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）

报告摘要:

- **近年来，概念、热点炒作频繁**

近年来，A 股市场上概念、热点的炒作频繁，例如最近的“沪港通”、“国企改革”、“互联网金融”、“一带一路”等，相关个股在某一概念、热点背景下，往往在中短期时间内有较好的市场表现。如何抓住这种热点、概念带来的投资机会，是当前量化研究的一个难点。本专题报告试图从大规模的互联网新闻的文本信息作为切入点，采用文本挖掘的方法研究热点、概念带来的投资机会。

- **基于互联网挖掘的热点选股策略构建**

基于大规模的热门网站关于个股的新闻文本信息，本专题策略通过动态地构建热词词库以及关于热词的热度指标，采用文本挖掘方法对热词的变化进行了量化。基于量化后关于热词热度的变化指标，我们进行策略的构建。

互联网挖掘的热点选股策略构建：在历史回测期间内，将资金等分为若干份。动态地监控在每个交易日内是否有热点概念出现，如果有热点概念出现，则将其中的一份没有持仓个股的资金配置当天出现的热点概念的个股，持有一段时间，不考虑个股在买进或者卖出时候出现涨停、跌停或者停牌的情况，个股在买进时以该个股当日的开盘价买进，卖出时以该个股在持有期末的收盘价卖出。

- **实证结果**

历史回测结果显示，热点选股策略在取得了较优异的结果。相对沪深 300 指数，在整个回测期内，实现了约 19.00%的年化超额收益率以及 112.08%的年化绝对收益率；在胜率方面胜率，取得了 65.85%的周胜率以及 81.82%的月度胜率；从回撤角度看，在没有用沪深 300 进行对冲前，策略累计净值的回撤为 6.27%，在用沪深 300 进行对冲后，策略对冲后的净值的回撤为 3.59%。整体上而言，基于对大规模的热门网站的文本信息挖掘构建的热点选股策略，在历史回测期内取得相对较为优异的表现。

- **风险提示**

本模型为采用纯量化方法，所推荐的个股未必具有实质性的利好，其股价表现还受到诸多因素影响，请结合基本面及自身判断进行恰当使用。

图 1 热点选股策略历史回测表现

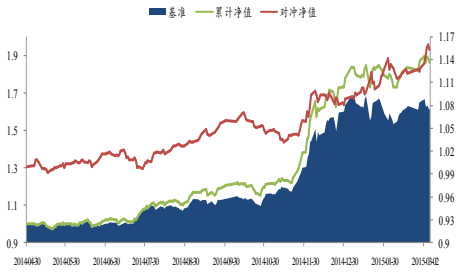


表 1 热点选股策略表现指标

指标	数值
年化绝对收益率	112.08%
年化相对收益率	19.00%
最大回撤	-6.27%
周胜率	65.85%
月胜率	81.82%

分析师: 史庆盛 S0260513070004
020-87555888-8618
sqs@gf.com.cn

相关研究:

基于网络新闻热度的择时策略—互联网大数据挖掘系列专题之（一）	2014-06-25
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）	2014-06-26
倾听股吧之声-互联网大数据挖掘系列专题之（三）	2014-06-27
那些年一起追过的财经小编策略-互联网大数据挖掘系列专题之（四）	2014-08-22
上市公司披露信息变更隐含的投资机会—事件驱动策略之(十四)>>	2014-12-26

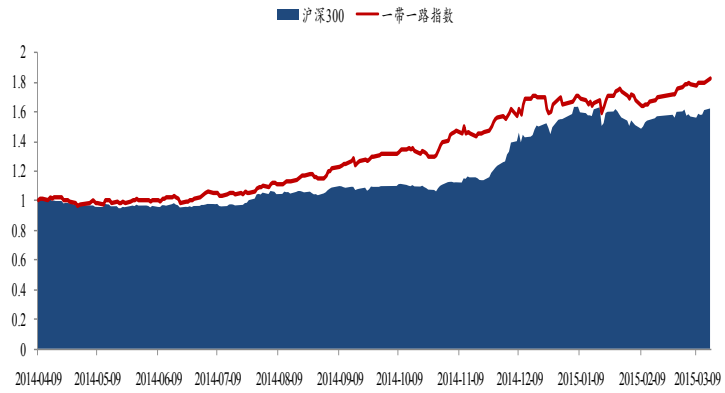
目录索引

图表目录

一、 前言

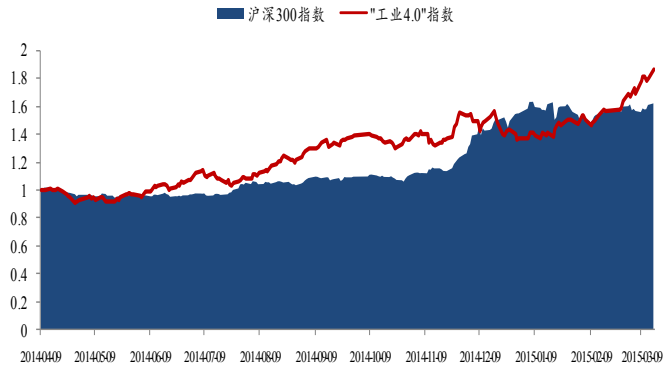
近年来，A 股市场上概念、热点的炒作频繁，例如最近的“沪港通”、“国企改革”、“互联网金融”、“一带一路”等，相关个股在某一概念、热点背景下，往往在中短期时间内有较好的市场表现。热点、概念的形成往往是因为某一类信息，例如国家政策方针、龙头企业的相关产品信息在投资者之间互相传播，演进，达成共识最后带动相关标的的上涨。这种热点、概念的兴起带来的投资机会在往往在中短期内的市场表现比市场上的整体表现要优异得多，而如何研究这种热点、概念变化带来的投资机会是本专题策略研究的出发点。

图 2 "一带一路"概念指数走势



数据来源：广发证券发展研究中心、Wind

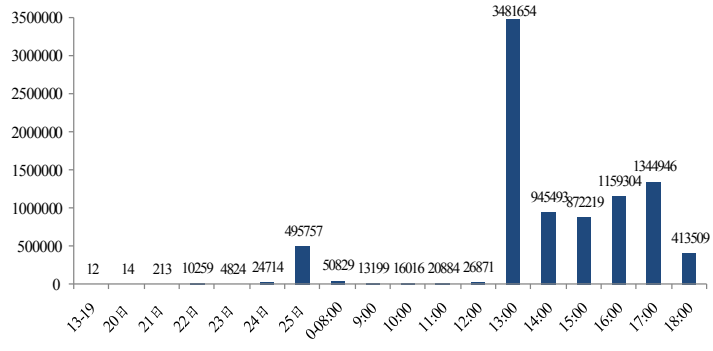
图 3 "工业 4.0"概念指数走势



数据来源：广发证券发展研究中心、Wind

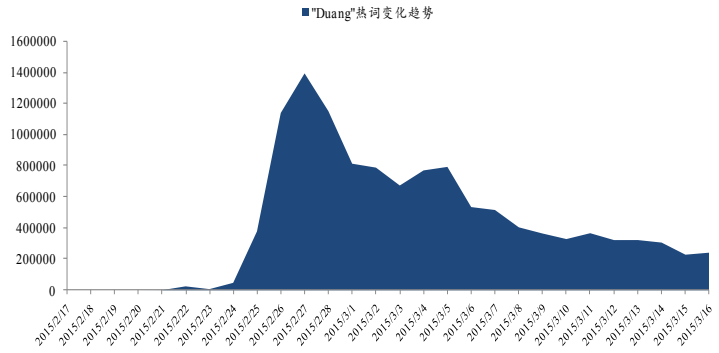
在当前互联网大数据时代下，信息的传播渠道已经越来越丰富，微博、微信、热门财经网站等已经成为了信息传播的主要途径。举个例子而言，最近兴起的一个热点词汇--"Duang"。"Duang"这个词好像一夜之间，其实就是一夜之间，"Duang"这个词火了，火得一塌糊涂，这个热词的形成毫无疑问是因为互联网的传播。简要讲述一下这个热词的形成过程，"Duang"这个词最初是来自一个网友在 Bilibili 上传的一段关于成龙代言的一段广告视频，视频比较搞笑，同时涉及成龙，又是在春节期间，因此就慢慢地就在网友之间传播起来了，其中在 2 月 26 日早上 9:00 至 19:00 这段时间内在新浪微博中被提及 826 万次，也即是说每秒平均被提及 229 次。从中可以看到互联网的传播对于热点、概念的形成起到了至关重要的作用。因此在针对研究热点、概念的产生、形成变化过程，可以从互联网上大规模的新闻文本信息作为切入点，而如何获取到大规模的关于个股相关资讯的新闻文本，并从这些文本信息中挖掘出当前可能的一些热点、概念，并识别出对应的标的带来的投资机会则是本专题策略研究的难点。

图 4 "Duang"关键字提及的微博数变化趋势



数据来源：广发证券发展研究中心、数据化管理

图 5 "Duang"热词变化趋势图



数据来源：广发证券发展研究中心、新浪

对大规模的互联网新闻文本信息的挖掘与研究，我们广发金工在该领域进行了比较深入的研究，在专题策略上也取得了一系列的研究成果。例如之前采用文本挖掘的方法对上市公司公告披露背后的投资机会进行了统计分析以及实证，得到了较好的就结果，具体可见《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）》专题报告；从股吧、个股的新闻热度、上市公司信息变更、财经频道的荐股信息等角度对文本信息进行挖掘，具体可见《基于网络新闻热度的择时策略—互联网大数据挖掘系列专题(一)》、《倾听股吧之声，洞察大盘趋势—互联网大数据挖掘系列专题(三)》、《那些年一起追过的财经小编选股策略—互联网财经频道文本挖掘策略》、《上市公司披露信息变更隐含的投资机会—事件驱动策略之(十四)》等相关的专题策略报告。

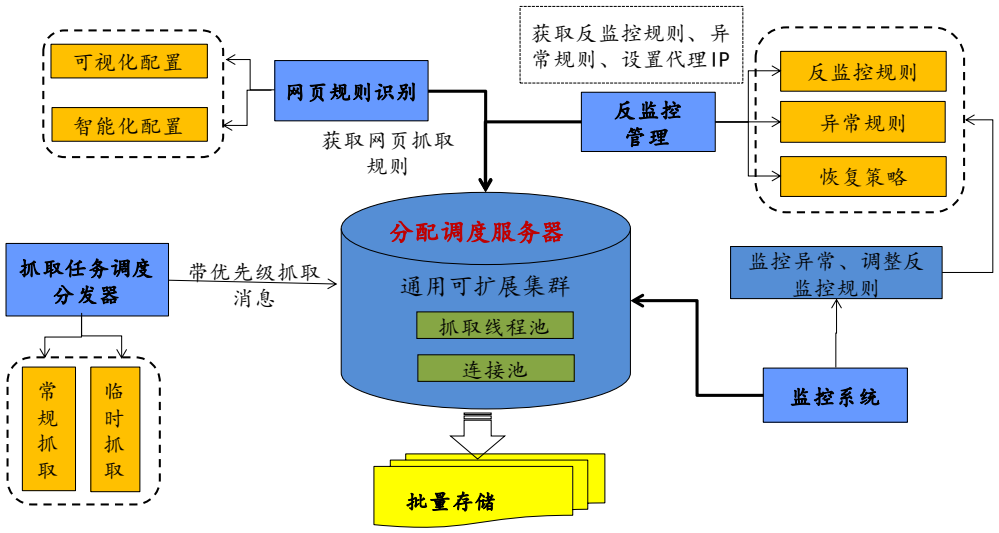
二、 文本挖掘框架流程

2.1 数据抓取平台组件框架

在之前的互联网大数据挖掘的专题系列报告中，我们广发金工搭建了一个完善的数据抓取平台框架。这里我们首先对整个数据抓取的平台框架进行详细地介绍，然后再介绍基于此平台上搭建的此次专题策略的框架流程。

首先，介绍一下我们金融工程小组搭建的完善的数据抓取平台框架。

图 6 数据抓取平台框架组件



数据来源：广发证券发展研究中心

上图刻画了我们金融工程小组搭建的完善的数据抓取平台组件的整个框架流程。**整个抓取平台的平台核心部分是中间的分配调度服务器，辅助搭建的模块有四个，分别为抓取任务调度器、网页规则识别、反监控系统以及监控系统。**

分配调度服务器功能为负责将所需数据从互联网上抓取下来，然后以指定形式的格式批量存储下来，这里可以将记录以特定的标题格式以 txt 的形式存储于本地或者以数据库的形式存储相关的文本信息。如果是以新闻文本的标题作为存储格式的一部分的话，对于在 Windows 系统下注意 txt 文件名中的一些非法字符的处理以及由于网页本身代码的特殊性，导致的一些非法字符的处理。抓取任务调度分发器负责分配抓取的网站的调度，包括一些常规的抓取以及一些临时的抓取(链接失效、断网重新抓取等任务)调度、网页规则识别负责根据抓取任务调度器分配到的网站，调用指定网站的网页内容获取规则。监控系统负责监控网页的异常（例如链接失效、断网、网页加载过慢等情况）、反监控系统负责监控网站的防盗链等问题(例如 IP 频繁访问造成的 IP 被禁等问题）。

2.2 大规模互联网新闻获取介绍

这里简要介绍一下，本专题策略对大规模互联网信息文本信息的抓取过程。因为本专题策略目的是研究热点、概念变化带来的投资机会，因此在数据获取上应该尽可能地获取到市场上关于个股的所有的新闻文本信息。然而，“理想是美好的，现实是残酷的”，获取互联网上所有关于个股的新闻文本信息，无论是在硬件条件以及效率上来说都是不现实的。由于热点、概念的传播、形成往往是通过一些比较热门的财经网站渠道，因为这些热门财经网站上每天都有大量的浏览用户，这些信息在这些浏览用户之间传播后，逐渐地形成共识，进而产生热点、概念。因此，在数据源上可以选择一些用户浏览量大的热门财经网站的，利用这些热门网站的新闻热点作为数据源切入点。**本专题策略的数据源分别选择了新浪网、和讯网、腾讯网、东方财富网这四个热门网站上的关于个股的新闻资讯，这几个网站上据我们广发金**

融工程小组对相关资讯栏目的统计，每一天关于全市场的个股新闻平均总共有 15000 条左右的新闻量。

以下以新浪网为例，简要介绍一下个股新闻资讯爬取的主要过程。

1) 首先获取到关于个股资讯的网址：
http://vip.stock.finance.sina.com.cn/corp/view/vCB_AllNewsStock.php?symbol=sz000001&Page=1;

2) 从新浪的个股资讯网站上可以查看到每天每个个股的相关的新闻的数据，而且从这个个股资讯网站上可以查看到该个股历史上所有的关于该个股的新闻的文本信息数据。

3) 通过2)中的解析后有两种方法可以获取到个股指定日期的新闻资讯，一种为通过每次查询后的结果获取到的结果URL，然后比较不同URL之间的固定参数以及可变参数的关系，找到URL之间的规律，另一种为通过模拟鼠标操作，例如翻页的操作等来获取到个股的相关资讯；
以第一种方法为例，解析查询后获取的网址的规律。
通过对网页编码的解析，我们可以获取到每次查询后的网址的实际网址信息，例如想获取到平安银行(000001.SZ)的个股资讯，可以看到实际的网址就是：
http://vip.stock.finance.sina.com.cn/corp/view/vCB_AllNewsStock.php?symbol=sz000001&Page=1，而如果在翻页操作后，例如翻到第二页，可以看到网址为：
http://vip.stock.finance.sina.com.cn/corp/view/vCB_AllNewsStock.php?symbol=sz000001&Page=2，因此可以看到，对于单个个股资讯的获取，可以通过page参数的变化来获取到历史上每一天的个股资讯新闻文本信息，而对于不同的个股，例如万科A(000002.SZ),个股资讯的网址为：
http://vip.stock.finance.sina.com.cn/corp/view/vCB_AllNewsStock.php?symbol=sz000002&Page=1，对不同个股资讯的查询可以通过symbol参数的变化来获取到不同个股历史上所有的新闻文本信息。通过以上的解析，可以知道这个URL的规律如下：
固定参数部分：
http://vip.stock.finance.sina.com.cn/corp/view/vCB_AllNewsStock.php
可变参数部分：symbol、Page
其中Page表示如果查询结果返回的是多页情况，可以通过Page的变化获取到不同页面的结果，symbol指的是不同的个股的代码，格式为："上市交易所(sz or sh)+个股上市代码"

4) 通过3)中对查询网址的规律的查询，我们可以根据URL的规律批量获取到市场上所有的个股在历史上的所有的个股资讯信息，并将将个股资讯信息提取出来。

基于上述对新浪网个股资讯获取规则的解析，我们可以得到不同个股在历史上所有的个股资讯信息，通过对需要提取的信息的网页编码格式进行解析后，可以获取到不同的个股新闻文本信息。

图 7 新闻网个股资讯栏目

数据来源：广发证券发展研究中心、新浪网

图 8 和讯网个股资讯栏目

数据来源：广发证券发展研究中心、和讯网

图 9 个股资讯获取示例

数据来源：广发证券发展研究中心、新浪网

图 10 新闻资讯提取一览

数据来源：广发证券发展研究中心、巨潮资讯网

通过对网站网页源代码解析的方式来获取新闻文本内容，有三种方式可以获取到需要提取的新闻文本信息。

1) 字符串匹配查找的方法

通过对特定网站的网页源代码的识别，利用字符串匹配的方式获取到需要提取的文本信息，该方法需要对字符的匹配查找能力有较高的要求，缺点是往往提取出来的文本信息包含一些干扰的文本信息，即信息提取不“纯”。

2) 正则表达式匹配方法

这种方法要求使用者对正则表达式的规则使用较熟练，该方法能够迅速地匹配到所需的文本信息，并将所需的文本信息提取出来。

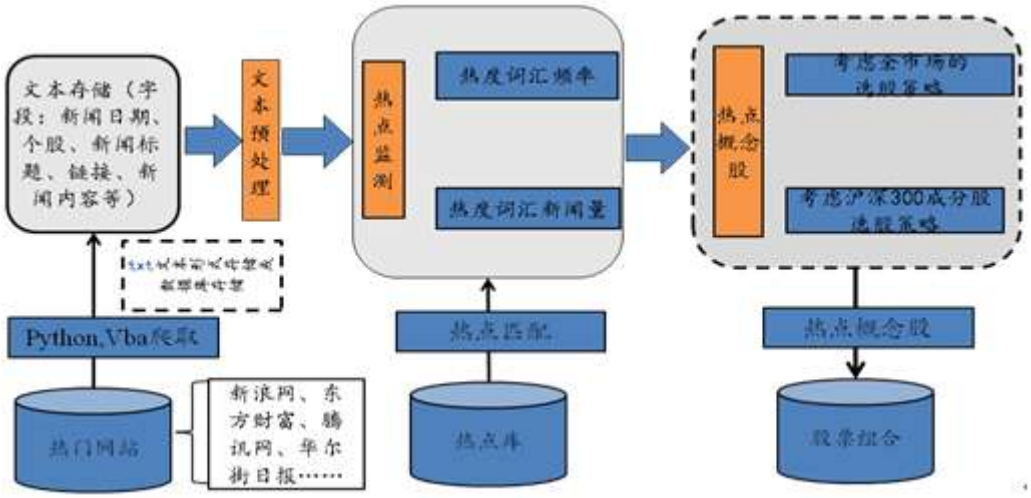
3) HTMLDOM结构解析

本质上，网站返回到本地电脑的是一个代码文件，代码文件经过浏览器的重新解析识别后，才会以各种非常规则漂亮的形式显示在我们的视野面前。网站返回的网页源代码其实是一种树状结构的，有根节点、子节点、元素、属性等。而HTMLDOM能够将这种树状结构的网页源代码存储到本地，从而能够方便快捷地利用HTMLDOM的相关语句以及语法对代码进行解析，从而获取到需要的文本信息。这对于大规模获取热门网站上的新闻的文本信息，效率上是有非常大的提高的。

2.3 基于互联网挖掘的热点选股策略框架

基于上述的完善的数据抓取平台，我们构建了此次专题策略的整体框架。

图 11 基于互联网挖掘的热点概念选股策略框架

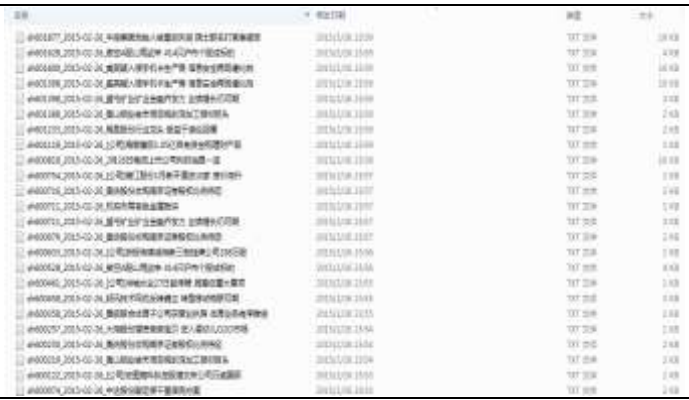


数据来源：广发证券发展研究中心

整个信息变更的事件驱动策略流程分为三大模块，分别为个股新闻数据的抓取存储、热词词库的构建以及热度指标的构建、策略构建。

个股新闻抓取与存储模块主要是利用 VBA+Python 编程从指定的热门网站上获取到所有的关于个股新闻的文本信息，并将获取到的文本信息以指定的格式存储到本地磁盘，供后续建模调用处理。

图 12 txt 存储格式:个股代码_新闻发表时间_新闻标题



数据来源：广发证券发展研究中心

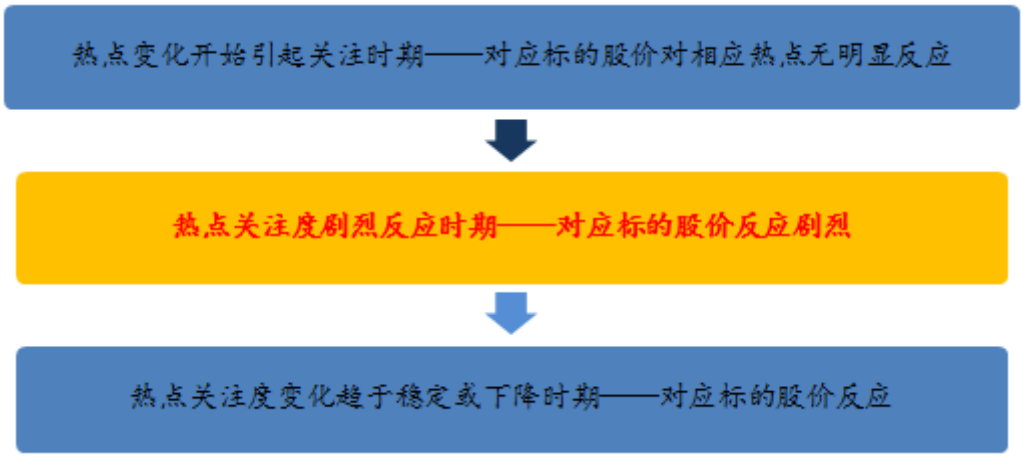
图 13 数据库存储字段:个股、时间、新闻标题、对应链接

Stock	Title	pub_date	Link
sz000063	中兴通讯创三年新高	2014/12/18	http://finance.sina.com.cn/stock/s/20141218011921102606.shtml
sz000078	海虹导航：12月18日产评报	2014/12/18	http://finance.sina.com.cn/stock/s/20141218084121103329.shtml
sz000089	上市公司公告（系列）	2014/12/18	http://finance.sina.com.cn/stock/s/20141218075921100059.shtml
sz000089	海海导航：12月18日产评报	2014/12/18	http://finance.sina.com.cn/stock/s/20141218084121103329.shtml
sz000089	深圳市机场股份有限公司关	2014/12/18	http://finance.sina.com.cn/stock/s/20141218011921102829.shtml
sz000096	截测基因有成品油价格并心	2014/12/18	http://finance.sina.com.cn/chanjing/cyyw/20141218081921104959.shtml
sz000150	北京光电股份有限公司补公	2014/12/18	http://finance.sina.com.cn/stock/s/20141218075921103986.shtml
sz000150	北京光电股份有限公司公告	2014/12/18	http://finance.sina.com.cn/stock/s/20141218075921103986.shtml
sz000150	北京光电股份有限公司复牌	2014/12/18	http://finance.sina.com.cn/stock/s/20141218011921102882.shtml
sz000150	（上接A23版）	2014/12/18	http://finance.sina.com.cn/stock/s/20141218005921102271.shtml
sz000158	海海导航：12月18日产评报	2014/12/18	http://finance.sina.com.cn/stock/s/20141218084121103329.shtml
sz000159	截测基因有成品油价格并心	2014/12/18	http://finance.sina.com.cn/chanjing/cyyw/20141218081921104959.shtml
sz000333	小米联手美的推出智能家居	2014/12/18	http://finance.sina.com.cn/stock/s/20141218023851103534.shtml
sz000401	冀东合作新方案或开业整合	2014/12/18	http://finance.sina.com.cn/stock/hyyj/20141218011921103559.shtml
sz000402	华泰柏瑞货币市场证券投资	2014/12/18	http://finance.sina.com.cn/stock/s/20141218075921100140.shtml
sz000407	山东胜利股份有限公司董事	2014/12/18	http://finance.sina.com.cn/stock/s/20141218075921100082.shtml
sz000407	海海导航：12月18日产评报	2014/12/18	http://finance.sina.com.cn/stock/s/20141218084121103329.shtml
sz000418	首届机器人大会临近 三五互	2014/12/18	http://finance.sina.com.cn/stock/hyyj/20141218024121103498.shtml
sz000426	上市公司公告（系列）	2014/12/18	http://finance.sina.com.cn/stock/s/20141218075921100062.shtml
sz000426	海海导航：12月18日产评报	2014/12/18	http://finance.sina.com.cn/stock/s/20141218084121103329.shtml
sz000426	内蒙古兴业矿业股份有限公	2014/12/18	http://finance.sina.com.cn/stock/s/20141218011921102789.shtml

数据来源：广发证券发展研究中心

热词词库的构建以及热度指标的构建模块整体上来说主要分类为两大部分，一部分是热词词库的构建，在热词词库构建方面，根据我们广发金工在文本挖掘领域的深入的研究，已经积累了大量的历史上相关的热点、概念词库，同时也基于每天的市场行情的变化，动态地往热词词库中添加新的可能的热点词汇或者删除部分已经失效的热点词汇，从而基于热词词库来动态地研究每一天可能出现的热点、概念，并通过建模来研究这种热点、概念对应标的的表现情况。热词、概念的热度指标的构建，可以通过扫描热词词库中每一个热词在指定的日期内的新闻量，如果出现的新闻量或者是热词的频率跟过去的一段时间内相比，有一个继续上升的趋势，并且上升的速率超过给定的阈值，则说明该热词可能成为当前的一个热点，相应的标的则可能存在的上升机会。

图 14 热点概念传播与对应的股价关系



数据来源：广发证券发展研究中心

图 15 热词词库示例

热点词库示例				
一带一路	人脑工程	海洋执法	人工角膜	安防服务
沪港通概念	基金三方销售	沪港通行情软件	易信	智能表
奶牛养殖	云停车场	土豆涨价	病毒检测芯片	上海金改
江北新区	碳海绵	GEP	航空租赁	百度金融
辉钼	新疆旅游	翡翠涨价	新疆电源项目	上海国资委重组
重庆水资源	五一出境游	聚四氟乙烯	浦东前滩	硅铁
海藻炼油	参股申万	殡葬	海上丝路	低碳经济
沪新通	铂矿	沪港通结算银行	京津冀	黄金概念
大麻概念	临近空间	四大行	粤港澳	燃料电池
帕拉米韦	烟花爆竹	长三角	次新股	水利建设
疫苗存储	老八股	珠三角	含H股	电子支付
农业综改	参股外资金融	武汉规划	含可转债	海水淡化
参股银河证	上海崇明岛	丝绸之路	军工航天	抗流感
维生素	智能交通	信息安全	免疫医疗	智能机器
文化振兴	空气治理	电商	特岗	央企改革
节能	污水治理	网贷	钛金属	全息概念
金融改革	固废治理	油气改革	海外工程	绿色照明
农村金融	装饰园林	苹果	赛马	高端设备
宽带提速	体育	生物智能	聚氨酯	菠菜

数据来源：广发证券发展研究中心

策略构建模块主要是根据通过动态地监控热词词库中每个热词当前时刻的热度值，然后根据当前时刻的热度值跟该热词在过去一段时间的热度值进行比较，如果当前的热词相比过去一段时间的热度值的变化率超过一定的阈值则将该热词选定为

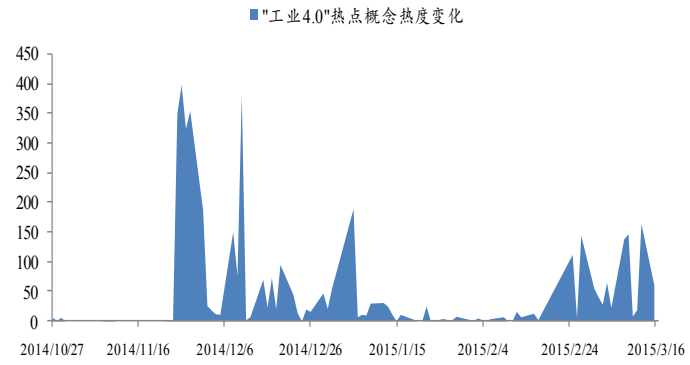
当前时点的热点，并对该热词对应的标的进行组合的配置，持有一段时间，在持有期末卖出的个股。

图 16 热点概念热度指标构建步骤

数据来源：广发证券发展研究中心

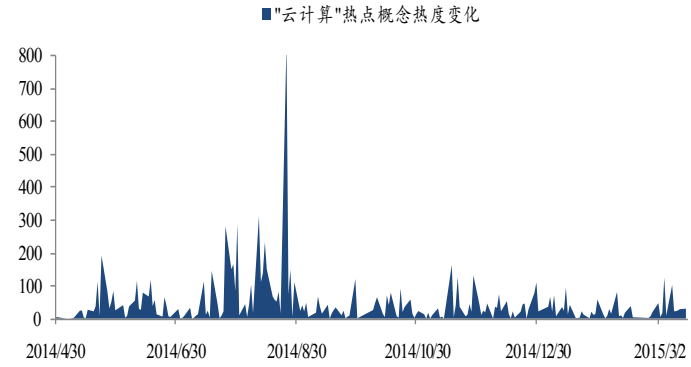
热点概念热度计算过程中，首先对抓取下来的新闻文本信息进行预处理：每个新闻对应的网站进行标注，对在之前抓取下来内容为空的新闻对应的链接进行重新抓取；其次对热词词库中的每一个热词计算热度指标，并将其存储到对应的热度词库矩阵中；最后根据热度词库中的每个热度的变化识别出每一个交易日对应的热点概念标的，然后配置对应的标的。

图 17 "工业 4.0"热点概念热度变化



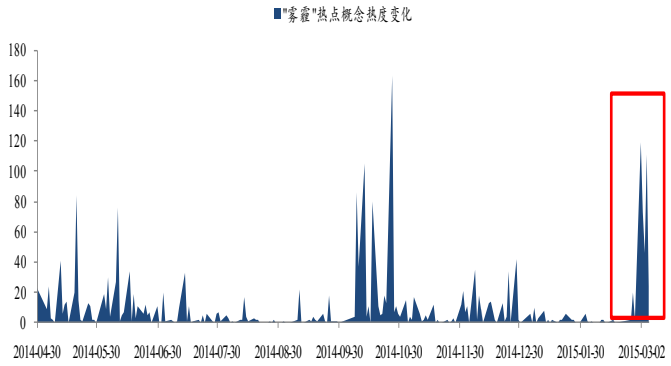
数据来源：广发证券发展研究中心

图 18 "云计算"热点概念热度变化



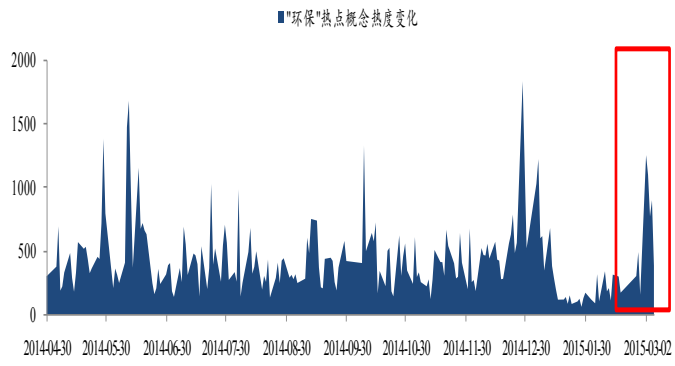
数据来源：广发证券发展研究中心

图 19 "雾霾"热点概念热度变化



数据来源：广发证券发展研究中心

图 20 "环保"热点概念热度变化

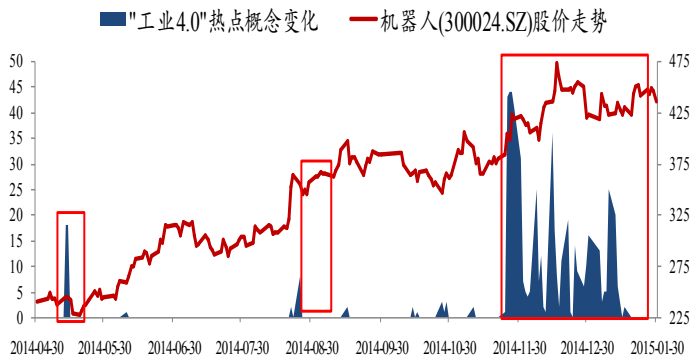


数据来源：广发证券发展研究中心

从图 19、图 20 可以看出，近期由于柴静的记录片“穹顶之下”中关于雾霾的报道，使得人们对“雾霾”、“环保”等的关注度在短期内就呈现一个剧烈上升的趋

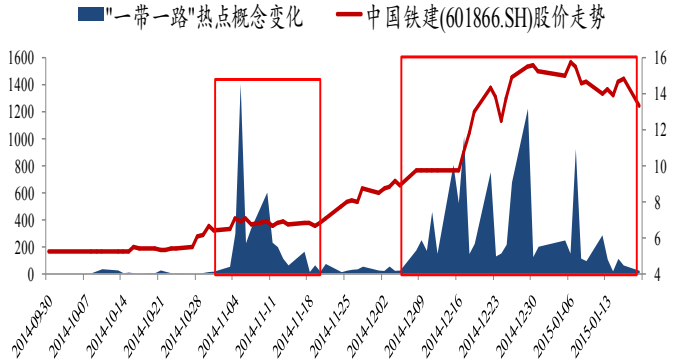
势，而相应的个股在短期内也可能比同期的大盘表现要更加地优异。

图 21 热点变化与对应标的股价变化示例 1



数据来源：广发证券发展研究中心

图 22 热点变化与对应标的股价变化示例 2



数据来源：广发证券发展研究中心

从图 21、22 可以看出，每当对应的热点改变呈现比较剧烈的一个上升趋势时，相对应的个股也往往会表现出一个上涨的趋势。

三、 基于互联网挖掘的热点选股策略实证

3.1 样本数据

由上述的策略框架可得，本专题策略的数据主要来源于新浪网、和讯网、腾讯网、东方财富网这四个热门网站。所需的数据包括热门网站上关于个股新闻资讯的历史文本信息以及 A 股市场上所有个股的历史交易数据，包括开盘价、收盘价、成交量、交易状态等数据。本专题策略选择实证分析的样本区间为 2014 年 4 月 30 日至今。策略的历史收益比较基准为沪深 300 指数。

3.2 策略构建原理

初始资金： 1;

策略原理： 在历史回测期间内，将资金等分为若干份。动态地监控在每个交易日内是否有热点概念出现，如果有热点概念出现，则将其中的一份没有持仓个股的资金配置当天出现的热点概念的个股，持有一段时间，不考虑个股在买进或者卖出时候出现涨停、跌停或者停牌的情况，个股在买进时以该个股当日的开盘价买进，卖出时以该个股在持有期末的收盘价卖出。

交易费用： 千分之二费用，在卖出时候收取。

资金投资权重： 等权投资于个股;

3.3 互联网挖掘的热点选股策略

基于以上的交易策略原理，将策略应用于沪深300指数中，具体的交易策略原理为：初始资金为1，在历史回测期内，将资金等分为10份，通过扫描每一日热词词库

中每个热词的热度，将该日热词的热度与过去一段交易日内该热词的热度进行比较，选择相对增长率最大的前两个热点作为当期的热点概念，并用没有配置个股的一份资金配置该热点所对应的个股，持有期为10个交易日，资金等权投资于该个股。个股在买进时以买进日的开盘价买进，在持有期末以收盘价卖出。

图 23 热点选股策略表现(相对沪深 300)

数据来源：广发证券发展研究中心

表 2 热点选股策略表现指标一览

指标	数据
年化绝对收益率	112.08%
年化相对收益率	19.00%
最大回撤	-6.27%
周胜率	65.85%
月胜率	81.82%

数据来源：广发证券发展研究中心

从上述图与表可以看出，热点选股策略在历史回测期内取得了较优异的结果。相对沪深 300 指数，在整个回测期内，实现了约 19.00%的年化超额收益率以及 112.08%的年化绝对收益率；在胜率方面胜率，取得了 65.85%的周胜率以及 81.82%的月度胜率；从回撤角度看，在没有用沪深 300 进行对冲前，策略累计净值的回撤为 6.27%，在用沪深 300 进行对冲后，策略对冲后的净值的回撤为 3.59%。整体上而言，基于对大规模的热门网站的文本信息挖掘构建的热点选股策略，在历史回测期内取得相对较为优异的表现。

四、 总结

4.1 总结

基于大规模的互联网财经文本信息，本专题策略通过采用文本挖掘的方法研究

这些文本信息中所包含的热点、概念变化带来的投资机会。通过构建热词的热度变化指标，建立量化投资策略，考虑了热点、概念变化带来的投资机会。

通过所构建的量化策略，对热点、概念变化所引起的相应的个股的价格的变化进行了历史回测，实证结果表明，所构建的量化策略在回测期间内，表现优异，主要结论有：

- 本专题对互联网大规模的新闻数据的文本信息进行了量化处理，建立了基于热点概念识别的文本挖掘选股策略；
- 基于所建立的量化策略，可以动态地识别出每日可能的热点、概念并筛选出对应个股；

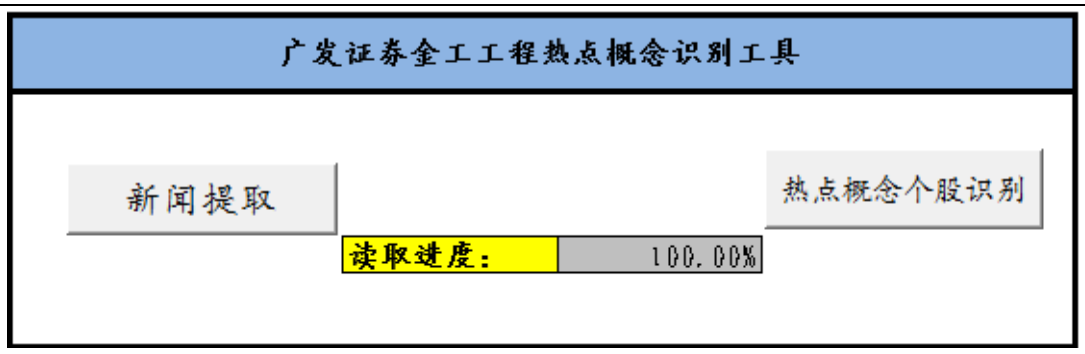
4.2 未来研究方向

- 动态更新并维护热点概念词库；
- 研究基于全市场的热点概念文本挖掘选股策略；
- 基于互联网大规模的新闻数据，每日定时推送相应的宏观热点(国内、国外)、热点概念以及对应的个股；
- 个性化需求定制；

4.3 工具推荐

基于对大规模的互联网热门网站上的个股信息相关网页的网页格式的研究，我们开发出了基于互联网热门网站的热点概念识别工具。这个工具能根据用户对不同网站的选择或者是基于所跟踪的所有的热门网站，根据本专题策略关于热点概念构建的算法，自动地识别出当前市场上可能的热点概念，并输入对应热点概念的标的个股，同时该工具能够显示用户自定义的热点概念在历史的走势情况，以图表的形式呈现。目前，该工具已经开发接近完善，即将外发给客户在本地使用，欢迎发邮件与我们索取或交流。

图 24 广发证券金融工程热点概念识别工具



数据来源：广发证券发展研究中心

风险提示

本模型为采用纯量化方法，所推荐的个股未必具有实质性的利好，其股

价表现还受到诸多因素影响，请结合基本面及自身判断进行恰当使用。

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	广州市天河北路 183 号大都会广场 5 楼	深圳市福田区金田路 4018 号安联大厦 15 楼 A 座 03-04	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	510075	518026	100045	200120
客服邮箱	gfyf@gf.com.cn			
服务热线	020-87555888-8612			

免责声明