

# 大数据的金融新创力

---

新浪财经 产品大数据中心

2016-11-18

## ► 4个V的应用



商业  
决策



财产  
安全



精准  
营销



猜你  
喜欢



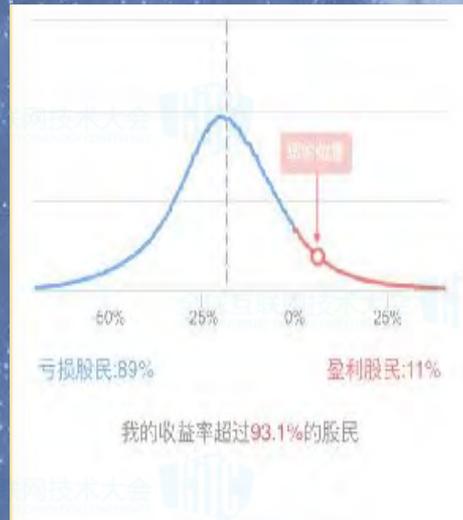
共同  
兴趣

大数据技术的战略意义不在于掌握庞大的数据信息，而在于对这些含有意义的数据进行专业化处理。

# ► My Portfolio Performance



今年跑赢上证指数  
34.18%



我的收益超过  
93.1%的股民



52.17% VS 25.47%



**新浪**影响主流中国，**微博**汇聚社会力量

# 新浪+微博 双核联动

注册用户超过  
**5亿**

人均好友超过  
**200人**

每天新发微博内容量  
超过**1.3亿**

用户每日时长超过  
**90分钟**

## 权威媒体

## 社交网络

## 开放平台

每天超过**100万**优质媒体及意见领袖贡献内容

每天超过**1亿**主流用户通过新浪网来了解世界

每天超过**40亿**页面访问及内容消费

超过**34万**平台开发者

超过**5万**应用和网站可以使用微博账号直接登录

每日超过**600万**站外分享至微博，超过**6000万**流量导出

近年来，随着微博的快速崛起，微博成为网民参与公共舆论的主要平台。微博用户加大对财经信息的关注力度，以及用户的消费能力的增长，都为微博在财经领域快速发展提供重要基础，其传播通路的价值不断深化。

2016年上半年

1.89亿条

财经博文

3752.5万

微博讨论用户

31.6万个

财经官方账号

2.69亿人次

财经订阅用户

财经相关博文总阅读数

1374.8亿次





新闻 行情 用户 微博 搜索 股吧





## 数据

社交  
移动  
用户行为  
资讯传播



## 技术

SMAC革命  
大数据处理  
智慧分析  
云计算



## 思想

理性精确  
探寻因果  
普遍相关  
逻辑发现



# INVESTMENT TOPIC

# 新闻数据—话题挖掘—概念潜在语义分析(LDA —— HDA)

LDA的概率图模型:



LDA(隐性语义分析)的目的是要从文本中发现隐含的语义维度——即“Topic”或者“Concept”。我们知道,在文档的向量空间模型(VSM)中,文档被表示成由特征词出现概率组成的多维向量,这种方法的好处是可以将query和文档转化成同一空间下的向量计算相似度,可以对不同词项赋予不同的权重

其中D代表文档, Z代表隐含类别或者主题, W为观察到的单词, 表示单词出现在文档的概率, 表示文档中出现主题下的单词的概率, 给定主题出现单词的概率。并且每个主题在所有词项上服从Multinomial 分布, 每个文档在所有主题上服从Multinomial 分布。整个文档的生成过程是这样的:

- (1) 以 $P(d_i)$ 的概率选中文档 $d_i$ ;
- (2) 以 $P(z_k|d_i)$ 的概率选中主题 $z_k$ ;
- (3) 以 $P(w_j|z_k)$ 的概率产生一个单词。

我们可以观察到的数据就是 $(d_i, w_j)$ 数据对, 而 $z_k$ 是隐含变量。 $(d_i, w_j)$ 的联合分布为:

$$P(d_i, w_j) = P(d_i) P(w_j|d_i),$$

$$P(w_j|d_i) = \sum_{k=1}^K P(w_j|z_k)P(z_k|d_i)$$

# 挖掘投资风口：主题聚合



## 今日最新

热度  
151789

### 房地产

时间：2015-12-25 13:57:31  
相关股票：大港股份 +10.01%

热度  
135987

### 万科概念

时间：2015-12-25 12:57:10  
相关股票：中兴商业 +2.76%

热度  
16826

### 中石油管道业务整合

时间：2015-12-25 08:39:10  
相关股票：东方热电 +9.42%

## 近期最热

热度  
328695

### 房地产

时间：2015-12-25 13:57:31  
相关股票：大港股份 +10.01%

热度  
187150

### 增持

时间：2015-12-25 13:52:10  
相关股票：中兴商业 +2.76%

热度  
178095

### 银行

时间：2015-12-25 13:55:10  
相关股票：怡亚通 +9.99%

## 今日异动

热度变化率  
860.50 %

### 航运

时间：2015-12-25 13:57:34  
相关股票：中海发展 +10.01%

热度变化率  
458.00 %

### 油品改革

时间：2015-12-25 13:13:16  
相关股票：龙宇燃油 +10.00%

热度变化率  
248.00 %

### 海工装备

时间：2015-12-25 13:41:25  
相关股票：中船防务 +5.36%

# ▶ 投资主题关注度与主题净值

正感情  
100.00%

负感情  
0.00%

近期最热

今日最新

2015-12-16 09:44:05  
新闻情感分析

中性  
0.00%

热度  
184978

医疗

时间：2015-12-16 10:57:31  
相关股票：宝莱特 +10.01%

热度  
175609

医疗

时间：2015-12-16 10:57:31  
相关股票：宝莱特 +10.01%





# TOPIC STOCK

## ► 投资主题下最值得关注的个股

序号	股票	最新价	涨跌额	涨跌幅	昨收	今开	最高	最低
1	明星电力 600101	15.95	+1.45	+10.00%	14.50	15.95	15.95	15.95
2	西宁特钢 600117	7.41	+0.67	+9.94%	6.74	7.36	7.41	6.91
3	冠农股份 600251	11.44	-0.14	-1.21%	11.58	11.58	11.68	11.34
4	鼎立股份 600614	12.34	+0.23	+1.90%	12.11	12.05	12.53	11.96
5	林洋电子 601222	39.99	-0.04	-0.10%	40.03	40.18	40.65	39.31

# 个股维度

荃银高科 300087.SZ

11.63 -0.03 -0.26%

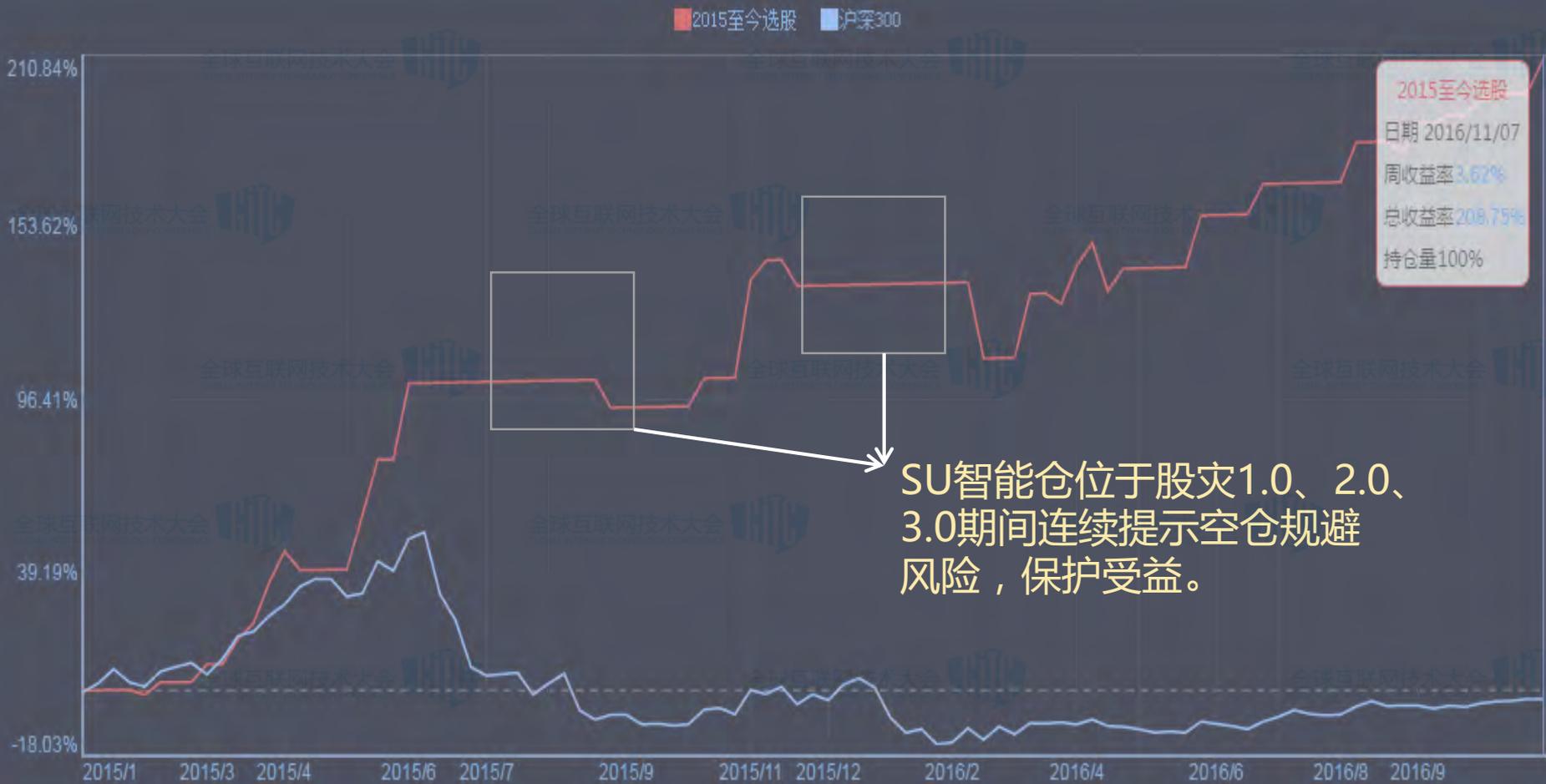
苯乙烯	甲醛检测	甘蔗废料利用	碳海绵
海洋渔业	有机硅	船舶航空	微电子
电解电容器纸	太阳帆	类稀土	碳海绵
海洋渔业	有机硅	船舶航空	微电子





# SMART USER

# ▶ Sina Smart User

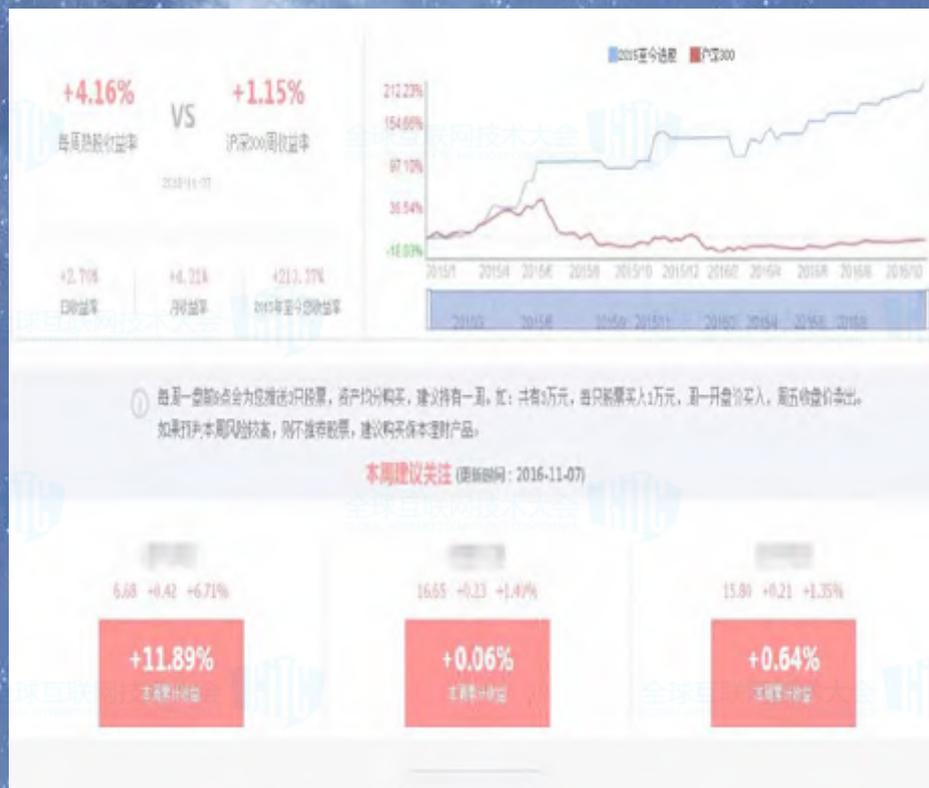




# BIG DATA STRATEGY

# ▶ Sina smart user + Big data portfolio

2015年以来回溯累计收益**210.37%**  
同期沪深300指数累计收益**-0.94%**





## ► 我们用到的技术



# 大数据 让投资更容易



# 2014年

- 9月12日 新浪南方**大数据i指数**在深交所挂牌
- 9月24日 大数据战略发布，**财经大数据研究中心**成立
- 11月11日 全新**数据中心** (finance.sina.com.cn/data) 上线

# 2015年

- 4月1日 基于新浪大数据的“**舆情牛股**” 上线
- 4月1日 **港股Level2**行情上线
- 4月22日 **大数据100指数基金**首发，有效认购近36亿，配售比27.88%，一天完成10亿元募集上限
- 5月28日 大数据100指数基金打开申购一天，截至6月30日，净值规模达**102.8亿元**
- 6月2日 与中金公司独家合作的投顾产品“**i牛股**” 上线
- 6月15日 **大数据300指数基金**首发
- 8月18日 **沪深Level2**行情普及版上线

# 2016年

- 1月15日 36支队伍中脱颖而出获得新浪集团**创新大赛一等奖**
- 5月10日 基于社交互动数据的**众智基金**首次募集
- 11月11日 大数据C端产品“**投资易**” 上线

新浪财经 产品大数据中心

Hey, Big  
Data

Thank You !

---