

金融市场文本情绪研究综述^{*}

李合龙¹ 任昌松¹ 柳欣茹¹ 汪存华²

¹(华南理工大学经济与金融学院 广州 510006)

²(北华航天工业学院经济管理学院 廊坊 065000)

摘要:【目的】通过文献梳理分析和总结金融市场文本情绪发展现状,为后续研究提供参考。【文献范围】以“金融市场”“文本情感分析”“文本情绪”“投资者情绪”等以及相应英文为关键词在中国知网、Web of Science、谷歌学术等学术平台进行检索,延伸阅读相关文献,共筛选115篇文献进行综述。【方法】根据金融文本数据类型对提取的文本情绪分类,介绍文本情感分析框架,最后梳理文本情绪对金融市场影响的相关研究成果。【结果】金融文本情绪分为信息报告情绪、新闻媒体情绪和社交媒体情绪三种,在构造情绪指标时,应用较多的分析方法有基于词典的方法和基于机器学习的方法,三种文本情绪都对金融市场产生了一定的影响。【局限】筛选文本情感分析框架相关文献时,由于文本分析方法在各领域具有一定通用性,这类文献不完全聚焦于金融市场。【结论】在构建金融文本情绪指标时,应根据文本特点、研究条件、研究目标等的不同选择合适的情感分析方法。

关键词: 金融市场 文本情感分析 文本情绪 投资者情绪

分类号: F832 G350

DOI: 10.11925/infotech.2096-3467.2022.0890

引用本文: 李合龙, 任昌松, 柳欣茹等. 金融市场文本情绪研究综述[J]. 数据分析与知识发现, 2023, 7(12): 22-39. (Li Helong, Ren Changsong, Liu Xinru, et al. Review of Textual Sentiment Research in Financial Markets[J]. Data Analysis and Knowledge Discovery, 2023, 7(12): 22-39.)

1 引言

金融市场是一个非线性动态系统,而投资者情绪在这个系统中扮演着重要角色。已有大量文献尝试准确提取投资者情绪信息,研究其对市场的影响并纳入预测模型。度量投资者情绪的方法很多,主要可以分为市场指标法^[1-4]、直接调查法^[5-7]和文本分析法^[8-11]。然而,市场指标法和直接调查法在情绪测度方面存在较多缺陷:市场指标对情绪的反映不够直接和纯粹,还受到许多市场噪声的干扰;直接调查法的数据获取过程存在耗时费力、成本大、代表性

差、时效性弱、频率低等问题,在应用中还受到不同市场、不同投资者、不同地域等差异的制约,难以反映投资者对行业、个股等市场更细粒度的情绪。并且,这两种方法得到的指标均存在滞后性,不能实时地用于市场预测。

在实证分析中,通过上述三种方法构建的投资者情绪在股票市场中的表现亦有差异:受调查方法、对象的限制,通过直接调查法构建的情绪变量通常只代表某一类投资者的情绪,如个人投资者与机构投资者,无法对整个市场进行系统分析^[12];市场指标法构建的投资者情绪比较依赖所选代理指标的质

通讯作者(Corresponding author): 汪存华(Wang Cunhua), ORCID: 0009-0009-4153-220X, E-mail: forum300@163.com。

^{*}本文系中央高校基本科研业务费专项(项目编号: ZDPY202209)、广州市哲学社科规划2022年度课题(项目编号: 2022GZYB08)和北华航天工业学院博士基金(项目编号: BKY-2018-30)的研究成果之一。

The work is supported by the Fundamental Research Funds for the Central Universities(Grant No. ZDPY202209), the Guangzhou Philosophy and Social Science Planning Project(Grant No. 2022GZYB08), the Doctoral Fund of North China Institute of Aerospace Engineering (Grant No. BKY-2018-30).

量,若能选取合适的代理指标,由此构建的投资者情绪在对股市收益、风险以及预测等方面的影响都能取得不错的效果^[13-14],但同时部分指标也会带来限制,如新增开户数,这一类指标通常为低频数据,在构建高频投资者情绪时需要进行特殊处理,导致信息失真^[15];文本分析法首先对文本进行情绪提取,提取的结果通常代表投资者相对看涨、看跌的程度,再基于该结果构建情绪变量,包括情感倾向性与情感分歧性^[16],鉴于新闻媒体、社交网站等文本内容的时效性,基于文本分析的投资者情绪在对日内或日间高频收益率的预测能力上表现更为突出^[16-18]。

网络文本大数据为反映和研究投资者心理提供了宝贵而丰富的样本。一方面,随着线上交易的普及和网上股民数量的增加,投资者越来越倾向于在论坛、微博、微信等社交平台交流股票走势预期的投资意见。另一方面,文本分析技术的进步使从社交平台文本、公司财报、财经新闻等各类文本中获取情绪信息并构建情绪指标成为可能。因此,可以测度包含投资者情绪在内的更加广泛的文本情绪,拓展行为金融学的研究范围。与市场指标法和直接调查法相比,文本分析法在测度和研究文本情绪方面有以下优势:第一,论坛留言、社交动态等是投资者意见的直接表达,情绪指标内涵更丰富;第二,利用数据挖掘技术可在较短时间内获取海量文本数据,保证了样本的代表性、时效性,还可以降低人力、时间等成本;第三,可根据研究目标自由选择特定来源或主题的文本,如仅针对个股或某个行业分析;第四,对在线文本自由设定采样时间间隔,为得到更高频的情绪序列提供可能;第五,可获取、分析的信息不只有文本本身,其他诸如发布者的性别、所在地、发布时间等数据可随文本一同被捕获,从而使研究内容更丰富;第六,文本分析法构建的情感指标比其他两种方法有更强的预测能力^[19]。

目前,围绕“文本情绪如何影响金融市场”这一问题展开的综述研究并不多。唐国豪等^[20]综述了“金融市场文本情绪研究”,但是未能纳入基于深度学习的文本情绪分析方法,也未梳理文本情绪对除收益率外其他市场变量的影响;欧阳资生等^[21]综述了“网络舆情对金融市场的影响”,但是对网络舆情测度方法的介绍不够深入全面,也未能对以网络舆

情为代表的投资者情绪做进一步细分;沈艳等^[22]在综述“文本大数据分析在经济学和金融学中的应用”时,对文本情绪测度方法的理论介绍较为丰富,但未对新兴的深度学习法进行拓展和比较。本文补充了对深度学习模型和常用情感分析工具的比较和归纳,以及文本情绪应用于市场预测的最新研究。

本文梳理了近年来国内外相关文献,其中中文文献主要从中国知网检索获得,检索方式为先以“金融市场”“金融”为关键词进行初步检索,然后在结果中以“情绪”“文本”“文本分析”“投资者情绪”等为关键词进一步检索;外文文献从 Web of Science、谷歌学术检索获得,检索方式与中文文献的检索方式类似,把各关键词替换成对应的英文进行检索。在检索出文献的基础上,本文于 2022 年以期刊来源是否为北大中文核心期刊要目总览、中文社会科学引文索引 (Chinese Social Science Citation Index, CSSCI)、社会科学引文索引 (Social Science Citation Index, SSCI)、科学引文索引 (Science Citation Index, SCI) 等 (各版本当时的最新版本) 为标准,同时延伸阅读相关文献,最终获得与金融市场文本情绪有关的中文文献 47 篇,外文文献 68 篇。本文从研究方法和研究内容两个角度进行综述。研究方法上,整理了文本情绪指标构建的几种主要方法并进行比较;研究内容上,总结了信息报告情绪、新闻媒体情绪、社交媒体情绪这三种不同文本情绪对不同市场变量的影响机制和结果,并归纳了文本情绪在优化预测模型、改善投资决策中的应用进展。本文旨在呈现对以文本形式表达和传递的情绪和金融市场之间关系的更全面、更前沿的认识,并为今后行为金融学领域的相关研究提供思路。

本文内容框架如图 1 所示。

2 文本大数据和文本情绪梳理及分类

在金融文本情绪研究中,涉及的文本信息可分为三类:上市公司信息披露及外部报告、财经媒体新闻报道和互动社区平台文本。相应地,文本情绪也可分为三种:信息报告情绪、新闻媒体情绪和社交媒体情绪。这三种文本数据及对应文本情绪的差异如表 1 所示。

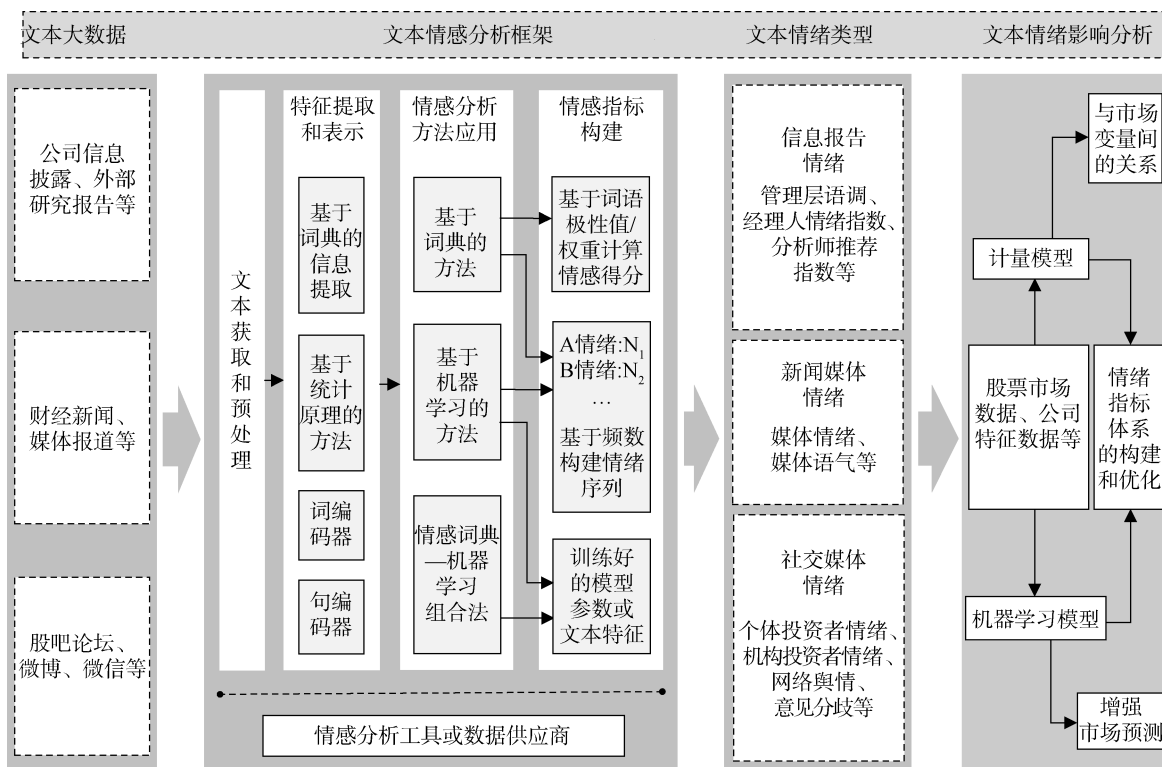


图1 文章内容框架

Fig.1 Framework of the Article

表1 三种文本情绪比较

Table 1 Comparison of Three Text Emotions

比较维度	信息报告情绪	新闻媒体情绪	社交媒体情绪
文本数据来源	公司信息披露和外部研究报告,如招股书、评级报告、盈余公告、电话会议记录、行研报告等	财经媒体新闻报道,如华尔街日报、彭博社、中国证券报等	股吧、微博等社交平台,如东方财富股吧、雪球网、淘股吧、新浪微博、Twitter、Facebook等
文本数据特点	格式规范、专业术语多、图表丰富、理解难度较大	有一定格式、标题高度概括、理解难度较小	短小精悍、格式不一、内容通俗易懂、符号丰富、含义多元
内涵情绪特点	偏正面积极、负面表达委婉、可读性不一	有正面有负面,主题多元,视角多维	有正面有负面、情感浓度高、表达直白
主要受众群体	专业研究员、机构投资者	机构投资者、个体投资者	个体投资者

信息报告情绪的文本来源有两种:上市公司信息披露和券商等金融中介机构的研究报告,这些文本按照一定格式和规范组织而成,偏好使用图表直观展示,常以独立的文档形式呈现。该类文本专业术语丰富、信息理解门槛高^[23],有时管理层会刻意使用晦涩语句或玩“文字游戏”以提高文本理解难度、隐藏负面信息、夸大利好^[24]。

从财经新闻得到的情绪称为新闻媒体情绪。财经新闻相对财务披露文本的专业性更低、术语更少,

表达重点突出且通俗易懂。新闻报道一般以篇章形式呈现,但在研究中有两种分析模式。一种模式是只分析新闻标题和摘要,这类研究认为标题是比全文更加简洁、精练的表达,文本噪声更少^[9],也能在保证信息质量的前提下提高运算效率。另一种模式是关注报道全文内容,这类研究认为全文包含了事件的背景信息和更详细的描述,可为预测提供额外有价值的线索^[25]。

本文将从互动社区平台提取的文本情绪定义为

社交媒体情绪,互动社区平台产生的文本主要指用户在财经论坛、微博、朋友圈等发布的留言、动态,便捷而即时的评论、转发行为在促进信息二次传播的同时也加强了金融市场参与主体之间的互动交流和情绪传染。已有研究表明,股价的动态变化与投资者之间的社会网络结构与交流互动过程密切相关^[26]。根据结构和形式不同,可进一步区分为论坛型文本和微博型文本^[27]。

综上,来源渠道或传播媒介的差异性使文本数据呈现多样性。从数据结构看,既有纯文本数据或以图表呈现的结构化数据,又有混杂着文本、图表、表情符号等非结构化或半结构化数据;从文本形式看,既有文档单位的长文本,又有受到平台规则限制(如微博、Twitter等)的短文本;从文本内容看,既有行文规范统一、负面语调规避、专业词汇频现的文本,又有风格形式多变、语言表达歧义、表情符号丰富的文本。多样化的文本数据和复杂的情感分析任务推动从不同角度对文本分析技术进行改进和创新。

3 文本情感分析框架

3.1 文本情感分析原理

文本情感分析(Sentiment Analysis, SA)是一个旨在分析文本中主观性信息(如观点、态度、情感等)的研究领域^[28],也是自然语言处理(Natural Language Processing, NLP)最活跃的研究领域之一。文本情感分析的子任务众多,金融市场研究中涉及较多的是极性分类和情绪分析。极性分类任务通常将文本分为正面、负面两类或正面、负面、中性三类^[29-30]。若考虑情绪强度,可进一步区分强、中、弱不同情绪^[31-32]。情绪分析任务对情绪的分类则更加复杂多元。Mittal等^[33]将Twitter文本情绪分为4类(Calm, Happy, Alert, Kind),发现Calm和Happy这两种情绪能更好地预测股指走势;王夫乐等^[34]将微博文本情绪分为悲伤、恐惧、喜悦、惊奇和愤怒。文本情感分析总体框架包括文本获取和预处理、文本特征提取和表示、情感分析方法应用,以及情感指标构建。

3.2 文本获取和预处理

目前,文本信息主要通过网络实现存储和传播,形成了海量的网络文本大数据。大多数网站都有应

用程序接口可供研究人员获取数据。原始文本通常需要进行预处理(也称为文本清洗),以达到缩减篇幅、提高后续处理效率的目的。常见的步骤有词性标注、停用词删减、否定词处理、短语切分等,这一过程会因语言的差异而不同。中文文本不以空格作为单词的自然分界,还需进行分词处理。常用的分词工具有中国科学院计算技术研究所汉语词法分析系统 ICTCLAS (Institute of Computing Technology, Chinese Lexical Analysis System)^[29,31,35]、jieba库^[36-37]等。

3.3 文本特征提取和表示

NLP任务中,通常需要对文本进行特征提取和表示,将其转化为计算机可理解的结构化数据。文本表示方法可分为基于词典的信息提取(如POS Tagger等)、基于统计原理的方法(如 n -gram、TF-IDF等)、词编码器(如Word2Vec、GloVe、FastText等)和句编码器^[38],金融领域应用中使用较为广泛的是前三种。基于词典的信息提取指识别和保留包含有用信息或情感相关的单词,一般是与基于词典的情感分析方法结合使用;基于统计原理的方法通常与传统机器学习模型结合使用,但这种文本特征表示方法存在数据稀疏、数据维度高、单词间语义关系容易被忽略等问题;词编码器通过词嵌入(Word Embedding)这一过程,实现在保留单词上下文信息的同时将离散的词汇映射到等密度、低维的实数向量,根据向量间的距离可判断不同词汇在语义或所处语境上的相似程度,从而快速有效地捕捉到更多复杂的语义信息,弥补了基于统计原理的方法的一些不足^[39],词编码器和深度学习模型结合最紧密。

3.4 情感分析方法应用

金融领域的文本情感分析方法可分为基于词典的方法、基于机器学习的方法和情感词典-机器学习组合法^[40-42]。其中,机器学习方法又可以分为传统机器学习和深度学习^[43],深度学习是在传统机器学习的基础上发展起来的,近年来发展迅速,已逐渐成为有别于传统机器学习的研究领域。4种情感分析方法的详细比较如表2所示。

(1) 基于词典的方法

基于词典的方法是将文档中识别出的词汇与情感词典中的词汇进行匹配,然后根据单词和短语的语

义方向或计算出的态度倾向值判定文本情感极性。目前,较权威且广泛应用的情感词典或语料库有7个,其中适用于英文文本的有Harvard IV-4情感词典^[44-45]、Loughran-McDonald金融情感词典^[19, 40, 44-48]、SentiWordNet^[49-50]、SenticNet^[50],适用于中文文本的有大连理工大学中文情感词汇本体库^[36, 51-52]、台湾大学简体中文情感极性词典NTUSD^[36, 52-53]、《知网》情感分析用词语集(beta版)HowNet^[36, 51-52, 54]。相比于中文适

用的情感词典资源,英文适用的情感词典资源更加丰富和完善。Price等^[55]研究发现,特定主题或语境相关的词典比通用词典的使用效果更好,然而,主流的中文情感词典的主题性、专业性严重不足。例如,国内并没有类似Loughran-McDonald词典这样专门以财务文本构建的金融情感词典,虽然可以利用该词典的中文翻译版,但由于翻译局限性及语言文化的差异,使用效果会大打折扣^[37]。

表2 4种情感分析方法比较

Table 2 Comparison of Four Sentiment Analysis Methods

比较维度	基于词典的方法	基于传统机器学习的方法	基于深度学习的方法	情感词典-机器学习组合法
基本原理	词汇匹配	特征表示,性能评估,模型优化	神经网络,模型优化	综合前两种
上手难度	简单	相对简单	复杂	复杂
主要优点	易于理解、操作简单	可直接套用成熟模型,便捷高效、省时省力;更精准捕捉文本语义;可处理文本大数据	特征提取和选择不需过多人工干预;适用于大型数据集	文本语义丢失少;可自主高效地构建合格训练数据集
主要缺点	易丢失、曲解部分文本信息;人工构建词典耗时费力;可直接使用的中文词典资源有限、专用性差	模型好坏高度依赖训练数据集的标注质量和数量;优质公开的训练数据集比较匮乏	涉及较多计算机知识,难以快速上手;依赖高端硬件资源,训练时间长;部分传统机器学习缺点	知识门槛高、理解和应用难度大;金融领域研究中的使用效果有待进一步验证
适用范围	篇幅短小、上下文语境联系较弱的文本	数据量相对较小的文本	数据量大的文本,针对不同文本特点可以选用、改进或融合不同模型	电影评论、产品评论等数据集
应用情况	在金融领域研究中应用广泛	在金融和计算机交叉领域研究中应用较多	在计算机视觉、自然语言处理领域应用较多	在非金融领域研究应用更多、效果更好

一些学者不满足于依赖现有词典,尝试针对所研究领域或主题构建专属词典或语料库。例如,Caparin等^[56]根据美国两家主流财经媒体的新闻,并针对标准普尔100指数成分股构建了专门的新闻情绪度量指标数据库。在国内,主流中文情感词典的缺陷和匮乏也迫使学者们不断尝试,如汪昌云等^[35]以《最新汉英经济金融常用术语使用手册》《现代汉语词典》以及Loughran-McDonald词典的中文翻译版为标准,构建了适用于中国财经媒体报道的金融词库;于琴等^[36]融合了《知网》情感分析用词语集、台湾大学简体中文情感极性词典、大连理工大学中文情感词汇本体库、《新华08汉英金融词典》《路透金融词典》和《高盛金融词典》,构建了专门度量财经新闻文本情绪的词典;Li等^[37]对东方财富网上沪深300指数成分股相关的样本帖进行人工标注并提取关键词,构建了“中国股票市场词典”。姚加权等^[57]借助词典重组和深度学习算法构建了适用于金融领域的中文情绪词典。目前,细分领域中文情感词典

的构建工作已经有了初步进展,但由于这些词典尚未公开,参照标准和构建方法也各不相同,无法对其做进一步的比较、改进或融合。因此,这些词典的价值和应用还有待发掘和推广。

基于词典的方法存在的缺陷有:忽略了单词出现的顺序和语法结构,造成对文本内容理解的偏差甚至背离;词加权方案因人而异且对上下文内容敏感,词典无法自动调整以反映变化的上下文和语义;人工构建词典耗时费力,且多针对特定领域制定取舍分类标准,适用范围小,推广性差。尽管如此,在金融领域文本情感分析中,基于词典的方法因易于理解、操作简单而得到广泛应用,特别是对于篇幅短小、上下文语境联系较弱的文本而言,基于词典的方法与基于机器学习的方法对文本情绪分类的准确性不相上下^[37, 58]。

(2) 基于传统机器学习的方法

情感词典法的缺陷促使一些学者开发基于传统机器学习的统计技术,可以从数据中学习复杂特征,

改善情感分析过程。传统机器学习在情感分析中的应用较早,通过文献梳理,发现较受欢迎的传统机器学习模型有支持向量机(Support Vector Machine, SVM)和朴素贝叶斯(Naive Bayes, NB)^[41],不同机器学习模型在对金融文本进行情感极性分类时的性能表现不同。若想更深入了解文本情感分析中常用的机器学习模型原理,可以参考 Liang 等^[59]、Yadav 等^[43]以及 Habimana 等^[60]的综述研究。

(3) 基于深度学习的方法

深度学习模型相比传统机器学习模型更加复杂,性能更佳,表现在:传统的机器学习方法中,特征提取这一过程非常耗时,但深度学习能够自动创建分类过程所需的特征,特征的提取和选择几乎不需要过多人工干预;对于大型数据集,传统的机器学习无法执行,而深度学习模型却可以被训练得更好;深度学习的多层架构模式可以很好地捕获数据中的非线性关系甚至更复杂的特性。

在金融文本情感分析领域,运用较多的深度学习模型有卷积神经网络(Convolutional Neural Networks, CNN)和循环神经网络(Recurrent Neural Networks, RNN)^[37, 61]。RNN 中引起广泛关注的是变换器(Transformer),它通过使用编码器和解码器进行 Sec2Sec 转换,进一步提高整体分类性能。例如,Devlin 等^[62]证明了 BERT 在 11 种 NLP 任务中均展现了独特优势,而 Zhao 等^[39]基于财务文本的实证表明, RoBERTa 模型在情感分析任务上的表现以微弱优势超过了 BERT 模型,精度达到 96.64%。Mishev 等^[38]基于两种人工标注的财经新闻数据集,对比了包括 BERT、RoBERTa、BART 等在内的多种 Transformer 系列模型的情感分类效果,结果表明 BART 模型最优。

大多数深度学习技术都依赖于人工标注数据集进行训练,即使是在预训练模型基础上进行迁移学习也离不开标注数据。因此,标注数据集的数量和质量对进一步提高模型性能和预测精度有很大影响。然而,从 Habimana 等^[60]整理的情感分析常用数据集来看,目前公开的标注数据集主要是基于电影、产品、服务等评论数据^[38],以及以 Twitter 为代表的外国社交媒体数据,鲜少有针对中国金融市场相关文本标注的数据集,即使存在这样的汉语标注数据

集,因其可公开获取的数据量有限,也达不到最前沿深度学习模型对训练数据量的基本要求。综上,国内基于深度学习的金融领域文本情感分析研究相比国外进展较慢。

(4) 情感词典-机器学习组合法

机器学习法依赖于从训练数据中学习有效特征,而情感词典法缺乏对上下文语义的考虑^[42],因此有研究提出情感词典-机器学习组合法(简称组合法):先借助情感词典优化特征表示(相当于发挥训练集的作用)或利用词典生成训练集,再在此基础上训练出性能更好的机器学习模型进行文本情感分析。然而, Mishev 等^[38]研究表明,金融领域研究中组合法的性能远不及单纯基于机器学习的方法。

(5) 情感分析工具或数据库

基于上述方法原理,国内外开发了不少情感分析程序或软件,可以更便捷地获取文本情绪分类结果。此外,还可以直接购买特定网络文本的情感数据来构建情绪序列,大大提高了研究效率。英文适用的情感分析工具或数据库有 OpinionFinder^[33, 53]、Bloomberg^[63]、DataSift^[64]、PsychSignal^[65-66]、FinSentS Web News Sentiment^[67]、Sentic Computing^[65]等,中文适用的情感分析工具或数据库有 ROST CM^[30, 68-69]、SnowNLP^[29]、NLPIR^[35]等。

综上所述,应该综合考虑文本数据的来源、类型、语言特点、篇幅长短、拟提取信息特征以及研究目的、研究条件、成本等因素选择合适的情感分析技术或工具。

3.5 情感指标构建

经过上述流程,通常可以得到三种结果。第一种是数值型情感词典法的特有输出,即依据词典中单词的极性值或情感权重,按照一定规则计算得到单位采样时间内所有样本文本的总情绪得分,进而生成情绪序列^[32, 36, 50]。第二种是非数值型情感词典法和机器学习均可以产生的输出,即单位采样时间内所有样本文本在不同情绪类别中的分布和频数,据此可进一步计算单位采样时间内样本文本的情感倾向^[40, 51, 70-71]。第三种是机器学习和组合法的特有隐含输出,即训练好的模型参数或文本特征,后续通常和其他机器学习模型结合用于预测任务^[25, 29, 51, 67],最常见的是以训练生成的词嵌入特征作为后续模型

的输入。

本文将现有研究中构建的指标按照三种文本情绪归类:将管理层语调、经理人情绪指数、分析师推荐指数等类似指标归入信息报告情绪;将以媒体语调、媒体情绪、媒体语气等类似指标归入新闻媒体情绪;将投资者情绪、意见分歧、网络舆情、社会情绪等指标(不包括单纯基于数量或搜索量构建的关注度指标)归入本文所划分的社交媒体情绪。

4 文本情绪影响金融市场的相关研究

按照交易对象不同,金融市场可以分为股票市场、债券市场以及衍生品市场。

在债券市场,相关研究主要围绕文本情绪和债券信用利差的关系展开,主要结论有:第一,以管理层语调为代表的信息报告情绪与债券信用利差显著负相关^[72];第二,从Twitter提取的社交媒体情绪与标准普尔等4种不同的绿色债券指数回报显著正相关^[73];第三,新闻媒体情绪与债券利差存在显著相关性,而这种相关关系的方向在经济繁荣和低迷时期、较短和较长时期内均有所不同^[74],并且,在美国、德国、英国、意大利、西班牙、法国等多个市场的实证研究均表明,新闻媒体情绪确实有助于解释和预测主权债券的利差走势和期限结构变动^[75-76]。在国内,庞有明等^[77]用违约事件新闻情绪优化信用债估值模型,并证实新模型的预测精度有显著提升。

在衍生品市场,林杰等^[78]根据期货公司发布的行情预测信息构建了信息报告情绪,发现乐观(悲观)情绪会使收盘价在短期内出现上涨(下跌)。刘桂芳等^[79]用公司年报提取的信息报告情绪对欧式脆

弱期权定价模型进行改进,发现期权价格对负面情绪更加敏感。

目前,对于文本情绪影响债券市场或衍生品市场的研究较少,可能是因为这两个市场的活跃度相对较低,投资者关注度较少,相关文本数据不够丰富,文本情绪对其影响较小。因此,接下来将重点梳理文本情绪影响股票市场的研究,具体又分为对市场收益影响、对市场风险影响、在市场预测的应用,以及情绪指标体系优化这4方面。其中,对市场收益影响的研究囊括了有关股票价格、收益率、成交量、IPO抑价(Initial Public Offering Underpricing)等变量的内容;而对市场风险影响的研究囊括了有关波动率、崩盘风险、系统性风险等变量的内容。按照三种不同的文本情绪分别阐述这4方面。此外,为了保证内容的全面性,还纳入了现有研究中投资者情绪范畴内关注度指标的相关研究。在研究文本情绪与股票市场间相互关系时通常使用的文本数据来源、市场指标构成、模型和方法等如表3所示。可以看出三类文本情绪的相关研究中最的区别在于文本数据来源不同,在选择市场指标时主要是根据所研究的市场主体进行选择,如研究文本情绪与A股市场收益率的关系,则选择的市場指标通常为上证指数或沪深300指数的收益率。大多数模型和方法对三种类型的文本情绪都适用,如VaR模型、多元线性回归模型等,这些模型主要用于检验文本情绪与市场指标之间的影响关系。

4.1 信息报告情绪

(1) 信息报告情绪对市场收益的影响

在对收益率影响的研究方面,段江娇等^[71]基于

表3 文本数据来源、市场指标构成及模型和方法

Table 3 Sources of Text Data, Market Indicators, Models and Methods

文本情绪类型	文本数据来源	市场指标构成	模型和方法	相关文献
信息报告情绪	分析师预测评级、公司财报、电话会议记录等	上证A指、标普500等指数或个股的各期值及收益率、波动率、成交量等指标	VaR模型、多元线性回归模型、	段江娇等 ^[71] 、Jegadeesh等 ^[44] 、Jiang等 ^[19] 、Tsai等 ^[80] 、徐高彦等 ^[81] 、姚加权等 ^[57] 、Iwasaki等 ^[61]
新闻媒体情绪	主流财经新闻媒体,如《中国证券报》《证券时报》、路透社、英国《卫报》等	沪深300、标普500、富时100等指数或个股的各期值及收益率、波动率、成交量等指标	资产定价模型、GARCH模型、格兰杰因果关	于琴等 ^[36] 、Fraiberger等 ^[48] 、尹海员 ^[82] 、吕华揆等 ^[83] 、Johnman等 ^[40] 、张天骄等 ^[84] 、王昶等 ^[85] 、Liu等 ^[86] 、Shi等 ^[87] 、Duan等 ^[25]
社交媒体情绪	东方财富网、雪球网、新浪微博、Twitter、Facebook等主流社交媒体	上证A指、上证指数、沪深300、标普500等指数或个股的各期值及收益率、波动率、成交量等指标	系检验、脉冲响应分析等	段江娇等 ^[71] 、王夫乐等 ^[34] 、张宁等 ^[88] 、黄润鹏等 ^[69] 、黄创霞等 ^[89] 、石善冲等 ^[30] 、Affuso等 ^[64] 、Siganos等 ^[90] 、Checkley等 ^[66] 、Teti等 ^[63]

新浪财经公布的证券分析师对个股的评级数据构建了一种分析师推荐指数,发现该指数的上涨会导致股指收益率的短期上涨。Jegadeesh 等^[44]从公司年报构建指标以衡量财务文本语调,发现积极情绪得分与股票收益正相关,与 IPO 抑价负相关。然而, Jiang 等^[19]得出了相反结论,根据美国多家公司的财报和电话会议记录构造了经理人情绪指数,发现高涨的经理人情绪预示着股票收益的降低。

(2) 信息报告情绪对市场风险的影响

在对波动率影响的研究方面, Tsai 等^[80]基于回归模型得出财务文本情绪对预测美国股票波动率具有不可忽视的作用,可利用这种联系分析公司的财务风险。Rekabsaz 等^[47]发现公司年度财务报告情绪可以有效预测市场的长期波动率。

在股价崩盘风险研究方面,徐高彦等^[81]发现分析师预测分歧度与股价崩盘风险显著负相关。杨七中等^[91]采用长短期记忆网络(Long Short-Term Memory, LSTM)模型构建了管理层语调指标,发现积极的管理层语调有助于化解未来股价崩盘风险。姚加权等^[57]基于自建词典构建了年报语调指标,发现该指标在预测上市公司的股价崩盘风险方面具有显著效果。

(3) 信息报告情绪在市场预测的应用

Iwasaki 等^[61]提取了涉及 800 多种股票的日本分析师报告中的情绪信息,发现主观的分析师意见比客观的新闻报道对股价的影响更大,考虑分析师意见后的资产定价模型对股票收益的样本外预测能力(调整后 R^2)提高了 11.8%。

综上,信息报告情绪和股票收益率、波动率、崩盘风险之间均存在相关关系,可以利用信息报告情绪增强对股票市场的预测。

4.2 新闻媒体情绪

网络新闻媒体特别是专业权威的财经媒体,由于发布渠道多、传播速度快、影响力持久等特性,极易引起投资者关注和情绪波动,研究发现大多数投资者会优先考虑购买那些吸引他们注意的股票^[92],因此媒体新闻对金融市场运行发挥着重要作用。

从财经新闻得到的情绪指标通常称为媒体语调、媒体情绪、媒体语气等。国内早期相关研究就发现:媒体情绪越强,股票价格与其内在价值的偏离程

度就可能越大;媒体情绪对资产误定价的影响并不对称,乐观的媒体情绪更容易催生股价泡沫^[88]。

(1) 新闻媒体情绪对市场收益的影响

在对收益率影响的研究方面,大多数学者认为积极的新闻媒体情绪对个股或股指收益有正面影响,可用于市场预测^[36, 48, 82, 93]。于琴等^[36]根据中证网的新闻构建情绪指标,得出积极的新闻情绪可以有效预测未来一周股票收益率的上涨。Fraiberger 等^[48]使用路透社国际新闻估计媒体情绪对股票价格的影响,发现不管是发达经济体还是新兴经济体,新闻情绪都可以强有力地预测市场每日收益。单个国家地方新闻积极(消极)情绪的增长预示着该国本地市场收益会有小而短暂的增长(下降)。全球新闻情绪变化比地方新闻情绪变化对股票收益的影响更加强劲持久,且全球新闻积极情绪可以比地方新闻积极情绪带来更永久性的外资流入。Seng 等^[93]基于 100 多种股票的研究得出,积极(消极)的新闻情绪与收益率存在正(负)相关关系。尹海员^[82]基于词汇匹配和人工判断构建了多个衡量媒体情绪的主、客观指标,发现其中的媒体报道乐观词汇比率和乐观报道比率两个指标对沪深 300 指数收益率产生正向冲击效应。然而,Johnman 等^[40]基于英国富时 100 指数的研究,认为新闻情绪不会影响该股指的超额收益。情绪分歧也同样对市场收益有一定的影响,吕华揆等^[83]运用 Word2Vec+k-Means 方法将财经新闻划分为股权及高管变动类、股市波动类和公司战略及其他类三类新闻,并基于构建的带权重情感词典对三类新闻中的情感进行量化,研究情感分歧与股市变动之间的关系,发现只有公司战略及其他类新闻的情感分歧能对股市收益率产生显著影响,而另外两类新闻的情感分歧会对成交量产生显著影响。

此外,也有个别研究关注到中国国有媒体情绪和市场化媒体情绪可能对股市产生不同影响,You 等^[94]发现股票价格对市场化媒体发布的新闻报道情绪更加敏感。通过文献搜集和梳理,本文发现对于国有媒体和市场化媒体的对比研究,更多聚焦于其对公司治理、经营业绩等微观层面的影响,如 You 等^[94]发现市场化媒体的负面媒体情绪显著增加了高管被迫离职的可能性,而国有媒体负面情绪不会。但是,很少有专门探讨两类媒体对股市影响的研究,

这可能是由于影响宏观市场的因素众多,对媒体进一步细分后,两类新闻情绪的差异化影响并不明显突出,分开研究的价值不高,因而绝大多数研究中不区分媒体性质,直接构建综合的新闻情绪变量。

在对IPO抑价(溢价)影响的研究方面,汪昌云等^[35]研究得出,负面语气每下降1%,IPO抑价率和超募比例就会上升0.22%和0.13%。南晓莉^[95]基于212家A股公司的研究发现,负面媒体报道与IPO溢价显著负相关。然而,也有研究认为媒体无法在IPO期间提供新信息,不会制造或破坏市场泡沫。Bhattacharya等^[96]研究了1996-2000年美国互联网股票IPO样本,发现媒体报道本身并不能解释IPO定价差异和互联网泡沫的形成。

(2) 新闻媒体情绪对市场风险的影响

在对波动率影响的研究方面,Caporin等^[56]基于带有连续波动率和跳跃的杠杆异质自回归(Leverage Heterogeneous Auto-Regressive with Continuous volatility and Jumps, LHAR-CJ)模型探究了收益公告、宏观经济公告、特定公司新闻报道等多种信息数据对标准普尔100成分股的影响,最终发现特定公司新闻报道对股票波动率预测的效果最好。Johnman等^[40]发现,从英国《卫报》中提取的新闻情绪会影响富时100指数的波动率,具体表现为负面情绪会增加波动率,而正面情绪会降低波动率。

在股价崩盘风险研究方面,张天骄等^[84]得出媒体过度乐观情绪与公司未来的股价崩盘风险显著正相关,而媒体合理乐观情绪不会产生显著影响。王昶等^[85]对创业板进行研究,发现正面和负面媒体情绪都显著提高了公司第二年股价崩盘风险。Li等^[97]的研究表明分析师意见和媒体新闻之间的分歧放大了新闻情绪对未来股价崩盘风险的影响。然而,Campbell等^[98]对英国铁路股票这一历史泡沫进行研究后发现,媒体新闻情绪并没有在繁荣时期提振股票收益,也没有引起市场崩盘。

(3) 新闻媒体情绪指标体系的构建与优化

此类研究多集中在国外,关注如何从新闻文本得到更有效的媒体情绪指标或信息以增强股市预测效果。一部分研究从文本分析技术角度进行优化,Liu等^[86]提出了一种新的神经网络架构——层次互补注意网络(Hierarchical Complementary Attention

Network, HCAN)以从新闻标题和新闻主体中捕获互补信息,避免冗余信息。该模型设计了新的度量标准计算注意力权重,并采用两级注意力机制以量化给定新闻中单词和句子的重要性。基于标准普尔500的实证结果表明,与BoW、FastText等文本特征表示方法相比,HCAN可以更好地表示新闻文本和预测股价。Shi等^[87]基于一种分层神经网络模型开发了Deep Clue系统,可以对主流财经新闻进行文本分析预测标准普尔500成分股的价格变动,并将预测模型中学到的关键因子直观展示给用户。

另一部分研究从新闻文本特征角度进行优化。出于省时高效的考虑,一些对新闻媒体情绪的研究只关注新闻标题中的情感信息^[9],但是也有研究认为仅利用新闻标题或摘要不足以实现精确预测,应该考虑整个新闻主体内容^[25, 40]。例如,Duan等^[25]开发了基于目标的新闻文档表示框架,无论目标是实体还是句子,均可以根据目标整合现有的表示模型。以路透社新闻为例,通过与其他模型结果对比,证明使用完整新闻内容与仅使用新闻标题和摘要相比在预测金融市场异常收益时的有效性。也有研究尝试兼顾新闻的标题和全文,如Caporin等^[56]优先从新闻标题中提取情感信息,如果检测不到积极或消极情绪,再对该篇新闻的全文进行情感分析。

此外,还有研究发现,新闻媒体报道具有普遍的积极情绪倾向,并且习惯用中性或正面词汇含蓄传递负面消息^[7],但一般的文本分析技术难以像人工阅读新闻那样准确地识别和捕捉这种隐含否定倾向表达,从而降低了金融新闻情绪分析的准确性和股市预测效果。针对此问题,Pröllochs等^[46]使用强化学习(Reinforcement Learning)方法增强文本的否定范围检测,可以使新闻情绪与股市收益之间的相关性提高10.63%。同样地,Caporin等^[56]在提取新闻情绪时也考虑了否定表达的影响。

4.3 社交媒体情绪

根据沪深两所统计年鉴,截至2019年,国内自然人开户总量超过4.5亿。在上交所A股市场,个体投资者持有约20%的市值,却贡献了80%以上的交易量。因此,中国股市容易受到个体投资者情绪的影响。

(1) 社交媒体情绪对市场收益的影响

在对收益率影响的研究方面,部分研究认为社

交媒体情绪与收益率存在正相关关系,也有研究认为存在负相关关系。段江娇等^[31]基于上证 A 股的个股研究发现,股票收益率与当日论坛情绪显著正相关。王夫乐等^[34]研究得出,社会情绪与股市收益显著正相关;社会情绪低落时,股市收益对社会情绪波动更为敏感。张宁等^[88]基于网络股评“发布者-关注者”数据分别构建了投资者情绪倾向指标与投资者情绪分歧指标,发现两类指标均与上证综合指数的价格和收益率显著相关。而段江娇等^[71]研究得出,论坛情绪指数的上涨会导致上证 A 指收益率的下降。

不同倾向的社交媒体情绪对收益的影响具有不对称性,一些研究认为正面情绪产生的影响更大^[69, 89, 99],但也有研究得出相反结论^[30, 64]。易洪波等^[99]认为投资者多方情绪比空方情绪对收益率的影响更明显。黄润鹏等^[69]将新浪微博情绪分为一般、中度、高度积极(消极)共 6 种,发现滞后 4 天的高度积极情绪变化是上证指数收盘价变化的格兰杰原因。黄创霞等^[89]得出,论坛积极情绪是上证指数收益率的格兰杰原因,但消极情绪对其影响不显著。然而,石善冲等^[30]基于 VaR 模型的研究认为,微信文本消极情绪倾向是上证指数收盘价变动的格兰杰原因,而积极和中性情绪不是。Affuso 等^[64]通过投资者地理位置等变量控制了模型内生性后,发现 Twitter 文本的正面和负面情绪对股票日收益的直接影响分别为 0.036% 和 0.078%,表明负面情绪对股票收益的影响更大。

在对成交量影响的研究方面,黄创霞等^[89]发现论坛情绪与上证指数成交量之间存在双向格兰杰因果关系。石善冲等^[30]发现微信积极和中性情绪变动会引发滞后一天的上证指数成交量变动,而消极情绪不是市场成交量变化的格兰杰原因。易洪波等^[99]指出投资者空方情绪比多方情绪对市场成交量的影响更明显。段江娇等^[31]发现论坛发帖的情绪分歧度越大,个股交易量就越大。Siganos 等^[90]也发现情绪分歧与同期股票交易量正相关。Li 等^[37]发现论坛文本情绪异常高涨或低落时,股票成交量会增加。

在对 IPO 抑价(溢价)影响的研究方面,石善冲等^[68]从企业公众号和非企业公众号文章中分别提取机构投资者情绪和个体投资者情绪,发现积极的机构投资者情绪能够显著正向影响 IPO 抑价,而消极

的机构投资者情绪无显著影响。此外,个体投资者情绪无论积极或消极对 IPO 抑价均没有显著影响。然而,南晓莉^[95]根据网络论坛文本构建了个体投资者意见分歧指标,发现其与 IPO 溢价存在显著正相关关系。

(2) 社交媒体情绪对市场风险的影响

在对波动率影响的研究方面, Li 等^[37]对沪深 300 成分股进行研究,发现论坛情绪对股票波动性的影响显著且具有不对称性。不同于一般研究使用日度或小时数据, Checkley 等^[66]以两分钟作为采样间隔收集 Twitter 和 StockTwits 上的文本情绪,以 5 种美股进行实证研究,这种更高频的情感指标和股票收益率、成交量和波动率之间也存在显著的因果关系。Siganos 等^[90]从 Facebook 获取 20 个国际市场的正面和负面情绪的每日数据,并基于正面情绪和负面情绪计算当日情绪分歧,发现情绪分歧与股价波动显著正相关,这一结论与 Banerjee 等^[100]的研究一致。

在股价崩盘风险研究方面,孙鲲鹏等^[101]发现乐观的股吧情绪会推高股价崩盘风险,并且这种影响随着情绪传播越广而变得越大,即存在“情绪传染”效应。而唐斯圆等^[102]认为无论看涨还是看跌的股吧情绪,都与未来股价崩盘风险显著正相关。姚加权等^[57]发现社交媒体情绪指标可以显著预测上市公司的股价崩盘风险。

在市场系统性风险研究方面,邓学斌等^[103]通过股吧情绪研究融资融券制度影响股市系统性风险的机制,得出股票系统性风险上升是由于融资融券放大了投资者情绪导致的。具体表现在,融资的杠杆机制放大了投资者情绪,导致系统性风险上升,而融券的卖空机制下,投资者乐观和悲观情绪都能得到充分表达,导致系统性风险下降。欧阳资生等^[104]发现股吧情绪对系统性金融风险的影响具有非对称性,积极情绪比消极情绪对系统性金融风险的影响更大。王白雪等^[105]实证发现雪球网投资者情绪可以很好地监测股市系统性风险,且效果明显优于 4 种传统的衡量系统性风险的方法。

(3) 社交媒体情绪在市场预测的应用

众多研究表明,社交媒体情绪在预测股票价格、收益率和波动率方面有显著贡献,不仅是基于计量模型的研究给出这一定性的结论^[37, 63, 106],一些基于

机器学习模型的预测研究还展示了量化的比较结果^[29, 43, 51, 70, 107]。Li等^[37]对沪深300成分股进行研究,发现隔夜论坛文本情绪可以显著预测市场回报,并且在几个月的时间范围内表现出“反应不足-反应过度”的模式,尤其是对于小市值股和成长股。Teti等^[63]以美国科技行业的个股为研究对象,经普通最小二乘法(Ordinary Least Squares, OLS)验证表明, Twitter文本情绪有助于股票价格预测,尤其是社交媒体覆盖率高的公司。丁肖丽^[106]发现基于股吧论坛的投资者情绪及其分歧对股票市场交易量与收益率均具有显著的预测能力。

Zhang等^[29]基于SVM和多层感知器(Multilayer Perceptron, MLP)模型预测A股股票走势,在纳入雪球网评论文本中的情绪特征变量后,模型的ACC和AUC值均有不同程度的提高。Wang等^[107]提出一种深度随机子空间集成的机器学习模型,可融合技术指标分析和微博情感分析,预测金融市场,以上证综合指数每日走势进行实证发现,该模型在F1值和AUC值上分别高出基准模型14.46%和14.2%。Ren等^[51]在预测股市涨跌时融合了股票论坛情绪信息并考虑了星期效应,对上证50指数的预测精度达到89.93%,比未考虑情绪变量的情况提高18.6%。Gupta等^[108]对5种美股进行价格预测,发现在纳入社交网站StockTwits的帖子情绪信息后,预测精度有了显著提升。Liu等^[70]以10种股票为例进行实证研究,表明纳入论坛情绪指标后的RNN模型对股票波动率的预测精度比原本提高了将近4%。

(4) 社交媒体情绪指标体系的构建和优化

考虑到社交网络的传导效应及社交平台的互动性,朱梦珩等^[54]在模型中不仅纳入了微博情感指标,还考虑了用户活跃度、内容影响力、创建时间、作者影响力等因素,显著提高了股价预测模型的拟合效果。同样地,Zhang等^[29]在基于雪球网上评论提取情绪特征时,根据数据集构建了每个交易日的用户动态转发网络计算不同用户的权限或影响力,并以此作为该用户所发布帖子情绪得分的权重。戴德宝等^[32]通过因子分析法将股吧情绪序列和股票交易数据融合,构建了投资者情绪综合指数,与未纳入情绪序列的两种指数相比,该情绪指数对上证指数的走势预测准确率最高。

4.4 关注度指标

一些研究将由文本数量特征构建指标称为“投资者关注度”“投资者情绪”“传播强度”等^[109],为了与前文特指的从文本内容出发、以情感分析方式构建的情绪指标进行区分,考虑到这类指标的构造是基于文本数量而非文本内容,本文将这类研究构建的指标统称为“*关注度”。数量型指标虽然不能明确指示情绪极性,但根据经验事实(被看涨的股票或牛市行情更容易引发市场关注和讨论),它在一定程度上可以反映市场情绪波动,因此可以纳入本文所讨论的文本情绪范畴。根据文本类型不同,可将关注度指标分为投资者关注和媒体关注。投资者关注指标通常基于对社区平台上用户发帖、关注、转发、评论等的数量统计或对目标关键词网络搜索量的统计来构建,主要反映投资者对股票信息的关注。媒体关注指标基于网络上发布和传播的包含目标关键词的新闻报道数量来构造,通常需要借助搜索引擎或人工词语匹配。

关于投资者关注的研究方面,Guo等^[110]基于投资者与财务顾问之间的Twitter发文频率数据研究投资者关注与市场回报之间的关系,发现每日股票收益率波动与投资者关注有关,异常关注会导致长达9周的暂时性收益波动。段江娇等^[31]研究发现,论坛关注度显著负向影响当日及未来两日的股票收益率,显著正向影响当日股价波动。孙书娜等^[111]发现雪球网用户关注度会显著提升市场交易量且在短期内对股价有显著的正向影响。Fang等^[112]研究发现,将反映投资者情绪的关键词搜索量数据(百度指数)纳入基准广义自回归条件异方差模型(Generalized AutoRegressive Conditional Heteroskedasticity, GARCH)可显著提高股市波动预测的准确性,尤其是在高波动时期。欧阳资生等^[104]研究发现股吧投资者关注度对系统性金融风险具有显著影响,表现为系统性风险随关注度增加而变高。

然而,有些研究认为投资者关注在预测股市方面作用有限。Guo等^[113]发现雪球网用户关注并不总是能有效预测股价,除非是对于那些本身就有较高关注度的股票。陈卫华等^[114]用深度学习模型预测沪深300指数的波动率,认为股吧发帖数增长率数据对提高预测精度的贡献有限,在不同损失函数下

最大仅提高 4.4%。

还有研究同时考查了投资者关注和媒体关注对市场的影响,如 Gao 等^[115]发现投资者关注比单纯的媒体关注更有助于降低中国股市的信息不对称;Xu 等^[11]根据百度搜索指数构建投资者关注指标并根据奇虎 360 搜索指数构建媒体关注指标,将两者纳入模型预测股市波动率。

总体来说,关注度指标构建过程简单且应用广泛,多用于研究股市波动性,但对投资者情绪的反映不够直接和具体,难以体现和探究其对股市的影响机制,而本文重点讨论的三种金融文本相关情绪指标均需要运用复杂的文本挖掘和情感分析技术,能更加细腻、准确地刻画不同类型的文本情绪。

5 总结和展望

5.1 总 结

对于文本所传递情绪的准确表征以及文本情绪与金融市场的关系研究,可以使上市公司、投资者和监管层更精确、迅速地了解市场情绪波动,及时采取风险防范措施。本文根据所使用文本来源渠道和性质特点的不同,将金融领域文本情绪分为信息报告情绪、新闻媒体情绪和社交媒体情绪三种。通过对大量文献的梳理归纳,本文认为,在对文本情绪进行极性分类时,应根据文本特点、研究条件、研究目标等的不同选择合适的情感分析方法。例如,基于词典的方法易于理解、操作简单,适用于篇幅短小、上下文语境联系较弱的文本;当需要更多考虑文本语境,涉及市场预测任务时,可采用传统机器学习模型;当样本数据量极大,对模型精度有较高要求,有条件获取充足且高质量的标注数据集时,可以考虑采用深度学习模型。

近年来,国内外对三种文本情绪的研究中,围绕社交媒体情绪展开的研究是一大热点,而对信息报告情绪的研究较少,且主要聚焦于国外市场。有关文本情绪与市场变量关系的研究大多采用计量分析的方法,因为传统回归分析以严格假设和充足先验为前提,可清晰直观地刻画出变量之间的关系。总体而言,对文本情绪影响股票收益率和成交量的研究较多,但对于波动率的影响分析较少;文本情绪确实会影响股票市场,并有助于构建更好的交易策略

和投资组合。但是,关于文本情绪对不同市场变量的影响方向、影响程度,以及变量之间的因果关系等问题,尚无统一的定论,甚至出现相悖的结论,这就需要对文本情绪影响市场的路径和机制进行更深入的探究。

有关利用文本情绪改善价格预测、优化投资策略的研究大多基于机器学习模型。这类研究主要集中在三个方向:一是从社交网络等信息传播学和心理学领域出发,探索从社交平台文本构建可以包含更多行为信息的情绪指标;二是从计算机、NLP 等领域出发,将新兴的文本分析技术迁移应用于金融市场,或是针对特定金融文本的分析任务对已有模型进行改善;三是从金融、经济学领域出发,在已有的基于技术或基本面的分析框架下纳入文本情绪信息,以实现资产组合和投资策略的优化。

5.2 建议与展望

对我国金融领域文本情绪相关研究的建议和展望将从以下三方面展开:文本数据源、文本情感分析框架、与金融领域结合的研究方向。

在文本数据源方面,除了本文提到的,还可以考虑政府规划和工作报告、热门公众号文章、知名人士微博、法院诉讼和判决书等其他文本数据源。目前,以这些文本为样本数据的研究还不多,从更广泛的文本数据源挖掘和获取有效信息,可发现和建立变量间新的联系。例如,可以通过对政府工作报告或政策法规的文本分析研究政府部门情绪对宏观市场或特定行业的影响。此外,可特别关注以上文本数据源中的非中文文本,也可以将数据源拓展至国外社交网站或财经媒体。在当今资本全球化的趋势下,随着越来越多国际资本的渗透参与,有必要研究国际投资者情绪和外国媒体情绪与本土市场之间的动态关系。同时,由于不同国家语言特点和文化环境相异,对文本的处理和情绪识别也有所不同,可以尝试比较不同文化语境下文本情绪与金融市场的关系。

在文本情感分析框架方面,第一,构建并推广特定领域、中文适用的情感词典或语料库。主流的几款中文情感词典具有普适性但缺乏专业性,单纯以词语的褒、贬义进行分类而不考虑文本风格和金融语境。这项工作的意义在于可以为词典法提供与研

究方向更匹配的标准数据,如围绕新闻报道类文本构建的情感词典有助于更好地识别新闻媒体情绪。第二,构建并丰富公开可用的标注数据集,加快NLP前沿技术在金融领域的迁移应用。目前鲜少有针对性中国金融市场相关文本标注的能满足模型训练体量要求的数据集。在计算机领域,已有学者尝试将最新NLP技术应用于财务文本和财经新闻的情感分类,得到了显著优越的结果。但是,基于此结果在金融领域的拓展和应用研究还很少,因此需要促进NLP前沿技术在金融领域的迅速迁移和应用。

在与金融领域结合的研究方向方面,第一,通过比较不同细分类型、不同时间频率、不同事件背景或经济周期下的文本情绪对金融资产价格及其波动的差异化影响,进一步深入对文本情绪影响金融市场的路径和机制研究。第二,建议更多地探究不同文本情绪间的关系,可以加深对文本情绪如何在上市公司、新闻媒体、社会公众之间传导的理解,从而有助于监管方对网络舆情可能引发的市场风险的控制。最后,文本情绪对金融市场的影响研究涉及语言学、传播学、计算机、心理学、经济金融等多个领域的知识和技能,需要综合运用多学科思维创新研究视角、增加研究深度、拓展研究边界。

参考文献:

- [1] de Long J B D, Shleifer A, Summers L H, et al. Noise Trader Risk in Financial Markets[J]. *Journal of Political Economy*, 1990, 98(4): 703-738.
- [2] Baker M, Wurgler J. Investor Sentiment and the Cross-Section of Stock Returns[J]. *The Journal of Finance*, 2006, 61(4): 1645-1680.
- [3] 伍燕然, 韩立岩. 不完全理性、投资者情绪与封闭式基金之谜[J]. *经济研究*, 2007, 42(3): 117-129. (Wu Yanran, Han Liyan. Imperfect Rationality, Sentiment and Closed End Fund Puzzle[J]. *Economic Research Journal*, 2007, 42(3): 117-129.)
- [4] 易志高, 茅宁. 中国股市投资者情绪测量研究: CICSII的构建[J]. *金融研究*, 2009(11): 174-184. (Yi Zhigao, Mao Ning. Research on the Measurement of Investor Sentiment in Chinese Stock Market: The CICSII's Construction[J]. *Journal of Financial Research*, 2009(11): 174-184.)
- [5] Brown G W, Cliff M T. Investor Sentiment and the Near-Term Stock Market[J]. *Journal of Empirical Finance*, 2004, 11(1): 1-27.
- [6] 王美今, 孙建军. 中国股市收益、收益波动与投资者情绪[J]. *经济研究*, 2004, 39(10): 75-83. (Wang Meijin, Sun Jianjun. Stock Market Returns, Volatility and the Role of Investor Sentiment in China[J]. *Economic Research Journal*, 2004, 39(10): 75-83.)
- [7] Lemmon M, Portniaguina E. Consumer Confidence and Asset Prices: Some Empirical Evidence[J]. *The Review of Financial Studies*, 2006, 19(4): 1499-1529.
- [8] Loughran T, McDonald B. When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks[J]. *The Journal of Finance*, 2011, 66(1): 35-65.
- [9] Nassirtoussi A K, Aghabozorgi S, Wah T Y, et al. Text Mining for Market Prediction: A Systematic Review[J]. *Expert Systems with Applications*, 2014, 41(16): 7653-7670.
- [10] Nassirtoussi A K, Aghabozorgi S, Wah T Y, et al. Text Mining of News-Headlines for FOREX Market Prediction: A Multi-Layer Dimension Reduction Algorithm with Semantics and Sentiment[J]. *Expert Systems with Applications*, 2015, 42(1): 306-324.
- [11] Xu Q F, Wang L K, Jiang C X, et al. A Novel UMIDAS-SVQR Model with Mixed Frequency Investor Sentiment for Predicting Stock Market Volatility[J]. *Expert Systems with Applications*, 2019, 132(C): 12-27.
- [12] 张同明, 张宁. 股票市场投资者情绪指数研究综述[J]. *计算机科学*, 2021, 48(S1): 143-150. (Zhang Tongming, Zhang Ning. Review of Research on Investor Sentiment Index in Stock Market[J]. *Computer Science*, 2021, 48(S1): 143-150.)
- [13] 何诚颖, 陈锐, 薛冰, 等. 投资者情绪、有限套利与股价异象[J]. *经济研究*, 2021, 56(1): 58-73. (He Chengying, Chen Rui, Xue Bing, et al. Investor Sentiment, Limited Arbitrage and Stock Price Anomalies[J]. *Economic Research Journal*, 2021, 56(1): 58-73.)
- [14] Fu J H, Wu X, Liu Y F, et al. Firm-Specific Investor Sentiment and Stock Price Crash Risk[J]. *Finance Research Letters*, 2021, 38: 101442.
- [15] 姚尧之, 王坚强, 刘志峰. 混频投资者情绪与股票价格行为[J]. *管理科学学报*, 2018, 21(2): 104-113. (Yao Yaoshi, Wang Jianqiang, Liu Zhifeng. Mixed-Frequency Investor Sentiment and Stock Price Behavior[J]. *Journal of Management Sciences in China*, 2018, 21(2): 104-113.)
- [16] 岑咏华, 谭志浩, 吴承尧. 财经媒介信息对股票市场的影响研究: 基于情感分析的实证[J]. *数据分析与知识发现*, 2019, 3(9): 98-114. (Cen Yonghua, Tan Zhihao, Wu Chengyao. Impacts of Financial Media Information on Stock Market: An Empirical Study of Sentiment Analysis[J]. *Data Analysis and Knowledge Discovery*, 2019, 3(9): 98-114.)
- [17] 部慧, 解峥, 李佳鸿, 等. 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. (Bu Hui, Xie Zheng, Li Jiahong, et al. Investor Sentiment Extracted from Internet Stock Message Boards and Its Effect on Chinese Stock Market[J]. *Journal of Management Sciences in China*, 2018, 21(4): 86-101.)

- [18] Antweiler W, Frank M Z. Is all that Talk Just Noise? The Information Content of Internet Stock Message Boards[J]. The Journal of Finance, 2004, 59(3): 1259-1294.
- [19] Jiang F W, Lee J, Martin X, et al. Manager Sentiment and Stock Returns[J]. Journal of Financial Economics, 2019, 132(1): 126-149.
- [20] 唐国豪, 姜富伟, 张定胜. 金融市场文本情绪研究进展[J]. 经济学动态, 2016(11): 137-147. (Tang Guohao, Jiang Fuwei, Zhang Dingsheng. Research Progress of Text Emotion in Financial Market[J]. Economic Perspectives, 2016(11): 137-147.)
- [21] 欧阳资生, 李虹宣. 网络舆情对金融市场的影响研究: 一个文献综述[J]. 统计与信息论坛, 2019, 34(11): 122-128. (Ouyang Zisheng, Li Hongxuan. A Review of Research on the Impact of Internet Public Opinion on Financial Markets[J]. Statistics & Information Forum, 2019, 34(11): 122-128.)
- [22] 沈艳, 陈赟, 黄卓. 文本大数据分析在经济学和金融学中的应用: 一个文献综述[J]. 经济学(季刊), 2019, 18(4): 1153-1186. (Shen Yan, Chen Yun, Huang Zhuo. A Literature Review of Textual Analysis in Economic and Financial Research[J]. China Economic Quarterly, 2019, 18(4): 1153-1186.)
- [23] Courtis J K. Annual Report Readability Variability: Tests of the Obfuscation Hypothesis[J]. Accounting, Auditing & Accountability Journal, 1998, 11(4): 459-472.
- [24] Lo K, Ramos F, Rogo R. Earnings Management and Annual Report Readability[J]. Journal of Accounting and Economics, 2017, 63(1): 1-25.
- [25] Duan J W, Ding X, Zhang Y, et al. TEND: A Target-Dependent Representation Learning Framework for News Document[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2019, 27(12): 2313-2325.
- [26] 王丽佳, 卢国祥. 基于社交网络的股票动态价格研究[J]. 投资研究, 2017, 36(5): 81-91. (Wang Lijia, Lu Guoxiang. Research on Stock Price Dynamics Based on Social Networks[J]. Review of Investment Studies, 2017, 36(5): 81-91.)
- [27] Ye M J, Li G Z. Internet Big Data and Capital Markets: A Literature Review[J]. Financial Innovation, 2017, 3(1): 6.
- [28] Lin C H, He Y L, Everson R, et al. Weakly Supervised Joint Sentiment-Topic Detection from Text[J]. IEEE Transactions on Knowledge and Data Engineering, 2012, 24(6): 1134-1145.
- [29] Zhang X, Shi J W, Wang D, et al. Exploiting Investors Social Network for Stock Prediction in China's Market[J]. Journal of Computational Science, 2018, 28: 294-303.
- [30] 石善冲, 朱颖楠, 赵志刚, 等. 基于微信文本挖掘的投资者情绪与股票市场表现[J]. 系统工程理论与实践, 2018, 38(6): 1404-1412. (Shi Shanchong, Zhu Yingnan, Zhao Zhigang, et al. The Investor Sentiment Mined from WeChat Text and Stock Market Performance[J]. Systems Engineering-Theory & Practice, 2018, 38(6): 1404-1412.)
- [31] 段江娇, 刘红忠, 曾剑平. 中国股票网络论坛的信息含量分析[J]. 金融研究, 2017(10): 178-192. (Duan Jiangjiao, Liu Hongzhong, Zeng Jianping. Analysis on the Information Content of China's Internet Stock Message Boards[J]. Journal of Financial Research, 2017(10): 178-192.)
- [32] 戴德宝, 兰玉森, 范体军, 等. 基于文本挖掘和机器学习的股指预测与决策研究[J]. 中国软科学, 2019(4): 166-175. (Dai Debao, Lan Yusen, Fan Tijun, et al. Stock Forecast with Investors Sentiment by Text Mining and Machine Learning[J]. China Soft Science, 2019(4): 166-175.)
- [33] Mittal A, Goel A. Stock Prediction Using Twitter Sentiment Analysis[OL]. [2022-01-13]. <http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf>.
- [34] 王夫乐, 王相悦. 社会情绪是否会影股市收益——来自新浪微博的证据[J]. 山西财经大学学报, 2017, 39(2): 35-46. (Wang Fule, Wang Xiangyue. Does Social Emotion Affect Stock Return—The Evidence from Sina Weibo[J]. Journal of Shanxi University of Finance and Economics, 2017, 39(2): 35-46.)
- [35] 汪昌云, 武佳薇. 媒体语气、投资者情绪与 IPO 定价[J]. 金融研究, 2015(9): 174-189. (Wang Changyun, Wu Jiawei. Media Tone, Investor Sentiment and IPO Pricing[J]. Journal of Financial Research, 2015(9): 174-189.)
- [36] 于琴, 张兵, 虞文微. 新闻情绪是股票收益的幕后推手吗[J]. 金融经济研究, 2017, 32(6): 95-103. (Yu Qin, Zhang Bing, Yu Wenwei. Are the Emotions of News a Wire-Puller of Stock Returns?[J]. Financial Economics Research, 2017, 32(6): 95-103.)
- [37] Li J, Chen Y, Shen Y, et al. Measuring China's Stock Market Sentiment[J]. Social Science Research Network, DOI: 10.2139/ssrn.3377684.
- [38] Mishev K, Gjorgjevikj A, Vodenska I, et al. Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers [J]. IEEE Access, 2020, 8: 131662-131682.
- [39] Zhao L Y, Li L, Zheng X H, et al. A BERT Based Sentiment Analysis and Key Entity Detection Approach for Online Financial Texts[C]//Proceedings of the 24th International Conference on Computer Supported Cooperative Work in Design. 2021: 1233-1238.
- [40] Johnman M, Vanstone B J, Gepp A. Predicting FTSE 100 Returns and Volatility Using Sentiment Analysis[J]. Accounting & Finance, 2018, 58(S1): 253-274.
- [41] Singh N K, Tomar D S, Sangaiah A K. Sentiment Analysis: A Review and Comparative Analysis over Social Media[J]. Journal of Ambient Intelligence and Humanized Computing, 2020, 11(1): 97-117.
- [42] Zucco C, Calabrese B, Agapito G, et al. Sentiment Analysis for Mining Texts and Social Networks Data: Methods and Tools[J]. WIREs: Data Mining and Knowledge Discovery, 2020, 10(1):

- e1333.
- [43] Yadav A, Vishwakarma D K. Sentiment Analysis Using Deep Learning Architectures: A Review[J]. Artificial Intelligence Review, 2020, 53(6): 4335-4385.
- [44] Jegadeesh N, Wu D. Word Power: A New Approach for Content Analysis[J]. Journal of Financial Economics, 2013, 110(3): 712-729.
- [45] Li X D, Xie H R, Chen L, et al. News Impact on Stock Price Return via Sentiment Analysis[J]. Knowledge-Based Systems, 2014, 69(1): 14-23.
- [46] Pröllochs N, Feuerriegel S, Neumann D. Negation Scope Detection in Sentiment Analysis: Decision Support for News-Driven Trading[J]. Decision Support Systems, 2016, 88: 67-75.
- [47] Rekabsaz N, Lupu M H, Baklanov A, et al. Volatility Prediction Using Financial Disclosures Sentiments with Word Embedding-Based IR Models[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2017: 1712-1721.
- [48] Fraiberger S P, Lee D, Puy D, et al. Media Sentiment and International Asset Prices[J]. Journal of International Economics, 2021, 133: 103526.
- [49] Khan F H, Qamar U, Bashir S. A Semi-Supervised Approach to Sentiment Analysis Using Revised Sentiment Strength Based on SentiWordNet[J]. Knowledge and Information Systems, 2017, 51(3): 851-872.
- [50] Pandarachalil R, Sendhilkumar S, Mahalakshmi G S. Twitter Sentiment Analysis for Large-Scale Data: An Unsupervised Approach[J]. Cognitive Computation, 2015, 7(2): 254-262.
- [51] Ren R, Wu D D, Liu T. X Forecasting Stock Market Movement Direction Using Sentiment Analysis and Support Vector Machine [J]. IEEE Systems Journal, 2019, 13(1): 760-770.
- [52] Liu S M, Chen J H. A Multi-Label Classification Based Approach for Sentiment Classification[J]. Expert Systems with Applications, 2015, 42(3): 1083-1093.
- [53] 乐国安, 董颖红, 陈浩, 等. 在线文本情感分析技术及应用[J]. 心理科学进展, 2013, 21(10): 1711-1719. (Yue Guoan, Dong Yinghong, Chen Hao, et al. Online Textual Sentiment Analysis Technology and It's Applications[J]. Advances in Psychological Science, 2013, 21(10): 1711-1719.)
- [54] 朱梦琨, 蒋洪迅, 许伟. 基于金融微博情感与传播效果的股票价格预测[J]. 山东大学学报(理学版), 2016, 51(11): 13-25. (Zhu Mengjun, Jiang Hongxun, Xu Wei. Weibo Moods and Propagation Factors Based Stock Prices Prediction[J]. Journal of Shandong University(Natural Science), 2016, 51(11): 13-25.)
- [55] Price S M, Doran J S, Peterson D R, et al. Earnings Conference Calls and Stock Returns: The Incremental Informativeness of Textual Tone[J]. Journal of Banking & Finance, 2012, 36(4): 992-1011.
- [56] Caporin M, Poli F. Building News Measures from Textual Data and an Application to Volatility Forecasting[J]. Econometrics, 2017, 5(3): 35.
- [57] 姚加权, 冯绪, 王赞钧, 等. 语调、情绪及市场影响: 基于金融情绪词典[J]. 管理科学学报, 2021, 24(5): 26-46. (Yao Jiaquan, Feng Xu, Wang Zanjun, et al. Tone, Sentiment and Market Impacts: The Construction of Chinese Sentiment Dictionary in Finance[J]. Journal of Management Sciences in China, 2021, 24(5): 26-46.)
- [58] Renault T. Intraday Online Investor Sentiment and Return Patterns in the U. S. Stock Market[J]. Journal of Banking & Finance, 2017, 84: 25-40.
- [59] Liang H, Sun X, Sun Y, et al. Text Feature Extraction Based on Deep Learning: A Review[J]. Eurasip Journal on Wireless Communications and Networking, 2017(1): 211.
- [60] Habimana O, Li Y H, Li R X, et al. Sentiment Analysis Using Deep Learning Approaches: An Overview[J]. Science China Information Sciences, 2019, 63(1): 111102.
- [61] Iwasaki H, Chen Y. Topic Sentiment Asset Pricing with DNN Supervised Learning[J]. SSRN Electronic Journal. DOI:10.2139/ssrn.3228485.
- [62] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1, 2019: 4171-4186.
- [63] Teti E, Dallochio M, Aniasi A. The Relationship Between Twitter and Stock Prices. Evidence from the US Technology Industry[J]. Technological Forecasting & Social Change, 2019, 149(C): 119747.
- [64] Affuso E, Lahtinen K D. Social Media Sentiment and Market Behavior[J]. Empirical Economics, 2019, 57(1): 105-127.
- [65] Xing F Z, Cambria E, Welsch R E. Intelligent Asset Allocation via Market Sentiment Views[J]. IEEE Computational Intelligence Magazine, 2018, 13(4): 25-34.
- [66] Checkley M S, Higón D A, Alles H. The Hasty Wisdom of the Mob: How Market Sentiment Predicts Stock Market Behavior[J]. Expert Systems with Applications, 2017, 77: 256-263.
- [67] Weng B, Lu L, Wang X, et al. Predicting Short-Term Stock Prices Using Ensemble Methods and Online Data Sources[J]. Expert Systems with Applications, 2018, 112: 258-273.
- [68] 石善冲, 康凯立, 赵志刚. 机构与个体投资者情绪对IPO抑价影响研究——基于微信文本挖掘[J]. 经济与管理, 2019, 33(1): 51-58. (Shi Shanchong, Kang Kaili, Zhao Zhigang. Research on the Influence of Institutional and Individual Investor Sentiment on IPO Underpricing[J]. Economy and Management, 2019, 33(1): 51-58.)
- [69] 黄润鹏, 左文明, 毕凌燕. 基于微博情绪信息的股票市场预测

- [J]. 管理工程学报, 2015, 29(1): 47-52, 215. (Huang Runpeng, Zuo Wenming, Bi Lingyan. Predicting the Stock Market Based on Microblog Mood[J]. Journal of Industrial Engineering and Engineering Management, 2015, 29(1): 47-52, 215.)
- [70] Liu Y F, Qin Z C, Li P Y, et al. Stock Volatility Prediction Using Recurrent Neural Networks with Sentiment Analysis[C]// Proceedings of International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems. 2017: 192-201.
- [71] 段江娇, 刘红忠, 曾剑平. 投资者情绪指数、分析师推荐指数与股指收益率的影响研究——基于我国东方财富网股吧论坛、新浪网分析师个股评级数据[J]. 上海金融, 2014(11): 60-64. (Duan Jiangjiao, Liu Hongzhong, Zeng Jianping. Research on the Influence of Investor Sentiment Index, Analyst Recommendation Index and Stock Index Return—Based on the Data of Analysts' Individual Stock Rating of China Oriental Fortune Network and Sina.com [J]. Shanghai Finance, 2014(11): 60-64.)
- [72] 姚潇, 吴冬晓, 庞守林. 基于文本挖掘的管理层语调对公司债券信用利差的影响[J]. 经济理论与经济管理, 2020(3): 99-112. (Yao Xiao, Wu Dongxiao, Pang Shoulin. The Influence of Management Tone on Corporate Bond Credit Spread Based on Text Mining[J]. Economic Theory and Business Management, 2020(3): 99-112.)
- [73] Piñeiro-Chousa J, López-Cabarcos M Á, Caby J, et al. The Influence of Investor Sentiment on the Green Bond Market[J]. Technological Forecasting and Social Change, 2021, 162: 120351.
- [74] Erlwein-Sayer C. Macroeconomic News Sentiment: Enhanced Risk Assessment for Sovereign Bonds[J]. Risks, 2018, 6(4): 141.
- [75] Gotthelf N, Uhl M W. News Sentiment: A New Yield Curve Factor[J]. Journal of Behavioral Finance, 2019, 20(1): 31-41.
- [76] Consoli S, Pezzoli L T, Tosetti E. Emotions in Macroeconomic News and Their Impact on the European Bond Market[J]. Journal of International Money and Finance, 2021, 118: 102472.
- [77] 庞有明, 蒋洪迅. 基于新闻舆情的信用债估值修正模型及其应用[J]. 山西大学学报(自然科学版), 2017, 40(1): 1-13. (Pang Youming, Jiang Hongxun. Media-Adjusted Credit Bond Valuation and Its Applications[J]. Journal of Shanxi University (Natural Science Edition), 2017, 40(1): 1-13.)
- [78] 林杰, 江晨曦. 基于 BERT 模型的投资者情绪指数建模及与价格关系分析[J]. 上海管理科学, 2020, 42(4): 75-80. (Lin Jie, Jiang Chenxi. Modeling of Investors' Sentiment Index Based on BERT Model and Analysis of Its Relationship with Price[J]. Shanghai Management Science, 2020, 42(4): 75-80.)
- [79] 刘桂芳, 徐维军, 黄静龙, 等. 考虑公司信息披露情绪的欧式脆弱期权定价[J]. 运筹与管理, 2021, 30(9): 164-171. (Liu Guifang, Xu Weijun, Huang Jinglong, et al. European Vulnerable Option Pricing Considering Corporate Information Disclosure Emotion [J]. Operations Research and Management Science, 2021, 30(9): 164-171.)
- [80] Tsai M F, Wang C J. On the Risk Prediction and Analysis of Soft Information in Finance Reports[J]. European Journal of Operational Research, 2017, 257(1): 243-250.
- [81] 徐高彦, 高歌. 分析师预测分歧是市场情绪洪流的“漫水桥”吗——基于股价崩盘风险的视角[J]. 现代经济探讨, 2022(2): 52-67. (Xu Gaoyan, Gao Ge. Is the Dispersion of Analysts' Forecast a “Water-Overflowing Bridge” in the Flood of Market Sentiment?: From the Perspective of Stock Price Crash Risk[J]. Modern Economic Research, 2022(2): 52-67.)
- [82] 尹海员. 新闻媒体报道对投资者情绪影响效应研究——来自我国股票市场的经验证据[J]. 厦门大学学报(哲学社会科学版), 2016(2): 92-101. (Yin Haiyuan. A Study on Effect of Media Reports on Investor Sentiment: Evidence from China's Stock Market[J]. Journal of Xiamen University (Arts & Social Sciences), 2016(2): 92-101.)
- [83] 吕华揆, 刘政昊, 钱宇星, 等. 异质性财经新闻与股市关系研究[J]. 数据分析与知识发现, 2021, 5(1): 99-111. (Lv Huakui, Liu Zhenghao, Qian Yuxing, et al. Relationship Between Financial News and Stock Market Fluctuations[J]. Data Analysis and Knowledge Discovery, 2021, 5(1): 99-111.)
- [84] 张天骄, 孙谦. 网络媒体过度乐观情绪与上市公司股价崩盘风险[J]. 管理现代化, 2022, 42(1): 34-39. (Zhang Tianjiao, Sun Qian. Sun Qian. Over-Optimistic Sentiment of Internet Media and Stock Price Crash Risk[J]. Modernization of Management, 2022, 42(1): 34-39.)
- [85] 王昶, 王敏, 龚铖. 媒体报道降低了股价崩盘风险吗?——来自创业板的证据[J]. 金融与经济, 2017(4): 74-80. (Wang Chang, Wang Min, Gong Cheng. Does Media Coverage Reduce the Risk of Stock Price Collapse? -Evidence from the Growth Enterprise Market[J]. Journal of Finance and Economics, 2017(4): 74-80.)
- [86] Liu Q K, Cheng X, Su S, et al. Hierarchical Complementary Attention Network for Predicting Stock Price Movements with News[C]//Proceedings of the 27th ACM International Conference on Information and Knowledge Management. 2018: 1603-1606.
- [87] Shi L, Teng Z Y, Wang L, et al. DeepClue: Visual Interpretation of Text-Based Deep Stock Prediction[J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 31(6): 1094-1108.
- [88] 张宁, 尹乐民, 何立峰. 网络股评“发布者-关注者”BSI与股票市场关联性研究[J]. 数据分析与知识发现, 2018, 2(6): 1-12. (Zhang Ning, Yin Lemin, He Lifeng. Impacts of “Poster-Follower” Sentiment on Stock Market Performance[J]. Data Analysis and Knowledge Discovery, 2018, 2(6): 1-12.)
- [89] 黄创霞, 温石刚, 杨鑫, 等. 个体投资者情绪与股票价格行为的互动关系研究[J]. 中国管理科学, 2020, 28(3): 191-200. (Huang Chuangxia, Wen Shigang, Yang Xin, et al. The Interactive Relationship Between Individual Investor Sentiment and Stock Price Behaviors[J]. Chinese Journal of Management Science, 2020, 28(3): 191-200.)

- 2020, 28(3): 191-200.)
- [90] Siganos A, Vagenas-Nanos E, Verwijmeren P. Divergence of Sentiment and Stock Market Trading[J]. *Journal of Banking & Finance*, 2017, 78: 130-141.
- [91] 杨七中, 章贵桥, 马蓓丽. 管理层语意与未来股价崩盘风险——基于投资者情绪的中介效应分析[J]. *中南财经政法大学学报*, 2020(1): 26-36, 159. (Yang Qizhong, Zhang Guiqiao, Ma Beili. Management Meanings and Stock Price Crash Risk: The Mediating Role of Investors Sentiments[J]. *Journal of Zhongnan University of Economics and Law*, 2020(1): 26-36, 159.)
- [92] Barber B M, Odean T. All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors[J]. *The Review of Financial Studies*, 2008, 21(2): 785-818.
- [93] Seng J L, Yang H F. The Association Between Stock Price Volatility and Financial News-A Sentiment Analysis Approach[J]. *Kybernetes*, 2017, 46(8): 1341-1365.
- [94] You J X, Zhang B H, Zhang L. Who Captures the Power of the Pen?[J]. *The Review of Financial Studies*, 2018, 31(1): 43-96.
- [95] 南晓莉. 新媒体时代网络投资者意见分歧对IPO溢价影响——基于股票论坛数据挖掘方法[J]. *中国软科学*, 2015(10): 155-165. (Nan Xiaoli. Effect of Internet Investors' Opinion Divergence on IPO Premiums in the New Media Age: Analysis Based on Data Mining of Stock BBS[J]. *China Soft Science*, 2015 (10): 155-165.)
- [96] Bhattacharya U, Galpin N, Ray R, et al. The Role of the Media in the Internet IPO Bubble[J]. *Journal of Financial and Quantitative Analysis*, 2009, 44(3): 657-682.
- [97] Li S Y, Kong J H. News Sentiment and the Risk of a Stock Price Crash Risk: Based on Financial Dictionary Combined BERT-DCA[J]. *Discrete Dynamics in Nature and Society*, 2022, 2022: 8305947.
- [98] Campbell G, Turner J D, Walker C B. The Role of the Media in a Bubble[J]. *Explorations in Economic History*, 2012, 49(4): 461-481.
- [99] 易洪波, 赖娟娟, 董大勇. 网络论坛不同投资者情绪对交易市场的影响——基于VAR模型的实证分析[J]. *财经论丛*, 2015 (1): 46-54. (Yi Hongbo, Lai Juanjuan, Dong Dayong. A Study of the Influence of BBS Investor Sentiments on the Trading Market——An Empirical Analysis Based on VAR Model[J]. *Collected Essays on Finance and Economics*, 2015(1): 46-54.)
- [100] Banerjee S, Kremer I. Disagreement and Learning: Dynamic Patterns of Trade[J]. *The Journal of Finance*, 2010, 65(4): 1269-1302.
- [101] 孙鲲鹏, 肖星. 互联网社交媒体对投资者情绪传染与股价崩盘风险的影响机制[J]. *技术经济*, 2018, 37(6): 93-102. (Sun Kunpeng, Xiao Xing. Influence of Social Media in Internet on Sentiment Contagion and Stock Price Crash[J]. *Technology Economics*, 2018, 37(6): 93-102.)
- [102] 唐斯圆, 李丹. 普通投资者关注度与股价崩盘风险——基于自媒体的研究[J]. *投资研究*, 2018, 37(4): 38-56. (Tang Siyuan, Li Dan. Ordinary Investors' Attention and Stock Price Crash Risk——Empirical Analysis Based on "We Media" [J]. *Review of Investment Studies*, 2018, 37(4): 38-56.)
- [103] 邓学斌, 胡凡. 融资融券制度、投资者情绪与股票市场系统性风险[J]. *暨南学报(哲学社会科学版)*, 2021, 43(9): 54-67. (Deng Xuebin, Hu Fan. Margin and Short-Selling System, Investor Sentiment and Systemic Risks of Stock Market[J]. *Jinan Journal (Philosophy & Social Sciences)*, 2021, 43(9): 54-67.)
- [104] 欧阳资生, 李虹宣, 杨希特. 网络舆情对中国上市金融机构系统性风险影响研究[J]. *系统科学与数学*, 2021, 41(5): 1339-1354. (Ouyang Zisheng, Li Hongxuan, Yang Xite. Research on the Influence of Network Public Opinion on the Systematic Risk of Listed Financial Institutions in China[J]. *Journal of Systems Science and Mathematical Sciences*, 2021, 41(5): 1339-1354.)
- [105] 王白雪, 郭琨, 孙毅, 等. 基于投资者情绪的股市系统性风险监测新方法[J]. *山西财经大学学报*, 2018, 40(S2): 30-34. (Wang Baixue, Guo Kun, Sun Yi, et al. A New Method of Systematic Risk Monitoring in Stock Market Based on Investor Sentiment [J]. *Journal of Shanxi University of Finance and Economics*, 2018, 40(S2): 30-34.)
- [106] 丁肖丽. 网络讨论、情绪分化与股票市场表现[J]. *湘潭大学学报(哲学社会科学版)*, 2019, 43(2): 75-81. (Ding Xiaoli. Internet Discussions, Sentiment Divergence and Stock Market Performance[J]. *Journal of Xiangtan University(Philosophy and Social Sciences)*, 2019, 43(2): 75-81.)
- [107] Wang Q L, Xu W, Zheng H. Combining the Wisdom of Crowds and Technical Analysis for Financial Market Prediction Using Deep Random Subspace Ensembles[J]. *Neurocomputing*, 2018, 299: 51-61.
- [108] Gupta R, Chen M. Sentiment Analysis for Stock Price Prediction [C] // *Proceedings of the 2020 IEEE Conference on Multimedia Information Processing and Retrieval*. 2020: 213-218.
- [109] 乔智, 耿志民. 股吧对个人投资者情绪的影响研究[J]. *金融理论与实践*, 2013(11): 77-81. (Qiao Zhi, Geng Zhimin. An Empirical Research on Individual Investor Emotional Effect from Stock BBS[J]. *Financial Theory & Practice*, 2013(11): 77-81.)
- [110] Guo T, Finke M, Mulholland B. Investor Attention and Advisor Social Media Interaction[J]. *Applied Economics Letters*, 2015, 22 (4): 261-265.
- [111] 孙书娜, 孙谦. 投资者关注和股市表现——基于雪球关注度的研究[J]. *管理科学学报*, 2018, 21(6): 60-71. (Sun Shuna, Sun Qian. Investor Attention and Market Performance: Evidence Based on "Xueqiu Attention" [J]. *Journal of Management Sciences in China*, 2018, 21(6): 60-71.)
- [112] Fang J C, Gozgor G, Lau C K M, et al. The Impact of Baidu

Index Sentiment on the Volatility of China's Stock Markets[J]. Finance Research Letters, 2020, 32: 101099.

- [113] Guo K, Sun Y, Qian X. Can Investor Sentiment be Used to Predict the Stock Price? Dynamic Analysis Based on China Stock Market[J]. Physica A: Statistical Mechanics and Its Applications, 2017, 469: 390-396.

- [114] 陈卫华, 徐国祥. 基于深度学习和股票论坛数据的股市波动率预测精度研究[J]. 管理世界, 2018, 34(1): 180-181. (Chen Weihua, Xu Guoxiang. Research on Prediction Accuracy of Stock Market Volatility Based on Deep Learning and Stock Forum Data [J]. Management World, 2018, 34(1): 180-181.)

- [115] Gao Y, Wang Y J, Wang C, et al. Internet Attention and Information Asymmetry: Evidence from Qihoo 360 Search Data on the Chinese Stock Market[J]. Physica A: Statistical Mechanics and Its Applications, 2018, 510: 802-811.

作者贡献声明:

李合龙:提出研究思路,设计研究方案,论文最终版本修订;

任昌松:文献总结归纳及论文修订;

柳欣茹:数据及文献搜集分析,论文起草;

汪存华:收集文献,论文最终版本审定。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

[1] 李合龙. 金融市场文本情绪研究综述文献列表及文献集. DOI:10.57760/sciencedb.07650.

收稿日期:2022-08-24

收修改稿日期:2023-03-06

Review of Textual Sentiment Research in Financial Markets

Li Helong¹ Ren Changsong¹ Liu Xinru¹ Wang Cunhua²

¹(School of Economics and Finance, South China University of Technology, Guangzhou 510006, China)

²(School of Economics and Management, North China Institute of Aerospace Engineering, Langfang 065000, China)

Abstract: [Objective] This paper analyzes and summarizes the current situation of the development of text sentiment in financial markets, and provides reference for subsequent related research. [Coverage] We used “financial market”, “text sentiment analysis”, “text sentiment” and “investor sentiment” as keywords to search on academic platforms such as CNKI, Web of Science and Google Academic, and extended the search for relevant literatures. A total of 115 papers were reviewed. [Methods] We classified the extracted text sentiment according to the type of the source financial text, then introduced the framework of text sentiment analysis, and finally sorted out the relevant research results on the impact of text sentiment on the financial markets. [Results] Text sentiment in financial markets can be divided into information reporting sentiment, news media sentiment and social media sentiment. In the construction of sentiment indicators, dictionary-based methods and machine learning-based methods are widely used. The above three text sentiments have a certain impact on the financial markets. [Limitations] Due to the universality of text analysis methods in various fields, the selected literature on the framework of text sentiment analysis is not entirely focused on the financial markets. [Conclusions] When constructing financial text sentiment indicators, we should choose the appropriate sentiment analysis method according to the text characteristics, research conditions and research objectives.

Keywords: Financial Market Text Sentiment Analysis Text Sentiment Investor Sentiment