CTA 系列二十四：机器学习在单品种的应用—玻璃

发布日期：2022 年 05 月 09 日　　分析师：彭鲸桥　　电话：023-86769675　　投资咨询从业证书号：Z0012925

摘要

- 当前机器学习越来越多的运用到了投资领域，本文的目的也是为了探索机器学习在单品种间的应用，当然，机器学习存在一定的缺陷，其暴力搜索的特性，可能会让整个模型的参数敏感性较高，所以，为了使该探索结果具有一定的可参考性，我们对模型挖掘做了一些限制，希望能降低最后模型挖掘出的因子的过拟合性。

- 在机器学习中，遗传规划是一个优秀的特征生成工具，其优势在于结合基础数据和预算符进行大规模的符号表达式挖掘。正是基于这种特点，遗传规划算法能够突破人类思维局限，挖掘出特异的对现有 CTA 基础策略优异补充的信号。

- 上一篇机器学习文章中，我们探讨了遗传规划在热卷这个品种中的应用，测试效果相对较好，本文将延续上一篇机器学习文章中的方式方法，继续对玻璃这个品种进行发掘，通过一段时间的观察，我们发现玻璃这个品种的日成交量较大，样本数据较长，可交易性较强，符合机器学期的模拟条件，所以通过遗传规划对高频量价数据进行深挖，我们得到了一系列的有效因子。

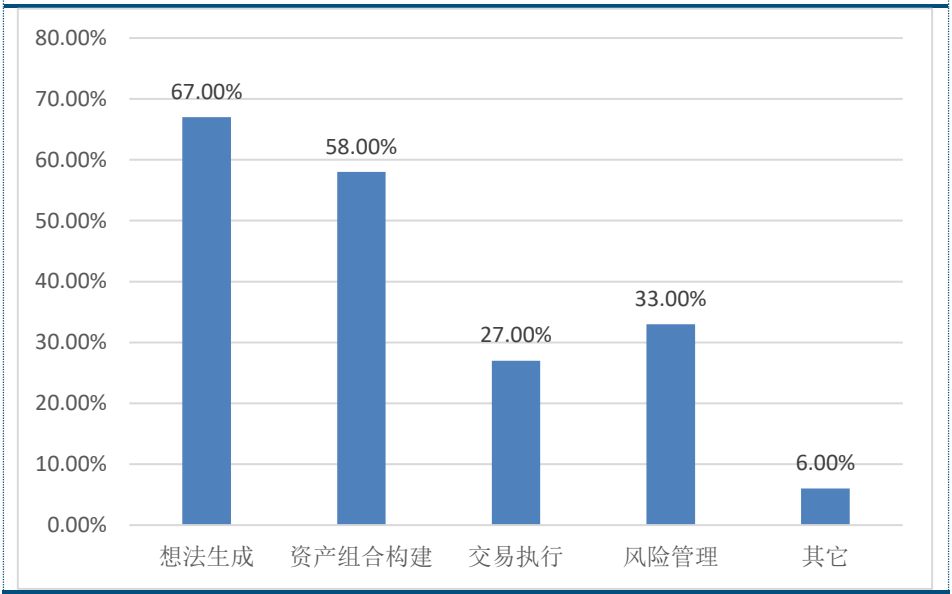
- **风险提示**：本研究主要基于历史数据统计，存策略失效风险、模型误设风险、历史统计规律失效等风险

一、概述

目前国内 CTA 市场上，量化 CTA 的数据利用主要还是以量化数据为主，量化 CTA 策略的核心在于利用不同周期的量价数据挖掘构建信号。遗传规划近年来收到越来越多的关注，在国内外已经成为一个比较热门的量化发展方向。我们将介绍我们在引入机器学习方法到数据挖掘方法上的探索，希望能够更好更高效的从数据中挖掘规律，利用规律。

目前关于 AI 在投资上的应用仍然存在一定争议，这一点我们也认为在可预见的未来一段时间，整个投资任务对于 AI 来说还是过于宽泛复杂了。但是 AI 的长处在于大量繁杂数据的高速处理，这一点已在包括投资领域证明了其巨大价值。正如我们前期报告“CTA 拥抱机器学习之一”中提到，如下图机器学习在投资领域的使用形式。

图 1：巴克莱 2018 人工智能/机器学习 使用形式



数据来源：中信建投期货

我们考虑的是，以人类投资成功经验为基础，利用 AI 作为工具帮助人类去高效完成一些特定环节，以便能够更高效以及相对客观的进行投资。因此，AI 的引入的目的应该是降低低端的人力成本，把研究人员从繁琐的、低边际效益的工作中解放出来，能够投入更多时间和精力来对对市场发展变化以及构建投资逻辑进行更深入的思考。

二、系统介绍

2.1 算法简介

遗传规划算法是优秀的特征生成工具，可以生成具备可解释可理解的显式特征。我们认为，特征的可理解性与可解释性应该是一个有效因子的必要条件。

遗传规划算法涉及的几个关键参数如下：

参数	说明
population_size	每轮多少个个体
generations	生成多少轮(代数)
stopping_criteria	停止进化条件
p_crossover	公式树发生交叉(Crossover)的概率
p_subtree_mutation	子树变异(Subtree Mutation)概率
p_hoist_mutation	子树抬升变异(Hoist Mutation)概率
p_point_mutation	点变异 (Point Mutation)概率
parsimony_coefficient	节俭系数，对复杂公式进行惩罚
random_state	随机种子
metric	评价函数
function_set	算子集合
init_depth	初始公式复杂程度
xdf_set	属性集合
ydf	标签

数据来源：gplearn，中信建投期货

根据遗传规划的运用场景，我们认为以下几个参数需要特别注意。

- **parsimony_coefficient**

模型信号应该是逻辑简洁清晰的，所以我们应当避免使用长度过长因子公式，我们这里对于过长的公式的适应度常用了 sigmoid 的惩罚函数，设置参数如下：

$$penalty = parsimony\_para * \frac{1}{1+\exp(-slop*(func\_length-threshold))}$$

在函数参数中，通过设置 threshold 参数即可对公式长度大于等于阈值的公式较大的适应度惩罚，从而控制得到的公式长度。

- **function_set**

算子集合中，我们也是考虑到公式简洁性需求，我们只选择一些逻辑清楚直接的算子进入模型中，目前考虑算子如下：

参数	说明
neg(x)	x 的相反数
sign(x)	x 的方向
delay(x, d)	d 天以前的 x 值
delta(x, d)	过去 d 天 x 的变化值
pct_change(x, d)	过去 d 天 x 的变化率。
ema(x, d)	span=d 天 x 构成的指数加权均值。
kama(x, d)	er_para=d x 构成的 kama 加权均值
ts_sum(x, d)	过去 d 天 x 值构成的时序数列之和。
ts_mean(x, d)	过去 d 天 x 值构成的时序数列均值。
ts_wmean(x, d)	过去 d 天 x 值构成的时序数列之线性加权均值。
ts_median(x, d)	过去 d 天 x 值构成的时序数列之中位数
ts_min(x, d)	过去 d 天 X 值构成的时序数列中 a 最小值
ts_max(x, d)	过去 d 天 X 值构成的时序数列中最大值
ts_argmin(x, d)	过去 d 天 x 值构成的时序数列中最小值出现的位置
ts_argmax(x, d)	过去 d 天 x 值构成的时序数列中最大值出现的位置
ts_arg_maxmin(x, d)	过去 d 天 x 值构成的时序数列之最大最小值位置之差
ts_maxmin_norm (x, d)	当前 x 值处于过去 d 天最大最小值区间相对位置
ts_zscore(x, d)	过去 d 天 x 值构成的时序数列 zscore
ts_rank(x, d)	过去 d 天 x 值构成的时序数列中当前 x 值所处分位数
ts_mean_return(x, d)	过去 d 天 x 值构成的时序数列之最大最小值位置之差
ts_cov(x,y,d)	过去 d 天 x 值构成的时序数列与 y 构成的时序数列的协方差
ts_corr(x,y,d)	过去 d 天 x 值构成的时序数列与 y 构成的时序数列的相关系数
ts_beta(x,y,d)	过去 d 天 x 值构成的时序数列与 y 构成的时序数列的 beta
ts_dema(x, d)	过去 d 天 x 值的双移动平均线，DEMA = 2 * EMA - EMA(EMA)
ts_midpoint(x,d)	过去 d 天 X 值构成的时序数列的最大值与最小值的平均值
ts_linearreg_angle(x,d)	过去 d 天 x 值序列为因变量，序列 1,..,d 为自变量的线性回归角度
ts_linearreg_intercept(x,d)	过去 d 天 X 值序列为因变量，序列 1,..,d 为自变量的线性回归截距
ts_linearreg_slope(x,d)	过去 d 天 X 值序列为因变量，序列 1,..,d 为自变量的线性回归斜率
ts_alpha(x, y, d)	过去 d 天 x 值构成的时序数列与 y 构成的时序数列的 alpha
s_log(x)	protected signed log（可对负数求对数，保留符号）

数据来源：中信建投期货

而算子参数中，回溯参数参数选择上我们按照逻辑含义，分别选择交易时点的的 20 根 K 线整数倍。

- **信号构建**

假设得到公式指标数据序列 Si(i=1,...,N),设置回溯参数为自然日 365/730 天...(1/2 年...),当前信号为 Si， 回溯期信号为 Sr(r=i-T...i)。

(1) 若当前信号 Si 向上突破回溯期 Sr 85 分位处，发出看多信号；当前信号状态为多，且 Si 向下突破 70 分位处，发出平多信号。

(2) 若当前信号 Si 向下突破回溯期 Sr 15 分位处，发出看空信号；当前信号状态为空，且 Si 向上突破 30 分位处，发出平空信号。

2.2 开发流程

- **Step1：数据清洗**
  a) 获取原始量价数据，确定数据的时间标签，空值处理方法。。。
  b) 根据最迟数据获取时间确定信号产生的时间；

- **Step2：遗传算法挖掘因子**
  a) 初始化种群，计算每个个体适应度，如果存在适应度无法计算的个体，则淘汰个体，重新生成新的个体；
  b) 种群逐个个体开始交叉、变异进行进化形成新的种群，个体在进化中若进化与上一代的最优个体相同或出现适应度无法计算情况，则该个体重新进化；记录新的种群中适应度最好的个体；
  c) 若出现 M 代种群中最佳个体不变的情况，则提高进化的变异概率，吸收更多新的基因进入跳出局部最优点；
  d) 若 M>3，则重新生成新的随机种群开始新的一轮进化。
  e) 一轮进化完成后，对每一代最佳个体进行评估，选择逻辑较为简洁，公式长度<=7 的因子。

三、开发范例

3.1 测试条件

在模拟交易过程中，我们选择 15 分钟作为信号周期，我们交易价格采用上期所热轧板卷主力合约的开盘价作为成交价格，交易时间选择技术面信号触发的后一天。测试时段选择 2016 年 1 月 1 日至 2021 年 6 月 28 日。

原始数据以及预处理分类如下：

参数	说明
OPEN	开盘价
HIGH	最高价
LOW	最低价
CLOSE	收盘价
PCT_CHANGE	收益率
VOLUME	成交量
AMOUNT	成交额
OPEN_INTEREST	持仓量
MFI	资金流指标(参数 20, 80)

数据来源：中信建投期货

3.2 信号测试

部分待评估的表现较好因子公式

公式编号	公式表达式
1	ts_linearreg_intercept(ts_mean(delta(ts_max(ts_argmin(LOW,20),20),20),80),80)
2	ts_wmean(ts_sum(ts_beta(ts_linearreg_slope(LOW,150), OPEN,80),150),80)
3	ts_dema(ts_max(ts_wmean(ts_arg_maxmin(ts_dev(CLOSE,150),20),20),20),80)
4	ts_dema(ts_linearreg_intercept(ts_max(ts_argmin(LOW,20),20),80),80)
5	ts_mean(ts_sum(ts_beta(ts_linearreg_slope(LOW,150), OPEN,80),150),80)
6	ema(kama(kama(ts_maxmin_norm(ts_maxmin_norm(ts_max(neg(LOW),20),20),20),20),20),20)
7	ts_mean(ts_sum(ts_beta(ts_linearreg_slope(LOW,150), HIGH,80),150),80)
8	kama(ts_dema(ts_dema(ts_arg_maxmin(CLOSE,20),150),20),20)
9	ts_dema(s_sqrt(ts_max(neg(ts_linearreg_angle(OPEN,20)),20)),80)
10	ts_wmean(ts_arg_maxmin(ts_midpoint(ema(ts_argmin(LOW,20),20),20),20),150)
11	ts_wmean(s_sqrt(s_sqrt(ts_argmax(ts_midpoint(CLOSE,20),20))),150)
12	ema(ts_max(kama(ts_arg_maxmin(ts_dev(CLOSE,150),20),20),20),80)
13	ts_min(ts_wmean(ema(ts_zscore(ts_arg_maxmin(square(ts_dev(neg(OPEN),150)),20),80),150),20),20)
14	ts_mean(ts_arg_maxmin(ts_dema(ts_wmean(OPEN,20),80),20),20)
15	kama(s_curt(s_curt(ts_maxmin_norm(ts_arg_maxmin(HIGH,20),150))),150)
16	ema(ts_argmin(ts_std(kama(CLOSE,20),150),150),80)
17	ts_dema(ema(s_sqrt(ts_arg_maxmin(CLOSE,20)),80),20)

数据来源：中信建投期货

以上信号测试结果如下图所示：

图 2：公式 1 净值图

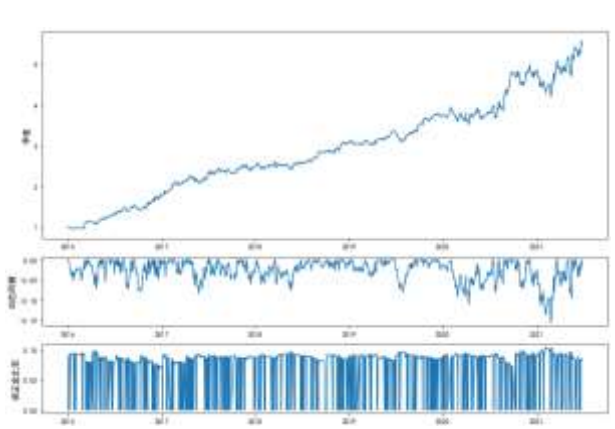


图 3：公式 2 净值图

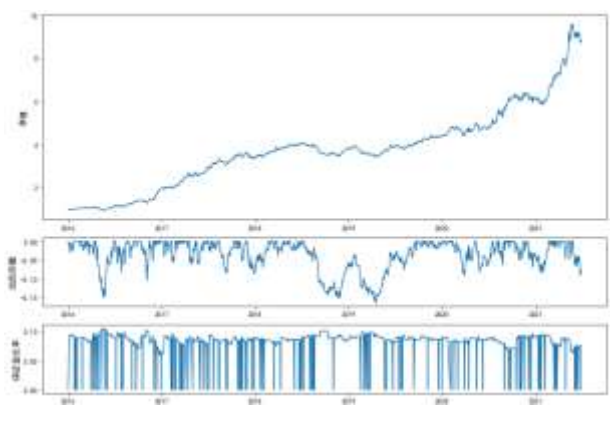


图 4：公式 3 净值图

图 5：公式 4 净值图

图 6：公式 5 净值图

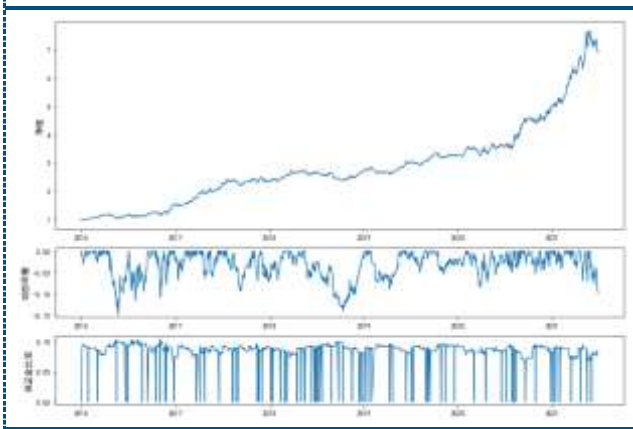


图 7：公式 6 净值图

图 8：公式 7 净值图

图 9：公式 8 净值图

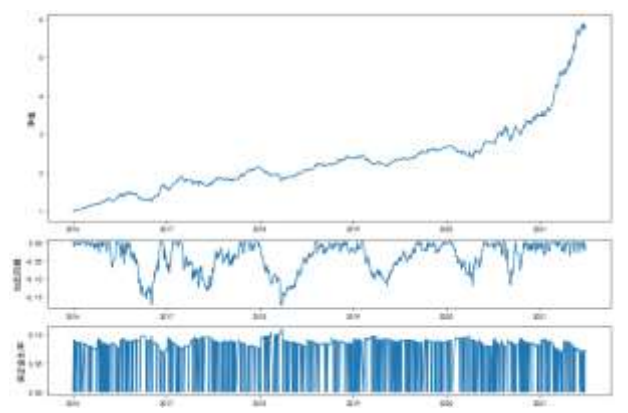


图 10：公式 9 净值图

图 11：公式 10 净值图

图 12：公式 11 净值图

图 13：公式 12 净值图

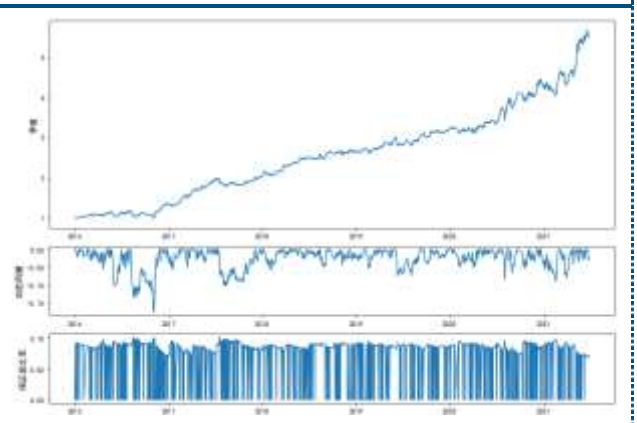


图 14：公式 13 净值图

图 15：公式 14 净值图

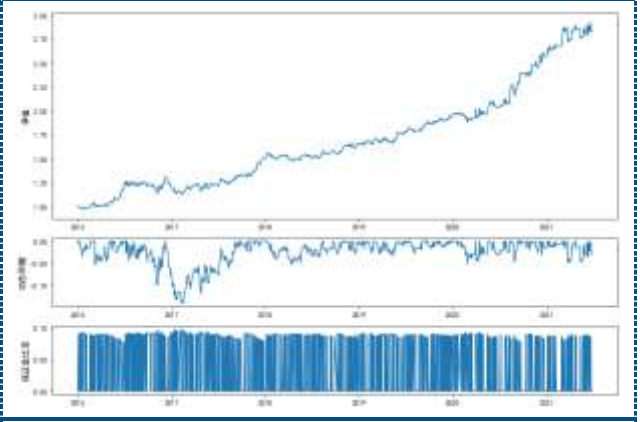


图 16：公式 15 净值图

图 17：公式 16 净值图

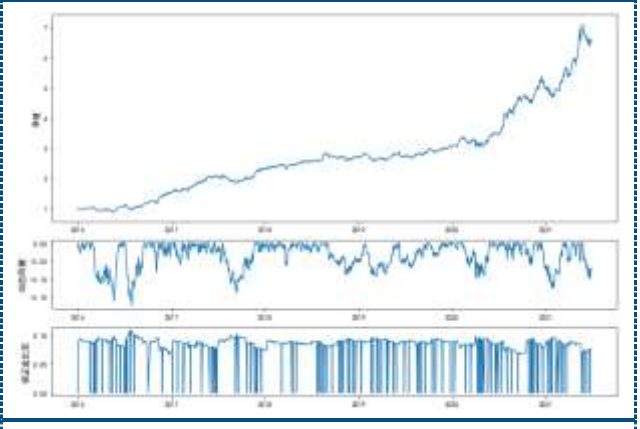


图 18：公式 17 净值图

数据来源：中信建投期货

公式	总收益	年化收益	波动率	最大回撤	夏普比率	卡玛比率	最大回撤周期
公式 1	449.05%	36.35%	19.48%	15.68%	1.763	2.317	109

公式 2	776.05%	48.45%	19.83%	16.54%	2.341	2.930	293
公式 3	639.87%	43.95%	18.20%	26.99%	2.304	1.628	109
公式 4	436.47%	35.77%	19.07%	20.59%	1.771	1.737	149
公式 5	590.38%	42.15%	20.27%	14.81%	1.981	2.845	156
公式 6	223.87%	23.85%	18.62%	15.50%	1.173	1.538	171
公式 7	319.40%	29.82%	16.61%	15.03%	1.674	1.983	195
公式 8	480.94%	37.75%	17.53%	17.32%	2.039	2.180	151
公式 9	443.39%	36.09%	18.20%	22.83%	1.873	1.580	109
公式 10	359.09%	31.98%	17.75%	18.16%	1.689	1.761	159
公式 11	301.92%	28.82%	18.02%	12.59%	1.489	2.289	136
公式 12	449.46%	36.36%	18.36%	17.79%	1.872	2.043	95
公式 13	364.56%	32.26%	16.34%	19.90%	1.851	1.621	165
公式 14	184.00%	20.93%	13.63%	14.22%	1.388	1.472	189
公式 15	378.23%	32.96%	16.22%	14.78%	1.909	2.231	139
公式 16	553.96%	40.76%	20.05%	17.51%	1.932	2.328	180
公式 17	244.23%	25.24%	16.97%	18.88%	1.369	1.337	237

数据来源：中信建投期货

3.3 信号合成

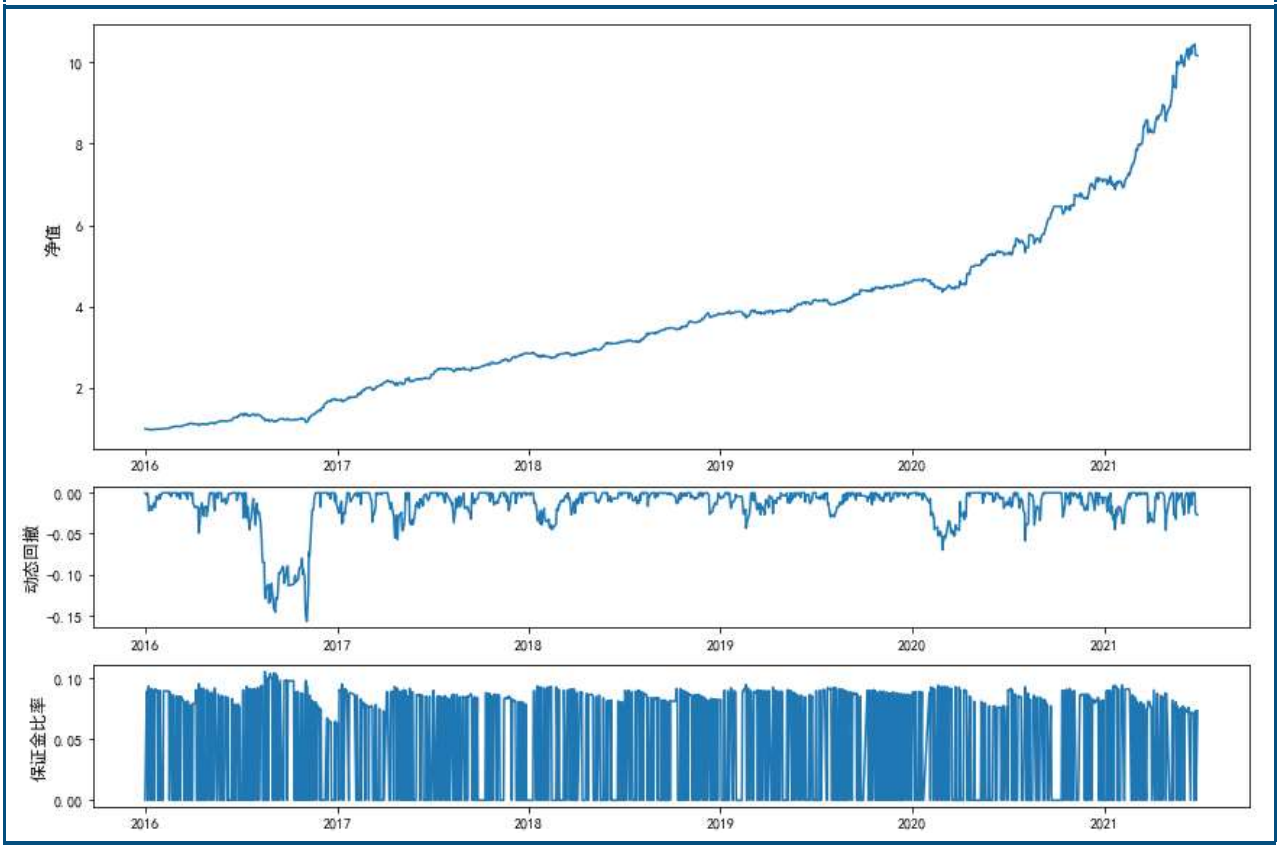
为了尽量保持多空区间一致，我们将信号区间均匀分布为 3 个区间：

信号阈值 = round(因子个数/3)

信号求和<(-信号阈值)=-1

信号求和>信号阈值=1

图 19：合成净值图



	总收益	年化收益	波动率	最大回撤	夏普比率	卡玛比率	最大回撤周期
因子合成	915.54%	52.50%	14.79%	15.77%	3.413	3.328	87

品种	交易次数	做多次数	做空次数	做多占比	做空占比
FG	965	522	443	32.44%	21.21%

四、总结

在上一篇文章中，我们对遗传规划在热卷中的应用进行了探讨，取得了比较好的绩效水平，本文继续在黑色板块中进行了探索，玻璃虽然属于传统上的建材，但从产业链上下游关系来看，其与黑色板块的品种存在较强的联系，上游受到煤的影响，下游房地产与钢矿也存在交叉，所以我们决定对玻璃进行挖掘，从最终的测试结果来看，玻璃的绩效水平较热卷相对稳定，这可能与玻璃自身波动率较低有关，从测试结果中，我们可以发现，净值在 2020 年以后增长较快，原因是玻璃近几年

玻璃的行情较以往更大，也比较符合市场特征，从信号分布情况来看，玻璃的交易次数相对较高，多空分布相对合理。后续我们将继续探索黑色板块中的其他品种。

联系我们

中信建投期货总部

地址：重庆市渝中区中山三路107号上站大楼平街11-B，名义层11-A，8-B4,C

电话：023-86769605

中信建投期货有限公司上海分公司

地址：中国（上海）自由贸易试验区浦电路 490 号，世纪大道 1589 号 8 楼 10-11 单元

电话：021-68765927

中信建投期货有限公司湖南分公司

地址：长沙市芙蓉区五一大道 800 号中隆国际大厦 903

电话：0731-82681681

南昌营业部

地址：南昌市红谷滩新区红谷中大道 998 号绿地中央广场 A1#办公楼-3404 室

电话：0791-82082702

中信建投期货有限公司河北分公司

地址：廊坊市广阳区吉祥小区 20-11 门市一至三层、20-1-12 号门市第三层。

电话：0316-2326908

漳州营业部

地址：漳州市龙文区九龙大道以东漳州碧湖万达广场 A2 地块 9 幢 1203 号

电话：0596-6161588

西安营业部

地址：西安市高新区高新路 56 号电信广场裙楼 6 层北侧 6G

电话：029-89384301

北京朝阳门北大街营业部

地址：北京市东城区朝阳门北大街 6 号首创大厦 207 室

电话：010-85282866

北京北三环西路营业部

地址：北京市海淀区中关村南大街 6 号 9 层 912

电话：010-82129971

武汉营业部

地址：武汉市江汉区香港路 193 号中华城 A 写字楼（阳光城 • 央座）1306/07 室

电话：027-59909521

中信建投期货有限公司杭州分公司

地址：杭州市上城区庆春路 137 号华都大厦 811、812 室

电话：0571-28056983

太原营业部

地址：太原市小店区长治路 103 号阳光国际商务中心 A 座 902 室

电话：0351-8366898

北京国贸营业部

地址：北京市朝阳区光华路 8 号和乔大厦 A 座向东 20 米

电话：010-85951101

中信建投期货有限公司济南分公司

地址：济南市历下区泺源大街 150 号中信广场 A 座六层 611、613 室

电话：0531-85180636

中信建投期货有限公司大连分公司

地址：辽宁省大连市沙河口区会展路 129 号大连国际金融中心 A 座大连期货大厦 2901、2904、2905 室

电话：0411-84806316

中信建投期货有限公司河南分公司

地址：郑州市未来大道 69 号未来大厦 2205、2211、1910 房

电话：0371-65612397

广州东风中路营业部

地址：广州市越秀区东风中路 410 号时代地产中心 20 层自编 2004-05 房

电话：020-28325286

重庆龙山一路营业部

地址：重庆市渝北区龙山街道龙山一路 5 号扬子江商务小区 4 幢 24-1

电话：023-88502020

成都营业部

地址：成都市武侯区科华北路 62 号（力宝大厦）1 栋 2 单元 18 层 2、3 号

电话：028-62818701

中信建投期货有限公司深圳分公司

地址：深圳市福田区深南大道和泰然大道交汇处绿景纪元大厦 11I

电话：0755-33378759

上海徐汇营业部

地址：上海市徐汇区斜土路 2899 甲号 1 幢 1601 室

电话：021-64040178

南京营业部

地址：南京市黄埔路 2 号黄埔大厦 11 层 D1、D2 座

电话：025-86951881

中信建投期货有限公司宁波分公司

地址：浙江省宁波市鄞州区和济街 180 号国际金融中心 F 座 1809 室

电话：0574-89071681

合肥营业部

地址：合肥市包河区马鞍山路 130 号万达广场 C 区 6 幢 1903、1904、1905 电话：0551-2889767

广州黄埔大道营业部

地址：广州市天河区黄埔大道西 100 号富力盈泰大厦 B 座 1406

电话：020-22922102

上海浦东营业部

地址：上海自由贸易试验区世纪大道 1777 号 3 楼 F1 室

电话：021-68597013

重要声明

本报告中的信息均来源于公开可获得资料，中信建投期货力求准确可靠，但对这些信息的准确性及完整性不做任何保证，据此投资，责任自负。本报告不构成个人投资建议，也没有考虑到个别客户特殊的投资目标、财务状况或需要。客户应考虑本报告中的任何意见或建议是否符合其特定状况。

全国统一客服电话：400-8877-780

网址：www.cfc108.com