

新媒体舆情大数据分析 投资策略简介

合作单位：

招商证券股份有限公司

新浪网络科技股份有限公司

北京大学网络与信息系统研究所

目录

一、 策略背景3

二、 典型投资案例分析4

三、 策略合作方案9

四、 策略实施流程.....10

五、 总结与展望.....13

六、 联系方式.....14

一、策略背景

随着以微博、微信为代表的社交网络在国内的崛起，新媒体时代的讯息生成与扩散的完整传播链条发生了本质性的巨大变化。社会热点不再是自顶向下地通过门户网站、报纸、电视等传统媒体进行有选择性地层层传播，而是在覆盖在整个社会之上的扁平自媒体网络中随机地自发爆发。目前，中国的散户投资者以 20~50 岁左右的白领和自由职业者为主，而这些人也正好是互联网和移动互联网，特别是社交网络的主力使用人群。因此，社交网络不但极其敏感地反映出这部分用户的整体价值观与所关注的兴趣点，更反过来凭借其强大的用户附着力，占据了这部分用户大量的注意力资源，潜移默化地改变着他们看待世界的看法，更进一步影响到他们投资决策的理念。（参考：马歇尔·麦克卢汉：《理解媒介——论人的延伸》）

在具有无尺度网络拓扑结构的互联网社交网络中，热点的流行和消退具有十分明显的幂律分布特征——瞬间爆发，一发不可收拾；随后快速消散，消失地无影无踪；之后又在别处有新的热点被引爆，席卷全网，大众的注意力也在这样的浪潮中不断发生着转移。而在每一个时点，整体网络社群的注意力却又高度集中在屈指可数的若干个话题上。新媒体时代社会热点的转换，就如同夜空中的礼花：瞬间爆炸、网状扩散，随后缓缓消失，别处又有礼花引爆，此起彼伏。（如图一）。（参考：艾伯特-拉斯洛·巴拉巴西：《链接——网络新科学》）



图一：社交网络中热点的爆发与转移结构图

应对新媒体时代社会热点的快速转移，传统的记者采集的媒体模式已无法跟上这种大众口味的快速变化。因此，当今的新闻媒体，不再可能仅依赖记者在现实世界中的资讯采集，而必须密切关注社交网络中随时可能爆发的话题热点。网站、报纸、电视评论等等，都以不同的速度反映着之前爆发的自媒体热点讯息，使得社交网络中的流行热点在退潮后很长时间内，仍能固化在门户网站、即时通讯软件、电子邮件、券商研究报告、电视证券评论、报纸、书籍甚至是口碑中，在社会上持续传播、演变，向不同的受众群体缓缓释放着巨大而持久的影响力，改变着他们的投资决策。这种传统媒介对新媒体热点跟进式的报道，会进一步延长该热点对股市的影响时间跨度并提升其公信力，从而形成一个自我加强的循环过程。（参考：马尔科姆·格拉德威尔：《引爆点——如何制造流行》）

二、典型投资案例分析

10月28日天安门吉普自杀性恐怖袭击事件对股市的冲击

2.1 事件概述

2013年10月28日中午，中午12时05分，一辆吉普车由北京市南池子南口闯入长安街便道，由东向西行驶撞向天安门金水桥护栏后起火。据初步统计，事件导致司机和两名乘客死亡，另有2名游客死亡（1名菲律宾籍女游客、1名广东省男游客），38名受伤人员中包括3名菲律宾籍游客（2女1男）及1名日本籍男游客。

2.2 媒体反应

面对上述突发事件，新浪微博再次充当了各种媒体系统中反应速度最快的自媒体先锋。依据“新浪微博热词指数”中的统计，在事件发生后的三个小时内，有关该事件的微博在网络中迅速爆发性扩散。包含该事件相关热词“天安门”的微博转发数量，从2013年10月27日的6469条，顷刻间飙升为2013年10月28日的82268条，增长12.7倍；包含相关热词“吉普”的微博的转发数量，更是从2013年10月27日的2626条，暴涨为2013年10月28日的58316条，增长22.2倍。具体热词关注度的几何脉冲形态，请参见下图：



图二：事件相关热词“天安门”的微博的转发数量变化示意图



图三：事件相关热词“吉普”的微博的转发数量变化示意图

依据前文所分析阐述的理论，该事件在新浪微博爆发流行后，会吸引其它各种慢速的媒体以各自不同的速度跟进报道该热点事件，将该事件的影响力推向不同的投资者群体。

果然，该事件在腾讯新闻客户端、网易新闻客户端上引起了广泛关注。在有选择性大量删帖的前提下，该事件相关的新闻条目仍然积攒了超过 1.4 万条评论，创近期腾讯新闻客户端、网易新闻客户端中新闻回复数量之最。

其余包括优酷视频、人民网、中央电视台、各大报纸等传统媒体也纷纷报道该事件。而在百度的用户搜索量和媒体关注度指数中可以看出，28~29 日，各大媒体对该事件进行了广泛的报道，但整体速度均慢于新浪微博，因此并不能充当相关股票价格变动的先兆指标，仅起到了推波助澜的作用，进一步加强了该事件对证券市场的影响。



图四：“天安门+吉普”的百度搜索量和媒体关注度变化示意图

2.3 交易策略

在相关投资决策层面，基于本部门对微博热点实时监控系统的报警，我部于 2013 年 10 月 28 日下午 14: 00 盘中发现该事件在社交网络中的传播路径已具备突破引爆点的形态，极有可能引发新一轮的社会热点。因此，交易员在分析后决定做空四只“北京旅游”相关概念股票：中国国旅（601888）、中青旅（600138）、北京旅游（000802）和首旅股份（600254），并做多等 β 的沪深 300 股指期货以对冲系统性风险。结果如预期所料，29 日全天，中国国旅（601888）下跌 5. 2%；中青旅（600138）下跌 3. 33%；北京旅游（000802）下跌 1. 89%；首旅股份（600254）下跌 2. 09%。

截止 29 日收盘，做空的四只股票：中青旅（600138）、中国国旅（601888）、北京旅游（000802）、首旅股份（600258）平均跑输沪深 300 指数 4. 005%。可以明显发现这是由“吉普在天安门前自杀性爆炸，5 死 38 伤”的突发事件引起的“北京旅游概念股”集体下跌。但是由于本条新闻的敏感性，其在微博等自媒体中爆发增长之后被官方明显压制，各大门户、报纸、电视都没做过多的传播扩散，分享和评论的功能也被屏蔽。由此可预见，本次事件的影响将会在自媒体脉冲结束后被掩盖。因此，在 30 日开盘后的合适时机进行了结平仓。整体操作跨越三个交易日，总收益率为 4. 3%，效果显著。

2.4 策略总结

类似的典型案例与投资机会，如“塑化剂风波”、723 甬温线动车事故、韩亚航空空难、光大证券乌龙指事件、鹏博士暴涨、蓝色光标暴涨。等等，不胜枚举。随着移动互联网、社交网络，以及互联网金融产业的蓬勃发展，网络热点信息的爆发式流行转变为证券市场投资行为和股价变动的路径将越来越短，成本也将越来越低。因此，研发基于新媒体舆情大数据分析的投资策略大有可为。

三、策略合作方案

本策略是一个新媒体环境下的产、学、研相结合的互联网金融研究与实践项目，聚合了包括招商证券、新浪微博、北京大学、高盛、斯坦福大学等企业、高校和媒体的科研力量与业务资源。相信本项目将成为互联网金融时代证券公司、互联网公司、高校科研力量合作的经典案例。

在管理模式上，本项目建立了互联网时代的新型项目团队，以达成“不同行业，不同城市，目标一致，节奏统一，分工协作，无缝连接”的管理目标。我们将充分运用互联网和移动互联网中的社交网络，保持整个研发团队的实时同步，确保项目整体流程流畅、高效，并以扁平化的团队管理结构充分降低沟通成本、提高合作效率。

本项目组将设立项目督导三人，由分别来自招商证券、新浪微博、北京大学的领导担任；并设立项目经理一人，全面统筹项目的整体运作。在此基础上，项目组将设立三个相对独立并相互依存的团队，分别是：“产品团队”（负责项目相关互联网数据产品的设计与开发工作）、“研究团队”（负责项目相关的交易策略和计算机算法的研究工作）以及“市场团队”（负责本项目成果在投资界的推广与营销工作）。不同的企业和机构依据自身特点和比较优势的不同，在项目组中承担各自不同的职责。具体分工如下：

3.1 招商证券

招商证券衍生投资部将为本项目提供基于金融证券行业的观点和资源，在包括金融交易数据、投资模型开发、交易系统搭建等方面提供全方位支持，并有能力在项目落地后调动足够的资金实现相关的投资策略，将项目成果转化为现实生产力。

3.2 新浪微博

新浪微博负责为本项目提供基于互联网行业的观点和资源，并在获取来自资本市场第一线的数据需求后，依据研究团队所设计的交易模型开发基于微博的舆情数据接口。本项目力求最大限度地高效利用微博舆情数据，开发出优美的互联网产品，并实现团队的共同目标。

3.3 北京大学

北京大学网络与信息系统研究所在搜索引擎、社交网络等与新媒体相关的前沿科研领域具有国际领先的研究实力，负责在本项目中提供来自高校的科研力量，并努力将研究成果转化为相关领域世界顶级学术会议的论文。

四、策略实施流程

本投资策略属于“新媒体与社交网络时代的新闻事件驱动型市场中性策略”。我们将在不同的抽象层面建立起网络热点分布的实时监控模型,并在及时发现爆发的热点后经过交易员的投资业务分析将事件映射到相关股票上,从而针对不同事件建立多头与空头头寸,自然对冲系统性风险。在具体的实施过程中,我们将项目拆分为四个独立的迭代子项目,每一个子项目的需求层层递进,对市场的建模与描述也逐渐细化,从而最终达到对全媒体舆情的全面掌控。

4.1 子策略一：社会网络热词实时监控

该子项目主要在宏观层面对新浪微博热点词汇进行统一监控,暂不考虑热点的传播网络形态,而是使用由新浪微博统计整理并实时发布的“热搜度”和“热议度”指标来对事件的流行程度进行描述。我们着力于以最快的速度发现具有如下流行形态的热点事件:该事件相关词汇在很长一段时间内未曾被过度关注,却突然在短时间内爆发式流行。在关注图上,该形态的表示就是一条低位笔直的横线之后突然出现一段脉冲式的陡峭上升的关注度高峰,并在之后迅速下降,却在之后的时间时有小规模突起。因为只有这样,才满足吸引新关注人群的策略要素。我们将我们所定义的这种热词筛选条件称之为“首度流行条件”。

该子项目所需数据接口主要有如下两个:

1. **getWordFrequency (keyword,date)**
返回某一日某一个词在所有微博中出现的词频;
2. **getTopWords(date,num)**
返回某天微博中提及次数超过一定阈值的词汇列表及词频。

在最初级的系统版本中,利用上述两个通用数据接口,每日开盘前返回如下热词表:假设昨天热度为 A,前天热度为 B,大前天热度为 C,大大前天热度为 D,则返回同时满足如下条件的热词:(一、 $A > 30000$; 二、 $A/B > 10$ 或者 $A/C > 15$ 或者 $A/D > 20$)。交易员在接收到上述热词表后,会进行细致的投资决策分析,将其中可能影响股价的信息映射到最具确定性投资价值的资产组合上去,并在开盘后择时建仓。

当最初的系统版本取得阶段性效果后,我们将进一步升级为实时监控系統,在盘中实时追踪微博热点变化,以确保第一时间发现盘中爆发的热点事件——如 10 月 28 日盘中发生的“天安门吉普自杀性恐怖袭击事件”,从而确保在全市场投资者做出反应之前提前建仓完毕,以将对热点事件的反应速度优势转化为证券市场中的交易优势。

4.2 子策略二：证券市场热点概念二部图的建立与实时监控

综合来说，分析股价变化就是对各种不同来源的买卖盘未来到达情况的分析和预测。而市场中的每一只股票价格的涨落，都与一系列与之相关的概念的流行与消退有着直接或间接的联系。每一家上市公司所属的行业、地区、商业模式、资本结构类型，以及与之相关的政策法规等讯息，如果携带足够的公信力而流行，成为市场的主流观点，并广泛覆盖了投资者群体，就会改变多空双方的实力对比，推动股票价格向着买卖盘力量新的均衡位置运动，从而实现事件对市场的冲击。

以贵州茅台（600519）为例，“白酒”、“贵州”、“反腐败”、“食品安全”等概念都会长期与其股价保持直接联系。因此，建立起概念向股票映射的二部图，并对影响力加权加总，可以对各种常规概念对贵州茅台（600519）股价的影响加以定量的描述。

4.3 子策略三：社会网络拓扑结构分析与流行预测

这一阶段的系统监控将不再仅仅停留在宏观词频统计的层面，而是进入对整个新浪微博社交网络拓扑结构的微观层面，监控在社交网络中流动的各类信息的传播路径和扩散形态。对信息传播的微观结构的研究，将会使我们更加准确地发现热点流行的内在规律。

社交网络中流行的爆发，主要由三方面力量共同作用而产生：

首先，概念的流行，关键人物起到至关重要的作用。大部分人是通过少数几个人脉网络中的关键结点与整个世界相连接。是这些人说服周围的人，引领并推动着热点的转移。关键人物包括：“联络员”（“社交高手、交友广泛并不局限于特定的一个圈子、经常涉足不同的领域并乐于传播信息”）“内行”（“在某一专业领域具有大量业务知识，并且在其擅长的领域公信力被周围的同伴所认可，大家在这一领域的专业问题都倾向于向他寻求解答和帮助的人”）以及“推销员”（“有很强的说服能力和情绪感染力、具有个人魅力、善于影响他人的想法和行为的人”）等。

其次，信息本身所具有的可传播性也是引爆流行的重要因素。这包括信息吸引眼球和促使人们行动的能力，特别是信息中蕴藏的暗示和情绪，对信息的附着力有着至关重要的作用——这一点与社会心理学中的相关研究不谋而合，具体内容可参考古斯塔夫·勒庞的《乌合之众》以及埃里克·霍弗的《狂热分子》等著作。

最后，“环境威力法则”指出，不但传播的人以及传播物本身对流行很重要，流行所处的环境也对信息的传播至关重要。这其中又分解为“破窗理论”和“150法则”等影响形式。

因此，我们将在三个领域进行社会网络流行的分析，包括：1. 整体社交图谱（Social Graph）的连接关系和关键拓扑节点发现；2. 社交网络中每个节点自身属性的发现，即兴趣社群的划分与话题领袖的识别；3. 微博文本的语义分析。

4.4 子策略四：证券市场冲击力定量研究

需要明确的是，在证券投资的流行传播网络中，节点是以人为单位；但真正作用于市场时，直接的影响力权重则是每个人控制的资金和证券的量。有人手上并无头寸，但对市场具有巨大的影响力（著名分析师、政府官员等），但这种影响是间接实现的。

因此，在这个网络上其实有两种重要节点存在：首先是大量存在的联系人、内行和推销员，我们称他们为“卖方业务人员”（社会名流、著名学者、券商研究员、媒体、财经话语领袖等），他们负责推动股票的流行。而另一部分人，并非推动流行，却是手中握有大量头寸的人，他们在传播网络中也许不是关键节点，却是传播网络链接金融市场的重要通道，流行通过他们的交易传导到证券市场当中，我们称他们为“买方业务人员”（如券商自营、资管、私募、公募基金经理、游资、大散户、大量的中小散户群体等）。

而事实上，金融流行网络中的每个节点，其实都兼有买方和卖方的身份：我们不但拥有一定的头寸可以买入和卖出，我们也在人脉网络和媒体（包括自媒体和社交网络）中去传播流行，影响他人的判断。因此，每个节点都是卖方网络（流行传播的信息网络）中的节点和买方网络（金融交易市场）中的节点的复合节点。更进一步来描述，任何一个节点的买方和卖方属性都要联合考虑才更加符合实际。这是因为，持有多头头寸的人（或和持有多头头寸的人有利益捆绑的人，如被大户收买的卖方研究员等），有动机去传播和说服他人去购买该股票而不是抛售；持有空头头寸的人（或和持有空头头寸的人有利益捆绑的人，同上），有动机去传播可说服他人去抛售该股票而不是买入。因此，一个节点作为在卖方网络中对待某种流行传播的立场和行为，与它作为买方网络中的头寸方向有很大关系。

存在买方和卖方的不同，就是金融流行网络与传统信息传播网络的最大不同。对于卖方，我们要用传播学的理论去分析他们的重要性；对于买方，我们要推测他们的头寸情况以及投资理念。信息经过网络传播到买方的头脑中后，会经过他们脑中的意识形态处理（如价值投资、短线投机、量化投资等投资理念）转化为对证券市场的影响。只有在这两个层面共同考虑，我们才能够真正明白：是什么力量在推动某只股票的流行，而这种流行又是通过什么力量转化为该股票价格的变动。

五、总结与展望

在新媒体时代，网民和散户这两个群体正在加速融合，这是因为：第一，在人口结构上，互联网时代的原著民正逐步进入投资需求最旺盛的年龄阶段，成为新时代投资的主力大军，而这种趋势正在日趋加速；第二，在技术层面上，目前互联网金融相关产业蓬勃发展，这导致目前散户获得信息、执行交易的几乎全部渠道都来自于互联网。随着新一轮互联网金融的产业变革，腾讯、百度、阿里巴巴等巨头的加入，必将导致未来接入网络交易和信息获取的通道更加顺畅——用户在看到信息后直接一个链接就可以执行相应的证券市场操作，传统券商甚至将会面临淘汰。互联网用户行为的一个基本定律就是：简化一点操作流程，减少一层链接或表单，点击和执行的人数就会几何级数增长。因此可以预想，到那时，互联网信息的流行对股价的影响力将呈十倍、百倍的增强。

可以预想，在不久的将来，社会热点事件在发生后的几分钟内，即可通过智能手机等移动终端进入社交网络并爆发式传播，几小时后传遍全网，而附着在这则讯息上同时传播的，将是该讯息所关联的证券的买卖链接，直接点击下单即可完成交易。到那时，互联网热点对证券市场的影响力将更加举足轻重——谁能拥有对社会网络热点流行的实时监控和快速反应能力，谁就能在新媒体与互联网金融时代的证券投资市场中占得先机。

六、联系方式

项目督导：

刘成彦	招商证券 衍生投资部	总经理
李晓明	北京大学	校长助理、教授、博士生导师
罗盎	新浪微博 数据部	总经理

项目经理：

汪沛
招商证券衍生投资部
投资经理

wangpei3@cmschina.com.cn
(+86) 185-2088-1816

产品组负责人：

张欢
新浪微博 数据部
项目经理

zhanghuan2@staff.sina.com.cn
(+86) 186-1013-6623

研究组负责人：

黄威靖
北京大学 网络与信息系统研究所
博士在读

huangwaleking@gmail.com
(+86) 152-1095-4593

项目组成员（部分）：

方罳
斯坦福(Stanford)大学 法学院
法学博士(JD)在读

zhaofang713@gmail.com
(+86) 139-1819-9907

侯笑笑
高盛(Goldman Sachs) 资产管理部
投资经理助理

danaxiaoxiao.h@gmail.com
(+86) 152-0147-7365

俞金辉
新浪微博 微博产品运营部
产品经理

jinhui2@staff.sina.com.cn
(+86) 134-6669-5043

张腊梅
北京大学 网络与信息系统研究所
硕士在读

citlmzhang@163.com
(+86) 134-6669-5043