

# 股票择时投资策略

汇真科技 量化金融组 2019.06.14

## 策略投研目录

### 1：量化择时

#### 1.1：量化择时算法

#### 1.2: 4/7 择时原则

### 2：甄选行业及股票池

#### 2.1：LSTM 算法

#### 2.2：股票池

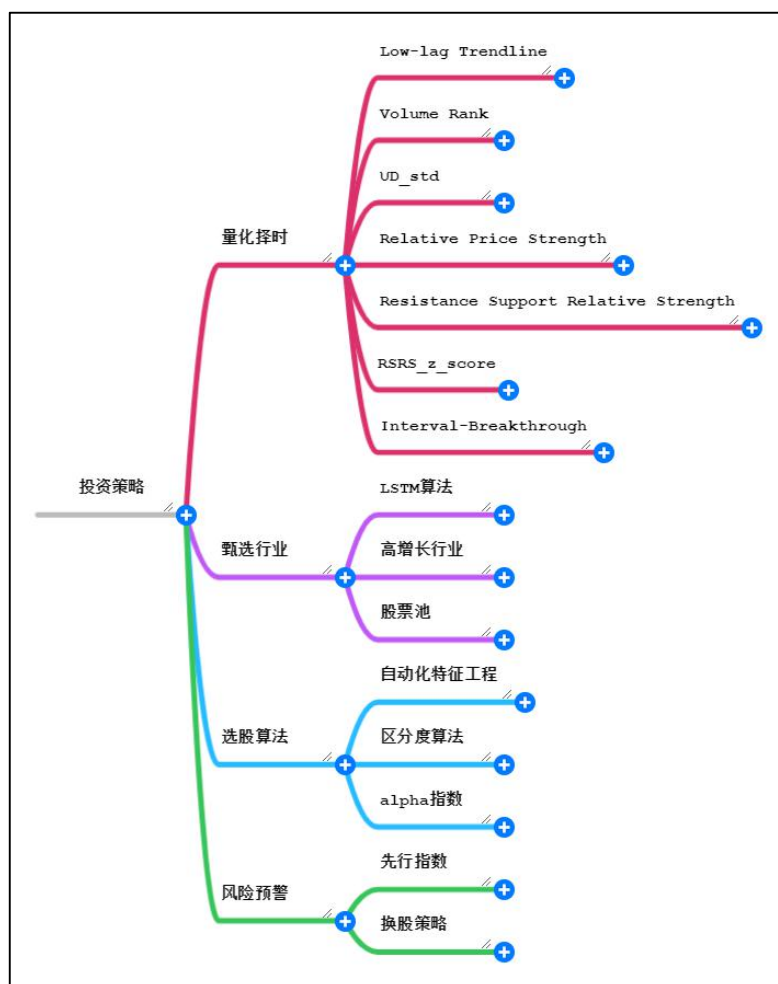
### 3：投资组合的算法及组合的表现

#### 3.1：自动化特征工程。

#### 3.2：区分度算法

#### 3.3：投资组合 19 年的表现。

### 4：风险预警及换股策略



#### 1.1:量化择时

本量化择时策略主要针对股票市场总体进行。结合七种不同的量化择时方法，构建基于股票市场总体的单方向做多策略。当七种方法中任意四种及以上给出持仓信号时，入场做多；反之出场或保持空仓。简称 4/7 择时策略。

这七种方法分别为：

(1) 低延迟趋势线 ( Low-lag Trendline ) ， LLT

原理类似于 EMA 均线。通过将价格信号中的高频分量进行有效的过滤，构造低延迟趋势线，相比传统均线延迟程度下降。由于 LLT 趋势线与其他趋势系统类似，在一定的参数条件下具有较好的平

滑性，因此可以将该趋势线近似看作一条处处可微的曲线。通过向前差分计算，可以在每个交易日结束后得到 LLT 趋势线在该点处切线的斜率。当斜率值大于某一临界值时，看多市场，给出持仓信号。

(2) 成交量排序法，Volume Rank

根据最近一日成交量在过去 N 日中的排序判定是放量还是缩量，依此作出开平仓决策。

(3) 向上、向下单向波动差，UD\_std

依据一段时间内指数值上升、下降分别计算波动值，然后计算二者差值并取平均，当平均值大于 0 时，则向上波动势头相对强，做出持仓决策。

(4) 相对价格强度 (Relative Price Strength) 单向波动差，RPS\_std

对 UD\_std 中取平均时间段做自适应处理。RPS 是指数当前点位在过去 N 日点位序列中所处高度的比例，由 RPS 值确定计算波动率差值移动均值所取天数，计算指数相应上行、下行波动率的差值，确定是否持仓。

(5) 阻力支撑相对强度 (Resistance Support Relative Strength)，RSRS

以窗口期日度最高价、最低价的相对变化程度，判定支撑强度与阻力强度的关系，决定是否持仓。

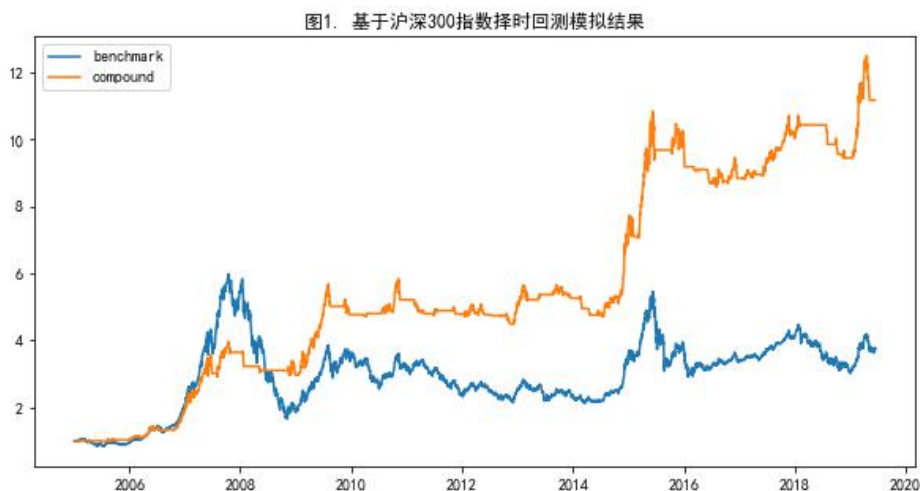
(6) RSRS 标准分，RSRS\_z\_score

在 RSRS 基础之上对阻力支撑相对强度进行 N 日标准化，即对指标进行尺度变换，作为判定是否持仓的指标。

(7) 区间突破指标，Interval-Breakthrough

根据窗口期价格波动宽度，弹性调整观察窗口宽度，确定向上、向下突破阈值，作为开平仓标准。

以沪深 300 指数作为中国股票市场代表性指数，采用上述 4/7 择时策略，做出持仓、空仓决策。为刻画择时策略效果，依据持仓、空仓决策对沪深 300 指数进行模拟交易，手续费参考对应场内 ETF 基金设为双边万五。回测模拟期间为 2015.01.04-2019.6.14(日频) 结果见下图 1。其中，benchmark 为沪深 300 指数值曲线，compound 为累积收益曲线，为使二者具有可比性，均进行初始时刻归一化处理。



对上述回测模拟结果进行统计分析，结果见下表 1、表 2：

表 1. 沪深 300 指数整体及分年度表现

	收益率	最大回撤	Sharpe 比率	Calmar 比率
<b>整体</b>	<b>2.7499</b>	<b>-0.7230</b>	<b>0.3353</b>	<b>0.2633</b>
<b>2005</b>	-0.0604	-0.2202	-0.3041	-0.2781
<b>2006</b>	1.1680	-0.1382	3.5941	8.5958
<b>2007</b>	1.5825	-0.2090	2.6421	7.7191
<b>2008</b>	-0.6625	-0.7160	-2.2747	-0.9277
<b>2009</b>	0.8990	-0.2526	2.0176	3.6077
<b>2010</b>	-0.1151	-0.2950	-0.4983	-0.3946
<b>2011</b>	-0.2646	-0.3164	-1.5299	-0.8478
<b>2012</b>	0.0975	-0.2241	0.4714	0.4389
<b>2013</b>	-0.0770	-0.2216	-0.3732	-0.3513
<b>2014</b>	0.5219	-0.1012	2.2317	5.1845
<b>2015</b>	0.0246	-0.4348	0.0627	0.0573
<b>2016</b>	-0.0458	-0.1938	-0.2260	-0.2391
<b>2017</b>	0.2060	-0.0607	1.9002	3.4429
<b>2018</b>	-0.2634	-0.3188	-1.4667	-0.8379
<b>2019</b>	0.2411	-0.1349	1.9716	4.0258

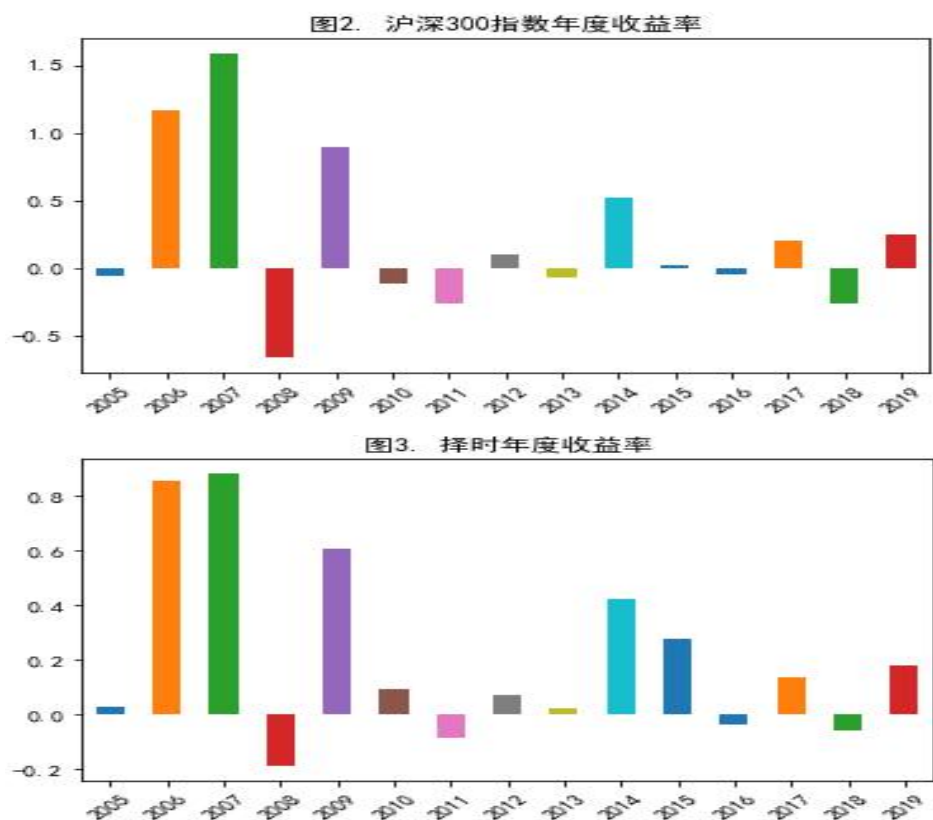
表 2. 累积收益整体及分年度表现

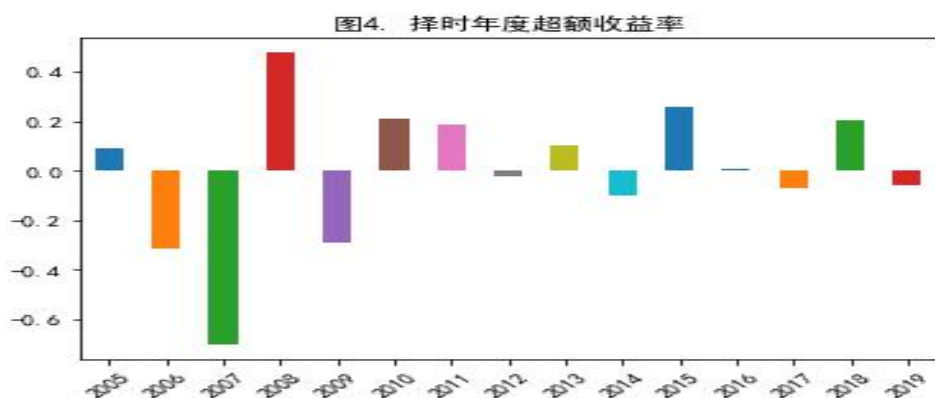
	收益率	超额收益率	最大回撤	Sharpe 比率	Calmar 比率
<b>整体</b>	<b>10.1645</b>	<b>7.4146</b>	<b>-0.2552</b>	<b>0.9773</b>	<b>2.7572</b>
<b>2005</b>	0.0308	0.0911	-0.0415	0.4172	0.7514
<b>2006</b>	0.8554	-0.3126	-0.0956	3.2523	9.0968
<b>2007</b>	0.8813	-0.7012	-0.1847	2.0567	4.8657
<b>2008</b>	-0.1870	0.4754	-0.1951	-1.0636	-0.9613
<b>2009</b>	0.6099	-0.2891	-0.1608	1.9775	3.8464

2010	0.0929	0.2080	-0.1073	0.6355	0.8758
2011	-0.0816	0.1829	-0.0911	-1.1237	-0.9089
2012	0.0726	-0.0249	-0.1174	0.5939	0.6241
2013	0.0243	0.1013	-0.0829	0.1898	0.2962
2014	0.4230	-0.0989	-0.1138	2.2044	3.7362
2015	0.2800	0.2554	-0.1339	1.0596	2.1208
2016	-0.0369	0.0089	-0.0656	-0.4486	-0.5689
2017	0.1354	-0.0707	-0.0624	1.5066	2.1985
2018	-0.0580	0.2054	-0.1193	-0.8422	-0.4930
2019	0.1821	-0.0590	-0.1063	1.7567	3.8615

从选取的整个时段来看，择时策略带来 7 倍多超额收益；且回撤（25.52%）远小于指数本身的下撤（72.30%）。同时 Sharpe 比率和 Calmar 比率也进一步说明采取择时策略，可获得更高收益、承受更小风险。

下面分年度来呈现市场择时的适用性与效果：





对比如上三幅年度收益率柱状图，显然在市场整体表现不佳的年份，如 2008、2018 年等，择时效果明显；而在市场处于牛市状态的年份，如 2007、2017 年等，择时策略的应用会降低当年总体收益，不过绝对收益依然为正。年度胜率（53%）虽不够高，但从整体来看，择时可以有效规避下行风险，并获得更高平均收益。

上述七种择时策略我们主要用于资产配置周期选择中，作为市场状态的判断依据。

### 3. 各方法持仓建议

择时方法	信号
LLT	空仓
Volume Rank	空仓
UD_std	空仓
RPS_std	空仓
RSRS	持仓
RSRS_z_score	持仓
Interval-Breakthrough	持仓

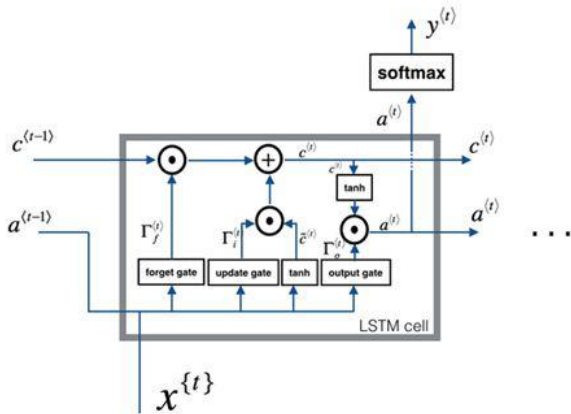
七种择时方法中，三种给出持仓建议，可判定当前市场处于震荡状态，此时我们会部分减持。

七种择时方法中，根据以上的回测效果，我们定义了买卖原则如下：

买入或持有	减持	卖出
$\geq 4/7$	$= 3/7$	$\leq 2/7$

## 1.2：甄选行业的算法-LSTM 算法

LSTM 算法工作原理：



LSTM 区别于 RNN 的地方 ,主要就在于它在算法中加入了一个判断信息有用与否的“处理器”，这个处理器作用的结构被称为 cell。一个 cell 当中被放置了三扇门，分别叫做输入门、遗忘门和输出门。一个信息进入 LSTM 的网络当中，可以根据规则来判断是否有用。只有符合算法认证的信息才会留下，不符的信息则通过遗忘门被遗忘。简单说就是一进二出的工作原理。

我们把每个行业的输入值（每个行业 29 个特征值）带入 LSTM 模型进行计算。

```
# ----- 训练模型 -----
def train_lstm(batch_size=6,time_step=12,train_begin=0,train_end=682):
    X=tf.placeholder(tf.float32, shape=[None,time_step,input_size])
    Y=tf.placeholder(tf.float32, shape=[None,time_step,output_size])
    batch_index,train_x,train_y=get_train_data(batch_size,time_step,train_begin,train_end)
    pred,_=lstm(X)
    #损失函数
    loss=tf.reduce_mean(tf.square(tf.reshape(pred,[-1])-tf.reshape(Y, [-1])))
    train_op=tf.train.AdamOptimizer(lr).minimize(loss)
    saver=tf.train.Saver(tf.global_variables(),max_to_keep=15)
    #module file = tf.train.latest_checkpoint()

    with tf.Session() as sess:
        sess.run(tf.global_variables_initializer())
        saver.restore(sess, module_file)
        #重复训练10000次
        for i in range(32000):
            for step in range(len(batch_index)-1):
                _loss_=sess.run([train_op,loss],feed_dict={X:train_x[batch_index[step]:batch_index[step+1]:],Y:train_y[batch_index[step]:batch_index[step+1]:]})
                print(i,_loss_)
            if i % 32000==0:
                base_path = saver.save(sess, "module/alpha_lstm1.model")
                print("保存模型: ",saver.save(sess,base_path,global_step=i))

train_lstm()

# ----- 预测模型 -----
def prediction(time_step=12):
    X=tf.placeholder(tf.float32, shape=[None,time_step,input_size])
    Y=tf.placeholder(tf.float32, shape=[None,time_step,output_size])
    mean,std,test_x,test_y=get_test_data(time_step)
    pred,_=lstm(X)
```

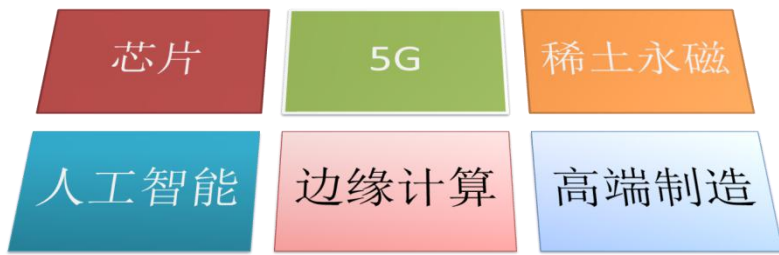
序号	行业	10 组涨幅	净利润同比	市净率	市盈率	主营收入	每股收益	MA	VR0C	GB	...
1	种业	16.02%	-35.31	3.23	43.32	244701	0.0638	2611.805	55.386	0.326	...
2	黄金股	9.34%	-89.654	1.91	58.52	12863970	0.0277	3252.402	129.176	0.067	...
3	稀缺资源	9.09%	2.839	1.56	21.99	29757468	0.0737	3415.824	49.982	0.002	...
4	国产乳业	7.06%	23.465	4.06	25.4	5222641	0.186	5699.748	38.605	0.395	...
5	水泥	5.24%	147.117	1.27	12.03	8606615	0.252	3284.83	-38.718	0.14	...
6	工业互联网	4.54%	36.597	2.54	17.84	31468042	0.21	3197.37	-40.351	0.152	...
7	新材料	4.31%	7.293	2.14	25.94	28011914	0.0931	4427.554	5.802	0.019	...
8	5G	3.98%	84.565	2.1	44.44	22223612	0.067	3952.584	-18.157	0.09	...
9	白酒	3.64%	22.013	6.56	19.12	7642085	1.536	13437.881	-51.068	0.435	...
10	小金属	3.44%	-3.354	1.54	26.75	24840654	0.0511	2343.08	40.113	0.063	...
11	半导体	3.21%	-22.367	2.51	32.24	18546280	0.093	2962.203	-18.121	0.102	...
12	乡村振兴	3.12%	-335.112	3.44	-180.35	11568785	-0.0076	4601.51	9.472	0.482	...
13	广电系	3.01%	-8.842	1.34	38.59	985821	0.0418	2878.005	-11.589	0.021	...
14	北斗导航	2.91%	-11.759	2.77	142.65	1728495	0.0201	2579.426	10.635	0.043	...
15	土地流转	2.54%	1056.443	1.62	22.37	4162557	0.0533	2217.517	16.063	0.021	...
16	次新股	2.49%	44.137	2.61	23.86	37738484	0.148	10847.566	3.916	0.006	...
17	自主可控	2.38%	2.125	2.88	138.99	9575264	0.0144	2566.285	-14.761	0.033	...
18	机械	2.35%	54.721	1.84	28.03	19738178	0.0708	2846.239	-14.311	0.048	...
19	征信概念	2.18%	109.798	2.92	43.04	1910214	0.0731	4852.508	-44.585	0.033	...
20	稀土永磁	2.10%	3.468	2.01	48.71	14979424	0.0272	2803.69	33.115	0.198	...
21	大飞机	2.07%	9.308	1.78	47.02	5009717	0.0407	1878.071	-7.456	0.054	...
...	...	...	...	...	...	...	...	...	...	...	...

计算公式如下：

$$\begin{aligned} \tilde{c}^{<t>} &= \tanh(W_c[a^{<t-1>}, x^{<t>}] + b_c) \\ \Gamma_u &= \sigma(W_u[a^{<t-1>}, x^{<t>}] + b_u) \\ \Gamma_f &= \sigma(W_f[a^{<t-1>}, x^{<t>}] + b_f) \\ \Gamma_o &= \sigma(W_o[a^{<t-1>}, x^{<t>}] + b_o) \\ c^{<t>} &= \Gamma_u * \tilde{c}^{<t>} + \Gamma_f * c^{<t-1>} \\ a^{<t>} &= \Gamma_o * \tanh c^{<t>} \end{aligned}$$

经过计算我们从 550 多个行业中，把行业分为两大类，高增长型和普通类型，然后我们从高增长型行业中甄选出我们重点关注的行业，截止到 2019-年 6 月份我们目前主要关注以下 6 种行业。





然后根据我们重点关注的行业及个股所属行业来挑选股票入选股票池，见下表：

代码	名称	所属行业
600487'	亨通光电	充电桩,半导体,新能源,新能源车,光通信,MSCI 中国,特高压,大数据,网络安全,量子通信,芯片概念,5G,一带一路
002415'	海康威视	机器视觉,智能硬件,智能交通,MSCI 中国,智慧城市,大数据,智能机器,人脸识别,人工智能,生物识别,车联网,安防
000063'	中兴通讯	物联网,足球概念,创投,MSCI 中国,智慧城市,工业 4.0,智能机器,亚投行,无线充电,体育产业,网络安全,人工智能,云计算,移动互联网,5G,智能家居,一带一路
300676'	华大基因	次新股,特色小镇,MSCI 中国,大数据,医疗器械,基因测序
002405'	四维图新	物联网,北斗导航,移动互联通信,半导体,航天军工,特斯拉,智能汽车,智能交通,智慧城市,大数据,证金扫货,军工动态,卫星导航,无人驾驶,知识产权,人工智能,地理信息,芯片概念,车联网,智能手机
002384'	东山精密	苏南示范区,OLED,新能源,特斯拉,苹果三星,苹果产业链,柔性电子,太阳能,5G,LED
002368'	太极股份	软件产业,智慧城市,数字中国,云计算,军民融合,雄安新区,汽车电子
002410'	广联达	电子商务,智慧城市,征信概念,体感 3D,装配建筑,供应链金融,雄安新区
300098'	高新兴	虚拟现实,物联网,增强现实,智能交通,智慧城市,PPP 概念,人脸识别,人工智能,云计算,5G,生物识别,车联网,安防,雄安新区
002236'	大华股份	智慧物流,新能源,新能源车,MSCI 中国,智慧城市,大数据,智能机器,PPP 概念,人脸识别,人工智能,二胎,智能家居,生物识别,安防
002230'	科大讯飞	粤港澳,智能音箱,智能硬件,智能汽车,人脑工程,智慧医疗,MSCI 中国,大数据,智能机器,数字中国,无人驾驶,教育产业,无人零售,人脸识别,在线教育,人工智能,云计算,智能家居,生物识别,车联网
002385'	大北农	智慧农业,MSCI 中国,猪肉概念,生态农业,供给侧改革,农村电商
002215'	诺普信	智慧农业,乡村振兴,电商,生态农业,供给侧改革,农村电商,互联网金融
300022'	吉峰科技	农机,巴菲特概念,智慧农业,乡村振兴,农业现代化,生态农业,农村电商
000555'	神州信息	智慧农业,信息安全,大数据,数字中国,土地流转,人工智能,量子通信,独角兽概念
002130'	沃尔核材	能源互联,新能源,特斯拉,风电,超导概念,两会相关,高送转,核电,新材料
002036'	联创电子	虚拟现实,增强现实,特斯拉,智能穿戴,无人机,无线充电,芯片概念,盖板玻璃
300088'	长信科技	能源互联,OLED,小米概念,新能源,特斯拉,新能源车,智能汽车,全面屏,三元锂电,大数据,证金扫货,苹果产业链,智能穿戴,触摸屏,柔性电子,高送转,LED,汽车零部件
000012'	南玻A	粤港澳,OLED,新能源,特斯拉,涨价预期,低碳经济,高股息,太阳能,节能环保,国资重组,险资举牌,雄安新区,汽车零部件
600458'	时代新材	轨道交通,OLED,新能源,特斯拉,柔性电子,高铁,铁路基建,新材料
000034'	神州数码	大数据,数字中国,教育产业,云计算
300017'	网宿科技	微信,虚拟现实,移动互联通信,电子商务,网络游戏,MSCI 中国,信息安全,大数据,混改,网络安全,高送转,云计算,一带一路



600845'	宝信软件	虚拟现实,智能汽车,国产软件,工业互联网,大数据,博鳌相关,车联网,央企改革
300024'	机器人	保险相关,智慧物流,新能源,航天军工,MSCI 中国,工业 4.0,智能机器,东北振兴,无人驾驶,引力波,传感器,人工智能,3D 打印,中国制造 2025,养老产业
300339'	润和软件	王亚伟概念,智能汽车,国产软件,智慧城市,智能电网,蚂蚁金服,高送转,区块链,云计算
300020'	银江股份	虚拟现实,物联网,增强现实,智能交通,智慧医疗,互联医疗,智慧城市,G20 峰会,大数据,博鳌相关,数字中国,无人驾驶,教育产业,PPP 概念,京津冀,地理信息,云计算,高铁,铁路基建,车联网
300036'	超图软件	物联网,北斗导航,智慧物流,智能交通,智慧城市,大数据,数字中国,卫星导航,地理信息,云计算,军民融合,安防,雄安新区,一带一路
600601'	方正科技	创投,依法治国,智慧城市,电子发票,云计算,5G
600446'	金证股份	创投,国产软件,余额宝概念,博鳌相关,区块链,互联网金融,独角兽概念
300663'	科蓝软件	次新股,电子支付,无人零售,区块链,互联网金融
603383'	顶点软件	次新股,知识产权
002657'	中科金财	银联概念,电子商务,国产软件,债转股,信息安全,NFC 概念,大数据,移动支付,博鳌相关,数字中国,彩票,征信概念,智能家居,互联网金融
300085'	银之杰	参股券商,软件产业,大数据,移动支付,征信概念,跨境电商,区块链,移动互联网,安防
300465'	高伟达	大数据,蚂蚁金服,区块链,云计算,互联网金融,独角兽概念
300368'	汇金股份	征信概念,京津冀
000049'	德赛电池	小米概念,新能源,恒大系,苹果三星,苹果产业链,锂电池,国资改革,移动互联网,电池管理,智能手机
300662'	科锐国际	次新股,小米概念,宁德时代概念
002123'	梦网集团	SAAS,充电桩,小米概念,新能源,共享经济,特高压,智能电网,节能环保,燃料电池,独角兽概念
002456'	欧菲光	虚拟现实,碳纳米管,纳米概念,增强现实,数字电子,OLED,小米概念,新能源,智能汽车,苹果三星,全面屏,MSCI 中国,智慧城市,无人驾驶,苹果产业链,智能穿戴,触摸屏,柔性电子,高送转,盖板玻璃
300083'	劲胜智能	创投,小米概念,苹果三星,苹果产业链,触摸屏,无线充电,盖板玻璃,智能手机
002429'	兆驰股份	OLED,半导体,小米概念,家用电器,高送转,芯片概念,智能电视,智能家居,独角兽概念,LED
002313'	日海智能	物联网,北斗导航,小米概念,工业互联网,智慧城市,体育产业,5G
002369'	卓翼科技	虚拟现实,物联网,创投,增强现实,小米概念,区块链,人工智能,量子通信,智能手机
002217'	合力泰	OLED,小米概念,苹果三星,全面屏,柔性电子,无线充电,锂电池,高送转,5G,盖板玻璃
002177'	御银股份	债转股,NFC 概念,无人零售,区块链,人脸识别,生物识别,独角兽概念
...	...	...

截止到 2019 年 6 月 15 号我们的股票池总共挑选了 460 只股票。每周调整一次，周期取决于

## 4/7 择时原则

2.1：选股算法分为两个步骤，第一步是自动化特征工程，这个过程主要目的是挑选个股的最优特征。他分两个步骤，步骤过程见下图



## 第一步：算法分别挑选最优特征

1：用 xgboos 算法对财务数据做特征提取

```
NewFeature1=XGBClassifier(input(finance[36]),output[])
```

2：用 lightGBM 算法对行情数据做特征提取

```
NewFeature2=lightGBMClassifier(input(market[32]),output[])
```

3：用 GradientBoosting 算法对行技术指标做特征提取

```
NewFeature3 =GradientBoostingClassifier (input(technology[42]),output[])
```

4：用 randomforest 算法对统计数据做特征提取

```
NewFeature4 =randomforest(input(statistic[30]),output[])
```

## 第二步：特征融合

用 lr 算法把以上算法挑选出的新特征进行融合，筛选出逻辑层的特征。

```
CombineFeature=LRClassifier(input[NewFeature1, NewFeature2,
```

```
NewFeature3,NewFeature4],output[ ])
```

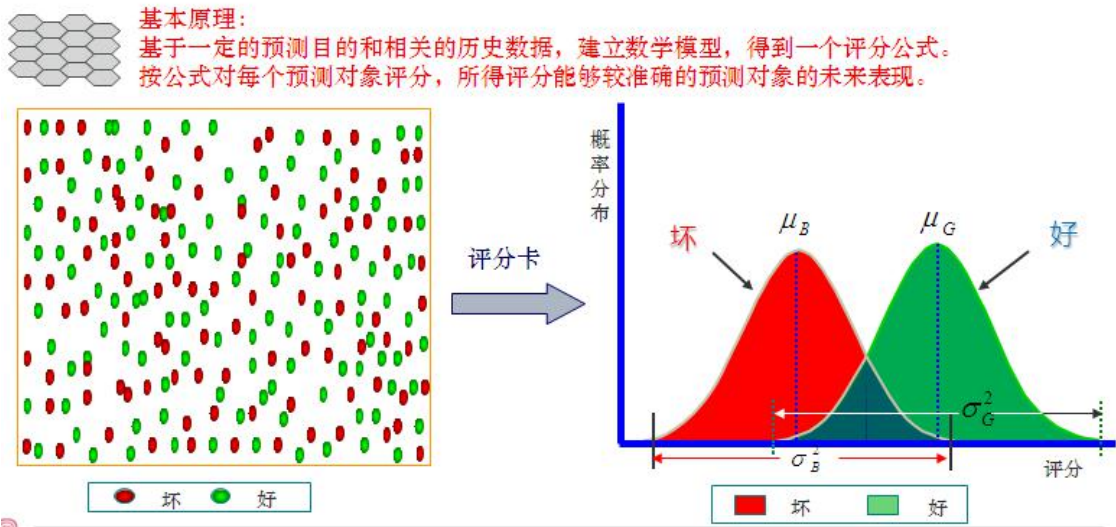
LRClassifier 是逻辑回归函数，NewFeature1，NewFeature2，NewFeature3 和 NewFeature4 是前四步选出来的新特征组

```
crystal-force - [E:\crystal-force] - \champion\360_stacking2.py - PyCharm 2017.1.4
File Edit View Navigate Code Refactor Run Tools VCS Window Help
crystal-force champion 360_stacking2.py
Project | 360_adaboost.py | 360_adaboost2.py | 360_adaboost3.py | 360_decisiontree.py | 360_decisiontree2.py | 360_feature_selection.py | 360_feature_selection2.py | 360_lightgbm.py | 360_lightgbm2.py | 360_logreg_prob.py | 360_mul_algorithm.py | 360_mul_algorithm2.py | 360_mul_algorithm3.py | 360_NB.py | 360_randomforest_feature.py | 360_randomforest_feature2.py | 360_randomforest_prob.py | 360_randomforest_prob2.py | 360_randomforest_prob3.py | 360_stacking2.py | 360_stackingClassifier.py | 360_tensorflow.py | 360_tensorflow_output.py | 360_test.py | 360_xgb.py | 360_xgb2.py | 360_xgb_standardization.py | 360finance.py | 360finance_70.py | 360finance_DeepRandomForest.py | 360finance_RandomForest.py | data.csv | financedata.py
59 #test_data = preprocessing.minmax_scale(test.iloc[:,1:].values, feature_range=(-1,1))
60 Xtest = list(test.columns.values)[4:6745]
61 x_data_output = dfp.iloc[:, 0:1].values
62 #print(data_train.iloc[:, 3:4])
63
64 predictors = list(cc.columns.values)[4:6745]
65 alg1 = lgb.LGBMClassifier(boosting_type='gbdt', num_leaves=42, max_depth=-1, learning_rate=0.054, n_estimators=490,
66                          subsample_for_bin=200, objective='binary', class_weight=None, min_split_gain=0.0,
67                          min_child_weight=1, min_child_samples=21, subsample=0.72, subsample_freq=1,
68                          colsample_bytree=0.63, reg_alpha=6.18, reg_lambda=2.718, random_state=142857, n_jobs=-1,
69                          silent=True, importance_type='split')
70
71 alg2 = XGBClassifier(n_estimators=60, max_depth=9, min_child_weight=2, gamma=0.9, subsample=0.8, learning_rate=0.02,
72                    colsample_bytree=0.8, objective='binary:logistic', nthread=-1, scale_pos_weight=1)
73
74 alg3 = GradientBoostingClassifier(learning_rate=0.01, n_estimators=600, max_depth=7, min_samples_leaf=60,
75                                  min_samples_split=1200, max_features=9, subsample=0.7, random_state=10)
76
77 lr = LogisticRegression()
78
79 pipe1 = make_pipeline(ColumnSelector(cols=predictors[4:2000]), lr)
80 pipe2 = make_pipeline(ColumnSelector(cols=predictors[2000:4000]), alg2)
81 pipe3 = make_pipeline(ColumnSelector(cols=predictors[4000:6745]), alg3)
82
83 scf = StackingClassifier(classifiers=[lr, pipe2, pipe3], meta_classifier=alg1)
84
85 # Compute the accuracy score for all the cross validation folds. (much simpler than what we did before!)
86 # kf=cross_validation.KFold(data_train.shape[0], n_folds=10, random_state=1)
87 kf = model_selection.KFold(n_splits=10, shuffle=False, random_state=1)
88 scores = model_selection.cross_val_score(scf, cc[predictors], cc['tag'], cv=kf)
89 print("scores: mean", scores.mean())
90
91 File = open("data/prob_stackingXg.txt", "w", encoding="utf-8", errors='ignore')
92 File.write("id* * * Target * \n")
93 classifier = scf.fit(cc[predictors], cc['tag'])
94 predictiontest = classifier.predict_proba(test[Xtest])[1:, 1]
95 for step in range(len(test)):
96     #
```

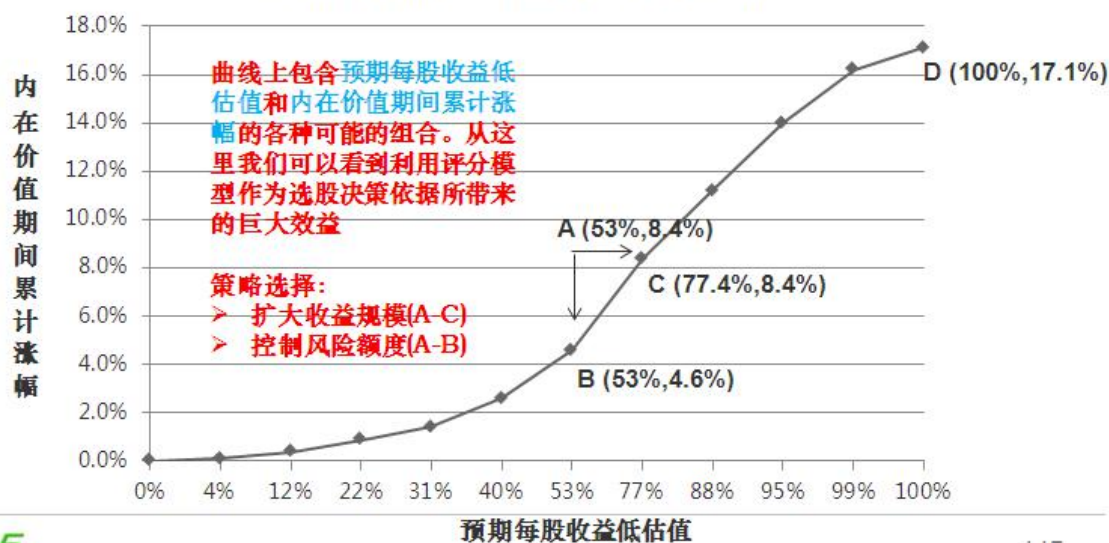
自动化特征工程是我们最耗时的地方，由于历史数据众多，算法复杂，目前做一次特征工程平均需要耗时 4-5 天，如果升级 GPU 后能缩短到 2-3 个小时。本轮周期经过融合后的特征从 140 个特征中甄选出了 25 个特征。每周期计算一次，周期取决于 4/7 择时原则。

2.2：选股算法第二步--区分度算法

区分度模型评分原理。



## 根据内在价值评分进行选股的策略



实际上我们是改造了区分度的部分算法的，改造后的算法一方面他剔除了市场外部因素对股价的影响，计算股票内在价值与当下股价的逻辑关系，另外一方面算法也从风险收益率上寻找最优的区分度。

计算公式如下：

每种不同部分协方差公式都是好坏协方差之和除以 2 即

$$S = \frac{\sigma_G^2 + \sigma_B^2}{2} \text{ (对于公式中 “} \sigma_G^2, \sigma_B^2 \text{” 指方差和)}$$

S	$X_1$	$X_2$	$X_3$	$y_1$	$y_2$	$z_1$	$z_2$
$X_1$	a	b		c			
$X_2$							
$X_3$							
$y_1$							
$y_2$							
$z_1$							
$z_2$							

浅灰部分协方差矩阵 S 中主对角线上 bin 内的方差

$$a = \frac{\sigma_{G1}^2 + \sigma_{B1}^2}{2}$$

$$\sigma_{G1}^2 = P_G(x_1) - P_G(x_1)^2$$

$$\sigma_{B1}^2 = P_B(x_1) - P_B(x_1)^2$$

深灰部分一个变量内不同 bin 之间的协方差

$$b = \frac{\sigma_{G1}^2 + \sigma_{B1}^2}{2}$$
$$\sigma_{G1}^2 = -P_G(x_1) \times P_G(x_2)$$
$$\sigma_{B1}^2 = -P_B(x_1) \times P_B(x_2)$$

虚线边框部分不同变量之间 bin 的协方差

$$c = \frac{\sigma_{G1}^2 + \sigma_{B1}^2}{2}$$
$$\sigma_{G1}^2 = P_G(x_1, y_1) - P_G(x_1) \times P_G(y_1)$$
$$\sigma_{B1}^2 = P_B(x_1, y_1) - P_B(x_1) \times P_B(y_1)$$

其中,

$$P_G(x_1, y_1) = \frac{\text{同时落在粗分箱 } x_1 \text{ 与 } y_1 \text{ 的好记录对应的样本权重之和}}{\text{原数据集中所有目标变量为 1 的记录对应的权重之和}}$$
$$P_B(x_1, y_1) = \frac{\text{同时落在粗分箱 } x_1 \text{ 与 } y_1 \text{ 的坏记录对应的样本权重之和}}{\text{原数据集中所有目标变量为 0 的记录对应的权重之和}}$$

把自动化特征工程甄选出来的 25 个特征值带入区分度算法的程序计算得到每个股票的

SCALINGSCORE 得分，见下表：

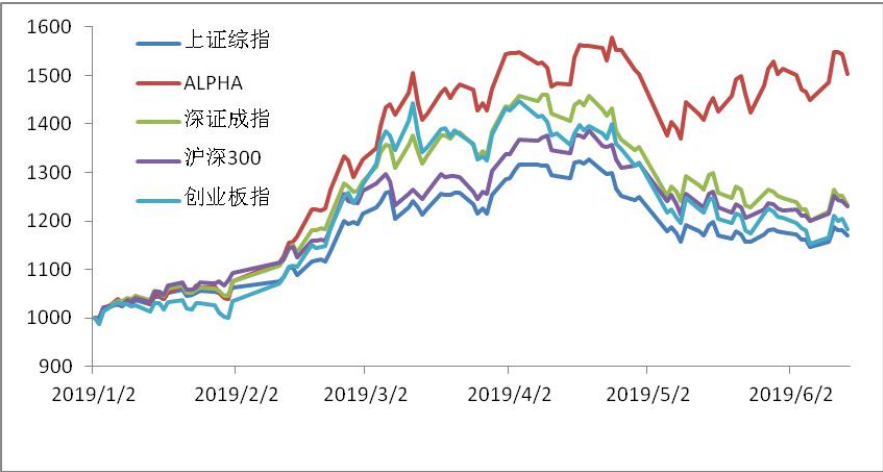
代码	名称	10 组涨幅	净利润同比	10 组换手	市净率	市盈率	主营收入	每股收益	净利润	MA	VROC	SCALINGSCORE
'300748'	金力永磁	214.65%	8.641	164.82%	22.73	250.54①	35886	0.06	2636	30.716	607.878	2.215
'002600'	领益智造	1.87%	47.852	151.99%	3.9	16.61①	464238	0.09	60205	6.442	-53.854	1.264
'300127'	银河磁体	115.67%	4.188	48.49%	6.97	51.63①	15125	0.12	3715	15.628	2674.269	1.231
'002547'	春兴精工	3.39%	-22.65	194.55%	5.06	1387.50①	138797	0.0022	243	11.88	-29.4	1.844
'002077'	大港股份	13.59%	-544.514	139.56%	1.58	-10.25①	12437	-0.21	-12000	8.164	-4.667	1.137
'002002'	鸿达兴业	18.03%	-24.613	41.91%	2.74	21.98①	140526	0.0737	18999	5.598	306.038	1.772
'000795'	英洛华	89.25%	82.846	61.79%	4.14	72.32①	48717	0.028	3140	6.176	944.23	1.286
'002467'	二六三	4.72%	165.845	132.70%	4.56	58.88①	26246	0.0273	3692	6.528	26.641	1.689
'000831'	五矿稀土	24.62%	-54.562	127.47%	8.27	357.54①	40990	0.0126	1235	16.822	130.287	1.835
'600198'	大唐电信	16.49%	-1.77	92.16%	271.88	-19.69①	34724	-0.177	-15585	14.97	-6.38	1.685
'300033'	同花顺	-2.64%	33.248	37.83%	14.13	109.78①	28664	0.19	10039	85.008	-40.788	1.214
'603679'	华体科技	0.64%	102.278	17.81%	7.33	48.08①	13227	0.23	2350	46.454	-12.382	1.553
'600460'	士兰微	-4.89%	-25.488	51.65%	6.21	226.94①	66020	0.018	2320	16.448	-65.157	0.84
'300397'	天和防务	-4.67%	272.245	65.51%	4.47	55.53①	16463	0.09	2237	21.058	-51.797	1.265
'300184'	力源信息	22.40%	-29.158	176.61%	2.15	32.10①	281010	0.0651	7674	9.084	-60.911	1.342
'600259'	广晟有色	11.67%	448.3	91.17%	7.85	184.25①	55927	0.06	1925	40.24	25.346	1.348
'603186'	华正新材	6.10%	31.086	67.86%	5.34	57.60①	43109	0.12	1496	27.348	-50.066	1.373
'600366'	宁波韵升	56.63%	432.951	42.84%	2.13	22.49①	36226	0.108	10644	7.99	556.734	1.492
'300325'	德威新材	-13.14%	-299.238	104.29%	4.59	-50.67①	24776	-0.03	-3358	5.948	-11.241	0.914

'300292'	吴通控股	4.15%	61.807	151.47%	5.66	29.25①	101845	0.06	7954	7.63	-66.784	0.993
'000969'	安泰科技	11.49%	464.178	34.10%	2.07	40.95①	125426	0.0533	5467	8.298	169.598	1.618
'600206'	有研新材	48.04%	532.288	33.48%	3.6	184.71①	160261	0.017	1472	9.445	388.031	0.738
'000066'	中国长城	-6.88%	-65.758	26.40%	4.23	125.97①	191962	0.018	5187	9.298	-52.675	0.743
'603701'	德宏股份	8.56%	-18.414	127.11%	4.48	45.11①	11962	0.0777	1587	15.868	125.074	1.578
'002902'	铭普光磁	64.31%	15.441	158.39%	5.28	149.66①	37973	0.0445	935	26.359	405.057	1.317
'300399'	京天利	19.62%	357.679	103.97%	7.34	102.13①	13243	0.04	759	16.258	13.87	0.994
'000993'	闽东电力	2.26%	142.077	96.81%	2.05	37.57①	13825	0.0571	2615	8.972	-19.17	1.235
'300046'	台基股份	11.31%	-0.726	42.05%	4.7	46.92①	8315	0.108	2302	19.834	-37.523	1.349
'300638'	广和通	5.14%	476.252	93.57%	12.26	36.77①	41513	0.317	3845	47.026	-33.198	1.455
'300615'	欣天科技	113.22%	6.891	231.71%	11.87	463.14①	5590	0.0156	297	20.908	533.638	1.432
'600410'	华胜天成	-6.51%	929.452	76.82%	2.38	14.99①	115148	0.175	19295	10.786	-41.177	1.1
'600980'	北矿科技	55.20%	160.892	47.79%	4.64	67.48①	12346	0.0652	1006	14.026	1106.739	1.068
'002378'	章源钨业	18.80%	7.587	63.30%	3.94	276.95①	45198	0.0077	714	8.262	975.401	1.504
'600745'	闽泰科技	4.97%	175.07	17.38%	6.11	124.46①	488638	0.07	4295	35.874	-9.148	1.244
'300079'	数码科技	-0.17%	2.023	40.38%	2.21	36.98①	30902	0.0388	5345	5.624	-31.963	0.834
'600331'	宏达股份	-6.30%	448.124	15.12%	2.89	27.62①	61908	0.0296	6015	3.268	-34.506	0.707
'300352'	北信源	16.28%	204.662	81.48%	3.79	576.92①	11019	0.0026	373	6.1	102.897	0.747
'300205'	天喻信息	19.49%	529.802	32.15%	4.65	23.04①	54283	0.153	6581	14.44	20.764	0.841
'300134'	大富科技	16.04%	-29.755	31.88%	2.17	126.58①	58975	0.03	2230	14.198	23.619	0.571
'300561'	汇金科技	1.67%	4.041	126.05%	5.55	92.96①	5357	0.0387	988	14.555	-9.688	0.566
'300333'	兆日科技	16.29%	-2.299	127.11%	4.73	241.95①	5380	0.0118	397	10.174	155.417	0.763
'002371'	北方华创	-4.30%	29.652	21.72%	8.09	362.01①	70794	0.0435	1991	63.826	-64.556	1.137
'002405'	四维图新	-5.16%	-42.68	39.26%	4.29	184.74①	52193	0.0321	4203	21.94	9.193	1.399
'000970'	中科三环	20.39%	4.806	24.85%	2.91	74.34①	92052	0.0415	4422	10.176	255.263	0.571
'300414'	中光防雷	20.28%	-22.193	44.73%	4.78	118.03①	8719	0.0262	852	11.838	432.343	0.674
...	...	...	...	...	...	...	...	...	...	...	...	...

以上是经过区分度算法排名前 50 的截止到 2019-05-30 的股票。

我们根据量化择时原则和甄选行业的算法，不定期的甄选高增长行业（平均 2-3 周），然后根据区分度模型评分最高的前 20-50 只股票，形成我们的投资组合并根据这个组合生成指数，叫 alpha 指数。目前股票 alpha 指数比上证 180，上证指数，中证 1000，ETF100，ETF 基金，深圳成指数，沪深 300 指数等绝大部分公开指数的收益都高。





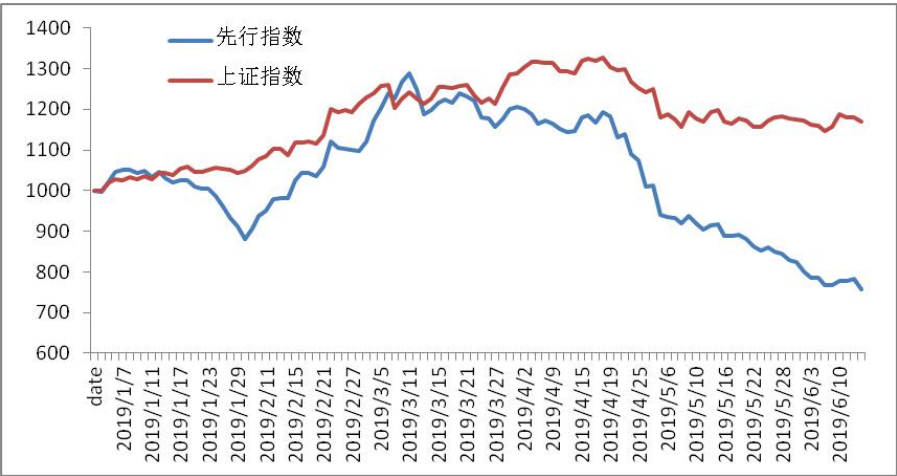
评测报告：

总交易次数	平均收益率	风险收益率	收益标准差
112	0.35	1.14	0.188
正确率	总收益率	夏普比率	最大回撤
58%	0.503	2.67	0.128

4：风险预警及换股策略

4.1：alpha 先行指数

根据区分度模型的评分，我们根据区分度算法排名最靠后的股票，挑选最有可能被高估的股票做成指数（这是 2019 年一季度的模型评分的后 22 名）。该指数的作用在于尽早预测出大盘下跌的可能性，有系统性风险预警作用。





从图上我们可以看出来,先行指数明显早于大盘指数下跌,表明我们的先行指数确有预测作用。所以我们在4月20号以后开始减持了我们的投入,减持的比例我们以凯利公式来计算,此处不做详述。

#### 4.2 : Alpha 换股策略

1 : 换股策略 : alpha 的换股周期取决于宏观择时的 3/7 择时或 2/7 择时原则。减持的比例以凯利公式的计算结果为准,具体换那个股票取决于重新计算的区分度的 SCALINGSCORE 得分,优先换掉哪些得分较低的股票。

2 : 换股周期出现后,区分度的目标变量要重新做一次调整,其他数据以最新数据带入,目标变量的变动如下 :

如果上一个周期的个股涨跌幅-大盘涨跌幅》平均涨跌幅  $gb : =1$

如果上一个周期的个股涨跌幅-大盘涨跌幅》平均涨跌幅  $gb : =0$