Best Investment Strategy: Prediction Based on AR-LSTM & Decision Based on SVM

Changhong Zhao, Kang Li

Shijiazhuang Tiedao University, Shijiazhuang 050043, China.

Abstract: In volatile asset market transactions, traders always want to maximize the total returns. This paper performs day-by-day forecasting of closing prices, quantitatively analyzing the market using economic indicators, and proposes the best daily investment strategy that maximizes total returns.

Firstly, we use Python's TensorFlow library to build an AR-LSTM model to make day-by-day predictions of the closing prices of gold and bitcoin. Secondly, we use the greedy algorithm, support vector machines(SVM) combined with three economic indicators, a total of two methods to analyze the market situation. On the one hand, we rely on the predicted results and use the greedy idea to get the maximum total return . On the other hand, we established the Support vector machine optimize specifications model. Firstly, 15 indicators that can express the characteristics of market oscillation, trend, and energy are selected for factor analysis. Secondly, three public factors were identified. The calculated factor scores were then fed into a SVM, which, combined with the linear programming results, can output one of 2 (trade) or 1 (continue to hold). Thirdly, the three technical indicators, EMA, RSI, and MACD, are calculated by applyin g the predicted closing price. Finally, the indicators are linearly weighted, and the weighted results are re-operated with the output of the support vector machine to obtain the trading intention for the day. Then, we use Monte Carlo and grid search methods to adjust the four parameters in the model, w1, w2, r, uplimit. We use grid search, fix w1, w2 and adjust r and uplimit. Subsequently, we fixed r, uplimit, and used the Monte Carlo method to adjust the size of w1, w2 and determined the optimal daily investment strategy.

***Keywords**:* AR-LSTM ; Investment Decision; SVM ; Factor Analysis; Parameter Optimization

1. Background

Many investors consider bitcoin as a new type of gold, which can replace gold for risk management, hedging against inflation. As an investor with $1000 and plans to trade for five years, choosing the right time to trade and allocating assets will directly affect the investment outcome. Therefore, we should follow the market rules, grasp the investment methods and find the best trading strategy.

1.1 Our Work

First, we obtain the trading prices at the data source before Sept 11th, 2016, used as the Warmup period data. In the forecasting phase, we use the AR-LSTM model to analyze the trading prices for the Warmup period thus forecast the trading prices for the window period (Sept 11th, 2016, to Sept 10th, 2021). First, in the decision definition phase, we analyze the best investment strategy trades during the forecast period , using it as a label. We use factor analysis to score the economic indicators for each trading day. Then we trained an SVM from the scores and labels. Finally, we get the investment strategy by combining the economic indicators and the SVM results for each trading day.

1.2 Data Viewing

We collected data from the official source, and it shows gold and bitcoin prices have fared in recent years. Gold rose overall around 2018-2020, maintaining small fluctuations at other points in time. Bitcoin closing below $20,000 in 2017 and broke $60,000 in 2021, with colossal volatility and high speculation relative to gold.

1.3 Linear programming solves for maximum gain

There are two categories of problems: First, predicting the trading day's closing price. We should use the daily price flow before predicting the trading day's closing price. After obtaining the actual closing price of the trading day, this data is used as historical data to update the prediction model. Second, decision daily trading strategy. Compare the current closing price with the closing price of the next trading day to determine whether to buy, hold or sell assets in the portfolio. If the purchase or sale is determined, and the new venture portfolio is determined. The mathematical model is as follows.

$$Max\ C_{1826}$$

$$s.t.\begin{cases} gh_1 = 0 \quad bh_1 = 0 \quad C_1 = 1000 \\ gh_i = gh_{i-1} + gan_i - gon_i \\ bh_i = bh_{i-1} + ba_i - bo_i \\ gon_i \leqslant gh_{i-1} \\ bo_i \leqslant bh_{i-1} \\ C_i = C_{i-1} - \dfrac{ga_i \cdot gv_i}{1 - galpha} - \dfrac{ba_i \cdot bv_i}{1 - balpha} + (1 - galpha) go_i gv_i + (1 - balpha) bo_i bv_i \\ C_i \geqslant 0 \\ gan_i = ax_i \times ga_i \\ gon_i = ax_i \times go_i \end{cases}$$

Where ga denotes gold purchases, go denotes gold sales, ba is bitcoin purchases, bo is bitcoin sales, and ch, gh, and bh are cash, gold, and bitcoin holdings. The analysis of the results shows that the initial $1,000 experienced 5 years of investment and increased in value to $34.5 billion. Furthermore, at each sale, all of it was sold off. Referring to this strategy, we choose to buy all and sell all when we make subsequent sales or buys.

2. Closing price prediction based AR-LSTM

In recent years, LSTM has been widely used in financial series forecasting in the stock market, and the forecasting effect has been improved. Therefore, we think of combining time-series and deep earning, i.e., AR-LSTM model for forecasting.[1]

Firstly, the input sequence X is preheated. Then enter the LSTM network to start the prediction. The general LSTM network has one input and directly outputs a set of prediction results. In order to improve the accuracy of prediction, we introduce the AR model here. It first uses all the historical data to predict the results of one period, then uses the new period results together with the historical data as the new historical data to predict the following period results, and so forth. Next, the output is corrected with Label to obtain better prediction results.

3. Common technical specifications

Technical market indicators can be divided into the oscillator, the energy, and trend categories. The oscillator category contains RSI, KDJ, WR. The energy category mainly includes VR, OBV , VOL. The common trend categories are MACD, MA, DMI, EXPMA. The following are the three most important indicators[2].

3.1 Moving Average(MA) & Exponential Moving Average(EMA)

Moving Average is the arithmetic average of the closing prices in a selected trading cycle. The calculation formula is as follows.

$$MA_i = \frac{\sum_n Value_{i-n}}{n}\ ,\ EMA_i = k \times value_i + (1-k) \times EMA_{i-1}\ ,\ k = \frac{2}{n+1}$$

Where n is the number of moving average periods, and i is the current period. The current market is more widely used averages depending on the trading period divided into short term, 5MA, 10MA, medium-term 20MA, 60MA, long term 120MA, and 250MA.

MA and EMA are very similar. At the same time, EMA is a cumulative calculation that includes all historical data. Therefore, EMA is used in this article instead of MA. If there is no intersection, the trade indication is 0, The ratio is taken to be positive if a buy is recommended. Otherwise, it is negative.

3.2 Moving Average Convergence and Diver gence (MACD)

MACD is obtained using the difference between the deviation of the two fast (DIF) and slow (DEA) smoothed moving averages, commonly calculated with the 26-day and 12-day EMA periods. The mathematical expression is as follows:

$$DIF_i = EMA^{12}{}_i - EMA^{26}{}_i\ ,\ DEA_i = \frac{\sum_j DIF_{i-j}}{j}\ ,\ MACD = 3(DIF - DEA)$$

The main application principle of this technical indicator is to make a buyer's entry when the DIF breaks upwards through the MACD. When the DIF falls below the MACD, it is time to sell gold or bitcoin. ed using a cubic polynomial, and the MACD for the next period is predicted. If the curve between the current and subsequent periods does not cross the 0 axes, the trade indication is 0, If it crosses the 0-axis from below 0-axis at this point, it has a buying bias, and the ratio is taken as positive otherwise, it is negative.

3.3 Relative Strength Index (RSI)

RSI predicts the market direction by analyzing the market buying and selling, supply and demand. The calculation formula is as follows.

$$RSI_i = \frac{upavg}{upavg + downavg} \times 100$$

Where upavg is the average of all the upside in a period and downavg is the average of the downside, which takes values in the range [0,100]. When RSI>80 is considered a buyer's market overflow, it is appropriate to sell gold or bitcoin. On the contrary, when RSI<20 means that the sellers' market is overflowing, the price is about to retrace, and you can buy. setting the overbought upper limit at 80, the oversold lower limit at 20 and the mean at 50. If there is no deviation from the mean for both the current and following periods, the trade indication is 0, If both are biased towards the oversold zone, then it tends to buy, and the ratio is taken to be positive; otherwise, it is negative.

4. Factor Analysis

We selected 15 technical indicators, including EMA5, EMA10, RSI5, MACD and MACDHIST, for factor analysis.[3] First, we normalize the zscore for the 15 indicators. Then, we used factor rotation to obtain a new loading matrix. From the loadings matrix, we attribute the common factors to EMA5, RSI5, and MACD based on each column of data's top and bottom comparisons. For bitcoin, the cumulative contribution of the three public factors reaches 84.59%, and for gold, the cumulative contribution of the three public factors reaches 87.97%, which can be accepted.

5. Training support vector machines

We use machine learning thinking to assist in the analysis. Before using the SVM, we must first use the data from the factor analysis and label them. For each investment, there is a total of three different states: doing nothing, buying the traded items, and selling the traded items, which we write down as class1, class2, and class3. Because class2 and class3 have good distinguishability and support vector machines can directly deal with the problem of binary classification, we combine class2 and class3 into a class, collectively referred to as a positive example (transaction). Record calss1 as the negative class (no transaction). We then use the binary support vector machine to learn the data parameters given by the factor analysis.

We quantify the trading principles used to make decisions for the three indicators, and map them between [-1,1], where 1 indicates the maximum probability of buying, -1 indicates the maximum probability of selling and 0 indicates no trade. The EMA, RSI, and MACD are linearly weighted, since the most appropriate weights cannot be derived initially, they are temporarily recorded as[w1, w2, w3].Finally, the calculation of W is as follows:

$$W_1 = w_1 EMA + w_2 RSI + w_3 MACD$$

For parameter adjustment stage. Firstly, we add the result of the support vector machine. If the support vector machine determines a trade, multiply W1 by 2. If the support vector machine determines no trade, divide W1 by 2. Becausew1 + w2 + w3 = 1and uplimit and lowlimit are opposite each other, so the actual number of parameters to be adjusted is only 4. Secondly, we control w1 and w2, change uplimit and r, and start the network search with the goal of maximum return on the final trading day. The result shows that the maximum value of the maximum return obtained on the final return day is $3714.94 when uplimit = lowlimit = 0.42, r = 0.3. Lastly, the optimal parameters of uplimit and r are fixed, and the values of w1, w2, w3 parameters are adjusted using Monte Carlo simulation to find the maximum value of the maximum return. We get the maximum return of $7086.4 for w1 = 0.0454, w2 = 0.9477, w3 = 0.0069. In summary, we finally determined the maximum return of $7086.4 and we can get the best daily investment strategy.

References

[1] Liu L, Lou L, Liu XJ, and Shi SZ. Comparative study of comex gold futures price forecasting based on lstm. Journal of Changchun University of Science and Technology: Natural Science Edition, 44(2):7, 2021.

[2] Fang PC. A machine learning based stock prediction and quantitative investment system. PhD thesis, Zhejiang University, 2018.

[3] Zhou H. Quantitative investment model and practice based on dual technical indicators. PhD thesis, South China University, 2019.