



系统工程  
Systems Engineering  
ISSN 1001-4098, CN 43-1115/N

## 《系统工程》网络首发论文

题目: 基于决策树的量化交易择时策略研究  
作者: 张茂军, 饶华城, 南江霞, 王国栋  
收稿日期: 2021-03-08  
网络首发日期: 2021-12-27  
引用格式: 张茂军, 饶华城, 南江霞, 王国栋. 基于决策树的量化交易择时策略研究 [J/OL]. 系统工程.  
<https://kns.cnki.net/kcms/detail/43.1115.N.20211224.1403.002.html>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

# 基于决策树的量化交易择时策略研究

张茂军<sup>1</sup>, 饶华城<sup>2</sup>, 南江霞<sup>1</sup>, 王国栋<sup>3</sup>

(1.苏州科技大学商学院, 苏州 215009; 2. 桂林电子科技大学数学与计算科学学院, 桂林 541004; 3. 苏高新集团纪委办公室, 江苏 苏州 215011)

**摘要：**量化择时策略是量化投资的核心问题。考虑特征因子的具体含义和资产价格涨跌幅，用决策树方法提取区间突破分类标签和值特征选择指标，构建了 CLBIB-VSD-CART 算法，形成了量化择时策略系统，并应用于螺纹钢期货交易的择时策略分析。研究表明：（1）提出的 CLBIB-VSD-CART 算法不仅呈现了预测金融资产价格特征的经济含义，而且可以有效控制价格涨跌幅限制；（2）该算法在累积收益率、夏普比率和比率三个绩效指标都显著高于两个基准模型和其他择时系统，投资绩效比较稳健；（3）提出的研究方法为投资者选择最佳交易时间提供了有效的决策依据，具有重要的应用价值。

**关键词：**决策树；量化交易；择时信号；商品期货；算法

**中图分类号：**G210, F832.5 **文献标识码：**A

## Timing Selection of Quantitative Trading Based on Decision Tree

ZHANG Maojun<sup>1</sup>, YAO Huacheng<sup>2</sup>, NAN Jiangxia<sup>1</sup>, WANG Guodong<sup>3</sup>

(1.School of Business, Suzhou University of Science and Technology, Suzhou, 215009, China; 2.School of Mathematics and Computer Science, Guilin University of Electronic technology, Guilin, 541004, China; 3.Discipline inspection and supervision room, Suzhou HighTech Group, Suzhou, 215011)

**Abstract:** Quantitative timing strategy is the core issue of quantitative investment. Considering the specific meaning of characteristic factors and the rise and fall of asset prices, the interval breakthrough classification label and value feature selection index are extracted by decision tree method, CLBIB-VSD-CART algorithm is constructed to setup a quantitative timing strategy system with an application to the timing strategy analysis of rebar futures trading. The research shows that: (1) the proposed CLBIB-VSD-CART algorithm not only presents the economic meaning of the price characteristics of financial assets, but also can effectively control the price rise and fall limit; (2) The three performance indexes of the algorithm including the cumulative return, Sharpe ratio and Sotino ratio are significantly higher than that of the two benchmark models and other timing systems, and the investment performance is relatively stable; (3) The proposed research method provides an effective decision-making basis for investors to choose the best trading time, and has important application value.

**Keywords:** Decision tree; Quantitative trading; Timing signal; Commodity futures; Algorithm

**收稿日期：**2021-03-08 **修改日期：**2021-12-14

**基金项目：**国家自然科学基金资助项目（71961004, 72061007）

**作者简介：**张茂军（1977-），男，山西忻州人；苏州科技大学商学院教授，硕士生导师；研究方向：金融风险；饶华城（1996-），男，广东梅州人，硕士，研究方向：金融风险；南江霞（1978-）（通讯作者），女，山西运城人，教授，研究方向：供应链金融；王国栋（1985-），男，山西永济人，高级经济师，硕士，研究方向：债务投融资管理。

## 1. 引言

量化择时是量化交易中的一个热点研究方向,它是使用数量化的方法对某种金融资产在特定的时刻进行买入或卖出操作。量化择时方法可以分为两大类:基本面分析和技术分析。基本面分析是分析影响金融资产未来价格走势的基本面因素,进而预测金融资产的未价格走势。对于不同的金融资产而言,影响价格走势的基本面因素也不相同。例如,影响股指期货的基本面因素有宏观经济、利率水平、货币供应等,影响商品期货的基本面因素是该商品在整个行业中的供给和需求以及政策因素等<sup>[1]</sup>。Pierdzioch 等<sup>[2]</sup>使用通货膨胀率、期限利差等基本面信息可以预测黄金期货超额收益。Enke 和 Thawornwong<sup>[3]</sup>利益基本面指标构建神经网络模型可以有效预测股票收益率。

技术分析使用计算出来的指标数据指导实际操盘。关于技术分析的定義,Pring<sup>[4]</sup>认为技术分析是在早期阶段识别趋势变化,并保持投资状态,直到有大量证据表明趋势已经反转。Murphy<sup>[1]</sup>总结了技术分析的假设条件:①政治、基本面以及供求关系等对标的价格上涨或者下跌的影响,都会在价格和交易量中反映出来;②价格存在趋势中波动;③历史会重演。在国内外的研究当中,技术分析应用于股票、期货以及外汇市场的有效性已经得到了广泛研究<sup>[5-9]</sup>。早期研究主要使用单一技术指标进行分析。但是随着金融市场的发展,交易也变得越来越复杂,使用单一技术指标已经无法分析市场规律<sup>[10]</sup>。因此,学者们开始探索使用多种技术指标结合的量化择时方法,将多种技术指标与特定方法或模型进行结合。Mabu 等<sup>[11]</sup>、林文修和蔡秉洲<sup>[12]</sup>、Kim 等<sup>[13]</sup>以及孔傲等<sup>[14]</sup>用遗传算法寻找最优技术指标的组合,梁洪俊等<sup>[15]</sup>使用 Cox 模型将指标信息进行组合分析。也有学者使用机器学习算法对大量技术指标进行挖掘,进而发现金融资产的价格变化趋势<sup>[16]</sup>。

神经网络和支持向量机等机器学习方法在金融量化交易中得到了广泛应用。Tenti<sup>[16]</sup>用平均动向指数、趋势运动指数和变化率等技术指标,借助于循环神经网络预测期货价格走势。Kamruzzaman 和 Sarker<sup>[17]</sup>发现神经网络模型的预测效果都要优于 ARIMA 模型,准确率达到 80%。禹建丽等<sup>[18]</sup>使用收盘价、成交量和移动平均值等技术指标作为神经网络模型的输入,对中国股票市场的价格趋势进行预测。Kourentzes 等<sup>[19]</sup>发现集成学习方法可以提高多个神经网络模型的鲁棒性和准确性。Kim<sup>[20]</sup>以动量指标、威廉指标和商品通道指数等技术指标作为输入特征,分别使用了支持向量机、神经网络和最近邻算法预测 KOSPI 股票价格走势。Tay 和 Cao<sup>[21]</sup>发现在期货预测方面支持向量机比神经网络具有更高的准确率。Trafalis 和 Ince<sup>[22]</sup>使用类似的方法研究了美国股票市场预测。陈荣达和虞欢欢<sup>[23]</sup>使用启发式算法提高支持向量机模型的准确率,该策略的年化收益率要远远大于基准组合。因此,基于机器学习的技术分析成为了预测金融资产价格的有效工具。

然而,现有的文献存在一定的不足之处:一是已有文献将技术指标数值直接作为机器学习模型的特征输入,并没有考虑技术指标的具体含义;二是构建机器学习中的标签是基于价格的涨跌,价格上涨将标签标记为“+1”,否则将标签标记为“-1”,这就造成了模型只是预测涨跌,并没有考虑到涨幅或者跌幅。为此,本文构建了基于多个技术指标的机器学习量化择时系统。该系统探讨了机器学习方法在金融资产价格预测中的两个问题:一是如何对技术指标进行处理,二是改进无监督学习的标签构造方法。对于第一个问题,从技术指标所代表的市场含义出发,对连续型技术指标进行二元离散化和三元离散化,并在此基础上利用信息系数方法对离散后的技术指标进行选择。对于第二个问题,我们在构造标签的过程中考虑了涨幅和跌幅两个重要因素。我们将原有固定时间区间方法和区间突破方法结合,其中固定时间区间可以理解为目标持有时间的限制,区间突破的上区间为止盈线、下区间为止损线。在时间限制、止盈线和止损线三个限制中,只要有一个限制被打破就退出市场。另外,我们通过计算从进入市场到退出市场整个持有期间的收益情况来构造分类标签,如果在标的持有期间收益为正则将标签标记为“+1”,否则将标签标记为“-1”。

本文的研究主要有以下三点贡献：第一，丰富了技术指标的处理方法。已有研究主要是将计算出来的技术指标进行二元离散化处理后，再将其作为机器学习的输入特征，这种方法并不能体现出技术指标的特性。为此，我们提出了值特征选择方法，探讨了三元离散化的处理方式并且用信息系数对离散结果进行选择，实验结果证明该方法的预测效果要优于二元离散化方法。第二，创新性地借助于决策树算法提出了基于区间突破的标签构造方法。已有固定时间区间方法的标签构造只考虑了金融资产价格的涨跌情况，本文构造标签还考虑了价格的涨跌幅度，即将价格波动幅度加入了上下区间的限制，价格突破上区间限制进行止盈操作，反之价格突破下区间限制则进行止损操作，这样构造出来的标签更加符合实际交易逻辑。第三，提出了螺纹钢商品期货交易的 CLBIB-VSD-CART 量化择时系统。实证结果显示 CLBIB-VSD-CART 量化择时系统在累积收益率、夏普比率和索提诺比率三个绩效指标都显著高于两个基准模型和其他择时系统。该系统为期货交易提供了有效的量化择时方法。

论文的其余部分安排如下：第二节为技术指标的构建与决策树模型的基本算法，第三节为量化择时系统的构建与评估指标体系，第四节检验择时系统的有效性并且对比分析其稳健性。

## 2 技术指标和分类决策树

### 2.1 技术指标

在预测金融资产价格时，用开盘价、最低价、最高价和成交量等历史交易数据可以计算技术指标。常用技术指标分为趋势类、动量类、能量类和区间突破类等，下面给出几类重要的技术指标。

(1) SMA (Sample Moving Average)。移动平均线计算一段时期内价格的平均值，是一种通过平滑数据消除金融资产价格的短期波动、分析价格长期趋势的技术指标。

$$SMA(N) = \frac{1}{N} \sum_{i=1}^N P_{i,close}$$

式中， $p_{i,close}$  为当日收盘价， $N$  为平移天数。

(2) EMA (Exponential Moving Average)。考虑到时间越近的价格影响越大，指数移动平均线通过加权使各期价格的影响力随着时间递减。

$$EMA(N) = \begin{cases} p_{1,close} & \text{if } N = 1 \\ \frac{2}{N+1} * p_{N,close} + \frac{N-1}{N+1} * EMA(N-1) & \text{if } N > 1 \end{cases}$$

(3) MACD (Moving Average Convergence and Divergence)。平滑异同移动平均线是一种判断金融资产买卖时点以及价格运行趋势的技术指标，其原理是运用短期指数移动平均线  $EMA(N_{fast})$  和长期指数移动平均线  $EMA(N_{slow})$  交叉和分离的信息进行平滑运算，其中  $\alpha$  为平滑系数，短期和长期的 EMA 最常用的是 12 和 26 日的 EMA。

$$\begin{aligned} DIF(i) &= EMA(N_{fast}) - EMA(N_{slow}) \\ DEA(i) &= \alpha * DEA(i-1) + (1-\alpha) * DIF(i) \\ MACD(i) &= 2 * (DIF(i) - DEA(i)) \end{aligned}$$

(4) MOM (Momentum Index)。动量指标是一种专门研究金融资产价格波动速度的技术指标，其原理是金融资产价格的涨幅或者跌幅会随着时间的推移而慢慢减小。

$$MTM(n) = p_{i,close} - p_{i-n,close}$$

式中， $p_{i,close}$  为当日收盘价， $p_{i-n,close}$  为  $n$  日前的收盘价， $n$  为计算参数，一般情况下  $n=6$ 。

(5) WR (Williams %R)。威廉指标是一种研究金融资产价格波动幅度的技术指标，其原理是分析一段时间内金融资产价格的最高价、最低价和收盘价的关系，通过三者的关系反应当前价格买卖气势强弱，借此判断当前价格与理性投资相背离的程度。

$$W\%R(n) = \frac{H_n - p_{i,close}}{H_n - L_n} \times 100$$

式中， $p_{i,close}$ 为当日收盘价， $H_n$ 和 $L_n$ 是 $n$ 个交易日来的最高价和最低价。

(6) RSI (Relative Strength Index)。相对强弱指标与威廉指标类似，是一种研究金融资产价格波动幅度的技术指标，其原理是计算一段时间内的涨幅与跌幅的比值，反映了多空双方的力量强弱程度。该技术指标可被投资者用来判断行情的变化以及预测未来金融资产价格的走势。

$$RS(n) = \frac{n\text{日内收盘价涨数和的均值}}{n\text{日内收盘价跌数和的均值}}$$

$$RSI(n) = 100 - \frac{100}{1 + RS(n)}$$

(7) OBV (On Balance Volume)。能量潮指标是一种专门研究成交量的技术指标，认为成交量往往要比价格先行，可以通过成交量的变化来判断价格变化。

$$OBV_i = \begin{cases} 0 & , i = 1 \\ OBV_{i-1} + \text{sgn}(p_{i,close} - p_{i-1,close}) \times vol_i & , i > 1 \end{cases}$$

(8) VR (Volume Ratio)。成交量变异率指标与能量潮指标类似，是一种研究量和价格关系的技术指标，其原理是成交量的变化会引发价格的改变，成交量增加价格随之上涨，成交量下跌价格随之下跌，所以观察成交量可以判断金融市场的行情变化。

$$ANS(n) = \sum_{i=1}^{n-1} I(p_{i,close} < p_{i+1,close}) \times vol_i$$

$$BNS(n) = \sum_{i=1}^{n-1} I(p_{i,close} > p_{i+1,close}) \times vol_i$$

$$CVS(n) = \sum_{i=1}^{n-1} I(p_{i,close} = p_{i+1,close}) \times vol_i$$

$$VR(n) = \frac{ANS(n) + \frac{1}{2} \times CVS(n)}{BNS(n) + \frac{1}{2} \times CVS(n)}$$

(9) MFI (Money Flow Index)。资金流量指标是一种测算金融资产供求关系和多空双方力量强度的技术指标。资金流量指标是威尔斯在相对强弱指标的基础上改进而来，相对强弱指标只考虑了价格的变化，而资金流量指标还加入了交易量变化的，综合上涨和下跌的天数以及成交量的增加和减少幅度四个变量。

$$RMF_i = \frac{(p_{i,high} + p_{i,low} + p_{i,close}) \times Vol_i}{3}$$

$$PMF_i = \begin{cases} RMF_i, & \text{if } p_{i,close} > p_{i-1,close} \\ 0, & \text{if } p_{i,close} \leq p_{i-1,close} \end{cases}$$

$$NMF_i = \begin{cases} RMF_i, & \text{if } p_{i,close} < p_{i-1,close} \\ 0, & \text{if } p_{i,close} \geq p_{i-1,close} \end{cases}$$

$$R_i = \frac{\sum_{j=1}^n PMF_{i-j+1}}{\sum_{j=1}^n NMF_{i-j+1}}$$



$$MFI_i = 100 - \frac{100}{1 - R_i}$$

(10) ROC (Rate of Change)。变动速率指标一种研究金融资产价格变化动力大小的技术指标,其原理是通过比较当期收盘价和  $n$  期前收盘价,计算金融资产在一段时间内收盘价的变动速率。使用该变动速率对金融资产价格的动量进行分析,判断价格运行的走势。

$$ROC(n)_i = \frac{p_{i,close} - p_{i-n,close}}{p_{i-n,close}}$$

## 2.2 分类决策树

如图 1 所示,分类决策树是一种描述对实体进行分类的树形结构<sup>[25]</sup>,具体实现如算法 1 所示。决策树由父节点  $t_p$  以及父节点的左右子结点  $t_l$  和  $t_r$  构成。规定数据集矩阵  $X$  有  $M$  个特征变量  $x$  和  $N$  个观测值,类别向量  $Y$  有  $N$  个观测值以及  $K$  个类。

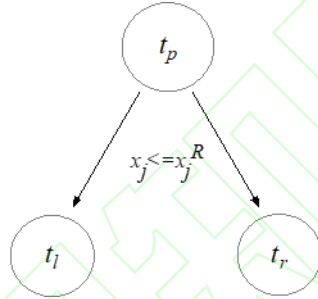


图 1 CART 分裂算法

分类决策树是一个递归过程,使用分裂规则递归地将训练样本划分为较小的部分,每一次递归将会把数据划分为两个部分,并且每个划分后的样本具有最大同质性。子节点的最大同质性是由杂质函数  $i(t)$  定义,对于所有的可能的分裂点  $x_j \leq x_j^R, j = 1, \dots, M$ , 其父节点的杂质都是恒定的,所以子节点的最大同质性等于杂质函数变化最大值  $\Delta i(t)$

$$\Delta i(t) = i(t_p) - E[i(t_c)]$$

式中,  $t_c$  为父结点  $t_p$  的左右子结点,  $E[\cdot]$  为均值算子。假设  $p_l, p_r$  为左右子结点的概率, 上述公式可以转化为:

$$\Delta i(t) = i(t_p) - p_l i(t_l) - p_r i(t_r)$$

因此在每一个节点, 决策树算法将搜索矩阵  $X$  中所有可能的变量  $x_j$ , 获得最佳分裂判断规则  $x_j \leq x_j^R$ , 使得杂质度量  $\Delta i(t)$  最大化, 即求解以下优化问题

$$\underset{x_j \leq x_j^R, j=1, \dots, M}{\operatorname{argmax}} [i(t_p) - p_l i(t_l) - p_r i(t_r)]$$

决策树算法非常多, 较为常用的决策树算法有 CART、ID3 和 C4.5, 这些决策树算法的区别主要是在杂质函数的选取。Breiman<sup>[26]</sup>提出的 CART 决策树算法使用基尼系数作为杂质函数、D3 决策树算法使用信息增益作为杂质函数, Quinlan<sup>[27]</sup>提出的 C4.5 决策树算法使用信息增益比作为杂质函数。本文选取业界较为常用的基尼系数作为杂质函数:

$$i(t) = \sum_{k \neq l} p(k|t)p(l|t)$$

其中,  $k, l = 1, \dots, K$  为类别的索引,  $p(k|t)$  是在结点为  $t$  的情况下类别  $k$  的条件概率。将上述基尼杂质函数带入优化问题, 可得以下式子

$$\Delta i(t) = -\sum_{k=1}^K p^2(k|t_p) + P_l \sum_{k=1}^K p^2(k|t_l) + P_r \sum_{k=1}^K P^2(k|t_r)$$

因此，基于基尼系数的决策树算法将需要求解以下优化问题

$$\underset{x_j \leq x_j^R, j=1, \dots, M}{\operatorname{argmax}} \left[ -\sum_{k=1}^K p^2(k|t_p) + P_l \sum_{k=1}^K p^2(k|t_l) + P_r \sum_{k=1}^K P^2(k|t_r) \right]$$

### 3 基于决策树的量化择时系统

本文提出的 CLBIB-VSD-CART 量化择时系统分为以下四个步骤：第一步、设置单个技术指标参数，第二步、根据技术指标自身特性进行特征离散化，第三步、构造决策树训练中使用的分类标签，第四步、择时信号绩效评估。具体流程如下图 2 所示：

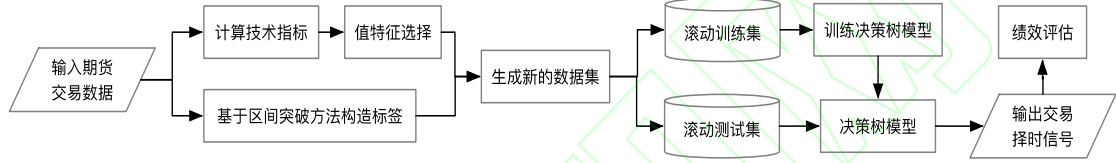


图 2 CLBIB-VSD-CART 择时系统预测流程

#### 3.1 单个技术指标的参数设置

技术指标利用金融资产过去一段时间内的价格和成交量等历史信息反映当前市场状态。技术指标的计算依赖于时间窗口的选择，不同的时间窗口计算出来的技术指标可能会产生截然相反的交易信号。本文使用的时间窗口最小为 5 天，最大为 30 天。除时间窗口外，本文使用的其他参数（如威廉指标的阈值参数等）取值范围如表 1 所示（详见附录），其具体取值为：威廉指标指标的阈值参数为 80，相对强弱指标的阈值参数为 35，成交量变异率的阈值参数为 40 和 160，资金流量指标的阈值参数为 25，变动速率指标的阈值参数为 5。

#### 3.2 构造分类标签

设  $X$  为某金融资产的数据集（如商品期货），该数据集中包括金融资产的开盘价、最低价、最高价和交易量。一般地，将第  $i$  天的开盘价、最低价、最高价、收盘价以及成交量分别记为  $p_{i,open}$ 、 $p_{i,low}$ 、 $p_{i,high}$ 、 $p_{i,close}$  和  $vol_i$ ，数据集  $X$  表示如下

$$X = \begin{bmatrix} p_{1,open} & p_{1,low} & p_{1,high} & p_{1,close} & vol_1 \\ p_{2,open} & p_{2,low} & p_{2,high} & p_{2,close} & vol_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{n,open} & p_{n,low} & p_{n,high} & p_{n,close} & vol_n \end{bmatrix} \quad (1)$$

本文使用的分类决策树模型为有监督学习。无监督学习和有监督学习最大的区别在于有监督学习训练过程中需要用到分类标签。Gerlein 等<sup>[28]</sup>、Chen 和 Hao<sup>[29]</sup>以及 Pang 等<sup>[30]</sup>构造分类标签均为固定时间区间方法（Fixed-time Horizon Method，记为 FHM），如式（2）计算金融资产在一段时间内收盘价的价差变化  $\Delta C_s^i$ ，其中  $m$  为时间间隔。假如  $\Delta C_s^i \geq 0$  将标签标记为“+1”，否则标记为“-1”。

$$\Delta C_s^i = p_{i+m,close} - p_{i,close}, \forall i = 1, 2, \dots, n - m \quad (2)$$

当标签为“+1”时，表明这段时间内资产价格上升，投资者应当买入资产；反之，当标签为“-1”时，表明资产价格下降，投资者应当卖出资产。

尽管，固定时间区间的方法简单直观，但这种方法并不科学。在实际运用中对某个金融资产持有时间不可能是固定的，还需要考虑到止盈止损等问题。因此本文在固定时间区间的基础上，提出了基于区间突破的构造标签方法（Construction Label Based on Interval Breakthrough，记为 CLBIB）。该方法通过区间突破规则对买入的资产在固定持有时间限制内加入了止盈和止损线。当资产价格突破持有时间限制内的止盈线或止损线其中之一，就退出市场。类似地，计算资产从进入到退出市场这段时间内的收益，假如收益为正，将标签标记为“+1”，否则将标签标记为“-1”。止盈线  $StopProfitLine_i$  和止损线  $StopLossLine_i$  的计算公式如下。

$$MID_i = \frac{1}{n} \sum_{j=i-n}^{i-1} p_{j,close}$$

$$TR_i = \max\{p_{i,high} - p_{i,low}, |p_{i,close} - p_{i,high}|, |p_{i,close} - p_{i,low}|\}$$

$$ATR_i = \sum_{j=i-n}^{i-1} TR_j$$

$$StopProfitLine_i = MID_i + k \times ATR_i$$

$$StopLossLine_i = MID_i - (k - 1) \times ATR_i$$

### 3.3 值特征选择

现有的大部分研究将计算出来的技术指标数值作为机器学习模型输入特征，使用训练好的模型预测未来金融资产价格涨跌。但是技术指标的数值本身有其特殊含义和性质，使用技术指标的特性能够更好的预测金融资产价格的未来走势。Patel 等<sup>[31]</sup>针对以上问题提出了趋势确定性数据层方法（Trend Deterministic Data Preparation Layer，记为 TDDPL），该方法将连续的技术指标数值进行离散化，以凸显每个技术指标的特性，进而提高机器学习模型的预测精度。

以 RSI 相对强弱指标为例：当 RSI 指标大于等于 70 的时候，认为金融资产的价格被高估，未来将会下跌，此时将数值离散化为“-1”；当 RSI 指标小于等于 30 的时候，认为金融资产被低估，价格将会上涨，此时将数值离散化为“+1”；当 RSI 指标数值在 30 和 70 之间，且 RSI 指标数值大于前一刻的 RSI 数值，也离散化为“+1”。当 RSI 指标数值在 30 和 70 之间，且 RSI 指标数值小于或等于前一刻的 RSI 数值，技术指标将离散化为“-1”。

TDDPL 方法将技术指标离散化为“+1”或“-1”，所以本文将 TDDPL 离散化方法命名为二元特征离散化（Binary Discretization, BD）。当技术指标为“+1”时，表示发出“买入”的交易信号；当技术指标为“-1”时，表示发出“卖出”的交易信号。但是，在实际交易过程中有“买入”、“卖出”和“观望”三种状态。因此，本文在二元特征离散化方法的基础上引入三元特征离散化（Ternary Discretization，记为 TD）。仍然以 RSI 相对强弱指标为例：当 RSI 指标大于等于 70 的时候，将数值离散化为“-1”，表示发出“卖出”信号；当 RSI 指标小于 30 的时候，将数值离散化为“+1”，表示发出“买入”信号；当 RSI 指标数值在 30 和 70 之间，将数值离散化为“0”，表示发出“观望”信号。图 3 为改进后的 TDDPL 特征离散化方法示意图。



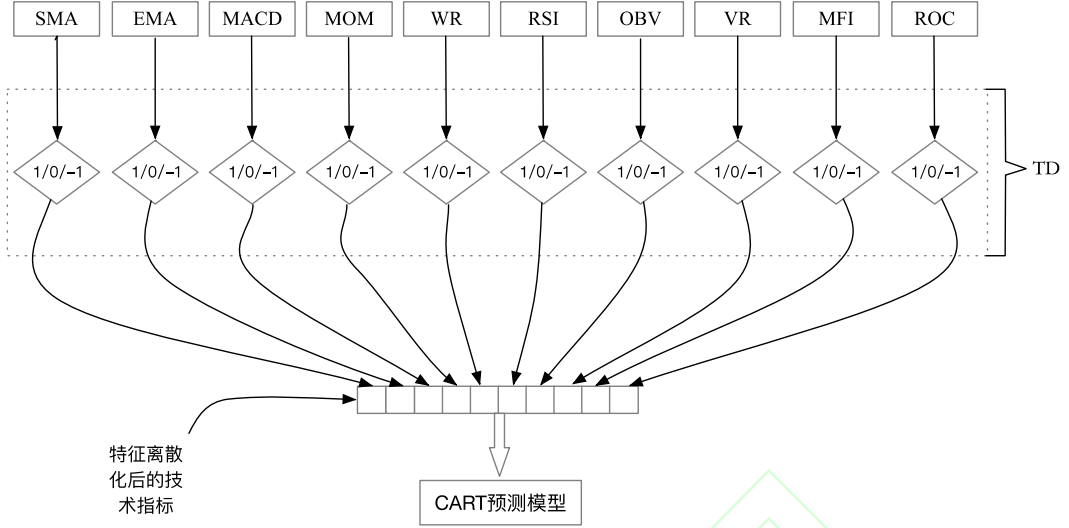


图 3 三值特征离散化方法流程图

本文在实证研究中发现：三元特征离散化在不同的期货品种中预测效果不一样，有些期货品种的预测效果优于二元特征离散化，有些期货品种的预测效果则较二元特征离散化差。因此，如图 4 所示，我们在两种特征离散化的基础上提出值选择特征离散化（Value Selection Discretization，记为 VSD），该方法计算二值特征和三值特征的信息系数（Information Coefficient，记为 IC）。信息系数是量化研究中用来衡量特征和收益率的相关性的指标，式（3）为信息系数的计算公式，信息系数的绝对值越大，表明特征预测收益的能力越强。

$$\text{Rank IC} = \frac{\text{cov}(\text{order}_{t-1}^f, \text{order}_t^r)}{\sqrt{D(\text{order}_{t-1}^f)D(\text{order}_t^r)}} \quad (3)$$

式中， $\text{order}_{t-1}^f$  为滞后一期的特征排名值， $\text{order}_t^r$  为收益率的排名值。

使用信息系数来衡量两种离散化方法得到的特征预测收益能力的强弱。因此对于每一个技术指标，通过比较两种不同的离散化方法计算的信息系数绝对值大小，保留较大的离散化结果。最后，类似文献<sup>[31]</sup>，我们在 10 个技术指标中按照信息系数绝对值大小选择 6 个技术指标作为模型的输入。

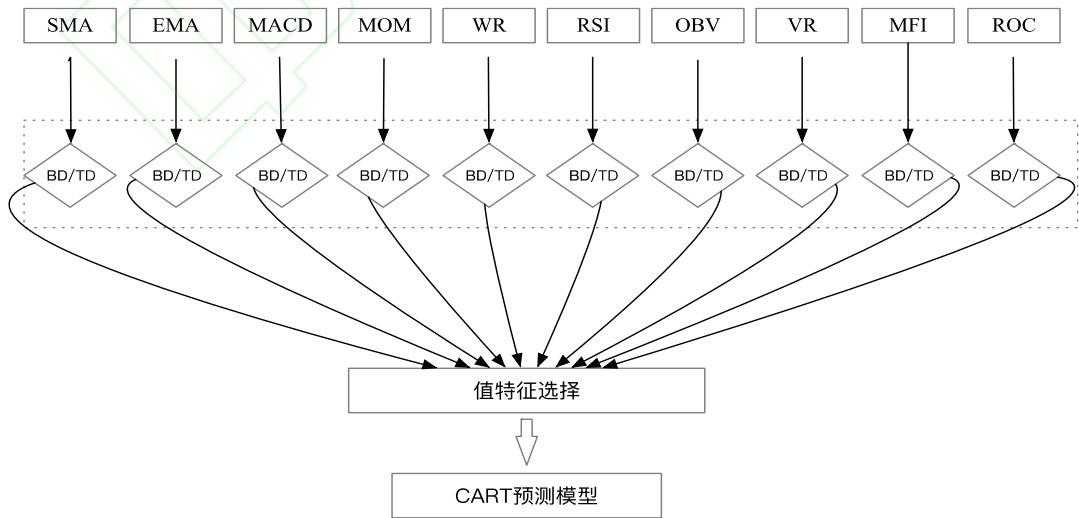


图 4 值选择特征离散化方法流程图

### 3.4 择时信号绩效评估

Henrique 等<sup>[32]</sup>总结了用机器学习预测金融资产价格时，不同预测类型所使用的绩效评估标准，如平均绝对误差（Mean Absolute Error，记为 MAE）和均方误差（Mean Square Error，记为 MSE）等。此外，因为预测精度高的模型并不意味着实际交易获得高的收益，所以使用收益率评价模型或者策略的绩效，如年化收益率（Annualized Yield）和夏普比率（Sharpe Ratio）等。本文的择时模型属于预测金融资产价格趋势，类似地使用择时准确率、累积收益、夏普比率和索提诺比率四个标准来衡量择时绩效。

择时准确率为整个交易期内预测正确的比率，也就是在整个交易期内预测信号为“买入”且预测结果为“买入”的次数加上预测信号为“卖出”且预测结果为“卖出”次数占全部信号的比率。累积收益是交易开始至交易结束期间的收益和，第  $i$  期的累积收益计算公式如下：

$$CumulativeRetrun_i = (1 + CumulativeRetrun_{i-1}) \times (1 + Retrun_{i-1}) - 1$$

夏普比率是一种综合考虑风险和收益的指标，其定义为超额收益的均值除以收益的标准差。假设  $r_f$  为无风险收益率，则夏普比率的计算公式如下：

$$SharpRatio = \frac{E(r - r_f)}{\sigma_r}$$

夏普率越高说明投资者承担单位风险获得的收益越高，此时的投资绩效表现越好。

索提诺比率是基于夏普比率改进的指标，在计算收益率的标准差（风险）时，只考虑下行风险，上涨所带来的波动不计入风险计算中。索提诺比率的计算公式如下：

$$SortinoRatio = \frac{E(r - r_f)}{\sqrt{\frac{1}{T} \sum_{t=0}^T (r_{pt} - r_f)^2}}, \text{其中 } (r_{pt} < r_f)$$

## 4 实验分析

### 4.1 实验数据

为了更好的追踪某一期货品种的行情变化，可以构造商品期货价格指数作为商品期货的交易信号。业界常用的商品期货价格指数有文华财经商品指数、Wind 商品品种指数和通达信商品指数等，本文研究使用 Wind 商品品种指数。本文使用持仓量较高且流动性较好的螺纹钢商品期货。时间范围是 2010 年 1 月 4 日至 2020 年 10 月 27 日。数据为日度数据，表 2 为螺纹钢指数收盘价的描述性统计分析，图 5 为螺纹钢指数收盘价的价格走势及日收益率折线图。

表 2 螺纹钢价格指数的描述性统计

	观测数	最大值	最小值	均值	标准差	偏度	峰度
开盘价	2589	5139.52	1625.86	3505.59	797.46	-0.40	-0.35
最高价	2589	5139.52	1637.24	3524.47	791.03	-0.43	-0.33
最低价	2589	5139.52	1615.80	3487.30	803.41	-0.37	-0.37
收盘价	2589	5139.52	1624.40	3505.44	797.39	-0.40	-0.35

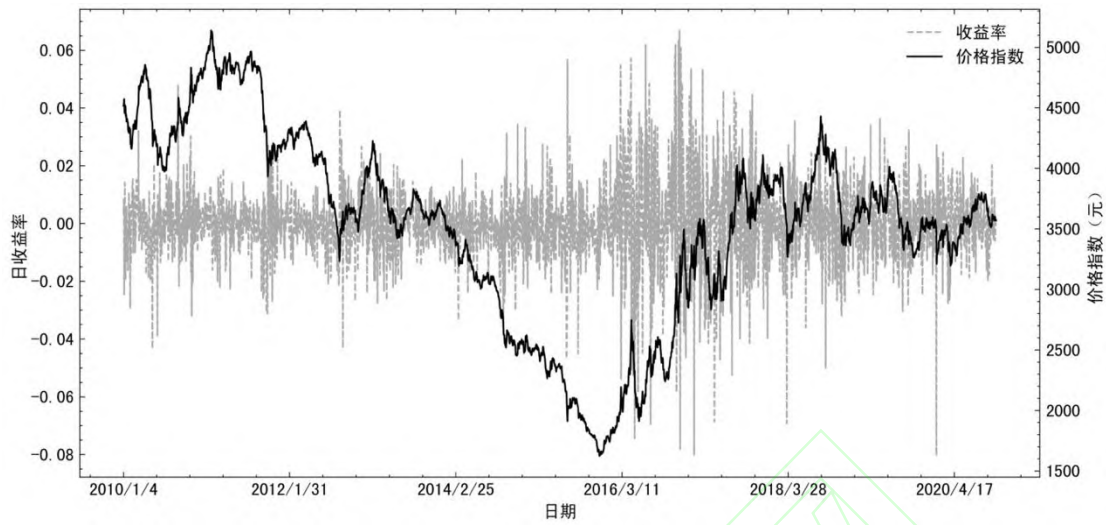


图5 螺纹钢价格走势及日收益率

由图5可知，价格趋势在2010年至2012年前小幅上涨，2012年以后，价格趋势下跌，直至2016年开始回升。Gerlein等<sup>[28]</sup>认为机器学习算法在金融时间序列预测的训练集并非越大越好。很久以前的市场数据特征可能与当前市场的数据特征相差较大，这会导致训练出来的模型在测试集中表现较差。类似于Gerlein等的工作，本文使用滚动训练集的方法，每隔五个交易日对模型进行重新训练，且每一个滚动训练集的大小为250。

## 4.2 基准策略

类似于文献<sup>[33]</sup>，为了更好地衡量CLBIB-VSD-CART择时系统的表现，本文将CLBIB-VSD-CART择时系统的实证结果与两个基准策略进行比较。如表3所示，这两个基准策略是多数投票组合策略（TA）和随机策略（RAND）。买入并持有策略是在交易期初持有标的直到交易期末；多数投票组合策略是将本文十个技术指标规则进行组合，计算十个规则中“买入”和“卖出”信号的个数，比较两个信号的个数，个数较多的信号为多数投票组合的综合信号。随机策略是随机产生交易信号，并根据随机信号进行买入和卖出操作。

表3 实证研究中使用的基准策略

模型	描述	策略
TA	多数投票组合策略	依据十个指标信号综合分析得出交易信号
RAND	随机策略	依据随机信号进行买入和卖出

## 4.3 CLBIB-VSD-CART 择时系统投资绩效

本节将3.1~3.4节提出的CLBIB-VSD-CART择时系统应用于螺纹钢商品期货的模拟交易中，择时系统发出择时交易信号，依信号模拟进行买入和卖出操作。我们对该择时系统构造标签的参数 $k$ 、滚动训练集大小以及重训周期三个参数的选择进行讨论，构造标签参数 $k$ 的取值为2和3，滚动训练集大小从200开始每次增加50个样本，重训周期的参数设置为5、10、15和20。表4是27种不同参数组合的实验结果。可以看出，在三种不同的参数组合下，CLBIB-VSD-CART择时系统的夏普比率0.5564至0.9468之间波动，索提诺比率在1.2743

至 2.3296 之间波动, 说明该择时系统可以获得正收益, 是一个有效的投资系统。在这些不同的参数组合中, 累积收益最低为 59.48%, 而最高达到了 101.98%, 说明该择时系统在不同参数中的绩效表现差异较大。

表 4 CLBIB-VSD-CART 择时系统在不同参数下的投资绩效

参数组合	$k$	训练集大小	重训周期	累积收益率	夏普比率	索提诺比率
组合 1	2	200	5	97.58%	0.9183	2.1706
组合 2	2	200	10	81.73%	0.7672	1.7668
组合 3	2	200	15	85.02%	0.8035	1.8878
组合 4	2	200	20	87.76%	0.8288	1.9226
组合 5	2	250	5	101.98%	0.9468	2.3296
组合 6	2	250	10	97.06%	0.8971	2.2062
组合 7	2	250	15	98.85%	0.9147	2.2659
组合 8	2	250	20	86.64%	0.7940	1.8824
组合 9	2	300	5	82.59%	0.7616	1.8126
组合 10	2	300	10	70.06%	0.6532	1.5137
组合 11	2	300	15	75.69%	0.7095	1.6850
组合 12	2	300	20	59.50%	0.5568	1.2760
组合 13	3	200	5	94.16%	0.8856	2.0829
组合 14	3	200	10	79.93%	0.7501	1.7239
组合 15	3	200	15	85.38%	0.8024	1.8367
组合 16	3	200	20	87.65%	0.8277	1.9193
组合 17	3	250	5	95.18%	0.8826	2.1559
组合 18	3	250	10	93.19%	0.8604	2.1070
组合 19	3	250	15	93.75%	0.8669	2.1345
组合 20	3	250	20	85.88%	0.7869	1.8642
组合 21	3	300	5	83.38%	0.7697	1.8408
组合 22	3	300	10	74.20%	0.6936	1.6257
组合 23	3	300	15	76.52%	0.7171	1.7034
组合 24	3	300	20	59.48%	0.5564	1.2743

组合 1-12 为构造标签参数  $k$  为 2 的 12 种组合的结果, 其中累积收益最高为组合 5 (训练集大小为 250、重训周期为 5), 其累积收益达到 101.98%。在训练集大小为 250 的组合中, 模型的绩效总体上随着重训周期的增大而下降的。这验证了 Gerlein 等<sup>[28]</sup>的结论: 较远的市场数据特征可能与当前市场数据特征相差较大。出于预测的角度, 应尽量地使用与当前市场特征较相似的数据, 因此在选择重训周期的大小上应越小越好。组合 1-12 的构造标签参数  $k$  为 2, 组合 13-24 的构造标签参数  $k$  为 3, 组合 1-12 的绩效表现在总体上要优于组合 13-24, 说明在设置止损和止盈的区间时, 小区间更加趋向得到较优的绩效表现。因此, 在之后的实验参数设置中, 我们选择绩效表现最好的参数组合 5。

表 5 CLBIB-VSD-CART 择时系统与基准策略投资绩效情况

择时系统	择时准确率	夏普比率	索提诺比率	累积收益率
CLBIB-VSD-CART	0.5834	0.9468	2.3296	1.0198
TA	0.5317	0.5229	0.9533	0.3793
RAND	0.5000	0.1363	0.2049	0.1208

表 5 是 CLBIB- VSD-CART 择时系统与两个基准策略模拟交易的投资绩效情况。通过表 5 数据可以看出, 本文提出的 CLBIB- VSD-CART 择时系统预测准确率为 58.34%, 高于多数投票组合策略和随机策略两个基准策略, 说明该择时系统可以提升技术指标对金融资产价格变化的预测能力。从夏普比率和索提诺比率来看, CLBIB-VSD-CART 择时系统的数值均为最大。在索提诺比率上 CLBIB-VSD-CART 择时系统要远大于两个基准策略, 说明较基准策略而言, 本文提出的择时系统承担单位风险获得的收益远大于两个基准策略。可以发现, CLBIB-VSD-CART 择时系统的累积收益最高, 达到了 101.98%。与本文提出的择时系统比较, 多数投票组合策略表现次之, 多数投票组合的累积收益为 37.93%, 随机策略的累积收益率最低, 仅有 12.08%。

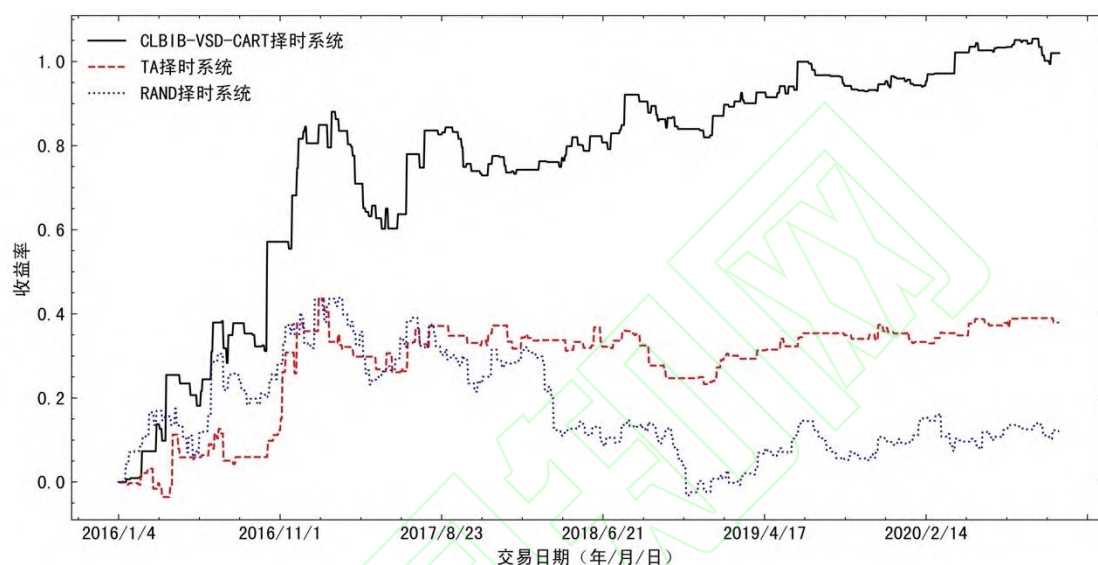


图 6 CLBIB-VSD-CART 择时系统与基准策略累积收益率曲线

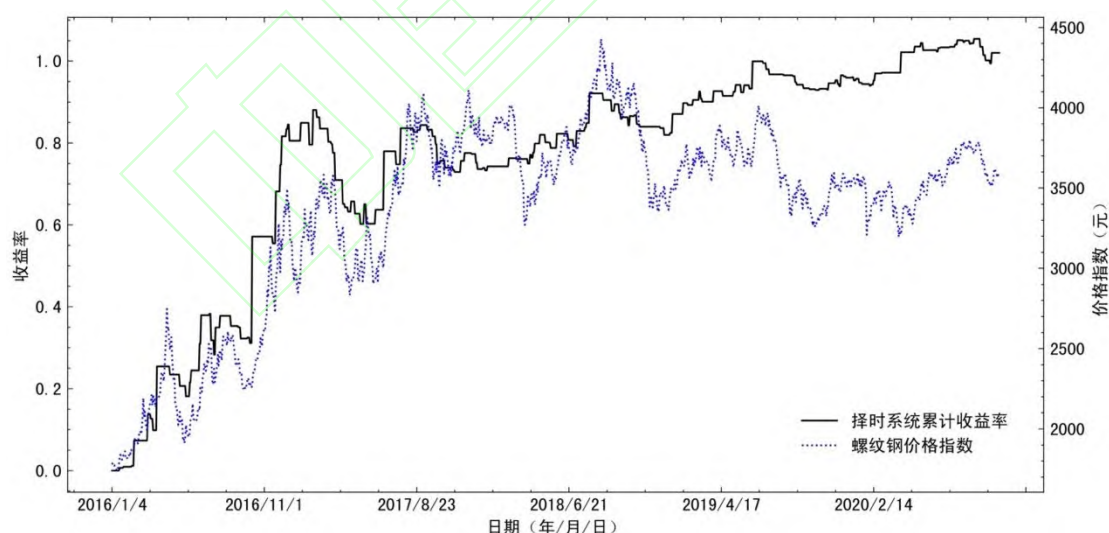


图 7 CLBIB-VSD-CART 择时系统累积收益率曲线与价格指数

图 6 绘制 CLBIB-VSD-CART 择时系统和基准策略的螺纹钢累积收益曲线。图 7 是 CLBIB-VSD-CART 择时系统累积收益率与螺纹钢价格指数曲线图。从该图中可知, 择时系统的累积收益曲线的方向与螺纹钢的价格走势非常相似。在 2016 年 1 月 4 日到 2016 年 12 月 13 日期间, 螺纹钢的价格从 1820 元一直涨到 3480 元, 涨幅接近一倍。在这时间段内,



择时系统的累积收益超过了 80%，说明 CLBIB-VSD-CART 择时系统能够较好地捕捉到价格上升趋势。此外，螺纹钢的价格指数从 2018 年 10 月的近 4200 元下降到 2020 年 2 月 3200 元的时候，CLBIB-VSD-CART 择时系统的累积收益率并没有随之下跌，而是稳中小幅增加，这说明 CLBIB-VSD-CART 择时系统能够较好地避免价格下行风险。

#### 4.4 择时系统比较分析

在现有文献中，标签构造方法常用的方法为固定时间区间法（Fixed-time Horizon Method, FHM），处理技术指标数据的方法为二元特征离散化（Binary Discretization, BD）。为了与现有方法对比，本节给出 CLBIB-BD-CART、FHM-VSD-CART、FHM-BD-CART 和 FHM-VSD-CART 四种择时系统，讨论三元特征离散化（Ternary Discretization, TD）与固定时间区间构造标签方法结合的 FHM-TD-CART 和 CLBIB-TD-CART 择时系统。表 6 是以上几种择时系统在螺纹钢商品期货上模拟交易的绩效表现。

表 6 CLBIB-VSD-CART 择时系统与其他择时系统投资绩效情况

择时系统	择时准确率	夏普比率	索提诺比率	累积收益率
CLBIB-VSD-CART	0.5834	0.9468	2.3296	1.0198
CLBIB-BD-CART	0.5713	0.6476	1.5842	0.6621
CLBIB-TD-CART	0.5216	0.6658	1.0572	0.5515
FHM-VSD-CART	0.6115	0.9582	1.9873	0.7010
FHM-BD-CART	0.6003	0.7020	1.5710	0.5765
FHM-TD-CART	0.6353	0.4088	0.6283	0.3353

由表 6 可知，在择时准确率方面，FHM-TD-CART 择时系统的择时准确率最高，达到了 63.53%；在夏普比率方面，FHM-VSD-CART 择时系统的夏普率要略微高于本文的 CLBIB-VSD-CART 择时系统，其夏普率为 0.9582；但是在索提诺比率方面，本文的 CLBIB-VSD-CART 择时系统反而要高于 FHM-VSD-CART 择时系统；最后从累积收益率上看，CLBIB-VSD-CART 择时系统的累积收益率最高。

在标签构造方法对比方面，使用本文提出的基于区间突破的构造标签方法（CLBIB）在择时准确率上要比固定时间区间方法（FHM）要低，但是从夏普比率、索提诺比率和累积收益率上看，基于区间突破的构造标签方法均远高于固定时间区间方法。这表明本文所构造标签的质量要高于现有文献中固定时间区间方法。另一方面，从处理技术指标数据的方法来看，本文提出的值选择特征离散化（VSD）方法的投资平均绩效优于已有的二元特征离散化方法。

#### 4.5 模型的稳健性检验

本节通过改变模型的训练集大小来检验 CLBIB-VSD-CART 择时系统的稳健性。一般情况下，训练集越大，使用的信息越多，择时系统的预测效果与投资绩效会更好。因此，在本节中我们讨论了 CLBIB-VSD-CART 择时系统在训练集在 200 至 290 的投资绩效情况，以检验训练集大小达到多少时模型趋于稳定。图 8 展示了训练集从 200 增加至 290 过程中，择时系统的累积收益、夏普比率、索提诺比率和择时准确率。

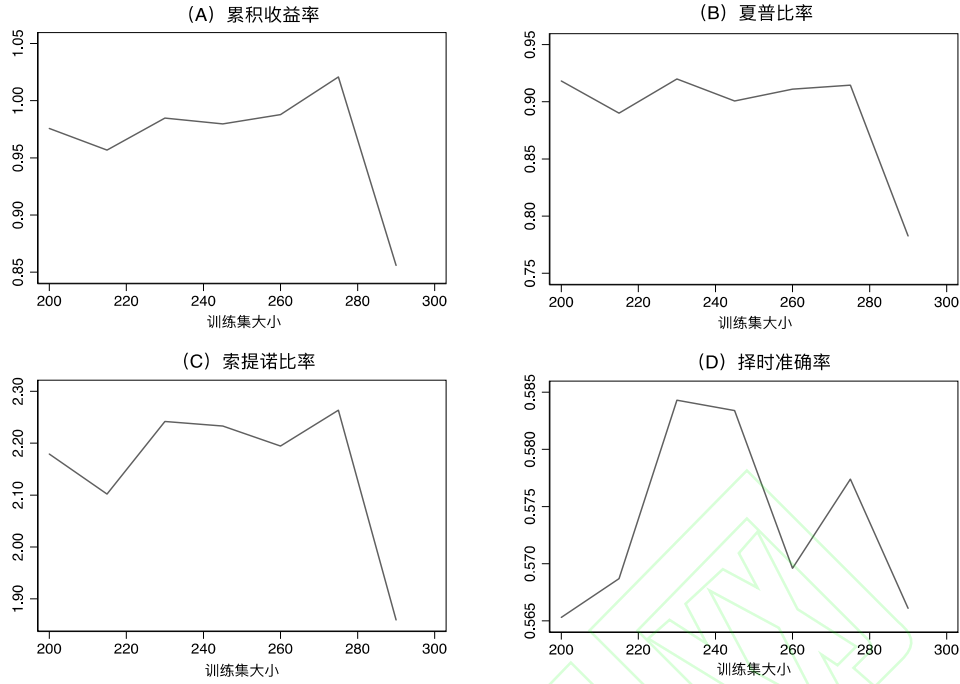


图 8 训练集大小与 CLBIB-VSD-CART 择时系统投资绩效

从该图中可以看出，累积收益率随着训练集的增大而提高，训练集大小在 240 至 260 的时候趋于稳定，当训练集增加到 270 左右，累积收益率开始下降；训练集大小在 200 至 240 的时候，夏普比率小幅波动。从 240 到 270，夏普率趋于稳定，270 之后夏普率则出现了大幅下降；训练集大小对索提诺比率的影响与夏普比率的影响类似；在择时准确率方面则存在不规则的波动，但是其数值可以认为稳定在 56.5%至 58.5%之间。综上分析可以认为：当训练集大小为 250 的时候，CLBIB-VSD-CART 择时系统有着稳健的投资绩效。

## 5 结论

随着量化交易逐渐成为了私募基金等机构投资者的主流投资模式，如何设计科学合理的量化择时系统成为了量化投资的核心问题。目前，机器学习方法已经成为了量化投资的重要工具，在预测金融资产价格方面具有重要作用。决策树是一类有监督学习的分类机器学习方法，可以对影响金融资产价格变化的特征因子进行分类。针对已有量化择时系统中存在的两个问题，即特征因子没有具体含义和没有考虑资产价格的涨跌幅度，本文提出了基于区间突破的分类标签和值特征选择方法，创新性地提出了 CLBIB-VSD-CART 量化择时系统。

首先，根据金融资产价格的开盘价、最低价、最高价和成交量等历史交易数据，构建了趋势、动量、能量和区间突破等技术指标，并且构建了决策树算法。其次，提出了基于区间突破的构造标签方法，通过区间突破规则对资产在持有时间限制内加入了止盈和止损线，并且利用三值特征的信息系数衡量特征和收益率的相关性，提出了 CLBIB-VSD-CART 择时系统。最后，用螺纹钢商品期货数据检验了 CLBIB-VSD-CART 择时系统的有效性和稳健性。

研究发现本文提出的量化择时系统在累积收益率、夏普比率和索提诺比率三个绩效指标都要显著高于两个基准模型和其他择择时系统的绩效。本文的研究丰富了量化投资的研究方法，为投资者选择最佳交易时间提供了有效的决策系统，具有重要的应用价值。

---

## 参考文献

- [1] Murphy J. Technical analysis of the futures market [M]. London: Prentice Hall Press, 1986.
- [2] Pierdzioch C, Risse M, Rohloff S. On the efficiency of the gold market: Results of a real-time forecasting approach[J]. International Review of Financial Analysis, 2014, 32(3):95-108.
- [3] Enke D, Thawornwong S. The use of data mining and neural networks for forecasting stock market returns[J]. Expert systems with applications, 2005, 29(4): 927-940.
- [4] Pring M J. Study Guide for Technical Analysis Explained Fifth Edition[M]. New York: McGraw Hill Professional, 2014.
- [5] Metghalchi M, Marcuci J, Chang Y H. Are moving average trading rules profitable? Evidence from the European stock markets[J]. Applied Economics, 2012, 44(12): 1539-1559.
- [6] Chiang Y C, Ke M C, Liao T L, Wang C D. Are technical trading strategies still profitable? Evidence from the Taiwan Stock Index Futures Market[J]. Applied Financial Economics, 2012, 22 (12): 955-965.
- [7] 张华, 任若恩. 基于 Dempster-Shafer 证据理论的外汇交易策略研究[J]. 数理统计与管理, 2013, 32(3): 452-461.
- [8] 周铭山, 冯新力, 林靛, 方旭赟, 周开国. A 股市场均线策略有效性与收益率随机特征研究[J]. 证券市场导报, 2013, (1): 58-64.
- [9] Mendens L, Codinho P, Dias J. A Forex trading system based on a genetic algorithm [J]. Journal of Heuristics, 2012, 18: 627-656.
- [10] Allen F, Karjalainen R. Using genetic algorithms to find technical trading rules[J]. Journal of Financial Economics, 1999, 51(2): 245-271.
- [11] Mabu S, Hirasawa K, Obayashi M. Enhanced decision making mechanism of rule-based genetic network programming for creating stock trading signals[J]. Expert Systems with Applications, 2013, 40(16): 6311-6320.
- [12] 林文修, 蔡秉洲. 集成基因表达规划法应用于动态股票交易策略探勘之研究[J]. 中国管理科学, 2015, 23(1): 510-517.
- [13] Kim Y, Ahn W, Oh K J. An intelligent hybrid trading system for discovering trading rules for the futures market using rough sets and genetic algorithms[J]. Applied Soft Computing, 2017, 55: 127-140.
- [14] 孔傲, 朱洪亮, 郭文旌. 一个基于技术指标规则的启发式量化择时系统[J]. 系统工程, 2019, 37(1): 111-122.
- [15] 梁淇俊, 郑贵俊, 徐守萍. 基于生存分析的择时策略择优体系研究——以技术指标交易信号为例[J]. 金融经济研究, 2015, 30(1): 96-106.
- [16] Tenti P. Forecasting foreign exchange rates using recurrent neural networks[J]. Applied Artificial Intelligence, 1996, 10(6): 567-582.
- [17] Kamruzzaman J, Sarker R. Comparing ANN based models with ARIMA for prediction of forex rates[J]. Asor Bulletin, 2003, 22(2): 2-11.
- [18] 禹建丽, 孙增圻, 刘治军. 基于 BP 神经网络的股市建模与决策[J]. 系统工程理论与实践, 2003, 23(5): 15-19.
- [19] Kourentzes N, Barrow D K, Crone S F. Neural network ensemble operators for time series forecasting[J]. Expert Systems with Applications, 2014, 41(9): 4235-4244.
- [20] Kim K. Financial time series forecasting using support vector machines[J]. Neurocomputing, 2003, 55(2): 307-319.
- [21] Tay F E H, Cao L. Application of support vector machines in financial time series forecasting[J]. omega, 2001, 29(4): 309-317.

- 
- [22] Trafalis T B, Ince H. Support vector machine for regression and applications to financial forecasting[C]// International Joint Conference on Neural Networks, IEEE Computing Society, 2000, 348-353.
- [23] 陈荣达, 虞欢欢. 基于启发式算法的支持向量机选股模型[J]. 系统工程, 2014, 32(2): 40-48.
- [24] Li J, Tsang E P K. Investment decision making using FGP: a case study[C] Proceedings of the 1999 Congress on Evolutionary Computation, 1999, 2: 1253-1259.
- [25] Quinlan J. Introduction of decision trees[J]. Machine Learning, 1986, 1(1): 81-106.
- [26] Breiman L, Friedman J, Stone C J. Classification and regression trees[M]. CRC press, 1984.
- [27] Quinlan J. C4.5: Programs for Machine Learning[M]. Morgan Kaufmann, 1992.
- [28] Gerlein E A, McGinnity M, Belatreche A, et al. Evaluating machine learning classification for financial trading: An empirical approach[J]. Expert Systems with Applications, 2016, 54: 193-207.
- [29] Chen Y, Hao Y. A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction[J]. Expert Systems with Applications, 2017, 80: 340-355.
- [30] Pang X, Zhou Y, Wang P, et al. An innovative neural network approach for stock market prediction[J]. The Journal of Supercomputing, 2020, 76(3): 2098-2118.
- [31] Patel J, Shah S, Thakkar P, et al. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques[J]. Expert systems with applications, 2015, 42(1): 259-268.
- [32] Henrique B M, Sobreiro V A, Kimura H. Literature review: Machine learning techniques applied to financial market prediction[J]. Expert Systems with Applications, 2019, 124: 226-251.
- [33] Gorenc Novak M, Velušček D. Prediction of stock price movement based on daily high prices[J]. Quantitative Finance, 2016, 16(5): 793-826.

## 附录

算法 1: CART 决策树生成算法

**输入:** 训练数据集  $D$ , 停止条件

- 1: **for**  $j = 1:n$  **do**
- 2: 对第  $j$  个连续型的特征  $X$  进行值特征选择离散化, 离散化后的结果记为  $X^*$
- 3: 从根节点开始
- 4: **while** 基尼系数最小 **do**
- 5:     **for**  $j = 1:n$  **do**
- 6:         对于第  $i$  个特征, 选择分裂点  $a$ , 计算该特征  $A=a$  时的基尼系数
- 7:     **end while**
- 8: 基尼系数最小的特征及其分裂点为最优特征与最优分裂点, 据此将训练集分为两个子节点
- 9: **if** 满足停止条件 **then** 退出
- 10: **else** 递归调用 4 - 9

**输出:** CART 决策树

表 1 技术指标的参数及其交易规则

指标名称	参数设置	交易规则
SMA	$n = 5 - 30$	价格上穿 SMA 时, 买入; 价格下穿 SMA 时, 卖出
EMA	$n = 5 - 30$	价格上穿 EMA 时, 买入; 价格下穿 EMA 时, 卖出
MACD	$fast = 12$ $slow = 26$	DIF 上穿 DEA 时, 买入 DIF 下穿 DEA 时, 卖出
MOM	$n = 5 - 30$	MOM 上穿 0 轴时, 买入 MOM 下穿 0 轴时, 卖出
WR	$n = 5 - 30$ $th = 60 - 90$	WR 上穿 $-th$ 时, 买入 WR 下穿 $-(100-th)$ 时, 卖出
RSI	$n = 5 - 30$ $th = 10 - 40$	RSI 上穿 $th$ 时, 买入 RSI 下穿 $100-th$ 时, 卖出
OBV	-	OBV 上穿 OBV_10MA, 买入 OBV 下穿 OBV_10MA, 卖出
VR	$th1 = 40 - 70$ $th2 = 160 - 450$	VR 下穿 $th1$ 时, 买入 VR 上穿 $th2$ 时, 卖出
MFI	$n = 5 - 30$ $th = 10 - 40$	MFI 下穿 $th$ 时, 买入 MFI 上穿 $100-th$ 时, 卖出
ROC	$n = 5 - 30$ $th = 5 - 30$	ROC 上穿 $th$ 时, 买入 ROC 下穿 $-th$ 时, 卖出