第 14 卷第 3 期 杭州电子科技大学学报(社会科学版) Vol. 14 No. 3
2018 年 6 月 JOURNAL OF HANGZHOU DIANZI UNIVERSITY(Social Sciences) Jun. 2018

DOI: 10.13954/j.cnki.hduss.2018.03.003

网络舆情与上证指数涨跌幅的关联性分析
——基于 LDA 主题模型的文本挖掘

徐 翔,靳 菁
(同济大学 艺术与传媒学院,上海 201804)

摘要: 基于对网络信息进行 LDA 主题模型的文本挖掘和实证分析,考察网络舆情的变化与上证指数的涨跌幅之间的关联性。编写网络爬虫抓取“今日头条”网站中的一千三百余万条帖子,通过 LDA 主题模型的方法将网络新闻和帖子分为 100 类主题,考察各类主题在每天的网络舆情中所占的比重及其变化情况。分析结果显示,网络舆情中部分主题所占比重的变化情况,与上证综指涨跌幅之间具有关联性。这既反映着网络舆情作为“社会传感器”与金融市场中的社会后果之间的关联,也从网络信息挖掘的角度为有限理性和有限注意力背景下的行为经济学研究提供有益路向。

关键词: 网络舆情;LDA 主题模型;文本挖掘;社会传感器
中图分类号: F83; G206 **文献标志码:** A **文章编号:** 1001 - 9146(2018)03 - 0018 - 07

信息时代,网络已经成为受众获取信息的主要途径。根据中国互联网信息中心(CNNIC)发布的第 41 次《中国互联网络发展状况统计报告》,截至 2017 年 12 月,我国网民规模已达 7.72 亿,全年共计新增网民 4 074 万人,互联网普及率为 55.8%。网络新闻用户规模为 6.47 亿,年增长率为 5.4%,占网民总体的 83.8%,其中手机网络新闻用户规模达到 6.20 亿,占手机网民的 82.3%,年增长率为 8.5%。针对网络舆情进行的监测与探究已成为了解社情民意不可或缺的重要途径。同时,我国股票市场经历了数十年的发展,成为反映宏观经济发展情况的重要指标。但是,我国股票市场起步较晚,证券投资环境不够完善、投资者素质有待提高、证券市场需要进一步加强监管。在这种情况之下,网络舆情缺少信息把关,容易产生混淆视听的谣言,会使投资者形成“羊群效应”等社会效果,导致股票市场异常,甚至引起动摇金融市场稳定的风暴。

行为经济学认为,金融市场中的投资者并不是完全理性和信息对称的,其在投资决策过程中可能存在认知偏差。随着 Web3.0 时代的发展,网络已经成为人们信息获取和共享的重要渠道。股票论坛、社交媒体、新闻网站和客户端都成为股民投资选择的重要参考依据。传播学的“议程设置”理论认为,大众传播具有为受众“设置议程”的功能,媒体通过信息传达活动赋予各种议程不同程度的显著性,以此影响人们对于周围事物及其重要性的判断。随着“人人都有麦克风”时代的到来,信息的提供者也不再仅仅局限于上市公司和监管机构,普通股民也成为股市信息的提供者、传播者和执行者。从海量网络舆情表现出的“议题”,反过来也可以代表受众的重大关切和社会的普遍感知。

那么,网络舆情主题特征是否与股市交易量存在关联性?或者说,每天通过海量网络信息表现出来的社会感知和社会“议题”分布及其波动,它们与短周期内的股市交易量之间,是否存在着作用力与关联性?目前有关网络舆情变化和股票市场关联性的研究主要集中在经济学、金融学领域,而且大多是静态分析舆情主题和金融市场的关联。本研究从新闻传播角度进行文本和数据挖掘,特别是从网络舆情

收稿日期: 2018 - 05 - 31
基金项目: 上海市哲学社会科学规划项目(2014FXW001);同济大学中央高校基本科研业务费资助项目(22120180186)
作者简介: 徐翔(1983 -),男,江西上饶人,博士,教授,网络传播研究.

动力学的角度考察舆情主题变化对股价涨跌的影响,是非常具有理论和现实意义的。

一、文献回顾与理论分析

网络舆情与股票市场的理论关联主要在于:舆情作为“社会传感器”与社会后果之间的反映和关联;由于“有限理性”,投资者的观念、偏见等对于投资和决策行为产生的影响;由于注意力的有限性,注意力分配和投资者行为之间的关系。这些因素使得在对投资者行为的分析中,难以忽视其信息层面的网络信息“拟态环境”和网络舆情“传感器”的传递和作用。

(一)媒体信息与股票市场的相关性

媒体信息与股票市场的关联,国内外学者进行了较多研究。大部分学者对网络舆情和股价预测性之间的关联持积极肯定态度。

国外研究方面,Wysocki(1999)[1] 最早开始对网络论坛讨论进行研究,他搜集了雅虎网站上网民投资者对于超过 3 000 只股票的 94.6 万条的评论,发现发帖量大的公司往往对应的是交易量大的、市值价值比高的、机构持股比例更小的公司。这表明前一日的股票评论数量对于后一日的股票收益率和成交量有一定的预测解释能力。

Antweiler(2004)[2] 等利用贝叶斯和向量机方法对 Raging Bull 和雅虎金融板块的论坛进行分析,研究了 45 支股票约 150 万条网络信息,所有的帖子被分成 3 类:看多、看空、持平,建立了投资者情绪指数和投资者意见分散度指数。发现股票收益率与当天论坛中的信息指标呈显著的相关关系。同时当天的讨论意见的分歧也与交易量相关,但在滞后一天的检验中,这个效应则不一定存在。

同样证明了媒体信息对股票市场具有导向作用还有 Tetlock(2007)[3-4] 等以及 Schumaker 和 Chen(2009)[5] 等。

随着社交媒体的发展,Blankespoor、Miller 和 White(2014)[6] 等人从市场有效性的角度研究社交媒体与股票交易的关联,他们发现,公司在 twitter 上发布新闻,能够减少信息不对称性,以降低异常的买卖差价。

国内研究方面,胡洋(2011)等[7],金雪军、祝宇等(2013)[8] 以及马俊伟等(2014)[9] 分别通过分析不同数据源,采用不同方法证明了媒体信息与股价的相关性。

(二)投资者情绪和关注度对股票价格的预测能力

除了媒体信息对股价造成的影响外,投资者的情绪和关注度也可能成为影响股价的重要变量。针对这一点,不同学者进行了以下研究。

Bollen、Mao(2010)[10] 等学者从行为经济学的角度出发,使用情绪挖掘工具 Opinion Finder 和 Google Profile of Mood States,对美国大型社交网站 twitter 近 1 000 万条网络信息中表现出的投资者的不同情绪进行挖掘分析,以检验网络情绪状态的日变化与证券收益指数之间的关系。结果表明,积极和消极情绪的比例对滞后一天的道琼斯指数变化产生显著影响。他们发现,把“冷静”态度的指数向后移动 3 天左右得到的结果,同道琼斯工业的平均指数非常相近,准确率甚至达到了 86.7%。但作者也指出,当股市波动较大时其预测能力会下降。

同样关注投资者情绪对股价影响的还有 Ljungqvist(2003)和 Sapienza(2004)。Ljungqvist(2003)[11] 等认为,投资者情绪使资产价格在短期内偏离了其内在价值,从而产生了 IPO 抑价。Sapienza(2004)[12] 等人关注情绪对市场收益率的预测,主要研究情绪对市场交易量、波动性、收益率等指标的预测作用或者滞后影响。

(三)搜索强度的股价预测能力

搜索实时数据是网络舆论重要晴雨表。大部分学者研究表明搜索数据对股价涨幅有预测能力。

Da 和 Gao 通过网民在互联网中的搜索指数来说明股票市场中投资者情绪的变化。他们(2009)[13] 通过搜集 Russell 3000 指数中所有公司股票简称的搜索强度数据,得出的结论是,规模较小的市场的搜索强度对于股价的预测能力更强。他们(2011)[14] 还发现,高搜索频次之后 2 周的收益率较高,并解释

了 IPO 当日估价过高的现象。

运用搜索引擎为数据依托进行研究的学者包括 Joseph 等(2011)[15]、Dzielinski(2012)[16]以及国内的宋双杰等人(2011)[17]。Joseph 利用 Google 检索得到股票的搜索强度,发现搜索强度最高的组平均周收益比最低的组高 0.17%。Dzielinski 认为个体在面对不确定信息时会寻求信息检索行为,他们从实证的角度证明了基于搜索数目构建的投资者信息不确定性指标,与股市的相关性较高。宋双杰利用谷歌搜索数据分析了国内市场新股异常现象,并提出发行前互联网的关注度火爆导致证券市场新股认购需求激增,上市首日收益率较高,但是这种关注度造成的影响会在中后期反转,新股在之后的交易日反而出现估计下降的情况。

(四)媒体信息与股票市场的不相关性

但是,也有学者认为,媒体信息与股票市场可能呈现不相关的情况。

Tumarkin 和 Whitelaw(2001)[18]分析了美国 Raging Bull 论坛,以互联网行业的 72 只股票为研究对象,用 Ranging Bull 自带的打分功能作为网络文本信息的替代变量,着重观察事件日前后各五天内证券市场变量的变化情况。作者最后得出发帖者的情绪并不能预测股票成交量和回报,论坛上的开源信息主要是"市场噪音"这一结论。

Das 和 Chen(2001)[19]研究发现,网络讨论能够迅速反映信息,但无法预测股票收益。Das(2007)[20]还以科技行业的 24 只股票为对象,发现前一天的股票评论所反映的投资者情绪和股市总指数具有显著的正相关性,但是对单只股票而言这种相关性并不成立,不具备预测能力。

Fang 和 Peress(2009)[21]以股票或公司的新闻报道数量衡量媒体关注度,将 New York Times、USA Today、Wall Street Journal 和 Washington Post 中个股新闻数量作为其媒体覆盖指标,研究发现未被媒体报道的股票比被媒体高度报道的股票获得更高收益。

国内研究方面,饶育蕾[22]等使用新浪网搜索引擎,通过因子模型证实了网络信息量对我国股票的收益率有显著影响。但是,饶育蕾和王攀(2010)[23]以 200 多只新发行股票在百度上的相关新闻数量表示个股的媒体关注度,以异常收益率为被解释变量,以媒体关注度和其他多个控制变量为解释变量进行回归分析。结果表明,短期内媒体关注度会对新股产生正向影响,但从长期来看媒体关注对新股影响是负向的。

二、研究设计与研究假设

结合研究目的以及前文的理论回顾分析,本研究的核心假设为:在将网络舆情用 LDA 主题模型分为 100 种主题的基础上,分析每天的舆情在这 100 个主题上所占比重的变化幅度;则某个交易日的上证指数涨跌幅,与该交易日前 1 天至前 7 天之间的舆情变化幅度存在的相关联性。

本研究通过编写网络爬虫,抓取了《今日头条》网站上的大量帖子,删除重复后包含 13 525 315 条。同时,选取了上证指数 2016 年 2 月 1 日到 2017 年 11 月 29 日的交易日涨跌幅数据,分析每个交易日的涨跌幅和其前 1 至 7 天的今日头条网络舆情之间的关联。

网络舆情的分析样本选定为"今日头条"网站(www.toutiao.com)作为来源,是因为今日头条网站是我国具有重要地位的大规模信息资讯网站,其中的信息具有综合性,来源于多种网站、媒体和自媒体用户,既有新闻也有其他非新闻类的资讯、知识和帖子,主题涉及到金融、经济、政治、社会、体育等丰富多样的内容。这较之单一的网站或媒体来源,具有更高的反映舆情、社会注意力和社会关注结构的程度。

对于今日头条,所抓取的发布者用户来自于今日头条网站首页的"热点"板块(www.toutiao.com/ch/news_hot),根据版块近 2 个月时间段的帖子中,抓取到这些帖子的发布者用户。对于这些发布者,从该用户的今日头条个人主页采集其历史发帖。页面的帖子都是从当前往过去的时间顺序排列。由于今日头条的页面具有 ajax 结构,采用 python + selenium 的工具,模拟人在浏览器的行为,往下拉。一般一个用户可以下拉到约一万条历史发帖子。平均每天的帖子数为 18 983.4 条。帖子的获取具有随机

性,每天帖子样本数的不同,不影响到最后的分析。

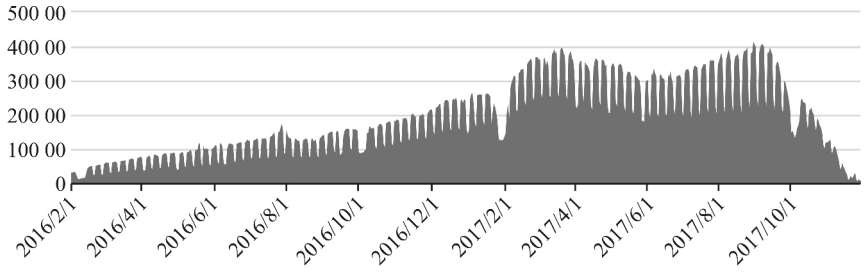

图 1《今日头条》每日帖子数量(2016 年 2 月 1 日 – 2017 年 11 月 29 日)

三、文本挖掘与分析过程

(一) 对于网络舆情内容的 LDA 主题模型分析

本文分析中所用的 LDA 主题模型是一种对于文本内容进行分析的方法。LDA 通过对离散数据集建立模型,分析概率主题。这种模型的核心思想是,一个文档包含了若干主题,而每一主题又包括若干个主题词,但是该模型不注重文档内部的语句和词语的出现顺序和上下文关联。LDA 主题模型主要通过以下过程建立的。首先,了解整个该文本的单词总数。其次,针对文本的每一个单词,抽样生成某一主题的概率分布。第三,针对该文本中的每一个单词,从该主题的分布中随机选择一个作为主题词,并且抽样生成主题词的概率分布。

表 1　LDA 主题模型示例(以截选的 20 个主题及其所包含的前 15 个主题词为例)

主题序号	关键词 1	关键词 2	关键词 3	关键词 4	关键词 5	关键词 6	关键词 7	关键词 8	关键词 9	关键词 10	关键词 11	关键词 12	关键词 13	关键词 14	关键词 15
主题 1	网友	结婚	夫妻	情侣	离婚	出轨	曝光	恩爱	爱情	老公	回应	女友	一对	幸福	王宝强
主题 2	男人	喜欢	女人	大家	怎么	看看	这样	一个	女朋友	你们	觉得	如果	女生	自己	什么
主题 3	银行	金融	投资	理财	平台	p2p	基金	如何	保险	风险	互联网	监管	网贷	资产	行业
主题 4	2016	中国	2017	全国	出炉	名单	十大	年度	公布	发布	最新	排行榜	最佳	全球	报告
主题 5	实拍	飞机	现场	航空	火车	场面	机场	视频	钓鱼	飞行	空中	震撼	挖掘机	国外	美国
主题 6	小伙	女友	男友	女孩	美女	结果	男子	自己	姑娘	女子	涂磊	相亲	二代	嘉宾	分手
主题 7	还是	真的	就是	其实	只是	但是	自己	有点	这个	一点	我们	虽然	以为	他们	觉得
主题 8	健康	健身	减肥	食物	身体	方法	每天	动作	养生	运动	可以	瑜伽	这些	效果	训练
主题 9	城市	中国	一个	世界	神秘	发现	历史	一座	地方	最大	这个	国家	这里	唯一	千年
主题 10	2016	增长	同比	亿元	上半年	净利润	年度	2017	营收	快讯	下降	今年	去年	收益	10
主题 11	中国	日本	美国	印度	国家	这个	韩国	越南	民族	实拍	俄罗斯	为何	生活	他们	直击
主题 12	工作	扶贫	推进	脱贫	建设	召开	精准	攻坚	调研	提升	开展	全面	群众	环保	做好
主题 13	手机	小米	华为	苹果	三星	曝光	发布	iphone	魅族	全面	oppo	新机	旗舰	骁龙	售价
主题 14	诈骗	网络	警方	快递	骗局	套路	提醒	朋友圈	骗子	男子	谣言	警惕	电信	网上	小心
主题 15	发现	回家	老板	结果	看到	男子	美女	老婆	女子	小伙	突然	尴尬	自己	晚上	上班
主题 16	汽车	上市	车展	全新	新能源	亮相	东风	发布	上海	suv	新车	2017	奥特曼	正式	大众
主题 17	高手	陈翔	六点半	功夫	孙悟空	厉害	大师	民间	西游记	武功	竟然	小伙子	传奇	太极	大战
主题 18	天气	准备	高温	预警	今天	来袭	广东	发布	台风	暴雨	今日	明天	气温	降温	影响
主题 19	出行	注意	高速	交通	交警	公交	期间	12	市民	部分	这些	日起	春运	高峰	铁路
主题 20	人生	自己	如何	马云	成功	成为	看看	孩子	职场	年轻人	社会	决定	改变	如果	男人

LDA 主题模型的优点主要包括以下三点。首先,LDA 模型简短描述文档,并且引入先验参数,减少了过度拟合的可能性。即使文档数量增加,主题参数也不会随之线性增加,而只会留下最本质的统计信息,这使得大规模处理文档信息,进行文本分类变得更加高效。其次,3 层贝叶斯结构的 LDA 模型具有清晰的层次划分,包括文档集合层、主题、主题特征词三层,这使得潜在语义分类更加科学化、智能化,减小了人工分类带来的主观偏见的影响,有助于提高聚类效果的质量。第三,本文中的 LDA 是动态化的主题模型,不再关注单个静态因素对因变量的影响,而是关注主题的动态变化与因变量的关联,这种动态演化过程很好地反映了主题的演化与文本特征变化。

本研究中在 python 编程语言的平台上进行,在对帖子标题通过常用的“jieba”分词模块进行分词的基础上,运用 sklearn 的 CountVectorizer 函数进行词频矩阵的统计,词频统计中设定的主要参数如下:min_df = 20,max_df = 0.1,也即最低出现的词频(min_df 参数)为 20 次,最高的文档出现比例(max_df 参数)为 0.1。因为有一些词语虽然出现很频繁,但是在过多的帖子中都会出现,因而实际意义不大,比如“的”“了”等词。LDA 主题模型的分析通过导入“lda”计算模块进行,其关键的命令为:model = lda. LDA(n_topics = 100, n_iter = 500, random_state = 1),也即设主题的分类个数(n_topics)为 100 个,迭代次数(n_iter 参数)为 500 次。分析之后的 20 类主题及每类中前 15 个主题词示例如下(LDA 分析的结果设定为 100 个主题,但由于篇幅所限,本处只展示其中的 20 个主题;每个主题存储了前 200 个主题词,本处只展示前 15 个)。本研究中的 LDA 主题模型分析中迭代了 500 次,迭代次数具有充分性。从表 1 的各行也可以看出,各个主题的解析效果和区分度较好。

每天的帖子在 100 个主题上的概率,其相对于前一天的变化差值如下所示(局部截选七天内的部分主题)。这些主题的分解及其变化情况,反映着社会事件、关注点、社会心理等方面的变化。需要强调的是,表 2 中每个单元格的数值,不是某天的某个主题所占比例的绝对情况,而只是该主题相对于前一天所占比值的变化值。由于各个主题具有不均衡性,有些主题天然比另一些主题所占的绝对比重大,因而这种比重的相对变化情况(或一阶差分)对于社会注意力和舆情关注度的结构可以很好地表现和反映。

表 2 主题分布概率及相对于前一天的变化差值(以 7 天内的部分主题为例)

发布日期	LDA 第 1 列	LDA 第 2 列	LDA 第 3 列	LDA 第 4 列	LDA 第 5 列	LDA 第 6 列	LDA 第 7 列	LDA 第 8 列	LDA 第 9 列	LDA 第 10 列
2016/8/14	0.000 36	-0.000 64	-0.000 43	-0.000 60	-0.000 24	0.000 79	-0.000 16	0.001 24	-0.000 17	0.000 23
2016/8/15	-0.000 04	0.000 14	0.000 24	0.000 03	0.000 17	-0.000 74	-0.000 16	-0.000 98	0.000 34	0.000 27
2016/8/16	0.000 00	0.000 18	-0.000 19	-0.000 02	-0.000 01	0.000 65	0.000 18	-0.000 23	-0.000 39	-0.000 23
2016/8/17	-0.000 28	-0.000 04	0.000 56	0.000 33	-0.000 19	-0.000 25	-0.000 03	0.000 42	-0.000 03	0.000 17
2016/8/18	0.000 04	-0.000 40	-0.000 20	-0.000 26	-0.000 06	0.000 24	-0.000 04	-0.000 27	0.000 47	0.000 27
2016/8/19	0.000 59	0.000 29	0.000 01	-0.000 27	0.000 08	-0.000 88	-0.000 41	0.000 11	-0.000 73	-0.000 49
2016/8/20	-0.000 51	0.000 04	0.000 28	0.000 61	0.000 50	0.000 67	0.000 44	-0.000 23	0.000 77	0.000 22

(二)上证指数涨跌幅与前 1 - 7 日的舆情主题变化幅度的多元线性回归分析

多元线性回归分析的因变量是交易日的上证指数涨跌幅,自变量是该交易日的前 1 天至前 7 天内每天的网络舆情在 100 个主题上相对于前一天的增减幅度。回归分析采用“逐步”策略,显著性 0.05 进入,0.1 退出,从而看整个模型的 F 统计量变化,从变化来看模型有效。分析的结果显示,前 1 到前 7 日的网络舆情主题的变化对于上证指数存在作用,调整 R 方为 0.637(见表 3);F 值为 13.583,回归方程的 Sig 值为 0.000,具有显著性(见表 4)。

表 3 多元线性回归分析的模型汇总

模型汇总[by]					
模型	R	R 方	调整 R 方	标准 估计的误差	Durbin-Watson
76	.829[bx]	.687	.637	.524 427 1	1.774

表 4 多元线性回归分析的方差分析

Anova[a]						
模型		平方和	df	均方	F	Sig.
76	回归	231.606	62	3.736	13.583	.000[by]
	残差	105.334	383	.275		
	总计	336.940	445			

a. 因变量:涨跌幅

从关联强度来看,调整 R 方为 0.637 具有一定程度的理想性,显现出自变量对因变量可以解释其中 63.7% 的变化。Durbin-Watson 值为 1.774(见表 3),接近于 2,残差项间不存在明显的相关关系。从

特征值和条件索引的结果来看,自变量之间不存在明显的共线性:其中特征值较为均衡地分布在3.341到0.24之间,并未出现接近于0的值;条件索引在1到3.734之间,远小于10。此外,回归分析的残差的均值为0(见表5),残差分布中(见图2)沿对角直线方向较为吻合地分布且只有微小、平稳的误差,说明回归模型满足正态要求。总体而言,特征值检验比较平稳,回归分析的"逐步"策略也保证了模型的拟合效果,结论基本稳健。

表5　多元线性回归分析的残差统计量

残差统计量[a]					
	极小值	极大值	均值	标准偏差	N
预测值	-5.029 458	3.429 564	.048 250	.721 431 4	446
残差	-1.851 254 8	2.187 445 2	.000 000 0	.486 524 3	446
标准预测值	-7.038	4.687	.000	1.000	446
标准残差	-3.530	4.171	.000	.928	446

a. 因变量:涨跌幅

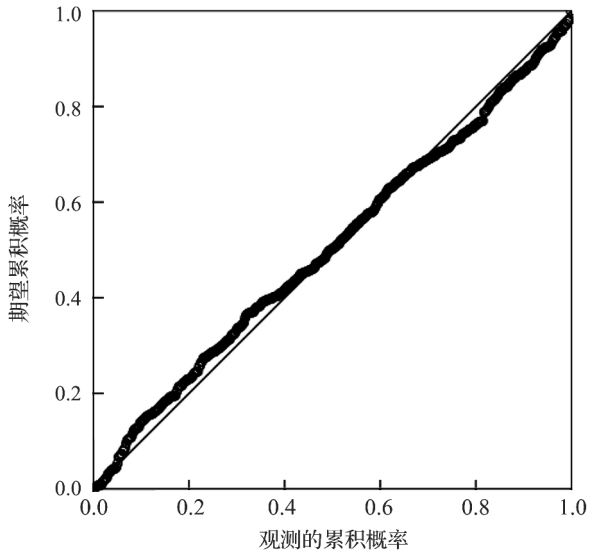

图2 回归标准化残差的标准P-P图

四、结语

本研究采用LDA主题模型的方法,对网络舆情的内容进行细粒度分解,并将之用于分析与股票市场的关联。对于金融研究而言,虽然从网络舆情的角度进行分析是重要研究路径之一,但从LDA主题模型的分解所进行的分析仍有待加强。这种分析方法排除了对于内容的分析中所受到的主观和人工分类的影响。通过LDA主题模型不是关注于情绪或信息量等因素,而是关注于内容和主题的变化及其与股市变化的关联性。同时,本研究不是舆情主题和金融市场的关联,而是研究舆情主题的变化和金融市场的变化之间的关联。

从网络舆情动力学的角度审视,舆情主题所占比重的变化反映了社会在关注的内容上所发生的结构性变化,而这种变化比之舆情绝对比重的本身更加反映着受众、用户的社会心理和社会注意力结构。分析的结果显现出网络舆情的变化与股票市场的涨跌变化之间存在着关联性。同时,这种关联也是非普适性的,是不断演变发展的系统,只在一定时期内可能有效。因为驱动股市涨跌的社会舆情因素,在不同的阶段可能会来自不同的因素。但是由于注意力的有限性以及信息面的作用,这种某部分舆情主题的作用,是可资利用的分析股市变化的有效因素。

本研究的不足和需继续改进之处主要包括以下:在网络舆情样本的覆盖上,还可进一步加大分析的样本数量,以提高舆情计算的精度;虽然今日头条网站具有多方面内容的覆盖性和综合性,但可以将今日头条与其他具有代表性的网站一并纳入分析框架,进一步完善媒体内容的覆盖;LDA的分析目前只

考察了分为 100 类的情况下舆情主题与股市的关联性,但并未详细比较不同的分类数量对于分析结果的影响,这些有待进一步详细计算与比较,确定最优参数。

参考文献

[1] Wysocki Peter D. Cheap Talk on the Web: The Determinants of Postings on Stock Message Boards [EB/OL]. [1999 - 04 - 20]. https://ssrn.com/abstract = 160170.

[2] Werner Antweiler, Murray Z Frank. Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards [J]. Journal of Finance, 2004, 59(3): 1259 - 1294.

[3] Tetlock Paul C. Giving Content to Investor Sentiment: The Role of Media in the Stock Market [J]. Journal of Finance, 2007, 37(4): 56 - 57.

[4] Tetlock P C, Tsechansky M S. More than Words: Quantifying Language to Measure Firms' Fundamentals [J]. The Journal of Finance, 2008, 63(3): 1437 - 1467.

[5] Schumaker R P, Chen H. A Quantitative Stock Prediction System Based on Financial News [J]. Information Processing & Management, 2009, 45(5): 571 - 583.

[6] Blankespoor E, Miller G S, White H, et al. The Role of Dissemination in Market Liquidity: Evidence from Firms' Use of Twitter [J]. The Accounting Review, 2014, 89(1): 79 - 112.

[7] 李玉梅,闫相斌,胡洋. 在线股评对股票市场的影响分析 [J]. 中国管理科学,2011(专辑):386 - 390.

[8] 金雪军,祝宇,杨晓兰. 网络媒体对股票市场的影响——以东方财富网股吧为例的实证研究 [J]. 新闻与传播研究,2013 (12) :36 - 51.

[9] 马俊伟,王铁军,李庆,等. 基于网络信息挖掘的股市影响因素分析 [J]. 吉林大学学报(信息科学版),2014,32 (2) : 195 - 200.

[10] Bollen J, Mao H, Zeng X, et al. Twitter Mood Predicts the Stock Market [J]. Journal of Computational Science, 2011, 2(1): 1 - 8.

[11] Ljungqvist A P, Wilhelm W J. IPO Pricing in the Dot-com Bubble [J]. Journal of Finance, 2003(58): 723 - 752.

[12] Polk C, Sapienza P. The Real Effects of Investor Sentiment [EB/OL]. [2002 - 11 - 30]. https://ssrn.com/abstract = 585885.

[13] Da Z, Engelberg J, Gao P, et al. In Search of Attention [EB/OL]. [2009 - 06 - 04]. https://ssrn.com/abstract = 1364209.

[14] Da Z, Engelberg J, Gao P, et al. In Search of Attention [J]. Journal of Finance, 2011, 66(5): 1461 - 1499.

[15] Kissan Joseph, M Babajide Wintoki. Forecasting Abnormal Stock Returns and Trading Volume Using Investor Sentiment: Evidence from Online Search [J]. International Journal of Forecasting, 2011(11) :123 - 158.

[16] Dzielinski M. Measuring Economic Uncertainty and its Impact on the Stock Market [J]. Finance Research Letters, 2012, 9(3): 167 - 175.

[17] 宋双杰,曹晖,杨坤. 投资者关注与 IPO 异象——来自网络搜索量的经验证据 [J]. 经济研究,2011(S1): 145 - 155.

[18] Tumarkin Robert, Robert F Whitelaw. News or Noise? Internet Message Board Activity and Stock Prices [J]. Financial Analysts Journal, 2001 (57): 41 - 51.

[19] Das Sanjiv R, Mike Chen. Yahoo! for Amazon: Sentiment Parsing from Small Talk on the Web [EB/OL]. [2001 - 08 - 05]. https://ssrn.com/abstract = 276189.

[20] Das S R, Chen M Y. Yahoo! for Amazon: Sentimental Extraction from Small Talk on the Web [J]. Management Science, 2007, 59(9): 1375 - 1388.

[21] Fang L, J Peress. Media coverage and the cross-section of stock returns [J]. Journal of Finance, 2009, 64 (5): 2023 - 2052.

[22] 饶育蕾,彭叠峰,成大超. 媒体注意力会引起股票的异常收益吗? ——来自中国股票市场的经验证据 [J]. 系统工程理论与实践,2010,30(2) : 287 - 297.

[23] 饶育蕾,王攀. 媒体关注度对新股表现的影响——来自中国股票市场的证据 [J]. 财务与金融,2010 (3): 1 - 7.

(下转第 31 页)

[29] Belaid S, Behi A T. The role of attachment in building consumer-brand relationships: an empirical investigation in the utilitarian consumption context [J]. Journal of Product & Brand Management, 2011, 20(1):38 - 39.
[30] Schmalz S, Orth U R. Brand attachment and consumer emotional response to unethical firm behavior [J]. Psychology & Marketing, 2012, 29(11):869 - 884.
[31] 薛海波,王新新. 品牌社群影响品牌忠诚的作用机理研究——基于超然消费体验的分析视角 [J]. 中国工业经济, 2009(10):96 - 107.
[32] 周健明,邓诗鉴. 品牌依恋对消费惯性与品牌忠诚的影响研究 [J]. 管理现代化,2015,35(6):73 - 75.
[33] 杨爽,郭昭宇. 品牌依恋对品牌对抗忠诚的影响研究——基于心理距离的调节作用 [J]. 消费经济,2017,33(3): 69 - 76.

A Research Review and Prospect of Brand Attachment

CAI Dan-hong, HU Jian
(*School of Management, Hangzhou Dianzi University, Hangzhou Zhejiang* 310018, *China*)

Abstract: The Attachment theory, originated from psychology and gradually entering into the marketing field, becomes a new perspective to explore the relationship between consumers and brands. However, the current domestic research on brand attachment is still at a preliminary stage, and the related research results need to be further explored and verified. Therefore, based on the relevant literature at home and abroad, this article sorts out the concept of the brand attachment, the dimension and the identification of similar concepts, as well as the related researches such as the measurement scale of brand attachment, the antecedent variables and the outcome variables. It also points out some problems in the current research and prospects future research directions in order to provide reference and enlightenment for the follow-up researches.

Key words: brand attachment; brand relationship; measurement scale; antecedent variables; outcome variables

(上接第 24 页)

A Correlation Analysis between Online Public Opinion and Shanghai Securities Composite Price Limit: A Text Mining Based on LDA Theme Model

XU Xiang, JIN Jing
(*School of Arts and Media, Tongji University, Shanghai* 201804, *China*)

Abstract: Based on the text mining and the empirical analysis of the LDA theme model of online information, the correlation between the change of the online public opinion and the price limit of Shanghai securities index is investigated. The web crawlers is written and captures more than one thousand and three million posts in "Today's Headlines" website. With the LDA theme model, the online news and posts are divided into 100 categories, and the proportion of various topics in daily online public opinion and its daily changes are examined. The result of the analysis shows that the change of the proportion of some themes in online public opinion is related with the price limit of Shanghai Securities Composite Index. This not only reflects the correlation between the online public opinion as a social sensor and the social consequences at the financial market, but also provides a useful way for the research of behavioral economics under the background of the bounded rationality and the limited attention from the aspect of the network information mining.

Key words: online public opinion; LDA theme model; text mining; social sensor