



开放科学  
(资源服务)  
标识码  
(OSID)

# 融合情感分析与多元时间序列的 区块链产业舆情监测研究

刘颖<sup>1,2,3</sup> 薛云龙<sup>1</sup>

- 吉林财经大学管理科学与信息工程学院 长春 130117;
- 吉林省金融科技重点实验室 长春 130117;
- 吉林省商务大数据研究中心 长春 130117

**摘要:** [目的/意义] 区块链技术被纳入“新基建”范畴后,其产业发展演进快、舆情热度高。本研究将情感因素纳入新兴产业网络舆情热度预测,探究区块链产业关注主题及发展态势。[方法/过程] 论文融合情感分析与多元时间序列特征提出舆情热度预测模型,采用 BERT-BiLSTM (Bi-directional Long Short-Term Memory, BiLSTM) 方法对舆情文本分类并赋值,挖掘情感极性类别的主题,将不同情感倾向的情感值分别取绝对值累加,构建基于情感因素的多元时间序列特征体系,并输入 LSTM (Long Short Term Memory, LSTM) 模型进行区块链产业舆情热度预测。[结果/结论] BERT-BiLSTM 在情感分类任务中准确率为 84%,其中消极和中性情感类属文本的成因主要为“对于区块链技术的信任”和“缺乏区块链相关概念的了解”。在热度预测模型中,模型均方根误差 (Root Mean Square Error, RMSE) 降低 17.67,平均绝对误差 (Mean Absolute Error, MAE) 降低 15.14,决定系数 (R-Square,  $R^2$ ) 提升 11%,模型总体性能良好。

**关键词:** 区块链; 产业舆情; 情感分析; 深度学习; 多元时间序列

**中图分类号:** G35; TP391

## Research on Public Opinion Monitoring of Blockchain Industry That Integrates Sentiment Analysis and Multiple Time Series

LIU Ying<sup>1,2,3</sup> XUE Yunlong<sup>1</sup>

- School of Management Science and Information Engineering, Jilin University of Finance and Economics, Changchun 130117, China;
- Jilin Provincial Key Laboratory of Fintech, Changchun 130117, China;
- Jilin Provincial Business Big Data Research Center, Changchun 130117, China

**基金项目** 国家社会科学基金项目“基于多源数据深度集成的供应链金融风险评估方法研究”(20BTJ062)。

**作者简介** 刘颖(1979-),博士,教授,研究方向为机器学习,文本挖掘, E-mail: lyaihua1995@163.com; 薛云龙(1997-),硕士研究生,研究方向为网络舆情分析,文本挖掘。

**引用格式** 刘颖,薛云龙.融合情感分析与多元时间序列的区块链产业舆情监测研究[J].情报工程,2023,9(1):3-14.

**Abstract:** [Objective/Significance] The blockchain technology has developed fast and attracted a lot of attentions after being included in the scope of “new capital construction”. The research integrated the emotional factor into the prediction for the degree of discussion of online public opinion about the emerging industry and explored the hot topics and development trends of the blockchain industry. [Method/Process] The research proposed the prediction model for the degree of discussion of online public opinion through combining the emotional analysis and the multivariate time series characteristic; made classification and assignment for the text of public opinions by the BiLSTM (Bi-directional Long Short-Term Memory) method, explored the topic under the category of sentiment polarity, cumulated the absolute values of the emotional value with different emotional tendency, created the multivariate time series characteristic system based on the emotional factor, and input the LSTM (Long Short Term Memory) module for predicting the degree of discussion of online public opinion about the blockchain industry.[Results/Conclusions] The empirical results revealed that the BERT-BiLSTM method could reach the accuracy rate of 84% in the emotion classification task, where the factors for negative and neutral emotion type text included “not trust in the blockchain technology” and “insufficient understanding for the blockchain related concept”. In the prediction model, the RMSE (Root Mean Square Error) decreased 17.67 and the MAE (Mean Absolute Error) decreased 15.14, while the R2 (R-Square) increased 11%, indicating good overall performance of the model.

**Keywords:** Blockchain; Industrial public opinion; Deep learning; Multivariate time series

## 引言

区块链是一种全新的去中心化基础架构与分布式计算范式，具有匿名性、防篡改以及高度可扩展等特点<sup>[1]</sup>。2020年4月20日，国家发改委将区块链技术纳入新技术基础设施，并与其他技术融合创新，推动“新基建”下的新型价值体系形成。区块链技术被划入“新基建”范畴后，其产业发展迅速。2021年3月11日，《关于国民经济和社会发展第十四个五年规划和2035年远景目标纲要》指出：“培育壮大人工智能、大数据、区块链、云计算、网络安全等新兴数字产业”“赋能传统产业转型升级，催生新产业新业态新模式，壮大经济发展新引擎”。规划首次将区块链纳入国家五年规划，将区块链产业列为“十四五”七大数字经济重点产业之一<sup>[2]</sup>。区块链产业在促进技术、资源

和市场跨时空、跨领域融合的同时，也对产业舆情监管提出了新要求、新挑战<sup>[3]</sup>。

舆情是指在一定的社会空间内，围绕中介性社会事件发生、变化以及民众所表达态度、意见和立场的总和，直观反映出社会情绪的变化<sup>[4]</sup>。新兴产业在产品研发与技术落地的过程中关注度较高，容易引发舆情事件，缺少必要的舆情监管可能导致产业发展受阻。概括原因可能包括：（1）新兴产业舆情受众面大，影响范围广，错误信息较易引发重大舆情事件或次生事件；（2）新媒体快速涌现，海量性、交互性、即时性等新媒体特征加速舆情传播和响应过程，增大了舆情管控难度。区块链作为近年来关注度较高的新兴产业，其发展过程涉及的网络舆情较多，准确识别和归纳舆情隐含的情感因素和意见组成，对促进区块链产业健康发展十分重要<sup>[5]</sup>。

本文提出 BERT-BiLSTM-LSTM 区块链产业舆情监测模型。模型主体由情感分析层和热度预测层两部分组成：在情感分析层，根据情感极性对舆情文本赋值并分类，利用 BERT-BiLSTM (Bi-directional Long Short-Term Memory, BiLSTM) 挖掘各情感类属文本的主题词；在热度预测层，提出使用转发数量、评论数量、总体博文数量，结合积极情感值、消极情感值和总的绝对情感值构建指标体系，并输入 LSTM (Long Short Term Memory, LSTM) 深度学习模型进行热度预测。旨在为监管部门判别舆情事件的发展和制定舆情引导策略提供实证基础，为区块链产业舆情监管提供模型依据。

## 1 相关研究

目前，网络舆情的研究对象包括公共事务、社会事件、社会现象、社会问题和政府管理机构等<sup>[6]</sup>，主要内容分为“舆情传播与演化”和“舆情监测与预警”两部分，以机器学习、文本挖掘为主要研究方法<sup>[7]</sup>。

舆情传播与演化研究侧重于舆情文本分析，关注描述和把控舆情发展规律和演变路径。学者们常使用的方法有机器学习与文本挖掘<sup>[8,9]</sup>。传统机器学习方法为提升训练的适用性，依赖大量人工标注以提高训练集质量，原因在于机器学习方法无法人工标注自然语言特征，当处理高阶、抽象的自然语言任务时，机器学习模型表现较差<sup>[10,11]</sup>。深度学习方法则以词或句的向量化为前提，不断学习语言特征，能够较好的掌握更高层次、更抽象的语言结构。以文本挖掘方法中的 LSTM 深度学习模型为例，LSTM 作为 RNN (Recurrent Neural Network, RNN) 改

进模型不断学习词、句向量，无需人工定义训练集，可自动学习高层次特征，并且通过自身存储单元记录任意时刻的细胞状态，增强句中远距离词汇的依赖关系<sup>[12]</sup>，广泛应用于自然语言处理任务中。而 BiLSTM 模型则是在 LSTM 模型的基础上结合 LSTM 输入序列前向和后向两个方向上的信息，赋予其捕获词汇上下文的能力<sup>[13]</sup>。陆敬筠等<sup>[14]</sup>提出 BiLSTM 结合 LDA (Latent Dirichlet Allocation, LDA) 模型，通过获取热门主题评论的情感极性，分析潜在舆情隐患。刘继等<sup>[15]</sup>基于 BERT-BiLSTM 模型剖析正负情感特征词网中的成团特征词，给出了关于“新型冠状病毒肺炎”事件网络舆情引导方向的建议。郝彦辉等<sup>[16]</sup>将 BERT 与 BiLSTM 结合构建深度神经网络模型，提升文本情感主题挖掘的准确性，为网络舆情的情感分析提供了实证基础。

舆情监测与预警研究侧重于对舆情状况预测和实时跟踪。由于网络舆情变化趋势复杂、规律性弱，相较于传统统计学方法，机器学习方法可以更好地完成预测任务<sup>[17-19]</sup>。曾子明等<sup>[20]</sup>在建立微博舆情热度评价指标体系的基础上，提出了基于 BP (Back-Propagation, BP) 神经网络的突发传染病舆情热度预测模型。林育曼等<sup>[21]</sup>基于 BP 神经网络结合 ARIMA (Autogressive Integrated Moving Average model, ARIMA) 模型预测微信舆情热度。靳春妍等<sup>[22]</sup>在使用 Doc2vec 和支持向量机对舆情文本实现情感分析后，利用 LSTM 进行时间序列分析预测舆情热度。通过文献梳理发现，在网络舆情监测研究中将情感倾向作为舆情热度预测关键要素的研究较少。基于此，本文以区块链产业为例，引入情感极

性特征，构建融合情感分析与多元时间序列的舆情监测模型。

## 2 舆情监测模型的构建

舆情监测模型包括数据处理层、情感分析

层和热度预测层。数据处理层主要获取区块链产业舆情文本数据的结构化语料集；情感分析层利用深度神经网络模型进行情感极性分析，获取情感倾向；最后，构建基于情感因素的时间序列特征，输入预测层进行舆情热度预测，模型框架如图 1 所示。

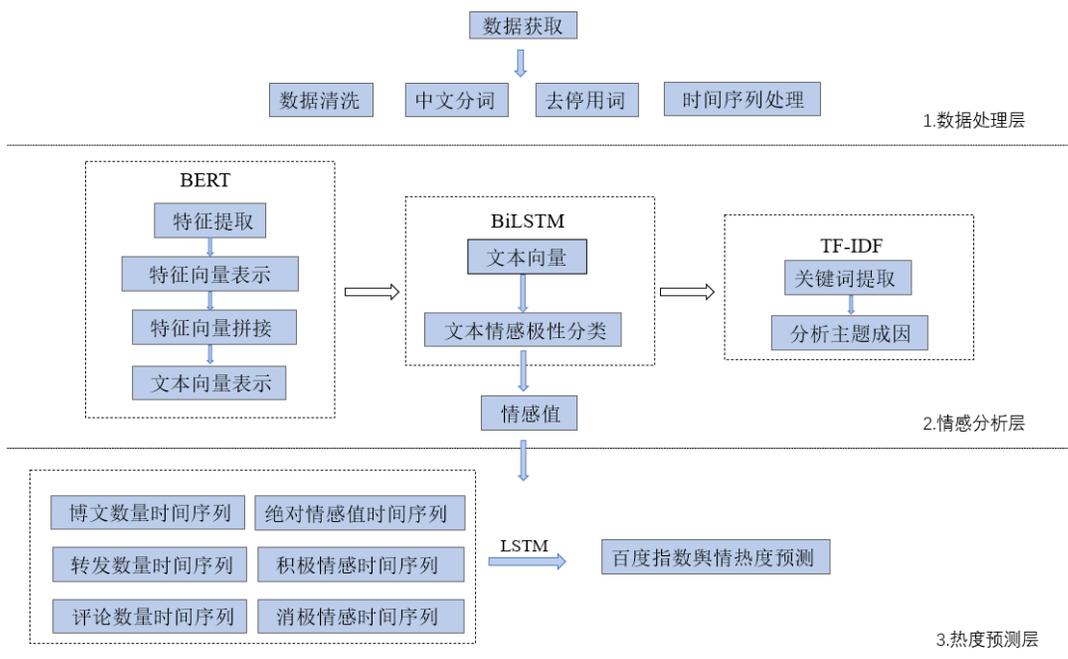


图 1 舆情监测模型框架图

### 2.1 数据处理层

在数据处理层中，对获取的区块链产业舆情文本进行数据预处理，包括数据清洗、jieba 中文分词、去除停用词和时间序列处理等，具体内容如下：

- (1) 数据清洗包括删除 URL 链接、标签、特殊符号与表情等；
- (2) 调用 python 中的 jieba 库进行中文分词；
- (3) 根据哈工大停用词表和四川大学机器智能实验室停用词库去除文本停用词；

(4) 对文本数据进行时间序列处理，按照文本数据发布时间进行排序和数量统计。

### 2.2 情感分析层

BERT 调用 Transformer 中的 Encoder 模块，将文本转换为词向量 (Token Embedding)、段向量 (Segment Embedding) 和位置向量 (Position Embedding)，通过连接残差加深网络深度，赋予模型捕捉句中双向语义能力。其内部结构包含一个多头注意力机制层，两个残差连接、正则化层和一个向前传播层，如图 2 所示。

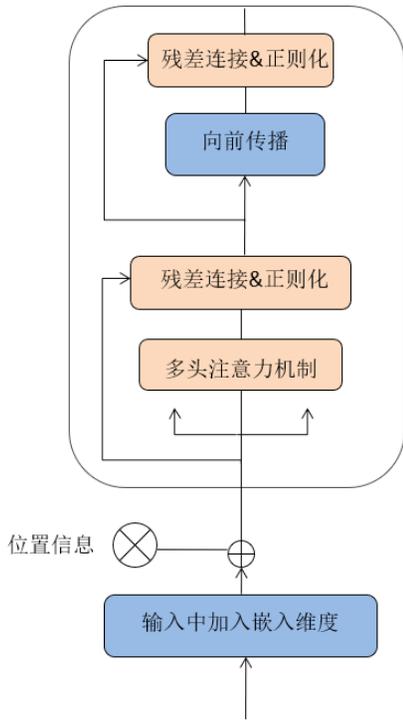


图2 Encoder内部结构图

输入的数据经过多头注意力机制层得到加权之后的特征向量，与输入层的结果进行残差连接后归一化处理，经过向前传播层后再次进

行归一化处理，输出特征向量。多头注意力机制层利用注意力机制帮助模型理解上下文语义，将句中任意词进行联系，其时间复杂度为  $O(n^2)$ ，具体计算方式如下：

$$Attention(Q, K, V) = softmax\left(\frac{QK}{\sqrt{d_k}}\right) \quad (1)$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (2)$$

$$MultiHead(Q, K, V) = Linear(W_i \text{concat}(head_1, head_2, \dots, head_n) + b) \quad (3)$$

其中，Query 向量 ( $Q$ )、Key 向量 ( $K$ )、Value 向量 ( $V$ ) 是通过 3 个不同的权值矩阵由嵌入向量乘以三个不同的权值矩阵  $W^Q$ 、 $W^K$  和  $W^V$  所得， $d^k$  是  $Q, K$  矩阵的列数，即向量维度， $W$  是输入数据  $x$  的权值， $b$  为偏置。正则化层 ( $Linear$ ) 的引入可以有效防止梯度爆炸、消失问题，具体计算方式：

$$Linear(x_i) = W \left( \frac{x_i - \mu_L}{\sqrt{\sigma^2 + \epsilon}} \right) \quad (4)$$

其中， $x_i$  为输入， $\mu_L$  为 Linear 层的平均数， $\sigma^2$  为 Linear 层方差， $\epsilon$  是一个极小数。

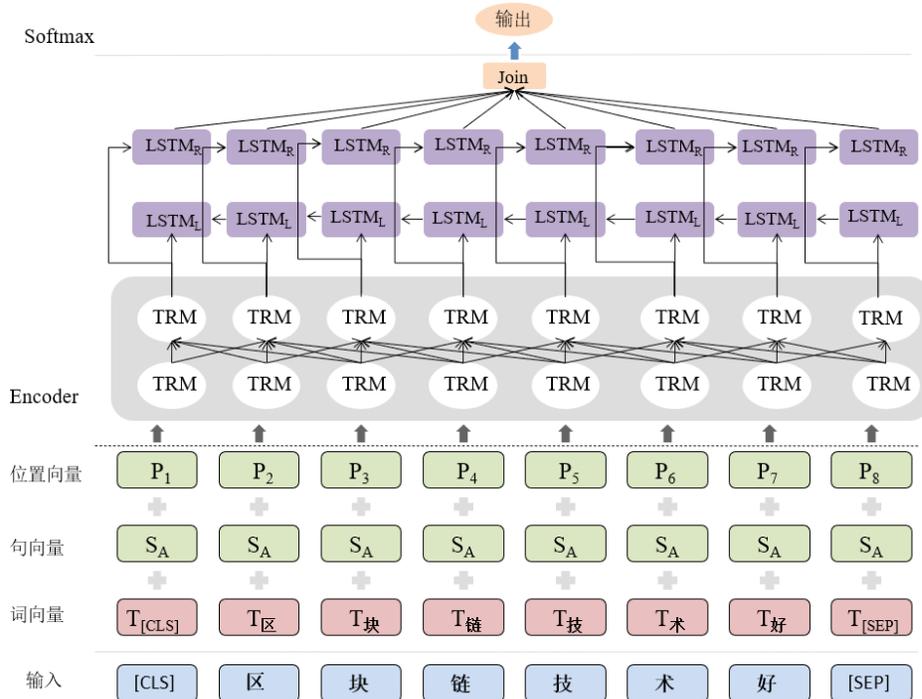


图3 情感分类模型图

模型经过词嵌入后将 BERT 的最后一层以特征向量的形式输入 BiLSTM, 如图 3 所示, 其中, TRM 代表的是 Transformer 的 Encoder 层, 在 BiLSTM 中将双向加权向量进行全连接, 输出新向量并利用 Softmax 函数分类。实验将输入文本情感极性划分为积极、中性和消极, 对应赋予情感值 1, 0, -1; 利用 TF-IDF (Term Frequency-Inverse Document Frequency, TF-IDF) 获取不同情感极性下类属文本的关键词, 重点归纳、总结各情感极性类属文本主题的成因。

### 2.3 热度预测层

传统网络舆情预测研究中, 多将原创博文量、转发量和评论量作为影响舆情热度的主要特征。鉴于作为社会属性的情感倾向对网络舆情传播的重要意义, 本文在热度预测层中利用情感分析层输出情感值, 将对应时间段内的情感值取绝对值后累加, 得到绝对情感值并形成时间序列, 并将对应时间段内的积极情感值与消极情感值分别累加取绝对值, 与原创博文量、转发量和评论量共同构成 LSTM 输入值, 将对应时间段的百度指数作为 LSTM 的输出值, 进行多元时间序列预测。

LSTM 内部隐含层由输入门、输出门、遗忘门和记忆单元构成, 通过“门”选择性的控制信息流动。当前时刻的输入数据  $x_t$  和相邻时刻隐含层输出  $h_{t-1}$  分别流向输入门、输出门和遗忘门, 通过激活函数 sigmoid 将数据映射到 0~1, 得到当前时刻隐含层输出  $h_t$ , 如公式 (5-7) 所示:

$$E_t = \delta(W_{he} * h_{t-1} + W_{xe} * x_t + b_e) \quad (5)$$

$$F_t = \delta(W_{hf} * h_{t-1} + W_{xf} * x_t + b_f) \quad (6)$$

$$G_t = \delta(W_{hg} * h_{t-1} + W_{xg} * x_t + b_g) \quad (7)$$

其中,  $W_{he}$ 、 $W_{hf}$ 、 $W_{hg}$  是相邻时刻  $h_{t-1}$  的权值,  $W_{xe}$ 、 $W_{xf}$ 、 $W_{xg}$  输入数据  $x_t$  的权值,  $\delta$  代表 sigmoid 激活函数,  $b_e$ 、 $b_f$ 、 $b_g$  是偏置。

### 2.4 评价指标

本文为评价情感分析模型拟合效果, 使用准确率 (Accuracy), 精确率 (Precision), 召回率 (Recall), F1 作为情感极性分析评价指标, 计算方式如公式 (8-10) 所示。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

$$F1 = \frac{1}{n} \sum_{i=1}^n \frac{2 * P * R}{P + R} \quad (11)$$

其中, TP 代表正例预测正确的个数, FP 代表负例预测错误的个数, TN 代表负例预测正确的个数, FN 代表正例预测错误的个数。F1 值是 Precision 和 Recall 的综合指标,  $0 \leq F1 \leq 1$ , 当 Precision=Recall=1 时, 理想情况下 F1 值达到最大, 在使用 F1 值作为评价分类器性能时, 认为其值越接近 1, 分类器性能越好。

为评价热度预测模型拟合效果, 文本引入均方根误差 (Root Mean Square Error, RMSE)、平均绝对误差 (Mean Absolute Error, MAE)、决定系数 (R-Square,  $R^2$ ) 作为指标。RMSE 是指预测值与真实值偏差的平方与观测次数  $N$  比值的平方根, 误差值越小, 模型的预测效果越好; MAE 是指预测值与真实值间的偏差, 一般数值越小代表预测值与真实值差值越

小，模型预测效果越好； $R^2$  取值范围为 [0,1]，一般数值越接近 1 代表模型性能越好，0 则相反，计算方式如公式 (12-14) 所示，其中， $N$  表示总样本数量， $\hat{y}_i$  表示根据自变量和拟合模型计算出的因变量， $y_i$  表示原始因变量。

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (12)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (13)$$

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y}_i - y_i)^2} \quad (14)$$

### 3 区块链产业舆情监测研究

#### 3.1 数据来源

百度指数是以百度搜索引擎为基础，通过海量网民行为数据进行统计分析的网络平台。将网民阅读、评论和转发等行为的数量进行加权求和，通过指数化处理的方式，衡量全网对于某一话题的关注度。论文检索“区块链”关键字，发现其百度指数在 2022 年 3—8 月期间呈不稳定的波动趋势，为较好探究区块链产业舆情发展态势和情感分布，爬取 2022 年 3 月 1 日至 2022 年 8 月 31 日连续 6 个月区间的微博博文及相关评论，共计 46 869 条。对文本数据进行预处理，去除特殊字符、数字、表情符号和 Web 链接，部分文本如表 1 所示。

#### 3.2 实验结果及分析

##### 3.2.1 网民舆情事件关注度分析

论文首先对原创博文、转发量和评论量的发布情况进行对比，发现 2022 年 3 月—2022 年 8 月期间发文量出现了 3 个明显的波动区间，如图 4 所示。第一波动区间为 6 月 13 日—6 月

18 日，该时间段，国际数字藏品领域 - 音乐版权数字藏品售价达 500 万元，刷新区块链产业成交额历史最大值；此外，网民对于“无锡区块链冤案”的关注度呈现上升趋势，事件最早可以追溯到 3 月 1 日，在社交媒体中引起广泛热议。第二波动区间为 7 月 9 日—7 月 15 日期间，中国信通院发布《基础设施研究报告》，提出随着区块链基础设施规模不断扩展、运行能力不断增强，区块链技术的应用将更加多元化；另一方面，链上产业区块链研究院在南京鼓楼高新区揭牌，预示着区块链全产业链落地正迈向高速发展时代。第三波动区间为 7 月 22 日—7 月 28 日，商务部等 27 个部门联合印发《关于推进对外文化贸易高质量发展的意见》指出促进大数据、云计算、人工智能、区块链等新技术应用，带动传统行业数字化转型，提升企业数字化运营能力；另外，中国联通为推进农业生产数字化转型，加速“数字蝶变”等也推动关注度的提升。通过网民对博文关注度分析发现，网民对于区块链产业的关注度可能与金融价值、产业落地和政策法规的颁布实施有关，并且随着其金融价值的起伏、政策的出台和产业落地变化<sup>[23]</sup>。

##### 3.2.2 网络舆情情感极性分析

本文使用 ChineseNlpCorpus 研究团队的开源数据集对 BERT-BiLSTM 模型训练，样本数量为 2 131 837 条，其中，训练样本集为 70%，测试样本集为 20%，验证样本集为 10%。文本长度设置为 512 个字符。为避免模型过拟合，模型将不同层赋值不同权重，在损失函数的基础上增加权重参数的绝对值，将模型学习率分别设置为  $1e-5$  和  $1e-4$ 。实验结果如表 2，在训

表 1 相关评论表

Date	Trans	Vote	Comment	Content
2022.3.21	1096	337	127	“i茅台”试运行两天，申购结果公示后，几家欢喜几家愁，因为引入了最顶尖的区块链技术，确保了申购的公平性，之前活跃在各电商平台的黄牛这次全部被狙击。另一方面，申购成功的真实用户已纷纷在各个平台晒出自己的战绩，在公平性的前提下，每天记得去申购，下一个好运的就是你！
2022.3.21	742	875	622	想你们了，将近一年了，你们还好吗？不知道你们在里面是“晴天”还是“雨天”？恳请社会各界人士关注转发无锡区块链冤案。
2022.6.15	2498	4159	2339	日前，知名音乐唱作人萧全 @萧全Seal 与贝多音乐平台 @贝多音乐科技联合发布的《海草舞》音乐版权数字藏品，以500万美金价格成交（前360枚）。这是国际数字藏品圈及区块链行业有史以来成交额最高的音乐版权数字藏品之一，同时也刷新了单曲数字藏品最高价值的世界纪录。
2022.7.28	962	219	223	广西联通与贺州市人民政府签订5G应用“扬帆”战略合作框架协议。

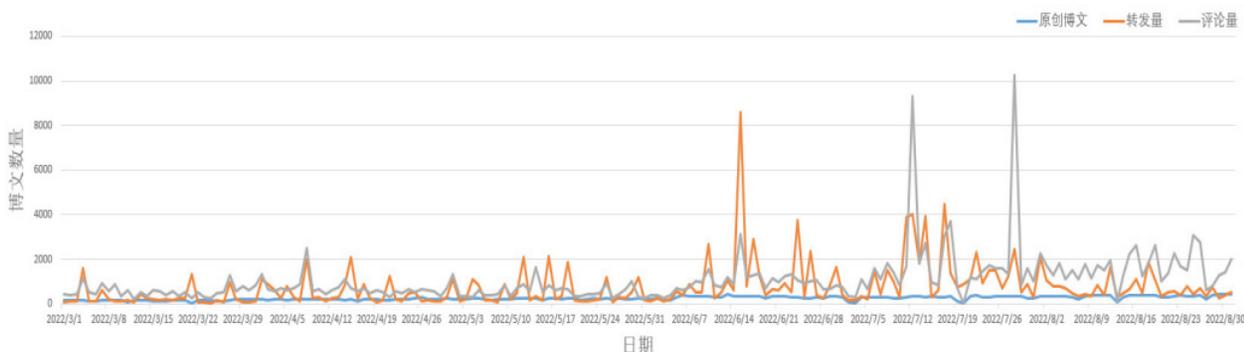


图 4 网民关注度变化

表 2 模型精准度表

	Precision	Recall	F1
Postive	0.84	0.87	0.85
Neutral	0.75	0.90	0.82
Negative	0.93	0.76	0.84
Accuracy	—	—	0.84

练集中的积极 (positive) 文本、中性 (neutral) 文本和消极 (negative) 文本的情感极性分析 F1 值分别为 0.85、0.82 和 0.84，模型整体准确率达到了 84%，模型情感极性分析效果较好。

接下来，将 2022 年 3—8 月采集到的博文评论导入已训练模型，进行情感极性分析，对各情感极性类属文本进行分类并赋值，具体分析结果如表 3 所示。2022 年 3—8 月期间，博文发布数量逐月上升，呈现出以积极情感为主

导、消极情感为次要地位的情感极性分布。其中，2022 年 8 月积极情感博文数量占当月博文数量的比重最高，为 82%；2022 年 5 月消极情感博文数量占当月博文数量的比重最高，为 22%，相应的情感极性分布如图 5 所示。

表 3 数据统表

日期	数据量	积极	中性	消极
2022年3月	5 019	4 014	84	851
2022年4月	6 344	5 099	107	1 138
2022年5月	7 111	5 439	135	1 537
2022年6月	9 441	7 356	206	1 879
2022年7月	8 852	7 236	147	1 469
2022年8月	10 908	8 965	193	1 750

### 3.2.3 TF-IDF主题词分析

本节使用 TF-IDF 方法，探究不同情感极性下主题词的变化与潜在成因。TF 值是一个词在

文本中出现的次数，出现次数越多，表示词对文本越重要；IDF 值是倒转文档频率，表示词的区分能力，词的区分能力越弱则主题代表性

越弱。最后，结合词与上下文之间的语义聚合程度，获得词在文本中权重指标并制作词云图，如图 6 所示。

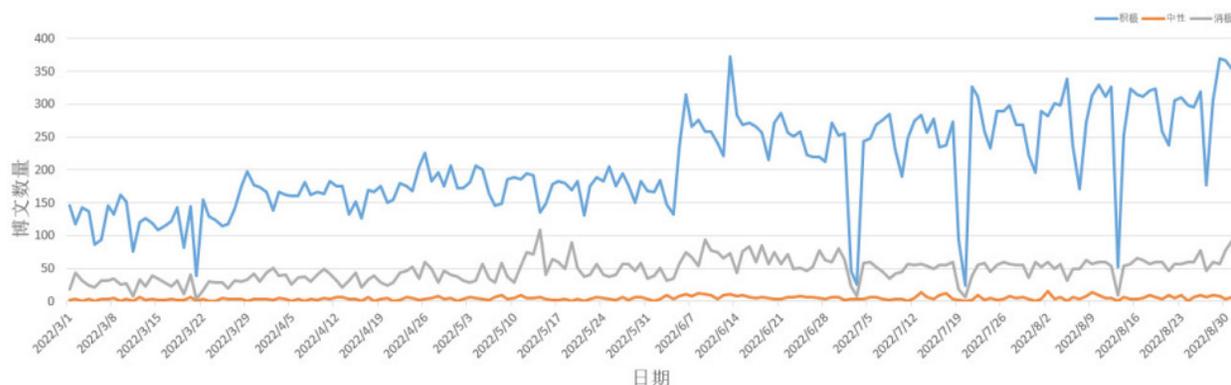


图 5 情感极性分布



图 6 不同情感极性下的词云图

在情感极性为积极属性的类属文本中，关键词主要为“数字藏品”“比特币”等。结合内容可知，文本主要对目前区块链产业的技术落地和产品转化进行积极回应。如“私募投资、房地产、零售餐饮商业等资产代币化，值得探索”“这个藏品价格真的好高啊，惊到我了，太厉害了”等。此类网民对区块链投资方面的关注度较高，舆情监管部门应着重关注以区块链概念股和数字货币为主题的投资经验贴，避免不实信息、虚假信息对网民进行错误引导。

在情感极性为中性的类属文本中，关键词主要为“元宇宙”“以太坊”等。此类文本主要是网民缺乏对于区块链相关概念的了解，对区块链产业展现出不同程度的投资热情。如“小

白也能懂吗，谁来说说元宇宙区块链怎么回事”“区块链还挺好玩的，想弄透”和“区块链没有财富密码，自己创作财富，冲以太坊”等。对于上述问题，相关部门应加强区块链相关概念、应用场景和投资属性等问题的宣传科普工作。在普及相关知识的同时，对于关注度较高的“投资经验贴”和“知识科普贴”加强舆情监管。

在情感极性为消极的类属文本中，关键词主要为“数字货币”“技术”等。文本数据显示网民一定程度上缺乏对于区块链技术的信任。如“就是为了发行虚拟货币来割韭菜”“虚拟货币本来就是传销工具国家早已出台政策，对所有和虚拟货币有关的犯罪行为，严厉打击，彻底

封杀”等。相关部门应加强网民对区块链金融诈骗的防范意识，减少风险发生的可能，同时加深政策解读，推动区块链产业的法制建设。

### 3.2.4 网络舆情热度预测

论文在网络舆情热度预测中，将转发量 (Trans)、评论量 (Comment)、原创博

文量 (Total)、绝对情感值 (Emotion)、积极情感值 (Positive) 和消极情感值 (Negative) 构建舆情热度预测模型指标体系，如表 4 所示。本文根据两组不同的特征并设立对照试验，按照 8 : 2 划分训练集与测试集。

表 4 网络舆情热度预测指标体系

Date	Trans	Comment	Total	Emotion	Positive	Negative
2022.7.11	3 898	1 695	306	305	548	57
2022.7.12	4 028	9 321	334	329	274	55
2022.7.13	1 816	1 830	354	340	283	57
2022.7.14	3 954	2 471	318	311	257	54
2022.7.15	322	979	330	327	278	49

本文采用 LSTM 进行多元时间序列预测，测定最佳滞后范围并对模型进行参数调优。在时间窗口的选择上设定最长滞后范围为 1-5 天，分别对比传统指标与融合情感因素的指标预测结果，实验表明当时间窗口滞后 5 天时模型预测精度最优，如表 5 所示。从精度来看，传统预测指标体系下模型 RMSE 为 210.14，MAE

为 163.10，R<sup>2</sup> 为 0.55；融合情感因素后的模型 RMSE 为 192.47，MAE 为 147.96，R<sup>2</sup> 为 0.66，相较于传统预测指标，RMSE 降低 17.67，MAE 降低 15.14，R<sup>2</sup> 提升 11%，真实值与预测值对比如图 7 所示。结果表明，加入情感倾向作为舆情预测模型的关键特征，在一定程度上缩小预测误差，提升了舆情预测的精度。

表 5 LSTM 舆情热度预测

传统预测指标体系				融合情感因素的预测指标体系			
窗口	RMSE	MAE	R <sup>2</sup>	窗口	RMSE	MAE	R <sup>2</sup>
W <sub>1</sub>	222.94	176.13	0.50	W <sub>1</sub>	219.39	174.12	0.52
W <sub>2</sub>	215.37	169.28	0.53	W <sub>2</sub>	211.86	167.53	0.55
W <sub>3</sub>	212.75	165.52	0.54	W <sub>3</sub>	196.43	174.73	0.62
W <sub>4</sub>	210.23	163.97	0.55	W <sub>4</sub>	203.45	158.43	0.59
W <sub>5</sub>	210.14	163.10	0.55	W <sub>5</sub>	192.47	147.96	0.66

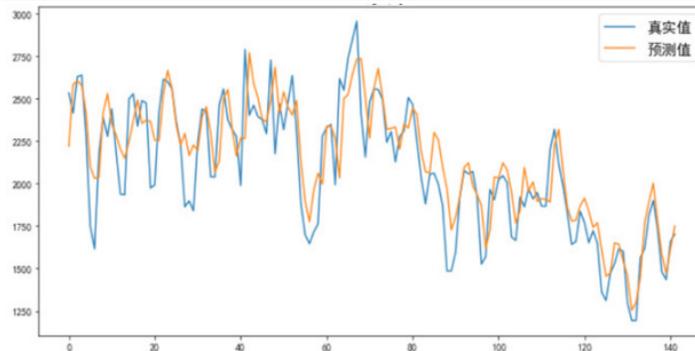


图 7 LSTM 舆情热度预测

## 4 结语

本文提出使用基于 BERT-BiLSTM-LSTM 模型的区块链产业舆情监测方法, 结合区块链产业热点舆情事件的数据进行实证研究。以情感极性分析与舆情热度预测作为切入点, 将情感因素纳入舆情预测指标中。研究表明, 通过 TF-IDF 方法准确识别各情感极性下类属文本主题词的变化, 通过引入积极情感值、消极情感值和绝对情感值作为舆情热度预测的关键因素, 有效提高了模型热度预测能力, 增强了区块链产业舆情监测效果。

为更好实现区块链产业舆情监测任务, 根据本论文实验结果提出以下政策建议:

(1) 加强舆论引导, 树立意见领袖。在技术投产和项目落地的过程中, 网民对区块链等新兴产业的基本概念和技术应用缺乏全面了解, 舆情监管部门应加强对热点话题、热点博主的监察力度, 防止不实信息传播; 通过官方媒体渠道树立“意见领袖”, 对区块链产业相关概念进行详尽阐述, 以自身影响力和政府公信力为背书进行舆论引导, 第一时间发布正面消息影响更多网民持正面观点, 以此形成良性循环。

(2) 围绕网民关注关切, 切实加强政策解读。目前, 我国区块链等战略性新兴产业的政策、法规尚不完善, 虚拟资产的法律权属关系界定不清晰, 容易使网民产生困惑, 激发舆情事件的出现。建议通过官方媒体发布通稿、举办新闻发布会、线上访谈等方式围绕网民关注点进行详细解读; 此外, 相关政策解读的工作时效性要求高, 应建立长效、便捷的信息获取途径加快推进政策解读工作整体进程。

(3) 建立区块链产业舆情预警机制。论文发现网民关注度在政策公布前后、金融价值发生大幅波动时舆情热度变化程度尤为明显。为防止网民受消极观点影响, 首先在舆论波动初期应进行相应疏导; 其次, 当网民关注度高度聚焦时, 应根据历史舆情事件发展规律, 以舆情事件热度为重要指标, 对舆情走势进行研判, 避免负面舆论引发舆情次生事件对产业造成不良影响。

(4) 加强区块链投资风险管控。随着区块链产业发展进程不断深化其金融价值不断提升, 区块链金融诈骗风险骤升。围绕此问题, 应建立以区块链投资为主要核心的风险管控机制, 加强对互联网平台相关舆论的管控, 通过官方媒体、APP 发布风险提示等方式增强网民对于区块链金融风险的防范意识, 避免财产损失。

## 参考文献

- [1] 袁勇, 王飞跃. 区块链技术发展现状与展望 [J]. 自动化学报, 2016, 42(4): 481-494.
- [2] 《中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要》[http://www.gov.cn/xinwen/2021-03/13/content\\_5592681.htm](http://www.gov.cn/xinwen/2021-03/13/content_5592681.htm).
- [3] 吴俊, 邵丹睿, 姜尚杨帆. 融合语义与情感分析的区块链产业新闻监测研究 [J]. 现代情报, 2020, 40(11): 22-33.
- [4] 龙玥, 刘译阳. 新媒体环境下高校负面网络舆情传播特征和路径研究 [J]. 情报科学, 2019, 37(12): 134-139.
- [5] 邓建高, 张璇, 傅柱, 等. 基于系统动力学的突发事件网络舆情传播研究: 以“江苏响水爆炸事故”为例 [J]. 数据分析与知识发现, 2020, 4(Z1): 110-121.
- [6] 王连喜. 网络舆情领域相关概念分布及其关系辨析 [J]. 现代情报, 2019, 39(6): 132-141.
- [7] 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.
- [8] 张志霞, 郝纹慧, 张二双. 网络舆情驱动下突发事件情景推演研究 [J]. 情报科学, 2020, 38(5): 141-147.
- [9] 王家辉, 夏志杰, 王诣铭, 等. 基于句法规则和社会网络分析的网络舆情热点主题可视化及演化研究 [J]. 情报科学, 2020, 38(7): 132-139.
- [10] Al-Hadhrani S, Al-Fassam N, Benhidour H. Sentiment Analysis Of English Tweets: A Comparative Study of Supervised and Unsupervised Approaches[M]. 10.1109/CAIS.2019.8769550, 2019.
- [11] Prediction of California Bearing Ratio from Index Properties of Soils Using Parametric and Non-parametric Models[J]. Geotechnical and Geological Engineering. 2018, 36(6): 3485-3498.
- [12] Xu G, et al. Sentiment analysis of comment texts based on BiLSTM[J]. IEEE Access, 2019(7): 51522-51532.
- [13] 曾子明, 万品玉. 基于双层注意力和 Bi-LSTM 的公共安全事件微博情感分析 [J]. 情报科学, 2019, 37(6): 23-29.
- [14] 陆敬筠, 胡舜奕, 俞建光. 基于 LDA-BiLSTM 模型的高校网络舆情监测方法及实证 [J]. 情报理论与实践, 2020, 43(11): 156-161.
- [15] 刘继, 顾风云. 基于 BERT 与 BiLSTM 混合方法的网络舆情非平衡文本情感分析 [J]. 情报杂志, 2022, 41(4): 104-110.
- [16] 郝彦辉, 王曦, 陈铎. 基于 BERT-BiLSTM 模型的舆情监测方法及实证研究——以研究生招生考试为例 [J]. 情报科学, 2021, 39(8): 78-85.
- [17] 傅雷鸣, 王兆焯, 陈一飞, 等. 大数据情境下网络舆情分众化演变趋势分析 [J]. 情报科学, 2022, 40(2): 113-117+126.
- [18] 郑步青, 邹红霞, 胡欣杰. 基于拐点的网络舆情预测研究 [J]. 计算机科学, 2018, 45(S2): 539-541+575.
- [19] 张艾宁. 基于 EEMD-GRU-XGB 的舆情热度预测模型 [J]. 信息与电脑 (理论版), 2022, 34(5): 56-61.
- [20] 曾子明, 黄城莺. 基于 BP 神经网络的突发传染病舆情热度趋势预测模型研究 [J]. 现代情报, 2018, 38(5): 37-44+52.
- [21] 林育曼, 文海宁, 饶浩. 基于 ARIMA-BP 神经网络模型的微信舆情热度预测 [J]. 统计与决策, 2019, 35(14): 71-74.
- [22] 靳春妍, 牟冬梅, 王萍, 等. 融入表情特征的网络舆情情感分析方法研究 [J]. 科技情报研究, 2020, 2(4): 13-22.
- [23] 凌晨, 冯俊文, 吴鹏, 等. 基于 SOAR 模型的高校网络舆情应急响应研究 [J]. 情报科学, 2019, 37(9): 145-152.