Intelligent Decisionmaking System through LSTM Prediction Model and DQN algorithm

Jiafeng Hu
School of Mathematics and Big Data, Anhui University of Technology, Huainan, Anhui, 232001, China
605281999@qq.com

2022 IEEE 2nd International Conference on Data Science and Computer Application (ICDSCA) | 978-1-6654-7200-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICDSCA56264.2022.9988461

***Abstract*—Today's market is trading more frequently, and traditional forecasting models are no longer applicable. Inspired by AlphaGo's game ideas, after consulting the economic and financial research literature related to artificial intelligence at home and abroad, in order to maximize traders' returns, minimize risks, and optimize decisions, this paper proposes an intelligent body decision model based on deep reinforcement learning and applies it to the actual columns of this paper. an LSTM prediction model was established to predict the price of gold and Bitcoin, and it was found that the data fit worked best when the number of iterations was 100. Then, the CPMA model is established to establish a risk assessment mechanism and make preconditions for the decision-making model. Finally, a reinforcement learning model is established based on artificial intelligence, and LSTM is used to extract data features and is combined with DQN algorithm to solve. Based on the CPMA model of problem one, we correlate this improved model with transaction costs. And conduct descriptive statistics on the relationship between yield rate and cost of revenue. It can be seen that the transaction cost adjustment has a large short-term impact on the market, and the long-term impact is relatively small .Therefore, we recommend reducing costs appropriately to improve profitability.**

Keywords—Reinforcement Learning, CPMA, LSTM, DQN, T-test

I. Introduction

The stock market is a dynamic, nonlinear and high-noise system. It is affected by many complex factors, such as political factors, economic conditions and investors' expectations. The change of stock price is often nonlinear and unsustainable. At the same time, with the continuous development of financial economy, the stock market situation has become more complicated, Because of advances in science and technology and the upgrading of information transmission systems, events take less time to be reflected in stock prices. As a result, the stock price is more sensitive to related events and fluctuates at a higher frequency. The traditional way of investment can no longer meet people's growing investment demand.

Stock price economic research based on traditional economics Wolfang Reitgruber Using ARMA model, analyzed data from stock trading in Vienna and the United States. The conclusion is that American stock market is more consistent with ARMA model [1].Pesaran and Timmermann applied the extended and general version of recursive modeling strategy to the prediction of UK stock market returns in the study of Stock Price Volatility based on machine learning Algorithm [2].In the research of stock volatility based on deep learning algorithm, Ugur Gudelek, Murat Ozbayoglu etal. adopted deep learning method based on convolutional neural network to predict stock prices by extracting features from time series. [3]

We build LSTM model for prediction, CPMA model for risk assessment and reinforcement learning model for intelligent decision making. Based on the established model, the prediction model and decision model are tested. Then the CPMA model is improved to explore the sensitivity of transaction costs. The overall work is shown in Fig 1.

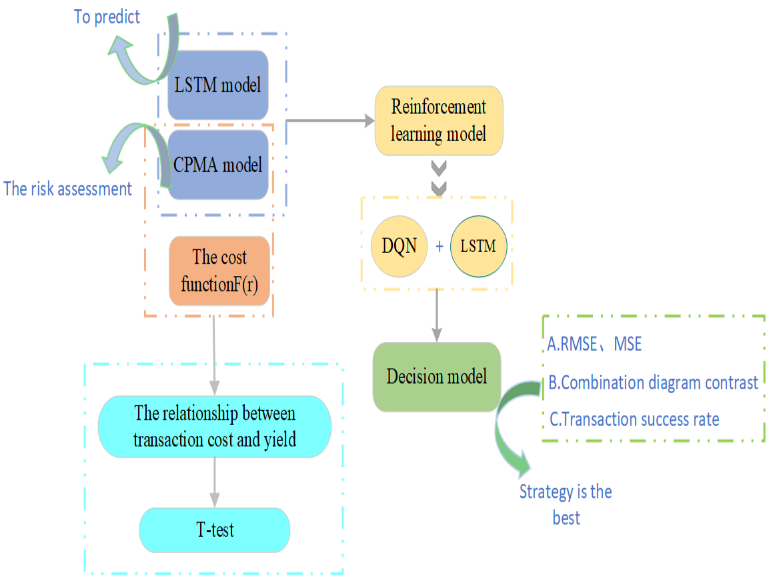

Fig. 1 The Flow Chart in This Paper

II. Notations

Table I shows the symbol description.

TABLE I SYMBOL DESCRIPTION

Symbol	Explanation
W_f	The weight of the forgotten door
b_f	The bias of the forgotten door
W_i	The appropriate weight
W_c	The appropriate weight
b_i	The corresponding bias
b_c	The corresponding bias
C^t	The current cell status value
W_o	The weight of the output gate
b_o	The bias of the output gate
h^t	The output value of the current cell
s	Any given state
a	Any given action
s'	The state of any next moment
$\boldsymbol{\pi}$	Specify the policy
t	Any time
$\boldsymbol{V^{\pi}}$	For the status value function for policy π
$V^{\pi}(s)$	Under Policy A, the value function with a status of s
$Q^{\pi}(s,a)$	The reward expectation function that is obtained from the beginning of this state until the end
Q^{π}	The action value function under policy π
$E(Ri)$	The expected rate of return of a stock or investment fund
Rm	The market's desired rate of return
Rf	Market risk-free interest rates
$F(r)$	Transaction cost function

III. Intelligent Decision Mode

A. LSTM prediction model

A prediction model with good performance is inseparable from the selection of basic prediction framework. Based on previous studies, the basic prediction framework of LSTM is presented in this paper.As shown in Fig 2.

LSTM is a special variant of recurrent neural network. In THE RNN model, it often occurs that the previous saved information will be forgotten during the training of the following neurons. In response to this problem, LSTM deep neural network can record information for a long time.Therefore, RNN hidden layer neurons are replaced by a special structure. LSTM model was first proposed by Schmidhuber J and Hochreiter S in 1996 [1].

There is only one TANH module in RNN, LSTM specially establishes an independent module, It consists of three gates, input, output and forgetting gates control the information flow in the memory block and control the cell level of CEC inflow, store the information in the cell, and then corresponding output. The differences are shown in Fig 3 and Fig 4.

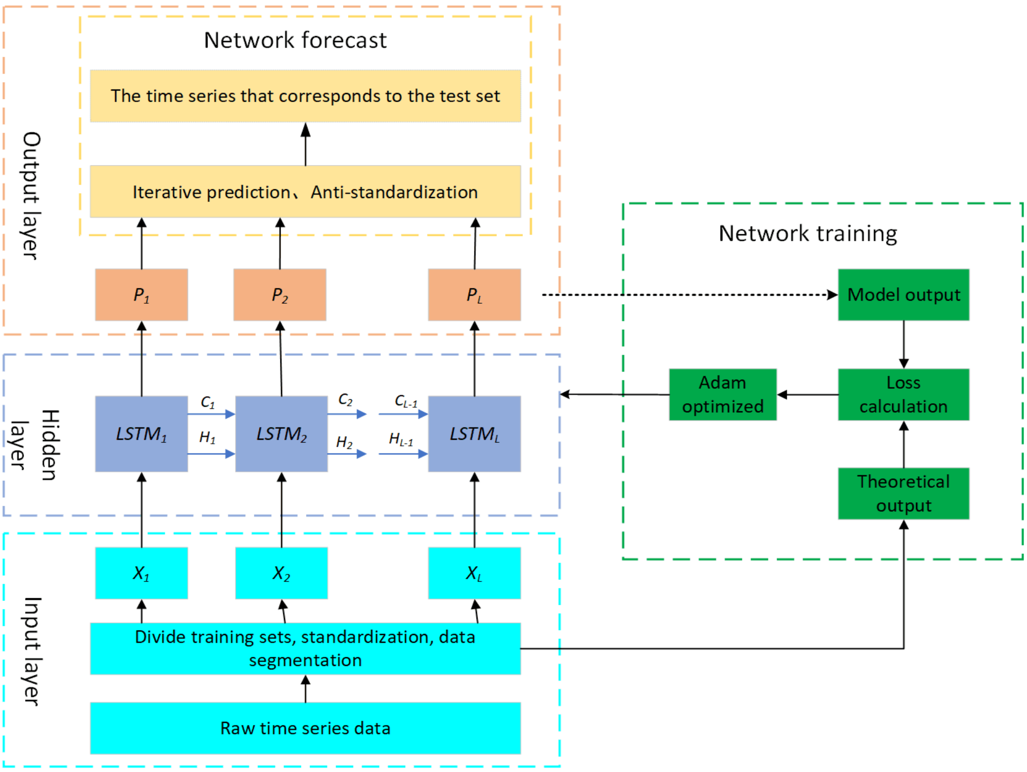

Fig. 2 LSTM prediction framework

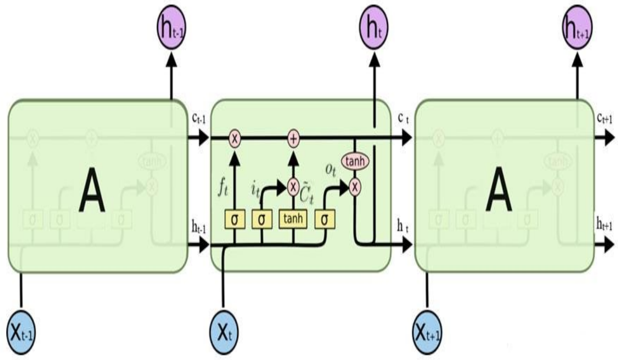

Fig.3 LSTM model diagram

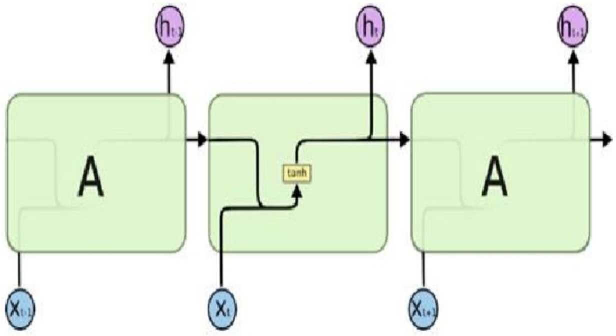

Fig.4 RNN model diagram

Forget the door: LSTM processes sequential data from left to right, retaining and discarding data through switch control processing.

$$f^{(t)} = \sigma\left(W_f\left[h^{(t-1)}, x^t\right] + b_f\right) \tag{1}$$

Enter the door: Choose to enter information into the gate and decide which tasks need to be updated and how much more

$$c^t = \sigma\left(W_c\left[h^{(t-1)}, x^t\right] + b_c\right) \tag{2}$$

Output the door: The information output is determined through the forgetting gate and the input gate and finally to the output gate.

$$o^t = \sigma\left(W_o\left[h^{(t-1)}, x^t\right] + b_o\right) \tag{3}$$

$$h^t = o^t * \tanh^{-1}(c^t) \tag{4}$$

Plugging N days of historical data into the model, the daily data is represented by a vector. For every X that goes through an input gate there's going to be a set of Y that goes out to form h, Finally, through the historical data of the previous N days, the price of the N+1 day is predicted. That is, only the price of the NTH day is taken as the final output, which will be converted in the neural network model. The closing price y on the NTH day is obtained as the output of the prediction model. The predictive performance of the model was tested by iterating 50, 75, and 100 times shown in Fig 5.

(a) The number of iterations is 50 (b) The number of iterations is 50

(c) The number of iterations is 100

Fig 5. Prediction graph of different iterations

As can be seen from Fig 5, The fitting ability of predicted data varies with the number of iterations. The larger the number of iterations, the better the model's ability to fit data. At the same time, data prediction is better at the turning point.

B. CPMA model

Since Markowitz[4] proposed the mean-variance model, People have been working on ways to expand and refine the selection model, To make models better suited to market conditions or better measure risk, As a trader who wants to maximize returns, Risk aversion.Based on this, the CPMA model is established.

$$E(Ri) = Rf + \beta(Rm - Rf) + F(r) \quad (5)$$

C. Reinforcement learning model

Reinforcement learning models[5] are not labeled supervised learning, Is rewarded by intelligence that improves through trial and error, And according to their own state and reward signals to learn a way, and then mature to make the best decision. The model consists of agents, environments, rewards and states See Fig 6.

Fig. 6 How reinforcement learning works

D. DQN and LSTM are solved

Establish the state transition matrix, given any state S and action A.

$$\mathcal{P}_{ss'}^{a} = \Pr\{s_{t+1} = s' \mid s_t = s, a_t = a\} \quad (6)$$

The state at the next moment corresponds to the expectation definition of the reward value function at the next moment.

$$\mathcal{R}_{ss'}^{a} = E\{r_{t+1} \mid s_t = s, a_t = a, s_{t+1} = s'\} \quad (7)$$

Define the rewards and expectations for the strategy.

$$V^{\pi}(s) = E_{\pi}\{R_t \mid s_t = s\} = E_{\pi}\{\textstyle\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s\} \quad (8)$$

Set up the expectation function of reward value from start to finish state.

$$Q^{\pi}(s,a) = E_{\pi}\{R_t \mid s_t = s, a_t = a\} = E_{\pi}\{\textstyle\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s, a_t = a\} \quad (9)$$

Take 6, 7, and 9 and substitute them into 8 to get behrman's equation. [2]

$$\begin{aligned} V^{\pi}(s) &= E_{\pi}\{R_t \mid s_t = s\} \\ &= E_{\pi}\left\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s\right\} \\ &= E_{\pi}\left\{r_{t+1} + \gamma \sum_{k=0}^{\infty} \gamma^k r_{t+k+2} \mid s_t = s\right\} \\ &= \sum_{a} \pi(s,a) \sum_{s'} P_{ss'}^{a}\left[R_{ss'}^{a} + \gamma E_{\pi}\left\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+2} \mid s_{t+1} = s'\right\}\right] \\ &= \sum_{a} \pi(s,a) \sum_{s'} P_{ss'}^{a}[R_{ss'}^{a} + \gamma V^{\pi}(s')] \end{aligned} \quad (10)$$

DQN and LSTM combined algorithm

Begin
Initialize the Qnet network
Collect experiences and add them to the memory pool
The network output of **LSTM** is taken as the state quantity of DQN
For i in range(epoch):
  if (random() > explore): # Decide whether to explore
  Select actions randomly
  else:
Select actions based on the Qnet network
Multi-agent trading
According to Bellman optimality principle

$$Q_M^*(s,a) = \mathbb{E}_{s' \sim P_{sa}}\left[R(s,a,s') + \gamma \max_{a' \in A} Q_M^*(s',a')\right]$$

Update the reward
q_eval = eval_net (s)
q_target =b_r + GAMMA * q_next
loss = loss_fun(q_next, q_eval)
Pair loss function $Q = \Sigma_{i=1}^{n} e_i^2 = \Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2$ executive gradient descent
END

The decision-making and evaluation of the DQN [6]lgorithm are shown in Table II.

TABLE II DQN DECISIONS

Q value	Decision making	Holding money
[7.8e-14,0.69,6.4e-18]	keep	3030.038
[1.89e-18,4.9e-22,1]	sell	3030.038
[7.72e-160.24,3.95e-16]	keep	3030.038
[5.91e-18,1,8.46e-22]	keep	3030.038
[6.47e-16,0.485,2.39e-16]	keep	3030.038
[7.27e-16,0.77,1e-15]	keep	3030.038
[9.64e-17,1,1.6e-22]	keep	3030.038
[6.79e-14,7.1e-2,9.53e-14][5e-2,7.2e-16,6.9e-14]	keep	3030.038
[2.59e-18,2.44e-19,1]	buy	3030.038
[3.58e-1,6.95e-8,4.26e-18]	sell	3108.077
[8.91e-16,7.52e-18,1]	buy	3108.077
[7.8e-14,0.69,6.4e-18]	sell	3218.221

The daily returns of gold and bitcoin in 5 years are calculated as shown in Fig 7.

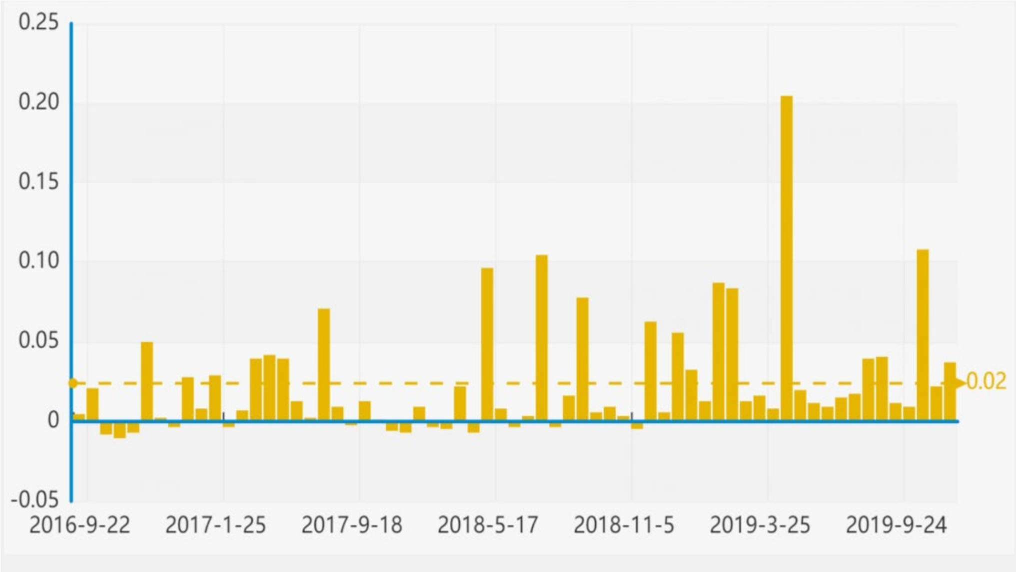

Fig.7 Daily returns

According to the decision model, traders' attitudes to gold, The frequency with which Bitcoin is bought, sold and held is shown in Fig 8.

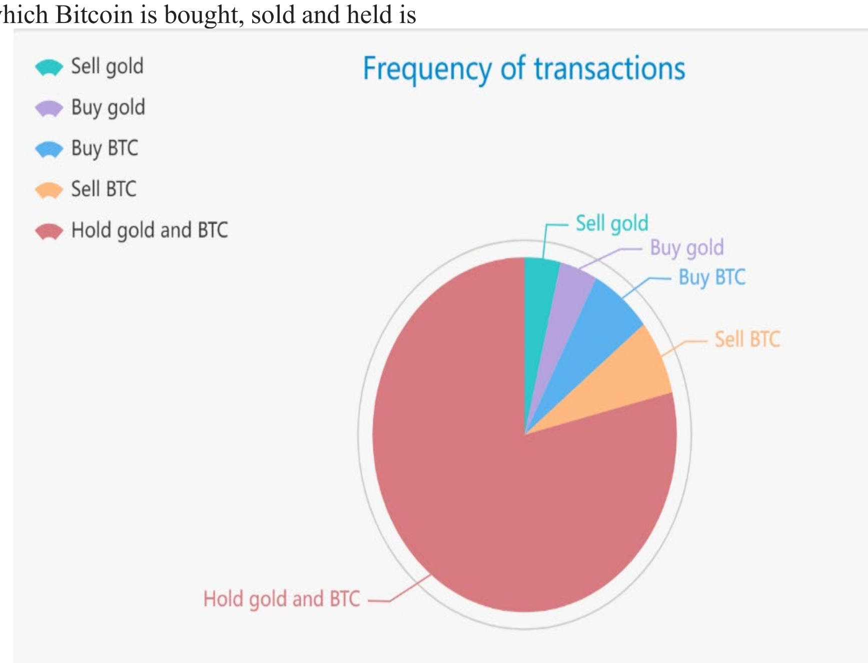

Fig. 8 Transaction frequency

As can be seen from the above table and pictures, This decision model has obvious changes in price. When prices go up, The model is biased towards buying and selling, The model is biased to hold when prices fall continuously. In five years, they bought gold 53 times, sold gold 60 times, bought bitcoin 127 times, bought bitcoin 127 times, The number of holds was 1,477. The original on September 10, 2021 A $1000 investment is worth $200,401.

IV. Model to improve

This problem studies the impact of transaction costs on our decision making, We can improve the CPMA model. The improved model is used to analyze the correlation between investment return rate and transaction cost. The change of return rate reflects the impact of transaction cost.

$$\mathrm{E(Ri)} = \mathrm{Rf} + \beta(\mathrm{Rm} - \mathrm{Rf}) + \mathrm{F}(r)(11)$$

Expected rate of return E(Ri) is positively correlated with transaction cost, Thus, the transaction price is negatively correlated with the transaction cost. Therefore, CPMA improved model can be used to study the sensitivity of transaction costs.

A. Descriptive statistics

a) Descriptive [7]analysis of the relationship between transaction cost adjustment and return rate.

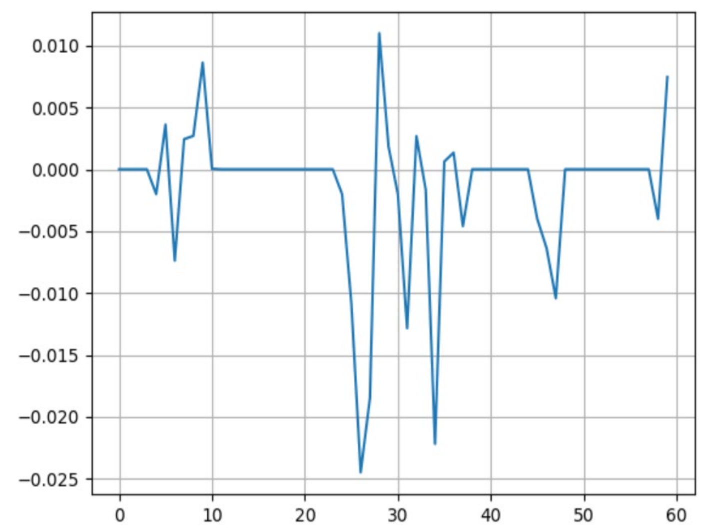

Fig. 9 Rising transaction costs

On September 18, 2016, the transaction cost tax rate increased from 0.001 to 0.004,Transaction cost tax rate increased from 0.001 to 0.004,Earnings fell sharply on the day, After turning negative, the index volatility increases, the sensitivity is strong, and the yield ratio decreases significantly.

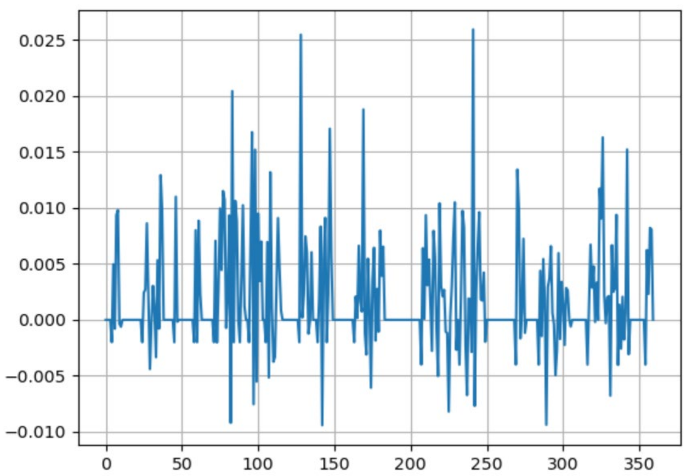

Fig. 10 Transaction costs fall

In June 2017, the tax rate on transaction fees was reduced from 0.004 to 0.002, The income rate fell and the index changed little.

b) Descriptive analysis of daily returns from 2016 to 2021

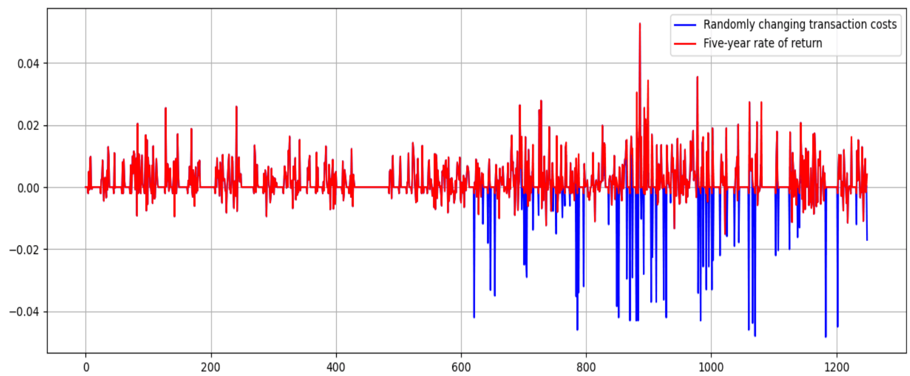

Fig. 11 Comparison of transaction costs over the past five years

Blue represents a five-year random change in transaction costs, In red, transaction costs are fixed and do not change, The year with transaction cost adjustment can be observed based on a),b).The sensitivity of yield varies widely, In years without transaction cost adjustment, the yield is more stable. In the short run, transaction cost adjustment has a great impact on the market in the short run, The long-term impact is relatively small.

B. Test and analysis of the impact of transaction cost on return rate

We use T-test to test the change of rate of return. The steps are as follows:

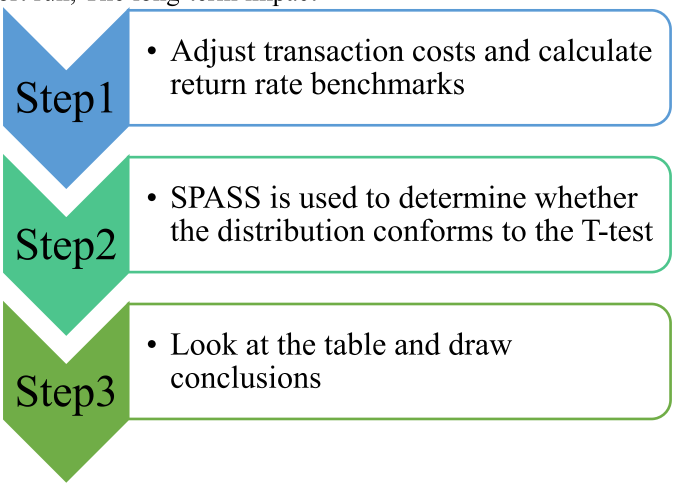

Fig 12. Steps

TABLE III T TEST

	Levene's Test for Equality of Variances			t-tesy for Equality of Means					
	F	Sig.	t	df	Sig.(2-tailed)	Means Difference	Std.Error Difference	95%Confidence Interval of the Difference	
								Lower	Upper
Days Equal variances assumed	.307	.581	.444	118	.658	.0003339837	.0007518397	-.001154864	.0018228310
Equal variances not assumed			.444	117.924	.658	.0003339837	.0007518397	-.001154874	.0018228409

As shown in the table III, Sig value of 0.581,Because 0.581 is greater than 0.05, It satisfies the homogeneity of variance, Then observe the corresponding Sig value, The value of Sig is 0.658,Greater than 0.05,This indicates that there are significant differences between the two groups of data, That is, you can see when transaction costs rise,交 Easy cost has significant change to yield rate, However, when the transaction cost is reduced, the impact on the yield rate is not obvious.

V. CONCLUSION

Strengths:

• Combining DQN algorithm with LSTM, it solves the problem that DQN can not be applied when used alone, so that it can adapt to the characteristics of market transaction decision.

• Compared with the traditional prediction model, the deep reinforcement learning model does not directly predict the price, but takes the decision of buying and selling as the output of the model, which has higher guiding significance for investors.

• Compared with the general neural network model, LSTM model can learn the rules of long-term dependence more effectively and improve the accuracy of prediction.

Weaknesses:

• Due to limited data, only technical analysis is considered, and the conclusions of the model are based on data results only. The fundamental information such as news policy, market competition and other factors are not considered to have certain limitations.

• The field of combining reinforcement learning with market buying and selling decision making is relatively new, and the understanding of the field of reform is still limited.

Prospect: With the development of artificial intelligence, AI is increasingly being used in economics and finance. Inspired by Alphago's game theory, this model adopts deep reinforcement learning. In order to realize the trading decision of agent without human participation. Therefore, this model has a good theoretical support and has a good guiding significance for traders.

ACKNOWLEDGEMENT

Supported by: Anhui Provincial Teaching Reform Research Project (2020JYXM0443, 2020JYXM0441), Anhui Provincial Ideological and Political Work Innovation Project (2020ZDXSJG094) natural Science Foundation of Anhui Province (2008085QD178); Excellent Talents Support Program of Anhui Universities (No. GxyqZD2020020)

REFERENCES

[1] Reitgruber W , Sterlina I . On the Forecastability of Share Prices on the Viennese Stock Exchange[J]. Empirical Economics, 1995, 20(3):415-433.

[2] M, Hashem, Pesaran, et al. A Recursive Modelling Approach to Predicting UK Stock Returns[J]. Economic Journal, 2000.

[3] Gudelek M U , Boluk S A , Ozbayoglu A M . A deep learning based stock trading model with 2-D CNN trend detection[C]// 2017 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, 2017.

[4] Chen K , Zhou Y , F Dai. A LSTM-based method for stock returns prediction: A case study of China stock market[C]// IEEE International Conference on Big Data. IEEE, 2015:2823-2824.

[5] Zubeldia A M , Zabalza L , Zubiaurre M Z . The Markowitz model for portfolio selection[J]. Cuadernos De Gestión, 2002, 2(1):33-46.

[6] Park, Frank, C, et al. Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies[J]. Expert Systems with Application, 2017.

[7] Parkhe A . Strategic Alliance Structuring: A Game Theoretic and Transaction Cost Examination of Interfirm Cooperation[J]. Academy of Management Journal, 1993, 36(4):794-829.