Finance 金融, 2020, 10(3), 158-170
Published Online May 2020 in Hans. http://www.hanspub.org/journal/fin
https://doi.org/10.12677/fin.2020.103016

Hans汉斯

Analysis of Investor's Investment Behavior Based on Python

An Hu, Shuang Li, Teng Wu, Yaxin Zeng, Huqin Yan

Xiamen National Accounting Institute, Xiamen Fujian
Email:15216739816@163.com, 2931754868@qq.com, zengyaxin2019@163.com, jxufewuteng@126.com, yanhuqin@xnai.edu.cn

Received: Apr. 6th, 2020; accepted: Apr. 30th, 2020; published: May 7th, 2020

Abstract

Investors' comments on investment on the Internet can often reflect their current investment behavior, investment preferences and market conditions. Therefore, this paper analyzes the investment behavior of investors in the stock market and futures market based on the powerful data mining function of Python. It analyzes the market situation, finds the hot topics that investors are paying attention to recently, and then analyzes the market performance of related industries to tap relevant comments and discuss relevance. Finally, it is concluded that the non-quantifiable text information is mined and processed by Python, and the obtained data information can help investors understand the market situation more intuitively before making a decision.

Keywords

Stock Market, Futures Market, Pneumonia, Gold, Crude Oil

基于Python分析投资者的投资行为

胡　安，李　爽，吴　腾，曾雅欣，阎虎勤

厦门国家会计学院，福建　厦门
Email:15216739816@163.com, 2931754868@qq.com, zengyaxin2019@163.com, jxufewuteng@126.com, yanhuqin@xnai.edu.cn

收稿日期：2020年4月6日；录用日期：2020年4月30日；发布日期：2020年5月7日

摘　要

投资者在互联网上留下的有关投资的言论，往往可以反映其目前的投资行为、投资偏好和市场情况等信息。因此，本文基于Python强大的数据挖掘功能，分析股票市场和期货市场下投资者的投资行为。分析

文章引用: 胡安, 李爽, 吴腾, 曾雅欣, 阎虎勤. 基于 Python 分析投资者的投资行为[J]. 金融, 2020, 10(3): 158-170.
DOI: 10.12677/fin.2020.103016

市场行情，发现近期投资者热点关注话题，再分析相关行业的市场表现，挖掘相关评论、讨论相关性。最后，得出通过Python对不可量化的文字信息进行挖掘处理，得到的数据信息，可以帮助投资者更直观了解市场行情再进行决策这一结论。

关键词

股票市场，期货市场，肺炎，黄金，原油

1. 绪论

1.1. 研究背景与意义

词云(Wordcloud)是 Python 语言所提供的一种数据搜索功能，它能够帮助我们对文本数据进行词汇挖掘，并按照词汇出现的频率进行统计。词云的文本搜索结果，常常可以按照可视化的形式表现出来，那些出现频次较高的词，将成为词云图反映的核心。利用 Python 的词云功能进行投资者行为分析，通过聚焦表现投资者行为的高频词汇，从而能够帮助我们分析投资者投资行为的某些特点。

例如，投资者在某论坛中发表言论提及某概念，则表示他近期特别关注与该概念有关的投资机会。又如，投资者经常阅读的投资信息中那些高频词汇，有可能在投资者投资时，影响到他的决策。如果说一个投资者对股票市场所发的评论反映了他对当前股市的投资态度，那么，与股票市场相关的所有评论则反映了大部分投资者对当前股市的投资态度。

本文将以 Python 词云功能的应用为例，以包括股票、基金、债券、外汇、期货等有关的金融市场作为研究对象，以有关网络论坛上所发表的与金融投资有关的评论文本作为搜索重点，通过对相关文本的高频词汇信息进行分析，来进一步分析投资者的投资行为，试图发现中国 A 股市场投资者的行为是否与外在影响有关以及这种关系是否会反应在其网络留言上等联系。

1.2. 国内外研究现状

随着互联网的迅猛发展，如今几乎已经是各行各业的基础设施之一。而互联网由于其特性，每天都产生了海量的信息数据，其中产生的 90%信息由非结构化数据 [1] 构成的，文本数据是非结构化数据的主要来源。而通过文本挖掘技术将难以量化的大规模文本数据整合转换为结构化数据，并抽取有价值的情报信息已广泛应用于商业、医疗和金融等多个领域。

Jun 等[1]研究人员致力于从 Google Insights 收集的搜索流量信息来进行消费者行为态度研究。目的是评估使用搜索流量信息来分析消费者对产品的真实态度的有效性，并试图预测消费者对产品的偏好。而 Preis 等[2]利用海量财经搜索文本内容和搜索频率建立股票市场波动性预警信号的判别模式，说明了结合大量行为数据进行分析可以更好地理解人类的一些集体性行为。

张博凯[3]认为将数据挖掘运用于股票预测项目中具有非常重要的现实意义；李源和李杰[4]通过行为金融学对个人投资者的心理、投资行为进行了分析，认为在中国 A 股市场下个人投资者的行为数据有一

[1] 非结构化数据是数据结构不规则或不完整，没有预定义的数据模型，不方便用数据库二维逻辑表来表现的数据。包括所有格式的办公文档、文本、图片、XML，HTML、各类报表、图像和音频/视频信息等等。——百度百科

定共性。

综上所述，在信息膨胀、大数据分析兴起的时代，文本挖掘作为分析互联网使用者的舆情关注、情感倾向等方面的工具和方法已经得到了广泛重视和应用。本文试图利用 Python 对互联网论坛中国 A 股投资者评论发言进行文本挖掘与分析。

1.3. 研究思路与方法

范珈瑜[5]通过 Rwordseg (基于 Java 的程序包)，对从旅游评论挖掘到的大量游客留言回复等共 6647 条评论进行分词处理，并使用了 Python 的 SnowNLP 组件对文本进行了情感分析，判断游客选择旅游景点的主要倾向。廖勇毅、丁怡心[6]使用 Python 实现股票定向爬虫，展示爬虫开发的基本思路以及 Python 的简洁高效。赵光亮等[7]探讨了 BBS 类论坛网页的文本数据的爬取与分析，他们对该类论坛网页的 HTML 结构进行了研究分析。

通过借鉴上述文献的研究思路，本文结合正则表达式、Xpath 函数和其他 Python 函数库构建算法，抓取了东方财富网股吧里面的前 2000 页评论[2]，以 Python 语言平台中专门处理汉字字库的库函数 Jieba 为工具，对这些网页进行词云搜索和词频排序，通过汇总分析，筛选掉无意义的词汇，最终得到频率最高的 50 个词，据此用 Wordcloud 生成了词云图。

因为进行词云制作时，英语语言由于其所有的句子均是一个一个单词构成的，且每个单词之间都存在空格，因此在进行分词统计词频时候较为简单，只需考虑抽取每两个空格直接的单词就能轻松完成分词；但是中文就不能如此操作，盖因为中文的语句构成都是由单音节词加上多音节词组成，还存在多音字、歧义、标点符号等诸多问题，因此对中文分词并不能简单效仿英文分词。而 Jieba 是专门用于中文分词的一个优秀 Python 语言第三方库函数。Jieba 分词原理主要是利用一个中文词库，确定汉字之间的关联概率；汉字间概率大的组成词组，形成分词结果；除了分词，用户还可以添加自定义的词组。本文采用 Jieba 的全模式(cut_all=True)分词，例如“股市跌企业就能复工吗”，分词效果为“股市”“企业”“就能”“能复”“复工”“股市跌”“吗”，即相比 Jieba 的默认模式分词会多出更多词汇，以防有分析价值的词汇遗漏。

2. 中国股票市场

从整体市场中的词云中，本文发现“疫情”、“新冠”、“肺炎”是出现频率最高的三个词，说明了疫情是当前投资者最为关注的因素，参见图 1。这是在不区分行业下，对整体市场的投资者情绪进行的刻画，由此可以发现疫情是近期股票市场最为重要的影响因子。

Figure 1. Oriental fortune net stocks first 2000 pages
图 1. 东方财富网股吧全部帖子前 2000 页图

[2]http://guba.eastmoney.com/，截止时间 2020 年 3 月 15 日 15:00。

股市之外，如图 2 所示，根据微博指数图可以看出，“疫情”、“湖北”、“全球”、“在家办公”等关键词在疫情席卷后，讨论度出现爬坡式提高，这说明了人们对疫情的关注度十分高。总体来看，此次疫情讨论度很高，影响很大。

受疫情影响，股市并未出现恐惧性大跌。相反，在年后 2 月 3 日开盘之后，大盘形势大好，一路高歌猛进；到 3 月 5 日，上证指数高达 3071.68。个股表现良好，尤其是医药股、无接触行业等表现亮眼，一路收红。因而，本文选择了受疫情影响而大涨的几个典型行业进行研究，分别是口罩股、医药行业及在线教育与云办公行业。

2.2. 疫情拉动口罩股跑赢大盘

口罩股是疫情在当前市场中投资情绪的集中体现，本文选取了泰达股份(N95 口罩)、道恩股份(口罩上游熔喷布)、容尚医疗(防护服)三支股票，将其与深证指数进行趋势对比分析，结果如口罩股与深证指数趋势对比图 3 所示。

结合口罩股与深证指数趋势对比图与微博指数图进行分析，从 2020 年 1 月 19 日始，肺炎引爆微博热搜，三只口罩股相对大盘均出现上涨。期间无论 A 股沉浮，或是疫情在全球蔓延，口罩供需缺口扩大。在 2020 年 1 月 1 日至 3 月 15 日，三只口罩股均跑赢了大盘，股价不断上涨。这说明在新冠肺炎疫情下，口罩等最基础的防护物资吸引了投资者大量的注意力，投资者对口罩股的讨论也越来越多。正是这样的投资情绪，使得资金大量涌入口罩股，拉动口罩股的股价一路飞涨并且跑赢大盘，A 股暴跌也难以阻其涨势。

2.3. 医药行业成为投资者关注的焦点

除了口罩这类最基础的防护物资之外，相关医药也成了大家关注的焦点。比如前段时间(2020 年 01 月 31 日深夜)，网传双黄连口服液可以预防新冠肺炎[3]，网上、线下药店的双黄连口服液几乎一瞬间售空，甚至连兽用的双黄连口服液也有人抢购(当然最后辟谣双黄连口服液对新冠肺炎没有作用)，可见国民对新冠肺炎的恐惧以及在特殊时期对能治疗或缓解病症的药物相当程度的信任和依赖。

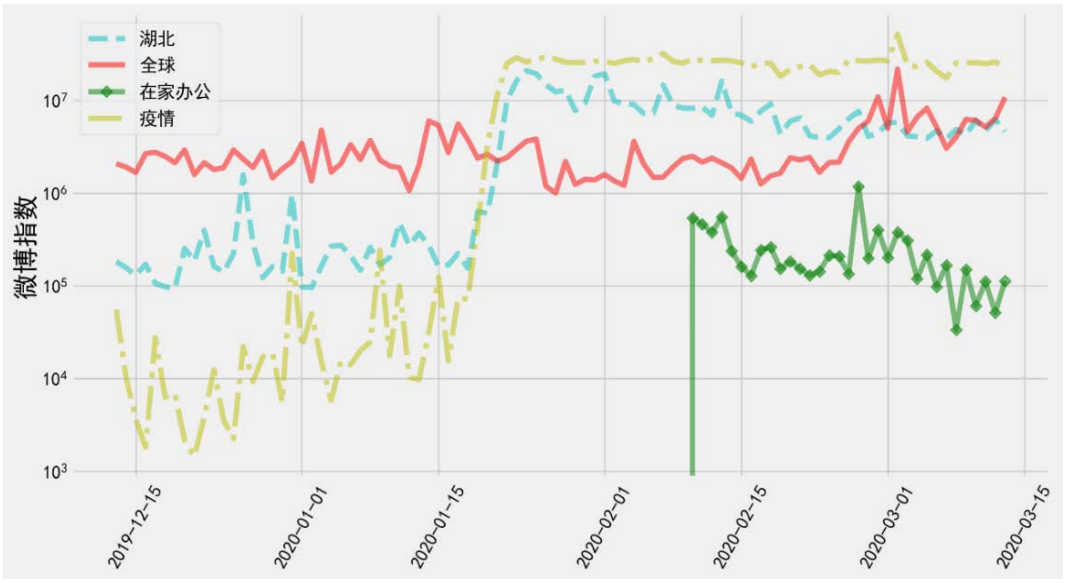

Figure 2. Weibo index chart
图 2. 微博指数图

[3]http://scitech.people.com.cn/n1/2020/0131/c1007-31566098.html。

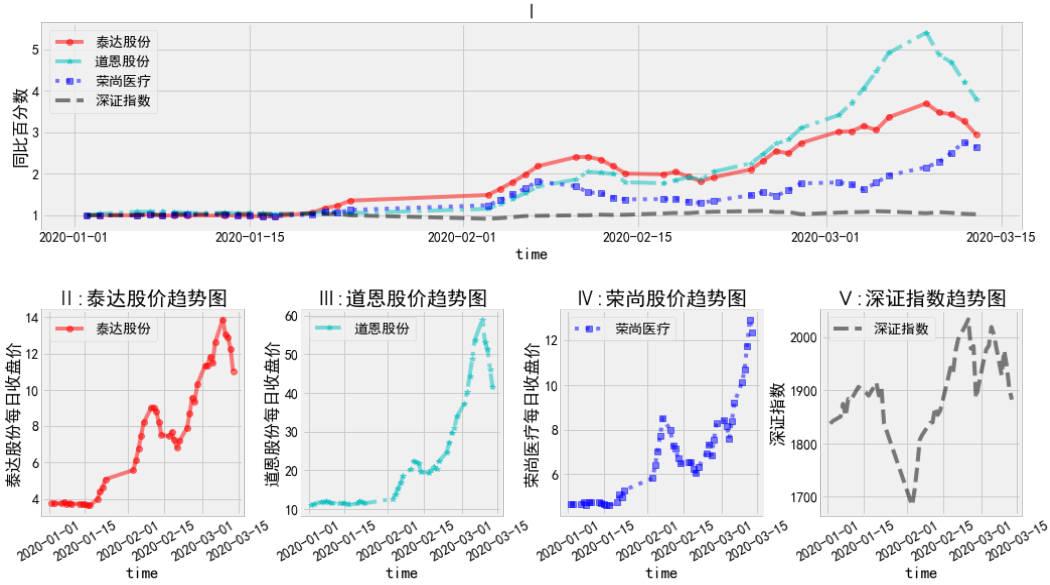

Figure 3. Contrast chart of mask stocks and shenzhen stock index
图 3. 口罩股与深证指数趋势对比图

由图 4 可知，医药行业指数(2019 年 12 月 1 日至 2020 年 3 月 13 日)自 2019 年 12 月份以来，总体是上升的。最高点在 2020 年 2 月 7 日，接近复工的时候，比 2019 年 12 月 1 日高出 5000 多点，涨势惊人。

Figure 4. Pharmaceutical manufacturing index chart
图 4. 医药制造指数图

为了进一步分析投资者关注医药行业的偏好，本文摘取以岭药业(002603)的评论数据，筛选掉无意义的高频词后，制成了以岭药业的词云，参见图 5。从图中可以很直观的发现，连花[4]清瘟胶囊是以岭药业投资者关注的焦点，被讨论的次数最多。

据以岭药业的统计数据显示，由该企业研发的连花清瘟胶囊(颗粒)先后被国家和 20 个省市列为诊疗方案临床推荐用药[5]。截至 2 月底，连花清瘟胶囊(颗粒)已在包括武汉方舱医院在内的湖北省 1600 余家医院(社区)应用。中药连花清瘟产品列入《新型冠状病毒感染的肺炎诊疗方案》，增强了人们抗击新型冠状肺炎疫情的信心，也给市场带来了利好消息。

[4]莲花 = 连花，部分评论出现的别字。
[5]以岭药业：驰援战“疫”一线 24 小时生产“不打烊”，http://he.people.com.cn/n2/2020/0313/c192235-33875160.html。

Figure 5. The first 1000 pages of Yiling pharmaceutical
图 5. 以岭药业股吧前 1000 页图

由图 6 所示，进入 2020 年以来，以岭药业股价持续走高，市值也从 2020 年初的 145.92 亿元最高上涨至 239.18 亿元。

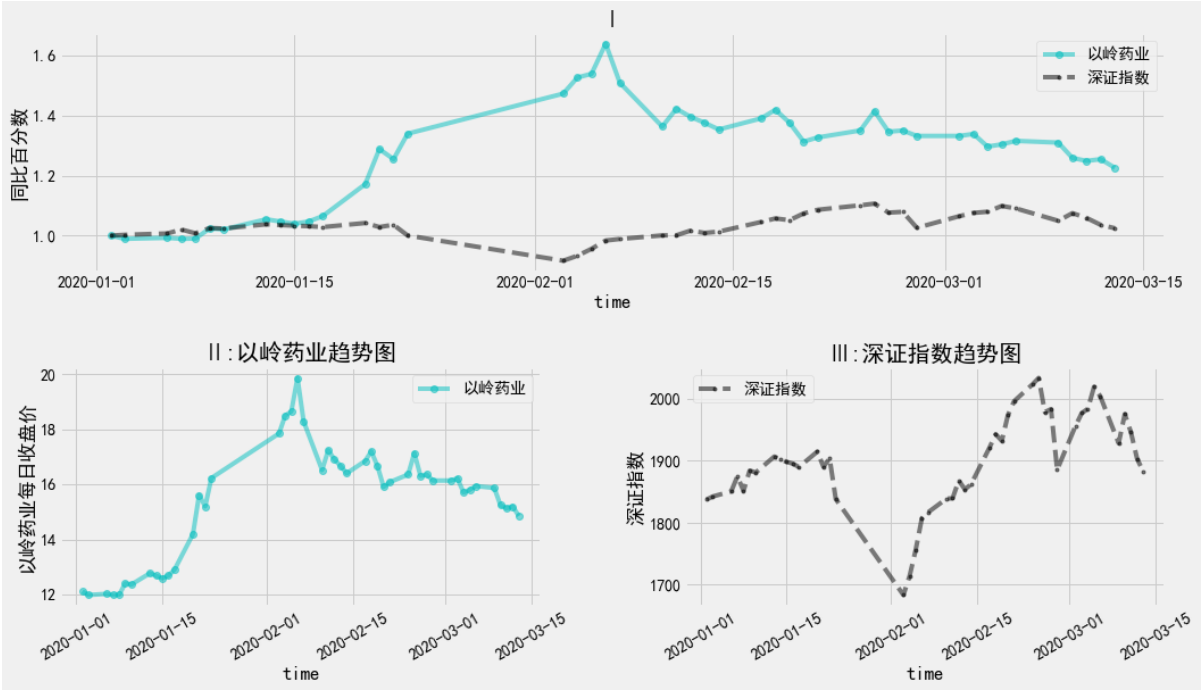

Figure 6. Yiling pharmaceutical stock chart
图 6. 以岭药业股票走势图

2.4. 在线教育与云办公异军突起

疫情不仅威胁着人们的身体健康，同时也阻碍了社会的正常运转。由于疫情愈发严重，国家不得不下命令推迟复工和开学时间。但是，职工不上班便没有工资；企业也因缺少员工无法正常运营，同时还需支出大量的固定成本；学生不上学则影响到其教育，于是，云办公和在线教育这些目前解决这些问题的最有效方法，理所当然地成为了投资者的宠儿。

由图 7 和图 8 所示，在线教育指数(2019 年 12 月 1 日至 2020 年 3 月 13 日)和云办公指数(2020 年 2 月 3 日至 2020 年 3 月 13 日)在疫情期间总体是上升的，而且云办公指数是在疫情发生之后才形成。在线教育指数极差大约为 500，云办公指数极差约为 800，医药指数极差约为 5000，前两者指数上涨绝对值比医药上涨的低很多，但是上涨幅度均超过医药，说明在线教育和云办公在疫情下热度很高，关注度甚至要超过医药。

Figure 7. Online education index chart
图 7. 在线教育指数走势图

Figure 8. Telecommuting index chart
图 8. 云办公指数走势图

以在线教育和云办公的概念股二六三为例，在 2020 年 1 月 1 日至 2020 年 2 月 1 日期间，二六三的股价是下降的，因为是春节期间，在线教育和云办公的重要性还没有体现出了；在 2020 年 2 月 1 日往后，因为即将复工，投资者开始意识到在线教育和云办公会成为发展趋势，二六三股价大幅上升，并成功赶超深证指数，参见图 9。说明，二六三受疫情的推动影响很大。

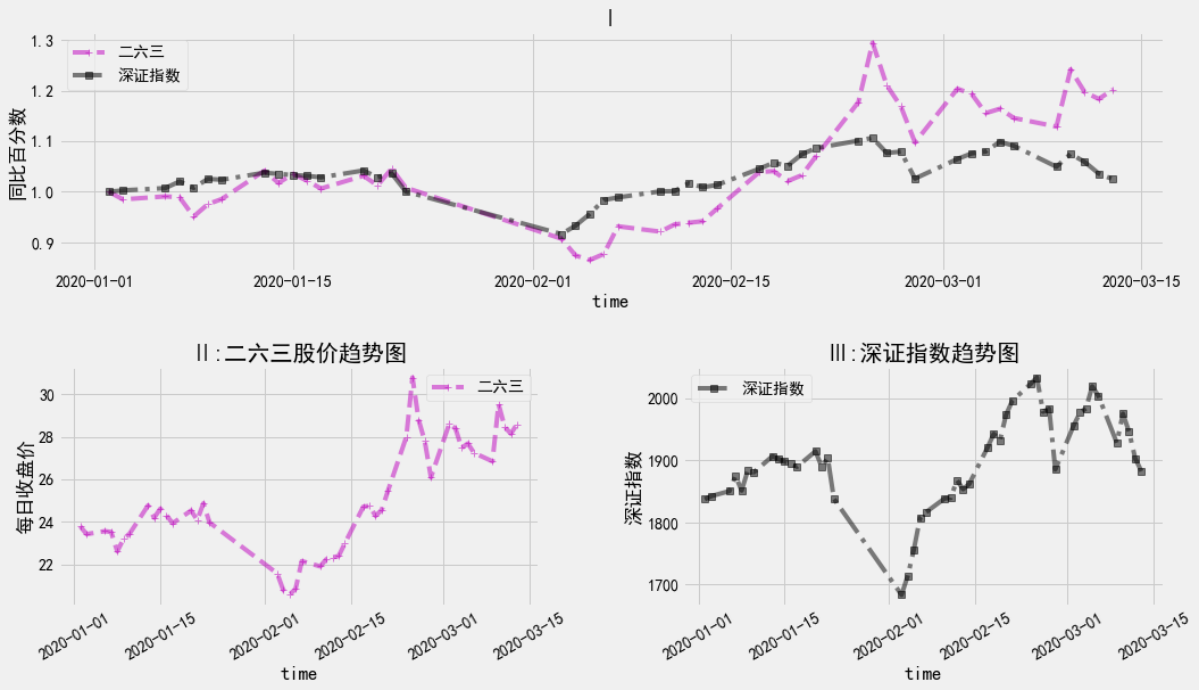

Figure 9. 263 stock chart
图 9. 二六三股票走势图

3. 期货市场

3.1. 原油和黄金存在相关性

3.1.1. 获取数据和预处理

从东方财富网股吧采集期货吧、黄金吧和原油吧评论词条组成数据。对采集到的评论分别进行处理，这里使用的是 Jieba 进行分词。分词结果见下表 1：

Table 1. Crude oil, gold, futures frequency analysis
表 1. 原油吧、黄金吧、期货吧词频分析

原油吧	频数	%	黄金吧	频数	%	期货吧	频数	%
原油	24,286	27.50%	黄金	46,201	37.09%	黄金	52,602	29.91%
黄金	18,589	21.05%	原油	14,102	11.32%	原油	19,825	11.27%
分析	4867	5.51%	分析	6225	5.00%	分析	11,344	6.45%
策略	4352	4.93%	策略	5319	4.27%	期货	9413	5.35%
行情	3603	4.08%	行情	4804	3.86%	走势	8715	4.95%
在线	2357	2.67%	继续	3136	2.52%	行情	6600	3.75%
直播	2017	2.28%	走势	3052	2.45%	震荡	6370	3.62%

Continued

走势	1926	2.18%	投资	3034	2.44%	策略	5564	3.16%
现价	1867	2.11%	震荡	2828	2.27%	市场	4188	2.38%
日内	1728	1.96%	非农	2751	2.21%	多头	4046	2.30%
继续	1712	1.94%	多头	2425	1.95%	反弹	3749	2.13%
晚间	1631	1.85%	反弹	2323	1.86%	交易	3574	2.03%
多单	1514	1.71%	多单	2299	1.85%	继续	3514	2.00%
空单	1505	1.70%	点金	2264	1.82%	非农	3458	1.97%
反弹	1433	1.62%	日内	2242	1.80%	生意	3447	1.96%
布局	1423	1.61%	现价	2192	1.76%	如何	3251	1.85%
EIA	1385	1.57%	晚间	2160	1.73%	价格	2961	1.68%
震荡	1380	1.56%	美元	2135	1.71%	投资	2860	1.63%
点金	1368	1.55%	空单	2116	1.70%	晚间	2630	1.50%
解析	1347	1.53%	在线	1817	1.46%	商品	2592	1.47%
实时	1250	1.42%	布局	1784	1.43%	美元	2508	1.43%
喊单	1210	1.37%	直播	1709	1.37%	后市	2341	1.33%
多头	1207	1.37%	解析	1615	1.30%	金价	2181	1.24%
非农	1198	1.36%	如何	1595	1.28%	国内	2135	1.21%
解盘	1083	1.23%	做空	1527	1.23%	螺纹	2081	1.18%
投资	1056	1.20%	实时	1467	1.18%	下周	1984	1.13%
美元	1008	1.14%	交易	1456	1.17%	上涨	1953	1.11%
合计	88,302	100.00%	合计	124,578	100.00%	合计	175,886	100.00%

根据分词结果制作词云，可以很直观地反映统计结果中的高频词情况，并从中寻找关联。图 10、图 11 和图 12 是运行出来的图云结果。

Figure 10. Gold bar word cloud
图 10. 黄金吧词云

Figure 11. Crude word cloud
图 11. 原油吧词云

Figure 12. Futures word cloud
图 12. 期货吧词云

3.1.2. 分析热点词关系

分析图 12 期货市场图云可以发现，人们对期货市场比较关注的是黄金和原油。进一步获取原油吧和黄金吧的评论信息，图 10 是从黄金吧获取文本进行处理后得到的云图，可以发现人们在讨论黄金相关信息的时候普遍也关注原油市场的情况；图 11 原油吧情况类似，人们讨论原油价格变动的同时明显关注黄金的交易情况。由此，可以初步断定，黄金与原油之间存在一定的相关性。

此外，三张云图中比较频繁的关键词是“震荡”，说明近期期货市场、黄金和原油状况不是很好，投资者可能遭受了损失。

3.1.3. 市场的真实情况

作为宏观经济的重要指标，黄金价格和石油价格一直受到人们重视，并且认为他们之间有着极其紧密的联系，尽管他们的涨跌幅度不尽相同。

由图 13 所示，在 2017 年至 2020 年之间，原油和黄金价格变化剧烈(黄金的价格波动相对原油小)，但是很明显，两者的变化趋势是相似的，当黄金(原油)价格下降是，原油(黄金)价格也在下降；反之亦然，当然不能排除偶然情况，但是相对概率偏低，所以大体上变化方向是一致的，因而两者变化存在正相关。

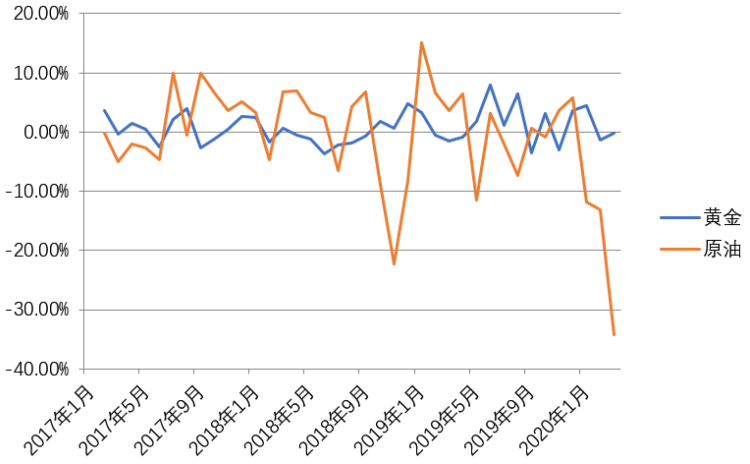

Figure 13. Gold and crude oil price trends
图 13. 黄金与原油价格走势

也就解释了为什么黄金吧和原油吧生成的词云图，黄金和原油都是高频词。

3.1.4. 存在相关性原因

出现以上现象，主要是因为影响两者价格的因素有共同点，当这些因素变动时，会影响黄金和原油的价格朝相同方向变动。这些因素包括美元汇率变动、通货膨胀、以及世界上重大(政治)事件等。首先，在世界交易市场上，美元是黄金和石油的报价货币，因此美元增值或贬值直接影响到黄金和原油的价格，这在词云图 7 和图 8 中都出现美元得到了印证。其次，原油价格随通货膨胀上涨又翻过了加剧通货膨胀，而黄金具有良好的抗通货膨胀性和保值性，因此高通货膨胀时期，黄金价格通常也会上涨。最后，原油价格和黄金价格为国际政局关系的晴雨表，两者都对国际政治局势的动荡或偶发事件的反应非常敏感。

例如，近期内黄金和原油价格波动很大，但是近期内美元汇率变动不大，也没有发生通货膨胀，因此可以推测发生了世界上重大的(政治)事件。搜集相关资料，可以推测近期发生的新冠肺炎产生的影响。新冠肺炎的不断蔓延，使其成为国际上重大事件。肺炎影响正常的生产运营，进而影响企业的运营，进一步作用于股市，股市没有资金，只能依靠变卖黄金来补充，因此黄金价格下跌，原油价格进一步也跟着下降。

3.1.5. 投资者的投资选择

从表 2 评论词条中寻找规律，原油吧和黄金吧投资者普遍讨论“策略、分析、行情、直播、反弹、震荡、非农、美元”等，说明投资者关注整个市场的行情走势，“震荡”表明对黄金原油市场不太看好，“策略、反弹”等表明在寻求新的投资策略或者等待反弹机会，“直播、非农”等表明他们比较关注在线直播、非农等行业。

Table 2. Comparison of high-frequency vocabulary for crude oil and gold
表 2. 原油吧、黄金吧高频词汇对比

	原油吧频数	%	黄金吧频数	%
原油	24,286	28.70%	14,102	11.75%
黄金	18,589	21.97%	46,201	38.50%
分析	4867	5.75%	6225	5.19%
策略	4352	5.14%	5319	4.43%

Continued

行情	3603	4.26%	4804	4.00%
在线	2357	2.79%	1817	1.51%
直播	2017	2.38%	1709	1.42%
走势	1926	2.28%	3052	2.54%
现价	1867	2.21%	2192	1.83%
日内	1728	2.04%	2242	1.87%
继续	1712	2.02%	3136	2.61%
晚间	1631	1.93%	2160	1.80%
多单	1514	1.79%	2299	1.92%
空单	1505	1.78%	2116	1.76%
反弹	1433	1.69%	2323	1.94%
布局	1423	1.68%	1784	1.49%
震荡	1380	1.63%	2828	2.36%
点金	1368	1.62%	2264	1.89%
解析	1347	1.59%	1615	1.35%
实时	1250	1.48%	1467	1.22%
多头	1207	1.43%	2425	2.02%
非农	1198	1.42%	2751	2.29%
投资	1056	1.25%	3034	2.53%
美元	1008	1.19%	2135	1.78%
合计	84,624	100.00%	120,000	100.00%

3.1.6. 对投资决策的建议

黄金和原油市场一般存在正向相关性，对投资者的投资决策有一定的帮助。以图 14 的 2019 年 12 月至今黄金原油走势图分析可以发现，原油的价格波动比黄金大，而且原油变化滞后于黄金。当前时点，在已知黄金价格已经开始回升的情况下，投资者可以预测原油价格未来有很大几率会上涨，此时可以考虑在原油的低点买入。

4. 总结与未来展望

本文在对股市、货币市场的评论采用 Python 文本挖掘与分析处理之后，很清晰明了地得出两点：其一是近期股市热点为“新冠”“肺炎”，进一步分析发现与之相关的口罩、医药、在线教育和云办公市场表现突出；其二是期货市场上黄金和原油大受关注，而且黄金和原油价格之间存在正向的相关性。

可以发现，在如今互联网迅猛发展的时代，投资者可以在互联网论坛、APP 评论区等地方进行投资者与投资者间交流、投资者与上市公司董事会秘书交流等，从而会导致投资者很容易在不经意间留下大量的语言文本数据，如评论、问答、分析等等。而这些数据完全可以作为大数据分析的原始材料。在进行数据挖掘与分析中，Python 因其强大的挖掘能力，能够使纷繁复杂的文本信息转化为简单易懂的图像，一目了然地展示了投资者的投资关注点。故而，本文认为，Python 分析股票市场上投资者的投资行为具有强大的分析作用，未来能够继续在文本挖掘、数据分析、数据可视化等方面得到广泛应用。

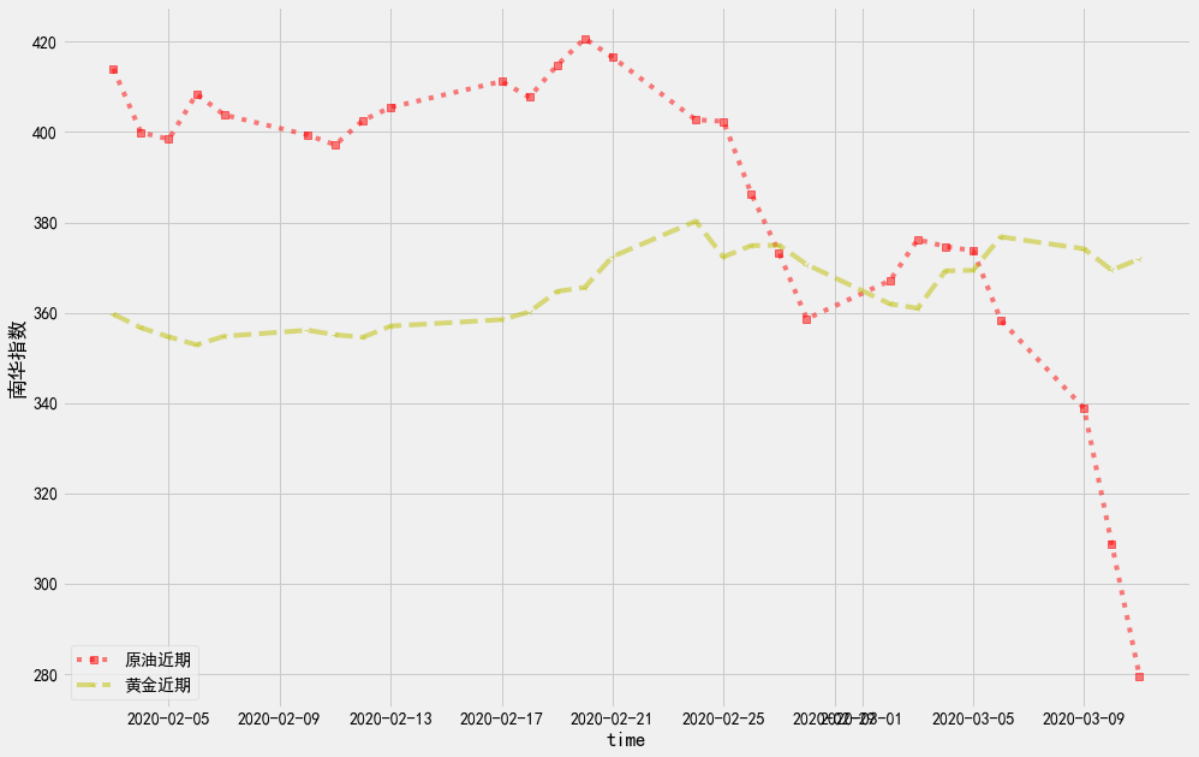

Figure 14. Recent charts of crude oil and gold
图 14. 原油、黄金近期走势图

综上所述，虽然互联网上文本资料信息量大、没有规律和重点不突出常常被人们忽视，但是通过 Python 等软件充分挖掘、量化之后，能较高程度地转化成有分析价值的信息，可以帮助投资者判断和决策。

基金项目

本论文得到了厦门国家会计学院 2019 年“云顶课题：YD20190101Python 财务数据分析”项目的支持。

参考文献

[1] Jun, S.P., Park, D.H. and Yeom, J. (2014) The Possibility of Using Search Traffic Information to Explore Consumer Product Attitudes and Forecast Consumer Preference. *Technological Forecasting & Social Change*, **86**, 237-253. https://doi.org/10.1016/j.techfore.2013.10.021

[2] Preis, T., Moat, H.S. and Stanley, H.E. (2013) Quantifying Trading Behavior in Financial Markets Using Google Trends. *Scientific Reports*, **3**, 1684. https://doi.org/10.1038/srep01684

[3] 张博凯. 数据挖掘在股票预测中的应用[J]. 当代经济, 2017(8): 46-47.

[4] 李源, 李杰. 股票市场个人投资行为分析及策略建议——基于行为金融学视角[J]. 金融经济, 2013(12): 91-92.

[5] 范珈瑜. 基于文本挖掘的游客对古镇旅游态度的分析[J]. 大数据, 2017(6): 95-103.

[6] 廖勇毅, 丁怡心. 基于 Python 的股票定向爬虫实现[J]. 电脑编程技巧与维护, 2019(5): 45-46.

[7] 赵光亮, 令狐雨薇, 朱德孙, 等. 基于 Python 的通用论坛正文提取研究[J]. 电脑知识与技术, 2018(2): 259-260.