第 11 卷 第 1 期
2022 年 1 月
集 成 技 术
JOURNAL OF INTEGRATION TECHNOLOGY
Vol. 11 No. 1
Jan. 2022

引文格式：
刘薇, 姜青山, 蒋泓毅, 等. 基于 FinBERT-CNN 的股吧评论情感分析方法 [J]. 集成技术, 2022, 11(1): 27-39.
Liu W, Jiang QS, Jiang HY, et al. A sentiment analysis method based on FinBERT-CNN for Guba stock forum [J]. Journal of Integration Technology, 2022, 11(1): 27-39.

基于 FinBERT-CNN 的股吧评论情感分析方法

刘 薇[1,2] 姜青山[1*] 蒋泓毅[1] 胡金帅[3] 曲 强[1]

[1]（中国科学院深圳先进技术研究院 深圳 518055）
[2]（中国科学院大学深圳先进技术学院 深圳 518055）
[3]（厦门大学 厦门 361005）

摘 要 我国股市波动受投资者情绪变化影响较大，通过对股吧等金融交流平台上投资者的评论进行情感分析，能够帮助投资者更好地了解股票市场的变化。现有的情感分析方法是利用模型对股票评论集进行分析，但缺少优质的股票评论标注数据集用于模型训练，且单一模型提取股票评论特征较为片面，模型的准确性有待提高。该文针对股吧平台上的评论数据，提出一种基于 FinBERT-CNN 的股吧评论情感分析方法，该方法通过 FinBERT 预训练模型学习股吧评论数据语义特征，解决缺乏股吧评论标注数据集的问题，并利用卷积神经网络学习股吧评论的局部特征，使模型充分学习股吧评论特征，提高模型情感分类的准确性。实验结果表明，基于 FinBERT-CNN 的股吧评论情感分析方法均优于现有情感分析方法。此外，通过基于股吧评论情感的股票市场关联分析，验证了股吧评论情感变化与股市波动存在相关性。

关键词 股吧评论；情感分析；预训练模型；FinBERT；卷积神经网络
中图分类号 TP 399 **文献标志码** A doi: 10.12146/j.issn.2095-3135.20210228001

A Sentiment Analysis Method Based on FinBERT-CNN for Guba Stock Forum

LIU Wei[1,2] JIANG Qingshan[1*] JIANG Hongyi[1] HU Jinshuai[3] QU Qiang[1]

[1](*Shenzhen Institute of Advanced Technology*, *Chinese Academy of Sciences*, *Shenzhen* 518055, *China*)
[2](*Shenzhen College of Advanced Technology*, *University of Chinese Academy of Sciences*, *Shenzhen* 518055, *China*)
[3](*Xiamen University*, *Xiamen* 361005, *China*)
[*]*Corresponding Author*: *qs.jiang@siat.ac.cn*

Abstract The fluctuation of the stock market greatly depends on investors' sentiment-based factors.

收稿日期：2021-02-28 **修回日期**：2021-05-10
基金项目：广东省自然科学基金项目（2018A030313943）；深圳基础研究（自由探索）项目（JCYJ20180302145633177）
作者简介：刘薇，硕士研究生，研究方向为机器学习和金融科技；姜青山（通讯作者），研究员，研究方向为数据挖掘，E-mail：qs.jiang@siat.ac.cn；蒋泓毅，研究员助理，研究方向为机器学习和金融科技；胡金帅，副教授，研究方向为公司治理及资本市场会计研究；曲强，研究员，研究方向为区块链技术。

Sentiment analysis of investors' reviews on financial exchange web platforms such as Guba stock forum (guba.com.cn), can help stockholders to understand the stock market more effectively. However, due to the unavailability of high-quality labeled datasets and deficient features of stock comments extracted by a single model, the accuracy of the existing sentiment methods still requires further improvements. This paper proposes a method that utilizes the FinBERT-CNN-based sentiment model for Guba comments. The semantic features of Guba comments are extracted by using the FinBERT pre-training model. Meanwhile, a convolution neural network (CNN) is applied to learn the local features of Guba comments. It enables the proposed method to learn features more precisely and improve the emotion classification's accuracy significantly. Experiments show that the proposed method outperforms the existing models. Furthermore, the correlation analysis on Guba comments and stock market data demonstrates a relationship between the investors' emotions and stock market volatility.

Keywords Guba stock forum; sentiment analysis; pre-training model; FinBERT; convolution neural network
Funding This work is supported by Natural Science Foundation of Guangdong Province of China (2018A030313943), and Shenzhen Basic Research Foundation (JCYJ20180302145633177)

1 引 言

我国的股票交易市场整体呈现出非有效性，舆论和政策对股票市场有较大影响[1]。根据行为金融学理论，投资者并不都是理性投资者，在投资的过程中会出现受他人情绪影响而改变投资倾向的现象[2]。随着网络的普及，越来越多的网民在以“东方财富网-股吧”（https://guba.eastmoney.com）为代表的金融网站上进行信息交流和传播。这些股票评论在传递大量市场信息的同时，还影响投资者的交易决策。因此，通过研究金融平台上的股票相关评论，分析投资者的情感倾向，对了解股票市场变化有重要意义。

情感分析是指利用自然语言处理和文本挖掘技术，对带有情感色彩的主观性文本进行分析、处理和抽取的过程[3]。国内外学者在股票评论情感分析领域的研究方法分为 3 种：基于词典的方法[4-5]、基于机器学习的方法[6-9]和基于深度学习的方法[10-14]。 基于词典的方法是指以情感词典作为情感倾向判断依据，该方法虽然对情感倾向判断简单，但是对词典依赖较高，且目前缺乏公开的股票情感词典。基于机器学习的方法是指利用评论信息特征进行情感分析，但是分析效果受评论特征的构建和训练语料影响较大。目前，基于深度学习的方法在计算机视觉、自然语言处理等领域应用广泛[15]，然而，对于金融评论情感分析，仍缺乏优质的金融评论标注数据集用于模型的训练。

本文通过研究股吧评论情感分析方法，构建更具准确性的情感分类模型，并对股票市场与股吧评论情感之间的关联性进行验证。具体地，针对真实的股吧评论数据，首先通过预训练模型 FinBERT 学习股吧评论的语义特征，然后采用卷积神经网络提取股吧评论的局部特征，最后利用“上证 50”成分股数据集验证股票市场与股票评论情感之间存在关联。

2 股票评论情感分析方法的研究现状

2.1 基于词典的股票评论情感分析方法

基于词典的方法的核心在于情感词典的构建，通过将评论数据与情感词典的正负情绪词进

行匹配，得到情感倾向分类结果。Maqsood 等[4]利用英文情感词典 SentiWordNet，将从 Twitter 股票相关评论数据中提取的 5 000 余个英文单词分为积极、中性和消极 3 类情感词，从而实现对英文股票评论进行情感分析。陈雷[5]提出一种基于情感词典的股票评论情感分析方法，首先通过对几种公开的情感词典合并去重后得到基础词典，然后引入股票情感词典对基础词典进行扩充。但是，目前在股票评论分析领域，缺少公开的股票情感词典，且该词典构建难度大、成本高，并且模型受情感词典的影响较大。

2.2　基于机器学习的股票评论情感分析方法

基于机器学习的方法，通过选择特征项、构造特征对评论数据进行特征提取，实现利用机器学习模型进行情感分类。Alkubaisi 等[6]利用一种混合朴素贝叶斯分类器的方法对 Twitter 股票评论数据集进行情感分析，并取得了较高的情感分类准确率。Yazdani 等[7]讨论了分别利用二进制编码、词频和词频-逆文本频率指数对金融新闻文章进行特征提取，并利用具有不同核函数的支持向量机作为分类器对金融文本进行情绪分类，实验结果表明，特征选择和特征加权在情感分类中起到重要作用。Salles 等[8]针对股票评论数据具有高维噪声的特点，提出延迟随机森林分类模型，通过最近邻样本投影得到与待分类样本相似的特征，具有较高的分类准确率。张对[9]通过对新浪股吧中的股评数据进行分词和过滤停用词，将剩下的名词、动词和副词等作为特征项进行词频统计，然后利用 SVM 分类器对股评数据进行情感分类。但是，基于机器学习的情感分析方法对特征项选取和特征构建的质量依赖较高，且缺少优质的股票评论标注数据集用于训练。

2.3　基于深度学习的股票评论情感分析方法

基于深度学习的方法，首先把输入映射到不同的特征空间来进行特征提取，然后持续地修正神经网络权重以学习评论特征。Jiang 等[10]将基于门控和关注机制的双向长短期记忆神经网络模型用于处理股票新闻和微博的情感分析任务，首先利用门控机制整合字符级别和单词级别的嵌入，然后利用双向长短期记忆神经网络组件将目标相关信息嵌入语句，最后利用线性回归层进行情感分类。Rao 等[11]提出将基于两个隐层的长短期记忆网络模型用于股票评论情感分析，其中，第一层用于学习句子语义，第二层用于学习句子关系。Sohangir 等[12]将长短期记忆网络和卷积神经网络这两个模型进行对比，结果表明，卷积神经网络在股票评论数据集上表现更好。Akhtar 等[13]利用卷积神经网络和词向量模型对股评数据进行情感分析——利用词向量模型将股评数据转化为向量，并通过卷积神经网络得到分析结果。直接采用卷积神经网络对股评数据进行情感分析能够实现自动提取特征，修正学习输出，但缺少优质的股票评论标注数据集用于卷积神经网络的训练，且卷积神经网络无法处理股评数据中的专业词汇。

Liu 等[14]提出基于 BERT 的预训练模型 FinBERT，并采用一个大型通用的金融语料集对其进行训练。FinBERT 首先在 TRC2-financial 语料库上训练语言模型，然后利用其权重初始化金融评论情感分析模型。FinBERT 由编码器和解码器堆叠形成的 Transformer 组成，每个 Transformer 利用编码器的多头自注意力机制将任意位置的词之间建立联系。FinBERT 能够解决目前金融领域缺乏优质股票评论数据集的问题，且能够解决无法处理金融专业领域词汇的问题，但是使用 FinBERT 仅能提取到股票评论的语义特征，而缺少股票评论的局部特征和空间特征。

针对现有模型存在缺少优质股票评论标注数据集用于模型训练，无法处理专业词汇，特征提取过于单一等问题，本文提出一种基于 FinBERT-CNN 的股吧评论情感分析方法，利用 FinBERT 预训练模型提取股吧评论语义特征和卷积神经网络模型来捕捉评论的局部特征和空间特征。

3 基于 FinBERT-CNN 的股吧评论情感分析方法

基于 FinBERT-CNN 的股吧评论情感分析方法将爬虫采集的股吧评论数据预处理后，分别利用 FinBERT 预训练模型提取语义特征，卷积神经网络提取局部特征和空间特征，最后使用全连接层输出情感倾向分类结果。基于 FinBERT-CNN 的股吧评论情感分析方法主要包括数据预处理和基于 FinBERT-CNN 的股吧评论情感分析模型 2 个部分，流程如图 1 所示。

3.1 数据预处理

利用 Scrapy 爬虫框架对“东方财富网-股吧”上所有 A 股股票的评论数据进行爬取，直接获取的股吧评论存在大量无用信息和冗余，需对其进行预处理。数据预处理包含以下 4 个步骤：

(1) 数据去重与合并。由于增量式爬虫会引起数据重复问题，所以在情感分析之前需进行评论去重。若评论正文和标题都没有内容则直接删除；若标题或正文没有内容，则合并标题与正文；

(2) 人工标注数据。在股吧评论数据集中随机选取 15 747 条股票评论并由金融专业人士对其标注情绪标签：“看涨”记作 1，“看跌”记作 −1，中性记作 0。看涨是指投资者表达价格将上涨的情绪，一般包括对当前股市积极的评价和对未来股市走势上涨的判断。看跌是指投资者表达价格将下跌的情绪，一般包括对当前股市的消极评价和对未来走势下跌的判断。标注数据示例如表 1 所示；

表 1 标注数据示例

Table 1 Examples of data annotation

股吧 A 股股评	情感倾向	标签
最近势头很好啊，是入市的大好时机，买买买，买了就等涨！	看涨	1
唯一跌的银行股，全场 100 多支跌的股票怎么就有你呢，你是怎么做到这么垃圾的呢？	看跌	−1
有没有高手预计一下这支股票今年年底能到什么价？有没有人指导一下我要跟进还是抛出？	中性	0

(3) 划分训练集和测试集。将步骤(2)获得的标注数据按照 7∶3 的比例划分训练集和测试集；

(4) 分词。为了保持中文评论信息的完整性，对于通过步骤(3)得到的训练数据以字作为划分单位进行分词。首先将股吧评论数据转为

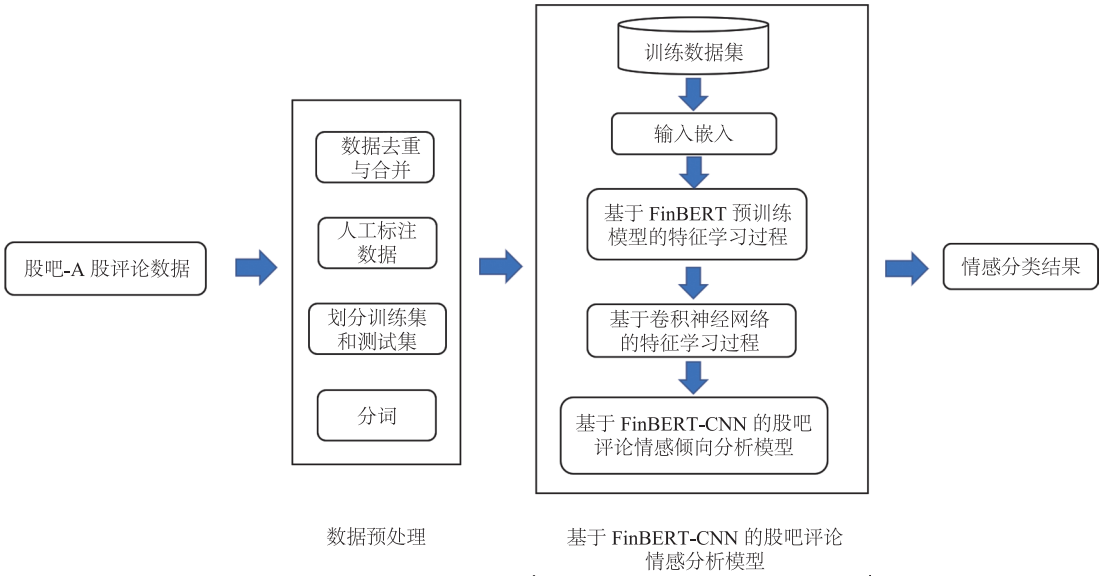


图 1 基于 FinBERT-CNN 的股吧评论情感分析方法

Fig. 1 Sentiment method for Guba reviews based on FinBERT-CNN

Unicode 格式编码，然后将其与 Unicode 特殊字符进行匹配，去除评论数据中的不合法字符和多余空格，最后判断数据中的每个字符是否为中文字符，若为中文字符就在字符后添加空格，否则就根据标点和空格对其进行分词。

从“东方财富网-股吧”上提取的 5 亿多条数据经预处理后，用于模型验证的数据共 15 747 条，用于基于股吧评论情感的股票市场关联分析的数据共 2 027 077 条。

3.2　基于 FinBERT-CNN 的股吧评论情感分析模型

基于 FinBERT-CNN 的股吧评论情感分析模型结构如图 2 所示，其中输入为预处理后的股吧评论数据，其包含数据嵌入层、FinBERT 预训练模型、特征整合、卷积层、池化层、全连接层、dropout 层和 Softmax 层，流程分为以下 3 个步骤：

(1)基于 FinBERT 预训练模型的特征学习过程：将预处理后的股吧评论数据经嵌入层输入 FinBERT 预训练模型，输出特征向量 $\boldsymbol{B}$，经过整合后得到语义特征 $\boldsymbol{F}$；

(2)基于卷积神经网络的特征学习过程：对步骤(1)中得到的语义特征 $\boldsymbol{F}$ 进行局部特征提取，通过卷积层、池化层、全连接层和 dropout 层输出特征向量 $\boldsymbol{H}$；

(3)情感倾向分析过程：对步骤(2)中得到的特征向量 $\boldsymbol{H}$，通过 Softmax 层得到情感倾向分类结果。

3.2.1　基于 FinBERT 预训练模型的特征学习过程

本文使用 FinBERT 预训练模型学习股吧评论数据的语义特征，其具体分为以下 3 个步骤：

(1)评论数据嵌入处理。对预处理后的股吧评论数据进行嵌入处理，即将文本数据向量化，其分为以下 3 个部分：

①token 嵌入：每条经过预处理分词后的股吧评论数据即为一条 token，将得到的 token 在词汇表中进行索引匹配得到 token 嵌入向量。在中文预训练模型 FinBERT 的词汇表中包含约 30 亿条 tokens。一条经过预处理后的股吧评论数据有 w 个字符，检查这 w 个字符在词汇表中是否出现。对出现的字符进行索引匹配，得到表示该字符的向量；对于未出现的字符，则生成新的字符索引。token 嵌入将每个字符用 n 维的向量表示，那么评论数据经 token 嵌入转换为 $w\times n\times 1$

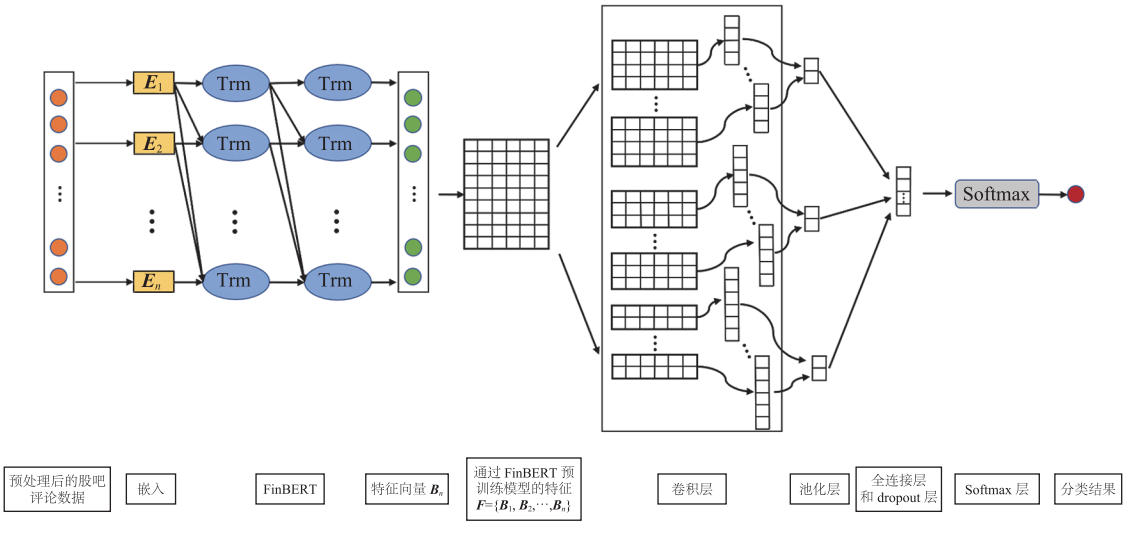


图 2　基于 FinBERT-CNN 的股吧评论情感分析模型

Fig. 2　FinBERT-CNN-based sentiment analysis model for Guba data

的张量 $\boldsymbol{T}$。为了得到模型进行分类任务的输入表示，在 tokens 的开始([CLS])和结束([SEP])处添加特定的 tokens；

②片段嵌入：本文输入评论数据的片段嵌入对应片段嵌入表的索引为 0 的向量，将其转换为 n 维的向量后得到 $w\times n\times 1$ 的张量 $\boldsymbol{S}$；

③位置嵌入：每个字符的位置信息编码都可用一个向量来表示，从而得到 $w\times n\times 1$ 的张量 $\boldsymbol{P}$，$\boldsymbol{P}$ 可表示因不同位置而带来不同语义信息的差异。

将这 3 个部分相加，得到股吧评论数据向量化表示，即：

$$\boldsymbol{E}_n=\boldsymbol{T}_n+\boldsymbol{S}_n+\boldsymbol{P}_n \tag{1}$$

其中，$\boldsymbol{E}_n$ 为股吧评论数据的嵌入；$\boldsymbol{T}_n$ 为 token 嵌入；$\boldsymbol{S}_n$ 为片段嵌入；$\boldsymbol{P}_n$ 为位置嵌入。最后对股吧评论嵌入 $\boldsymbol{E}$ 进行维度变换，使其成为大小为 $ms\times n\times b$ 的张量，b 为一次训练选取的样本量，ms 为股吧评论的最大长度。

(2)特征提取。将股吧评论数据嵌入 $\boldsymbol{E}$ 作为 FinBERT 预训练模型输入，并在隐藏层对其进行语义特征提取。一个隐藏层包括一层注意力层，两层线性映射层，两层正则化和 dropout 层，一层非线性映射层。首先对 $\boldsymbol{E}$ 通过注意力层后的输出 $\boldsymbol{E}_{\mathrm{A}}$ 进行维度变换，使其与嵌入维度 n 一致；然后通过 dropout 层后将 $\boldsymbol{E}$ 和 $\boldsymbol{E}_{\mathrm{A}}$ 相加，输入正则化层得到 $\boldsymbol{E}_{\mathrm{N}}$ 并经过维度变化，使其与嵌入维度 n 一致；接着通过 dropout 层得到 $\boldsymbol{E}_{\mathrm{D}}$；最后将 $\boldsymbol{E}_{\mathrm{N}}$ 与 $\boldsymbol{E}_{\mathrm{D}}$ 相加，输入正则化层得到语义特征 $\boldsymbol{B}$。

(3)特征整合。输入股吧评论数据嵌入 $\boldsymbol{E}$ 经过一层隐藏层后得到股吧评论数据语义特征 $\boldsymbol{B}$，每层隐藏层的输出都作为下一层隐藏层的输入，那么 FinBERT 预训练模型中 12 层隐藏层可得到 12 个输出，将其进行组合和维度变换后，通过一层池化层得到语义特征 $\boldsymbol{F}$，作为卷积神经网络的输入。

3.2.2 基于卷积神经网络的特征学习过程和情感倾向分类

本文使用的卷积神经网络分为输入特征层、卷积层、池化层、全连接层、dropout 层和 Softmax 层。卷积层有 3 种大小的卷积核，每种大小的卷积核有 128 个卷积滤波器进行 $\boldsymbol{F}$ 的局部特征学习。由于股吧评论数据为一维数据，假设输入特征层为 $d\times n\times 1$ 的张量，那么其卷积操作如公式(2)所示[16]：

$$\boldsymbol{c}_i=f\left(\boldsymbol{\omega}\cdot\boldsymbol{x}_{i:i+h-1}\right) \tag{2}$$

其中，$\boldsymbol{c}_i$ 为卷积运算的结果；$\boldsymbol{\omega}$ 为卷积核；$\boldsymbol{x}_{i:i+h-1}$ 为第 i 行到第 $i+h-1$ 组成的滑动窗口；f 为非线性函数。卷积核的宽度和语义特征 $\boldsymbol{B}$ 的维度一致，卷积核的高度 h 代表每次滑动窗口覆盖用于卷积的词数。$\boldsymbol{\omega}$ 和 $i+h-1$ 维度一致，即 $\boldsymbol{\omega}$ 的大小为 $h\times n\times 1$。经过卷积层输出的特征向量大小为 $(d-h+1)\times n\times 1$。

池化层采用的是最大池化方法，首先在保持特征的情况下进行参数的压缩，然后将这些特征拼接起来用向量表示，最后通过全连接层对池化层的输出进行整合。为了防止过拟合，加入 dropout 层和正则化层，再通过 Softmax 分类器得到分类结果，其计算公式如公式(3)所示[16]：

$$Softmax\left(z_i\right)=\frac{\mathrm{e}^{z_i}}{\sum_{i=1}^{n}\mathrm{e}^{z_i}} \tag{3}$$

其中，$Softmax$ 为分类函数；z_i 为第 i 个节点的输出值；n 为输出节点的个数；e 为自然底数。损失函数采用的是交叉熵损失函数，其计算公式如公式(4)所示[16]：

$$Loss=\frac{1}{N}\sum_i L_i=\frac{1}{N}\sum_i\sum_{c=1}^{C}y_{ic}\log\left(p_{ic}\right) \tag{4}$$

其中，$Loss$ 为损失；N 为样本个数；i 为第 i 个样本；L_i 为第 i 个样本的损失；C 为类别个数；c 为第 c 类；y_{ic} 为分类准确性，若样本 i 的类别和样本 i 的真实类别相同则为 1，否则为 0；p_{ic} 为

样本 i 属于类别 c 的概率。

4　股吧评论情感分析结果与评估

4.1　数据集与评价指标

“东方财富网-股吧”是国内金融领域最活跃的股票讨论平台。首先从“东方财富网-股吧”的股票评论中随机选取 15 747 条评论，然后经金融专业人士对其进行标注后用于测试和训练。本文使用 4 种常用的评价指标：准确率(Accuracy，Acc)、查准率(Precision，P)、召回率(Recall，R)和 F1 值[17]，其中查准率、召回率和 F1 值均为宏平均。评价指标的计算公式如公式(5)~(8)所示[18]：

$$Acc=\frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

$$P=\frac{TP}{TP+FP} \tag{6}$$

$$R=\frac{TP}{TP+FN} \tag{7}$$

$$F1=\frac{2P\bullet R}{P+R} \tag{8}$$

其中，Acc 为准确率；P 为查准率；R 为召回率；$F1$ 为 F1 值；TP 为真阳性的数量；TN 为真阴性的数量；FP 为假阳性的数量；FN 为假阴性的数量。查准率、召回率和 F1 值的宏平均计算公式如公式(9)～(11)所示[17]：

$$Macro_P=\frac{1}{n}\sum_{i=1}^{n}P_i \tag{9}$$

$$Macro_R=\frac{1}{n}\sum_{i=1}^{n}R_i \tag{10}$$

$$Macro_{F1}=\frac{1}{n}\sum_{i=1}^{n}F1_i \tag{11}$$

其中，$Macro_P$ 为查准率的宏平均；$Macro_R$ 为召回率的宏平均；$Macro_{F1}$ 为 F1 值的宏平均；n 为目标任务的类别数。

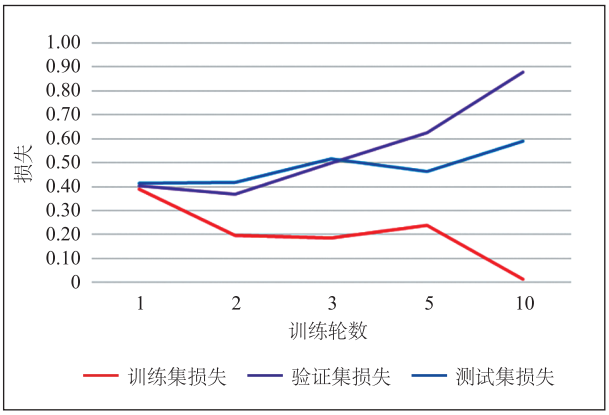
4.2　实验结果

本实验首先使用股吧评论标注数据集进行模型参数敏感性和有效性分析；然后将批大小、丢弃率、学习率、卷积核个数、训练次数和隐藏层个数作为变量，探究模型对其的敏感性以及其对模型的性能的影响；最后选择 5 种针对金融领域评论数据情感倾向分析的方法模型与基于 FinBERT-CNN 的股吧评论情感分析方法进行对比，包括基于词典的方法[5]、SVM[9]、TextCNN[16]、BERT[19]和 FinBERT[14]。本实验基于 FinBERT 和卷积神经网络的集成模型，模型搭建基于 TensorFlow 框架利用 Python 语言进行程序编写，程序运行环境为 Linux Ubuntu 18.04。

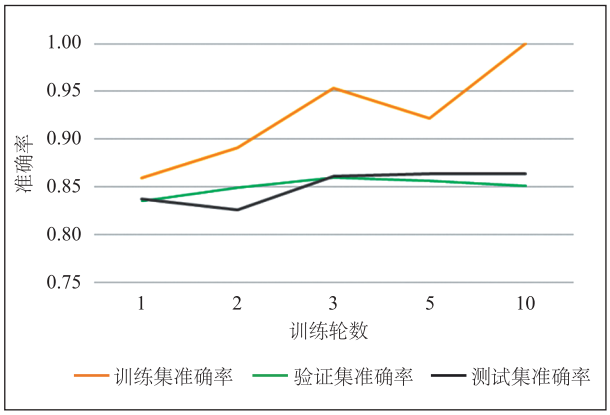
图 3 是在固定丢弃率为 0.5、卷积核个数为 128、隐藏层层数为 128 的条件下，训练轮数和批大小对模型的损失和准确率的影响结果示意图。模型训练轮数越多，学习数据特征就会越充分，但是可能会因为过度学习训练数据特征而产生过拟合现象，训练轮数太少则可能会出现学习数据特征不充分的情况[18]。批大小最大可以选择整个数据集的大小，但是可能会导致运行内存不足等问题，反之则可能导致数据不收敛[17]。

由图 3(a)和图 3(b)可知，训练轮数越多，训练集损失越小，测试集损失越大；训练集、验证集和测试集的准确率均随着训练轮数的增加而略有提升。由图 3(c)可知，训练集损失随着批大小的增加而增大，而测试集损失在批大小为 256 的时候最小。由图 3(d)可知，批大小越小，训练集准确率越高，而测试集准确率在批大小为 64 时最高。

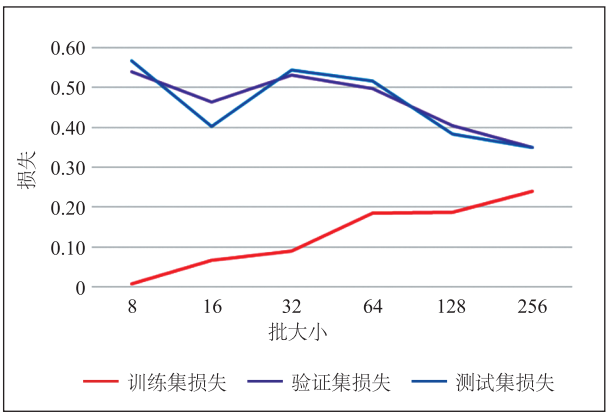
表 2 为丢弃率、卷积核个数、隐藏层层数对模型的影响结果。当改变丢弃率时，固定卷积核个数为 128，隐藏层层数为 128；当改变卷积核个数时，固定丢弃率为 0.5，隐藏层层数为 128；当改变隐藏层层数时，固定丢弃率为 0.5，



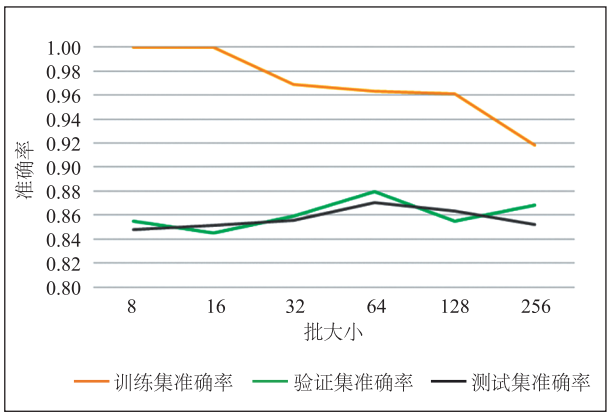
(a) 训练轮数对模型损失的影响



(b) 训练轮数对模型准确率的影响



(c) 批大小对模型损失的影响



(d) 批大小对模型准确率的影响

图 3 训练轮数和批大小对模型的损失和准确率的影响

Fig. 3 The effects of batch size and epochs on the model loss and accuracy

卷积核个数为 128。从表 2 中可以看出，丢弃率为 0.5 的时候，模型的效果最好；卷积核个数为 128 的时候，模型效果最好；隐藏层层数为 128 的时候，模型效果最好。这 3 个参数均和模型拟合程度有关[17]，由表 2 可知，当模型丢弃率为 0.5、卷积核个数为 128，隐藏层层数为 128 时模型效果最好。

表 3 为不同情感分析方法的测试数据集结果，其中 FinBERT-CNN 方法的测试参数为：批大小为 64，丢弃率为 0.5，卷积核个数为 128，隐藏层层数为 128，训练轮数为 3 轮，卷积核的大小分别是 2、3、4。

表 2 丢弃率、卷积核个数和隐藏层层数对模型的影响

Table 2 The influence of dropout, kernel number and hidden layer on the model

评价指标	丢弃率			卷积核个数			隐藏层层数		
	0.2	0.5	0.7	64	128	256	64	128	256
测试损失	0.500 0	0.516 9	0.367 5	0.378 9	0.516 9	0.369 1	0.474 4	0.516 9	0.384 4
测试准确率	0.833 3	0.870 4	0.856 8	0.840 3	0.870 4	0.848 8	0.850 4	0.870 4	0.851 7
测试查准率	0.848 4	0.868 6	0.854 3	0.841 3	0.868 6	0.844 9	0.857 7	0.868 6	0.850 5
测试召回率	0.826 5	0.868 7	0.859 1	0.836 8	0.868 7	0.855 3	0.846 2	0.868 7	0.851 6
测试 F1 值	0.827 9	0.865 0	0.850 8	0.831 6	0.865 0	0.842 8	0.844 9	0.865 0	0.845 6

由表 3 可知，基于词典的方法效果最差，4 个指标均不到 0.5，究其原因是其没有公开全面的高质量股票情感词典。SVM 的实验结果与基于词典的方法相比准确率提升大约 0.3，但由于其依赖特征质量，且需要已标注的数据，所以仅优于基于词典的方法。TextCNN 相较于基于词典的方法和 SVM 在 4 个评价指标上均有大幅提升，但稍逊于 BERT 和 FinBERT。FinBERT-CNN 在所有的模型中表现最好，相较于基于词典的方法准确率提升大约 0.4；与 BERT 和 FinBERT 的实验结果相比，FinBERT-CNN 的效果略有提升。

表 3　不同情感分析方法的测试数据集结果

Table 3　Results of different methods

方法	准确率	查准率	召回率	F1 值
基于词典的方法[5]	0.484 9	0.443 8	0.451 5	0.445 9
SVM[9]	0.799 9	0.821 2	0.790 6	0.800 7
TextCNN[16]	0.810 3	0.807 8	0.808 2	0.807 1
BERT[19]	0.855 4	0.843 8	0.846 4	0.844 9
FinBERT[14]	0.865 6	0.854 3	0.840 2	0.841 3
FinBERT-CNN	0.870 4	0.868 6	0.868 7	0.865 0

5　基于股吧评论情感的股票市场波动关联分析

国内外已有大量研究[22-28]表明投资者情感对股票收益变化具有一定影响，且对股市波动具有一定的预测作用。通过对投资者评论情感的有效分析，能够更好地了解我国股票市场波动。目前，国内外针对股票评论情感和市场波动之间的相关性研究，主要基于 A 股市场，缺乏对单只股票的研究。本文以“上证 50”成分股的股吧评论情感和市场波动为研究对象，探究单只股票的评论情感与市场波动之间的关系。

5.1　基于单只股票的股吧评论情感与股票波动关联分析

第 4.2 节验证了基于 FinBERT-CNN 的股吧评论情感分析方法的有效性，利用该模型对“上证 50”成分股的评论进行投资者投资情感倾向分析，并对分析结果进行市场波动关联讨论。“上证 50”，即挑选上海证券市场最具代表性的 50 只股票组成样本股，经一定规则计算后得到能反映上海证券市场的最具有市场影响力的一批龙头企业的整体状况指数[20]。本文选取 2018 年 1 月 1 日～2019 年 12 月 31 日作为研究时间，针对该时间段内“上证 50”成分股的股评和股票行情数据进行关联分析。其中，股评分析包括发帖量（CA）、情绪指数（BI）和意见分歧度（DI）3 个方面；股票行情数据统计包括收盘价（CP）、前一日收盘价（PCP）、交易量（Vol）、成交总额（Am）和涨跌幅（pctChg）6 种。发帖量代表了这只股票的评论热度。情绪指数采用 Antweiler & Frank 情绪指数构建方式[21]，计算公式如下：

$$BI=\frac{M^{\text{bull}}-M^{\text{bear}}}{M^{\text{bull}}+M^{\text{bear}}} \tag{12}$$

其中，M^{bull} 为当日股吧中看涨的评论数量；M^{bear} 为当日股吧中看跌的评论数量；BI 为情绪指数。意见分歧度则表示投资者们发表的意见之间表达的不同情绪，计算公式如下：

$$DI_t=\sum_{i\in D(t)}\left(e_i-BI_t\right)^2 \tag{13}$$

其中，DI_t 为 t 日的意见分歧度；$D(t)$为交易日 t 的评论发表时间集合；e_i 为第 i 时间的评论的情感情绪；BI_t 为第 t 日的情绪指数；i 为评论发表的时间；t 为交易日。

图 4 为发帖量、情绪指数和意见分歧度与股票行情数据之间的相关性热力图。图中颜色越接近红色，表明变量间的正相关性越强，反之则表明负相关性越强。从图 4 中可以看到，每日发帖量与交易量和成交总额之间呈强正相关性，与每日涨跌幅之间存在负相关性；情绪指数与涨跌幅存在强正相关性，意见分歧度与交易量和成交总额之间存在正相关性。

以贵州茅台（600519）为例，在 2018 年 1 月 1 日～2019 年 12 月 31 日期间共计发帖数为 88 542

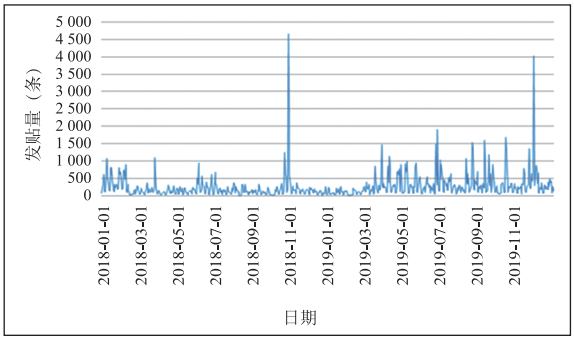
	CA-CP	CA-PCP	CA-Vol	CA-Am	CA-pctChg	BI-CP	BI-PCP	BI-Vol	BI-Am	BI-pctChg	DI-CP	DI-PCP	DI-Vol	DI-Am	DI-pctChg
宝钢股份	0.23	0.24	0.59	0.54	-0.08	0.20	0.17	0.18	0.17	0.33	0.31	0.31	0.30	0.29	0.03
中国平安	0.30	0.33	0.50	0.55	-0.19	0.27	0.21	0.11	0.19	0.52	0.20	0.18	0.19	0.24	0.17
保利地产	0.29	0.31	0.22	0.26	-0.11	0.27	0.17	0.16	0.18	0.45	0.08	0.08	0.25	0.24	0.01
北京银行	-0.22	-0.21	0.14	0.09	-0.09	0.07	0.05	0.04	0.06	0.16	0.20	0.20	0.33	0.34	0.03
大秦铁路	0.24	0.30	0.48	0.45	-0.29	0.00	0.05	0.05	0.03	0.26	0.35	0.37	0.42	0.42	-0.09
复星医药	-0.06	-0.03	0.49	0.39	-0.22	0.02	0.04	0.11	0.09	0.51	0.14	0.13	0.16	0.20	0.13
工商银行	0.42	0.45	0.57	0.60	-0.15	0.06	0.01	0.02	0.02	0.36	0.22	0.19	0.39	0.38	0.16
工业富联	0.36	0.36	0.76	0.75	0.00	0.33	0.24	0.36	0.35	0.49	0.15	0.10	0.38	0.34	0.26
光大银行	0.25	0.27	0.50	0.49	-0.06	0.07	0.03	0.08	0.07	0.20	0.26	0.25	0.36	0.35	0.01
国泰君安	0.40	0.38	0.65	0.64	0.08	0.10	0.05	0.04	0.04	0.24	0.34	0.33	0.32	0.33	0.07
海尔智家	0.15	0.19	0.47	0.43	-0.22	0.12	0.04	0.08	0.10	0.43	0.34	0.32	0.30	0.32	0.11
海螺水泥	0.03	0.05	0.46	0.47	-0.16	0.24	0.18	0.07	0.14	0.40	0.09	0.09	0.20	0.22	0.04
恒瑞医药	0.38	0.40	0.51	0.67	-0.14	0.26	0.19	0.15	0.06	0.39	0.24	0.23	0.14	0.23	0.08
华泰证券	0.30	0.33	0.57	0.57	-0.12	0.05	0.04	0.09	0.09	0.43	0.27	0.24	0.32	0.32	0.14
华夏幸福	0.18	0.21	0.63	0.55	-0.12	0.09	0.04	0.06	0.07	0.29	0.11	0.11	0.35	0.32	0.00
建设银行	0.18	0.19	0.51	0.50	-0.08	0.07	0.00	0.05	0.05	0.34	0.10	0.10	0.38	0.36	0.02
交通银行	0.23	0.24	0.48	0.46	-0.04	0.16	0.12	0.14	0.15	0.22	0.13	0.13	0.21	0.21	-0.03
洛阳钼业	0.46	0.46	0.65	0.68	0.03	0.12	0.08	0.34	0.32	0.48	0.09	0.07	0.30	0.24	0.22
绿地控股	0.72	0.72	0.89	0.89	0.03	0.01	0.05	0.03	0.04	0.36	0.29	0.26	0.35	0.33	0.13
民生银行	0.32	0.33	0.19	0.23	-0.12	0.06	0.07	0.03	0.02	0.24	0.08	0.08	0.26	0.28	0.06
南方航空	0.37	0.38	0.76	0.73	-0.09	0.02	0.02	0.10	0.07	0.40	0.22	0.21	0.28	0.27	0.11
农业银行	0.25	0.27	0.53	0.51	-0.11	0.03	0.03	0.11	0.10	0.31	0.23	0.23	0.30	0.29	0.01
浦发银行	0.13	0.15	0.59	0.58	-0.14	0.09	0.04	0.00	0.00	0.28	0.15	0.14	0.23	0.22	0.07
三安光电	-0.34	-0.33	0.76	0.64	-0.06	0.10	0.05	0.29	0.32	0.46	-0.05	-0.08	0.37	0.36	0.27
三六零	0.78	0.79	0.70	0.86	-0.09	0.05	0.01	0.07	0.03	0.40	0.18	0.17	0.39	0.33	0.16
山东黄金	0.62	0.61	0.77	0.82	0.10	0.25	0.19	0.21	0.22	0.42	0.37	0.34	0.38	0.39	0.20
上海银行	0.32	0.32	0.45	0.51	-0.01	0.03	0.05	0.04	0.01	0.23	0.32	0.32	0.27	0.33	0.05
上汽集团	-0.30	-0.28	0.51	0.41	-0.19	0.25	0.21	0.08	0.13	0.39	0.22	0.22	0.17	0.21	0.00
万华化学	0.08	0.09	0.71	0.69	-0.05	0.05	0.00	0.02	0.00	0.31	0.25	0.24	0.53	0.52	0.08
新华保险	-0.08	-0.03	0.48	0.39	-0.25	0.12	0.06	0.13	0.14	0.31	-0.08	-0.08	0.13	0.11	-0.01
兴业银行	0.08	0.10	0.68	0.66	-0.11	0.15	0.10	0.04	0.05	0.34	0.20	0.19	0.34	0.33	0.10
药明康德	0.42	0.40	0.66	0.75	0.10	0.07	0.01	0.01	0.00	0.32	0.27	0.25	0.24	0.23	0.04
伊利股份	0.03	0.09	0.73	0.69	-0.31	0.23	0.15	0.01	0.06	0.40	0.08	0.06	0.09	0.10	0.12
招商银行	0.11	0.14	0.59	0.60	-0.21	0.22	0.17	0.02	0.07	0.30	0.23	0.23	0.22	0.26	0.01
中国国旅	0.34	0.37	0.41	0.61	-0.26	0.05	0.05	0.07	0.04	0.03	0.25	0.25	0.07	0.17	-0.01
中国建筑	-0.11	-0.10	0.19	0.11	-0.07	0.06	0.04	0.18	0.16	0.37	-0.09	-0.09	0.14	0.10	0.10
中国交建	-0.53	-0.53	0.37	0.23	-0.01	0.30	0.30	0.20	0.24	0.01	0.04	0.01	0.27	0.26	0.19
中国联通	0.51	0.49	0.75	0.75	0.07	0.19	0.09	0.28	0.27	0.45	0.28	0.22	0.46	0.45	0.27
中国人寿	0.28	0.28	0.42	0.42	0.01	0.13	0.09	0.14	0.14	0.28	0.17	0.17	0.40	0.39	0.03
中国神华	0.28	0.32	0.56	0.53	-0.22	0.13	0.08	0.01	0.03	0.28	0.21	0.21	0.33	0.31	-0.01
中国石化	0.08	0.12	0.47	0.41	-0.29	0.09	0.03	0.01	0.02	0.43	0.17	0.16	0.33	0.33	0.05
中国石油	-0.38	-0.37	0.48	0.37	-0.10	0.05	0.09	0.01	0.00	0.34	-0.01	-0.02	0.30	0.28	0.14
中国太保	-0.07	-0.01	0.50	0.45	-0.24	0.15	0.07	0.00	0.02	0.33	0.02	0.01	0.22	0.21	0.02
中国铁建	0.16	0.16	0.61	0.59	-0.02	0.17	0.09	0.14	0.14	0.36	0.17	0.14	0.34	0.33	0.14
中国银行	0.38	0.40	0.54	0.54	-0.11	0.07	0.12	0.02	0.03	0.26	0.26	0.25	0.36	0.35	0.09
中国中车	0.51	0.53	0.66	0.66	-0.20	0.14	0.09	0.19	0.18	0.42	0.23	0.21	0.36	0.34	0.24
中国中铁	-0.06	-0.03	0.65	0.63	-0.21	0.17	0.15	0.02	0.00	0.17	0.11	0.11	0.26	0.27	0.01
中国重工	0.58	0.58	0.78	0.81	0.04	0.01	0.07	0.29	0.27	0.39	0.10	0.07	0.38	0.35	0.23
中信证券	0.31	0.34	0.51	0.54	-0.12	0.13	0.20	0.16	0.13	0.45	0.25	0.20	0.33	0.31	0.27
贵州茅台	0.25	0.26	0.43	0.56	-0.17	0.06	0.02	0.14	0.10	0.49	0.05	0.03	0.03	0.04	0.17

注：CA-CP 表示发帖量与收盘价之间的关系，CA-PCP 表示发帖量与前一日收盘价之间的关系，CA-Vol 表示发帖量与交易量之间的关系，CA-Am 表示发帖量与成交总额之间的关系，CA-pctChg 表示发帖量与涨跌幅之间的关系；BI-CP 表示情绪指数与收盘价之间的关系，BI-PCP 表示情绪指数与前一日收盘价之间的关系，BI-Vol 表示情绪指数与交易量之间的关系，BI-Am 表示情绪指数与成交总额之间的关系，BI-pctChg 表示情绪指数与涨跌幅之间的关系；DI-CP 表示意见分歧度与收盘价之间的关系，DI-PCP 表示意见分歧度与前一日收盘价之间的关系，DI-Vol 表示意见分歧度与交易量之间的关系，DI-Am 表示意见分期度与成交总额之间的关系，DI-pctChg 表示意见分歧度与涨跌幅之间的关系(同表 4)

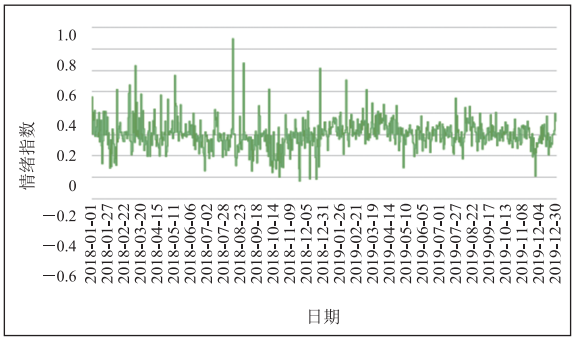
图 4 发帖量、情绪指数和意见分歧度与股票行情数据之间的相关性热力图

Fig. 4 Heap map of correlations between three indexes and stock data

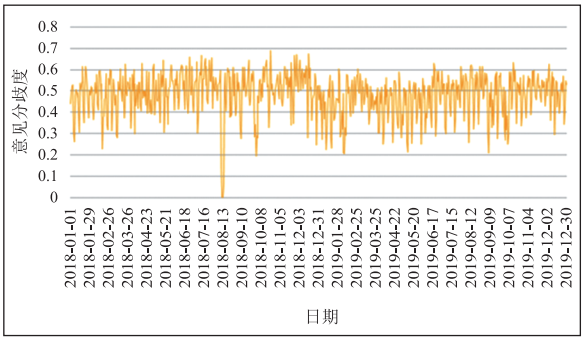
条。该时间段内贵州茅台吧的每日发帖量、情绪指数和意见分歧度统计结果如图 5 所示，由图 5(a)可知，在 2018 年 10 月 29 日和 2019 年 11 月 29 日发帖量剧增，分别达到了 4 659 条和 4 021 条。由图 5(b)可知，在 2018 年 8 月 11 日情绪指数达到最高，而在 2018 年 11 月 24 日投资者们的情绪最为低落。由图 5(c)可知，在 2018 年 10 月 22 日投资者们意见分歧最大。



(a) 贵州茅台(600519)发贴量



(b) 贵州茅台(600519)情绪指数



(c) 贵州茅台(600519)意见分歧度

图 5　贵州茅台吧的每日发帖量、情绪指数和意见分歧度

Fig. 5　Daily posts, BI and DI of Guizhou Moutai

表 4 为贵州茅台吧在 2018 年 1 月 1 日～2019 年 12 月 31 日期间的发帖量、情绪指数和意见分歧度与股票变动的相关系数表。由表可知，成交总额与发帖量正相关性最强，其次是情绪指数与涨跌幅，而涨跌幅与发帖量呈负相关性。

表 4　贵州茅台吧的发帖量、情绪指数和意见分歧度与股票变动的相关系数表

Table 4　Correlation coefficient table of correlations between three indexes of Guizhou Moutai

CA-CP	CA-PCP	CA-Vol	CA-Am	CA-pctChg
0.25	0.26	0.43	0.56	−0.17
BI-CP	BI-PCP	BI-Vol	BI-Am	BI-pctChg
0.06	0.02	0.14	0.10	0.49
DI-CP	DI-PCP	DI-Vol	DI-Am	DI-pctChg
0.05	0.03	0.03	0.04	0.17

综上所述，股市的波动与股吧发帖量具有强相关性，其中，股票的交易量和成交总额与发帖量呈正相关，涨跌幅与发帖量呈负相关。股票每日的涨跌幅与情绪指数和意见分歧度具有比较明显的相关性，而收盘价、前一日的收盘价、成交量和成交总额与情绪指数和意见分歧度的相关性并不大。

5.2　讨论与分析

实验结果表明，本文提出的基于 FinBERT-CNN 股吧评论情感分析方法能有效解决现有模型缺少优质的评论标注数据集用于训练、难以处理专业词汇和提取特征不全面的问题。与现有情感分析方法相比，该模型能够自动学习股吧评论中的语义特征和局部特征，处理股票评论中的专业词汇，全面地学习股票评论特征，从而提高对股吧评论情感分类的有效性。目前，针对投资者情感与股市波动关系的研究大多基于 A 股市场，本文则以“上证 50”成分股评论情感作为研究对象，证明了单只股票投资者情感与股市波动也存

在相关性，为研究投资者情感对股票市场预测、股票交易策略的构建等任务具有重要意义的结论提供了佐证。

6 结 论

本文针对现有模型存在缺乏优质的股票评论标注数据集用于模型训练，无法处理专业词汇，特征提取过于单一等问题，提出一种基于 FinBERT-CNN 的情感分析方法对股吧评论进行投资者情感分析。系统首先利用预训练模型 FinBERT 对评论提取语义特征，然后利用卷积神经网络提取局部特征，最后通过 Softmax 分类器输出情感倾向。实验结果表明，基于 FinBERT-CNN 的情感分析方法的准确率达到了 0.870 4，均优于现有模型。此外，本文还对单只股票评论情感与市场波动的相关性进行了验证。在下一步的研究中，将结合股票评论数据情感分析结果进行股票交易方法的构建，并提出高效的股票交易策略方法。

参 考 文 献

[1] 王道平, 贾昱宁. 投资者情绪与中国股票市场过度波动 [J]. 金融论坛, 2019, 24(7): 46-59.
Wang DP, Jia YN. Investor sentiment and excess volatility of Chinese stock markets [J]. Finance Forum, 2019, 24(7): 46-59.

[2] Hudson R, Muradoglu YG. Personal routes into behavioural finance [J]. Review of Behavioral Finance, 2020, 12(1): 1-9.

[3] Pang B, Lee L. Opinion mining and sentiment analysis [J]. Foundations and Trends in Information Retrieval, 2008, 2(1-2): 1-135.

[4] Maqsood H, Mehmood I, Maqsood M, et al. A local and global event sentiment based efficient stock exchange forecasting using deep learning [J]. International Journal of Information Management, 2020, 50: 432-451.

[5] 陈雷. 面向股票评论的情感分析系统研究与实现 [D]. 杭州: 浙江工商大学, 2017.
Chen L. Research and implementation of sentiment analytic system for stock reviews [D]. Hangzhou: Zhejiang Gongshang University, 2017.

[6] Alkubaisi G, Kamaruddin SS, Husni H. Stock market classification model using sentiment analysis on Twitter based on Hybrid Naive Bayes Classifiers [J]. Computer and Information Science, 2018, 11(1): 52-64.

[7] Yazdani SF, Murad M, Share FNM, et al. Sentiment classification of financial news using statistical features [J]. International Journal of Pattern Recognition and Artificial Intelligence, 2017, 31(3): 165-176.

[8] Salles T, Goncalves M, Rodrigues V, et al. Improving random forests by neighborhood projection for effective text classification [J]. Information Systems, 2018, 77(9): 1-21.

[9] 张对. 网络股评影响股市走势吗——基于股票情感分析的视角 [J]. 现代经济信息, 2015, (1): 355-357.
Zhang D. Does online stock review affect the trend of stock market——based on the perspective of stock sentiment analysis [J]. Modern Economic Information, 2015, (1): 355-357.

[10] Jiang M, Wang J, Man L, et al. An effective gated and attention-based neural network model for fine-grained financial target-dependent sentiment analysis [C] // International Conference on Knowledge Science, Engineering and Management, 2017: 42-54.

[11] Rao G, Huang W, Feng Z, et al. LSTM with sentence representations for document-level sentiment classification [J]. Neurocomputing, 2018, 308: 49-57.

[12] Sohangir S, Wang D, Pomeranets A, et al. Big Data: deep learning for financial sentiment analysis [J]. Journal of Big Data, 2018, 5(1): 1-25.

[13] Akhtar MS, Kumar A, Ghosal D, et al. A Multilayer Perceptron based Ensemble technique for fine-grained financial sentiment analysis [C] // Proceedings of the 2017 Conference on Empirical

Methods in Natural Language Processing, 2017: 540-546.

[14] Liu Z, Huang D, Huang K, et al. FinBERT: a pre-trained financial language representation model for financial text mining [C] // Twenty-Ninth International Joint Conference on Artificial Intelligence and Seventeenth Pacific Rim International Conference on Artificial Intelligence, 2020: 4315-4519.

[15] Medhat W, Hassan A, Korashy H. Sentiment analysis algorithms and applications: a survey [J]. Ain Shams Engineering Journal, 2014, 5(4): 1093-1113.

[16] Kim Y. Convolutional neural networks for sentence classification [J]. arXiv Preprint, arXiv: 1408.5882, 2014.

[17] Liu C, Wang W, Wang M, et al. An efficient instance selection algorithm to reconstruct training set for support vector machine [J]. Knowledge-Based Systems, 2017, 116(1): 58-73.

[18] Bishop CM. Pattern Recognition and Machine Learning (Information Science and Statistics) [M]. Springer-Verlag New York, Incorporated, 2006.

[19] Devlin J, Chang MW, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [J]. arXiv Preprint, arXiv: 1810.04805, 2018.

[20] 上海证券交易所. 上海证券交易所 [DB/OL]. 2015-12-19[2021-01-24]. http://www.sse.com.cn/.
Shanghai Stock Exchange. Shanghai Stock Exchange [DB/OL]. 2015-12-19[2021-01-24]. http://www.sse.com.cn/.

[21] Antweiler W, Frank MZ. Is all that talk just noise? The information content of internet stock message boards [J]. The Journal of Finance, 2004, 59(3): 1259-1294.

[22] 孟志青, 郑国杰, 赵韵雯. 网络投资者情绪与股票市场价格关系研究——基于文本挖掘技术分析 [J]. 价格理论与实践, 2018, (8): 127-130.
Meng ZQ, Zheng GJ, Zhao YW. The research on the relationship between network investor emotion and stock market price empirical analysis based on text mining technology [J]. Price: Theory and Practice, 2018, (8): 127-130.

[23] 张信东, 原东良. 基于微博的投资者情绪对股票市场影响研究 [J]. 情报杂志, 2017, 36(8): 81-87.
Zhang XD, Yuan DL. Research on the impact of investor sentiment on stock market based on microblog [J]. Journal of Intelligence, 2017, 36(8): 81-87.

[24] 查文媛. 我国投资者情绪对股票收益的影响研究——基于上证A股市场研究 [C] // 第三届社会科学与经济发展国际学术会议, 2018: 544-547.
Zha WY. Research on effects of Chinese investor sentiment on stock return—study on Shanghai A-Share market research [C] // 2018 3rd International Conference on Society Science and Economics Development, 2018: 544-547.

[25] 蒋钰慧. 投资者情绪对我国股票市场收益率的影响研究——基于微博文本数据的视角 [D]. 上海: 上海外国语大学, 2019.
Jiang YH. Research on the the influence of investor sentiment on the return rate of China's stock market return——based on the perspective of Weibo text data [D]. Shanghai: Shanghai International Studies University, 2019.

[26] Li Y, Ran J. Investor sentiment and stock price premium validation with Siamese twins from China [J]. Journal of Multinational Financial Management, 2020: 57-58.

[27] Chen RD, Bao WW, Jin CL. Investor sentiment and predictability for volatility on energy futures markets: evidence from China [J]. International Review of Economics and Finance, 2021, 75(2): 112-129.

[28] Yang CP, Wu HH. Investor sentiment with information shock in the stock market [J]. Emerging Markets Finance and Trade, 2021, 57(2): 510-524.