别贡献度较大的特征指标具有较好的参考性：正文中关键词覆盖率对虚假健康信息识别贡献最大，因为虚假性词汇从虚假类信息数据中提取，具有较强针对性，与虚假类信息的匹配度高；发布者粉丝数在很大程度上代表发布者的影响力，粉丝数多说明读者对其认可度高；信息的情感得分是内容情感倾向的表现，使用较多的情感词汇是虚假类信息吸引读者的一种方式；信息点赞数代表读者对内容的认可度，认可度高的信息较为可靠；周新榜指数用于衡量公众号的传播能力，数据客观权威，周新榜指数高的公众号发布的信息通常更可靠。

4 总结与展望

实验结果表明，就单维度而言，发布者特征对微信虚假健康信息识别的影响最大；就具体特征而言，正文中关键词覆盖率、情感得分等特征更有助于识别出虚假健康信息，粉丝数、点赞数、周新榜指数等特征更有助于识别出真实健康信息。实验结果还表明，融合多特征的虚假健康信息识别的准确率

和精确率较高，该方法在实验数据集上的准确率能达92.22%。这也充分说明单一维度或单一特征在判定健康信息真假方面存在一定缺陷，只有综合多维度属性特征，才能提高虚假健康信息识别的准确度。

本文方法不仅在一定程度上弥补了人工识别方法存在的及时性差、效率低、数量有限等缺陷，还优化了自动识别方法存在的特征维度片面性、准确率低的问题，实现了更全面、更接近人工识别准确率的自动化识别。本文的局限性是：实验数据量相对较少，且涉及的范围有限，所以根据其创建的关键词表较为局限，不够全面；设计的虚假健康信息的指标体系不够全面，且未充分考虑不同维度、不同指标的权重及其影响。未来可在现有数据的基础上，增加健康信息的种类和数量，并通过调整和完善虚假健康信息指标体系，优化识别方法。□

参考文献

[1] 於张闲，冒宇清，胡孔法．基于深度学习的虚假健康信息识别［J］．软件导刊，2020，19（3）：16-20.

[2] 邓胜利，孙瑾杰．图书馆参与虚假健康信息治理的价值、阻滞因素和实现路径［J］．图书情报工作，2022，66（9）：14-22.

[3] 侯筱蓉，付扬，陈娟．基于微信平台的健康信息用户感知和效用研究［J］．现代情报，2016，36（10）：89-93.

[4] GHENAI A. Health misinformation in search and social media［C］//Proceedings of the 2017 International Conference on Digital Health. ACM，2017：235-236.

[5] 李月琳，张秀．大学生社交媒体健康信息甄别能力研究［J］．图书情报知识，2018（1）：66-77，43.

[6] 张帅．社交媒体虚假健康信息特征识别［J］．图书情报工作，2021，65（9）：70-78.

[7] 李月琳，张秀，王姗姗．社交媒体健康信息质量研究：基于真伪健康信息特征的分析［J］．情报学报，2018，37（3）：294-304.

[8] 朱梦蝶，付少雄，郑德俊，等．文献视角下的社交媒体健康谣言研究：特征、传播与治理［J］．图书情报知识，2022，39（5）：131-143.

[9] ZHAO Yuehua，DA Jingwei，YAN Jiaqi. Detecting health misinformation in online health communities：Incorporating behavioral features into machine learning based approaches［J］. Information Processing and Management，2021，58（1）：102390.

[10] 曾祥敏，王孜．健康传播中的虚假信息扩散机制与网络治理研究［J］．现代传播（中国传媒大学学报），2019，41（6）：34-40.

[11] 郭泽萍．微信平台健康谣言的特征与治理思路——基于微信小程序“微信辟谣助手”的样本分析［J］．现代视听，2019（6）：44-48.

[12] 陈蕾，邹仪．面向微博平台的谣言识别技术［J］．科技与创新，2022（8）：156-159.

[13] 刘悦．健康类社交媒体不实信息检测方法研究与应用［D］．北京：北京邮电大学，2020.

[14] 邓莎莎，张朋柱，张晓燕，等．基于欺骗语言线索的虚假评论识别［J］．系统管理学报，2014，23（2）：263-270.

[15] JINDAL N，LIU Bing，LIM E P. Finding unusual review patterns using unexpected rules［C］//Proceedings of the 19th ACM International Conference on Information and Knowledge Management. Toronto. ACM，2010：1549-1552.

[16] AHMED H，TRAORE I，SAAD S. Detection of online fake news using N-gram analysis and machine learning techniques［C］//Proceedings of the International Conference on Intelligent，Secure，and Dependable Systems in Distributed and Cloud Environments. Cham：Springer，2017：127-138.

[17] 周娅，吴昱翰．基于HDXG算法的虚假评论识别方法［J］．计算机仿真，2020，37（1）：473-477.

[18] 陈燕方，李志宇．基于评论产品属性情感倾向评估的虚假评论识别研究［J］．现代图书情报技术，2014（9）：81-90.

[19] 张小旭．基于PageRank的垃圾评论用户群组检测［D］．杭州：浙江大学，2018.

[20] RODRIGO B，OSCAR A，CARLOS A I. A framework for fake review detection in online consumer electronics retailers［J］. Information Processing and Management，2019，56（4）：1234-1244.

[21] BAKER M A，MAGNINI V P，PERDUE R R. Opportunistic customer complaining：causes，consequences，and managerial alternatives［J］. International Journal of Hospitality Management，2012，31（1）：295-303.

[22] HARRIS R. Evaluating internet research sources［EB/OL］.［2021-10-27］. https://www.virtualsalt.com/evaluating-internet-research-sources/.

[23] 吴连伟，饶元，樊笑冰，等．网络传播信息内容的可信度研究进展［J］．中文信息学报，2018，32（2）：1-11，21.

[24] 沈威，聂卓，廖莉莉．网络谣言标题的特征研究——以健康养生类网络谣言标题为例［J］．四川文理学院学报，2020，30（4）：61-69.

[25] 张克永，李贺．健康微信公众平台信息质量评价指标体系研究［J］．情报科学，2017，35（11）：143-148，155.

[26] 王雯玉，田瑞．微信公众号健康养生类文章特点分析［J］．医学信息学杂志，2022，43（5）：51-55.

[27] 吴佳芬，马费成．产品虚假评论文本识别方法研究述评［J］．数据分析与知识发现，2019，3（9）：1-15.

[28] GUO Chuan，CAO Juan，ZHANG Xueyao，et al. DEAN：learning dual emotion for fake news detection on social media

[EB/OL]. [2020-11-02]. https://arxiv.org/abs/1903.01728v2.
[29] 余秋文. 微博信息可信度评价指标体系 [D]. 武汉：华中师范大学，2014.
[30] 孙茜甜. 在线评论内容特征及其效价的感知有用性研究 [D]. 济南：山东大学，2018.
[31] 李少愚. 微博虚假信息鉴别技术验证平台的设计与实现 [D]. 北京：北京邮电大学，2019.
[32] 金燕，闫婧. 基于用户信誉评级的UGC质量预判模型 [J]. 情报理论与实践，2016，39（3）：10-14.
[33] 徐琳宏，林鸿飞，潘宇，等. 情感词汇本体的构造 [J]. 情报学报，2008，27（2）：180-185.
[34] 徐善山. 基于关键句语义规则和领域情感词典的影评情感分析 [D]. 淮南：安徽理工大学，2020.
[35] 杨秀璋，郭明镇，侯红涛，等. 融合情感词典的改进BiLSTM-CNN+Attention情感分类算法 [J]. 科学技术与工程，2022，22（20）：8761-8770.
[36] 张雁，林英，吕丹桔. 集成学习在遥感分类中的应用 [J]. 计算机与数字工程，2013，41（5）：697-699.
[37] FREUND Y，SCHAPIRE R E. A decision-theoretic generalization of on-line learning and an application to boosting [J]. Journal of Computer and System Sciences，1997，55（1）：119-139.
[38] CHEN Tianqi，GUESTRIN C. XGBoost：A scalable tree boosting system [C] //Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM，2016：785-794.
[39] 王清楠. 基于旋转森林和LightGBM分类算法的高校实践教学数据分析 [D]. 长春：吉林大学，2020.
[40] 朱丽云. 基于LightGBM算法的个人信用风险评估研究 [D]. 广州：华南理工大学，2020.
[41] 曾子明，张瑜，李婷婷. 多特征融合的突发公共卫生事件潜在谣言传播者识别 [J]. 图书情报工作，2022，66（13）：80-90.
[42] 王泱. 社交媒体虚假信息及其辟谣信息探析——以2018年微博虚假信息为例 [J]. 新闻战线，2019（20）：113-118.
[43] 成俊会. 微博舆情传播中用户的参与行为研究 [D]. 哈尔滨：哈尔滨工程大学，2015.

作者简介：金燕，女，1977年生，博士，教授。研究方向：信息治理与健康信息学。**徐何贤**，女，1999年生，硕士生。研究方向：信息治理与健康信息学。**毕崇武**（通信作者，Email：767818984@qq.com），男，1994年生，博士，讲师。研究方向：文本挖掘与知识发现。

作者贡献声明：金燕，确定选题，提出研究思路，论文审定与修改。**徐何贤**，论文撰写与修改，数据收集与整理，实验设计与实现。**毕崇武**，论文框架设计，论文修订。

录用日期：2023－03－21

（上接第137页）

[18] 王哲，何飞. 高校网络舆情危机演化及管理机制研究 [J]. 西南民族大学学报（人文社会科学版），2022，43（8）：157-162.
[19] 朱斌，刘新成. 媒体融合环境下高校网络舆情传播引导模式研究 [J]. 理论视野，2020（11）：77-81.
[20] BERGE C. Graphs and hypergraphs [M]. New York：Elsevier，1973.
[21] NAGURNEY A，DONG J. Supernetworks：Decision-Making for the Information Age [M]. [S. l.]：Elgar，Edward Publishing，Incorporated，2002.
[22] 杨湘浩，阚顺玉，叶旭，等. 基于超网络的突发事件网络谣言传播模型研究 [J]. 情报理论与实践，2021，44（10）：129-136.
[23] 马宁，刘怡君. 基于超网络中超边排序算法的网络舆论领袖识别 [J]. 系统工程，2013，31（9）：1-10.
[24] 梁晓贺，田儒雅，吴蕾，等. 基于超网络的微博舆情主题挖掘方法 [J]. 情报理论与实践，2017，40（10）：100-105.
[25] 张连峰，周红磊，王丹，等. 基于超网络理论的微博舆情关键节点挖掘 [J]. 情报学报，2019，38（12）：1286-1296.
[26] 梁晓贺，田儒雅，吴蕾，等. 基于超网络的微博相似度及其在微博舆情主题发现中的应用 [J]. 图书情报工作，2020，64（11）：77-86.

作者简介：周欢，博士，副教授，硕士生导师。**张培颖**，硕士生。**王嘉仪**，本科生。**邹筱**（通信作者，Email：27887182@qq.com），博士，教授，硕士生导师。

作者贡献声明：周欢，研究指导，研究设计，研究方法的提出，论文初稿部分内容的撰写与修改，论文返修过程中的实验指导与修改。**张培颖**，数据采集，研究方法的优化，实验指导，实验结果分析，论文初稿修改，返修论文修改，论文修改详细说明的撰写。**王嘉仪**，数据采集，实验分析，论文初稿部分内容的撰写与修改。**邹筱**，研究指导，论文初稿修改指导，论文返修指导。

录用日期：2023－03－03