IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 14, 2022, accepted May 12, 2022, date of publication May 17, 2022, date of current version June 6, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3175842

Real-Trading-Oriented Price Prediction With Explainable Multiobjective Optimization in Quantitative Trading

TAO YIN[1], XINGBO DU[1], WEIPENG ZHANG[1], YUNAN ZHAO[2], BING HAN[2], AND JUNCHI YAN[1], (Senior Member, IEEE)

[1]School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[2]MYbank, Ant Group, Hangzhou 310063, China

Corresponding author: Junchi Yan (yanjunchi@sjtu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018AAA0100704; in part by the Shanghai Municipal Science and Technology Major Project (2021SHZDZX0102); and in part by MYbank, Ant Group.

ABSTRACT Price prediction using machine learning is a heated topic and an effective task in quantitative trading. However, the objectives that existing researches focus on are typically naive and can hardly express the diverse properties in real trading, by using which the results of price prediction cannot precisely reflect the returns in real trading. To alleviate this problem, we first formulate the characteristics of transactions and then propose a multi-objective method under a real-trading-oriented perspective, where a tree model inducing explainable trading characteristics is proposed. With lots of considerations of real trading characteristics, the results of price prediction can better obtain the properties in trading, and especially perform quite well in real trading. Multiple experiments on the Chinese Index Future Market display the effectiveness of the proposed model, where the performance of real trading has reached the industry-leading level. In particular, the difference between model prediction and returns in real trading can be diminished.

INDEX TERMS Machine learning, price prediction, quantitative trading, real trading, the Chinese index future market.

I. INTRODUCTION

Price prediction has been a remarkably hot issue in quantitative trading [1], [2], while recently tends to be more often applied using machine learning [3], [4]. Since regarded as a prediction task, price prediction typically aims to minimize the difference between true price and predicted one. The difference can be described as an evaluation metric, e.g., Mean Square Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), etc. With the reduction of the given evaluation metric's value, the true and predicted price can be approximated, and thus an investor is supposed to buy a stock that would gain and then obtain returns. Under this perspective, existing researchers mainly pay attention to devising a prediction model, no matter using traditional time series forecasting methods, such as ARIMA [1], or neural networks [3], to promote the accuracy of the prediction.

However, we argue that the reduction of the evaluation metric's value in price prediction doesn't directly achieve the return in real trading.

For instance, in quantitative trading, not each prediction needs to be traded, and we generally do not perform any trading when given most predictions with low values, so it is even not important to calculate the accuracy and expectation for these predictions.

For relatively large or relatively small predicted values, although tradings in the same direction, i.e. buying when gaining or selling when falling, are usually carried out according to the forecast value, we are more concerned about the accuracy of the forecast value direction because whether the actual price matches the forecast value has little impact on the results, due to the order filled probability, slippage, etc., as illustrated in this paper.

Therefore, common evaluation metrics, e.g. MSE and MAE, do not match the profit and loss of the actual trading. Considering to simulate the actual return, we can divide the market into several cases: 1) violently positive fluctuations;

The associate editor coordinating the review of this manuscript and approving it for publication was Omid Kavehei.

2) violently negative fluctuations; 3) ordinarily positive fluctuations; 4) ordinarily negative fluctuations; 5) constant or minute fluctuations. The diagram that illustrates the above cases is shown in Fig. 1. Then we sort the classes according to the predictions.

To solve the mismatch of model predictions and returns in real trading, we propose a multi-objective tree model, whereby inducing the diverse characteristics of transactions.

The main contributions are as follows:

- To our best knowledge to existing literature, we are the first to reveal the insufficiency of traditional evaluation metrics when executing price predictions using machine learning in Chinese quantitative trading market, where the accuracy of price predictions can hardly match the one in real trading.
- We formulate the characteristics of transactions, based on which we propose a multi-objective tree model inducing explainable trading characteristics to enable the model prediction to match the return in real trading.
- Multiple experiments on real-world trading data are performed to show the industry-leading effectiveness of the devised model, whereby we particularly find that the difference between model prediction and returns in real trading can be diminished under the proposed model.

In this paper, we first introduce some preliminaries and define the problem formulation in Sec. II to clarify the background of the proposed method. Besides, related works are introduced in Sec. III, including quantitative trading and price prediction with machine learning, to show the current studies and the main challenge. Then, we show the proposed method in Sec. IV and display the experimental results in Sec. V. Finally, we conclude the whole paper and suggest some future works in Sec. VI.

II. PRELIMINARIES AND PROBLEM FORMULATION

In this section, we first introduce the problem formulations, including price prediction and order filled probability prediction, to depict the problems we should solve. Then, some common factors in financial trading are illustrated to enable them to be more understandable.

A. PROBLEM FORMULATION

Price prediction defined in Def. 2.1 is a common task in financial trading, especially in quantitative trading. For a specific subject matter, such as a stock, if we can successfully predict its price, we will be able to buy or sell it in advance to obtain returns or avoid losses.

Definition 2.1 (Price Prediction): Given a subject matter s, we define its percent change of price as $s_{price}^{(T)}$ at time T. We further define the attributes of s as $s^{(T)} = (s_1^{(T)}, s_2^{(T)}, \ldots)$ (historical price, volume, external message, etc.) at time T, as well as the future moments t that need to be predicted at. Price prediction is a mapping $f : (S, time) \to \mathbb{R}$, satisfying $s_{price}^{(T+t)} = f(s^{(T)}, t)$, where S and $time$ are sets of subject matters and time points respectively.

Actually, there is an implicit assumption in price prediction, that is, the trading probability is 1 and we can complete the trading as long as we want. However, in real trading, it is of much difficulty to trade at the current price, since the price always fluctuates under a large number of tradings, thus the price we predicted to gain cannot stand for the real returns. Therefore, we exploit trading probability prediction in Def. 2.2 to consider the above mismatch.

Definition 2.2 (Order Filled Probability Prediction): Given a subject matter s and its attributes $s^{(T)} = (s_1^{(T)}, s_2^{(T)}, \ldots)$ at time T, as well as the future moments t that need to be predicted at, order filled probability prediction is a mapping $g : (S, time) \to [0, 1]$. Higher $g(s^{(T)}, t)$ stands for a higher probability to complete the trading.

In particular, the future time point t is typically fixed in price prediction and order filled probability prediction, since we do not care about the impact of the change of t in this paper. Besides, in the real trading, even if t changes, the order filled probability actually only depends on the cancellation time, not the prediction time point t. Therefore, we just use $g(s^{(T)})$ and $f(s^{(T)})$ instead of $g(s^{(T)}, t)$ and $f(s^{(T)}, t)$ respectively.

B. TERMINOLOGIES

In this section, we will give the definitions of some financial terminologies, especially those who play a decisive role.

First, we define *LastPrice*, which means the latest transaction price of tick data, and then we define *BidPrice*1 and *AskPrice*1 as the best bid price and the best ask price respectively. In short, *BidPrice*1 is the highest bid price while *AskPrice*1 is the lowest ask price.

However, in high-frequency trading, *LastPrice* is not as accurate, since the data often has a huge spread. For example, in the Chinese Index Future Market, the contract is about 7000 points, and there is often a spread of about one point. If a single transaction earns 0.7, the spread is 1/10000. Considering the margin ratio of 14%, the return is about 7/10000. The comprehensive annualized rate of return is 7/10000 $\times$ 2 $\times$ trading days, which is roughly calculated as 35%, and is a very considerable returns. Therefore, spread has a relatively great impact and we need to further define *FairPrice*, which means the actual transaction price if you intend to trade, and should be as close as possible to the real price of the transactions in the future.

III. RELATED WORK

A. QUANTITATIVE TRADING

Quantitative trading has been widely studied. For example, work [5] based on a Markov renewal model solve an optimal high-frequency trading problem to balance the accuracy and tractability. Work [6] employs algorithmic trading for index prediction, while work [7] utilizes deep learning with a stack of Restricted Boltzmann Machines (RBMs) [8] to achieve portfolio management to reduce risk [9]. Furthermore, works [10], [11] use deep learning

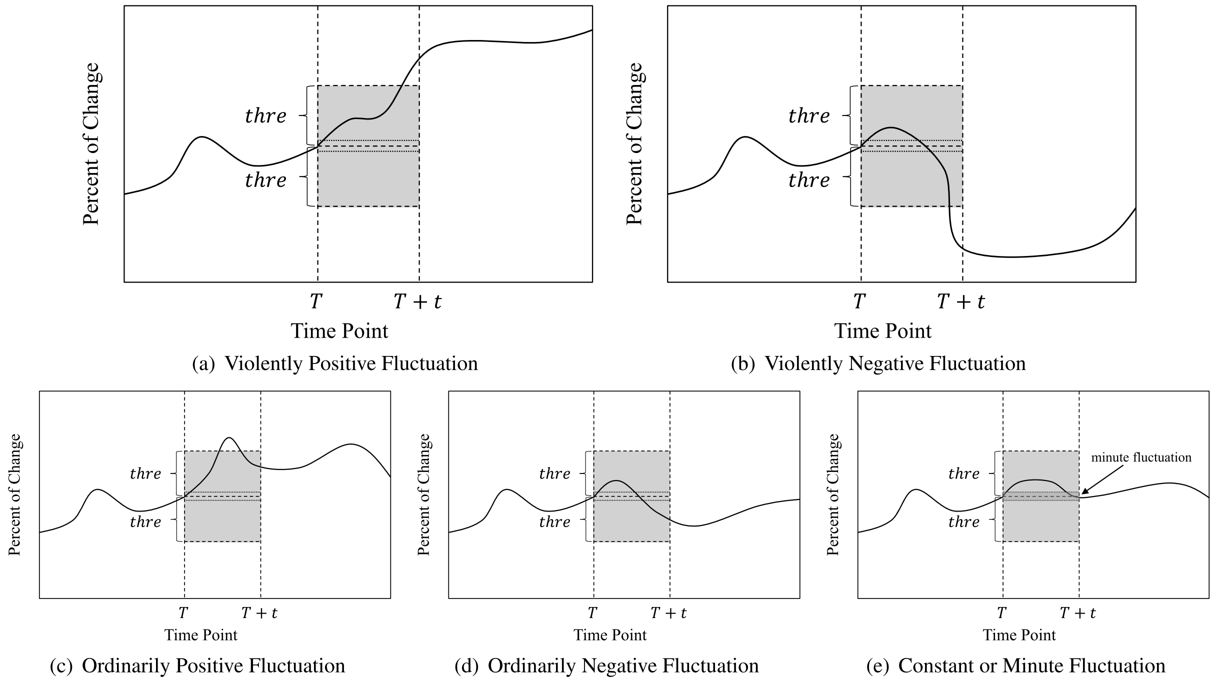


FIGURE 1. Different Cases of fluctuations, including: a) violently positive fluctuations; b) violently negative fluctuations; c) ordinarily positive fluctuations; d) ordinarily negative fluctuations; e) constant or minute fluctuations. The cases are mainly affected by the threshold *thre* (the dark regions) that decides whether the price changes tremendously. In (a), when the predicted percent of price's change is over *thre* at time $T + t$, we define it as a violently positive fluctuation. In (b), when the predicted percent of price's change is below $-$*thre* at time $T + t$, we define it as a violently negative fluctuation. Cases (a) and (b) show the situations that the price fluctuates greatly, where we will bid or ask a price actively, and as a result, it is more likely for us to make a successful transaction, while a slippage has to exist here. In (c), when the predicted percent of price's change is below *thre* and positive at time $T + t$, we define it as an ordinarily positive fluctuation. In (d), when the predicted percent of price's change is over $-$*thre* and below 0 at time $T + t$, we define it as an ordinarily negative fluctuation. Cases (c) and (d) show the situations that the price fluctuates ordinarily, where we will bid or ask a price passively, and as a result, there exists a probability of failing transaction. In (e), when the predicted percent of price's change is not notable, we define it as a minute fluctuation. In particular, we define it as a constant when the predicted percent of price's change is equal to 0. Since the fluctuation is not obvious and few returns can be obtained in this case, we tend not to trade.

methods to perform asset pricing and probability statistics. In particular, work [12] constructs $J - K$ portfolios proposed by [13] to show the inefficiency of the Chinese Stock Market in the weak form.

To sum up, quantitative trading contains multiple aspects, including price prediction, trading strategies [5], algorithmic trading [6], portfolio management [7], asset pricing [10], [11], etc. In this paper, we mainly care about price prediction, whose related work will be introduced in the next part.

B. PRICE PREDICTION WITH MACHINE LEARNING

Price prediction plays an important role in trading and multiple machine learning techniques have been employed to perform price prediction tasks. Of all the techniques, time series models [1], [14] are the most basic and important ones, due to the fact that prices with respect to time are essentially time series. At an early stage, work [15] utilized statistical methods to predict the Korean Stock Price Index 200 (KOSPI200), and compared the returns among some traditional machine learning techniques, including linear regression, logistic regression, neural networks, and support vector machine. With the development of deep learning [8], some trials [16], [17] are proposed to achieve better returns.

Price prediction methods are developing rapidly with more returns and efficiency. However, current studies still face challenges that they don't fully consider the circumstances faced with in real trading, including the order filled probability, the coordination of the trading strategy, etc., and thus can hardly match the returns in real trading, even though they perform well in accuracy of the model prediction. To overcome the challenge, we propose a multi-objective tree model inducing multiple characteristics to match the prediction and returns in real trading.

IV. THE PROPOSED METHOD

We propose a tree model, where we present multiple objectives to induce explainable trading characteristics, as such our price prediction method is devised with trading probabilities to diminish the difference of model prediction and returns in real trading.

A. TRANSACTION STRATEGY

As mentioned before in Fig. 1, we divide the market into several cases to simulate the actual return, due to the fact that we cannot always complete the transaction at the latest price.

For simplicity, let $f(s^{(T)}) \triangleq p$ and $s_{price}^{(T+t)} \triangleq l$ be the predicted and true percent of price change respectively, and we define $thre > 0$ as the threshold to decide whether the price changes tremendously. Namely, when p is larger[1] than $thre$, the price is considered to be a violent fluctuation.

To alleviate the negative impact brought by commission, capital utilization, and the inconsistency of current price and transaction price, it is a natural idea that we only buy or sell actively under a violent fluctuation. To enable the prediction to be more statistically significant, $T + 1$ trading market like the Chinese A-share Market is not proper due to its short frequency, thus we choose stock index futures, which is a $T + 0$ trading market, to validate our strategy, where we typically use 'bid' and 'ask' instead of 'buy' and 'sell' respectively. Thus, we formulate our strategy as follows:

- When $|p| > |thre|$, the strategy will open the position and bid/ask a price aggressively.
- Otherwise, the strategy will only bid/ask a price passively.

This strategy will lead to some characteristics:

1) The order filled probability will be very high when actively opening a position. In particular, if the transaction speed is high enough, the order filled probability will be close to 1. We hope to employ a projection to enable the order filled probability to be close to 1 when actively opening a position, to reduce the uncertainty caused by computational performance.
2) There will be a slippage, i.e. the deviation between $BidPrice1$ ($AskPrice1$) and $FairPrice$, when actively opening a position. On the contrary, when passively bidding or asking a price, the slippage will transform to returns, because we will bid or ask a price between $BidPrice1$ and $AskPrice1$. For example, when passively bidding a price, $BidPrice1$ is lower than $FairPrice$, and we will obtain a return of about $FairPrice - BidPrice1$. This characteristic present a novel perspective that a higher change of price does not mean more returns in real trading, while the approximation of predicted and true price does not decide the final returns. Conversely, current price prediction methods [18], [19] regard a higher change as more returns, which is not in line with the real trading.
3) There is an equivocal order filled probability when passively bidding or asking a price, however, there will be $g(s^{(T)}, t)$, i.e. the order filled probability will be 1, if the predicted direction is contrary to the real direction. For example, we predict the price will gain and bid a high price, however, if the true price falls, we will complete to bid a price, and vice versa. This characteristic shows a trading situation with the order filled probability of 1 and remind us that it is of importance to have a high accuracy of prediction when p is smaller than $thre$ (the order filled probability is not necessarily 1 if predicted wrongly, while the one is not 1 if predicted correctly).

Based on the above characteristics, it is obvious that traditional price prediction methods cannot obtain more returns due to the slippage and uncertain order filled probability, even though they have a perfect approximation between model prediction and real price. To solve this problem, we present our multi-objective models.

[1] We use 'larger' or 'smaller' to indicate the absolute value in this paper.

B. MULTIPLE OBJECTIVES

Except for the extraordinary perspective, through the analysis in Sec. IV-A, we can also summarize returns into three cases, including: 1) Return expectations of bidding at a higher price when the predicted price is close to or larger than $thre$; 2) Return expectations of asking at a lower price when the predicted price is close to or larger than $thre$; 3) Return expectations of quoting.

As is mentioned in the second characteristic in Sec. IV-A, we simply define the slippage as c, which is consistent to the difference between $FairPrice$ and $BidPrice1$ ($AskPrice1$).

1) RETURN EXPECTATIONS OF BIDDING AT A HIGHER PRICE WHEN THE PREDICTED PRICE IS CLOSE TO OR LARGER THAN *THRE*

When bidding at a higher price and the predicted price is close to or larger than $thre$, the true percent of price change l is close to the real return, except for the slippage, thus we use $l - c$ to estimate the real return.

On the other hand, the probability of this situation should be considered. Intuitively, when p is higher, it is more likely to fall in this situation. Formally, when predicting the price is close to or larger than $thre$, we define the probability of bidding at a higher price as

$$\begin{aligned}\mathbf{adp}(p) &= \frac{1}{1 + \exp[-\alpha(p - thre)]} \\ &= \sigma(\alpha(p - thre)),\end{aligned} \tag{1}$$

where α is a hyperparameter to control the distribution, and $\sigma(\cdot) = 1/[1+\exp(\cdot)]$ is a common activation function, whose gradient can be easily calculated by

$$\frac{\partial \sigma(x)}{\partial x} = \sigma(x) \cdot [1 - \sigma(x)]. \tag{2}$$

In Eq. 1, $\mathbf{adp}(p)$ is monotonically increasing with respect to $p - thre$ and tends to 1 with the increase of $p - thre$, while $\mathbf{adp}(p)$ tends to 0 with the reduction of $p - thre$.

Then, when the predicted price is close to or larger than $thre$, the first objective that we maximize the return expectations of bidding at a higher price is equal to minimizing its opposite one, where the objective to be minimized can be formulated as

$$\mathcal{L}_{bid} = -\mathbf{adp}(p) \times (l - c). \tag{3}$$

2) RETURN EXPECTATIONS OF ASKING AT A LOWER PRICE WHEN THE PREDICTED PRICE IS CLOSE TO OR LARGER THAN *THRE*

Since we only ask the price when we predict the price will fall, l is definitely less than 0. Similar to the above, when asking at a lower price and the predicted price is close to or larger than *thre*, the true percent of price change $-l$ is close to the real return. In addition, there is still a slippage when asking the price actively, thus we use $-l-c$ to estimate the real return.

On the other hand, the probability should still be considered and we hope that the probability should be higher when $-p$ is higher. Formally, when predicting the price is close to or larger than *thre*, we define the probability of asking at a lower price as

$$\begin{aligned}\mathbf{adm}(p) &= \frac{1}{1+\exp\left[-\alpha(-p-thre)\right]} \\ &= \sigma(\alpha(-p-thre)), \end{aligned} \tag{4}$$

where $\alpha > 1$ is the same as the one in Eq. 1.

Similarly, in Eq. 4, $\mathbf{adm}(p)$ is monotonically increasing with respect to $-p-thre$ and tends to 1 with the increase of $-p-thre$, while $\mathbf{adm}(p)$ tends to 0 with the reduction of $-p-thre$.

Then, when the predicted price is close to or larger than *thre*, the second objective that we maximize the return expectations of asking at a lower price is equal to minimizing its opposite one, where the objective to be minimized can be formulated as

$$\mathcal{L}_{ask} = -\mathbf{adm}(p) \times (-l-c). \tag{5}$$

3) RETURN EXPECTATIONS OF QUOTING

The above two show the return expectations when actively bidding or asking a price. In this part, we will depict the return expectation of quoting when passively bidding or asking a price.

In passive trading, we can obtain a return when bidding or asking a price no matter p is positive or negative. Also, considering the slippage, the return should be $|p|+c$.

Since $|p|+c$ is not differentiable, we employ:

$$\begin{aligned}\mathbf{ad}(p) &= \frac{2}{1+\exp(-\alpha p)} - 1 \\ &= 2\sigma(\alpha p) - 1, \end{aligned} \tag{6}$$

to perform the relaxation. In this way, it can be guaranteed that Eq. 6 is monotonically increasing and we can ensure that

$$\begin{cases} 0 < \mathbf{ad}(p) < 1 & \text{if } p > 0 \\ -1 < \mathbf{ad}(p) < 0 & \text{if } p < 0 \\ \mathbf{ad}(p) = 0 & \text{if } p = 0. \end{cases} \tag{7}$$

Thus, if traded successfully, the real return of bidding or asking a price can be formulated as

$$\mathbf{Ex}(p) = \mathbf{ad}(p) \cdot l + c. \tag{8}$$

As is mentioned in the third characteristic in Sec. IV-A, when predicting wrongly, the order filled probability is 1. In addition, when the predicted price is close to or larger than *thre*, we hope the order filled probability to be decreasing with the increase of p. To be specific, we define the order filled probability of bidding or asking a price passively as

$$\begin{aligned} g(s^{(T)}) &= \mathbf{adE}(p) \cdot \left[1 - \frac{r}{\exp\left(\frac{\mathbf{ad}(p)\cdot p}{thre}\right)}\right] + \frac{r}{\exp\left(\frac{\mathbf{ad}(p)\cdot p}{thre}\right)} \\ &= \mathbf{adE}(p) + \frac{r\,(1-\mathbf{adE}(p))}{\exp\left(\frac{\mathbf{ad}(p)\cdot p}{thre}\right)}, \end{aligned} \tag{9}$$

where r is a basic order filled probability to control the minimum probability of transaction, and $\mathbf{adE}(p, l)$ is a coefficient to adjust the order filled probability:

$$\begin{aligned}\mathbf{adE}(p) &= \frac{1}{1+\exp(\alpha^2 pl)} \\ &= \sigma(-\alpha^2 pl). \end{aligned} \tag{10}$$

In Eq. 10, $pl > 0$ when the predicted price is close to or larger than *thre*, and $pl < 0$ when predicting wrongly. In order to let $\mathbf{adE}(p, l)$ tend to be 0 when the predicted price is close to or larger than *thre*, we use α^2 ($\alpha > 1$) to adjust the coefficient. When predicting wrongly, the coefficient will decay with the increase of $|p|$, which is a specify of the third characteristic in Sec. IV-A.

Thus, the final objective that we maximize the return expectations of quoting is equal to minimizing its opposite one, where the objective to be minimized is formulated as

$$\mathcal{L}_{pend} = -\mathbf{Ex}(p) \cdot g(s^{(T)}) \cdot (1 - \mathbf{adp}(p) - \mathbf{adm}(p)), \tag{11}$$

where $1-\mathbf{adp}(p)-\mathbf{adm}(p)$ is an estimation of the probability of quoting, except for the probability of bidding or asking a price actively. Also, it can be proved in Theorem 4.1 that $0 < \mathbf{adp}(p) + \mathbf{adm}(p) < 1$, as such satisfying the range of probability.

Theorem 4.1: Given $\alpha > 0$ and $thre > 0$, we have the following constraints on the two added terms:

$$0 < \mathbf{adp}(p) + \mathbf{adm}(p) < 1. \tag{12}$$

Proof: In this theorem, we first prove $\mathbf{adp}(p) + \mathbf{adm}(p) > 0$, and then prove $\mathbf{adp}(p) + \mathbf{adm}(p) < 1$, as such illustrating that $1 - \mathbf{adp}(p) - \mathbf{adm}(p)$ in Eq. 11 fits the range of probability.

On the one hand, it is obvious that $\mathbf{adp}(p) > 0$ and $\mathbf{adm}(p) > 0$, thus $\mathbf{adp}(p) + \mathbf{adm}(p) > 0$ can be easily obtained. On the other hand,

$$\begin{aligned} &\mathbf{adp}(p) + \mathbf{adm}(p) \\ &= \frac{1}{1+\mathrm{e}^{-\alpha p+\alpha\cdot thre}} + \frac{1}{1+\mathrm{e}^{\alpha p+\alpha\cdot thre}} \\ &= \frac{\left(1+\mathrm{e}^{\alpha p+\alpha\cdot thre}\right) + \left(1+\mathrm{e}^{-\alpha p+\alpha\cdot thre}\right)}{\left(1+\mathrm{e}^{-\alpha p+\alpha\cdot thre}\right)\cdot\left(1+\mathrm{e}^{\alpha p+\alpha\cdot thre}\right)} \\ &= \frac{2+\mathrm{e}^{\alpha p+\alpha\cdot thre}+\mathrm{e}^{-\alpha p+\alpha\cdot thre}}{1+\mathrm{e}^{2\alpha\cdot thre}+\mathrm{e}^{\alpha p+\alpha\cdot thre}+\mathrm{e}^{-\alpha p+\alpha\cdot thre}}. \end{aligned} \tag{13}$$

Since $\alpha > 0$ and $thre > 0$, then $\mathrm{e}^{2\alpha \cdot thre} > 1$, hence $2 < 1 + \mathrm{e}^{2\alpha \cdot thre}$. Therefore, we have

$$\begin{aligned} 0 &< 2 + \mathrm{e}^{\alpha p+\alpha \cdot thre} + \mathrm{e}^{-\alpha p+\alpha \cdot thre} \\ &< 1 + \mathrm{e}^{2\alpha \cdot thre} + \mathrm{e}^{\alpha p+\alpha \cdot thre} + \mathrm{e}^{-\alpha p+\alpha \cdot thre}. \end{aligned} \tag{14}$$

Hence, Eq. 13 is lower than 1, thus Eq. 12 is proved. ■

Remark 4.1: In the above proof, we mainly prove that $1-\mathbf{adp}(p)-\mathbf{adm}(p)$ follows the range of probability. Not only that, we ensure that $\mathbf{adp}(p)$, $\mathbf{adm}(p)$, and $1-\mathbf{adp}(p)-\mathbf{adm}(p)$ in Eq. 3, 5, 11 are all between 0 and 1 to satisfy the range of probability. These constraints depict the different impacts of the return expectations of bidding at a higher price or asking at a lower price when the predicted price is close to or larger than *thre*, and the return expectations of quoting. This setting is in line with the real trading.

In a nutshell, the comprehensive objectives are given by:

$$\mathcal{O} = \mathcal{L}_{bid} + \mathcal{L}_{ask} + \mathcal{L}_{pend}. \tag{15}$$

However, we find that $\mathcal{O}$ in Eq. 15 varies little when p is very high. Furthermore, when l and p are very large, the loss will further decrease as p continues to rise. This phenomenon leads to overfitting, which is not in line with the actual situation and not good for position control. To solve this, we add a penalty term to control p and avoid falling into local optimization in the initial stage of training. Formally,

$$\mathcal{L}_p = \beta(l-p)^2, \tag{16}$$

where β is a penalty coefficient to control the impact of the penalty term. Accordingly, Eq. 15 is extended by:

$$\mathcal{O} = \mathcal{L}_{bid} + \mathcal{L}_{ask} + \mathcal{L}_{pend} + \mathcal{L}_p. \tag{17}$$

Note here we essentially assign different weights to the terms for $\mathcal{L}_{bid}$, $\mathcal{L}_{ask}$, and $\mathcal{L}_{ask}$, as illustrated in Remark 4.1, which is in line with the returns in real trading. In addition, $\mathcal{L}_p$ is actually decided by β in Eq. 16.

We regard different components in Eq. 17 together constitute a whole trading skeleton, which is the essence of 'multi-objective'. For this reason, it is meaningless to measure one of the objectives alone.

C. ILLUSTRATION OF EXPLAINABILITY

We regard the core objective to optimize is 'explainable' in Eq. 15, because $-L_{bid}$, $-L_{ask}$, $-L_{pend}$ respectively represent the return expectations of bidding at a higher price when the predicted price is close to or larger than *thre*, asking at a lower price when the predicted price is close to or larger than *thre*, and quoting, which corresponds to the real trading atmosphere and illustrate the ways to obtain returns instead of getting a better approximation between the true price and the predicted one.

We affirm that this perspective is important, because it reveals that higher accuracy of the model is not equivalent to more returns in quantitative trading. We provide a new insight of matching the real returns and model prediction under this 'real-trading-oriented' perspective.

D. OPTIMIZING WITH LIGHTGBM

In short period predictions, especially for high-frequency trading, we prefer tree models to deep neural networks, due to tree models' fast deployment and invocation in real trading.

In particular, there is a large amount of quantitative data and the market changes rapidly, so it is difficult to judge whether the distribution of market data has changed. Accordingly, we devise a LightGBM-based tree model, in common with [20], [21], to keep a balance between efficiency and reducing losses. Specifically, Light Gradient Boosting Machine (LightGBM) [22] is a state-of-the-art framework of gradient-boosting with the capability of fast training, high effectiveness, and low memory usage.

These characteristics are suitable for price prediction tasks. On the one hand, trading data has high randomness. For example, small differences in features may be caused by many factors, including our definition of price, the processing of time slice, etc., which do not affect the features of the data themselves. To a certain extent, histogram [22] in LightGBM has little loss for trading data.

The gradients of Eq. 17 are computed by

$$\begin{aligned} \frac{\partial \mathcal{L}_{bid}}{\partial p} &= (c-l) \cdot \frac{\partial \mathbf{adp}(p)}{\partial p}, \\ \frac{\partial \mathcal{L}_{ask}}{\partial p} &= (c+l) \cdot \frac{\partial \mathbf{adm}(p)}{\partial p}, \\ \frac{\partial \mathcal{L}_{pend}}{\partial p} &= \frac{\partial \mathbf{Ex}(p)}{\partial p} \cdot g(s^{(T)}) \cdot (\mathbf{adp}(p) + \mathbf{adm}(p) - 1) \\ &\quad + \mathbf{Ex}(p) \cdot \frac{\partial g(s^{(T)})}{p} \cdot (\mathbf{adp}(p) + \mathbf{adm}(p) - 1) \\ &\quad + \mathbf{Ex}(p) \cdot g(s^{(T)}) \cdot \left(\frac{\partial \mathbf{adp}(p)}{\partial p} + \frac{\partial \mathbf{adm}(p)}{\partial p} \right), \\ \frac{\partial \mathcal{L}_p}{\partial p} &= 2\beta(p-l), \end{aligned} \tag{18}$$

where the gradients in Eq. 1, 4, 8, 9 are calculated by

$$\begin{aligned} \frac{\partial \mathbf{adp}(p)}{\partial p} &= \alpha \cdot \sigma(\alpha(p-thre)) \cdot [1-\sigma(\alpha(p-thre))], \\ \frac{\partial \mathbf{adm}(p)}{\partial p} &= -\alpha \cdot \sigma(\alpha(-p-thre)) \cdot [1-\sigma(\alpha(-p-thre))], \\ \frac{\partial \mathbf{Ex}(p)}{\partial p} &= 2\alpha l \cdot \sigma(\alpha p)\,[1-\sigma(\alpha p)], \\ \frac{\partial g(s^{(T)})}{\partial p} &= r \cdot \frac{(\mathbf{adE}(p)-1) \cdot \frac{\partial \mathrm{e}^{\frac{\mathbf{ad}(p)\cdot p}{thre}}}{\partial p} - \frac{\partial \mathbf{adE}(p)}{\partial p} \cdot \mathrm{e}^{\frac{\mathbf{ad}(p)\cdot p}{thre}}}{\mathrm{e}^{\frac{2\mathbf{ad}(p)\cdot p}{thre}}} \\ &\quad + \frac{\partial \mathbf{adE}(p)}{\partial p}, \\ \frac{\partial \mathbf{adE}(p)}{\partial p} &= -\alpha^2 l \cdot \sigma(-\alpha^2 pl) \cdot \left[1-\sigma(-\alpha^2 pl)\right], \\ \frac{\partial \mathrm{e}^{\frac{\mathbf{ad}(p)\cdot p}{thre}}}{\partial p} &= \frac{\mathrm{e}^{\frac{\mathbf{ad}(p)\cdot p}{thre}}}{thre} \cdot [2\alpha p \cdot \sigma(\alpha p)\,(1-\sigma(\alpha p)) + \mathbf{ad}(p)]. \end{aligned} \tag{19}$$

Obviously, the gradients in Eq. 18 are differentiable with respect to p, thus the second derivative in Eq. 17 can be obtained to train the LightGBM model.

V. EXPERIMENTS AND ANALYSIS

In this section, experiments are conducted to show the effectiveness of the proposed approach. All the experiments are carried out on a machine with 12 128G memory modules, and 2 physical 64-core CPUs (AMD 7H12).

A. DATASETS AND PROTOCOL

In order to explore the performance of the proposed approach in financial data more clearly, we choose stock index futures as the datasets, due to its $T+0$ model. The reason is that the Chinese Stock Index Future Market is one of the varieties with the largest trading volume of a single variety, as shown in WIND,[2] with an average trading volume of more than 100 billion per day. The trading of stock index futures is relatively sufficient in the short term, and the cross-sectional impact is very simple. In addition, the push frequency of stock index futures market data is once $500ms$, while the one of stock market data is once $3s$. Therefore, stock index futures contract is a better variety to study short-term prediction.

Of all the trading days, we choose June 22 to September 1 in 2021, which supports our real trading. In addition, we also choose November 1 to December 31 in 2021 to exclude the randomness of the experiment.

We compare the models from three perspectives, including 1) correlation of all data; 2) the maximum and minimum 0.1% return; 3) the maximum and minimum 0.1% accuracy. The details are illustrated as follows:

- **Correlation**: Correlation means the correlation coefficient, namely, given two variables X, Y, correlation coefficient of X and Y is defined as

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var[X] \cdot Var[Y]}}, \tag{20}$$

  where $Cov(\cdot, \cdot)$ and $Var[\cdot]$ are covariance and variance respectively. In our experiment, X and Y represent the truth price and the predicted one, respectively. The reason why we choose the correlation of the total dataset is that we focus on using correlation to evaluate the overall fitness of the model, which can calculate the comprehensive conditions, and is more conducive to our analysis of the over-fitting situation.
- **Return**: Return is the net profits earned through transactions. The reason why we use the maximum and minimum 0.1% is that it is more in line with our actual transactions. Specifically, in our actual transactions, it is general to open positions dozens of times a day. The market push mechanism of stock index futures is to push twice a second, but it will not push if the market does not change. Thus, there are up to (4 hours) $\times$ (3, 600 seconds/hour)$\times$(2 ticks/second) $=$ 28, 800 ticks in a day. Considering the fact that constant tick does not push, we usually receive about 25, 000 ticks, so the maximum and minimum 0.1% correspond to about 50 signals, just in line with our expectations.
- **Accuracy**: Accuracy is a metric that evaluates the performance of predictions, and the reason why we use the maximum and minimum 0.1% is the same as above.

We use Root Mean Squared Error (RMSE), commonly used in price prediction tasks [3], as a baseline to validate the effectiveness of the proposed model. Formally, the objective of RMSE is to minimize

$$\mathcal{L}_{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(l_i - p_i)^2}, \tag{21}$$

where m is the number of samples in dataset. While l_i and p_i represent the i-th true and predicted percent of price respectively.

B. PARAMETER SETTING

We execute experiments to choose the hyperparameters with the best performance of the proposed model. Specifically, α in Eq. 1, 4, 6, 10 is set to 10, while β in Eq. 16 is set to 0.04.

In addition, the parameters in LightGBM[3] include:

- the number of leaves: $num\_leaves = 31$,
- the minimum number of each leaf: $min\_data\_in\_leaf = 50$
- the maximum number of bins: $max\_bin = 320$
- learning rate: $learning\_rate = 0.05$
- the feature fraction required for each iteration: $feature\_fraction = 0.5$
- the data fraction required for each iteration: $bagging\_fraction = 0.7$
- the number of iterations required to perform bagging once: $bagging\_freq = 3$
- the threshold to control when not to split: $lambda\_l1 = 0.5$
- the constant to avoid splitting when there are few node samples: $lambda\_l2 = 0.2$
- the rate of DARTs (Dropouts meet Multiple Additive Regression Trees): $drop\_rate = 0.1$

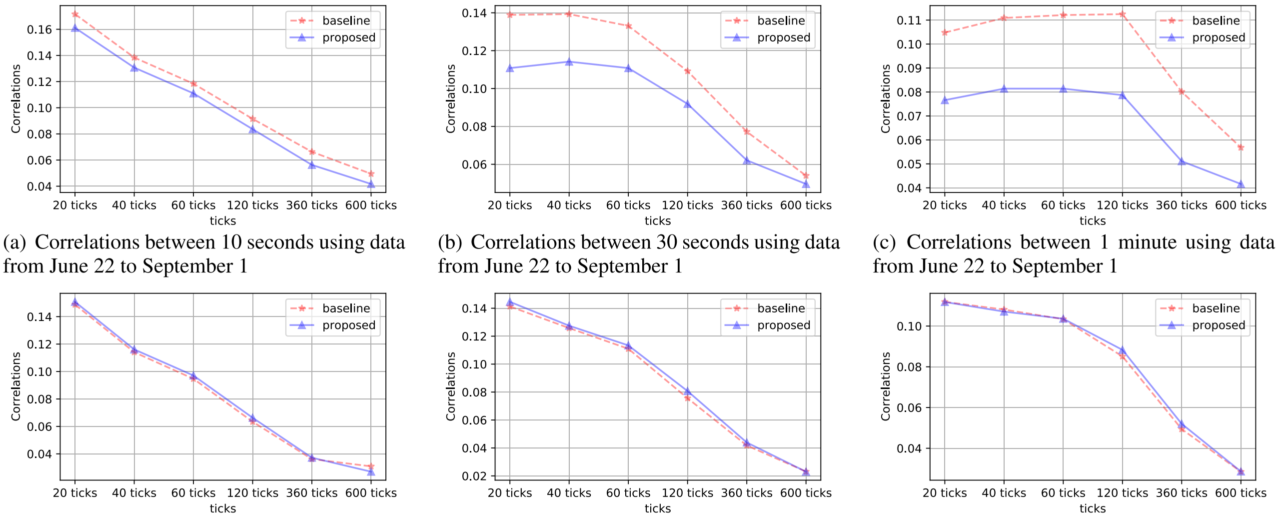
C. EFFECTIVENESS ANALYSIS

As depicted in Fig. 2, we can find that the real time matching our predicted time performs better. Note that the returns decline compared with the data from June 22 to September 1, due to the reason that the intraday price volatility is low during this period, so it still does not affect the experimental conclusion.

Generally, the proposed multi-objective tree model achieves better returns in most of the cases as shown in Fig. 3, while it performs well in both long and short trading terms.

[2]https://www.wind.com.cn/, Wind Data Feed Service, a leading financial data and analysis tool service provider.

[3]The names of parameters are in line with the ones in https://github.com/Microsoft/LightGBM



(a) Correlations between 10 seconds using data from June 22 to September 1

(b) Correlations between 30 seconds using data from June 22 to September 1

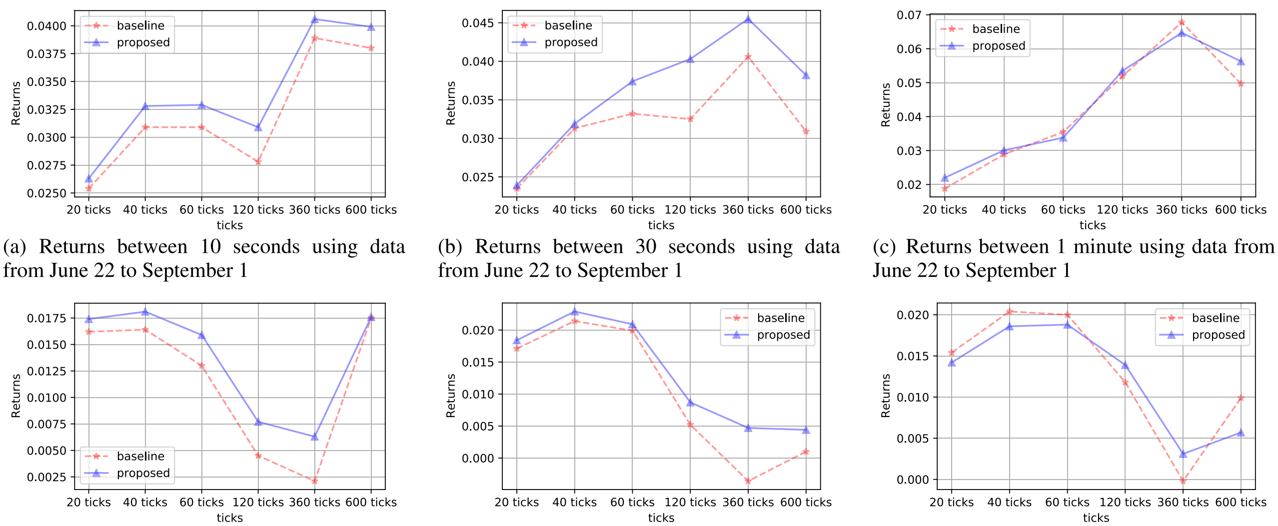
(c) Correlations between 1 minute using data from June 22 to September 1

(d) Correlations between 10 seconds using data from November 1 to December 31

(e) Correlations between 30 seconds using data from November 1 to December 31

(f) Correlations between 1 minute using data from November 1 to December 31

FIGURE 2. Correlations after 10 seconds / 30 seconds / 1 minute. (a)–(c) are from June 22 to September 1, while (d) and (e) are from November 1 to December 31. The referred '10 seconds,' '30 seconds,' and '1 minute' are predicted times, while the horizontal axis represents the real value of different times. Comparing these figures, we can find that the real time matching our predicted time performs better. The correlation using data from November 1 to December 31 differs little between proposed method and baseline, due to the reason that the intraday price volatility is low during this period, so it still does not affect the experimental conclusion.



(a) Returns between 10 seconds using data from June 22 to September 1

(b) Returns between 30 seconds using data from June 22 to September 1

(c) Returns between 1 minute using data from June 22 to September 1

(d) Returns between 10 seconds using data from November 1 to December 31

(e) Returns between 30 seconds using data from November 1 to December 31

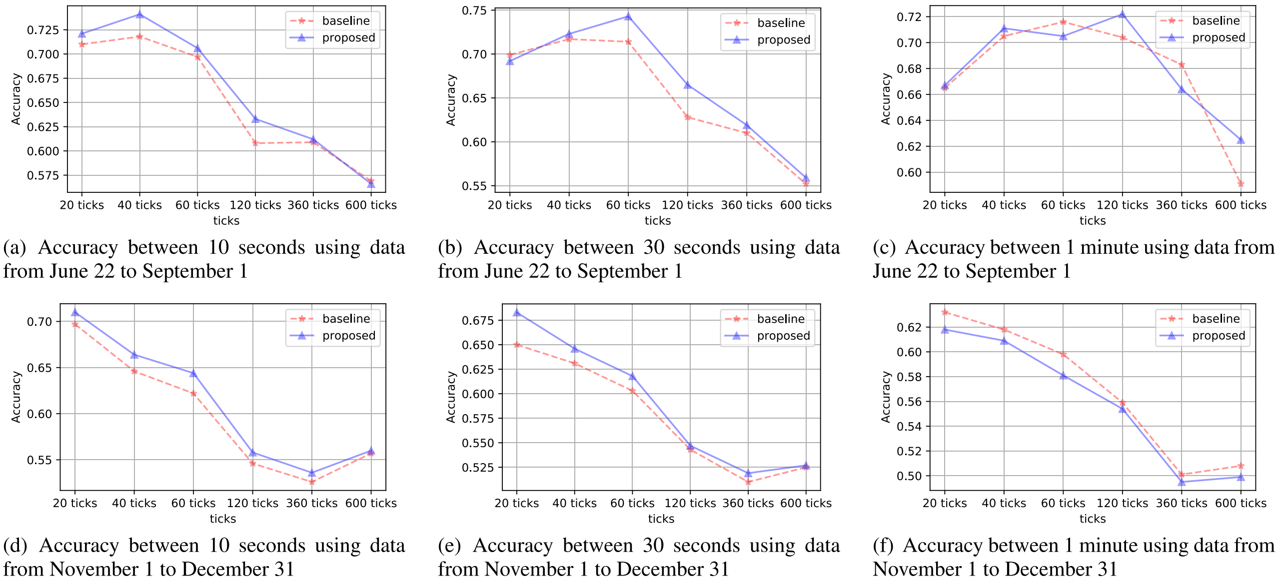
(f) Returns between 1 minute using data from November 1 to December 31

FIGURE 3. Returns between 10 seconds / 30 seconds / 1 minute. (a)–(c) are from June 22 to September 1, while (d) and (e) are from November 1 to December 31. The referred '10 seconds,' '30 seconds,' and '1 minute' are predicted times, while the horizontal axis represents the real value of different times. The proposed multi-objective tree model achieves better returns in most of the cases, while it performs well in both long and short trading terms. The horizontal axis is the time determined by the label, and our experiment performs better when the horizontal axis time is more consistent with the observed time.

In addition, the performance of accuracy, displayed in Fig. 4, is similar to that of returns. This is reasonable because accuracy and return obtain the same dataset, where a similar effect is expected. This phenomenon also helps us further verify the reliability of return.

The above analysis mainly display the performance from the perspective of machine learning. Then, as shown in Fig. 5, we validate our approach on backtest. The return curves in backtest are shown in Fig. 5, where daily results are obtained by rolling training with the data of the previous two months. According to the curves, it is obvious that the proposed one outperforms the baseline.

We further compare the proposed approach with baseline in several financial metrics, as shown in Table 1, including Returns, Drawdown (the maximum rate of return decline when the net value reaches the lowest point), Sharpe Ratio



(a) Accuracy between 10 seconds using data from June 22 to September 1

(b) Accuracy between 30 seconds using data from June 22 to September 1

(c) Accuracy between 1 minute using data from June 22 to September 1

(d) Accuracy between 10 seconds using data from November 1 to December 31

(e) Accuracy between 30 seconds using data from November 1 to December 31

(f) Accuracy between 1 minute using data from November 1 to December 31

FIGURE 4. Accuracy between different 10 seconds / 30 seconds / 1 minute. (a)–(c) are from June 22 to September 1, while (d) and (e) are from November 1 to December 31. The referred '10 seconds,' '30 seconds,' and '1 minute' are predicted times, while the horizontal axis represents the real value of different times. The proposed multi-objective tree model achieves better accuracy in most of the cases, while it performs well in both long and short trading terms. The horizontal axis is the time determined by the label, and our experiment performs better when the horizontal axis time is more consistent with the observed time.

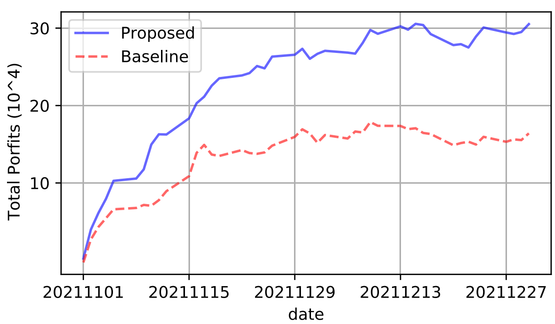


FIGURE 5. Return curves of the proposed method and baseline. The proposed one outperforms baseline.

FIGURE 6. Futures strategy return in real trading. The period is the same as the one in backtest.

TABLE 1. Comparison between proposed approach and baseline in several financial metrics. Bold numbers indicate the better performances. Note that our proposed approach surpass the baseline in all metrics except 'DrawDown,' and there is little difference in 'DrawDown.'

Metric	proposed	baseline
Returns	**6.10%**	3.28%
DrawDown	-0.611%	**-0.599%**
Sharpe Ratio	**9.9006**	6.1155
Annualized Rate of Returns	**34.96%**	18.79%
TotalProfit Fee Rate	**2.5817**	2.0148
Daily Winning Rate	**0.7045**	0.5682

TABLE 2. Financial metrics in real trading, including return, drawdown, sharpe ratio, annualized rate of returns, and daily winning rate.

Metric	real trading
Return	1.53%
DrawDown	-0.427%
Sharpe Ratio	3.1448
Annualized Rate of Returns	8.56%
TotalProfit Fee Rate	2.0162
Daily Winning Rate	0.5556

(Ratio of excess return to volatility), Annualized Rate of Returns, TotalProfit Fee Ratio and Daily Winning Rate. We find that the proposed approach achieves better performance in all metrics except 'DrawDown', and there is little difference in 'DrawDown'.

Besides the backtest performance, we also validate the performance on real trading, as shown in Fig. 6. Likewise, we give the financial metrics in real trading in Table 2. Note that we use 5 million CNY in both backtest and real trading to calculate the returns and drawDown. There exists a distance between backtest returns and real trading returns, because

1) We always assume that we can send orders with the fastest speed in the whole market in backtest, while it is impossible in real trading; 2) Order is affected by the model and current position in real trading, while there is no effect on backtest;

3) Real trading can affect the market, while backtest can not. Therefore, it is commonplace that returns in real trading are lower than those in backtest, especially in high-frequency trading.

VI. CONCLUSION AND OUTLOOK

This paper proposes a multi-objective tree model to solve the problem that the accuracy of price prediction are inconsistent with the real returns. We deeply discuss the essential reason and devise a quantitative strategy to induce the complex characteristics of real trading. Through multiple experiments, we find the proposed model achieves better returns and accuracy than the baseline, and especially the difference between model prediction and returns in real trading can be diminished. However, there are still lots of ways to improve. For example, the design of the objectives in Sec. IV-B are too complicated, which leads to the high computing overhead in Eq. 18. It is a challenge for us to think about how to simplify the objectives and better match the real trading.

For future work, we believe there are two important technical areas worth further study.

First, there is still much space for introducing advanced temporal data learning models. Specifically, temporal point process is an expressive framework for learning the event sequence with different time intervals between events. The application scenario includes clustering [23], forecasting [24], and casual or relation mining [25]. On the other hand, classic time series data is still an important area and anomaly detection [26] as well as forecasting [27] can be advanced, especially by deep networks.

Second, apart from temporal learning, the relation modeling among different entities is also an important aspect whereby graph structure learning, and especially graph neural networks [28] and network embedding [29], [30] can show their potential.

Last but not least, operation research, which is especially performed in a data-driven manner via machine learning, is another promising direction [31] since trading is essentially a decision making process. We have witnessed fast development in this area in recent years especially for combinatorial optimization [32] which relates to portfolio optimization and management in finance.

REFERENCES

[1] A. A. Ariyo, A. O. Adewumi, and C. K. Ayo, "Stock price prediction using the ARIMA model," in *Proc. 16th Int. Conf. Comput. Modeling Simulation (UKSim-AMSS)*, Mar. 2014, pp. 106–112.

[2] M. Sharaf, E. E.-D. Hemdan, A. El-Sayed, and N. A. El-Bahnasawy, "StockPred: A framework for stock price prediction," *Multimedia Tools Appl.*, vol. 80, no. 12, pp. 17923–17954, May 2021.

[3] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon, and K. P. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2017, pp. 1643–1647.

[4] W. Lu, J. Li, J. Wang, and L. Qin, "A CNN-BiLSTM-AM method for stock price prediction," *Neural Comput. Appl.*, vol. 33, no. 10, pp. 4741–4753, May 2021.

[5] P. Fodra and H. Pham, "High frequency trading and asymptotics for small risk aversion in a Markov renewal model," *SIAM J. Financial Math.*, vol. 6, no. 1, pp. 656–684, Jan. 2015.

[6] B. Xiang, M. Rozaini, and A. S. Abdullah, *A Stock Market Trading System Using Deep Neural Network*. Singapore: Springer, 2017.

[7] L. Takeuchi and Y.-Y. A. Lee, "Applying deep learning to enhance momentum trading strategies in stocks," Stanford Univ., Stanford, CA, USA, Tech. Rep., 2013.

[8] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[9] E. Platanakis and A. Urquhart, "Portfolio management with cryptocurrencies: The role of estimation risk," *Econ. Lett.*, vol. 177, pp. 76–80, Apr. 2019.

[10] R. Culkin and S. R. Das, "Machine learning in finance: The case of deep learning for option pricing," *J. Investment Manage.*, vol. 15, no. 4, pp. 92–100, 2017.

[11] P.-Y. Hsu, C. Chou, S.-H. Huang, and A.-P. Chen, "A market making quotation strategy based on dual deep learning agents for option pricing and bid-ask spread estimation," in *Proc. IEEE Int. Conf. Agents (ICA)*, Jul. 2018, pp. 99–104.

[12] H.-L. Shi, Z.-Q. Jiang, and W.-X. Zhou, "Profitability of contrarian strategies in the Chinese stock market," *PLoS ONE*, vol. 10, no. 9, Sep. 2015, Art. no. e0137892.

[13] N. Jegadeesh and S. Titman, "Returns to buying winners and selling losers: Implications for stock market efficiency," *J. Finance*, vol. 48, no. 1, pp. 65–91, Mar. 1993.

[14] H. V. Nguyen, M. A. Naeem, N. Wichitaksorn, and R. Pears, "A smart system for short-term price prediction using time series models," *Comput. Electr. Eng.*, vol. 76, pp. 339–352, Jun. 2019.

[15] Y. Son, D.-J. Noh, and J. Lee, "Forecasting trends of high-frequency KOSPI200 index data using learning classifiers," *Expert Syst. Appl.*, vol. 39, no. 14, pp. 11607–11615, Oct. 2012.

[16] M. M. Patel, S. Tanwar, R. Gupta, and N. Kumar, "A deep learning-based cryptocurrency price prediction scheme for financial institutions," *J. Inf. Secur. Appl.*, vol. 55, Dec. 2020, Art. no. 102583.

[17] S. Mehtab, J. Sen, and A. Dutta, "Stock price prediction using machine learning and LSTM-based deep learning models," in *Proc. Symp. Mach. Learn. Metaheuristics Algorithms, Appl.* Cham, Switzerland: Springer, 2020, pp. 88–106.

[18] N. Ren, M. Zargham, and S. Rahimi, "A decision tree-based classification approach to rule extraction for security analysis," *Int. J. Inf. Technol. Decis. Making*, vol. 5, no. 1, pp. 227–240, Mar. 2006.

[19] R. Choudhry and K. Garg, "A hybrid machine learning system for stock market forecasting," *World Acad. Sci., Eng. Technol.*, vol. 39, no. 3, pp. 315–318, 2008.

[20] J. Zhong, X. Zhang, K. Gui, Y. Wang, H. Che, X. Shen, L. Zhang, Y. Zhang, J. Sun, and W. Zhang, "Robust prediction of hourly $PM_{2.5}$ from meteorological data using LightGBM," *Nat. Sci. Rev.*, vol. 8, no. 10, Nov. 2021, Art. no. nwaa307.

[21] X. Zhao and Q. Zhao, "Stock prediction using optimized LightGBM based on cost awareness," in *Proc. 5th IEEE Int. Conf. Cybern. (CYBCONF)*, Jun. 2021, pp. 107–113.

[22] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–9.

[23] W. Wu, J. Yan, X. Yang, and H. Zha, "Discovering temporal patterns for event sequence clustering via policy mixture model," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 2, pp. 573–586, Feb. 2022.

[24] L. Li, J. Zhang, J. Yan, Y. Jin, Y. Zhang, Y. Duan, and G. Tian, "Synergetic learning of heterogeneous temporal sequences for multi-horizon probabilistic forecasting," in *Proc. AAAI*, 2021, pp. 1–9.

[25] Y. Zhang and J. Yan, "Neural relation inference for multi-dimensional temporal point processes via message passing graph," in *Proc. 13th Int. Joint Conf. Artif. Intell.*, Aug. 2021, pp. 3406–3412.

[26] L. Li, J. Yan, H. Wang, and Y. Jin, "Anomaly detection of time series with smoothness-inducing sequential variational auto-encoder," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 3, pp. 1177–1191, Mar. 2021.

[27] L. Li, J. Yan, X. Yang, and Y. Jin, "Learning interpretable deep state space model for probabilistic time series forecasting," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 2901–2908.

[28] L. Chen, Z. Qiao, M. Wang, C. Wang, R. Du, and H. E. Stanley, "Which artificial intelligence algorithm better predicts the Chinese stock market?" *IEEE Access*, vol. 6, pp. 48625–48633, 2018.

[29] X. Du, J. Yan, R. Zhang, and H. Zha, "Cross-network skip-gram embedding for joint network alignment and link prediction," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 3, pp. 1080–1095, Mar. 2022.

[30] H. Xiong and J. Yan, "BTWalk: Branching tree random walk for multi-order structured network embedding," *IEEE Trans. Knowl. Data Eng.*, early access, Oct. 6, 2020, doi: 10.1109/TKDE.2020.3029061.
[31] A. K. Sangaiah, A. Goli, E. B. Tirkolaee, M. Ranjbar-Bourani, H. M. Pandey, and W. Zhang, "Big data-driven cognitive computing system for optimization of social media analytics," *IEEE Access*, vol. 8, pp. 82215–82226, 2020.
[32] W. T. Scherer, S. Adams, and P. A. Beling, "On the practical art of state definitions for Markov decision process construction," *IEEE Access*, vol. 6, pp. 21115–21128, 2018.

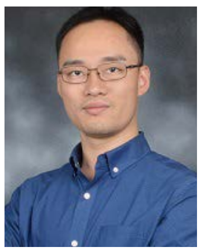
TAO YIN received the bachelor's and M.S. degrees in computer science from Shanghai Jiao Tong University, Shanghai, China, in 2011 and 2014, respectively, where he is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering. His research interests include machine learning and financial data analytics.

XINGBO DU received the B.S. degree from the Department of Mathematics, East China Normal University, Shanghai, China, and the M.E. degree from the School of Software Engineering Computer Science and Software Engineering, East China Normal University. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai. His current research interests include drug discovery, biological/social/financial networks mining, machine learning, and semi-supervised learning.

WEIPENG ZHANG received the B.E. and M.S. degrees from the Department of Computer Science and Engineering, Shanghai Jiao Tong University, in 2012 and 2015, respectively, where he is currently pursuing the Ph.D. degree in artificial intelligence with the Department of Computer Science and Engineering. His research interests include machine learning and data mining.

YUNAN ZHAO received the B.E. and M.E. degrees in automation from Shanghai Jiao Tong University, Shanghai, China, in 2013 and 2016, respectively. He was an Engineer with China Merchants Bank, Shanghai. He is currently a Staff Researcher with MYbank, Ant Group, Shanghai. His research interests include machine learning and computer vision.

BING HAN is currently a Senior Staff Engineer and the Head of the Intelligent Engine Department, MYbank, Ant Group. Her research interests include machine learning and data intelligence, especially in recommender systems and financial technology.



JUNCHI YAN (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Shanghai Jiao Tong University, China, in 2015. He is currently an Associate Professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University. Before that, he was a Senior Research Staff Member and a Principal Scientist with IBM Research—China, where he started his career, in April 2011. His research interests include machine learning and computer vision. He serves as the Area Chair for ACM-MM 2021/22, CVPR 2021, AAAI 2022, and ICML 2022; a Senior PC for IJCAI21/22; and an Associate Editor for *Pattern Recognition* and IEEE ACCESS.