

基于大数据挖掘的行业轮动策略

互联网大数据挖掘系列专题之（十）

报告摘要:

● 行业轮动策略思想

互联网的快速发展和普及为人们获取信息提供了更为方便快捷的渠道。随着互联网的普及和运用，信息也呈现爆炸式增长的态势。信息早已不再是一种稀缺的资源，相反，投资者的注意力成为一种稀缺的认知资源。在金融市场上，当投资者面临海量信息时，不得不将有限的注意力进行分配，从而对投资决策产生影响的往往是那些被关注的信息。我国A股市场是散户为主的市场，众多中小投资者对A股的关注和热情是股市上涨的重要推动力。根据搜索引擎等媒介构建的舆情指数能很好地反映投资者对各行业的关注。本专题通过对舆情数据和行情数据之间的关系进行研究，分析其在量化研究中的应用效果。

● 策略构建

相关性分析结果显示，申万一级行业指数的历史行情数据和对应的滞后5期历史舆情数据具有较强的正相关关系，大部分行业的正相关系数均值处于0.4~0.6之间，即中度相关关系。当投资者对某个行业的关注度急剧上升时，说明该行业可能是近期投资热点，同时若对应市场行情还未上涨或者涨幅很小，接下来市场行情上涨的可能性较大。

当某个行业满足舆情指数涨幅大于阈值A，对应的行业指数涨幅大于阈值B且小于阈值C时，发出买入信号。为了更有效地利用资金，对筛选行业数目设置阈值，若一周内选出的热点行业数目太多，对选出的行业做进一步筛选，固定持仓一段时间后平仓。

● 实证结果

实证结果表明，在实证区间2011年至2016年内，策略年化收益率为19.93%，相对于行业等权指数的年化超额收益率为20.83%，胜率为53.57%。在原策略上加入交易区间内舆情与行情的相关性判断，改进策略自2011年以来，实现了约25.55%的年化收益率，相对于行业等权指数的年化超额收益率为22.21%，策略胜率为54.95%。

● 风险提示

本报告提出的行业轮动策略基于行业指数舆情数据，舆情是投资者投资情绪的一个方面表现，市场行情受到各种综合因素影响，注意模型失效的风险。

图1行业轮动策略累计净值

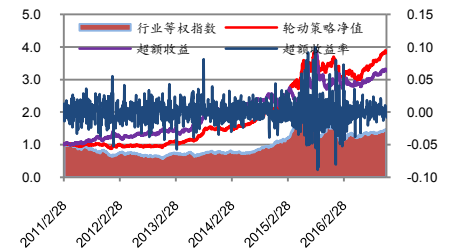


表1行业轮动策略历史表现

累计净值	3.82
年化收益率	25.55%
年化超额收益率	22.21%
胜率	54.95%

分析师：史庆盛 S0260513070004

02087577060

sqs@gf.com.cn

相关研究:

《基于互联网挖掘的热点选股策略——互联网大数据挖掘系列专题之（五）》	2015-04-09
《多维数据系的大数据择时策略研究——互联网大数据挖掘系列专题之（八）》	2016-05-17
《基于大数据挖掘的概念轮动策略-互联网大数据挖掘系列专题之（九）》	2016-08-22

联系人：陈原文 020-87576976
chenyuanwen@gf.com.cn

目录索引

一、前言	4
1.1 大数据背景下网络舆情数据与量化投资	4
1.2 互联网大数据研究体系	4
1.3 行业轮动策略介绍	6
二、行业轮动策略构建	9
2.1 样本数据	9
2.2 相关性分析	11
2.3 策略原理	13
三、实证分析	14
3.1 实证说明	14
3.2 实证结果	14
3.3 参数敏感性测试	16
3.4 策略改进	19
四、总结	21
4.1 总结	21
4.2 未来研究方向	21

图表索引

图 1 行业轮动策略累计净值	1
图 2: 广发金工数据抓取体系	5
图 3: 广发金工大数据挖掘策略框架	5
图 4: 2015 年和 2016 年各行业涨跌幅	6
图 5: 历史年度行业涨幅前三分布	6
图 6: 行业轮动策略研究框架	7
图 7: 相似性匹配行业轮动策略历史回测结果	7
图 8: 羊群效应行业轮动策略历史回测结果	8
图 9: 因子极值行业轮动策略历史回测结果	9
图 10: 百度指数——趋势研究	10
图 11: 360 指数——走势分析	10
图 12: 谷歌趋势——热度变化趋势	11
图 13: 申万一级行业正相关系数分布情况	12
图 14: 申万一级行业指数与对应舆情指数相关性分析（采掘行业）	12
图 15: 申万一级行业指数与对应舆情指数相关性分析（汽车行业）	13
图 16: 热点行业筛选方法	14
图 17: 行业轮动策略历史表现	15
图 18: 行业轮动策略分年度表现	16
图 19: 舆情指数涨幅阈值 A 敏感性测试结果	17
图 20: 行业指数涨幅阈值 B 敏感性测试结果	17
图 24: 行业指数涨幅阈值 C 敏感性测试结果	18
图 22: 每周持有行业指数数目选择	18
图 23: 改进的行业轮动策略历史表现	19
图 24: 改进的行业轮动策略分年度表现	20
图 28: 原策略与改进策略分年度表现比较	20
表 1 行业轮动策略历史表现	1
表 2: 互联网挖掘小工具一览	5
表 3: 行业月度表现一览	6
表 4: 相似性匹配行业轮动策略分年度表现	8
表 5: 羊群效应行业行业策略分年度表现	8
表 6: 因子极值行业轮动策略分年度表现	9
表 7: 舆情数据来源一览	10
表 8: 申万一级行业指数与舆情数据相关性分析	11
表 9: 行业轮动策略历史表现一览	15
表 10: 行业轮动策略分年度表现一览	15
表 11: 每周持有行业指数一览	16
表 12: 参数敏感性测试结果	18
表 13: 改进的行业轮动策略历史表现一览	19
表 14: 改进的行业轮动策略分年度表现一览	20

一、前言

1.1 大数据背景下网络舆情数据与量化投资

2017年1月22日中国互联网络信息中心（CNNIC）发布了第39次《中国互联网络发展状况统计报告》，报告显示，截至2016年12月，中国网民规模达7.31亿，相当于欧洲人口总量，互联网普及率达到53.2%。互联网技术的快速发展和普及为人们获取信息提供了更为方便快捷的渠道，信息的表现形式不断丰富，内容也逐渐增多，甚至呈现爆炸式增长，互联网上沉淀的数据规模呈现指数型增长。

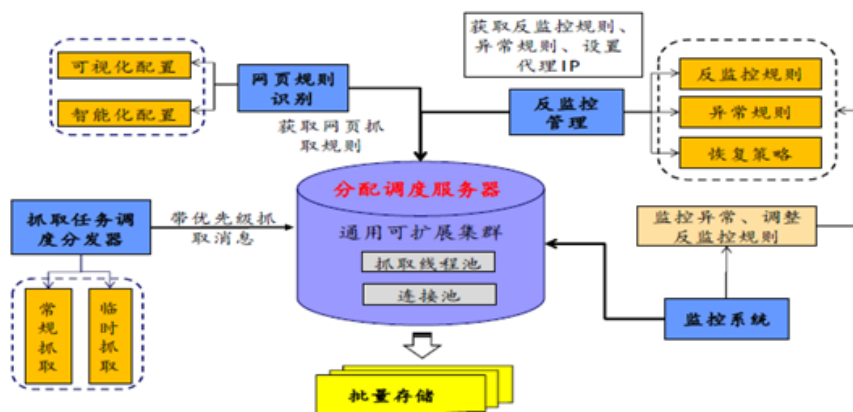
大数据技术的发展使得信息获取成本下降，获取信息更加全面。显然，信息早已不再是一种稀缺的资源，相反，人的注意力成为一种稀缺的认知资源，信息的丰富性与个体注意力的有限性之间形成一种矛盾。美国政治学家赫伯特·西蒙（Herbert Simon）在其著作《管理行为》中提到，伴随着信息量的不断增多，个体的注意力逐渐变得稀缺，因为海量信息会消耗个体的注意力，导致注意力匮乏。在金融市场上，当投资者面临海量信息时，注意力的稀缺性决定他们必须将有限的注意力进行分配，而配置的结果将会对投资决策产生影响，在决策过程中发挥作用的往往是那些被关注的信息，而被忽略的信息无法及时反应到市场中。因此，在大数据时代，如何有效地从海量数据中挖掘投资者行为信息，并运用于投资决策中，具有重要意义。

我国A股市场是散户为主的市场，众多中小投资者对A股的关注和热情是股市上涨的重要推动力，而投资者对市场的热情往往体现在热门财经网站的新闻阅读量、搜索网站的搜索量等媒体介质上，对这些“网络热度”即时的分析和利用，可以从新的视角上了解投资者情绪，指导投资。第39次《中国互联网络发展状况统计报告》显示，截止2016年12月，我国搜索引擎用户规模6.02亿，使用率82.4%，搜索引擎是网民了解信息的主要渠道。百度是世界上规模最大的中文搜索引擎，在国内有着巨大的用户基数，每天处理来自一百多个国家的超过一亿人次的搜索请求。以百度的海量网民行为数据为基础百度推出了百度指数。2014年奇虎360以国内第二大搜索引擎360搜索数据为基础推出了360指数；此外全球最大搜索引擎Google在2006年推出了谷歌趋势（Google Trends）。根据搜索引擎构建的舆情指数覆盖面广、时效性强，能够科学地反映某个关键词的搜索规模有多大，一段时间内的涨跌态势以及相关的新闻舆论变化，拥有强大的趋势和舆情洞察功能，能够很好地反映了广大用户，尤其是广大投资者对股票市场各个行业、上市公司等的关注以及舆情变化趋势。若能有效地组织和利用这些舆情数据，进行量化分析，将给传统的投资决策等带来新的机遇。

1.2 互联网大数据研究体系

在互联网数据挖掘与研究方面，广发金工建立了完善的数据抓取平台以及对互联网大数据各个数据维度的研究视角。

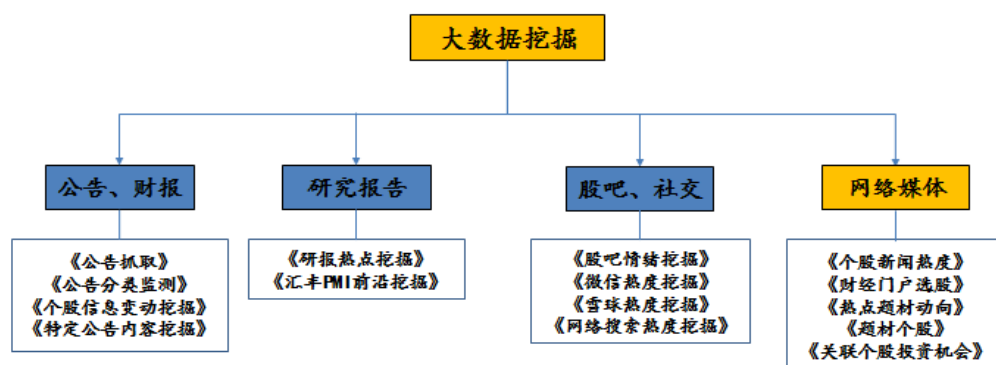
图2：广发金工数据抓取体系



数据来源：广发证券发展研究中心

广发金工研究了多个维度的互联网数据，比如从公告、财报角度研究的报告《公告披露背后隐藏的投资机会——互联网大数据挖掘系列专题之（二）》等；从股吧、社交角度研究的报告《倾听股吧之声，洞察大盘趋势——互联网大数据挖掘系列专题之（三）》等；以及从网络媒体角度分析的报告《基于互联网挖掘的热点选股策略——互联网大数据挖掘系列专题之（五）》、《基于大数据挖掘的关联个股投资机会——互联网大数据挖掘系列专题之（六）》、《基于大数据挖掘的Smart Beta策略——互联网大数据挖掘系列专题之（七）》、《多维数据下的大数据择时策略研究——互联网大数据挖掘系列专题之（八）》、《基于大数据挖掘的概念轮动策略——互联网大数据挖掘系列专题之（九）》等。

图3：广发金工大数据挖掘策略框架



数据来源：广发证券发展研究中心

基于对大数据的研究积累，广发金工也开发了一系列的互联网挖掘小工具，包括上市公司公告抓取、研究报告抓取、搜索量抓取、关注度抓取等小工具。

表2：互联网挖掘小工具一览

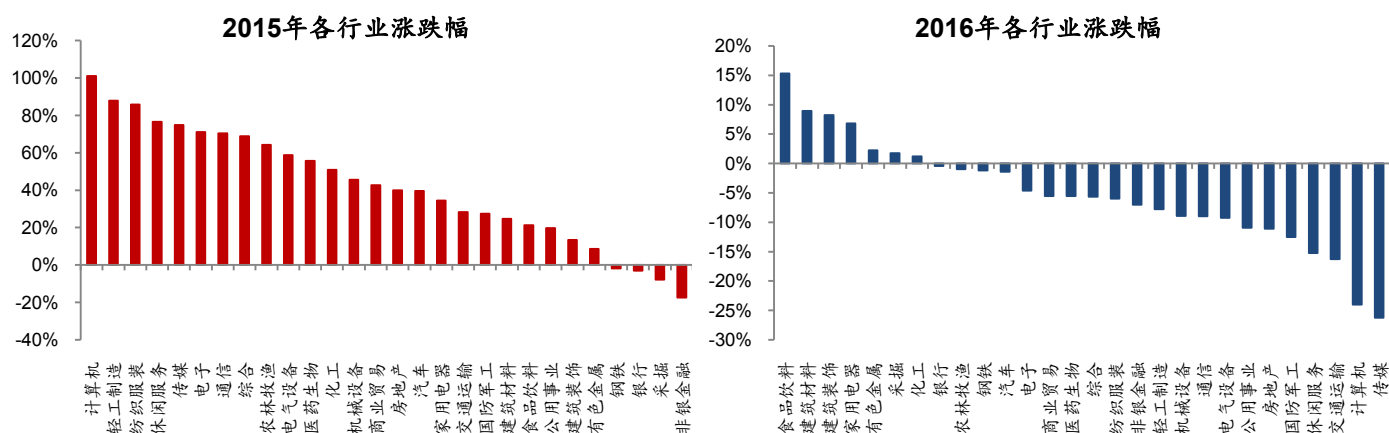
- | | |
|---------------|----------------|
| 1、A股新闻热度搜索工具 | 2、A股上市工具公告抓取工具 |
| 3、上市公司信息变更抓取 | 4、文本信息批量识别及处理 |
| 5、汇丰PMI实时监测工具 | 6、个股研报热点监测工具 |
| 7、特定公告实时监测工具 | 8、财经小编选股工具 |
| 9、舆情指数搜索工具 | |

数据来源：广发证券发展研究中心

1.3 行业轮动策略介绍

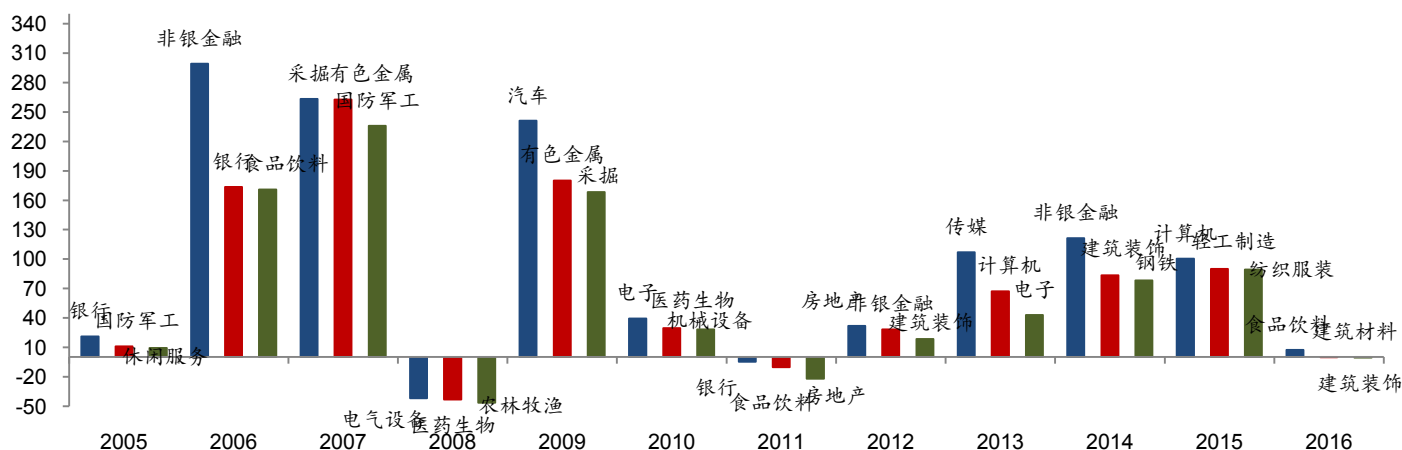
从历史各个行业的表现来看，行业板块业绩分化严重。2015年计算机行业涨幅最高约为100.29%，垫底行业为非银金融，跌幅-16.90%。2016年食品饮料行业涨幅最高约为7.43%，而传媒行业涨幅约为-32.39%。分年度看，历年涨幅排名前三的行业中，银行、非银等出现频率较大，而大多行业仅出现一次甚至未出现；分月度看，从2005年到2016年，单月涨幅前三的行业中，银行、非银、食品饮料频率最多，而化工、轻工制造、公用事业最少，连续两个月涨幅前三的行业中，银行、有色金属、计算机频率最多，机械设备、化工、轻工制造最少，行业轮动切换快。如何把握行业轮动成权益配置的关键！

图4：2015年和2016年各行业涨跌幅



数据来源：Wind、广发证券发展研究中心

图5：历史年度行业涨幅前三分布



数据来源：Wind、广发证券发展研究中心

表3：行业月度表现一览

	银行	非银金融	食品饮料	房地产	休闲服务	国防军工	有色金属	计算机	采掘	传媒	医药生物	通信	家用电器	综合
单月前三	34	28	25	25	23	23	22	22	20	19	18	18	16	15
连续两个	14	5	5	1	3	4	7	6	4	6	5	1	1	2

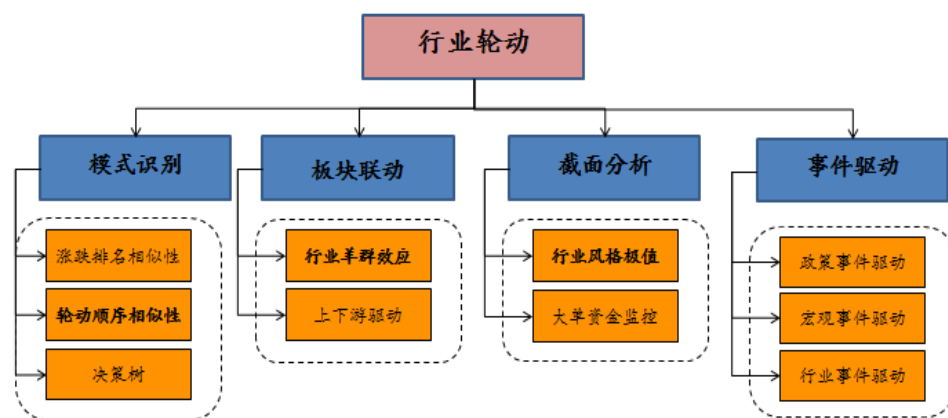
月前三

	农林 牧渔	钢铁	电子	建筑 装饰	电气 设备	建筑 材料	汽车	机械 设备	商业 贸易	纺织 服装	交通 运输	公用 事业	化工	轻工 制造
单月前三	14	14	13	13	13	11	11	8	7	6	6	4	3	1
连续两个 月前三	4	2	2	3	2	2	1	0	1	2	1	0	0	0

数据来源：Wind、广发证券发展研究中心

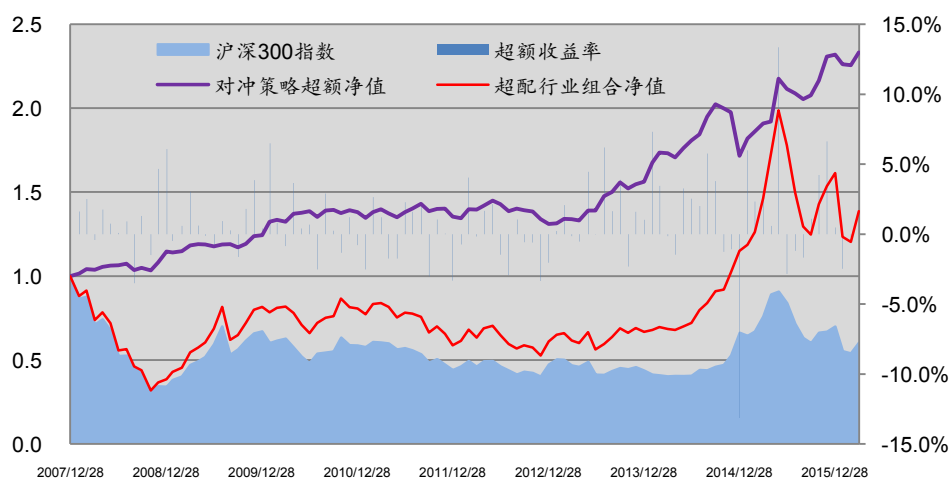
目前，行业轮动策略的研究主要从模式识别、板块联动、截面分析、事件驱动等角度出发，广发金工对此进行了一系列的深入研究，比如：从模式识别角度研究的报告《寻找行业“似曾相识”的轮动规律》通过观察近期行业之间的涨跌顺序，与历史样本进行相似性匹配，寻找“似曾相识”的样本时期，并以随后的强势行业作为当前配置的依据；从板块联动角度研究的报告《捕捉羊群效应下的行业轮动机会》根据行业在过去一周是否存在羊群效应，并且根据龙头股数量是否适宜来选择行业进行轮动；从截面分析角度出发的研究报告《基于个股极值比例的行业轮动策略》根据行业内成分股的各因子是否达到创新高或创新低，计算各行业极值比例，来挖掘行业板块机会。

图6：行业轮动策略研究框架



数据来源：广发证券发展研究中心

图7：相似性匹配行业轮动策略历史回测结果



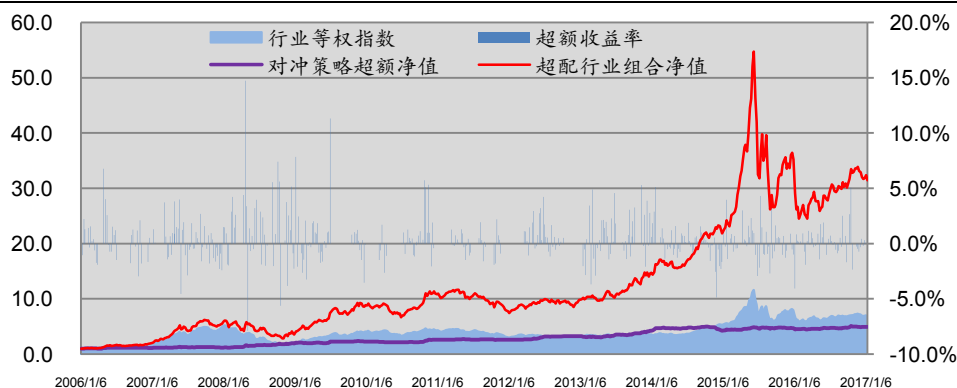
数据来源：Wind、广发证券发展研究中心

表4：相似性匹配行业轮动策略分年度表现

时间	超额收益率	胜率	最大回撤
全样本	137.63%	62.04%	15.18%
2008	14.75%	75.00%	3.75%
2009	8.39%	66.67%	1.63%
2010	11.02%	66.67%	2.52%
2011	-1.88%	58.33%	5.24%
2012	-3.23%	33.33%	9.59%
2013	19.19%	66.67%	2.32%
2014	9.81%	58.33%	15.18%
2015	35.11%	75.00%	5.62%
2016	2.50%	58.33%	2.69%

数据来源：广发证券发展研究中心

图8：羊群效应行业轮动策略历史回测结果



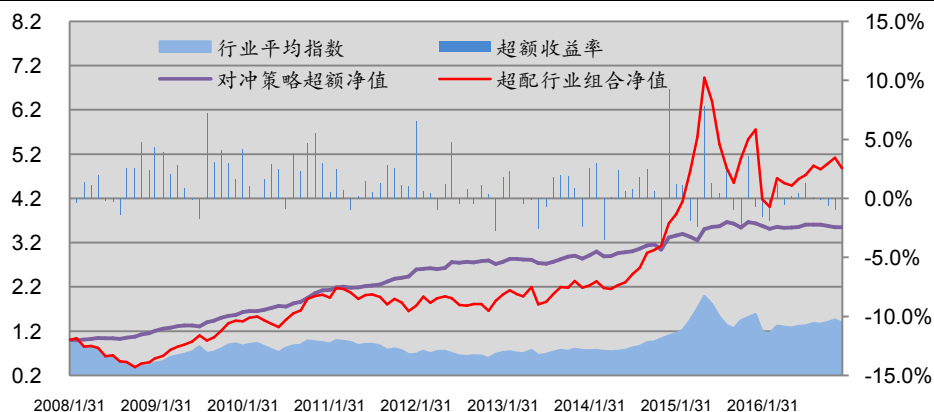
数据来源：Wind、广发证券发展研究中心

表5：羊群效应行业行业策略分年度表现

时间	超额收益率	胜率	最大回撤
全样本	392.00%	56.30%	14.70%
2006	13.40%	56.30%	4.50%
2007	3.90%	51.10%	10.80%
2008	70.80%	65.00%	7.60%
2009	17.70%	52.80%	6.30%
2010	17.30%	61.30%	4.10%
2011	0.40%	51.90%	6.00%
2012	23.80%	65.50%	1.90%
2013	32.20%	57.10%	5.50%
2014	1.10%	56.00%	14.50%
2015	10.80%	52.10%	6.00%
2016	5.70%	54.20%	4.30%

数据来源：广发证券发展研究中心

图9：因子极值行业轮动策略历史回测结果



数据来源：广发证券发展研究中心

表6：因子极值行业轮动策略分年度表现

时间	超额收益率	胜率	最大回撤
全样本	255.11%	71.70%	4.26%
2008	15.40%	63.64%	1.91%
2009	35.85%	83.33%	1.81%
2010	35.31%	91.67%	0.89%
2011	14.74%	91.67%	0.95%
2012	11.61%	66.67%	2.77%
2013	4.41%	54.55%	3.83%
2014	17.06%	83.33%	3.76%
2015	9.61%	58.33%	4.26%
2016	-2.43%	50.00%	3.49%

数据来源：广发证券发展研究中心

目前行业轮动策略的主要方法是从因子价量数据及指数的价量数据出发寻找可能的行业轮动规律；在当前大数据时代下，行业的舆情数据提供了新的研究视角，这些舆情数据能够更为及时地反应当前市场上投资者情绪，对投资者的投资决策起到潜移默化的作用，我们可以从舆情数据出发来研究行业轮动的规律。

二、行业轮动策略构建

2.1 样本数据

行业数据来源于wind的申万一级行业指数的数据，一共包含采掘、传媒等28个行业，其中综合行业历史舆情数据难以获取，在此不予考虑，各行业的行情数据来源于wind的申万一级行业指数的开盘价和收盘价数据，样本时间区间为2011年至2016年。相关行业的历史舆情数据则通过搜索引擎、财经网站等互联网渠道获得。目前以搜索引擎的搜索数据为基础构建的舆情指数主要有百度指数、360指数、谷歌趋势（Google Trends）等。百度指数以全球最大的中文搜索引擎百度的搜索数据为基础构建，360指数以搜索引擎360搜索数据为基础构建，据最新的《中国网民搜索行为调查报告》，百度搜索的常用率遥遥领先（手机端88.6%，PC端84.3%），360搜索次之（手机端2.6%，PC端11.9%），两者占据了国内搜索引擎市场的90%以上；谷歌趋势以全球最大搜索引擎Google的搜索数据为基础构建，但在国内常用率较低。

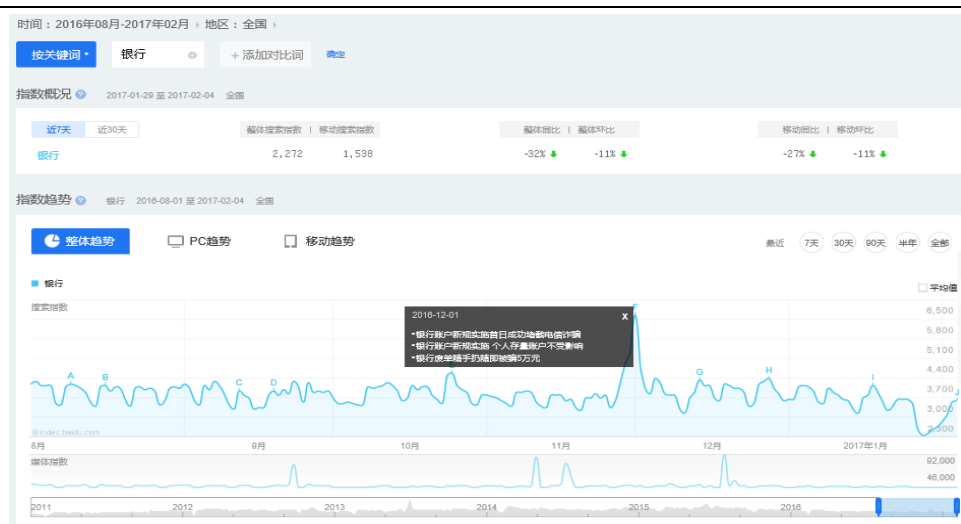
可见根据搜索引擎构建的搜索指数覆盖面广，时效性强，能够很好地反映投资者对各行业的关注度及其变化趋势。

表7：舆情数据来源一览

指数	含义	功能模块	数据区间	上线时间
百度指数	以网民在百度的搜索量为数据基础，以关键词为统计对象，科学分析并计算出各个关键词在百度网页搜索中搜索频次的加权和。	趋势研究、需求图谱、舆情洞察、人群画像	2011 至今	2007.11
360 指数	以网民在 360 搜索的搜索量为数据基础，计算出各个关键词在 360 搜索网页搜索中搜索频次的加权和。	走势分析、需求图谱、人群特征	2013 至今	2014.5
谷歌趋势	基于搜索日志分析的应用产品，通过分析 Google 全球数以十亿计的搜索结果，给出某一搜索关键词各个时期下在 Google 被搜索的频率和相关统计数据。	热度变化趋势、区域搜索热度、相关主题/查询	2004 至今	2006.5

数据来源：广发证券发展研究中心

图10：百度指数——趋势研究



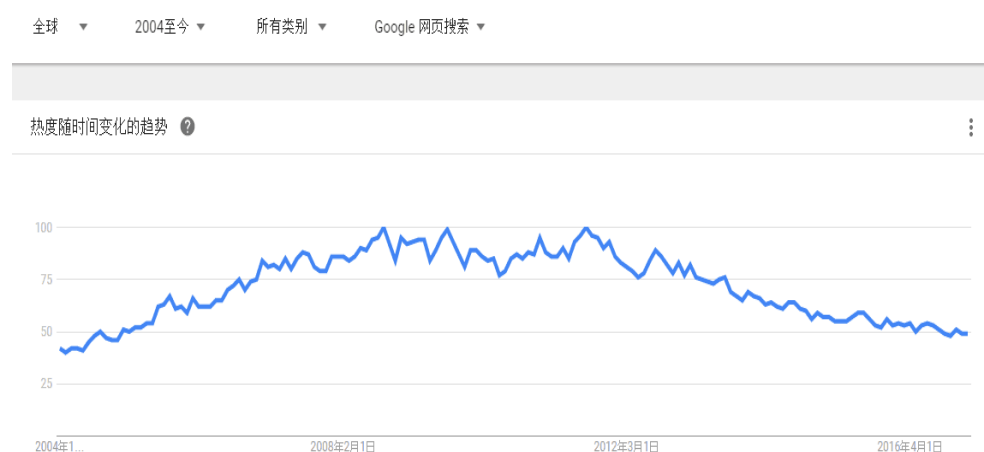
数据来源：百度指数、广发证券发展研究中心

图11：360指数——走势分析



数据来源：360指数、广发证券发展研究中心

图12：谷歌趋势——热度变化趋势



数据来源：谷歌趋势、广发证券发展研究中心

2.2 相关性分析

对申万一级行业指数的历史行情数据和对应的历史舆情数据进行相关分析，滞后期为5天，得到历史行情数据与舆情数据的相关系数。大部分行业的正相关均值处于0.4~0.6之间，存在中度的相关性。从相关系数分布来看，行情数据与舆情数据存在中度相关、强相关、极强相关的仍占50%以上，说明行情与滞后5阶舆情数据存在较强的正相关关系，采掘、汽车、化工等行业存在强、极强相关关系更多。

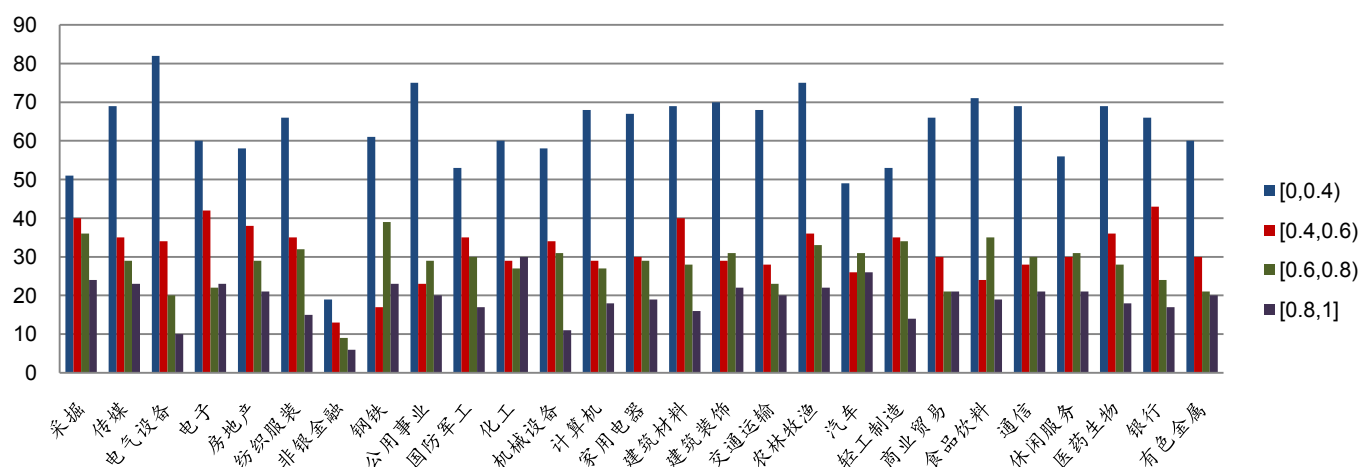
表8：申万一级行业指数与舆情数据相关性分析

申万一级行业	正相关均值	正相关个数	[0,0.4)	[0.4,0.6)	[0.6,0.8)	[0.8,1]
采掘	0.51	151	51	40	36	24
传媒	0.46	156	69	35	29	23
电气设备	0.38	146	82	34	20	10
电子	0.46	147	60	42	22	23
房地产	0.47	146	58	38	29	21
纺织服装	0.45	148	66	35	32	15
非银金融	0.45	47	19	13	9	6
钢铁	0.48	140	61	17	39	23
公用事业	0.44	147	75	23	29	20
国防军工	0.48	135	53	35	30	17
化工	0.49	146	60	29	27	30
机械设备	0.45	134	58	34	31	11
计算机	0.44	142	68	29	27	18
家用电器	0.46	145	67	30	29	19
建筑材料	0.46	153	69	40	28	16
建筑装饰	0.46	152	70	29	31	22
交通运输	0.44	139	68	28	23	20
农林牧渔	0.46	166	75	36	33	22
汽车	0.51	132	49	26	31	26

轻工制造	0.47	136	53	35	34	14
商业贸易	0.45	138	66	30	21	21
食品饮料	0.46	149	71	24	35	19
通信	0.46	148	69	28	30	21
休闲服务	0.5	138	56	30	31	21
医药生物	0.44	151	69	36	28	18
银行	0.45	150	66	43	24	17
有色金属	0.46	131	60	30	21	20

数据来源：Wind、广发证券发展研究中心

图13：申万一级行业正相关系数分布情况



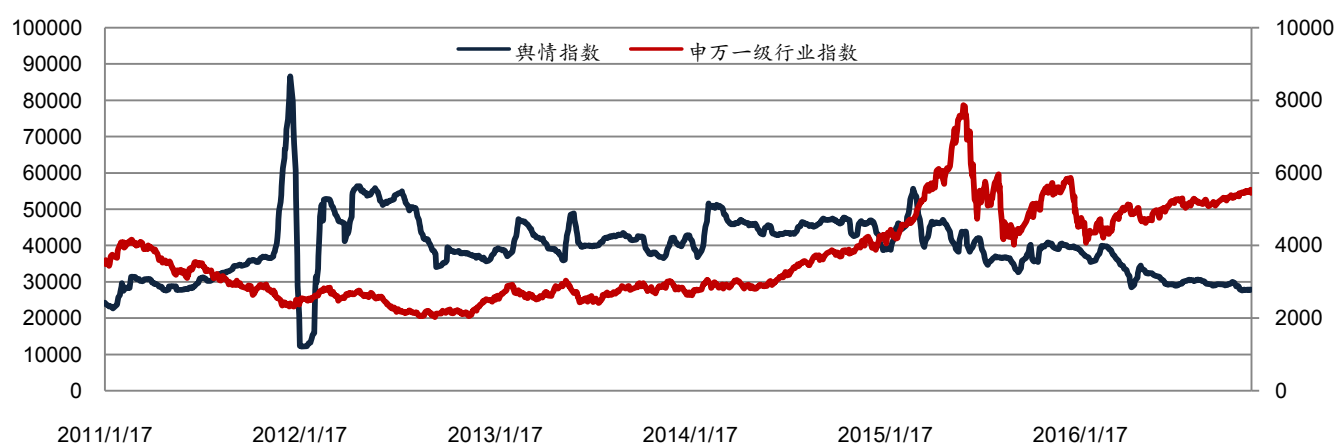
数据来源：Wind、广发证券发展研究中心

图14：申万一级行业指数与对应舆情指数相关性分析（采掘行业）



数据来源：Wind、广发证券发展研究中心

图15：申万一级行业指数与对应舆情指数相关性分析（汽车行业）



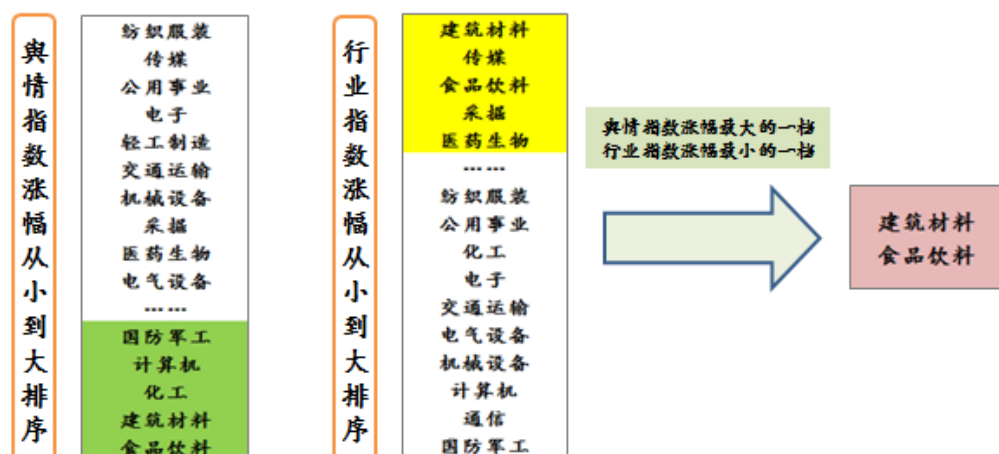
数据来源：Wind、广发证券发展研究中心

2.3 策略原理

通过相关性分析发现各行业的滞后5阶舆情数据和行情数据有较强的正相关关系，且舆情信息的变化要先于申万一级行业指数的变化，当投资者对于某个行业的关注度急剧上升时，说明该行业可能是近期投资热点。那么当某个行业指数的舆情涨幅较大，同时对应市场行情还未上涨或者涨幅很小，未来几天市场行情上涨的可能性较大。

本策略以周为统计区间，分别计算各个行业的舆情指数和行情指数的涨跌幅。为了更好地度量统计区间内各行业的舆情指数和行情指数是否上涨、涨幅多大，对涨幅设置阈值。当某个行业的舆情指数和行情指数满足阈值条件：舆情指数涨幅大于阈值A，对应的行业指数涨幅大于阈值B且小于阈值C时，发出买入信号。为了更加有效地利用资金，将资金分配给未来存在较大上涨可能性的行业，对筛选行业数目设置阈值n，如果在一个周内选出的热点行业数目太多，对选出的行业做进一步筛选，将这些行业按舆情指数涨幅排序分成三档，同时按对应的行业指数涨幅分成三档，选择同时满足舆情指数涨幅处于最大的一档，对应行业指数涨幅处于最小的一档的行业。如果筛选不出热点行业，放宽对行业指数涨幅的限制，直至筛选出热点行业，停止筛选过程。筛选出满足各个条件的行业后，在本周的第一个交易日等权买入，固定持仓一段时间后平仓，平仓后持有资产，等待下一个买入信号。

图16：热点行业筛选方法



数据来源：广发证券发展研究中心

三、实证分析

3.1 实证说明

实证区间：2011——2016

数据来源：行业指数历史数据：申万一级行业指数开盘价、收盘价；

舆情指数历史数据：对应行业的舆情指数。

参数设置：舆情指数涨幅阈值A；

申万一级行业指数涨幅阈值B；

申万一级行业指数涨幅阈值C；

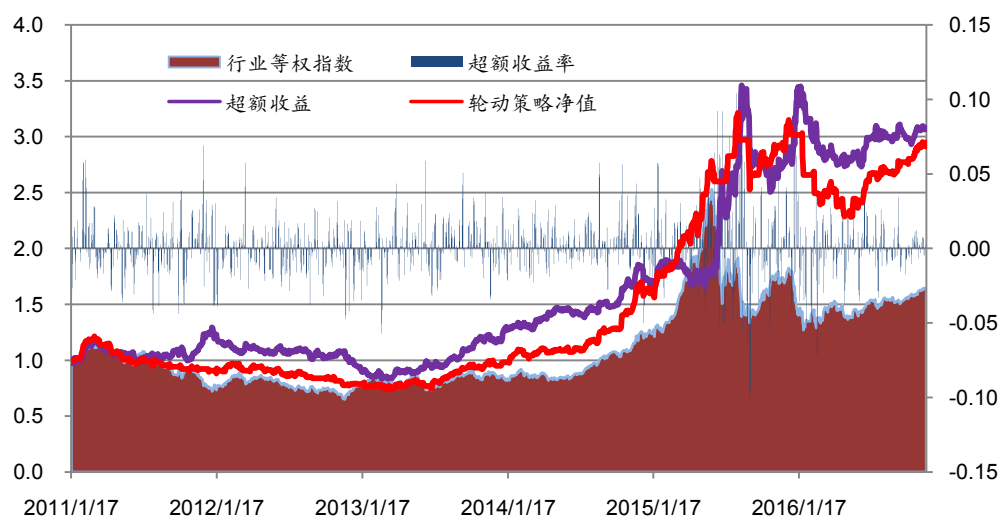
筛选行业个数阈值n。

资金投资权重：等权投资

3.2 实证结果

选择了2011年到2016年的历史数据进行策略的实证分析，并与行业等权指数进行比较，得到实证结果如下：

图17：行业轮动策略历史表现



数据来源：广发证券发展研究中心

表9：行业轮动策略历史表现一览

实证区间	累计净值	年化收益率	年化超额收益率	胜率
2011-2016	2.95	19.93%	20.83%	53.57%

数据来源：Wind、广发证券发展研究中心

策略在实证区间内，累计净值达2.95，年化收益率为19.93%，相对于行业等权指数的年化超额收益率为20.83%，策略胜率为53.57%。

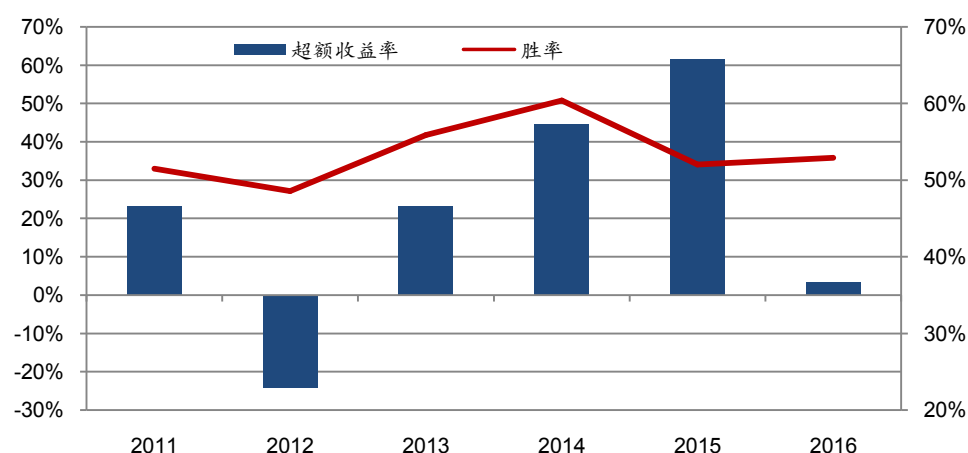
策略分年度表现如下表所示：自2011年以来，策略获得了208.20%的累计超额收益率；除2012年外，其余年份的超额收益率均为正，其中2015年超额收益率达61.39%。

表10：行业轮动策略分年度表现一览

时间	超额收益率	胜率
全样本	208.20%	53.57%
2011	23.01%	51.49%
2012	-24.30%	48.56%
2013	23.25%	55.88%
2014	44.60%	60.41%
2015	61.39%	52.05%
2016	3.35%	52.91%

数据来源：广发证券发展研究中心

图18：行业轮动策略分年度表现



数据来源：广发证券发展研究中心

表11：每周持有行业指数一览

时间	持有行业指数				
2016/6/6	建筑材料				
2016/6/13	农林牧渔	商业贸易	医药生物		
2016/6/20					
2016/6/27	轻工制造				
2016/7/4	非银金融	建筑材料			
2016/7/11	纺织服装	非银金融	建筑装饰		
2016/7/18	机械设备	交通运输			
2016/7/25					
2016/8/1	食品饮料				
2016/8/8	有色金属				
2016/8/15	非银金融	公用事业	家用电器	食品饮料	
2016/8/22	非银金融	机械设备	建筑材料	食品饮料	
2016/8/29	纺织服装				
2016/9/5	房地产				
2016/9/12	纺织服装	计算机	建筑材料	农林牧渔	
2016/9/19	房地产				
2016/9/26	纺织服装	商业贸易			

数据来源：广发证券发展研究中心

3.3 参数敏感性测试

为观察轮动策略的表现对于参数的敏感性，对于涨跌幅阈值，固定其余两个参数；对于筛选行业个数，固定三个涨跌幅阈值，策略的累计净值随变动参数的变化而变化。得到测试结果如下。

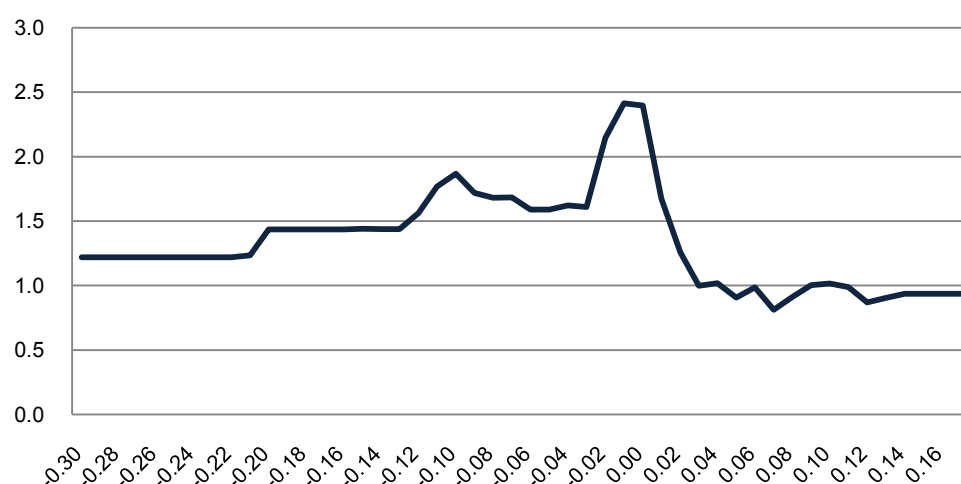
图19：舆情指数涨幅阈值A敏感性测试结果



数据来源：广发证券发展研究中心

上图为在固定阈值B和C、阈值A变动的情况下，累计净值随阈值A的变化情况，可以看到累计净值在0.5到3.0之间变动，阈值A处于0.09到0.4之间时波动较大，阈值A大于0.4后较为平稳。

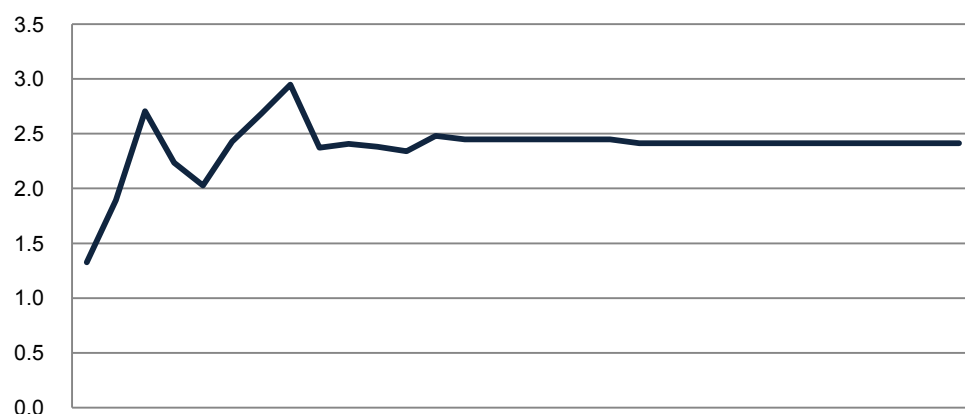
图20：行业指数涨幅阈值B敏感性测试结果



数据来源：广发证券发展研究中心

上图为在固定阈值A和阈值C、阈值B变动的情况下，累计净值的变动情况。在-0.03到0.03之间累计净值的波动幅度较大，其他区间较为平稳。当阈值B大于0.03，累计净值在1附近变动且略低于1。

图21：行业指数涨幅阈值C敏感性测试结果



数据来源：广发证券发展研究中心

上图为固定阈值A和B、阈值C变动的情况下，策略累计净值的变动情况。当阈值C在0到0.08之间，累计净值波动较大。当阈值C>0.08，累计净值在2.5左右，当C>0.18，累计净值为2.41并保持不变。

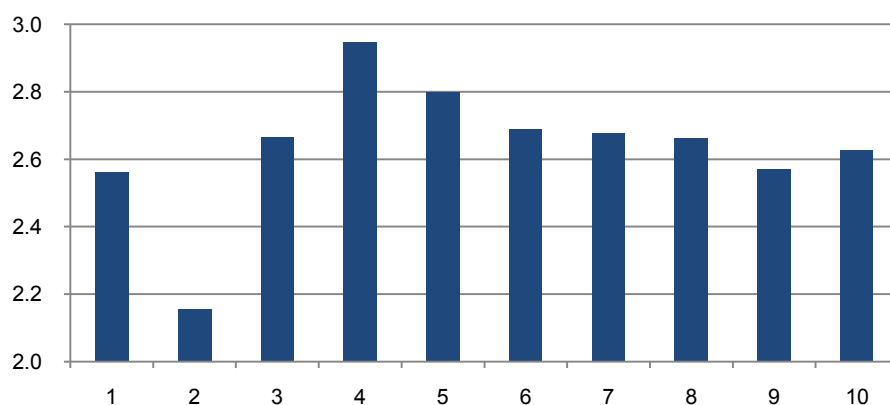
表12：参数敏感性测试结果

变动参数	固定参数	公差	累计净值最大	累计净值最小	策略净值变化标准差
阈值A	B, C	0.01	2.95	0.72	0.47
阈值B	A, C	0.01	2.41	0.81	0.38
阈值C	A, B	0.01	2.95	1.33	0.26

数据来源：广发证券发展研究中心

对不同的参数，按照其自身的特点设置了不同的变化范围，考察不同情况下的累计净值变动情况。可以看到策略对阈值C的变动最不敏感，阈值B次之，对阈值A的变动较为敏感。

图22：每周持有行业指数数目选择



数据来源：广发证券发展研究中心

为了将资金更为有效地分配在行情上涨可能性更大的行业上，我们对持有行业指数的数目n进行选择。当筛选出的行业数据不超过4时，策略累计净值最大。总的来看，随筛选行业的数目的变动，策略累计净值在2.1到3.0之间变动，变动幅度较小。

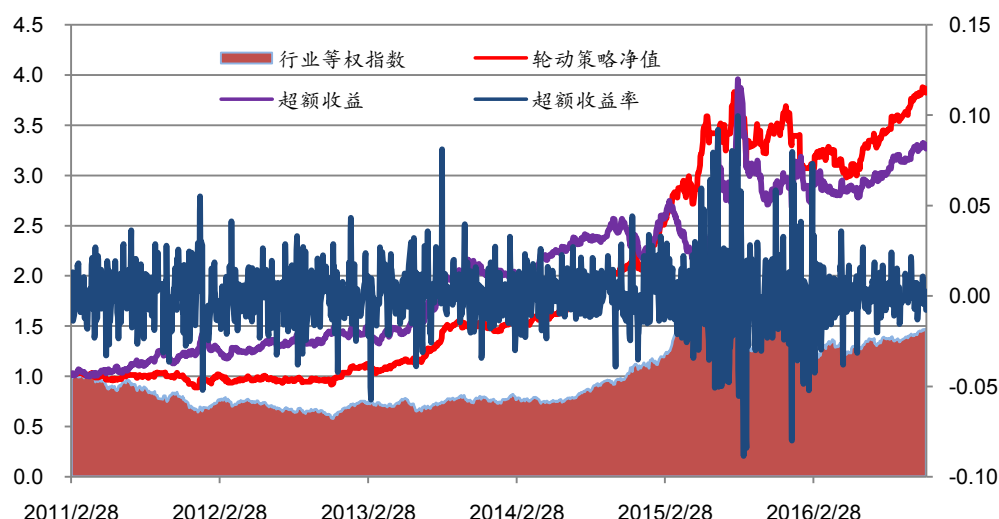
3.4 策略改进

前面的行业轮动策略中根据相关性分析结果（各行业的滞后5阶舆情数据和行情数据有很强的正相关关系），以两者存在很强的正相关关系为前提，在历史区间内具有良好的表现。但是这种正相关关系是否在每个交易区间都存在？若不存在，是否会影响策略表现，这里我们对其进行讨论。

首先对舆情数据与行情数据的相关关系进行判断，原有策略的其他设定不变。具体来说，根据过去一定时期内的舆情指数与对应行情指数的相关系数方向，估计未来一周两者的相关系数方向。若为正相关，则根据过去一周舆情指数以及对应的行业指数的涨跌幅，当某个行业满足舆情指数涨幅大于阈值A，对应的行业指数涨幅大于阈值B且小于阈值C时，发出买入信号，在本周的第一个交易日以持有资金等权买入满足条件的行业，固定持有一段时间后平仓。

同样地，以2011年到2016年为实证区间进行分析，得到策略实证结果如下：

图23：改进的行业轮动策略历史表现



数据来源：广发证券发展研究中心

表13：改进的行业轮动策略历史表现一览

实证区间	累计净值	年化收益率	年化超额收益率	胜率
2011-2016	3.82	25.55%	22.21%	54.95%

数据来源：广发证券发展研究中心

改进的策略自2011年以来，累计净值达3.82，实现了约为25.55%的年化收益率；相对于行业等权指数的年化超额收益率为22.21%，策略胜率为54.95%。无论是收益率还是胜率，均高于前面原策略的回测结果，从总体来看，改进的策略表现优于原策略。

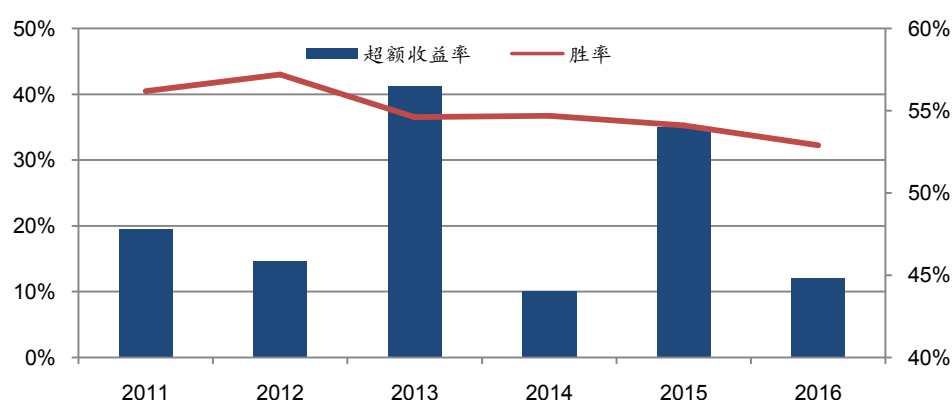
改进的策略分年度表现如下表所示：自2011年以来，策略获得了226.63%的累计超额收益率；所有年份的超额收益率均为正，2013年和2015年的超额收益率均超过30%，从胜率来看，所有年份胜率均大于50%，除2014年外，其余年份均大于50%。

表14：改进的行业轮动策略分年度表现一览

时间	超额收益率	胜率
全样本	226.63%	54.95%
2011	19.48%	56.19%
2012	14.63%	57.20%
2013	41.32%	54.62%
2014	10.15%	54.69%
2015	35.08%	54.10%
2016	12.00%	52.91%

数据来源：广发证券发展研究中心

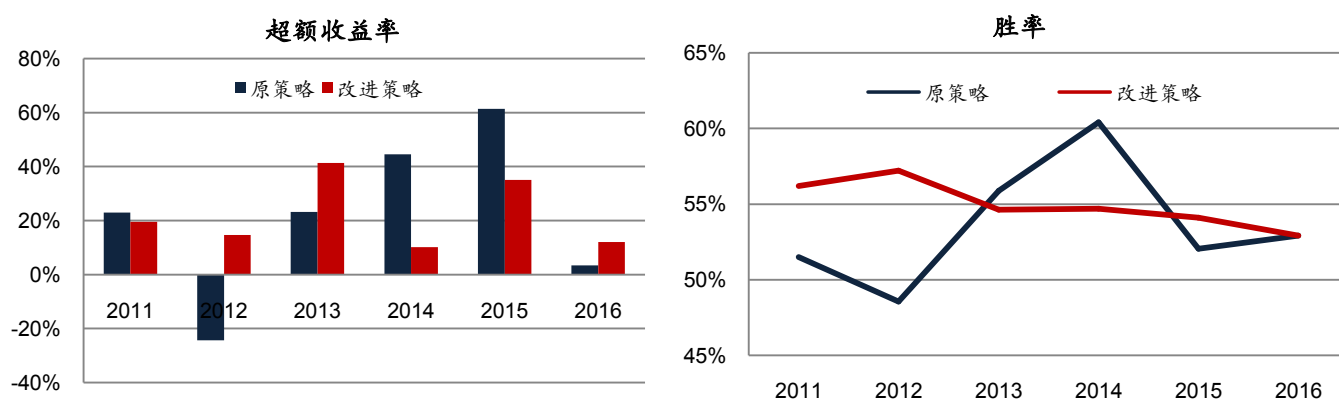
图24：改进的行业轮动策略分年度表现



数据来源：广发证券发展研究中心

将改进的策略与原策略的分年度表现进行比较，见下图。从超额收益率来看，原策略分年度超额收益率波动较大，2012年出现负值；改进的策略分年度超额收益率均为正，2012年、2013年和2016年改进策略的高于原策略，而原策略在2014年、2015年具有较高的超额收益率。从胜率看，原策略不同年份之间差异较大，2012年低于50%；除2013和2014年外，改进的策略胜率高于原策略；并且改进的策略胜率较为稳定。

图25：原策略与改进策略分年度表现比较



数据来源：广发证券发展研究中心

四、总结

4.1 总结

基于舆情的行业轮动策略根据舆情变化和行情变化有较强的正相关关系，且舆情变化会领先于行情变化构建了量化择时策略，实证结果表明基于舆情数据的行业轮动策略在历史区间内具有良好表现。策略要点：

1、行业舆情能够很好地代表了舆情信息。利用网络爬虫程序抓取了各个行业相关的历史舆情数据，从Wind提取了行业指数的历史数据用于策略构建。

2、策略中对各行业的舆情指数和行业指数涨幅设置阈值，当舆情指数涨幅和行业指数涨幅满足阈值条件时，发出看多信号，买入行业指数，固定持有一段时间后平仓。对阈值参数组合进行敏感性测试选择出了最优参数组合，同时发现策略对舆情指数涨幅阈值A变动最为敏感，行业指数涨幅阈值B、C次之。

3、策略中根据相关性分析结果，以舆情指数与行情指数存在较强的正相关性为前提，考虑到并非每个区间均存在这种相关关系，在原策略基础上加入相关性判断，得到的改进策略整体上优于原策略。

4.2 未来研究方向

以一周为时间区间判断舆情和指数行情是否出现上涨，选择出行业后持仓一段时间，导致策略中可能每周会有几天的空仓日期。未来可以改进策略，在空仓日期再做相同判断，选择性买入指数，减少空仓日期。

风险提示

本报告提出的行业轮动策略基于行业指数舆情数据，舆情是投资者投资情绪的一个方面表现，市场行情受到各种综合因素影响，注意模型失效的风险。

广发证券—行业投资评级说明

- 买入： 预期未来 12 个月内，股价表现强于大盘 10%以上。
- 持有： 预期未来 12 个月内，股价相对大盘的变动幅度介于-10%~+10%。
- 卖出： 预期未来 12 个月内，股价表现弱于大盘 10%以上。

广发证券—公司投资评级说明

- 买入： 预期未来 12 个月内，股价表现强于大盘 15%以上。
- 谨慎增持： 预期未来 12 个月内，股价表现强于大盘 5%-15%。
- 持有： 预期未来 12 个月内，股价相对大盘的变动幅度介于-5%~+5%。
- 卖出： 预期未来 12 个月内，股价表现弱于大盘 5%以上。

联系我们

	广州市	深圳市	北京市	上海市
地址	广州市天河区林和西路9号耀中广场A座1401	深圳福田区益田路6001号太平金融大厦31楼	北京市西城区月坛北街2号月坛大厦18层	上海市浦东新区富城路99号震旦大厦18楼
邮政编码	510620	518000	100045	200120
客服邮箱	gfyf@gf.com.cn			
服务热线				

免责声明

广发证券股份有限公司具备证券投资咨询业务资格。本报告只发送给广发证券重点客户，不对外公开发布。

本报告所载资料的来源及观点的出处皆被广发证券股份有限公司认为可靠，但广发证券不对其准确性或完整性做出任何保证。报告内容仅供参考，报告中的信息或所表达观点不构成所涉证券买卖的出价或询价。广发证券不对因使用本报告的内容而引致的损失承担任何责任，除非法律法规有明确规定。客户不应以本报告取代其独立判断或仅根据本报告做出决策。

广发证券可发出其它与本报告所载信息不一致及有不同结论的报告。本报告反映研究人员的不同观点、见解及分析方法，并不代表广发证券或其附属机构的立场。报告所载资料、意见及推测仅反映研究人员于发出本报告当日的判断，可随时更改且不予通告。

本报告旨在发送给广发证券的特定客户及其它专业人士。未经广发证券事先书面许可，任何机构或个人不得以任何形式翻版、复制、刊登、转载和引用，否则由此造成的一切不良后果及法律责任由私自翻版、复制、刊登、转载和引用者承担。