A decision level identity fusion based prediction model for smooth fluctuation products

Sixu Han[1] ,Chi Wang[1], Longting Xu[1]

[1]Donghua University Shanghai, College of Information Science and Technology, P.R.China

(Corresponding author: Chi Wang and Longting Xu)

2022 ICEICT

Abstract

Nowadays, various indicators with different features are used to analyze the price trend and forecast the future prices for investment decisions. However, for investment products that do not have drastic price trend changes, such as gold, when making decisions on whether to buy or sell based on the predicted price usually does not have good returns. Therefore, the prediction model for investment products with smooth fluctuation (PIPSF) is proposed based on the idea of decision level identity fusion. In this model the 'ideal buy and sell points' is established to decide the buying or selling of investment products. In the case of the large amount of data and feature vectors, the elastic net regression was used for feature selection, it combines the advantages of two regularization methods to ensure both the goodness of fit and the generalization ability of the model. The experimental results show that the model proposed in this paper has higher returns when compared with direct price prediction.

Keywords—decision level identity fusion, machine learning, ideal buy and sell points, elastic net regression, correlation coefficients, regularization

INTRODUCTION

Among them, ARIMA model and LSTM model have good prediction effects for the investment products with more significant price fluctuations. However, they often do not provide good forecasting results for investment products with relatively stable prices.

For such investment products, since decisions of buy and sell can't not be made based on the prices, the 'ideal buy and sell points' should be established, while the result of the prediction is the decision itself. The size of the profitability index is used to indicate how good the timing is for buying and selling. Gold is a typical investment product with smooth price fluctuations.

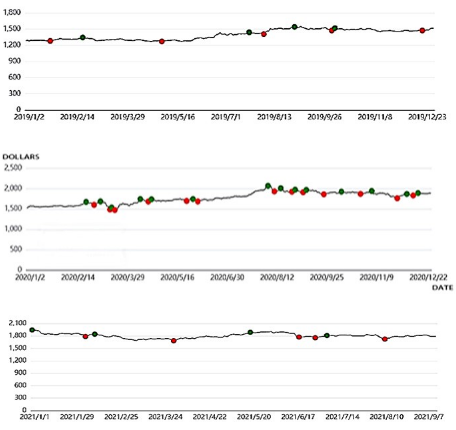

Fig. 1. Ideal buy points and sell points of gold in 2019-2021

RELATED WORK

A. ARIMA Model

After the difference operation, the series has constant mean and variance, which makes the prediction results not excessively biased. However, it cannot learn the nonlinear relationship, which may lead to a poor prediction effect.

B. LSTM Model

In the case of a large number of training sets and enough epochs, the training effect of LSTM model will be good, with a high prediction accuracy, but it will also cause overfitting. Meanwhile, the robustness of the model is not good enough.

ESTABLISHMENT OF IDEAL BUY AND SELL POINTS

A. Derivation of the Rate of Return

One of the main determinants of the ideal buy and sell points is the rate of return.

$$T_0 = b_1 \cdot m \cdot (1+\alpha) \tag{1}$$

$$P_1 = s_1 \cdot m \cdot (1-\alpha) - T_0 = \frac{s_1 \cdot (1-\alpha)}{b_1 \cdot (1+\alpha)} - T_0 \tag{2}$$

$$T_1 = T_0 + P_1 = \frac{s_1 \cdot (1-\alpha)}{b_1 \cdot (1+\alpha)} T_0 \tag{3}$$

$$T_2 = T_1 + P_2 = \frac{s_2 \cdot (1-\alpha)}{b_2 \cdot (1+\alpha)} T_1 = \frac{s_2 \cdot s_1}{b_2 \cdot b_1} \left(\frac{1-\alpha}{1+\alpha}\right)^2 T_0 \tag{4}$$

$$T_n = \prod_{i=1}^{n} \left(\frac{s_i}{b_i} \cdot \frac{1-\alpha}{1+\alpha}\right) T_0 \tag{5}$$

$$P_n = \frac{T_n}{T_0} = \prod_{i=1}^{n} \left(\frac{s_i}{b_i} \cdot \frac{1-\alpha}{1+\alpha}\right) \tag{6}$$

It is related to the bid-ask ratio, the number of trades and the trading rate.To achieve the effect of incremental increase in total assets, it should satisfy the following two conditions respectively

$$\frac{s_i}{b_i} \cdot \frac{1-\alpha}{1+\alpha} > 1 \tag{7}$$

$$s_i(1-\alpha) - b_{i+1}(1+\alpha) > 0 \tag{8}$$

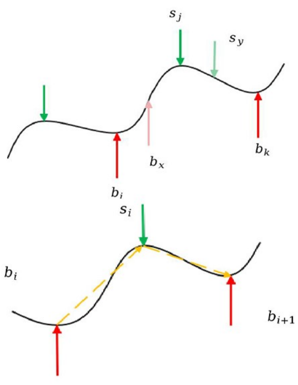

Fig. 2. Diagram of profitability index

B. Calculation of the Profitability Index ϵ

The profitability index ϵ is defined to indicate the timing of buying and selling. Asset appreciation when $\epsilon>0$,when $\epsilon<0$, the asset shrinks. As shown in the Fig.2 , the following analysis is performed.

There are four situations, the following is the analysis of these.

1) Ideal Sell Point Calculation ε

Sell at point s_j and sell at point b_k to calculate the maximum loss.

2) Calculate ϵ when it is a non-optimal buy point

Calculated with b_x and s_y, similar to case 1.

3) Calculate ϵ when it is a non-optimal sell point

Calculated with s_y and b_x, similar to case 2.

4) For the last partial data status, there may be two situations: holding position and short position

a) Holding positions:

Set the seed end point as the ideal sell point, and the calculation method is similar to the ideal sell point.

b) Short position:

Let the seed end point be the ideal buy point, and the calculation method is similar to the ideal buy point.

FEATURES SELECTION OF PIPSF MODEL

Lasso regression achieves feature sparsity by increasing the penalty term of L1 norm, which can solve the problems of over-fitting and multicollinearity.

$$Cost(w) = \sum_{i=1}^{N} (y_i - w^T x_i)^2 + \lambda ||w||_1$$

However, when multiple features are correlated, Lasso regression will only select one of them, which will lead to poor goodness of fit of the model. Therefore, add the penalty term of L2 norm and appropriately reduce the coefficient of L1 norm, that's the elastic net regression

$$Cost(w) = \sum_{i=1}^{N} (y_i - w^T x_i)^2 + \lambda\rho ||w||_1 + \frac{\lambda(1-\rho)}{2} ||w||_2^2$$

The correlation coefficient of each feature was calculated according to the least Angle regression method , as shown in Fig.3.

FEATURE	Nflag	D_4	AGR_4	S_8	RCR_5
f_i	f_1	f_{11}	f_6	f_{21}	f_{16}

TABLE I. THE CORRESPONDING RELATIONSHIP OF THE SELECTED FEATURES

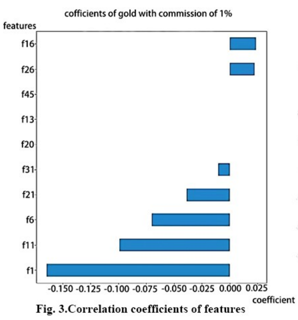

Fig. 3.Correlation coefficients of features

Regression is performed according to the selected features, and the final model is

$$f_{gold} = 4.11 + (-0.215) \times D_4 + (-25.43) \times AGR_4 + 0.191 \times S_8 + (-0.382) \times Nflag + (-342.764) \times RCR_5$$

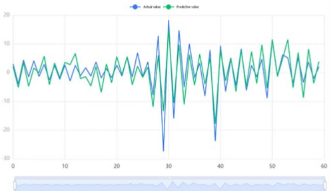

Fig. 4.Elastic Net Regression

EXPERIMENTS AND RESULTS

A. Prediction Experiments

The PIPSF model is obtained by predicting the ideal buying and selling points and selecting the features of a large number of data and feature vectors. It combines the advantages of the two regularization methods and ensures the goodness of fit of the model as well as the generalization ability. In order to prove that the proposed model has better results on the investment products with smooth fluctuations than the ARIMA model and the LSTM model, a prediction experiment was carried out on gold. In this experiment, the initial capital is \$1,000, the transaction cost is =1%, and the trading period is 5 years.

B. Analysis of Results

Based on the models' prediction results, the daily return is obtained. The return curves of PIPSF model, ARIMA model and LSTM model are shown in Fig.5. The average final return of the proposed PIPSF model is \$18,257, the average final return of the ARIMA model is \$12,325, and the average final return of the LSTM model is \$15,133. Obviously, the PIPSF model obtained the highest final benefit. In addition to the final return, the PIPFS model has better stability and fault tolerance during the trading period.

1) In the middle of the return curve, it is observed the return of PIPFS model rises steadily and the return of LSTM model almost remains stable, while that of ARIMA Model fluctuated significantly. The prediction of PIPSF model is more accurate, which obviously brings a more stable return.

2) In the latter part of the return curve,it is observed the returns of all models decline to some extent, but the decline of PIPFS model is the least because it avoids frequent trading.

The proposed PIPFS model is a more stable and more effective model for investment products with smooth fluctuations.

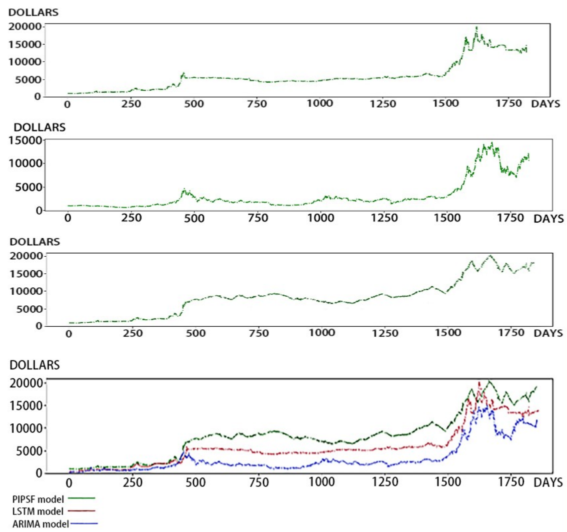

Fig. 5.The return curve of PIPSF model, ARIMA model and LSTM model

CONCLUSION

In this work, we propose the PIPSF model, which has a better prediction effect than the ARIMA model and the LSTM model for investment products with stable price fluctuations, such as gold. Experiments have also proved that the final gain is better. The model has good robustness that can select different effective features for different investment products. The future work will focus on extending the model to fit all investment products, which will make the investment more intelligent and reliable.