

# 情感分析在股票市场中的应用研究综述

胡冰玲

天津商业大学, 天津

收稿日期: 2021年11月9日; 录用日期: 2021年11月23日; 发布日期: 2021年12月7日

---

## 摘要

随着信息技术的发展, 利用情感分析方法挖掘股票评论信息, 通过研究股民对股票涨跌趋势做出的评论来研究股票市场的发展规律, 可以更好的了解股票市场。本文首先从影响股票市场走势的几大因素中, 提出了股票评论对股票市场的重要作用, 而分析股票评论则要用到情感分析技术。然后分别介绍了股票市场中, 传统的情感分析方法与情感分析方法的现状。最后, 发现在对股票评论的情感分析方法中存在自动化程度较低、股评情感词典较少等有待完善之处。

## 关键词

情感分析, 股票价格, 文本挖掘, 股票市场

---

# A Summary of Research on the Application of Sentiment Analysis in the Stock Market

Bingling Hu

Tianjin University of Commerce, Tianjin

Received: Nov. 9<sup>th</sup>, 2021; accepted: Nov. 23<sup>rd</sup>, 2021; published: Dec. 7<sup>th</sup>, 2021

---

## Abstract

With the development of information technology, the use of sentiment analysis methods to mine stock comment information, and to study the development laws of the stock market by studying the comments made by stockholders on the stock rise and fall trends, can better understand the stock market. This article first puts forward the important role of stock commentary on the stock

market from several factors that affect the trend of the stock market, and sentiment analysis technology is used to analyze stock commentary. Then respectively introduced the current situation of traditional sentiment analysis methods and sentiment analysis methods in the stock market. Finally, it is found that there is a low degree of automation in the sentiment analysis method for stock reviews, and there are fewer stock review sentiment dictionaries that need to be improved.

## Keywords

Sentiment Analysis, Stock Price, Text Mining, Stock Market

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 前言

情感分析可以基于不同的文本语料类型，可以是新闻文章也可以是评论对话等，在不同的领域需要达到的效果不同。由此，不同的文本类型和不同的文本长度篇幅，都有着不同的研究方法。由于，词和短语往往不能表示完整的语义，相对而言，句子是我们处理的主要对象。投资者作为股票市场的重要参与者，其情绪变化会迅速影响到股市的变化。进一步了解投资者情绪，对股市行情进行判断并做出投资，也有助于解释股票市场趋势以及走向。相比于长文本来讲，评论文本更加精练，因为要在简短的文字中表达出自己的想法，所以蕴含了更丰富的情感。目前，网络媒体已经成为人们发表意见看法交流的主要平台之一，对网络评论进行预测和分析对我们做出决策有着重要的作用。我国股票市场经过了二十多年的飞速发展，各方面实力都在稳步上升。但对于投资者来说无法准确的获取股市的预期走势以及公司主体的发展趋势，对股市的消息主要来源于网络社交平台，信息的真假难以辨别，对股票市场产生了一定的负面影响。但是，从股票网站中的股民评论可以发现，评论信息很大程度上反映了股市行情，也对股市的涨跌有一定影响。因此，快速有效的分析股民的股票评论，掌握他们的情感观点对股市的预测有着重要的指导作用。

影响股市走势的因素包括：宏观经济因素、行业发展因素、公司内部因素、股民情感因素[1]。因为受消息、股票走势、股民自身心理素质等影响，股民影响因素相比于其他几种来说更具有易变性。尤其是在微博、微信等社交媒体的兴起，股民在线参与股评的意愿越发强烈。网络中股民的评论能够反映整个社会对股市的看法[2]，用情感分析方法对评论进行语义分析，利用文本挖掘技术识别股评的情感倾向性[3]。股民则不仅是信息的使用者，也成为了信息的发布者。由此，股票评论包含了丰富的情感信息，为股市的研究提供了技术基础[4]。早期研究主要依据经典资本资产定价理论[5] [6]，随着现代经济的发展，对股票评论的研究从股票市场的发展趋势逐步转变为股民情感对市场的影响[7] [8]。研究通过选取直接指标和间接指标来衡量投资者情感，间接指标有交易量、流动性等[9]，而直接指标则多以问卷调查的形式来收集信息[10]。中国股民情感研究起步较晚，直接情感指数多反映机构对股市大盘的看法，面向广大股民的情感统计较少，无法代表市场。间接情感指数选取的间接指标难以评价其真实性，而且相关指标的数据也无法准确定义[11]。因此，利用情感分析方法应用于股票市场成为重要研究方向，股票市场中的主要情感分析方法如图 1 所示：

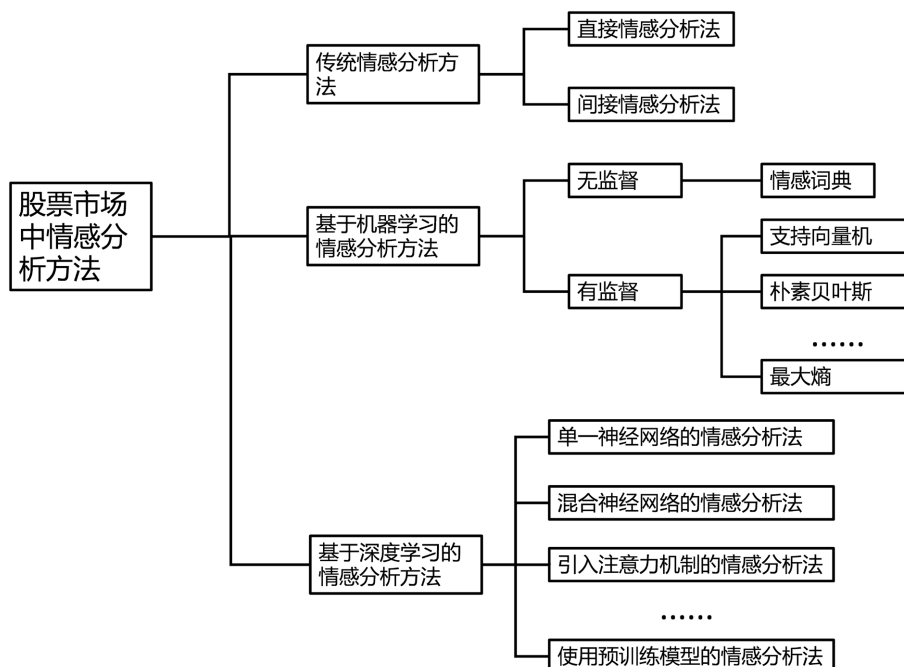


Figure 1. The sentiment analysis method of stock market  
图 1. 股票市场情感分析方法

## 2. 传统的股票市场情感分析方法

人类社会的出现伴随着语言文字的诞生，随着人类文明的发展，语言数据形式和内容越来越多样，互联网的产生让语言数据有了量上的质变。日常生活中产生的数据大约 80% 的信息是用语言负载的，没有用语言负载的信息也需要语言来帮助阐释，比如图画、音乐、建筑等艺术[12]。就股票投资而言，股民通过网络社交媒体获取股票相关信息，并据此对股市行情进行判断并做出投资决策，同时，网络股评这种形式不仅只是专业金融人士发表的专业意见，更多的还有普通股民对股市的看法，更多的反映了社会对整个股市的情感。所以，想要从不同的股评文本数据中获得有用的信息，就需要成熟的自然语言处理技术。

情感分析中，传统的文本分类方法，是基于规则的无监督网络情感分类方法，不依赖于人工标注的语料集，这类方法主要是采用正负情感词典作为种子词典，通过与情感词典进行情感词匹配来判断情感极性，然后通过相关规则对评论文本进行分类，但是股票评论没有标准的词库，在股票领域的情感词典研究相对较少，情感词库也不够全面，无法达到很好的分类效果。

已有文献中传统的投资者情感极性分析方法大致有三种[13]：第一种，通过市场调查得到股民对股票市场未来走势的看法，例如我国的央视看盘。市场调查是统计学中常见的直接反映事实的方法，但问卷调查十分主观，不能成为主要的分析方法。第二种，采用可以度量的经济指标来观测投资者情绪指数，例如首日评论收益率等，在美国的股票市场研究中应用十分广泛。第三种，基于社交网站、新闻媒体、股吧留言板等网络平台，利用文本挖掘方法，提取和构建投资者情绪指数判断情感极性。Antweiler 和 Frankl [14]选择了雅虎财经和 RagingBull.com 的股票评论近 150 万条，挖掘并构建了股民情感指数。Chen [15]、Sprenger [16]和 Bartov [17]利用社交网络上发布的帖子构建了投资者情绪指标。大量研究表明，当市场收益率高时，股票评论信息乐观高涨，反之，股票评论情感信息又进一步作用于股市变化。

我国对股票评论的研究大多是对单一变量进行分析，或者采用 Baker 和 Wurgler [18]的方法，对股票市场的多个变量合成股评情感指数进行分析。张强[19]、文凤华等[20]则是基于多个反映国内股票市场股

民情绪的指标,应用主成分分析、因子分析、对应分析等传统统计学分类方法,构建股评情感指数,研究了股民情感对市场收益的影响。目前,我国市场上利用文本挖掘方法对投资者情绪构建情感分析的研究较少,有待研究补充。

### 3. 股票市场情感分析方法现状

从研究现状来看,自然语言理解和处理的理论逐渐完善,形成了科学的体系,在应用方面产生了很好的社会效益和经济效益。情感分析在自然语言处理中是最为重要的一个方向之一,这种对情感态度或文本倾向性识别的处理,可以针对不同用户评价对分析某一事件、政策法规或社会现象,最重要的是情感分析结果可以结合实时数据对未来某一事件做出预测。因此,股票市场的波动与股民的情感态度有着很大的关联,通过对投资者的评论数据进行情感分析与股票数据指标结合可以更好的为股民和公司提供预测与决策。

现有的有监督情感分析技术有,支持向量机(SVM)、朴素贝叶斯(NB)、最大熵(ME)、Logistic 回归等分类算法[21]。利用机器学习算法对标注后的文本数据进行分类器训练,得到相应的分类结果。这种情感分析方法依赖于分类标注的准确性,而不同的行业领域需要对不同的语料进行标注。但是,在众多领域都标注训练集十分的复杂困难,需要大量的时间成本和人力物力。所以,大规模语料标注的自动化开发与实现成为目前有监督情感分类技术的瓶颈问题。

部慧、解峥等[22]在投资者情绪对我国股票市场影响的研究中对数据进行人工标注,通过邀请 10 位金融从业人员进行情绪标注,确保每条股评有两位人员标注,对于标注结果相异的股评,请第三位金融从业人员标注,取三次标注结果中占多数标签作为该股评最终标签。这样的标注方法更具有专业性,分类结果也会相对准确。随着社会的发展互联网信息数据急剧增加,海量的数据积累可以用于学习,需要有效的处理大量的非结构化数据对文本情感分类。因此,产生了基于深度学习的循环神经网络(RNN)、卷积神经网络(CNN)和长短期记忆网络(LSTM)等分类技术。深度模型可以随着模型复杂度的增大而增大,更好的贴近数据的本质映射关系,达到更优的效果。股票评论者来自各行各业纷繁复杂,无法区分是否专业人士,也无法准确判定情感的倾向性。首先在获得股票评论后要要进行一定的预处理,然后利用不同的机器学习算法对处理后的数据进行训练构建分类器,得到评价模型后计算模型的准确率召回率,最后进行模型的评估与股票的预测。除了通过股票评论对进行股票市场整体走势进行预测外,还有直接使用统计分析方法对大量的股票数据建立模型,通过分析股票的历史数据来预测股市的波动和走势。将机器学习方法应用于情感分析是近年来的主流研究趋势,股票评论的情感分析一般过程如图 2 所示:

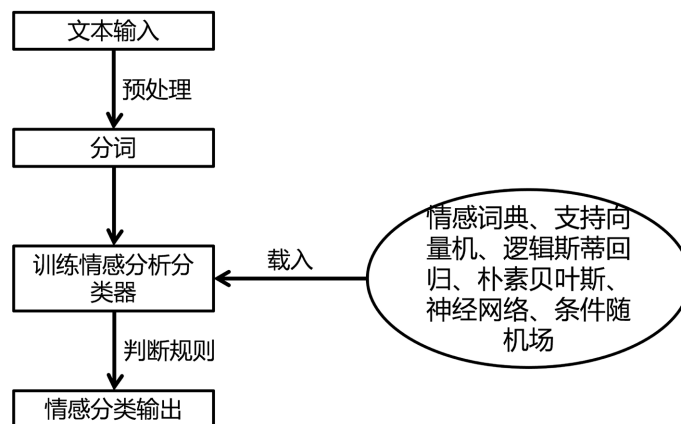


Figure 2. The general process of sentiment analysis methods  
图 2. 情感分析方法一般流程

在我国股票市场上, 普通股民受信息渠道以及知识水平等的限制, 很难做到完全理性投资。此外, 由于股评情感无法直接观测, 尤其是如何刻画投资者情感成为近年来一个重要的论题。针对网络股评影响股民决策以及股市走势的问题, 张对[23]利用新浪股吧这一网络股评发表及传播的载体进行研究分析, 通过收集相关的网络股评并提取情感指数, 建立 ARMAX-GARCH, 分析股评中情感因素与股市走势之间的关系, 将投资者情感作为 ARMAX-GARCH 模型的一个输入项来预测股票价格的波动。由此可知, 将情感分析方法应用于股票市场分析模型, 可以更好的分析股市走势, 对未来股票市场进行预测。

在股票市场中, 研究股票的涨跌往往是通过观察股票的波动与走势, 在网络上获取股票数据的基础上进行评论和预测。单一的对股票评论信息进行情感分析不能得到完整的股票涨跌趋势, 利用文本挖掘方法获取股评数据信息, 融合不同的数学模型, 才能有效的对股票市场进行分析。Bu 和 Pi [24]对学术文献中出现的刻画投资者情绪的方法进行了归类, 发现 Baker 和 Wurgler [18]提出的利用几个市场变量合成投资者情绪指数的方法是学术研究中使用最多的。在如今“大数据”背景下, 如何将情感分析技术与股票市场有机的结合起来, 值得继续深入研究。部慧、解峥等[22]探讨了投资者情绪对我国股票市场的影响, 基于东方财富网股吧帖文与朴素贝叶斯方法, 提出融合股评看涨看跌预期和投资者关注程度的投资者情绪度量指标。进一步, 利用 Granger 因果检验、瞬时 Granger 因果检验、跨期回归分析等方法, 探讨了投资者情绪对我国股票收益率、交易量和波动性是否具有预测能力及影响。肖亭、林玲[25]等通过研究分析不同专业人士发布股评的情感极性来预测股票上涨与下跌趋势。提出了一种综合金融词组词典和结尾段加权的情感分析方法, 能解决情感字典分析方法对领域依赖性问题, 有效地提高了情感分析准确度。另外, 论文还提出了一种加窗的股票预测模型, 可用于分析预测事件窗口的最佳值。实验结果表明, 基于股评情感分析来预测特定股票上涨或下跌趋势具有较好效果。

#### 4. 结论

各种情感分析方法虽然可以实现对股票评论的分类, 应用到股票市场趋势的大致预测, 但是预测准确率有待提高, 且自动化程度较低。这是由于目前我国股评情感分析仍处于不成熟的发展状态, 短期内的国民经济状况、宏观政策以及投资者的情绪等众多因素都会对股市造成一定程度的影响。因此, 在今后的研究中, 应综合考虑各方面因素, 实现自动化程度更高的情感分析方法。股票评论信息会由网络的发展而不断增加, 在后续情感分析方法的发展过程中, 需要不断地去探索, 深入研究股票市场的特征, 使得分析方法准确率更高, 扩大方法的适用性, 得到更好的预测精度。

#### 基金项目

天津市研究生科研创新项目人工智能专项(2020YJSZXS23)。

#### 参考文献

- [1] 宋敏晶. 基于情感分析的股票预测模型研究[D]: [硕士学位论文]. 哈尔滨: 哈尔滨工业大学, 2013.
- [2] Gillmor, D. (2004) *We the Media: Grassroots Journalism by the People, for the People*. O'Reilly Media, Inc., Sebastopol.
- [3] Baker, M.P. and Wurgler, J. (2007) Investor Sentiment in the Stock Market. *Journal of Economic Perspectives*, **21**, 129-152. <https://doi.org/10.1257/jep.21.2.129>
- [4] 马春林, 倪苏云, 吴冲锋. 股评家关注股票基本面因素: 股市信息供应量的影响因素分析[J]. 上海经济研究, 2002(5): 47-51.
- [5] 刘乃嘉, 陈泽龙, 杨宇. 时间序列分析在股票中的应用及其研究[J]. 财税金融, 2015(27): 30.
- [6] 池丽旭, 庄新田. 我国投资者情绪对股票收益影响——基于面板数据的研究[J]. 管理评论, 2011, 23(6): 41-48.
- [7] Baid, P., Gupta, A. and Chaplot, N. (2017) Sentiment Analysis of Movie Reviews Using Machine Learning Tech-



- niques. *International Journal of Computer Applications*, **179**, 45-49.
- [8] 黄少安, 刘达. 投资者情绪理论与中国封闭式基金折价[J]. 南开经济研究, 2005(4): 76-80+112.
- [9] Saunders, E.M. (1993) Stock Prices and Wall Street Weather. *The American Economic Review*, **83**, 1337-1345.
- [10] 刘维奇, 刘新新. 个人和机构投资者情绪与股票收益——基于上证 A 股市场的研究[J]. 管理科学学报, 2014, 17(3): 70-87.
- [11] Das, S.R. and Chen, M.Y. (2007) Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web. *Management Science*, **53**, 1375-1388. <https://doi.org/10.1287/mnsc.1070.0704>
- [12] 杨妥, 李万龙, 郑山红. 融合情感分析与 SVM\_LSTM 模型的股票指数预测[J]. 软件导刊, 2020, 19(8): 14-18.
- [13] 孙翌博, 濮泽堃, 徐玉华, 等. 应用 SD-LS-SVM 算法的评论情感分析模型[J]. 软件导刊, 2021, 20(4): 43-48.
- [14] Antweiler, W. and Frank, M.Z. (2004) Is All That Talk Just Noise? The Information Content of Interact Stock Message Boards. *The Journal of Finance*, **59**, 1259-1294. <https://doi.org/10.1111/j.1540-6261.2004.00662.x>
- [15] Chen, H., De, P., Hu, Y.J., et al. (2014) Wisdom of Crowds: The Value of Stock Opinions Transmitted through Social Media. *The Review of Financial Studies*, **27**, 1367-1403. <https://doi.org/10.1093/rfs/hhu001>
- [16] Sprenger, T.O., Tumasjan, A., Sandner, P.G., et al. (2014) Tweets and Trades: The Information Content of Stock Microblogs. *European Financial Management*, **20**, 926-957. <https://doi.org/10.1111/j.1468-036X.2013.12007.x>
- [17] Bartov, E., Faurel, L. and Mohanram, P.S. (2016) Can Twitter Help Predict Firm-Level Earnings and Stock Returns? Rotman School of Management Working Paper No. 2631421. <https://doi.org/10.2139/ssrn.2782236>
- [18] Baker, M. and Wurgler, J. (2006) Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance*, **61**, 1645-1680. <https://doi.org/10.1111/j.1540-6261.2006.00885.x>
- [19] 张强, 杨淑娥. 噪音交易、投资者情绪波动与股票收益[J]. 系统工程理论与实践, 2009, 29(3): 40-47.
- [20] 黄德龙, 文凤华, 杨晓光. 投资者情绪指数及中国股市的实证[J]. 系统科学与数学, 2009, 29(1): 1-13.
- [21] Kamal, A. and Abulaish, M. (2014) Statistical Features Identification for Sentiment Analysis Using Machine Learning Techniques. 2013 *International Symposium on Computational and Business Intelligence*, New Delhi, 24-26 August 2013, 178-181. <https://doi.org/10.1109/ISCBI.2013.43>
- [22] 部慧, 解峥, 李佳鸿, 吴俊杰. 基于股评的投资者情绪对股票市场的影响[J]. 管理科学学报, 2018, 21(4): 86-101.
- [23] 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.
- [24] Bu, H. and Pi, L. (2014) Does Investor Sentiment Predict Stock Returns? The Evidence from Chinese Stock Market. *Journal of Systems Sciences and Complexity*, **27**, 130-143. <https://doi.org/10.1007/s11424-013-3291-y>
- [25] 肖亭, 林玲, 黄永峰. 一种基于股票情感分析的股市趋势预测方法[J]. 电子技术应用, 2019, 45(3): 13-17.