

# 股指单边做多策略：非线性择时模型中的因子线性优化

2020年2月19日 星期三

兴证期货·研发中心

金融工程研究团队

韩惊

从业资格编号：F3010931

投资咨询编号：Z0012908

王锴

从业资格编号：F3063220

联系人

王锴

021-68982740

[wangkai@xzfutures.com](mailto:wangkai@xzfutures.com)

## ● 内容提要

回顾上一期报告《股指期货量化展期与择时对冲》中基于‘中低频基本面因子+量价指标’的择时模型，**非线性的 boosting 和 bagging 算法相较于传统多因子模型在特征选择和构建方面有显著优势**，择时对冲的胜率也提高到 65% 以上。通过进一步分析模型中各大类因子对于因变量的贡献度，我们发现相较于由宏观、估值和资金指标等组成的中低频因子，由日度平滑数据构成的市场行情指标在因子重要性排名中总体处于较低水平，并且在单因子测试当中预测正确率也不如基本面因子。尽管这些指标在数量上并不占多数，但是我们在因子构建过程中已经将时间序列上的滞后多阶差分转置为新的横截面因子。我们推测，基于决策树的非线性模型在构建过程中丢失了部分因子间存在的线性关系，并因此错过某些短线的交易机会。**本文旨在讨论如何将线性关系融入到非线性模型中，从而提高单边做多的择时收益。**

## ● 模型介绍

我们将原有的 Adaboost 择时模型由周频推广到日频信号。为了弥补高频数据对于中长线基本面因子的弱化作用并进一步提升模型的预测能力，我们考虑优化部分因子，把更为有效的关系加入到模型，同时考虑是否对预测结果有提升作用并具有可解释性。**技术面指标作为原始素材显然符合这方面的要求。**由于线性变换后的因子共用了某些变量作为动态参考系，在建模过程中需要考虑多重共线性，而非线性特征选择的模型则可以有效地避免这一类问题。

## ● 实证分析

加入线性化因子的 boosting 模型在保持回撤基本不变的情况下对比基准策略，以 IF 合约为标的的择时收益提升近 20%，IC 合约为标的的择时收益提升 34%。**证实线性化因子与非线性模型结合使用的有效性。**

**标题目录**

1.原有模型分析.....	3
1.1 单边做多策略下的局限性及其原因 .....	3
2.技术指标的线性优化.....	4
2.1 因子构成 .....	4
2.2 非线性特征选择 .....	5
3.模型构建.....	5
2.1 算法特点 .....	5
2.2 交易策略 .....	6
4.实证研究.....	6

**图目录**

图 1: 单位决策树.....	3
图 2: Adaboost 算法流程.....	4
图 3: IF 合约策略净值曲线 .....	6
图 4: IC 合约策略净值曲线.....	7

**表目录**

表 1: 部分因子重要性排名 .....	3
表 2: 线性化过程举例.....	4
表 3: IF 合约策略收益表现 .....	6
表 4: IC 合约策略收益表现.....	7

## 1. 原有模型分析

### 1.1 单边做多策略下的局限性及其原因

上一期报告中我们将包含技术、资金、宏观、估值，市场情绪、期货特色指标等方面的共计 64 个因子放入到 Adaboost 模型中，以周频信号的方式预测标的指数的跌幅是否会超过阈值。但是，如果我们将该模型应用于单边做多策略，显然周频信号会损失一些调仓间歇期内的交易机会。因此我们首先考虑将输出信号改为日频，胜率表现并不理想。我们对日频的模型进行集成学习算法的重要性排名，并将其与周度模型下的排名比较：

表 1：部分因子重要性排名

周频模型		日频模型	
排名	因子	排名	因子
1 ~ 20	Libor 一周、指数成分股上涨/下降比例	20 ~ 30	开盘主力净流入量, 主力持仓量变化、OBV 能量潮
.....	.....	.....	.....
40 ~ 50	开盘主力净流入量、换手率	30 ~ 40	Libor 一周、融资融券余额

数据来源：Wind，兴证期货研发部

显然，在周频模型中平滑日度数据的市场行情指标在因子重要性排名中总体处于较低水平，而在日频模型中，由宏观市场、估值指标等组成的中长线因子的重要性显著下滑，市场行情指标有所提升。因此，如何进一步优化以量价指标为代表的短线因子对于提升单边择时收益至关重要。

## 2. 技术指标的线性化

### 2.1 因子构成

技术指标分析自上世纪以来越来越广泛地被运用到投资交易当中，至今发展 200 多个种类，又可分为‘趋势型’、‘能量型’、‘超买超卖型’等多个大类，为交易市场的中短线操作提供了很好的指导作用。我们考虑在因子放入模型前进行线性变换并生成新的因子，这样更为丰富的线性关系已经被代入到了非线性模型中。如果我们把这一概念运用到量价指标中，那么就必须要考虑是否对预测结果有提升作用并具有可解释性。然而在量化交易尤其是程序化层面，技术指标也仅仅是进行简单的转译之后加入到条件判断中，也很少出现技术指标加入到机器学习模型中的案例。究其原因，还是因为技术指标的影响并非依赖于其绝对数值，而是往往要与一个同样动态变化的参照系进行比较，从而得到对因变量直接作用的关系。例如常用的 KDJ 线需要比较 K 线和 D 线的相对大小，如 K 线上穿 D 线为买入。反之则为卖出信号。单独的 K 线数值对于解释价格并没有实际含义，对模型预测能力也没有提升作用。因此，我们需要对部分技术指标进行线性变换。

表 2：部分线性化过程

原技术指标	线性化的因子
KDJ	$\text{Log}(K/D), \text{log}(K/J), \text{log}(D/J)$
BOLL	P/upper_bound、P/lower_bound
MACD	$\text{Log}(DIF/DEA)$

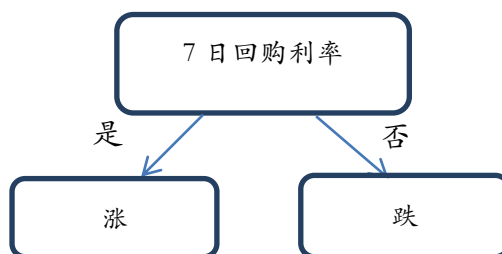
数据来源：Wind，兴证期货研发部

转换后的因子量化了技术指标中的相对关系。通过对一些一阶滞后变量取 log，就能反映出实际指标的特殊含义。如  $\text{log}(K_t/D_t)$  为正， $\text{log}(K_{t-1}/D_{t-1})$  为负，则代表了 KDJ 线的买入信号。

## 2.2 非线性特征选择

虽然有大量技术指标可以作为自变量放入模型中，但是数据维度的增加会带来过拟合的问题。仍然需要通过特征选取的方式来降低维度，选择有效因子。由于线性变换构建的因子包含了某些共同的变量，如价格、D 线数值等等，那么在建模过程就需要考虑多重共线性的问题，即因子间较强的自相关性会导致无法区分各自对因变量产生的影响。常见的因子选择方法包括了相关性检验、方差信息熵检验以及逐步递回归法在选择过程都无法有效避免这一问题。然而，基于决策树的集成学习算法则不用考虑这个问题。单位决策树是一个非线性区间的识别过程，以 7 日逆回购利率为例：

图 1：单位决策树



数据来源：兴证期货研发部

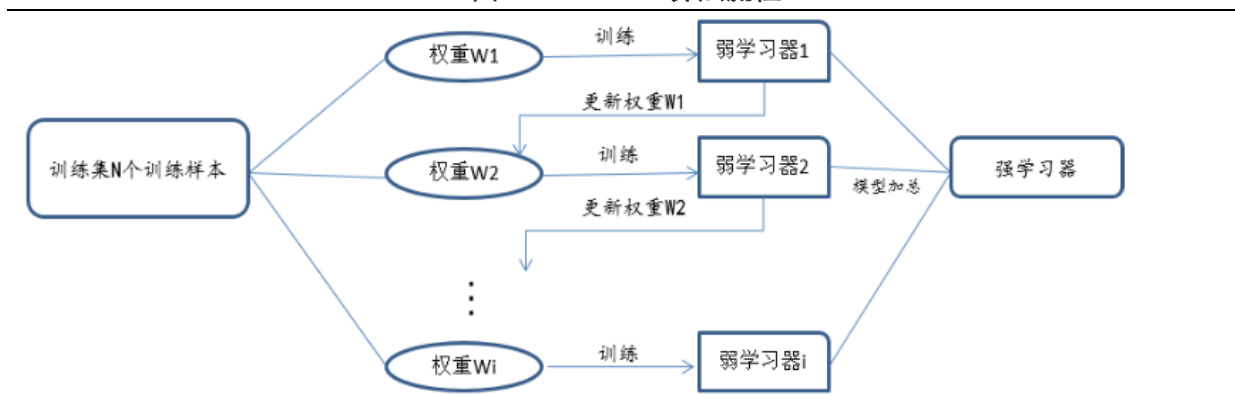
假设一个较为极端的情况，自变量 A 和 B 具有完全一样的特征，假设 A 对解释因变量有显著提升已经被放入到模型中，由于另一个变量 B 会达到完全一致的分类结果，将不会加入模型。因此，集成学习算法的模型可以有效避免这一类的问题。

### 3. 模型构建

#### 3.1 算法特点

由于 boosting 算法具有较强的泛化能力，结合以往的检验分析，我们仍然以 Adaboost 算法为基础：

图 2: Adaboost 算法流程



数据来源：兴证期货研发部

在构建过程中，Adaboost 模型根据错误率不断调整样本的权值，提高每个分类正确的弱分类器的权重，并以此循环进行迭代计算，后一个函数参数依据前一轮模型的结果生成，因此具有以下优势：

- 结合多个弱分类器，不易出现过拟合现象，模型较稳健。
- 具有处理非线性因子能力。
- 通过自适应改变样本权重实现高相关性因子的自动剔除。
- 不需要对特征进行人工筛选，可以加入大量因子

对于以 IF 合约为预测标的的模型，我们采用了共计 69 个因子以及对应的一阶和二阶差分，其中就包括了资金、宏观、估值以及新构建的线性化技术指标。对于以 IC 合约为预测标的的模型，我们采用了类似的 62 个因子以及对应的一阶和二阶差分。因变量方面，将标的次日涨幅是否超过 1.5%作为阈值进行判断，若超过则记为正号，反之则为负号。

对于 Adaboost 中所用到的参数，如

- 迭代次数
- 决策树终端节点数量、深度
- 权重缩减系数
- 划分的最大特征数
- ...

我们首先采用交叉验证(cross validation)的方法确定参数范围，即对随机抽样的子集进行循环测试，选取平均误差较小的参数组合，在限制过拟合的同时，保证模型具有泛化能力。

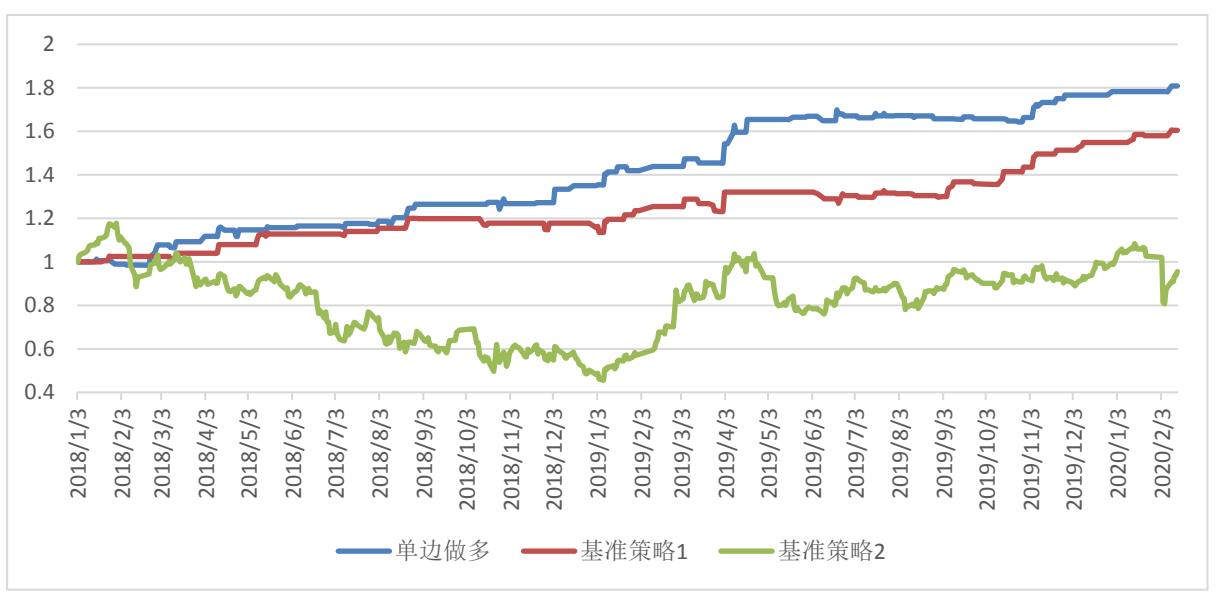
### 3.2 交易策略

每日开盘后预测下一交易日的走势。如果已经多开，预测为正号，则继续持有；预测为负号，则平仓。此外，如果当月合约将在 3 天后到期则使用次月合约开仓或展期到次月。

## 4. 实证研究

我们将回测参数设置为 0.03%冲击成本，0.002%手续费，20%保证金。回测区间为 2018 年 1 月至 2020 年 1 月的 IF 和 IC 合约。其中，单边做多策略即本文讨论的线性化技术指标的模型，基准策略 1 为不加入线性化技术指标的原 Adaboost 模型，基准策略 2 为持有合约多头的非择时策略。

图 3: 以 IF 合约为标的策略净值曲线



数据来源：Wind，兴证期货研发部

表 3: 以 IF 合约为标的策略收益表现

	累计收益率	年化收益率	最大回撤	夏普比率	择时胜率	开仓次数
单边做多	79.30%	34.90%	9.01%	2.10	69.45%	58
基准策略 1	60.12%	26.14%	8.45%	1.89	54.19%	56
基准策略 2	-5.12%	-1.96%	61.35%	-0.23	-	-

数据来源：Wind，兴证期货研发部

图 3：以 IC 合约为标的策略净值曲线



数据来源：Wind，兴证期货研发部

表 4：以 IC 合约为标的策略收益表现

	累计收益率	年化收益率	最大回撤	夏普比率	择时胜率	开仓次数
单边做多	66.59%	32.12%	19.01%	1.38	65.36%	67
基准策略 1	32.07%	14.96%	20.45%	1.06	59.12%	62
基准策略 2	-25.18%	-12.96%	61.35%	-0.19	-	-

数据来源：Wind，兴证期货研发部

实证研究显示加入线性变换的技术面指标后对于日频模型的预测能力提升显著。IF 合约方面，新的单边做多策略在有效控制回撤的情况下较基准策略 1 收益提升近 20%，开仓次数略少于基准策略，择时胜率有所提高。IC 合约方面，新的单边做多策略在控制回撤的情况下较基准策略 1 收益提升 34%，开仓次数 67 次略高于基准策略。总体来看，在 5 倍杠杆下新的单边做多策略对于 IF 合约能够保持较低的回撤，风险收益比较高。做多策略对于 IC 合约存在一定的回撤风险，我们推测这可能与标的指数本身的波动幅度有一定关系，但仍然证实了在原有非线性模型中加入线性化因子对预测指数走势有显著的提升作用，未来也可以推广到其它类别的因子和指数板块。

### 分析师承诺

本人以勤勉的职业态度，独立、客观地出具本报告。本报告清晰准确地反映了本人的研究观点。报告所采用的数据均来自公开资料，分析逻辑基于本人的职业理解，通过合理判断的得出结论，力求客观、公正，结论，不受任何第三方的授意影响。本人不曾因也将不会因本报告中的具体推荐意见或观点而直接或间接接收到任何形式的报酬。

### 免责声明

本报告的信息均来源于公开资料，我公司对这些信息的准确性和完整性不作任何保证，也不保证所包含的信息和建议不会发生任何变更。文中的观点、结论和建议仅供参考。兴证期货可发出其它与本报告所载资料不一致及有不同结论的报告。本报告及该等报告反映编写分析员的不同设想、见解及分析方法。报告所载资料、意见及推测仅反映分析员于发出此报告日期当日的独立判断。

客户不应视本报告为作出投资决策的惟一因素。本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。本公司未确保本报告充分考虑到个别客户特殊的投资目标、财务状况或需要。本公司建议客户应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。

在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的损失负任何责任。

本报告的观点可能与资管团队的观点不同或对立，对于基于本报告全面或部分做出的交易、结果，不论盈利或亏损，兴证期货研究发展部不承担责任。

本报告版权仅为兴证期货有限公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。如引用、刊发，需注明出处兴证期货研究发展部，且不得对本报告进行有悖原意的引用、删节和修改。