

doi: 10.19920/j.cnki.jmsc.2022.09.006

股票网络论坛中的意见领袖：慧眼识珠还是吸引眼球^①

张 科¹, 李心丹¹, 方 晓², 李小琳³

(1. 南京大学工程管理学院, 南京 210093; 2. 美国特拉华大学阿尔弗莱德勒纳商学院, 纽瓦克 19716, 美国; 3. 南京大学管理学院, 南京 210093)

摘要: 互联网是投资者获取信息的主要来源, 其中意见领袖的观点能在投资者群体中广泛传播并影响其行为。本研究基于社会网络分析技术识别股票网络论坛中的意见领袖, 并利用文本分析技术判断其帖子类型, 系统研究意见领袖情绪的收益可预测性及其原因和投资者反应。利用2011年~2017年的样本数据, 研究发现意见领袖情绪对未来1个月至6个月的个股收益率都具有预测能力, 但并非是因其掌握价值信息。进一步研究表明, 投资者对意见领袖的认同程度与其判断准确率无关, 投资者更为关注发表正面观点的意见领袖而忽视其准确性, 同时排斥能够准确预测市场下跌的意见领袖。本研究揭示了意见领袖对市场的影响及其形成机制, 并为进一步完善投资者教育保护工作提供理论依据。

关键词: 意见领袖; 投资者情绪; 收益可预测性; 投资者教育

中图分类号: F830.91 文献标识码: A 文章编号: 1007-9807(2022)09-0090-18

0 引 言

信息是投资者制定交易策略的关键, 随着通信技术的发展, 网络已经成为投资者获取信息最主要的渠道^[1]。与现实生活中人际交流不同, 信息在互联网中的传播速度更快, 内容也更庞杂, 虽然人人都可以发表观点, 但并非所有人的声音都能被公众听到。社会网络领域的研究发现, 群体中存在少数被称为意见领袖的特殊个体, 他们对网络信息传播具有关键枢纽作用并能广泛影响他人态度行为^[1]。在现实生活中, 部分意见领袖对公众的影响力甚至超越了不少财经媒体, 成为互联网上重要的信息传播者, 如在新浪微博上一些知名财经评论家的个人粉丝数量可以达到500万以上, 而作为专业新闻媒体的国内四大证券报粉丝数量之和仅约为489万。通常, 投资者关注意见领袖是因其具有丰富经验并能判断市场走势, 但却

有一些意见领袖利用其知名度, 先买入股票再公开推荐该股, 误导投资者并非法获利, 例如, 2018年廖英强因操纵股市而被证监会罚没1.29亿元人民币。那么, 由此产生的问题是这些意见领袖能否准确预测市场走向? 意见领袖何以能够预测市场? 普通投资者能否鉴别出意见领袖的判断能力并做出合理反应?

金融学研究中通常以投资者发布观点所体现的情绪来衡量其态度观点, 目前国外关于投资者情绪的研究主要涉及投资者情绪影响股票价格的理论模型^[2, 3]、投资者情绪的计算方法^[4, 5], 以及实证检验了投资者情绪对股票收益率^[6, 7]、股价波动性^[8, 9]、交易量^[10]和上市公司盈利水平^[11]的预测作用。我国学者也从帖子情绪和情绪一致性的角度分析了投资者情绪是否包含能够预测市场的信息^[12-15], 得出了许多有益结论, 但是, 现有研究大多是将投资者视为同质化的个体, 在衡量投

① 收稿日期: 2019-10-19; 修订日期: 2021-10-28。

基金项目: 国家自然科学基金资助项目(71720107001; 72073063; U1811462)。

作者简介: 张 科(1982—), 男, 江苏南京人, 博士, 助理教授。Email: kezhang@nju.edu.cn

资者情绪时对所有投资者的观点都赋予相同的权重,并未考虑到意见领袖和普通公众对他人行为影响力差异。通过考察少数意见领袖情绪的作用,不仅可以进一步揭示投资者情绪影响资产价格的运作机理,还可以检验普通投资者能否理性对待网络舆论的影响,合理制定投资决策,为投资者教育的研究提供有益补充。

本研究基于社会网络分析技术,检验股票网络论坛中意见领袖的帖子情绪对未来个股收益率的预测性及其原因,并通过普通投资者对意见领袖判断能力反应的研究分析投资者是否理性。研究结果表明:意见领袖情绪对未来1个月至6个月的个股收益率都具有预测能力,而非意见领袖情绪和整体投资者情绪则不能预测市场,并且意见领袖能够预测市场的原因并非其掌握了上市公司的价值信息,可能是由于其观点能够引导投资者行为进而使股价同向变化。在此基础上,进一步分析投资者对意见领袖预测准确程度的反应,发现投资者对意见领袖的认同程度与其判断准确率无关,投资者更为关注发表正面观点的意见领袖而忽视其观点准确性,同时排斥能够准确预测市场下跌的意见领袖。

本研究的主要贡献在于:第一,现有研究发现意见领袖在新产品营销^[16]、网络众筹^[17]等领域具有重要作用,但在投资者情绪这一领域,据笔者有限知识所知,尚没有文献研究互联网社区中意见领袖情绪对证券市场的影响,本研究通过社会网络分析技术识别股票网络论坛中的意见领袖,发现其帖子情绪对市场的预测能力相比普通投资者更为长期持久,拓展了社会网络在金融市场中的研究;第二,目前关于投资者行为的研究主要分析了投资者自身因素和亲友等社会关系的影响,缺乏对网络上社会互动影响个人行为决策的关注,本研究则基于网络发帖回帖数据,检验了普通投资者对意见领袖言论的反应,分析普通投资者能否识别不同意见领袖的预测准确度,研究结论对投资者情绪传染的控制干预,以及监管部门完善投资者教育保护工作具有一定借鉴意义。

1 文献综述

1.1 投资者情绪影响市场的研究

行为金融学认为,投资者情绪是影响股票价格的重要因素^[18]。Barberis等^[2]认为,投资者在决策时存在两种心理偏差:一是由于投资者心理存在保守倾向,不能及时根据新信息调整自身对股价的观点,导致股票价格对新信息反应不足,使得投资者情绪能够正向影响股票未来价格;另一种则是由于投资者存在代表性启发偏差,过分高估偶然事件的发生概率,造成股票价格对新信息过度反应,此时投资者情绪对股票未来价格为负面影响。Baker和Wurgler^[19]则认为,不同类型股票对投资者情绪的敏感度存在差异,小市值、成立时间较短、股价波动性较高或亏损的上市公司更易于受到投资者情绪的影响,而大市值且具有稳定盈利水平的上市公司股票则具有类似债券的特征,较少受到投资者情绪的影响。刘晓星等^[20]也提出,当存在卖空约束时,投资者情绪变化引起其产生有偏的认知并影响市场流动性,他们还通过实证检验发现,融资融券业务的开展进一步增强了投资者情绪对市场流动性的影响。

近年来,学者们越来越关注互联网上用户发表内容所体现的情绪对市场的影响。Das和Chen^[9]则利用多种方法综合判断Yahoo论坛上的帖子情绪,构建了高科技股票的情绪指数,结果发现该指数虽然对未来个股收益率和科技板块指数收益率并不具有显著影响,但是却与股价波动性显著正相关。Bollen等^[21]检验了Twitter情绪对道琼斯指数收益率的影响,他们将用户发表内容的情绪分为六类:平静、警惕、确信、热情、和善和愉悦,研究发现只有平静类型的情绪指标能够显著预测未来道琼斯指数收益率。Sabherwal等^[22]在排除了基本面消息的情况下检验了帖子情绪对论坛热门股票价格的影响,结果发现帖子情绪与当期股票收益率正相关,但与未来两日的收益率负相关,表明这些热门股票价格存在“拉高一出货”的模式,股价在成为热门股的当天上涨到最高值,随后就开始下跌。Chen等^[6]分析了Seeking

Alpha 网站上用户发布内容的情绪与个股收益率的关系,研究发现用户发布内容所体现的负面情绪与未来 3 日到 60 日的累积超额收益率显著相关. Li 等^[23] 比较了多种机器学习算法衡量投资者情绪的差异,发现基于 LSTM 的投资者情绪指标具有较强的预测能力. Ge 等^[24] 研究了投资者情绪与股价崩盘风险之间的关系,发现“恐惧”情绪与股价崩盘风险之间相关性较强.

国内学者在此方面也做了一些研究,杨晓兰等^[25] 根据东方财富网股吧中创业板上市公司的相关帖子内容构建投资者情绪指标,研究发现本地投资者关注对上市公司股票收益率的影响依赖于论坛帖子情绪. 金德环和李岩^[26] 也利用东方财富网股吧的发帖数据检验了投资者情绪对股票未来收益率的影响,研究发现投资者情绪能够显著预测股票未来收益,并且还对上市公司意外盈余具有解释力. 段江娇等^[13] 利用 SVM 算法判断论坛帖子的投资者情绪,研究发现投资者情绪与当日个股收益率正相关,但对未来两日收益率则没有显著影响,同时,他们还发现当日帖子数量与个股未来两日收益率负相关. 部慧等^[14] 同样基于东方财富网股吧的发帖数据构建了投资者情绪指标,检验发现投资者情绪对未来市场收益率、交易量和波动率均无预测能力,但可以影响当期收益率和交易量,他们还发现,个股收益率导致投资者情绪变化的原因. 钱宇等^[27] 研究了网络论坛用户的情绪一致性对收益率和波动性的影响. 姚加权等^[28] 构建了金融情绪词典,并基于上市公司年报和社交媒体文本构建了情绪指标,研究发现这些情绪指标对股票收益率、波动率和股价崩盘风险等都具有显著预测作用. 陈荣达等^[29] 则研究了互联网金融中的投资者情绪问题,发现该指标与互联网理财产品收益率呈负相关.

从上述文献回顾中可以看出,国内外学者对于网络用户发布内容所体现的投资者情绪进行了一定研究,但主要是将投资者不加区分的作为一个整体以衡量其情绪指标,在构建投资者情绪指标时,缺乏对投资者个体影响力的关注. 另外,目前投资者情绪对市场影响的研究尚未形成一致观点,有学者发现投资者情绪与未来收益率显著正

相关,但其他实证研究结果则表明投资者情绪与未来收益显著负相关,有必要针对中国市场投资者情绪的影响展开进一步研究.

1.2 意见领袖相关研究

意见领袖概念最早是由拉扎斯菲尔德等于 1948 年在《人民的选择》一书中提出的,是指在信息传递和人际互动过程中具有影响力的人. 该概念提出后,引起了不同领域学者的广泛关注,在营销学研究中,不少文献表明意见领袖能够促进新产品的推广^[16, 30, 31]、消费者对现有产品的认知^[32]. 在传播学领域,也有学者研究了意见领袖对于促进信息传递的作用, Liu 等^[33] 研究了社会网络中的信息传播现象,结果发现,相比没有意见领袖的情况,当意见领袖存在时社会网络中的信息传播速度大为增加. 在金融领域, Apesteguia 等^[34] 发现意见领袖的交易盈利状况会导致刺激其他投资者过度配置风险资产. Ouimet 和 Tate^[35] 则发现同事中掌握更多信息的人能够影响他人投资决策. 杨晓兰等^[36] 发现意见领袖能促进信息传播,放大货币政策对股市收益率的正向影响.

在发现意见领袖对于他人具有重要影响的基础上,学者们研究了其为何能够产生如此作用. 一部分文献认为信任感是导致意见领袖具有强大影响力的原因,如 Liu 等^[33] 发现,意见领袖能够促进产品口碑传播的原因是其言论被消费者所信赖. Zhao 等^[30] 通过模拟实验也表明意见领袖能够加速信息传播,不过前提是其具有较高的可信度. Kim 和 Viswanathan^[17] 发现,在网络众筹市场中具有相关专业经验的投资者更能获得他人的信任和跟随. 另一部分文献认为,意见领袖影响力的来源是其社会联系众多,他们认为,当个体拥有大量社会联系时,其能够与他人产生更多互动并具有较强影响力^[37, 38]. 另外,还有一些学者认为,意见领袖能够改变他人观点是因为他们之间具有相似性,如学历、籍贯或经历等,并且这种因相似而发生的改变具有持久性^[39, 40].

由于意见领袖在信息传递过程中具有重要作用,如何识别意见领袖也成为学者关心的问题. 不少学者采用社会网络分析中的中心性(centrality) 算法来识别意见领袖^[41],该类算法的特点是根据

网络节点对外连接的数量来判断其影响力,如度中心性(degree centrality)、接近中心性(closeness centrality)、特征向量中心性(eigencentrality)等都被用于意见领袖的识别研究中^[16, 37, 38, 42]。第二类识别意见领袖的方法是基于随机游走的算法,其典型代表为Brian 和 Page^[43]提出的PageRank 算法,该算法在计算网络节点影响力时不仅考虑了对外连接数量,同时还将其连接节点的影响力纳入计算,Shafiq 等^[44]比较了多种算法识别意见领袖的效果,结果表明PageRank 算法的表现优于中心性算法。另外,Kitsak 等^[45]认为,意见领袖并非网络中连接最多的节点,而是那些处于网络核心位置的节点,并提出用K-shell 算法识别意见领袖。

总体来说,目前这一领域的研究主要涉及意见领袖在产品营销和舆论传播方面的作用,分析其产生的原因和识别方法,缺乏对意见领袖在金融市场中作用的分析,特别是投资者对意见领袖言论的反应。

2 研究假设与研究设计

2.1 研究假设及相关指标构建

2.1.1 研究假设提出

如前文所述,现有文献认为意见领袖能够通过信任感^[33]、社会联系^[37]和相似性^[46]这3种渠道影响投资者行为。当意见领袖发表关于某上市公司的正面观点时,投资者受其影响也会提高对该公司的评价,进而买入并推高股票价格。反之,当意见领袖对该公司表示出负面态度时,投资者则倾向于卖出该公司股票,导致股价下跌。但是,与普通投资者的不同之处在于,意见领袖在投资者群体中具有强大影响力,其观点会被大量投资者获悉并采纳,对未来股价造成冲击。另外,部分投资者可能在不同途径转发意见领袖的言论,使其观点在投资者群体中反复传播,对股价产生持久影响。因此,本研究提出如下假设:

H1 相比普通投资者,意见领袖情绪能够对

未来股价收益产生显著并且长期的影响。

借鉴现有研究的结论,意见领袖情绪能够预测市场的原因可能是在于其比普通投资者掌握了更多价值信息,在这些市场并未发现的信息融入股价的过程中,意见领袖情绪就体现出预测能力。但是,另一方面,意见领袖也可能并不掌握价值信息,仅仅是凭借其强大的影响力引导投资者整体行为方向与自身观点趋同,从而表现出意见领袖情绪对未来股价的预测能力。因此,本研究提出以下假设。

H2a 意见领袖情绪比非意见领袖掌握更多价值信息;

H2b 意见领袖情绪并不比非意见领袖掌握更多价值信息。

关于意见领袖影响力的来源,现有文献认为信任感是意见领袖能够获得大众认同的重要原因^[30, 33],当意见领袖能够准确预测未来股价时,投资者会报以持续关注和追捧,而一旦意见领袖对市场预测失灵,则其影响力将会下降。另外,Barber 和 Odean 认为,由于卖空约束,个人投资者普遍更为关注正面消息而忽略负面消息^[3],相对于负面帖子,正面帖子能够为意见领袖带来更高的关注度和影响力。因此,本研究提出以下假设。

H3 对未来股价的判断准确率能提高意见领袖的影响力,并且正面帖子判断准确率对意见领袖影响力的作用要高于负面帖子判断准确率。

2.1.2 识别意见领袖

本研究选取我国访问量最大的股票网络论坛——东方财富网股吧^②为研究对象,首先根据论坛账户的发帖和回帖数据构建社会网络,将回帖视为回帖者对发帖者的连接,然后利用社会网络分析中PageRank 算法来计算账户的影响力,并选择其中最高的少量账户作为意见领袖。

选择PageRank 算法是基于以下两点考虑:一方面,PageRank 算法被广泛用于社会网络的研究中,文献表明该算法能准确识别意见领袖。另一方面,该算法非常适用于股票网络论坛这一场景,能够排除投资者利用多个账户回复自己对结果的干

^② 中国最热门的前3个股票网络论坛分别为:东方财富网股吧、和讯、金融界,其中东方财富网股吧的用户数和访问量最大^[13]。

扰. 在网络论坛中,不少投资者会注册多个账户,这些账户中通常只有一个主要使用,其余偶尔登录的账户则影响力极低,投资者可能在用一个主要账户发帖后再用其他次要账户回帖,人为提高帖子的关注度. 基于连接数量的算法仅能根据与账户联系的邻居数量确定其影响力,而 PageRank 算法在计算中也会考虑邻居自身的影响,从而降低了同一投资者次要账户对结果的影响.

本研究将论坛账户作为社会网络的节点,并根据发帖和回帖的记录构建社会网络连接,具体分为以下步骤.

1) 统计了每月所有用户在所有 A 股讨论区中的发帖和数据,依据以下 4 条规则清理回帖数据: (1) 考虑到论坛帖子信息具有时效性,参考 Chen 等^[6]的做法,本研究只保留发帖之后 2 天内的回帖记录; (2) 若账户 j 对账户 i 的同一个帖子回复多次,则忽略重复回帖,只保留第一次回帖记录; (3) 若账户 j 多次回复了账户 i 的多个不同帖子,则保留每个帖子下的第一个回帖记录; (4) 删除账户 i 对自身发布帖子的回帖.

2) 确定每个自然月内账户之间的连接及权重. 若账户 j 回复了账号 i 在当月发布的帖子,则从账户 j 到账户 i 建立一个单向链接. 当月内账户 j 回复账户 i 的次数就是链接的权重. 由此,构建每个月的有向带权重图.

3) 考虑到部分投资者可能间隔一段时间后才会再次发帖讨论,但公众对其印象并不会因其短暂的沉默而立即消失,在计算论坛账户影响力时不仅要考虑其当月被他人关注的情况,还应当考虑其以往受关注度. 因此,本研究借鉴现有文献的做法^[47],将账户 j 到账户 i 以往连接的权重乘以衰减函数 $e^{-\gamma(t-t')}$,即第 t 月时,账户 j 到账户 i 连接的权重 $w_{j,i}^t$ 计算公式为

$$w_{j,i}^t = \sum_{t' \in T}^t w_{j,i}^{t'} \times e^{-\gamma(t-t')} \quad (1)$$

其中 $t - t'$ 代表以往连接对应时间周期与当期的间隔,间隔越大,则以往连接权重衰减的程度就越

大,对当期的影响就越小. γ 代表衰减系数,该数值越大,则以往连接权重衰减的速度越快,这里取 $\gamma = 1$.^③ 需要说明的是,本研究只考虑最近 6 个月的连接权重数据,这是因为 6 个月前连接的权重值乘以衰减函数后已经趋近于 0. 同时,排除更久之前的数据可以避免在社会网络中引入大量长期不活跃的账户,增加计算难度.

根据上述 3 个步骤,本研究利用股票网络论坛的发帖回帖数据,建立了论坛账户的社会网络,构成一个动态有向带权重的社会网络. 在此基础上,本研究采用个性化 PageRank 算法计算网络中各个节点(账户)的影响力. 由于在股票网络论坛中,意见领袖能够获得认同的重要原因是公众认为其能够预测市场未来趋势. 因此,本研究以每个账户发帖的判断准确程度作为重启概率. 具体公式如下

$$PR_i = \alpha \sum_{j \in M_i} \frac{PR_j}{L_i} + (1 - \alpha) \times acc_i \quad (2)$$

其中 PR_i 代表账户 i 的影响力; M_i 代表所有连接到账户 i 的账户集合; L_i 代表账户 i 连接到其他账户的权重之和; N 是网络中账户总数; α 代表重启概率,一般取值为 0.85; acc_i 代表账户 i 发帖判断准确程度,以该账户 5 个月内总体发帖准确率衡量.

2.1.3 判断意见领袖情绪

在识别出意见领袖的基础上,还需要根据其发布帖子内容判断意见领袖的情绪倾向. 由于金融领域用语与日常交流有所不同,直接根据语言学通用词典统计词频或其他通用情感判断模型会产生较大误差^[48],本研究采用 Stack 集成学习的思路,综合利用 THULAC 中文分词工具^④、Word2vec 工具和 LSTM 算法判断论坛意见领袖和普通用户的帖子情绪,通过分为文本分词、计算词向量和训练模型判断帖子情绪 3 个步骤,构建 Stack-LSTM 模型. 具体过程如下.

1) 利用 THULAC 中文分词工具将所有帖子的内容切分为各个词组,并删除了不影响语句含

③ 本研究也尝试了 0.5 和 2 等其他数值,对检验结果没有显著影响.

④ 该工具是由清华大学自然语言处理与社会人文计算实验室推出的,分词准确率达到 94% 以上.

义的停用词,以避免无关信息对算法模型的干扰.

2) 利用 Word2vec 工具在随机抽取的 500 万条论坛帖子文本词组中进行训练,用得到的模型计算所有帖子中每个词组的词向量,从而将每一个帖子的文本内容都用词向量矩阵的形式表达.

3) 采用 Stack 集成学习的思路构建帖子情绪判断模型,由于帖子情绪分为中性、正面和负面 3 类,为了提高判断准确率,本研究将判断帖子情绪这样一个三分类问题转化为 3 个二分类问题,利用 LSTM 算法构建了 3 个初级模型,分别用于判断正面或其他情绪、负面或其他情绪,以及中性或其他情绪这 3 个二分类问题,每个 LSTM 模型均包含 2 个隐藏层且每层含有 100 个神经元. 在此基础上,根据 Stack 集成学习的思路,将这 3 个模型的输出结果输入线性回归模型,得到对帖子情

绪的最终判断结果.

关于模型训练方面,本研究随机选取了 10 000 条帖子并由 5 名具有股票市场交易经验的投资者判断其蕴含情绪^⑤. 若投资者的判断结果存在差异,则以多数人的意见作为最终判断结果. 在此基础上,为了评估 Stack-LSTM 模型的判断性能,本研究采用 5 倍交叉验证的方式,将 10 000 个帖子分为 5 等份,每次取其中 4 份作为训练数据建立模型,剩余 1 份作为验证数据检验模型的表现. 然后取 5 次验证结果的均值作为 Stack-LSTM 模型判断性能的最终结果,并与其它主流算法建立的模型进行对比. 发现本研究所采用 Stack-LSTM 模型的判断准确率达到 76%,并且在准确率、精确率和 F1 得分等方面均优于随机森林、SVM 和 XGBoost 等主流算法. 具体结果如表 1 所示.

表 1 各类模型在验证数据中的判断性能

Table 1 The performance of the classification in the validation data

Panel A: 各类模型的准确率				
交叉验证编号	Stack-LSTM	随机森林	SVM	XGBoost
1	0.753 5	0.486 0	0.584 5	0.658 5
2	0.766 5	0.499 5	0.600 0	0.659 0
3	0.760 5	0.507 0	0.587 0	0.667 5
4	0.746 5	0.503 0	0.586 0	0.663 5
5	0.763 5	0.490 0	0.613 5	0.667 5
平均值	0.758 1	0.497 1	0.594 2	0.663 2
Panel B: 各类模型的精确率(Precision)				
1	0.747 6	0.527 5	0.832 6	0.648 0
2	0.759 5	0.544 1	0.840 4	0.648 1
3	0.745 6	0.601 3	0.835 6	0.662 1
4	0.737 2	0.539 9	0.831 3	0.650 1
5	0.745 0	0.506 9	0.850 5	0.646 7
平均值	0.747 0	0.543 9	0.838 1	0.651 0
Panel C: 各类模型的 F1 得分				
1	0.741 5	0.373 9	0.494 6	0.635 9
2	0.750 1	0.370 2	0.505 4	0.634 3
3	0.742 8	0.380 5	0.508 2	0.646 4
4	0.733 8	0.376 8	0.502 7	0.639 9
5	0.740 9	0.368 7	0.515 6	0.638 6
平均值	0.741 8	0.374 0	0.505 3	0.639 0

注: 准确率是指模型预测正确的样本占总体样本的比例; 精确率是指 3 个类型帖子判断精确率的均值; F1 得分是指 3 个类型帖子 F1 得分的均值.

^⑤ 在这 10 000 个帖子中,包含正面帖子 2 234 条、负面帖子 3 211 条以及中性帖子 4 555 条.

在利用 Stack-LSTM 判断每个帖子情绪的基础上,本研究参考现有文献^[6, 8],采用以下公式计算意见领袖和非意见领袖的情绪指标

$$\begin{aligned} pos_kol &= \frac{(\text{意见领袖正面帖子数}-\text{意见领袖负面帖子数})}{(\text{意见领袖正面帖子数}+\text{意见领袖负面帖子数})} \\ pos_other &= \frac{(\text{其他人正面帖子数}-\text{其他人负面帖子数})}{(\text{其他人正面帖子数}+\text{其他人负面帖子数})} \end{aligned}$$

2.2 样本选取与数据来源

本研究选取东方财富网股吧论坛中关于沪深两市 A 股的帖子,时间自 2011 年 1 月 1 日~2017 年 12 月 31 日,通过网络爬虫程序收集帖子 1.6 亿余条,其中 7 392 万条是发帖,其余是回帖,发帖与回帖用户数超过 850 万。考虑到媒体新闻对论坛帖子内容的影响,本研究还收集了《上海证券报》、《中国证券报》、《证券日报》、《证券时报》和

《人民日报》在 2011 年到 2017 年期间的所有新闻报道共 419 907 条,剔除不含上市公司股票代码和名称的新闻后获得 204 120 条新闻报道。上市公司股价、财务和分析师报告等数据均来自 CSMAR 数据库。

3 实证分析及结果

3.1 描述性统计

表 2 展示了个股收益率、意见领袖情绪和非意见领袖情绪的描述性统计结果。其中,意见领袖情绪(pos_kol) 和非意见领袖情绪(pos_other) 的均值分别为 0.044 6 和 0.018 9,说明投资者整体态度略微偏正面。另外,为降低极端值的影响,本研究对个股收益率(ret) 在 1% 和 99% 分位上做了缩尾处理。

表 2 主要变量的描述性统计
Table 2 The statistical analysis of main variables

变量	观测值	均值	标准差	最小值	最大值
ret	171 798	1.414 8	13.870 1	-37.689 6	47.695 8
pos_kol	181 612	0.044 6	0.440 7	-1.000 0	1.000 0
pos_other	181 612	0.018 9	0.334 5	-1.000 0	1.000 0

3.2 意见领袖情绪对股票收益率的预测作用

本研究将个股收益率作为因变量,检验意见领袖情绪对其影响。

借鉴现有文献^[5, 13]的做法,本研究根据以下模型检验意见领袖情绪对股票收益率的影响。

$$ret_{i,t+n} = \alpha + \beta_1 pos_kol_{i,t} + \beta_2 pos_other_{i,t} + \sum_{j=1}^n \delta_j X_{i,j,t} + \varepsilon_{i,t} \quad (3)$$

其中 $ret_{i,t+n}$ 代表股票 i 未来收益率,包括未来 1 个月、未来 1 个月~3 个月和未来 4 个月~6 个月累积收益率等 3 个指标; $pos_kol_{i,t}$ 和 $pos_other_{i,t}$ 分别代表意见领袖和非意见领袖在 t 月关于股票 i 的帖子情绪,本研究以上月论坛社

会网络中影响力最高的前 1% 账户作为当月意见领袖,其余当月发帖账户则作为非意见领袖,从而避免选择偏差问题; $X_{i,j,t}$ 是一系列控制变量,包括: 媒体情绪($media_senti$)、个股收益率的滞后项、市场收益率的滞后项、分析师正面评价数量($upgrade$)、分析师负面评价数量($downgrade$)、账面市值比(bm)、换手率($turnover$)。另外,还控制了月份($month$) 和行业($industry$) 的影响,并在公司层面做了聚类($Cluster at firm$) 以控制个体效应。

本研究分别以未来 1 个月、未来 1 个月~3 个月和未来 4 个月~6 个月的累积收益率为因变量,利用模型(3) 检验意见领袖情绪的作用,具体如表 3 所示。

表3 意见领袖情绪对股票收益率的影响

Table 3 The impact of opinion leaders' sentiment on stock returns

变量	(1)	(2)	(3)	(4)	(5)	(6)
	ret_{t+1}	$ret_{t+1,t+3}$	$ret_{t+4,t+6}$	ret_{t+1}	$ret_{t+1,t+3}$	$ret_{t+4,t+6}$
pos_kol	0.479 0 *** (6.91)	1.194 0 *** (8.55)	0.713 4 *** (4.90)			
pos_other	-0.575 5 *** (-6.54)	-1.154 3 *** (-5.98)	-1.097 3 *** (-5.77)			
pos_all				-0.477 6 *** (-5.39)	-0.996 7 *** (-4.97)	-0.941 3 *** (-4.78)
ret_t	-0.067 2 *** (-20.70)	-0.258 8 *** (-43.93)	-0.031 9 *** (-5.65)	-0.066 0 *** (-20.39)	-0.255 6 *** (-43.59)	-0.030 6 *** (-5.43)
ret_{t-1}	-0.025 1 *** (-8.14)	-0.176 1 *** (-28.94)	0.145 6 *** (27.67)	-0.024 6 *** (-7.99)	-0.175 2 *** (-28.79)	0.146 0 *** (27.74)
mkt_ret_t	0.371 5 *** (71.60)	0.337 9 *** (37.42)	0.180 5 *** (19.22)	0.370 7 *** (71.48)	0.335 6 *** (37.21)	0.179 8 *** (19.15)
mkt_ret_{t-1}	-0.157 9 *** (-33.26)	0.051 7 *** (5.43)	0.022 3 *** (2.60)	-0.158 7 *** (-33.40)	0.049 9 *** (5.24)	0.022 1 *** (2.58)
$upgrade$	0.336 8 *** (6.41)	0.722 7 *** (7.53)	0.087 9 (0.92)	0.335 8 *** (6.39)	0.720 8 *** (7.50)	0.086 9 (0.91)
$downgrade$	-0.341 4 *** (-6.85)	-0.946 1 *** (-9.08)	-0.265 5 ** (-2.54)	-0.340 3 *** (-6.84)	-0.942 7 *** (-9.08)	-0.263 2 ** (-2.52)
$media_senti$	0.021 3 (0.40)	-0.366 3 *** (-3.47)	-0.412 0 *** (-3.78)	0.021 3 (0.40)	-0.365 8 *** (-3.46)	-0.414 5 *** (-3.81)
bm	1.773 2 *** (12.02)	5.491 4 *** (11.66)	7.691 1 *** (12.35)	1.797 4 *** (12.13)	5.555 1 *** (11.75)	7.779 7 *** (12.36)
$turnover$	-0.114 1 *** (-8.14)	-0.020 2 (-0.52)	0.041 4 (1.22)	-0.115 0 *** (-8.19)	-0.023 8 (-0.61)	0.039 4 (1.16)
截距项	6.434 0 *** (28.45)	9.272 4 *** (13.62)	-2.788 6 *** (-4.51)	6.472 4 *** (28.60)	9.365 6 *** (13.75)	-2.749 7 *** (-4.43)
$month$	Yes	Yes	Yes	Yes	Yes	Yes
$Industry$	Yes	Yes	Yes	Yes	Yes	Yes
$Cluster at firm$	Yes	Yes	Yes	Yes	Yes	Yes
样本数	161 438	146 515	134 769	161 438	146 515	134 769
R^2	0.080 2	0.082 6	0.056 6	0.079 9	0.082 1	0.056 4

注:括号中为 t 值,***、**、* 分别表示 1%、5% 和 10% 显著性水平。

根据上表结果,在控制影响个股累积收益率的相关变量后,列(1)、列(2)和列(3)中意见领袖情绪(pos_kol)的系数分别为 0.479 0、1.194 0 和 0.713 4,并且均在 1% 水平上显著,表明意见领袖情绪不仅能够在预测短期股价收益率,并且对长期股价收益率也具有预测能力。非意见领袖情绪(pos_other)的系数为 -0.575 5、-1.154 3 和 -1.097 3,并且也都在 1% 水平上显著,表明非

意见领袖情绪与股票未来收益率负相关。根据列(2)和列(3)结果,股票网络论坛中意见领袖正面负面帖子之差所占比例每提高 10%,相应个股未来 1 个月~3 个月和未来 4 个月~6 个月的累积收益率将平均提高 0.119% 和 0.071%,说明意见领袖情绪能够预测个股未来收益,支持了假设 H1 的推断。

在列(4)、列(5)和列(6)展示了股票网络论

坛整体投资者情绪(*pos_all*)对未来股票收益率的影响。结果表明,整体投资者情绪(*pos_all*)与股价未来1个月、未来1个月~3个月和未来4个月~6个月收益率都是负相关,说明投资者整体表现出过度反应的现象,当情绪高涨时,买入者增加导致股价高估,并随后发生反转。该结论与金德环和李岩^[26]的发现并不一致,这可能是由于本研究的因变量为累积收益率,而他们以累积超额收益率作为因变量,同时,本研究判断意见领袖情绪的模型算法也与其有所不同。

3.3 意见领袖情绪能够预测市场的原因

意见领袖情绪能够预测股价可能是因为其观点中包含了尚未被市场发现的信息,也可能仅仅是其能够引导投资者行为,从而表现出自我实现预期(self-fulfilling prophecies)的特点。本研究以上市公司发布季度业绩公告为研究事件,检验公告发布前的意见领袖情绪能否预测意外盈余,由此判断意见领袖情绪中是否包含价值信息。

参考Chen等^[6]、金德环和李岩^[26]的思路,本研究采用以下模型研究意见领袖情绪对意外盈余的影响。

$$SUE_{i,t} = \alpha + \beta_1 pos_kol_{i,t-3,t-90} + \beta_2 pos_other_{i,t-3,t-90} + \sum_{j=1}^n \delta_j X_{i,j,t} + \varepsilon_{i,t} \quad (4)$$

其中 $SUE_{i,t}$ 代表上市公司 i 在 t 日业绩公告的意外盈余,是上市公司季度业绩公告中的每股盈余减去分析师预测均值,再除以该季度末股价;

$pos_kol_{i,t-3,t-90}$ 代表业绩公告发布日前3日~前90日内意见领袖的帖子情绪; $pos_other_{i,t-3,t-90}$ 代表业绩公告发布日前3日~前90日内非意见领袖帖子情绪; $X_{i,j,t}$ 是控制变量,包括: 媒体情绪(*media_senti*)、上一年末的账面市值比(*bm*)、上一季度市值(*size*)、业绩公告发布前3日~前90日期间的股票累积收益率(*cret_{t-3,t-90}*)、上一季度意外盈余(*SUE_{i,t-1}*),还有分析师意见分歧(*ana_disper*),是指业绩公告发布日前3日~前90日期间,经股价调整后的分析师预测每股盈余的标准差。另外,还控制了年份(*year*)和月份(*month*)的影响,并在公司层面做了聚类(*Cluster at firm*)以控制个体效应。

模型(4)的检验结果如表4所示。其中,列(1)结果表明,意见领袖情绪(*pos_kol*)和非意见领袖情绪(*pos_other*)对意外盈余的影响均不显著。列(3)和列(4)中,为了排除无关帖子的影响,在剔除了不包含“业绩”这一关键词的帖子后重新计算了意见领袖情绪(*pos_kol*)和非意见领袖情绪(*pos_other*),此时,意见领袖情绪(*pos_kol*)和非意见领袖情绪(*pos_other*)都与意外盈余显著负相关。列(5)和列(6)排除了ST、创业板和金融行业上市公司样本,结果表明意见领袖情绪(*pos_kol*)仍然与意外盈余显著负相关。同时,在列(4)和列(6)中,整体投资者情绪(*pos_all*)也均与意外盈余显著负相关。上述结果说明,意见领袖并不比其他投资者掌握更多关于上市公司的价值信息,假设H2b得到了印证。

表4 意见领袖情绪对意外盈余的预测能力

Table 4 The prediction ability of opinion leaders' sentiment on earnings surprise

变量	(1)	(2)	(3)	(4)	(5)	(6)
	<i>SUE</i>	<i>SUE</i>	<i>SUE</i>	<i>SUE</i>	<i>SUE</i>	<i>SUE</i>
<i>pos_kol</i>	-0.000 3 (-0.36)		-0.005 4 *** (-5.74)		-0.005 6 *** (-5.09)	
<i>pos_other</i>	-0.000 9 (-1.17)		-0.002 3 *** (-5.16)		-0.002 7 *** (-5.34)	
<i>pos_all</i>		0.000 1 (0.07)		-0.004 2 *** (-6.95)		-0.004 7 *** (-6.76)
<i>media_senti</i>	-0.000 8 *** (-3.40)	-0.000 7 *** (-3.26)	-0.001 1 *** (-3.32)	-0.001 1 *** (-3.65)	-0.001 1 *** (-2.98)	-0.000 9 *** (-3.01)
<i>SUE_{t-1}</i>	0.333 2 *** (7.12)	0.336 5 *** (7.20)	0.209 4 *** (4.74)	0.272 2 *** (5.31)	0.206 3 *** (3.80)	0.258 1 *** (4.26)

续表4

Table 4 Continues

变量	(1)	(2)	(3)	(4)	(5)	(6)
	SUE	SUE	SUE	SUE	SUE	SUE
<i>bm</i>	-0.002 1 (-1.40)	-0.002 2 (-1.44)	0.003 2 (1.55)	0.000 6 (0.33)	0.001 8 (0.80)	-0.000 8 (-0.44)
<i>ana_disper</i>	-6.550 6 *** (-8.53)	-6.570 0 *** (-8.54)	-6.640 3 *** (-6.37)	-6.098 8 *** (-7.76)	-6.580 4 *** (-6.02)	-5.866 3 *** (-7.92)
<i>size</i>	0.006 2 *** (13.54)	0.006 1 *** (13.68)	0.005 5 *** (11.43)	0.005 6 *** (11.40)	0.005 0 *** (9.58)	0.005 1 *** (9.73)
<i>cu_ret</i>	0.006 6 *** (6.85)	0.006 9 *** (7.17)	0.003 7 *** (2.90)	0.004 1 *** (3.98)	0.004 7 *** (3.31)	0.004 8 *** (3.79)
截距项	-0.090 0 *** (-12.40)	-0.089 6 *** (-12.79)	-0.074 7 *** (-9.70)	-0.078 6 *** (-9.89)	-0.066 1 *** (-7.99)	-0.070 3 *** (-8.32)
<i>year</i>	Yes	Yes	Yes	Yes	Yes	Yes
<i>Industry</i>	Yes	Yes	Yes	Yes	Yes	Yes
<i>Cluster at firm</i>	Yes	Yes	Yes	Yes	Yes	Yes
样本数	48 163	48 897	24 828	30 628	19 408	23 900
<i>R</i> ²	0.190 2	0.191 2	0.141 4	0.166 7	0.110 2	0.133 2

注: 括号中为 *t* 值, ***、**、* 分别表示 1%、5% 和 10% 显著性水平.

3.4 投资者对意见领袖预测准确程度的反应

根据意见领袖影响力的计算方式,若投资者不再认同该意见领袖,则其影响力应下降,反之,其影响力将会提高,因此,投资者对意见领袖的整体反应可以通过其未来影响力的变化来衡量. 本研究利用模型(5)检验意见领袖判断准确率对其未来影响力的作用. 同时,由于卖空的成本较高或门槛较高,个人投资者普遍更为关注正面消息而忽略负面消息^[3],本研究利用模型(6)检验正确或错误的正面帖子和负面帖子数量对意见领袖影响力的作用.

本研究利用以下模型检验预测准确率能否提高意见领袖影响力

$$PR_{j,t+1} = \alpha + \beta_1 acc_{j,t} + \sum_{j=1}^n \delta_j X_{i,j,t} + \varepsilon_{i,t} \quad (5)$$

其中因变量 $PR_{j,t+1}$ 代表意见领袖 j 在 $t+1$ 月基于个性化 PageRank 算法得出的影响力; $acc_{j,t}$ 代表意见领袖 j 在 t 月所发帖子的判断准确率,分为总体发帖准确率(*acc_all*)、正面帖子准确率(*acc_pos*)和负面帖子准确率(*acc_neg*)3类指标,其中

总体发帖准确率(*acc_all*)代表当月判断正确帖子占全部发帖的比例,若帖子情绪为正面且次日相应个股收益率为正^⑥,则该帖子视为正确帖子,否则就是错误帖子; 正面帖子准确率(*acc_pos*)和负面帖子准确率(*acc_neg*)指标计算方法与之类似; $X_{i,j,t}$ 为控制变量,包括: 当月意见领袖发帖总数(*msg_num*)、当月论坛发帖人数(*poster_num*)、当月市场累计收益率(*mkt_ret*). 另外,本研究还控制了年份(*year*)和月份(*month*)的影响.

本研究利用以下模型检验判断正确或错误的正面帖子、负面帖子数量对意见领袖影响力的作用

$$PR_{j,t+1} = \alpha + \beta_1 pos_r_{j,t} + \beta_1 pos_w_{j,t} + \beta_3 neg_r_{j,t} + \beta_3 neg_w_{j,t} + \sum_{j=1}^n \delta_j X_{i,j,t} + \varepsilon_{i,t} \quad (6)$$

其中 $pos_r_{j,t}$ 和 $pos_w_{j,t}$ 分别代表意见领袖 j 在 t 月所发正面帖子中判断正确和错误的数量; $neg_r_{j,t}$ 和 $neg_w_{j,t}$ 分别代表意见领袖 j 在 t 月所发负面帖子中判断正确和错误的数量; $X_{i,j,t}$ 为控制变量,与模型(5)相同.

模型(5)和模型(6)的检验结果如表 5 所示

⑥ 本研究根据未来 3 日累积收益率来判断帖子是否正确,统计检验结果是一致的.

表 5 投资者对意见领袖预测能力的反应

Table 5 The investors' reaction upon the prediction power of opinion leader

变量	(1)	(2)	(3)	(4)	(5)	(6)
	PR_{t+1}	PR_{t+1}	PR_{t+1}	PR_{t+1}	PR_{t+1}	PR_{t+1}
acc_all	0.060 0 (1.25)			0.004 6 (1.56)		
acc_pos		0.051 7 (1.52)			0.011 8 ** (2.36)	
acc_neg		-0.053 8 (-1.38)			-0.000 3 (-0.04)	
pos_r			0.039 8 *** (9.02)			0.053 4 *** (14.17)
pos_w			0.034 1 *** (6.90)			0.049 1 *** (13.88)
neg_r			0.011 6 *** (3.77)			0.018 3 *** (7.87)
neg_w			0.011 5 *** (3.23)			0.014 6 *** (4.03)
msg_num	0.764 4 *** (28.43)	0.872 2 *** (25.89)		0.502 4 *** (68.23)	0.679 8 *** (57.90)	
$poster_num$	-3.104 2 *** (-19.13)	-3.043 7 *** (-19.96)	-3.060 4 *** (-19.00)	-0.860 9 *** (-32.14)	-1.197 2 *** (-29.07)	-0.842 3 *** (-31.64)
mkt_ret	0.005 3 *** (4.48)	0.007 0 *** (5.05)	0.005 2 *** (4.26)	0.001 6 *** (8.87)	0.003 0 *** (7.76)	0.001 1 *** (5.34)
截距项	57.468 4 *** (27.49)	56.970 4 *** (27.05)	58.589 5 *** (27.81)	14.262 5 *** (42.94)	19.699 1 *** (36.91)	14.591 5 *** (43.59)
$year$	Yes	Yes	Yes	Yes	Yes	Yes
$month$	Yes	Yes	Yes	Yes	Yes	Yes
样本数	366 452	315 802	366 452	3 077 369	1 532 542	3 077 369
R^2	0.274 9	0.335 4	0.282 3	0.146 3	0.201 5	0.156 8

注: 括号中为 t 值, ***、**、* 分别表示 1%、5% 和 10% 显著性水平.

根据上表结果,列(1) 和列(2) 表明,不论是总体发帖准确率(acc_all)、正面帖子准确率(acc_pos) 或是负面帖子准确率(acc_neg) 都对其下个月影响力(PR_{t+1}) 没有显著影响.列(3) 中结果表明,不论是判断正确的正面帖子(pos_r) 还是判断错误的正面帖子(pos_w) 都能显著提高其下个月的影响力,并且这两个变量的系数大致相当,另一方面,判断正确的负面帖子(neg_r) 和判断错误的负面帖子(neg_w) 也都能显著提高意见领袖下个月的影响力,但是其影响系数远低于两类正面帖子.列(4)、列(5) 和列(6) 则检验了论坛所有用户的预测准确率与其下个月影响力的关系,结果与列(1) 至列

(3) 基本一致.这说明,股票网络论坛中的意见领袖受到广泛关注并非因为其能够准确预测市场,而是因为其在论坛中的活跃表现,特别是发布关于股价看涨的帖子能够极大提升其影响力.这一结论反应出投资者并不能理性判断他人观点正确与否,而是片面认同发布正面消息的意见领袖,对能够准确判断市场下跌趋势的观点则较为忽视,假设 H3 部分得到了支持.

另外,相关性分析表明,意见领袖当月发帖准确率与其下个月影响力相关性较为微弱,而当月正面帖子准确率和负面帖子准确率与其下个月影

响力则显著为正和负^⑦,这与表5结果一致.

3.5 稳健性检验

3.5.1 改变论坛账户影响力计算方法

考虑到前文采用的个性化PageRank算法对结果的影响,本研究根据无个性化的PageRank算法计

算论坛账户的影响力,并仍选取上月影响力居前的1%账户作为当月意见领袖,表6展示了其情绪与股票收益率关系的检验结果.结果表明,意见领袖情绪与个股未来收益率正相关,而非意见领袖情绪和整体投资者情绪则与个股未来收益率负相关.

表6 按其他方法选出意见领袖的情绪对股票收益率的影响

Table 6 The impact of sentiment from opinion leaders using alternative method on stock returns

变量	(1)	(2)	(3)	(4)	(5)	(6)
	ret_{t+1}	$ret_{t+1,t+3}$	$ret_{t+4,t+6}$	ret_{t+1}	$ret_{t+1,t+3}$	$ret_{t+4,t+6}$
pos_kol	0.432 1 *** (6.18)	1.009 2 *** (7.22)	0.611 5 *** (4.16)			
pos_other	-0.531 3 *** (-6.06)	-1.121 6 *** (-5.87)	-1.069 9 *** (-5.70)			
pos_all				-0.477 6 *** (-5.39)	-0.996 7 *** (-4.97)	-0.941 3 *** (-4.78)
截距项	6.445 9 *** (28.50)	9.308 3 *** (13.68)	-2.755 7 *** (-4.47)	6.472 4 *** (28.60)	9.365 6 *** (13.75)	-2.749 7 *** (-4.43)
控制变量	Yes	Yes	Yes	Yes	Yes	Yes
<i>month</i>	Yes	Yes	Yes	Yes	Yes	Yes
<i>Industry</i>	Yes	Yes	Yes	Yes	Yes	Yes
<i>Cluster at firm</i>	Yes	Yes	Yes	Yes	Yes	Yes
样本数	161 438	146 515	134 769	161 438	146 515	134 769
R^2	0.080 1	0.082 5	0.056 5	0.079 9	0.082 1	0.056 4

注:括号中为t值,***、**、*分别表示1%、5%和10%显著性水平.

3.5.2 改变意见领袖的选取标准

考虑到意见领袖的选取标准可能对结果产生影响,本研究改变了意见领袖的选取标准,重新运行模型1,结果如表7所示.在列(1)、列(2)和列(3)中,本研究将意见领袖的选取标准扩大为上月影响力最高的前5%账户,结果表明,前5%意见领袖情绪(pos_kol5)同样与个股未来1个月和

3个月累积收益率显著正相关,但影响系数小于表3中的前1%意见领袖(pos_kol),说明意见领袖对未来股票收益率的影响随着其影响力的提高而增强,这可能是由于影响力较高的意见领袖能够引导更多投资者行动.在列(4)~列(6)中,本研究将上月影响力最高的前10%账户作为意见领袖(pos_kol10),结果与前文一致.

表7 前5%或前10%意见领袖情绪对股票收益率的影响

Table 7 The impact of sentiment from top 5% or 10% as opinion leaders on stock returns

变量	(1)	(2)	(3)	(4)	(5)	(6)
	ret_{t+1}	$ret_{t+1,t+3}$	$ret_{t+4,t+6}$	ret_{t+1}	$ret_{t+1,t+3}$	$ret_{t+4,t+6}$
pos_kol5	0.277 7 *** (3.61)	0.622 7 *** (4.10)	0.360 9 ** (2.38)			
$pos_other95$	-0.606 6 *** (-7.11)	-1.138 3 *** (-6.28)	-1.134 7 *** (-6.42)			
pos_kol10				0.250 6 *** (3.25)	0.362 3 ** (2.37)	0.311 6 ** (2.01)

⑦ 限于篇幅,本研究没有展示相关性分析的结果,欢迎联系作者索取.

续表 7

Table 7 Continues

变量	(1)	(2)	(3)	(4)	(5)	(6)
	ret_{t+1}	$ret_{t+1,t+3}$	$ret_{t+4,t+6}$	ret_{t+1}	$ret_{t+1,t+3}$	$ret_{t+4,t+6}$
<i>pos_other90</i>				-0.613 1 *** (-7.23)	-1.061 5 *** (-6.13)	-1.197 6 *** (-7.21)
截距项	-0.114 5 *** (-8.17)	-0.017 8 (-0.46)	0.041 3 (1.22)	-0.114 3 *** (-8.15)	-0.017 9 (-0.46)	0.041 1 (1.21)
控制变量	Yes	Yes	Yes	Yes	Yes	Yes
<i>month</i>	Yes	Yes	Yes	Yes	Yes	Yes
<i>Industry</i>	Yes	Yes	Yes	Yes	Yes	Yes
<i>Cluster at firm</i>	Yes	Yes	Yes	Yes	Yes	Yes
样本数	161 438	146 515	134 769	161 438	146 515	134 769
R^2	0.080 0	0.082 3	0.056 6	0.080 0	0.082 3	0.056 6

注: 括号中为 t 值, ***、**、* 分别表示 1%、5% 和 10% 显著性水平.

3.5.3 用百度 Senta 工具^⑧衡量意见领袖情绪

前文中的意见领袖情绪是采用本研究自建的 Stack-LSTM 模型判断分类, 训练数据的选择和算法优化建模过程可能影响研究结果, 表 8 中采用了百度 Senta 工具分别判断意见领袖和非意见领袖的帖子情绪类型, 从而重新计算出意见领袖情绪指标(*pos_kol*) 和非意见领袖情绪指标(*pos_other*), 并根据模型(3)再次检验了意见领袖情绪与股票收益率的关系. 列(1) 和列(3) 中, 意见领袖

情绪(*pos_kol*) 的系数分别为 0.438 1 和 0.743 7, 并且均在 1% 水平上显著, 说明意见领袖情绪与未来短期和长期股价收益率显著相关. 在列(2) 和列(3) 中, 非意见领袖情绪(*pos_other*) 与股价未来收益率显著负相关. 另外, 在列(4) 至列(6) 中, 整体投资者情绪(*pos_all*) 与股价未来收益率呈现无关或负相关关系. 这些结论与前文一致, 表明相比非意见领袖, 意见领袖情绪能够显著且持久的预测股价未来收益率.

表 8 基于百度 Senta 的意见领袖情绪对股票收益率的影响

Table 8 The impact of opinion leaders' sentiment based on Baidu Senta on stock returns

变量	(1)	(2)	(3)	(4)	(5)	(6)
	ret_{t+1}	$ret_{t+1,t+3}$	$ret_{t+4,t+6}$	ret_{t+1}	$ret_{t+1,t+3}$	$ret_{t+4,t+6}$
<i>pos_kol</i>	0.438 1 *** (5.93)	0.229 6 (1.53)	0.743 7 *** (5.18)			
<i>pos_other</i>	-0.202 0 (-1.63)	-1.369 5 *** (-5.13)	-0.749 6 *** (-2.86)			
<i>pos_all</i>				0.196 5 (1.51)	-0.892 8 *** (-3.07)	-0.182 1 (-0.63)
截距项	6.484 5 *** (28.24)	8.792 2 *** (12.85)	-2.976 3 *** (-4.71)	6.506 4 *** (28.28)	8.906 4 *** (12.99)	-2.981 5 *** (-4.71)
控制变量	Yes	Yes	Yes	Yes	Yes	Yes
<i>month</i>	Yes	Yes	Yes	Yes	Yes	Yes
<i>Industry</i>	Yes	Yes	Yes	Yes	Yes	Yes
<i>Cluster at firm</i>	Yes	Yes	Yes	Yes	Yes	Yes
样本数	161 825	147 794	136 007	161 825	147 794	136 007
R^2	0.079 9	0.081 8	0.056 3	0.079 7	0.081 7	0.056 2

注: 括号中为 t 值, ***、**、* 分别表示 1%、5% 和 10% 显著性水平.

⑧ Senta 是百度公司于 2018 年推出的自然语言处理工具, 通过深度学习算法在大规模语料库上训练得到, 在百度自建数据集上的效果分类准确率为 90%. 来源: <https://github.com/baidu/Senta>.

3.5.4 用以往 30 天帖子检验意见领袖情绪与季报意外盈余的关系

考虑到以往 90 天的信息过于陈旧,投资者和分析师尚不能准确评估上市公司意外盈余,表 9 使用了业绩公告发布前 3 日 ~ 前 30 日期间的数

表 9 最近 30 天意见领袖情绪对意外盈余的预测能力

Table 9 The prediction ability of opinion leaders' sentiment in recent 30 days on earnings surprise

变量	(1)	(2)	(3)	(4)
	SUE	SUE	SUE_diff	SUE_diff
pos_kol	-0.005 6 *** (-5.09)		-0.003 1 *** (-2.84)	
pos_other	-0.002 7 *** (-5.34)		-0.002 1 *** (-4.82)	
pos_all		-0.004 7 *** (-6.76)		-0.003 7 *** (-6.20)
截距项	-0.066 1 *** (-7.99)	-0.070 3 *** (-8.32)	0.011 2 *** (3.18)	0.010 4 *** (3.15)
控制变量	Yes	Yes	Yes	Yes
year	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes
Cluster at firm	Yes	Yes	Yes	Yes
样本数	19 408	23 900	19 400	23 890
R ²	0.110 2	0.133 2	0.095 9	0.120 6

注: SUE_diff = (实际 EPS - 去年同季度 EPS) / 股价.

3.5.5 投资者对前 5% 或前 10% 意见领袖预测能力的反应

表 10 以上月影响力最高的前 5% 或前 10% 账户作为意见领袖,检验其发帖准确率与未来收益率的关系. 其中,列(1) ~ 列(3) 是以影响力最高的前 5% 账户作为意见领袖,总体发帖准确率(acc_all) 对其下个月影响力没有显著影响,而正面帖子准确率(acc_pos) 则与其下个月影响正相关,同时负面帖子准确率(acc_neg) 与其下个月影响负相关. 另外,判断正确的正面帖子(pos_r) 和判断错误的正面帖子(pos_w) 都能显著提高其下个月影响力,并且其系数都大于判断正确的负面帖子(neg_r) 和判断错误的负面帖子(neg_w) . 列(4) ~ 列(6) 将意见领袖标准调整为前 10% 账户,结果与列(1) ~ 列(3) 一致. 上述结果表明,普通投资者对正面和负面消息的关注程度存在明显的不对称性,甚至抵制能够准确预测市场下跌的意见领袖,反而会认同发表错误积极言论的意见领袖.

本研究揭示了投资者情绪对个股未来收益率

据计算意见领袖情绪,并以分析师盈余预测的平均值和去年同一季度实际盈余(随机游走模型)两种方法来构造意外盈余. 列(1) 和列(3) 中,意见领袖情绪(pos_kol) 与因变量均为负相关,结论与前文一致.

具有一定预测能力,这一结果与 Das 和 Chen^[9]、部慧等^[14]不同,主要原因在于本研究根据社会影响力对投资者进行了区分,只有少数具有影响力的意见领袖能够预测市场未来变化,若考察投资者总体情绪的影响,则结果与现有研究一致. 进一步研究发现,这些意见领袖并没有掌握价值信息,他们能够预测市场可能是由于其强大的影响力引导投资者整体行为方向与自身观点趋同,从而使未来股价与其预期相符. 同时,本研究还发现,无论判断正确或错误,意见领袖发布的正面帖子都能够增强其影响力,甚至相比判断正确的负面帖子,错误的正面帖子能够产生更大影响. 上述研究结果说明,由于卖空约束,网络上的投资者群体明显更为偏好乐观的言论,而意见领袖可以通过发布夸张的看涨信息来吸引公众关注,即便这些信息事后被发现错误,也无损其影响力. 从这些行为和行为产生的结果看,本研究认为我国股吧论坛中的投资者群体存在非理性倾向,即他们宁可接收错误的乐观信息,也不愿看到正确的悲观信息.

表 10 投资者对前 5% 或前 10% 意见领袖预测能力的反应

Table 10 The investors' reaction upon the prediction power of top 5% or 10% as opinion leader

变量	(1)	(2)	(3)	(4)	(5)	(6)
	PR_{t+1}	PR_{t+1}	PR_{t+1}	PR_{t+1}	PR_{t+1}	PR_{t+1}
<i>acc_all</i>	0.003 1 (0.28)			0.002 6 (0.38)		
<i>acc_pos</i>		0.023 3 ** (2.39)			0.019 6 *** (2.82)	
<i>acc_neg</i>		-0.027 6 ** (-2.10)			-0.017 6 * (-1.84)	
<i>pos_r</i>			0.046 4 *** (12.33)			0.048 9 *** (13.10)
<i>pos_w</i>			0.041 0 *** (10.66)			0.043 7 *** (12.03)
<i>neg_r</i>			0.014 6 *** (5.69)			0.015 7 *** (6.47)
<i>neg_w</i>			0.013 4 *** (4.01)			0.013 9 *** (4.10)
截距项	29.969 1 *** (37.15)	31.489 9 *** (34.14)	30.524 7 *** (37.61)	23.114 7 *** (39.99)	25.429 7 *** (35.83)	23.566 7 *** (40.50)
控制变量	Yes	Yes	Yes	Yes	Yes	Yes
<i>year</i>	Yes	Yes	Yes	Yes	Yes	Yes
<i>Month</i>	Yes	Yes	Yes	Yes	Yes	Yes
样本数	1 204 670	894 387	1 204 670	1 758 291	1 179 238	1 758 291
R^2	0.217 5	0.262 6	0.228 4	0.190 5	0.233 8	0.202 4

注: 括号中为 t 值, ***、**、* 分别表示 1%、5% 和 10% 显著性水平.

4 结束语

本研究基于社会网络分析技术, 以 2011 年 1 月 1 日到 2017 年 12 月 31 日东方财富网股吧论坛的发帖数据和对应的上市公司特征数据为样本, 研究意见领袖情绪对未来个股收益率的预测性和原因, 以及其他投资者对意见领袖判断能力的反应. 研究结果发现: 1) 意见领袖情绪对未来 1 个月 ~6 个月个股收益率均有显著影响, 而非意见领袖情绪和整体投资者情绪则不能预测市场, 说明相比普通投资者, 意见领袖对未来股价收益率的判断更为准确. 2) 为了研究意见领袖能够预测市场的原因, 本研究以上市公司季度业绩公告为事件, 考察意见领袖和非意见领袖情绪与意外盈余的关系, 研究发现, 意见领袖情绪与意外盈余无关, 意见领袖并未比普通投资者掌握更多关于上市公司的价值信息, 他们对股价的影响力可能

是由于其观点能够引导投资者行为进而使股价同向变化. 3) 本研究进一步探究了投资者对意见领袖预测准确程度的反应, 研究发现, 意见领袖影响力程度与其总体判断准确率无关, 投资者群体更为关注发表正面观点的意见领袖而忽视其观点准确性, 同时排斥能够准确预测市场下跌的意见领袖.

本研究具有重要的理论与现实意义. 首先, 在理论上, 研究了意见领袖对资本市场的影响及其影响力成因, 从社会网络的角度拓展了投资者情绪的研究. 其次, 监管部门应该制定相关法律法规, 加强网络舆论监管, 规范意见领袖在网络上的言论, 并在必要时采取措施引导舆论方向, 互联网平台也应当配合监管部门进行舆情管理, 及时干预意见领袖的夸张言论, 以避免非理性情绪在网络上的蔓延. 最后, 监管部门有必要加强投资者教育工作, 使投资者特别是中小投资者对其所投资的金融产品风险有更充分了解, 避免盲目听信他人, 提高投资者的自我保护能力.

参考文献:

- [1] 于李胜,王成龙,王艳艳.分析师社交媒体在信息传播效率中的作用——基于分析师微博的研究 [J]. 管理科学学报, 2019, 22(7): 107–126.
Yu Lisheng, Wang Chenglong, Wang Yanyan. The role of analysts' social media in information dissemination: Evidence from analysts' microblogs [J]. Journal of Management Sciences in China, 2019, 22(7): 107–126. (in Chinese)
- [2] Barberis N, Shleifer A, Vishny R. A model of investor sentiment [J]. Journal of Financial Economics, 1998, 49(3): 307–343.
- [3] Barber B M, Odean T. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors [J]. Review of Financial Studies, 2008, 21(2): 785–818.
- [4] Huang D, Jiang F, Tu J, et al. Investor sentiment aligned: A powerful predictor of stock returns [J]. Review of Financial Studies, 2015, 28(3): 791–837.
- [5] Tetlock P C. Giving content to investor sentiment: The role of media in the stock market [J]. Journal of Finance, 2007, 62(3): 1139–1168.
- [6] Chen H, De P, Hu Y J, et al. Wisdom of crowds: The value of stock opinions transmitted through social media [J]. Review of Financial Studies, 2014, 27(5): 1367–1403.
- [7] Luo X, Zhang J, Duan W. Social media and firm equity value [J]. Information Systems Research, 2013, 24(1): 146–163.
- [8] Antweiler W, Frank M Z. Is all that talk just noise? The information content of internet stock message boards [J]. Journal of Finance, 2004, 59(3): 1259–1294.
- [9] Das S R, Chen M Y. Yahoo! For Amazon: Sentiment extraction from small talk on the web [J]. Management Science, 2007, 53(9): 1375–1388.
- [10] Tirunillai S, Tellis G J. Does chatter really matter? Dynamics of user-generated content and stock performance [J]. Marketing Science, 2012, 31(2): 198–215.
- [11] Bartov E, Faurel L, Mohanram P S. Can twitter help predict Firm-Level earnings and stock returns? [J]. The Accounting Review, 2017, 93(3): 25–57.
- [12] 张永杰,张维,金曦,等.互联网知道的更多么?——网络开源信息对资产定价的影响 [J].系统工程理论与实践, 2011, 31(4): 577–586.
Zhang Yongjie, Zhang Wei, Jin Xi, et al. Does the Internet know more? Open source information and asset pricing [J]. Systems Engineering: Theory & Practice, 2011, 31(4): 577–586. (in Chinese)
- [13] 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. (in Chinese)
- [14] 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. (in Chinese)
- [15] 张宗新,吴钊颖.媒体情绪传染与分析师乐观偏差——基于机器学习文本分析方法的经验证据 [J].管理世界, 2021, 37(1): 170–185.
Zhang Zongxin, Wu Zhaoying. Media's emotional contagion and analyst optimistic bias: Evidence based on the technique of machine learning [J]. Management World, 2021, 37(1): 170–185. (in Chinese)
- [16] Iyengar R, Van den Bulte C, Valente T W. Opinion leadership and social contagion in new product diffusion [J]. Marketing Science, 2011, 30(2): 195–212.
- [17] Kim K, Viswanathan S. The experts in the crowd: The role of experienced investors in a crowdfunding market [J]. MIS Quarterly, 2019, 43(2): 347–372.
- [18] 李双琦,陈其安,朱沙.考虑消费与投资者情绪的股票市场资产定价 [J].管理科学学报, 2021, 24(4): 86–108.
Li Shuangqi, Chen Qi'an, Zhu Sha. Asset pricing in the stock market with consumption and investor sentiment [J]. Journal of Management Sciences in China, 2021, 24(4): 86–108. (in Chinese)

- [19] Baker M, Wurgler J. Investor sentiment in the stock market [J]. *Journal of Economic Perspectives*, 2007, 21(2) : 129 – 151.
- [20] 刘晓星, 张 旭, 顾笑贤, 等. 投资者行为如何影响股票市场流动性? —— 基于投资者情绪、信息认知和卖空约束的分析 [J]. *管理科学学报*, 2016, 19(10) : 87 – 100.
Liu Xiaoxing, Zhang Xu, Gu Xiaoxian, et al. How does investor behavior affect stock market liquidity? Analysis of investor sentiment, information cognition and short-sale constraints [J]. *Journal of Management Sciences in China*, 2016, 19(10) : 87 – 100. (in Chinese)
- [21] Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market [J]. *Journal of Computational Science*, 2011, 2(1) : 1 – 8.
- [22] Sabherwal S, Sarkar S K, Zhang Y. Do internet stock message boards influence trading? Evidence from heavily discussed stocks with no fundamental news [J]. *Journal of Business Finance & Accounting*, 2011, 38(9 – 10) : 1209 – 1237.
- [23] Li X, Wu P, Wang W. Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong [J]. *Information Processing & Management*, 2020, 57(5) : 102212.
- [24] Ge Y, Qiu J, Liu Z, et al. Beyond negative and positive: Exploring the effects of emotions in social media during the stock market crash [J]. *Information Processing & Management*, 2020, 57(4) : 102218.
- [25] 杨晓兰, 沈翰彬, 祝 宇. 本地偏好、投资者情绪与股票收益率: 来自网络论坛的经验证据 [J]. *金融研究*, 2016, (12) : 143 – 158.
Yang Xiaolan, Shen Hanbin, Zhu Yu. The effect of local bias in investor attention and investor sentiment on stock markets: Evidence from online forum [J]. *Journal of Financial Research*, 2016, (12) : 143 – 158. (in Chinese)
- [26] 金德环, 李 岩. 群体智慧: 同伴观点与价值发现——来自社交媒体的经验证据 [J]. *经济管理*, 2017, 39(12) : 157 – 173.
Jin Dehuan, Li Yan. Wisdom of crowds: Peer opinions and value discovery: Empirical evidences of social media [J]. *Economic Management Journal*, 2017, 39(12) : 157 – 173. (in Chinese)
- [27] 钱 宇, 李子饶, 李 强, 等. 在线社区支持倾向对股市收益和波动的影响 [J]. *管理科学学报*, 2020, 23(2) : 141 – 155.
Qian Yu, Li Zirao, Li Qiang, et al. Impact of online community support tendencies on returns and volatility in Chinese stock market [J]. *Journal of Management Sciences in China*, 2020, 23(2) : 141 – 155. (in Chinese)
- [28] 姚加权, 冯 緣, 王贊钧, 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. (in Chinese)
- [29] 陈荣达, 林 博, 何诚颖, 等. 互联网金融特征、投资者情绪与互联网理财产品回报 [J]. *经济研究*, 2019, 54(7) : 78 – 93.
Chen Rongda, Lin Bo, He Chengying, et al. Internet finance characteristics, internet finance investor sentiment and the return of internet financial products [J]. *Economic Research Journal*, 2019, 54(7) : 78 – 93. (in Chinese)
- [30] Zhao Y, Kou G, Peng Y, et al. Understanding influence power of opinion leaders in e-commerce networks: An opinion dynamics theory perspective [J]. *Information Sciences*, 2018, (426) : 131 – 147.
- [31] Van Den Bulte C, Joshi Y V. New product diffusion with influential and imitators [J]. *Marketing Science*, 2007, 26(3) : 400 – 421.
- [32] Godes D, Mayzlin D. Firm-created word-of-mouth communication: Evidence from a field test [J]. *Marketing Science*, 2009, 28(4) : 721 – 739.
- [33] Liu S, Jiang C, Lin Z, et al. Identifying effective influencers based on trust for electronic word-of-mouth marketing: A domain-aware approach [J]. *Information Sciences*, 2015, 306(10) : 34 – 52.
- [34] Apesteguia J, Oechssler J, Weidenholzer S. Copy trading [J]. *Management Science*, 2020, 66(12) : 5608 – 5622.
- [35] Ouimet P, Tate G. Learning from coworkers: Peer effects on individual investment decisions [J]. *The Journal of Finance*, 2020, 75(1) : 133 – 172.
- [36] 杨晓兰, 王伟超, 高 媚. 股市政策对股票市场的影响——基于投资者社会互动的视角 [J]. *管理科学学报*, 2020, 23(1) : 15 – 32.
Yang Xiaolan, Wang Weichao, Gao Mei. The impact of stock market policies on stock market: From the perspective of investor social interaction [J]. *Journal of Management Sciences in China*, 2020, 23(1) : 15 – 32. (in Chinese)

- [37] Bonacich P. Some unique properties of eigenvector centrality [J]. *Social Networks*, 2007, 29(4): 555–564.
- [38] Albert R, Jeong H, Barabási A. Error and attack tolerance of complex networks [J]. *Nature*, 2000, 406(6794): 378–382.
- [39] Anagnostopoulos A, Kumar R, Mahdian M. Influence and Correlation in Social Networks [C]. Las Vegas: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2008: 7–15.
- [40] Crandall D J, Dan C, Huttenlocher D P, et al. Feedback Effects between Similarity and Social Influence in Online Communities [C]. Las Vegas: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2008: 160–168.
- [41] Pozzi F A, Fersini E, Messina E, et al. Challenges of Sentiment Analysis in Social Networks: An Overview [M]. Boston: Morgan Kaufmann, 2017.
- [42] Gómez D, Figueira J R, Eusébio A. Modeling centrality measures in social network analysis using bi-criteria network flow optimization problems [J]. *European Journal of Operational Research*, 2013, 226(2): 354–365.
- [43] Brin S, Page L. The anatomy of a large-scale hypertextual Web search engine [J]. *Computer Networks and ISDN Systems*, 1998, 30(1): 107–117.
- [44] Shafiq Z, Ilyas M, Liu A, et al. Identifying leaders and followers in online social networks [J]. *IEEE Journal on Selected Areas in Communications*, 2013, 31(9): 618–628.
- [45] Kitsak M, Gallos L K, Havlin S, et al. Identification of influential spreaders in complex networks [J]. *Nature Physics*, 2010, 6(11): 888–893.
- [46] Fang X, Hu P J. Top persuader prediction for social networks [J]. *MIS Quarterly*, 2018, 42(1): 63–82.
- [47] Van Vlasselaer V, Eliassi-Rad T, Akoglu L, et al. GOTCHA! Network-Based fraud detection for social security fraud [J]. *Management Science*, 2016, 63(9): 3090–3110.
- [48] 马长峰, 陈志娟, 张顺明. 基于文本大数据分析的会计和金融研究综述 [J]. 管理科学学报, 2020, 23(9): 19–30.
Ma Changfeng, Chen Zhijuan, Zhang Shunming. A survey on accounting and finance research based on textual big data analysis [J]. *Journal of Management Sciences in China*, 2020, 23(9): 19–30. (in Chinese)

Opinion leader at the online stock forum: Having the discerning eye or eye-catching

ZHANG Ke¹, LI Xin-dan¹, FANG Xiao², LI Xiao-lin³

1. School of Management and Engineering, Nanjing University, Nanjing 210093, China;

2. Alfred Lerner College of Business & Economics, University of Delaware, DE 19716, USA;

3. School of Management, Nanjing University, Nanjing 210093, China

Abstract: Internet is a major information source for investors. The opinion leaders' views can spread widely on the Internet among investors and alter their behavior. This paper uses social network analysis technique to identify the opinion leader at the online stock forum and the text analysis technique to judge the type of the opinion leader's posted message. The mechanism of the return predictability of the opinion leaders' sentiment and the subsequent investors' reaction are systematically investigated. Based on a sample data from 2011 to 2017, it is found that opinion leaders' sentiment can predict individual stock return in the next month and up till the next 6 months. The reason of return predictability is not that the opinion leaders know the value-relevant information of listed firms. Further studies show that the investors' identification on opinion leaders is irrelevant with their prediction accuracy. Investors place more attention on opinion leaders with positive opinion and ignore their correctness. In addition, investors repel opinion leaders who make correct negative predictions. Our study reveals the impact of opinion leaders on the market and their formation mechanism, and also provides theoretical basis for investor education and protection.

Key words: opinion leader; investor sentiment; return predictability; investor education