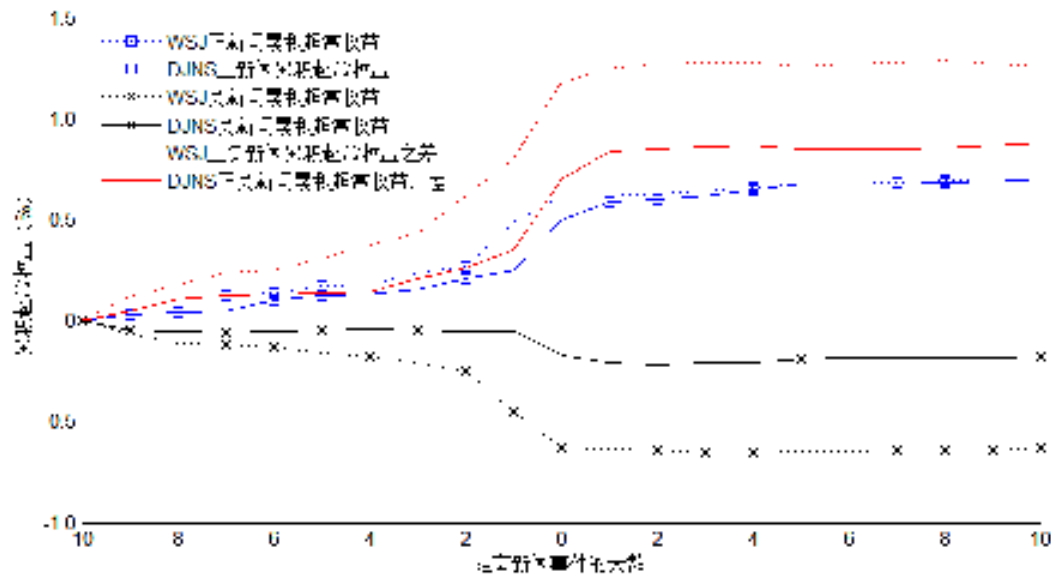


面向证券市场的智能舆情 跟踪系统一简介

2018

智能舆情跟踪系统—背景

- 新闻资讯已被广泛证实会对证券市场投资者的情绪，市场理解及预期等均会产生显著影响，从而进一步影响证券市场的资产价格波动^[1-2]。



1984至2004年S&P500个股的新闻事件研究结果^[1]

[1] Tetlock, 2007. Giving content to investor sentiment: The role of media in the stock market. Journal of Finance 62(3), 1139-1168.

[2] Barber, B.M., Odean, T., 2008. The effect of attention and news on the buying behavior of individual and institutional investors. Review of Financial Studies, 21, 785-818.

智能舆情跟踪系统一趋势

- 近年来，越来越多主流金融机构已开始布局如何结合AI和大数据等技术准确实时地收集和处理市场上的新闻并指导投资策略

重磅 | 微软首席人工智能科学家邓力离职，加盟对冲基金巨头Citadel

2017-05-19 11:58

人工智能 / 微软

华尔街再吸机器学习大牛, Pedro Domingos加入对冲基金巨头

2018-08-18 11:48

14 Active Competitions



TWO SIGMA

Two Sigma: Using News to Predict Stock Movements

Use news analytics to predict stock price performance

Featured - 3 months to go - news agencies, time series, finance, money

\$100,000

1,031 teams

智能舆情跟踪系统—现状

- 当前证券市场上的资讯分析产品的供应商等信息如下
 - ⊠ 供应商：1) Bloomberg; 2) Thomson Reuters; 3) RavenPack.
 - ⊠ 提供的数据（字段）：
 - ✓ 新闻情感评分 ✓ 新闻事件归类 ✓ 新闻热度
 - ✓ 新闻相关证券及相关性 ✓ 新闻新颖性（Novelty） ✓
- 然而，上述公司所提供的资讯分析产品服务
 - ⊠ 信息来源仅包括新闻机构所发布的新闻
 - ⊠ 只关注与海外证券市场交易标的英文资讯
 - ⊠ 未结合知识图谱做拓展推理

智能舆情跟踪系统一挑战

□ 面对日益丰富但庞杂的新闻资讯，智能舆情跟踪系统目前仍面临以下挑战

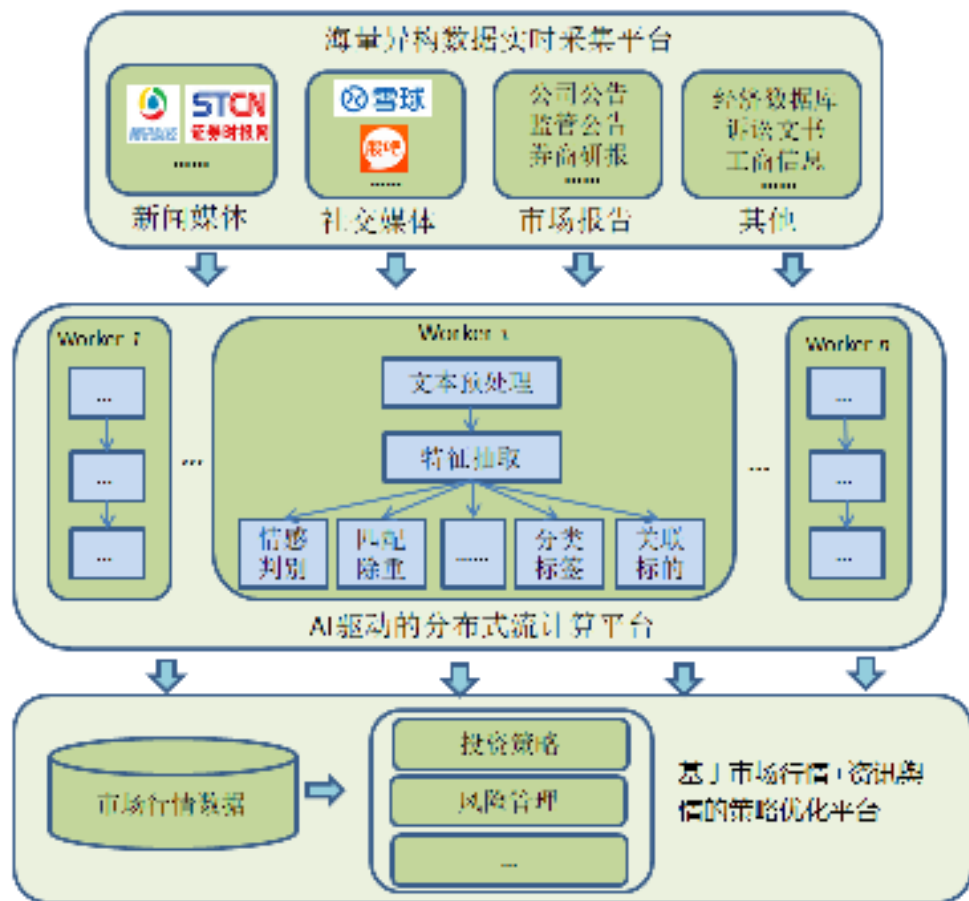
■ 海量实时异构信息源（文本, PDF, HTML, 图片, 语音, 视频）



智能舆情跟踪系统—基础架构

□ 一个基本的智能舆情跟踪系统通常包括：

- ▣ 海量异构数据采购平台
- ▣ AI驱动的分布式流计算平台
- ▣ 基于市场行情+舆情的策略优化平台



智能舆情跟踪系统—资讯采集清理

□ 面向异构资讯源及载体的实时采集平台：

- ⊠ 实时稳定的分布式爬虫
- ⊠ 面向异构文档的信息抽取：XBRL,HTML, PDF, 图片及表格等
- ⊠ 资讯除噪：去除咨询中的广告、行情评述、天气预报等噪声
- ⊠ 内容清理：繁简转换，

《苏宁“万人空巷抢彩电”大促销八月来袭》



《湖北长江电气有限公司招聘简章》



《浦发行高位企稳带大盘回升》



《苏宁海尔昨突然宣布联手组建销售公司》



“【新浪訊】香港華潤集團今年中收購深萬佳百貨，屆時兩家的年銷售額加起來將近40億元，其目標就是要進軍全國零售業的三甲。（更多報道見A5版）”



“香港华润集团今年中收购深万佳百货，届时两家的年销售额加起来将近40亿元，其目标就是要进军全国零售业的三甲。”

智能舆情跟踪系统—相关标的的分析

□ 抽取新闻的相关交易标的（比如个股和行业）以及相关度：

▣ 完整的历史标的的更名和常用名数据库+实体消歧

“东南汽车与三菱汽车合作加**深发展**自主品牌”



深发展(SZ000001)

“经济形势进入良性循环轨道 **老百姓**就业机会增多”



老百姓(SH603883)

“**老百姓**拟募集加码主业”



老百姓(SH603883)

▣ 标的相关度分析

任性停牌遭监管"祭旗" 厦华电子首家被公开谴责

7月31日晚间，上交所官网发布了对厦华电子进行公开谴责的监管函。监管函表示，厦华电子办理停牌事项不审慎.....

据东方财富Choice终端数据显示，截至7月31日，A股市场上有45家上市公司的连续停牌天数超过了100天，包括*ST新亿（600145.SH）、*ST华泽（000693.SZ）.....



厦华电子(SH600870)

↓
*ST新亿（SH600145）

||
*ST华泽（SZ000693）

↓
东方财富 (SH300059)



智能舆情跟踪系统—情感分析

- 分析新闻对相关标的（比如个股）的情感倾向：
 - ⊠ 面向金融领域的语料及模型：“跳水”：财经新闻vs体育报道
 - ⊠ 提供连续量化情感评分（如-1到1之间）：而非简单的三级评分（正面，中性和负面）

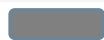
路翔股份半年净利剧减760% 沥青退出锂业难挑大梁

路翔股份今日晚间发布2014年半年报。报告期内，公司实现营业收入3.5亿元，同比减少0.37%；归属于上市公司股东的净利润为-4300万元，较上年同期剧减763.89%，基本每股收益-0.3元。

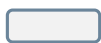
对于利润剧减，路翔股份公告称，为加快业务转型和战略转型的步伐，使公司能集中有限的资源投放到更有前景的锂业行业，公司在报告期内确定了沥青业务全面退出的转型方案。报告期内，公司先后转让了广州路翔100%股权、西安路翔100%股权、重庆路翔100%股权、北京路翔100%股权，截至报告期末，公司已经逐步完成了沥青业务退出计划，目前公司已不具备改性沥青生产能力。

同时公告称，报告期内，融达锂业因为征地问题尚未解决，导致融达锂业采选作业在冬歇期结束后尚未复工，主要是有计划销售上年度剩余库存产品，导致收入下降，报告期内销售收入为4,360,641.03元，同比下降85.03%。

资产收购方面，公告称，公司以现金1.69亿元向时代投资和陶广购买其合计持有的东莞德瑞65%股权，东莞德瑞已于2014年6月30日完成工商变更登记，成为公司持股65%的控股子公司。



负面文本



中性文本



正面文本

智能舆情跟踪系统—事件分析

□ 识别新闻所报道的具体事件类型

▣ 不断完善的事件分类体系

- ✓ 分析师预测评级
- ✓ 信用（债券）评级
- ✓ 债务危机
- ✓ 关联交易
- ✓ 法律诉讼
- ✓ 裁员
- ✓ 政府监察
- ✓ 股权质押
- ✓ 对赌风险
- ✓ 资产（股权，账户，抵押品）冻结
- ✓ 法律顾问变动
- ✓ 会计事务所变动
- ✓ 担保事项
- ✓ ...

▣ 基于海量历史数据的事件重要性分析（事件+公司/行业）

复星医药旗下公司被食药监部门调查



外汇局通报23起典型案件罚款8235万 涉交通等8家银行

智能舆情跟踪系统—关联新闻分析

□ 分析当前新闻与其他历史的相似性

- ▣ 新闻相似性分析：文本相似性 vs. 事件相似性
- ▣ 结合阅读量等构建新闻原创性及热度指标，并追踪新闻事件的发展演进等

【商报北京专讯】新组成的北京市第10届政协委员会与上届相比，又有14位港澳人士被增补为政协委员，令港澳人士总数达45人。

据介绍，新增补的14位港澳委员中，突出了专业人士和部分香港知名人士，如香港财经小说作家梁凤仪……

文本相似度
0.91



【新华社北京卅一日电】新组成的政协北京市十届委员会与九届相比，又有十四位港澳人士被新增补为北京市政协委员，十届北京市政协港澳委员总数已达四十五人。

据介绍，新增补的十四位港澳委员中，突出了专业人士和部分香港知名人士事业继承者，如香港畅销财经小说作家梁凤仪……

智能舆情跟踪系统—其他

- 针对不同资讯源的多源信息融合：不同资讯源的可信度和丰富度不一致



可信度：专业新闻报社/官方机构 > 网络新闻媒体 > 社交自媒体 > 论坛网贴
丰富度：专业新闻报社/官方机构 < 网络新闻媒体 < 社交自媒体 < 论坛网贴

- 知识图谱的构建与推理：看山不是山

- 人物关系知识图谱，持股关系知识图谱，担保关系知识图谱，产业链知识图谱
- 拓展新闻相关标的：公司 vs. 公司高管 vs. 母公司/子公司 vs. 供应链上下游 vs. 属地
- 知识推理：泰国洪水 → 硬盘产能下降 → 硬盘价格上涨 → 计算机价格上涨

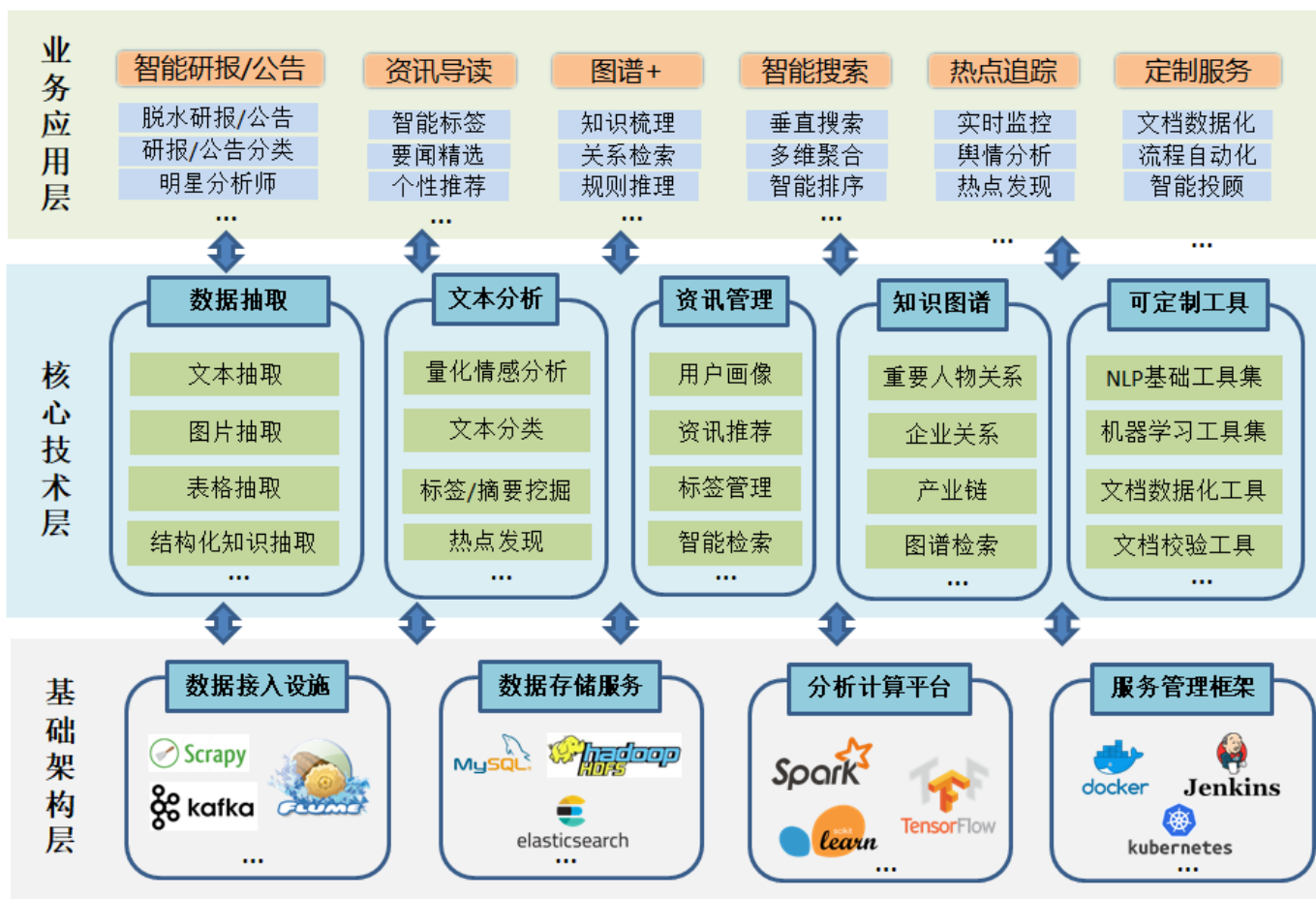
智能舆情跟踪系统—AI+金融资讯

- 文本分类任务：如事件分类，舆情分类，除噪等
 - ▣ 常用模型：CNN, FastText, 预训练+迁移学习模型等
- 结构化知识抽取：实体识别，实体消歧，关系抽取等
 - ▣ 常用模型：图像识别，有限自动状态机模型，BiLSTM+CRF，远程监督模型等
- 文本摘要及主体分析：话题分析，新闻/公告摘要等
 - ▣ 常用模型：LDA话题模型，基于注意力Seq2Seq深度学习模型

新的技术突破带来新的机遇与挑战

NLP历史突破！谷歌BERT模型狂破11项纪录，全面超越人类！
新智元 2018-10-13 17:37:28

智能舆情跟踪系统—我们的进展(1)



智能舆情跟踪系统—我们的进展(2)

基于Docker微服务架构的分布式实时采集平台

24小时内热门公司榜单

代码	名称	舆情范围	新闻数	涨跌	主要事件
000333	美的集团	-0.27-0.95	59	停牌	收购兼并
601198	建设银行	0.89-0.95	50	0.14%	财报披露
600687	伊利股份	-0.92-0.76	49	-2.40%	其他事件

24小时内负面新闻榜单

新闻日期	新闻标题	舆情分数	新闻来源
2018-10-25 09:50:00	市场监管总局公布7起食品保健食品欺诈和虚假宣传案件	-0.96	证券时报网
2018-10-24 18:02:35	近300笔合计56.23亿元逾期债务 仙股*ST凯迪资产出售将受影响	0.97	新京报网
2018-10-25 04:19:00	股吧热议*ST凯迪：限售股占比逾一成 减持利空影响有限	-0.97	证券时报

AI+大数据驱动的资料流处理平台

近300笔合计56.23亿元逾期债务 仙股*ST凯迪资产出售将受影响 2018-10-24 18:02:35

10月24日，*ST凯迪发布公告，对公司到期债务未偿债务情况进行更新。据新京报记者不完全统计，截至公告发布日，*ST凯迪到期未偿债务数量高达281笔，合计金额约56.23亿元，占其最新一期经审计公司净资产的52.88%。

记者查询相关财报得知，*ST凯迪最新一期经审计的公司净资产为10.63亿元，逾期债务占最新一期经审计净资产的比例为52.88%。而根据2018年三季度业绩预告显示，2018年三季度归属净利润预计亏损1.8亿元/4亿元，相对上年同期的4648.06万元，预计将比上年同期减少89.72.58%-5263.44%。

舆情分数: -0.97

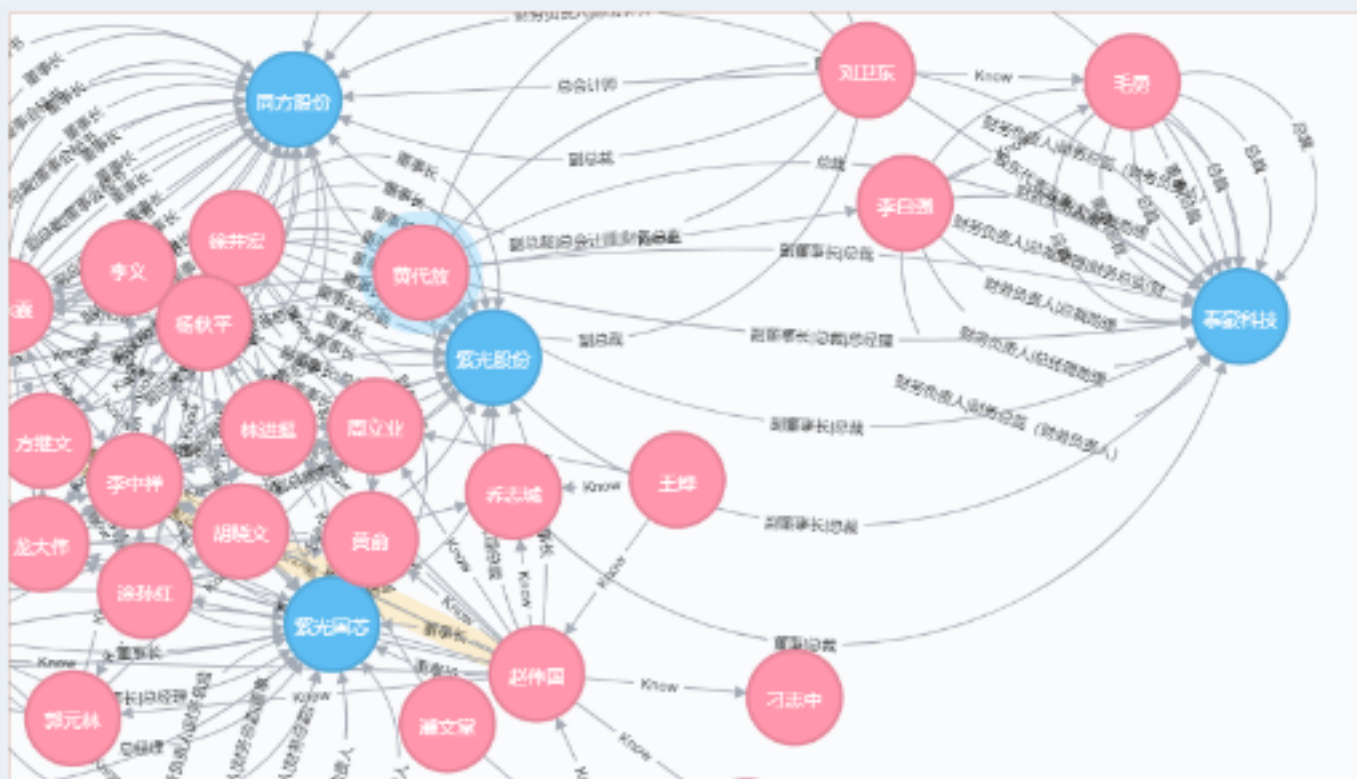
视频利空

相关文件: 15_0.6069, 查看更多>>

相关行业: 000000, 0.9941, 非行业新闻

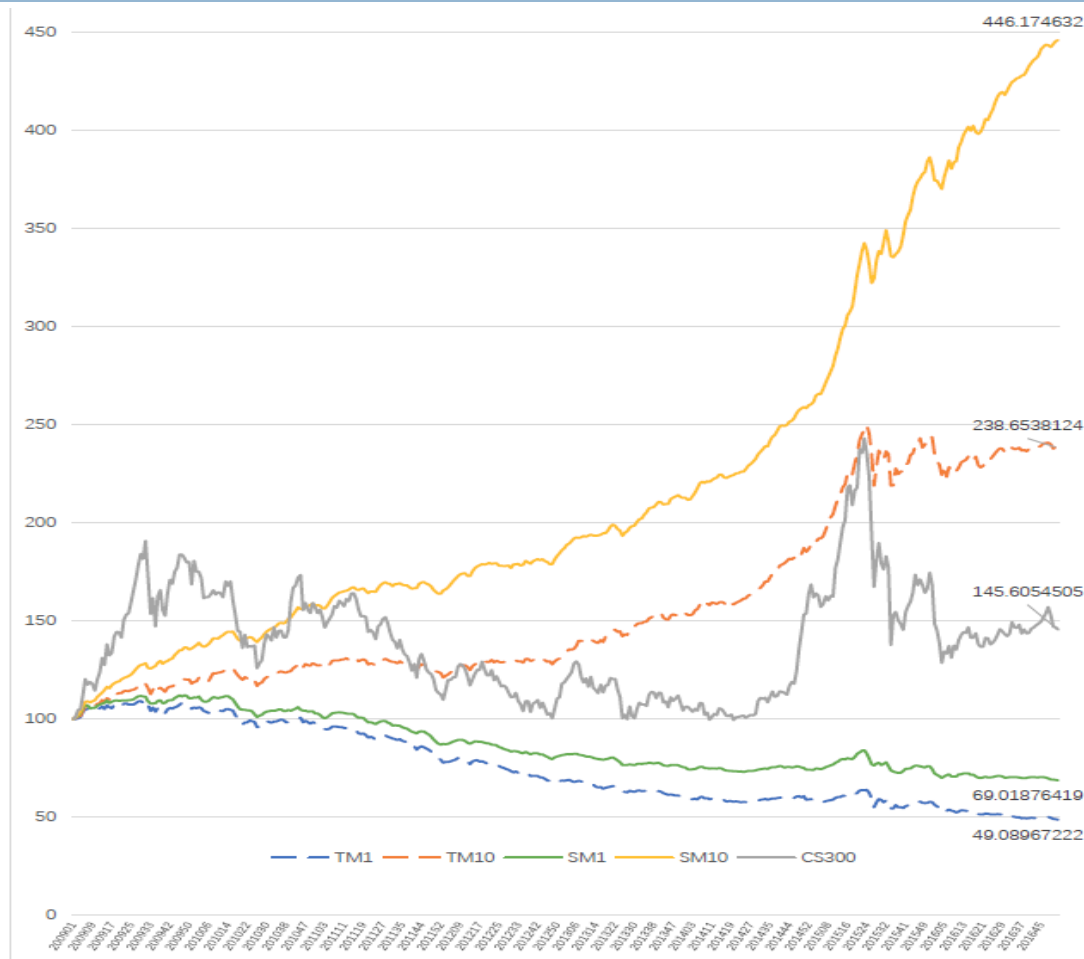
智能舆情跟踪系统—我们的进展(3)

- 不断完善的基础知识图谱（人物，公司，股东等）



智能舆情跟踪系统—我们的进展(4)

- 市场行情+资讯数据驱动的投资收益研究



图：2000年-2015年基于报刊新闻和股吧评论情绪的投资收益研究

谢谢！



Email: info@datago.com.hk

Site: <http://datago.com.hk>