证券研究报告

兴业证券 INDUSTRIAL SECURITIES

分析师：
于明明
yumingming@xyzq.com.cn
S0190514100003

研究助理：
宫民
gongmin@xyzq.com.cn
S0190119020038

系统化资产配置系列之三：基于 AdaBoost 机器学习算法的市场短期择时策略

2019 年 10 月 17 日

报告关键点

本篇是系统化资产配置系列报告的第三篇，对如何利用机器学习算法进行短期市场择时进行了系统介绍。本报告中，我们将股市未来的涨和跌定义为一个分类问题，利用机器学习算法来对 Wind 全 A 指数的未来涨跌建模。基于决策树的 AdaBoost 算法解决了构建择时模型面临的多个问题，它实现了：1.有效因子的自动筛选。2.非线性因子的建模。3.通过自适应调整样本权重解决因子相关性问题。4.不易过拟合，模型较稳健。最后，我们通过叠加水晶球择时模型，充分利用期权市场信息，形成了表现更加出色的双塔奇兵择时模型。

相关报告

《系统化资产配置系列之二：行业的重新分类以及行业轮动策略》2019-09-19

《跨资产的系统性配置策略之一：另类风险溢价的分类以及系统化的配置方法》2019-06-28

《抽丝剥茧，去芜存菁：水晶球择时模型之 3.0》2018-09-26

团队成员：

投资要点

- 本篇是系统化资产配置系列报告的第三篇，对如何利用机器学习算法进行短期市场择时进行了系统介绍。全球金融市场每天产生海量的各类数据，如何筛选并有效利用这些数据来预测股票市场走势一直是一个重要但棘手的问题。短期择时面临的主要困难包括：1.短期市场走势受情绪等因素影响较大；2.如何筛选有效因子；3.非线性因子如何建模；4.因子相关性问题如何解决；5.因子较多时如何避免过拟合等。幸运的是，机器学习技术的发展给我们提供了一条有效利用并筛选大量因子数据的途径。本报告中，我们将股市未来的涨和跌定义为一个分类问题，利用机器学习算法来对 Wind 全 A 指数的未来涨跌建模。
- 我们利用 51 种日频因子数据构建基于决策树的 AdaBoost 分类器，从而对下一交易日 Wind 全 A 指数的涨（1）跌（-1）做出预测。51 种因子中包含回购利率，信用利差、南华商品指数收益率、金油比、标普 500 指数等多种类型的市场信息。回测结果显示，若不考虑交易成本，滚动测算的 AdaBoost 多空择时策略在 2014 年 10 月 27 日至 2019 年 8 月 30 日获得了 41.31%的年化收益率和 1.41 的收益风险比，纯多头策略的年化收益率达到 24.67%，收益风险比达到 0.98，而同期简单持有策略的年化收益率和收益风险比仅有 7.66%和 0.26。
- 基于决策树的 AdaBoost 算法解决了构建择时模型面临的多个问题，它实现了：1.有效因子的自动筛选。2.非线性因子的建模。3.通过自适应调整样本权重解决因子相关性问题。4.不易过拟合，模型较稳健。
- 最后，我们通过叠加水晶球择时模型，充分利用期权市场信息，形成了表现更加出色的双塔奇兵择时模型。在单边万五的交易成本假设下，多空策略在 2015 年 6 月 1 日至 2019 年 8 月 30 日实现了 15.90%的年化收益率和 0.72 的收益风险比，纯多头策略年化收益率和收益风险比也分别达到 13.16%和 0.70，而同期 Wind 全 A 收益率仅有-11.18%。

风险提示：结论基于历史数据，在市场环境转变时模型存在失效的风险。

目 录

报告正文

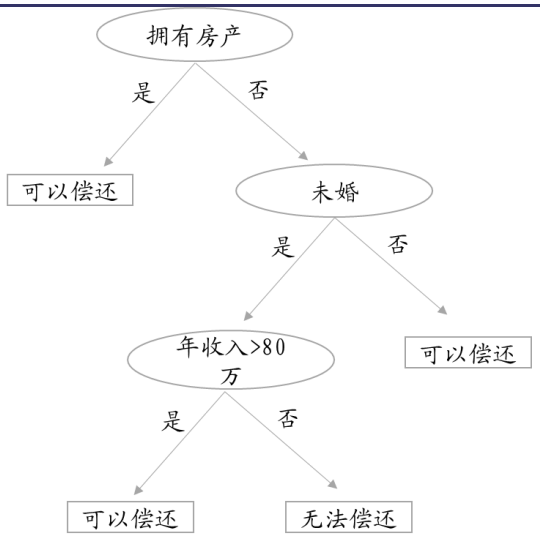
本篇是系统化资产配置系列报告的第三篇，对如何利用机器学习算法进行短期市场择时进行了系统介绍。全球金融市场每天产生海量的各类数据，如何筛选并有效利用这些数据来预测股票市场走势一直是一个重要但棘手的问题。短期择时面临的主要困难包括：1.短期市场走势受情绪等因素影响较大；2.如何筛选有效因子；3.非线性因子如何建模；4.因子相关性问题如何解决；5.因子较多时如何避免过拟合等。幸运的是，机器学习技术的发展给我们提供了一条有效利用并筛选大量因子数据的途径。本报告中，我们将股市未来的涨和跌定义为一个分类问题，利用机器学习算法来对 Wind 全 A 指数的未来涨跌建模。滚动回测结果显示，我们基于机器学习的短期择时策略能够获得显著超额收益，收益风险比、最大回撤等指标也远优于简单持有策略。

1、基于决策树的择时模型

1.1、决策树简介

决策树（Decision Trees）是一种常用的非参数监督学习模型（Nonparametric Supervised Learning）。它是一种树形结构，其中每个内部节点表示一个属性上的判断，每个分支代表一个判断结果的输出，最后每个叶节点代表一种分类或回归结果。一般来说，我们将目标变量为一系列离散值的决策树称为分类树（Classification Trees），将目标变量为连续值的决策树称为回归树(Regression Trees)。下图给出了一个银行根据客户信息判断是否批准贷款的决策树示例。

图表 1、贷款偿付能力决策树



资料来源：兴业证券经济与金融研究院整理

我们以著名的 CART（Classification and Regression Trees）算法为例来说明如何构建分类树（python 的 sklearn 包是使用优化后的 CART 算法来实现分类树和回归树模型）。

我们首先定义基尼系数，基尼系数代表了模型的不纯度，基尼系数越小，则不纯度越低，模型的分类能力越强。假设在分类问题中有 K 个类别，第 k 个类别的概率为p_k，则基尼系数表达式为：

$$Gini(\mathrm{p}) = \sum_{k=1}^{K} p_k(1-p_k) = 1 - \sum_{k=1}^{K} p_k^2$$

对于给定的样本 D，假设第 k 个类别的数量为C_k，则样本的基尼系数表达式为：

$$Gini(\mathrm{D}) = 1 - \sum_{k=1}^{K} (\frac{C_k}{D})^2$$

特别的，对于二分类问题的样本 D，如果根据特征 A 的某个值 a，把 D 分成D_1和D_2两部分，则在特征 A 的条件下，D 的基尼系数为：

$$Gini(\mathrm{D},\mathrm{A}) = \frac{D_1}{D} Gini(D_1) + \frac{D_2}{D} Gini(D_2)$$

其中 D、D_1和D_2表示样本的数量。

CART 分类树构建流程：

用 CART 算法构建分类数的流程如下：

输入：训练集，基尼系数阈值，样本个数阈值。

输出：分类树 T

从根节点开始，我们用训练集递归的建立分类树。

1. 对于当前节点的数据集 D，如果样本个数小于阈值或没有特征，则返回决策子树，当前节点停止递归。
2. 计算样本集 D 的基尼系数，如果基尼系数小于阈值，则返回决策树子树，当前节点停止递归。
3. 计算当前节点现有各个特征 A 的每个特征值 a 对数据集 D 的条件基尼系数。
4. 在计算得到的所有特征的各个特征值对应的基尼系数中，选择基尼系数最小的特征 A 和对应的特征值 a。根据这个最优特征和最优特征值，把数据集划分为D_1和D_2。
5. 对左右子节点递归的调用 1-4 步，生成决策树。

使用生成的决策树做预测时，假如测试集里的样本 X 落到了某个叶节点，而节点里有多个训练样本，则对于 X 的类别预测采用的是这个叶节点里概率最大的类别。

在 sklearn 的决策树实现中，除了基尼系数外，也提供了使用信息熵（Entropy）作为不纯度度量来生成决策树的选项，即在每个节点选择样本分裂后信息增益（Information Gain）最大也就是熵减最多的属性。假设在分类问题中有 K 个类别，样本 D 中第 k 个类别的概率为p_k，则信息熵定义为：

$$Entropy(D) = -\sum_{k=1}^{K} p_k log_2(p_k)$$

特别的，对于二分类问题，我们有：

$$Entropy(D) = -p_k log_2(p_k) - (1 - p_k) log_2(1 - p_k)$$

样本 D 关于属性 A 的条件信息熵为：

$$Entropy(D, A) = \frac{D_1}{D} Entropy(D_1) + \frac{D_2}{D} Entropy(D_2)$$

属性 A 带来的信息增益为：

$$Gain(A) = Entropy(D) - Entropy(D, A)$$

其中 D、D_1和D_2表示样本的数量。

1.2、数据说明

本报告择时模型的预测标的为 Wind 全 A 指数，使用了包含资金流动性、风险偏好、技术指标等 51 种因子数据。数据的时间范围是 2007 年 6 月 1 日至 2019 年 8 月 30 日。因子具体定义见下表：

图表 2、51 种短期因子定义

1	银行间同业拆借加权利率:1 天	27	南华商品指数
2	银行间同业拆借加权利率:1 天：过去五天的变化率	28	南华商品指数：过去五天的变化率
3	7 天期回购利率	29	CRB 现货指数:综合
4	7 天期回购利率：过去五天的变化率	30	CRB 现货指数:综合：过去五天的变化率
5	银行间质押式回购加权利率:7 天	31	期货收盘价(连续):COMEX 黄金
6	银行间质押式回购加权利率:7 天：过去五天的变化率	32	期货收盘价(连续):COMEX 黄金：过去五天的变化率
7	shibor 利率（ON）	33	期货结算价(连续):WTI 原油
8	shibor 利率（1W）	34	期货结算价(连续):WTI 原油：过去五天的变化率
9	shibor 利率（2W）	35	COMEX 黄金/WTI 原油：过去五天的变化率
10	shibor 利率（1M）	36	COMEX 黄金/WTI 原油：过去五天的变化率
11	shibor 利率（3M）	37	标普 500
12	shibor 利率（6M）	38	标普 500：过去五天的变化率
13	shibor 利率（ON）：过去五天的变化率	39	市场动量指标
14	shibor 利率（1W）：过去五天的变化率	40	市场交易活跃指标
15	shibor 利率（2W）：过去五天的变化率	41	市场动量指标：过去五天的收益率
16	shibor 利率（1M）：过去五天的变化率	42	市场交易活跃指标：过去五天成交量的变化率
17	shibor 利率（3M）：过去五天的变化率	43	医药行业超额：过去五天的变化率
18	shibor 利率（6M）：过去五天的变化率	44	食品饮料行业超额：过去五天的变化率
19	中债国债到期收益率（0 年）	45	防御性行业超额:过去五天的变化率（医药和食品饮料）
20	中债国债到期收益率（1 月）	46	Beta 分离度指标

21	中债国债到期收益率（2月）	47	Beta分离度指标：过去五天的变化率
22	中债国债到期收益率（0年）：过去五天的变化率	48	Wind全A过去60日的波动率
23	中债国债到期收益率（1月）：过去五天的变化率	49	Wind全A过去60日的波动率:过去五天的变化率
24	中债国债到期收益率（2月）：过去五天的变化率	50	Wind全A过去120日的波动率
25	中债国开债到期收益率:3年	51	Wind全A过去120日的波动率:过去五天的变化率
26	中债国开债到期收益率:3年：过去五天的变化率		

资料来源：Wind，兴业证券经济与金融研究院整理

1.3、决策树择时模型

使用决策树模型来构建择时策略有很多优势。首先，决策树模型可以选择多个因子作为预测节点，具有结合多个因子信息的能力。其次，决策树模型具有拟合非线性因子的能力。最后，决策树的算法保证了它会优先选取预测能力强的因子作为节点，实现因子的自动筛选。

1.3.1、决策树择时模型回测流程

下面我们给出本文的第一个择时模型：**决策树择时模型**。本模型基于51种短期因子，通过多层决策树来预测标的下一交易日的涨跌并构建交易策略。多层决策树模型的构建流程见下图。

图表3、多层决策树模型构建流程图

资料来源：兴业证券经济与金融研究院整理

基于多层决策树的决策树择时模型回测流程如下：

回测数据的时间范围是2007年6月1日至2019年8月30日，使用扩展窗口法（Expanding）：

1. 取得至少1800天的样本数据作为训练集。利用训练集进行分层KFold交叉验证(k=5)以选取预测准确率最高的决策树模型。超参数的搜索范围为：

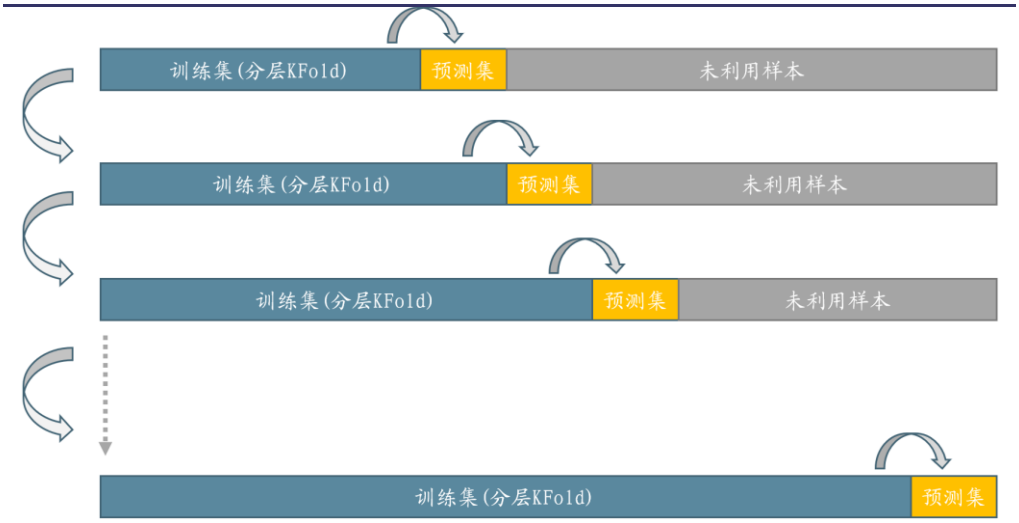
   a. 决策树不纯度度量：{信息熵，基尼系数}
   b. 决策树深度：{5,10，…，25，30}

2. 利用第一步得到的最优模型预测Wind全A指数下一交易日的涨跌。若预测信号为涨，则假设在当日收盘买入Wind全A；若预测信号为跌，则在当日收盘做空Wind全A。

3. 连续使用 2 中的最优模型 20 个交易日后再次进入步骤 1，并将此 20 个交易日数据加入训练集。

注意到由于首次训练模型时使用了 1800 个样本数据，因此实际的交易信号是在 2014 年 10 月 27 日首次给出。

图表 4、决策树择时模型回测流程图



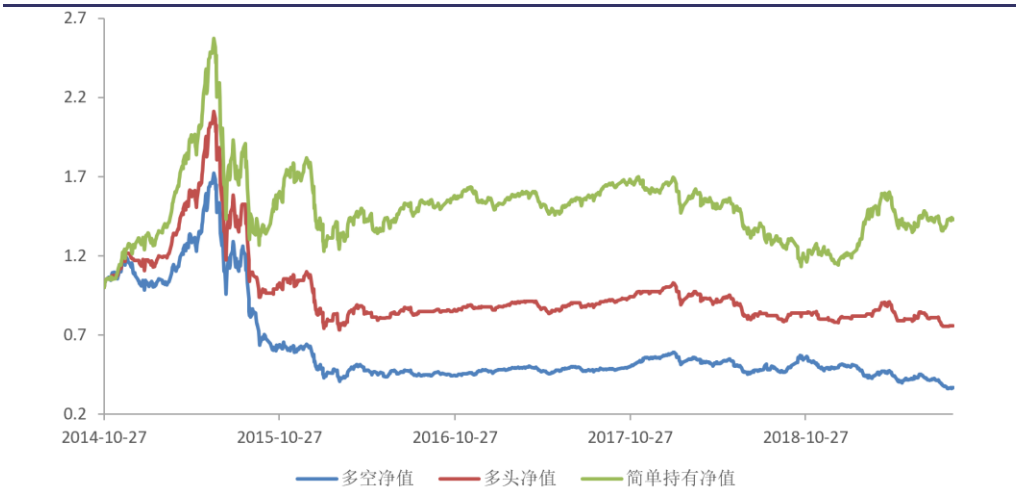
资料来源：兴业证券经济与金融研究院整理

1.3.2、决策树择时模型回测结果

假设在换仓信号发出当日收盘时进行交易且无交易成本，我们给出决策树择时模型多空策略和纯多头策略的表现。

从下图看，单纯使用决策树作为预测模型效果较差，策略收益甚至显著低于简单持有策略的表现。多空策略年化收益率仅有-18.72%，多头策略为-5.56%，远小于简单持有策略 7.66%的年化收益率。

图表 5、决策树择时模型策略净值（无交易成本，当日收盘时交易）



资料来源：Wind，兴业证券经济与金融研究院整理

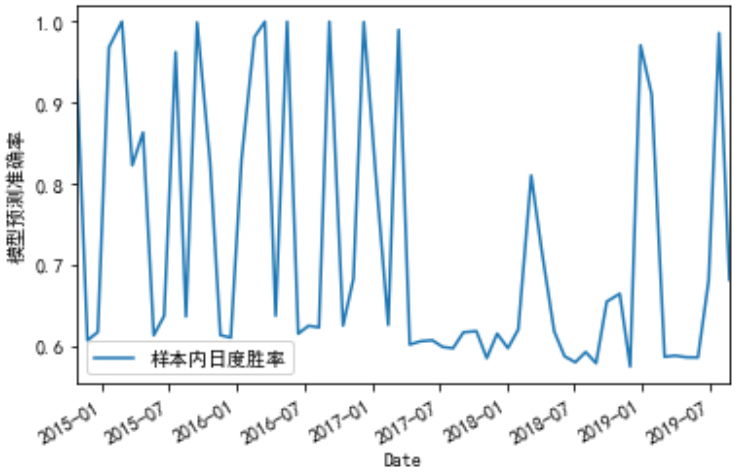
图表 6、决策树择时模型策略表现（无交易成本，当日收盘时交易）

	多空策略	多头策略	简单持有
年化收益率	-18.72%	-5.56%	7.66%
年化波动率	29.37%	25.81%	29.38%
收益风险比	-0.64	-0.22	0.26
最大回撤	-78.94%	-65.33%	-55.99%
每笔胜率	47.15%	47.56%	-
年均交易次数	101.37	50.79	-

资料来源：Wind，兴业证券经济与金融研究院整理

模型样本外表现较差可能是多层决策树极易出现过拟合的特性导致的。下图给出了每次训练得到的决策树样本内预测准确率。可以看到，多层决策树模型的样本内预测准确率相当高，平均达到 72.35%，远远高于其样本外 50%左右的胜率。这意味着模型出现了一定程度的过拟合。

图表 7、决策树择时模型样本内预测准确率



资料来源：Wind，兴业证券经济与金融研究院整理

1.4、优选决策树择时模型

1.4.1、优选决策树择时模型回测流程

前文介绍的基于多层决策树的择时模型表现非常不理想，这可能是由于多层决策树极易出现过拟合现象，从多层决策树的算法来看，该算法样本内输入的因子较多，缺乏对因子进行优选的过程，这也许是过拟合存在的原因。为了解决这个问题，我们尝试基于多个单层决策树构建择时模型，即**优选决策树择时模型**。

简单来说，优选决策树择时模型通过等权叠加多个单层决策树的预测值来获得最终交易信号。具体来说，我们首先分别对 51 种因子构建单层决策树，筛选出在样本内能够带来显著超额收益的因子及其决策树。然后使用筛选后的决策树预测未来市场涨跌，最后将多个决策树预测结果叠加，构建交易策略。

图表 8、优选决策树择时模型构建流程图

资料来源：兴业证券经济与金融研究院整理

基于多个单层决策树的优选决策树择时模型回测流程如下：

回测数据的时间范围是 2007 年 6 月 1 日至 2019 年 8 月 30 日，同样使用扩展窗口法（Expanding）。

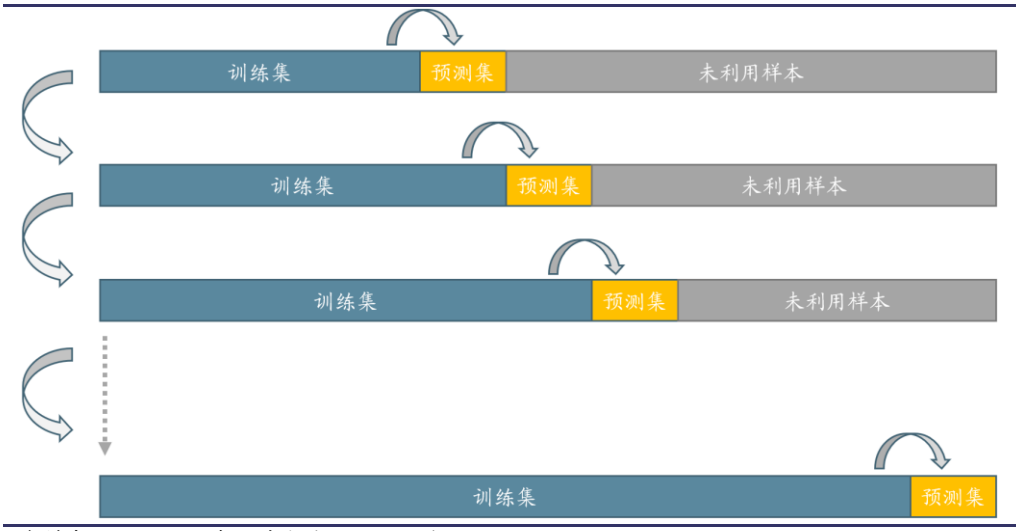
1. 取得至少 1800 天的样本数据作为训练集。对训练集中的 51 个因子分别构建单层决策树。

2. 计算 51 个单层决策树对应的样本内择时策略收益率序列。对每一组策略收益率序列和简单持有策略的超额收益序列做 t 检验，筛选其中在 5%显著性水平上显著的决策树。

3. 利用筛选出的若干决策树预测 Wind 全 A 指数下一交易日的涨跌，对所有决策树预测结果取均值，均值大于 0 则做多，小于 0 则多空，等于 0 则平仓。若未筛选出任何一个决策树，则直接给出做多信号。

4. 连续使用 2 中的模型 20 个交易日后再次进入步骤 1，并将此 20 个交易日数据加入训练集。

图表 9、优选决策树择时模型回测流程图



资料来源：兴业证券经济与金融研究院整理

1.4.2、优选决策树择时模型回测结果

假设在换仓信号发出当日收盘时进行交易且无交易成本，我们给出优选决策树择时模型多空策略和纯多头策略表现。

从下图看，结合多个单层决策树的择时模型效果比单个多层决策树择时模型要好。多空策略年化收益率达到 12.47%，多头策略达到 12.31%，显著高于简单持有 7.66%的年化收益率，收益风险比，最大回撤等指标也优于简单持有策略。

图表 10、优选决策树择时模型策略净值（无交易成本，当日收盘时交易）

资料来源：Wind，兴业证券经济与金融研究院整理

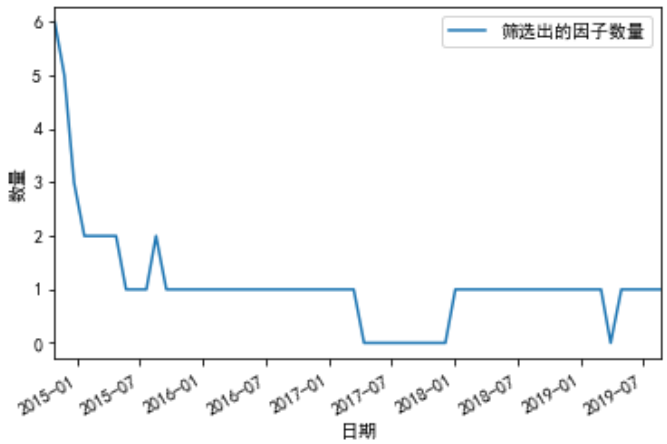
图表 11、优选决策树择时模型策略表现（无交易成本，当日收盘时交易）

	多空策略	多头策略	简单持有
年化收益率	12.47%	12.31%	7.66%
年化波动率	28.22%	25.03%	29.38%
收益风险比	0.44	0.49	0.26
最大回撤	-44.41%	-44.41%	-55.99%
每笔胜率	57.43%	57.43%	-
年均交易次数	39.02	20.64	-

资料来源：Wind，兴业证券经济与金融研究院整理

下图给出了本模型每次训练时筛选出的有效因子数量。可以看出，除了 2015 年初，大部分时间只有 1 个因子能带来显著样本内超额收益，这说明单个因子的预测能力并不理想。另外，单层决策树也难以解决因子与收益的非线性问题，所以我们需要进一步改进模型，考虑将单因子按照某种结合方式形成多因子模型。

图表 12、优选决策树择时模型有效因子数量



资料来源：Wind，兴业证券经济与金融研究院整理

2、AdaBoost 择时模型

2.1、AdaBoost 算法简介

决策树择时模型基于多层决策树，容易出现过拟合问题。优选决策树择时模型基于多个单层决策树虽不易过拟合，但它过于简单，无法充分利用因子信息以及捕捉因子的非线性性质。幸运的是，以上这些问题可以被 AdaBoost 算法完美解决。简要来说，基于 AdaBoost 算法的择时模型具有以下特点：

1. 结合多个弱分类器，具有较高精度。
2. 不易出现过拟合现象，模型较稳健。
3. 具有处理非线性因子能力。
4. 通过自适应改变样本权重实现高相关性因子的自动剔除。
5. 不需要对特征进行人工筛选，可以在模型中纳入大量因子。

具体而言，Boosting 是集成学习方法（Ensemble Method）的一种，在实践中有着广泛的应用，而 AdaBoost 模型则是其最为流行的一种实现，最早由 Yoav Freund 和 Robert Schapire 在 1995 年提出。AdaBoost 提供了一种框架，在框架内可以使用多种弱分类器，理论上不需要对特征进行人工筛选，训练得到的模型也不易出现过拟合现象。AdaBoost 是“Adaptive Boosting”（自适应增强）的缩写，它的自适应性体现在当前基分类器（通常是弱分类器）分类错误的样本权重会增大，而正确分类的样本权重会减小，从而在训练下一个基分类器时会着重拟合之前分类错误的样本。

AdaBoost 的具体算法如下：

假定我们有 N 个样本$(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$，其中 x 表示属性，y 表示类别。则基分类器数量设定为 M 个的 AdaBoost 算法可以归纳为以下几个步骤：

1. 初始化样本权重，令$w_i = \frac{1}{N}, \mathrm{i} = 1, 2, 3, \dots, \mathrm{N}$。

2. 对$m=1$到M:

a. 利用样本权重w_i训练一个基分类器$G_m(x)$。

b. 计算此基分类器的样本加权错误率:

$$err_m = \frac{\sum_{i=1}^{N} w_i I(y_i \neq G_m(x_i))}{\sum_{i=1}^{N} w_i}$$

c. 计算基分类器信心度:

$$\alpha_m = \ln\left(\frac{1-err_m}{err_m}\right)$$

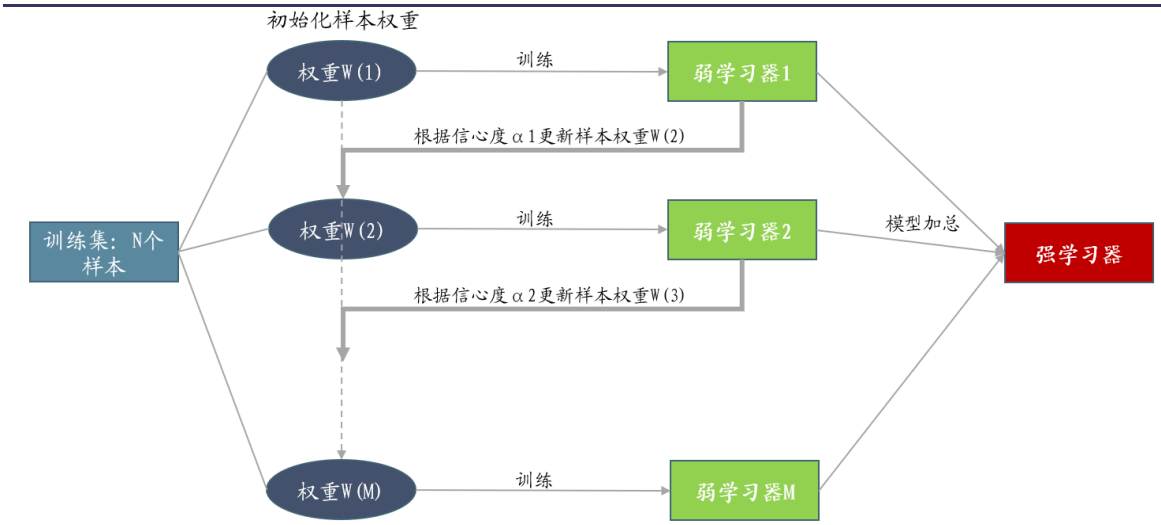
d. 根据信心度更新样本权重:

$$w_i \leftarrow w_i \cdot \exp\left[\alpha_m \cdot I\left(y_i \neq G_m(x_i)\right)\right], i = 1, 2, 3, \ldots, N$$

3. 结合 M 个基分类器得到最终分类器:

$$G(x) = sign[\sum_{m=1}^{M} \alpha_m G_m(x)]$$

图表 13、AdaBoost 算法流程图



资料来源：兴业证券经济与金融研究院整理

2.2、AdaBoost 择时模型回测流程

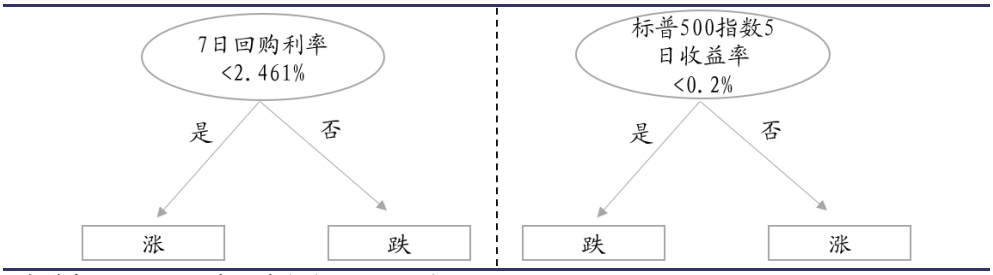
我们利用 51 种日频因子数据构建基于决策树的 AdaBoost 分类器，对下一交易日 Wind 全 A 指数的涨（1）跌（-1）做出预测。为了模型的泛化性能，我们将决策树的深度设置为 1，也就是仅有一个根节点和两个叶节点。

在这种模型设定下，AdaBoost 的每一个基分类器都只会选择 51 种因子中的某一个作为决策树节点来对 Wind 全 A 下一交易日涨跌做出预测。图表 15 给出了 AdaBoost 算法可能选取的两个基分类器（基于 2007 年 6 月 1 日-2019 年 8 月 22 日的数据）。举例来说，AdaBoost 的某个基分类器可能选择 7 日回购利率作为节点，当 7 日回购利率低于某一阈值给出涨的预测，7 日回购利率高于某一阈值则给出跌的预测。另一个基分类器可能选取标普 500 指数过去 5 日收益率作为节点，当 5 日收益率高于某一阈值给出涨的预测，否则给出跌的预测。而下个交易日的最终预测结果由所有基分类器共同决定。

图表 14、AdaBoost 择时模型构建流程图

资料来源：兴业证券经济与金融研究院整理

图表 15、AdaBoost 择时模型基分类器示例



资料来源：兴业证券经济与金融研究院整理

AdaBoost 择时模型回测流程如下：

回测数据的时间范围是 2007 年 6 月 1 日至 2019 年 8 月 30 日，我们同样使用扩展窗口法（Expanding）来训练和交叉验证模型并发出交易信号。具体来说，步骤如下：

1. 取得至少 1800 天的样本数据作为训练集。利用训练集进行分层 KFold 交叉验证(k=5)以选取预测正确率最高的模型超参数。超参数的搜索范围为：

   a. 决策树不纯度度量：{信息熵，基尼系数}
   b. AdaBoost 基分类器数量：{20, 25, 30}

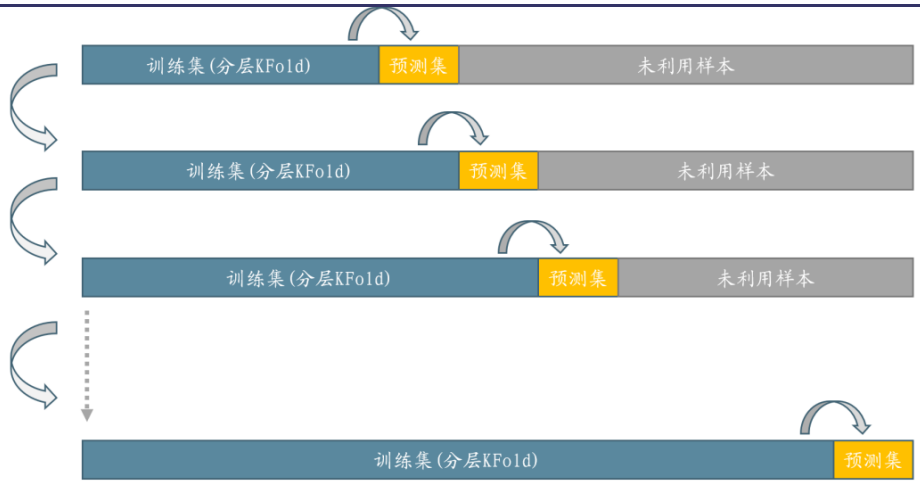
2. 利用第一步得到的最优模型预测 Wind 全 A 指数下一交易日的涨跌。若下一交易日预测信号为涨，则假设在当日收盘买入 Wind 全 A，若下一交易日预测信号为跌，则在当日收盘做空 Wind 全 A。

3. 连续使用 2 中的模型 20 个交易日后再次进入步骤 1，并将此 20 个交易日

数据加入训练集。

注意到由于首次训练模型时使用了 1800 个样本数据，因此实际的交易信号是在 2014 年 10 月 27 日首次给出。

图表 16、AdaBoost 择时模型回测流程图



资料来源：兴业证券经济与金融研究院整理

2.3、AdaBoost 择时模型回测结果

同样假设在换仓信号发出当日收盘时进行交易且无交易成本，我们给出 AdaBoost 择时模型多空策略和纯多头策略的表现。

从下图看，多空策略表现>纯多头策略表现>简单持有策略表现。纯多头的年化收益率接近 25%，收益风险比达到 0.97，远远高于简单持有 Wind 全 A 指数获得的 7.6%的年化收益率和 0.26 的收益风险比。AdaBoost 模型的表现也远远优于决策树和优选决策树模型。从换手率看，纯多头策略年均交易 43 次，平均 1 周多 1 次，交易频率比较适中。

图表 17、AdaBoost 择时模型策略净值（无交易成本，当日收盘时交易）

资料来源：Wind，兴业证券经济与金融研究院整理

图表 18、AdaBoost 择时模型策略表现（无交易成本，当日收盘时交易）

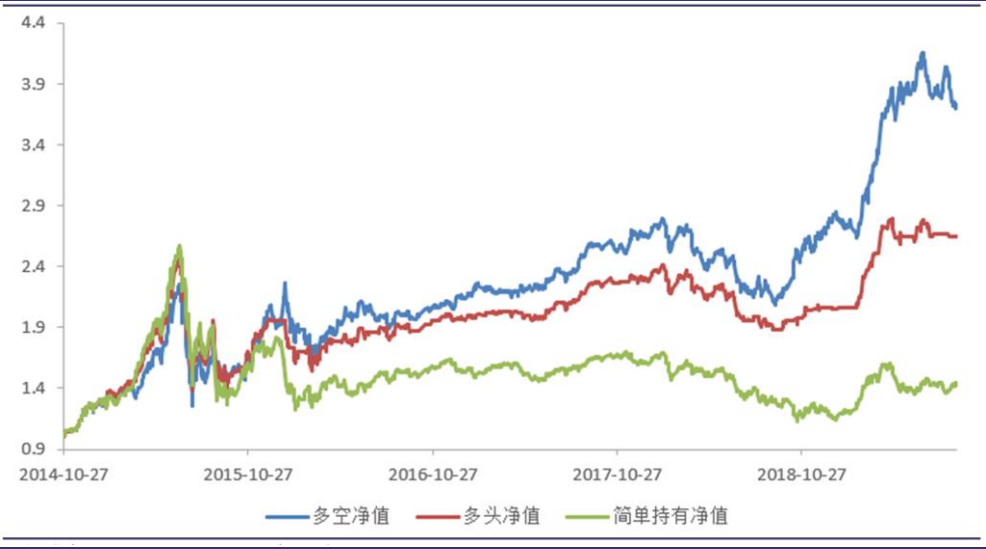
	多空策略	多头策略	简单持有
年化收益率	41.31%	24.67%	7.66%
年化波动率	29.28%	25.30%	29.38%
收益风险比	1.41	0.98	0.26
最大回撤	-44.41%	-44.89%	-55.99%
每笔胜率	59.13%	59.62%	-
年均交易次数	85.68	42.94	-

资料来源：Wind，兴业证券经济与金融研究院整理

为了让回测结果更贴近现实，我们进一步假设在换仓信号发出次日开盘时进入新的仓位以及单边万分之五的交易成本。

在新的假设下，我们的择时策略表现略有下降，但依然远远优于简单持有策略和优选决策树择时策略的表现。AdaBoost 择时模型纯多头年化收益率依然在 22%以上，收益风险比接近 0.9。这进一步验证了我们 AdaBoost 模型择时策略的有效性和稳健性。

图表 19、AdaBoost 择时模型策略净值（万五交易成本，次日开盘时交易）



资料来源：Wind，兴业证券经济与金融研究院整理

图表 20、AdaBoost 择时模型策略表现（万五交易成本，次日开盘时交易）

	多空策略	多头策略	简单持有
年化收益率	31.24%	22.25%	7.66%
年化波动率	28.86%	24.98%	29.38%
收益风险比	1.08	0.89	0.26
最大回撤	-44.41%	-44.41%	-55.99%
每笔胜率	57.21%	60.58%	-
年均交易次数	85.68	42.94	-

资料来源：Wind，兴业证券经济与金融研究院整理

3、双塔奇兵择时模型

3.1、双塔奇兵择时模型介绍

由于我国上证 50ETF 期权数据历史较短，因此前文介绍的机器学习择时模型并没有使用期权相关的择时因子。而兴业期权水晶球择时模型通过挖掘期权市场隐含的市场预期，充分利用期权相关信息，取得了极佳的历史择时表现（关于水晶球模型的详细介绍请参考《抽丝剥茧，去芜存菁：水晶球择时模型之 3.0》）。另外，水晶球择时策略收益率与 AdaBoost 择时策略收益率相关性较低，仅有 0.02。因此我们尝试将 AdaBoost 择时模型与水晶球择时模型结合，形成新的双塔奇兵择时模型。

双塔奇兵择时模型的信号生成方式和回测流程如下：

水晶球择时信号在 2015 年 6 月 1 日首次发出，因此我们回测的时间范围为 2015 年 6 月 1 日至 2019 年 8 月 30 日。

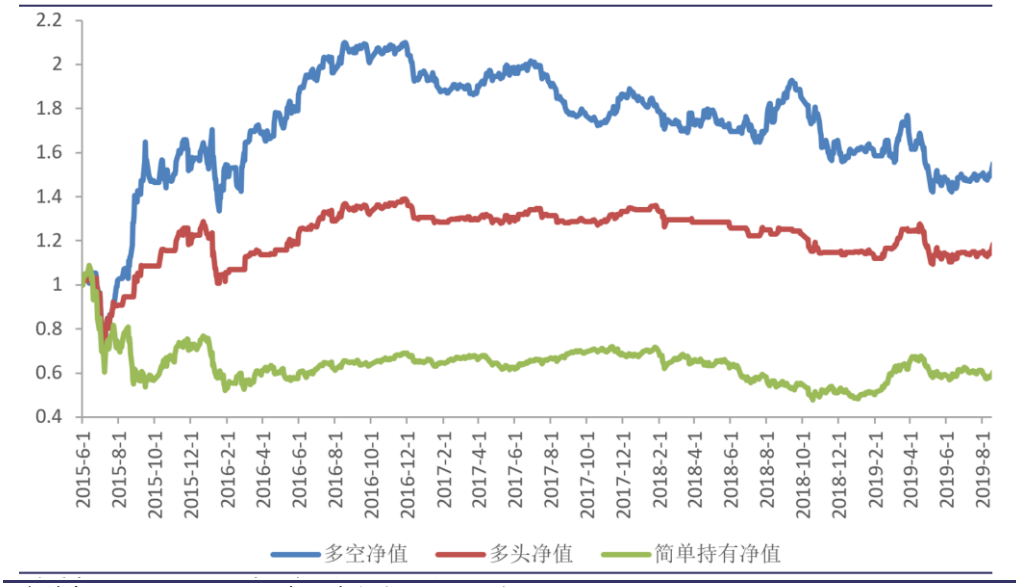
1. 获取水晶球择时模型多空策略信号 A。
2. 获取 AdaBoost 择时模型多空策略信号 B。
3. 令 C=A+B。若 C>0，则发出做多 Wind 全 A 的信号；若 C<0，则发出做空信号；若 C=0，则发出平仓信号。

3.2、水晶球择时模型用于 Wind 全 A 指数

原水晶球择时模型的标的为上证 50ETF，我们给出将其发出的交易信号直接用于 Wind 全 A 指数的策略表现。假设在换仓信号发出次日开盘时进入新仓位以及单边万分之五的交易成本，下文给出了 2015 年 6 月 1 日至 2019 年 8 月 30 日的水晶球择时策略表现。

从下图看，水晶球择时模型表现远远优于简单持有策略。纯多头策略年化收益率达到 4.58%，收益风险比达到 0.24，而同期简单持有策略收益率仅为-11.18%，收益风险比为-0.38。

图表 21、水晶球择时模型策略净值（万五交易成本，次日开盘时交易）



资料来源：Wind，兴业证券经济与金融研究院整理

图表 22、水晶球择时模型策略表现（万五交易成本，次日开盘时交易）

	多空策略	多头策略	简单持有
年化收益率	11.61%	4.58%	-11.18%
年化波动率	25.19%	18.89%	29.46%
收益风险比	0.46	0.24	-0.38
最大回撤	-34.69%	-33.65%	-55.99%
每笔胜率	49.30%	52.65%	-
年均交易次数	162.14	84.48	-

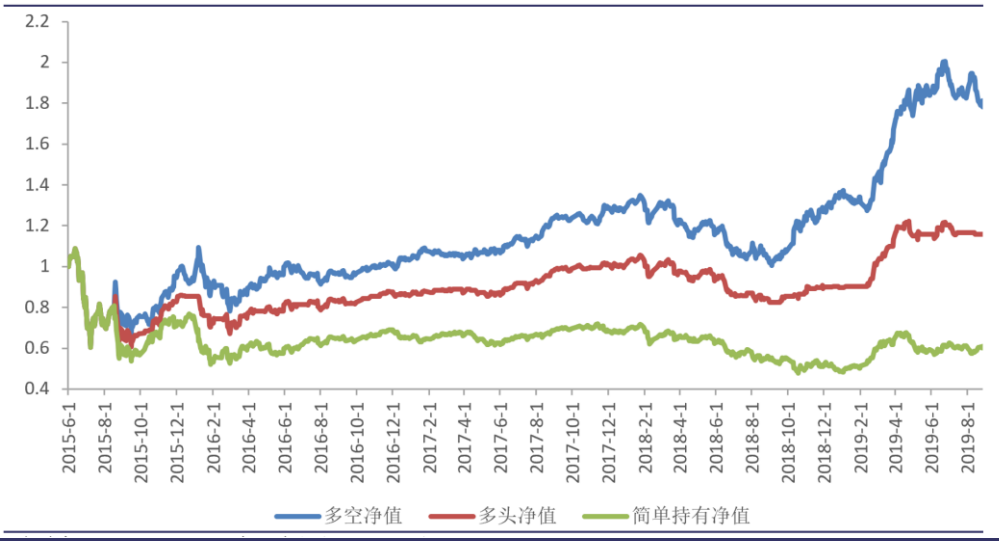
资料来源：Wind，兴业证券经济与金融研究院整理

3.3、AdaBoost 择时模型同期表现

为了便于后续比较，我们这里也给出 AdaBoost 择时模型在 2015 年 6 月 1 日至 2019 年 8 月 30 日的表现。同样假设在换仓信号发出次日开盘时进入新仓位以及万分之五的交易成本。

从下表看，AdaBoost 模型纯多头策略在 2015 年 6 月 1 日以来实现 3.49%的年化收益率和 0.14 的收益风险比。其表现优于简单持有策略，但比水晶球择时模型纯多头策略 4.58%的年化收益率和 0.24 的收益风险比略差。

图表 23、AdaBoost 择时模型同期策略净值（万五交易成本，次日开盘时交易）



资料来源：Wind，兴业证券经济与金融研究院整理

图表 24、AdaBoost 择时模型同期策略表现（万五交易成本，次日开盘时交易）

	多空策略	多头策略	简单持有
年化收益率	14.82%	3.49%	-11.18%
年化波动率	29.09%	24.71%	29.46%
收益风险比	0.51	0.14	-0.38
最大回撤	-44.41%	-44.41%	-55.99%
每笔胜率	58.16%	61.22%	-
年均交易次数	92.01	46.13	-

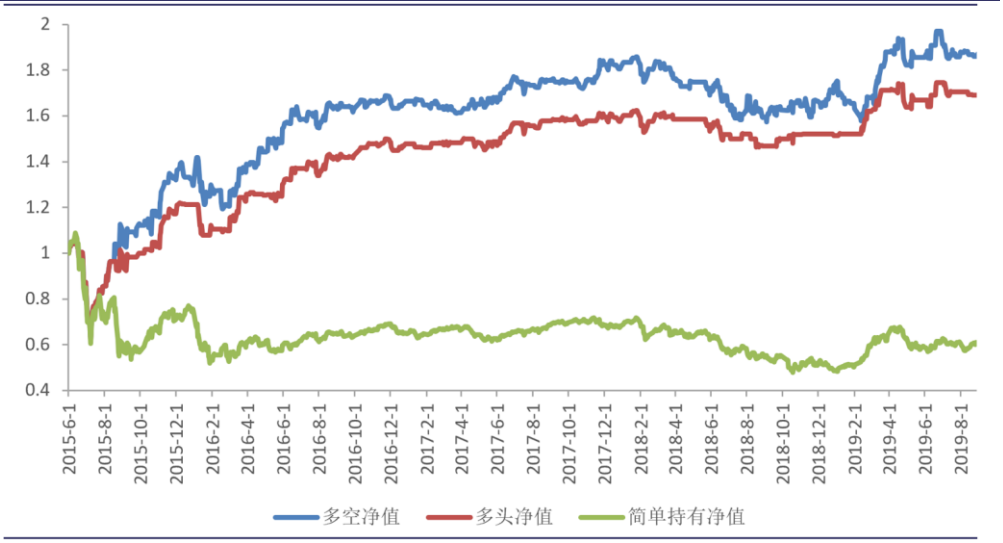
资料来源：Wind，兴业证券经济与金融研究院整理

3.4、双塔奇兵择时模型回测结果

最终我们给出结合 AdaBoost 与水晶球择时模型的双塔奇兵择时模型 2015 年 6 月 1 日以来的策略表现。同样假设在换仓信号发出次日开盘时进入新仓位以及万分之五的交易成本。

从图表 27 看，双塔奇兵择时模型的纯多头策略净值曲线持续高于另外两种模型。本策略在 2015 年 6 月 1 日以来实现了 13.16%的年化收益率，收益风险比达到 0.7，不仅远优于简单持有策略，也优于水晶球和 AdaBoost 择时模型的表现。

图表 25、双塔奇兵择时模型策略净值（万五交易成本，次日开盘时交易）



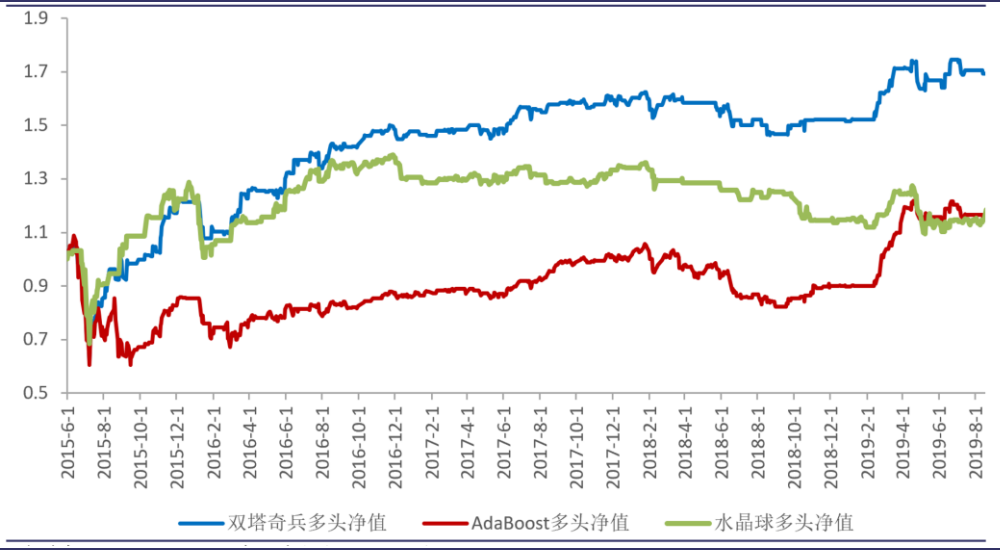
资料来源：Wind，兴业证券经济与金融研究院整理

图表 26、双塔奇兵择时模型策略表现（万五交易成本，次日开盘时交易）

	多空策略	多头策略	简单持有
年化收益率	15.90%	13.16%	-11.18%
年化波动率	21.99%	18.83%	29.46%
收益风险比	0.72	0.70	-0.38
最大回撤	-40.94%	-40.94%	-55.99%
每笔胜率	52.62%	58.45%	-
年均交易次数	122.37	69.66	-

资料来源：Wind，兴业证券经济与金融研究院整理

图表 27、双塔奇兵、AdaBoost、水晶球择时模型多头策略净值比较（万五交易成本，次日开盘时交易）



资料来源：Wind，兴业证券经济与金融研究院整理

4、结论

本报告试图将机器学习用于市场短期择时，先后尝试了多层决策树、优选决策树以及 AdaBoost 择时模型。其中，基于决策树的 AdaBoost 算法从理论和实证上解决了构建择时模型面临的多个问题，降低了过拟合的可能性，样本外择时表现亮眼。更值得一提的是 AdaBoost 择时模型的表现与基于期权市场择时的水晶球模型相关性很低，于是我们通过叠加水晶球择时模型，充分利用期权市场信息，形成了表现更加出色的双塔奇兵择时模型，大幅提高了短期择时的胜率和稳健性。

风险提示：结论基于历史数据，在市场环境转变时模型存在失效的风险。

分析师声明

本人具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，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。本人不曾因，不因，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收到任何形式的补偿。

投资评级说明

投资建议的评级标准	类别	评级	说明
报告中投资建议所涉及的评级分为股票评级和行业评级(另有说明的除外)。评级标准为报告发布日后的12个月内公司股价(或行业指数)相对同期相关证券市场代表性指数的涨跌幅。其中：A股市场以上证综指或深圳成指为基准，香港市场以恒生指数为基准；美国市场以标普500或纳斯达克综合指数为基准。	股票评级	买入	相对同期相关证券市场代表性指数涨幅大于15%
		审慎增持	相对同期相关证券市场代表性指数涨幅在5%～15%之间
		中性	相对同期相关证券市场代表性指数涨幅在-5%～5%之间
		减持	相对同期相关证券市场代表性指数涨幅小于-5%
		无评级	由于我们无法获取必要的资料，或者公司面临无法预见结果的重大不确定性事件，或者其他原因，致使我们无法给出明确的投资评级
	行业评级	推荐	相对表现优于同期相关证券市场代表性指数
		中性	相对表现与同期相关证券市场代表性指数持平
		回避	相对表现弱于同期相关证券市场代表性指数

信息披露

本公司在知晓的范围内履行信息披露义务。客户可登录 www.xyzq.com.cn 内幕交易防控栏内查询静默期安排和关联公司持股情况。

使用本研究报告的风险提示及法律声明

兴业证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

“本公司”）的客户使用，本公司不会因接收人收到本报告而视其为客户。本报告中的信息、意见等均仅供客户参考，不构成所述证券买卖的出价或征价邀请或要约。该等信息、意见并未考虑到获取本报告人员的具体投资目的、财务状况以及特定需求，在任何时候均不构成对任何人的个人推荐。客户应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求，必要时就法律、商业、财务、税收等方面咨询专家的意见。对依据或者使用本报告所造成的一切后果，本公司及/或其关联人员均不承担任何法律责任。

本报告所载资料的来源被认为是可靠的，但本公司不保证其准确性或完整性，也不保证所包含的信息和建议不会发生任何变更。本公司并不对使用本报告所包含的材料产生的任何直接或间接损失或与此相关的其他任何损失承担任何责任。

本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可升可跌，过往表现不应作为日后的表现依据；在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告；本公司不保证本报告所含信息保持在最新状态。同时，本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

除非另行说明，本报告中所引用的关于业绩的数据代表过往表现。过往的业绩表现亦不应作为日后回报的预示。我们不承诺也不保证，任何所预示的回报会得以实现。分析中所做的回报预测可能是基于相应的假设。任何假设的变化可能会显著地影响所预测的回报。

本公司的销售人员、交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。本公司没有将此意见及建议向报告所有接收者进行更新的义务。本公司的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。

本报告并非针对或意图发送予或为任何就发送、发布、可得到或使用此报告而使兴业证券股份有限公司及其关联子公司等违反当地的法律或法规或可致使兴业证券股份有限公司受制于相关法律或法规的任何地区、国家或其他管辖区域的公民或居民，包括但不限于美国及美国公民（1934 年美国《证券交易所》第 15a-6 条例定义为本「主要美国机构投资者」除外）。

特别声明

在法律许可的情况下，兴业证券股份有限公司可能会持有本报告中提及公司所发行的证券头寸并进行交易，也可能为这些公司提供或争取提供投资银行业务服务。因此，投资者应当考虑到兴业证券股份有限公司及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突。投资者请勿将本报告视为投资或其他决定的唯一信赖依据。

兴业证券研究

上 海	北 京	深 圳
地址：上海浦东新区长柳路36号兴业证券大厦15层	地址：北京西城区锦什坊街35号北楼601-605	地址：深圳市福田区皇岗路5001号深业上城T2座52楼
邮编：200135	邮编：100033	邮编：518035
邮箱：research@xyzq.com.cn	邮箱：research@xyzq.com.cn	邮箱：research@xyzq.com.cn