

# Analysis of the Rise and Fall of International Futures Based on Xgboost Algorithm

Jinqiang Li, Lei La

College of Information, University of International Business and Economics, Beijing

Email: 994114485@qq.com, lalei1984@aliyun.com

Received: Aug. 22<sup>nd</sup>, 2018; accepted: Sep. 7<sup>th</sup>, 2018; published: Sep. 14<sup>th</sup>, 2018

---

## Abstract

Based on an efficient and complex xgboost algorithm, a classification prediction model was constructed to train and test the daily transaction data of international futures in the past three years. The model traverses all parameter combinations through the reference tool to obtain the optimal parameters. Then, it is compared with decision tree, random forest, support vector machine algorithm, and combined with multiple evaluation indicators for comprehensive evaluation. Experiments show that the indicators of the model constructed by the xgboost algorithm are higher than other algorithms, and the overall prediction ability is better. At the same time, it also provides an effective new method for forecasting futures prices.

## Keywords

Futures, Xgboost, Change Forecast

---

# 基于Xgboost算法的国际期货涨跌预测分析

李进强, 喇磊

对外经济贸易大学信息学院, 北京

Email: 994114485@qq.com, lalei1984@aliyun.com

收稿日期: 2018年8月22日; 录用日期: 2018年9月7日; 发布日期: 2018年9月14日

---

## 摘要

基于高效复杂的xgboost算法构建了分类预测模型, 对最近三年国际期货的日交易数据进行了训练测试。该模型通过调参工具遍历所有参数组合, 得出最优参数。然后, 对比于决策树、随机森林、支持向量机算法, 结合多个评价指标进行综合评价。实验表明, xgboost算法构建的模型各项指标均高于其他算法,

综合预测能力更好。同时, 也为期货价格预测提供了一种有效的新方法。

## 关键词

期货, Xgboost, 涨跌预测

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

期货作为基础金融产品的重要衍生品, 在经济活动中具有转移风险和价格发现的功能, 因此, 期货市场是市场经济的重要组成部分, 对于促进社会资源的有效配置具有重要作用。期货市场又分为国际期货市场和国内期货市场, 国际期货市场相比国内期货市场品种更为齐全, 覆盖面更广。期货价格的涨跌反映了期货市场的运行状况, 在一定程度上也体现出当前经济状况, 所以对期货价格变化的准确预测, 对国家经济发展研究具有重要作用。同时, 期货交易者能否取得较高收益, 关键也在于对期货涨跌趋势的预判是否准确。相对于国内期货, 国际期货涨跌走势的预测更为复杂, 也更为重要。传统的预测期货价格的方法主要有基本分析法、技术分析法和组合分析法。但是, 随着科技金融化的到来, 以及投资者对量化投资策略的青睐, 传统的期货预测方法已无法满足海量数据及数据多样化, 而机器学习在金融经济领域的应用正好解决了这一问题。

## 2. 相关研究文献评述

金融市场的一切数据都具有随机性, 概率论与数理统计可以从数学角度来处理随机数据, 而精通金融理论的人进行技术分析时又往往会陷入习惯思维, 所以数理统计和金融理论的结合建模可以挖掘到更深层次的信息[1]。魏宇[2]以沪深 300 股指期货仿真交易的 5 分钟高频数据为例, 证明了已实现波动率模型以及加入附加解释变量的扩展随机波动模型预测精度高于基于日收益数据的历史波动率模型。Chunyang Wang [3]以大豆期货为例, 基于自回归相关系数、差分时间、及移动平均数构建 ARIMA 预测模型, 实验表明, 该模型对大豆期货收盘价的趋势预测较为准确, 并将其运用于其他农产品预测, 取得了不错的效果。董贺[4]制定了一套具有理论支撑的量化投资策略, 设定合理的开仓规则以及警戒点, 筛选出了国内期货市场农产品和金属大类的协整关系, 为期货价格预测提供了有价值的参考。刘海玥和白艳萍[5]构建 AR 模型、RBF 模型和 GRNN 模型滚动预测了上证指数开盘价、最高价、最低价和收盘价, 并与实际价格做对比, 发现三种模型预测误差均很小。Weina Wang [6]结合了模糊集合论和经典时间序列方法来预测股票价格, 在股票价格预测理论上做出了贡献。

但是, 大数据时代的到来, 期货市场积累了大量的数据, 依靠数学及金融理论构建的预测模型的效率已经无法满足需求。随着人工智能理论的发展成熟, 以机器学习算法构建量化投资预测模型的方法备受关注。人工智能以其强大的数据抓取能力和惊人的计算速度, 将进军证券投资领域, 促进金融市场资源配置的效率[7]。Sheng-Hsun Hsu [8]结合自组织映射(SOM)算法和支持向量机(SVM)算法构建两阶段架构模型来预测股票价格, 显著地提高了股票预测的性能。赵永进[9]运用决策树 ID3 算法选取具有代表性的股票相关财务指标, 并把关联规则应用到股票时间序列的发现上。结果表明, 比传统分析方法具有更

高的准确率。B Qian 和 K Rasheed [10]使用自动互信息和假近邻法生成训练模型所需的参数,并通过神经网络、决策树和 K 近邻算法构建模型的相互集成,实现了较高的股票价格预测精度。王刚和许晓兵[11]基于小波分析方法对股票每日最高价、最低价以及开盘价进行小波去燥处理,然后再使用神经网络进行预测分析,发现精度获得了大幅度提高。Karathanasopoulos A [12]等基于支持向量机算法进行了算法优化,引入了一种股票预测的混合方法,并与四个传统的投资策略进行了对比,结果表明,提出的方法具有更好的表现。可以看出,近年来,国内外学者将机器学习应用于量化投资领域更多的是预测股票走势及选出较好的量化选股方案等,将其应用于期货领域的较少,国际期货的预测研究更是缺乏,期货市场对经济发展也起着重要作用,所以对于国际期货的预测研究是很必要的。

Xgboost算法是2016年由陈天奇[13]博士提出的,该算法已经在众多领域取得了优异的成绩。张昊[14]将 xgboost 算法应用于商品推荐中,基于阿里巴巴的真实用户数据进行建模,得到了较准确的预测结果,对个性化推荐系统的完善具有重要的意义。蒋晋文和刘伟光[15]将 xgboost 算法应用与制造业质量预测中,处理了制造业生产过程中积累的大量数据,实现了准确预测产品质量的目的。李想[16]基于 xgboost 算法成功设计出了超越沪深 300 指数的超额收益率的多因子量化选股方案,并与随机森林、支持向量机进行了对比,验证了 xgboost 算法的稳定性与效果。

由此看出,数理统计和金融理论构建的传统预测模型在数据量很大的情况下,已经无法达到准确预测的效果。机器学习在金融经济领域的应用为股票期货价格预测开启了新篇章,这几年很多学者将机器学习算法构建的模型应用于股票价格预测领域,都取得了优异的成绩,但是,应用到国际期货价格预测领域的研究很少。传统期货预测方式已经在准确性和效率上都已经不满足要求,应用机器学习算法构建预测模型已成为趋势。xgboost 算法作为新出现的高效机器学习算法,在各个领域备受关注。所以,本文将使用 xgboost 算法构建模型预测国际期货价格涨跌。

### 3. 数据分析

从公开数据库下载的近三年期货日交易原始数据无法直接用于建模预测,必须对原始数据进行预处理,比如应用特征筛选、缺失值填充、量纲化等方式对原始数据进行加工处理。数据处理的好坏也影响着模型精度的高低,所以数据处理是分类建模的关键第一步。

#### (一) 数据描述

本文所使用的数据来自国泰安数据库,该数据是2015年1月1日到2018年3月30日的国际期货日交易数据,数据量为70,950条。数据集每条数据记录包含14个特征属性,分别是交易日期、GTA代码、合约名称、交易品种、日开盘价、日最高价、日最低价、日收盘价、涨跌值、涨跌幅度、成交量、持仓量、持仓量变化、标签,其中标签属性是在后续数据处理中新增加属性列。通过随机划分函数将数据集划分为训练集和测试集,训练集和测试集的样例个数比例为2比1,随机划分数据集避免了因人工划分所带来的误差,准确性更高。下面对数据集进行特征描述,见表1。

#### (二) 缺失值填充

缺失值的存在通常会影响到建模和预测质量,该数据集中关键属性列存在缺失值的数据记录占总数据集的7%左右。在缺失值比例不是很高的情况下,一般缺失值处理的方式是删除缺失值,但本文为了进一步提高模型精度,选择缺失值填充的方式,其中对于非数值型数据不进行填充。经过验证,该数据集缺失值填充为平均值最为合适。

由于该数据集时间跨度较大,所以采用分时间段均值填充。分别计算出2015年1月1日到2018年3月30日期间,每3个月即一个季度的平均值,将其作为对应时间段的缺失值填充值。比如,首先计算2015年1月到3月时间段中每个属性列的均值,将其作为这个时间段内缺失值填充值。

**Table 1.** Introduction to the characteristics of the data set  
**表 1.** 数据集特征属性介绍

属性列名称	含义	值
Trddt	交易日期	时间, 如 2018-03-30
Agmtcd	GTA 代码	如: TOPL1810
Agmtnm	合约名称	如: 白金 1810
Trdvar	交易品种	如: 白金、原油
Exgnur	交易所代码	如: TOCOM
Ffdt001	日开盘价	数值, 范围: 0.0~66710.0
Ffdt002	日最高价	数值, 范围: 0.0~67900.0
Ffdt003	日最低价	数值, 范围: 0.0~66480.0
Ffdt004	日收盘价	数值, 范围: 0.0~67900.0
Ffdt005	涨跌值	数值, 范围: -4530.0~4570.0
Ffdt006	涨跌幅度	数值, 范围: -6.73~6.8
Ffdt007	成交量	数值, 范围: 0.0~115834.0
Ffdt008	持仓量	数值, 范围: 0.0~90242.0
Ffdt008	持仓量变化	数值, 范围: -29662.0~55710.0
Label	标签	数值: 0 和 1

### (三) 特征工程

原始数据的所有属性都用于统计建模是不切合实际的, 必须根据模型要求和属性值特征进行特征工程构建。本文中特征工程构建分别通过删除无关属性列、增加新属性列、相关性分析完成。

#### 1) 特征筛选

特征的筛选需要能够充分反映国际期货涨跌趋势的情况, 同时, 基于模型的要求, 作为模型输入属性的数据必须是数值型数据。因此, 将数据集中 GTA 代码、合约名称、交易品种这三个属性列进行删除, 对于交易日期属性列, 被应用到缺失值填充过程中, 但不作为模型输入因子, 也将其删除。

本文对国际期货涨跌进行预测, 属于典型的二分类问题, 但原始数据集中并无分类标签, 所以对该数据增加新的属性列“label”, 作为分类标签。新属性的取值依据涨跌值或涨跌幅度的正负, 其中将涨跌为正的取值为 1, 将涨跌为负的取值为 0。

#### 2) 相关性分析

相关性分析是指对两个或多个具备相关性的变量因素进行分析, 从而衡量两个变量因素的相关密切程度。本文为了进一步进行特征筛选, 提高模型精度, 采用相关性分析对属性之间的关系进行分析处理。相关性分析主要通过数据集各个属性两两间的相关系数来进行, 相关系数是统计学中常用的指标, 用于研究变量之间的线性相关程度, 本文使用 Pearson (皮尔逊) 相关系数。公式如下:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (1)$$

其中,  $(X_i, Y_i)$  是样本点,  $\bar{X}$ ,  $\bar{Y}$  是样本均值。我们只关注结果属性列“label”与其他属性之间的相关系数大小, 如表 2 所示。

**Table 2.** Correlation coefficient between attributes**表 2.** 属性间相关系数

	Ffdt001	Ffdt002	Ffdt003	Ffdt004	Ffdt005
label	0.029345	0.032552	0.033062	0.036252	0.362373
	Ffdt006	Ffdt007	Ffdt008	Ffdt009	
label	0.415400	0.005370	0.020427	0.002786	

相关系数的绝对值在 0.3 以下是无直线相关, 0.3 以上是直线相关。由相关系数的具体数值可以看出, 涨跌值和涨跌幅属性列与标签属性列直线相关, 所以将其删除。同时, 我们再删除相关系数最小的属性, 即持仓量变化。因此, 本文保留相关系数在 0.005 到 0.1 之间的属性, 用于作为模型输入因子。

### (三) 归一化处理

由于各个指标的量纲不同, 开盘价、最高价等因子都是价格单位, 而持仓量和成交量都是数量单位, 因子之间数字量级有差别, 消除量纲有助于加速优化过程, 以及可以避免数值出现错误, 所以本文对模型输入因子做归一化处理, 将其数值转化到[0,1]区间上。这里我们采用最常用的线性函数归一化, 同时, 为了纠正误差, 参考了梁元浩[17]的误差处理方法, 取  $\sigma = 0.00001$ 。

归一化公式如下:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} + \sigma \quad (2)$$

## 4. 分类建模

### (一) Xgboost 算法

Xgboost 的全称是 eXtreme Gradient Boosting, 即极端提升树, 是梯度提升机器算法(Gradient Boosting Machine)的扩展版本。Boosting 集成算法可以将多个准确率较低的弱分类器经过加权叠加后形成准确率较高的强分类器, 以达到降低误差、提高精度的目的[18]。eXtreme Gradient Boosting 算法是在 Boosting 算法的基础上进行改进, 通过对生成的每一棵树采用梯度下降的思想, 以上一步生成的所有树为基础, 达到目标函数最小化。同时, Xgboost 算法中对损失函数进行了二阶泰勒展开, 在代价函数里增加了正则项, 用于控制模型复杂度[13]。并且, xgboost 算法可以通过 CUDA 实现 GPU 环境搭建, 相比于 CPU 运行环境, 处理大数据集的性能得到了大幅度提高[19]。除此之外, xgboost 算法在进行完一次迭代后, 会削弱每棵树的影响, 让后面可以有更大的学习空间, 还借鉴随机森林的列抽样, 不仅降低了过拟合, 而且减少了计算量。

Xgboost 算法实现步骤如下:

#### 1) Boosting 模型构建

$$\varnothing(X_i) = \sum_{k=1}^K f_k(X_i), f_k \in F \quad (3)$$

#### 2) 构建目标函数

$$L(\varnothing) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (4)$$

where  $\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2$

#### 3) 训练目标函数



$$L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}^{(t-1)} + f_t(X_i)) + \Omega(f_t) \quad (5)$$

4) 目标函数二阶泰勒展开

$$L^{(t)} \approx \sum_{i=1}^n \left[ l(y_i, \hat{y}^{(t-1)}) + g_i f_t(X_i) + \frac{1}{2} h_i f_t^2(X_i) \right] + \Omega(f_t) \quad (6)$$

where  $g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)})$  and  $h_i = \partial_{\hat{y}^{(t-1)}}^2 l(y_i, \hat{y}^{(t-1)})$

5) 去掉常数项

$$\tilde{L}^{(t)} = \sum_{i=1}^n \left[ g_i f_t(X_i) + \frac{1}{2} h_i f_t^2(X_i) \right] + \Omega(f_t) \quad (7)$$

6) 正则项展开

$$\tilde{L}^{(t)} = \sum_{i=1}^n \left[ g_i f_t(X_i) + \frac{1}{2} h_i f_t^2(X_i) \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \quad (8)$$

进一步转化为:

$$\tilde{L}^{(t)} = \sum_{j=1}^T \left[ \left( \sum_{i \in I_j} g_i \right) \omega_j + \frac{1}{2} \left( \sum_{i \in I_j} h_i + \lambda \right) \omega_j^2 \right] + \gamma T \quad (9)$$

解出使目标函数最小的  $\omega$  为:

$$\omega_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (10)$$

7) 目标函数最优解

$$\tilde{L}^{(t)}(q) = - \frac{1}{2} \sum_{j=1}^T \frac{\left( \sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \lambda} + \lambda T \quad (11)$$

## (二) 其他分类算法

为了验证 xgboost 算法的高效性, 本文还使用决策树、随机森林、支持向量机构建模型作为对比。决策树(DT)算法是机器学习中最基础且应用最广泛的算法模型, 它擅长处理非数值型数据, 模型易于理解, 运行速度也较快[20]。随机森林(RF)是在决策树的基础上改进的算法, 它的算法原理是从样本中随机选取大量样本, 并随机选择分裂属性, 以此构建一定数量的分类树形成随机森林, 最终通过投票决定预测结果[21]。它由于样本选取和特征选取的随机性, 所以能够处理高纬度数据, 并不需要做特征选择, 而且可以减小不平衡数据集的误差。支持向量机(SVM)算法是一种分类算法, 在很多二分类问题上, 都表现出很优秀的预测能力。它通过寻求结构风险最小化来提高模型泛化能力, 适用于解决小样本、非线性和高维模式识别问题[22]。Pedregosa F [23]等介绍了 Scikit-learn 这个模块, 它集成了一系列用于监督学习和无监督学习问题的最先进的机器学习算法, 具有极高的易用性, 以及极小的依赖性, 为我们的机器学习模型构建带来了极大的便利。

## (三) 评价指标

本文使用多个指标来评价模型, 分别是 ROC 曲线下的面积(AUC)、准确率(ACC)、召回率(Recall)、精确率(precision)、F1 值、时间(time)。它们的定义如下:

$$\text{AUC} = \frac{1}{2} \sum_{i=1}^{m-1} (x_{i+1} - x_i) * (y_i + y_{i+1}) \quad (12)$$

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \quad (13)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (14)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (15)$$

$$\text{F1} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (16)$$

$(x_i, y_i)$  表示 ROC 曲线上某个点的坐标, TP 表示被判定为正样本且事实也是正样本的数量, TN 表示被判定为负样本且事实也是负样本的数量, FP 表示被判定为正样本但事实是负样本的数量, FN 表示被判定为负样本但事实是正样本的数量。同时, 对于数据量较大的数据集, 运行时间也是评价指标。本实验均在个人 PC 上完成, 用 python 语言的 time 函数记录运行时间。由于每次代码运行时间有差别, 所以本文取 10 次运行时间的均值。

## 5. 实验

### (一) 模型调参优化

Xgboost 算法的参数分为三类, 分别是一般参数、提升参数、学习参数。提高模型的关键是优化该算法的提升参数。为了避免人为尝试调参带来的疏漏, 本文采用 sklearn.model\_selection 库中内置函数 GridSearchCV 来进行辅助调参, GridSearchCV 用于遍历所有参数组合, 通过交叉验证确定最佳效果参数, 这里采用十折交叉验证。具体的调参步骤如下:

1) 通过 GridSearchCV 遍历设定的学习率取值范围, 得出最佳“学习率”, 然后根据十折交叉验证来确定“最佳树的个数”。

2) 确定好“学习率”和“最佳树个数”后, 调节树的相关参数。首先, 给出 max\_depth 和 min\_child\_weight 参数组合范围, 遍历其所有取值, 得出其最佳参数组合。然后, 给出 colsample\_bytree 和 subsample 参数组合范围, 遍历其所有值, 得出其最佳参数。最后固定以上参数, 再得出“最佳树个数”。

3) 利用以上调节好的参数, 为了进一步避免过拟合, 调节正则化参数, 这里我们调节 L1 正则化参数“alpha”, 同样, 设定范围, 遍历其所有取值, 得出最佳参数。然后, 再得出“最佳树个数”。

4) 利用步骤 3 得出的“最佳树个数”, 以及之前调好的所有参数, 再将学习率变小。得出最佳参数组合。

最佳参数组合, 见表 3。

### (二) 模型结果对比

本文利用 sklearn 库中的决策树、随机森林、支持向量机构建预测模型, 与调参后的 xgboost 算法构建的模型进行对比。四种模型用相同训练数据集训练模型后, 采用相同测试数据集样本对建模结果进行预测, 四种算法的对比结果, 见表 4。

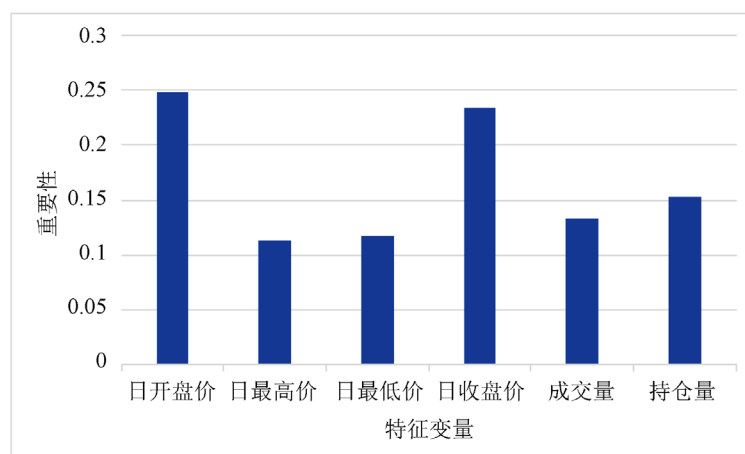
从结果可以看出, 在这四种算法构建的预测模型中, xgboost 的预测结果除了运行速度低于决策树和随机森林之外, 其他各项指标均高于其他算法, 这是因为 xgboost 分类树的数量远远高于这两种算法, 而相比于支持向量机, 则 xgboost 在速度上还是具有较大的优势。其中, xgboost 算法的 AUC 指标远远高于其他三种算法, 说明模型的综合预测能力很高。因此, xgboost 相对于其他算法, 具有准确性高、速度快的优势, 在国际期货预测中是一种较为有效的方法。

**Table 3.** List of best parameters**表 3.** 最佳参数列表

类别	参数名称	参数含义	取值
一般参数	booster	选择提升器	gbtree
提升参数	eta	学习率	0.95
提升参数	gamma	控制叶子个数	0.1
提升参数	Max_depth	每棵树最大深度	5
提升参数	Min_child_weight	每个叶子的最小权重和	1
提升参数	subsample	样本采样比率	0.86
提升参数	colsample_bytree	列采样比率	0.84
提升参数	alpha	L1 正则化参数	4
学习参数	objective	损失函数	binary:logistic
学习参数	eval_metric	评估指标	error
	num_boost_round	最佳树个数	396

**Table 4.** Comparison of modeling results**表 4.** 建模结果对比

模型名称	AUC	ACC	Recall	Precision	F1	time
xgboost	0.9111	0.8240	0.8293	0.8263	0.8278	约 21.98s
决策树	0.8044	0.8046	0.8169	0.8032	0.8100	约 4.434s
随机森林	0.8167	0.8166	0.8129	0.8250	0.8189	约 5.772s
SVM	0.5262	0.5221	0.3207	0.5542	0.4063	约 317.259s

**Figure 1.** Characteristic variable importance score**图 1.** 特征变量重要性得分

### (三) 特征重要性分析

通过 xgboost 算法建模可以判断每个模型因子对模型的贡献程度,从而判断哪些特征变量对于国际期货涨跌趋势的影响更为显著。

如图 1 所示,日开盘价和日收盘价对模型的贡献最大,对国际期货价格变动的影响较为显著。而同为价格变量的日最高价和日最低价对模型的贡献最小,这是因为这两个变量反映了日交易价格的极端情



况, 变动比较大。持仓量和成交量对模型的程度基本相同。

## 6. 结语

本文采用 `xgboost` 分类算法, 基于国际期货三年的日交易数据进行了特征提取和分类建模, 并与决策树、随机森林、支持向量机算法进行了对比, 得到了理想的预测结果。同时, 通过特征重要性分析, 我们获得了对国际期货涨跌影响较大的变量。可以看出, 基于 `xgboost` 算法构建的模型对国际期货涨跌趋势进行预测具有明显的优点, `xgboost` 算法是基于决策树的集成算法的改进版本, 理论上结合了决策树和集成算法的优点。该算法具有高度的灵活性, 允许自定义优化目标和评价标准, 并且参数较多, 可调整的范围大。同时, 利用自动化调参工具遍历所有参数组合, 对于模型调参带来了极大的便利。此外, 此论文的研究还将进一步深入, 比如优化改进 `xgboost` 算法, 以及收集更多地影响期货价格涨跌的因子, 从而获得精度更好的期货预测模型。

## 基金项目

课题名称: 北京市社会科学基金项目——多社会网络中的北京市科技创新引领要素研究, 序列号: 16GLC067。

## 参考文献

- [1] 李绍. 运用概率论和数理统计理论研究股市的内在规律[J]. 统计与决策, 1999(2): 16-17.
- [2] 魏宇. 沪深 300 股指期货的波动率预测模型研究[J]. 管理科学学报, 2010, 13(2): 66-76.
- [3] Wang, C.Y. (2016) Forecast on Price of Agricultural Futures in China Based on ARIMA Model. *Asian Agricultural Research*, 8, 9-12, 16.
- [4] 董贺. 基于国内商品期货的量化投资优化策略[D]: [硕士学位论文]. 杭州: 浙江大学, 2017: 1-31.
- [5] 刘海玥, 白艳萍. 时间序列模型和神经网络模型在股票预测中的分析[J]. 数学的实践与认识, 2011, 41(4): 14-19.
- [6] Wang, W. (2017) A Big Data Framework for Stock Price Forecasting Using Fuzzy Time Series. *Multimedia Tools & Applications*, 77, 10123-10134.
- [7] 梁楠. 人工智能交易“战队”悄然集结[N]. 期货日报, 2017-03-08(2).
- [8] Hsu, S.H., Hsieh, P.A., Chih, T.C. and Hsu, K.C. (2009) A Two-Stage Architecture for Stock Price Forecasting by Integrating Self-Organizing Map and Support Vector Regression. *Expert Systems with Applications*, 36, 7947-7951. <https://doi.org/10.1016/j.eswa.2008.10.065>
- [9] 赵永进. 基于数据挖掘的股票分析与预测研究[D]: [硕士学位论文]. 郑州: 郑州大学, 2005: 1-52.
- [10] Qian, B. and Rasheed, K. (2007) Stock Market Prediction with Multiple Classifiers. *Applied Intelligence*, 26, 25-33. <https://doi.org/10.1007/s10489-006-0001-7>
- [11] 王刚, 许晓兵. 基于小波分析与神经网络时间序列的股票预测方法[J]. 金融经济, 2013(12): 161-163.
- [12] Karathanasopoulos, A., Sermpinis, G., Dunis, C., et al. (2016) Stock Market Prediction Using Evolutionary Support Vector Machines: An Application to the ASE20 Index. *European Journal of Finance*, 22, 1145-1163. <https://doi.org/10.1080/1351847X.2015.1040167>
- [13] Chen, T.Q. and Guestrin, C. (2016) XGBoost: A Scalable Tree Boosting System. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, 13-17 August 2016, 785-794. <https://doi.org/10.1145/2939672.2939785>
- [14] 张昊, 纪宏超, 张红宇. XGBoost 算法在电子商务商品推荐中的应用[J]. 物联网技术, 2017, 7(2): 102-104.
- [15] 蒋晋文, 刘伟光. XGBoost 算法在制造业质量预测中的应用[J]. 智能计算机与应用, 2017, 7(6): 58-60.
- [16] 李想. 基于 XGBoost 算法的多因子量化选股方案策划[D]: [硕士学位论文]. 上海: 上海师范大学, 2017: 1-90.
- [17] 梁元浩. 基于支持向量机的股指期货超短期预测研究[J]. 信息与电脑(理论版), 2015(5): 82-83.
- [18] Taieb, S.B. and Hyndman, R.J. (2014) A Gradient Boosting Approach to the Kaggle Load Forecasting Competition. *International Journal of Forecasting*, 30, 382-394. <https://doi.org/10.1016/j.ijforecast.2013.07.005>

- [19] Mitchell, R. and Frank, E. (2017) Accelerating the XGBoost Algorithm Using GPU Computing. *PeerJ Computer Science*, **3**, e127. <https://doi.org/10.7717/peerj-cs.127>
- [20] 唐华松, 姚耀文. 数据挖掘中决策树算法的探讨[J]. 计算机应用研究, 2001, 18(8): 18-19.
- [21] 罗知林, 陈挺, 蔡皖东. 一个基于随机森林的微博转发预测算法[J]. 计算机科学, 2014, 41(4): 62-64.
- [22] 丁世飞, 齐丙娟, 谭红艳. 支持向量机理论与算法研究综述[J]. 电子科技大学学报, 2011, 40(1): 2-10.
- [23] Pedregosa, F., Gramfort, A., Michel, V., et al. (2016) Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research*, **12**, 2825-2830.

#### 知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>  
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-0967, 即可查询
2. 打开知网首页 <http://cnki.net/>  
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: [fin@hanspub.org](mailto:fin@hanspub.org)