

Prediction of Stock Price Based on LSTM Model

Yangtian Yan

Wuhan university, Wuhan 430072, China
Email: 869702656@qq.com

ABSTRACT

With the development of the stock market, listing has become one of the best ways for companies to develop. These companies are able to raise money in the equity market, while different types of investors come with an expectation of benefiting from potential stocks. The evidence suggests that this is the case stock investment is risky as well as large-duty, so investing wisely remains to be obviously crucial. As progress of artificial intelligence technology moves far ahead, it enables to introduce study about artificial intelligence into stock detecting. In this paper, combined with the actual stock market, a wealth of experiments are carried out on the relevant data sets. First of all, this paper collects the daily market data of China's A-share market. The data are preprocessed and the correlation coefficient characteristics of technical indicators are extracted. After the feature engineering processing of the stock data set, the relevant algorithm model is built, and a detailed comparative experiment is done. The prediction process of stock is a relatively complex process. The prediction based on LSTM model can be used as a reference for the factors of stock index and market analysis, and cannot accurately predict the trend.

Keywords: *LSTM, stock, stock index, doe theory, AI, listed company*

1. INTRODUCTION

Prediction of stock price is through the statistical analysis of the historical stock market from the mathematical point of view, the corresponding statistical law is obtained, and then the stock price is predicted according to the statistical law. However, the deeper the research on the stock trend prediction and the continuous improvement of the market, the more obvious the shortcomings of the traditional financial model are found, among which the most fatal- -The point is that the stock trend prediction of traditional basic financial model is completely based on the category of mathematical statistics, and the statistical model, to a certain extent, involves hypothesis testing, and the hypothesis test has a direct and important impact, that is, stock trend prediction. the existing stock forecasting research is mainly about a single point, such as specialized research on historical trading data, or related technical indicators of the stock (such as MACD, Ma index, etc.), or focus on the news of listed companies themselves), so as to make a judgment on the stock fluctuation.

Because of this, this paper intends to combine the previous experimental content, and carry out a broader research. First of all, the basic indicators of the stock market are introduced, and then combined with the

LSTM model, through data acquisition and data normalization processing, and prediction process, the stock price is predicted hoping to make a beneficial exploration on the research direction of stock trend prediction.

1.1. Introduction and Analysis of LSTM

To study long-term patterns, particular RNNs, also widely known as "LSTM"- Long term and short-term memory networks - are used. In 1997, Hochreiter schmiduber put forward the conception without precedent. Later it is improved and popularized by lots of professors. They are nicely applied to many problems, and LSTM is obviously designed to avoid the problem of long-term dependence. [1].

1.1.1. The difference between RNN (Recurrent Neural Networ) and LSTM

As can be seen from figure 1 and figure 2, the LSTM structure shows to be a lot more complicated. In RNN, two inputs both with single output are included. the past output together with the current input are connected end to end or in tandem, and the output of which is ruled by tanh. Only the newest status will be take into account [2].

There is an increase of LSTM that an input and an output are added to RNN, which is an crucial way in the process. In fact, the entire LSTM could be told as below in chronological order:

- 1) Which cell states should be forgotten
- 2) What new states should be added

3) What should the output be based on the current state and current input

In fact, LSTM is based on RNN, filtering the past state, so it is easy to select the state that makes a huge difference on the current state, instead of straightly selecting the latest state.

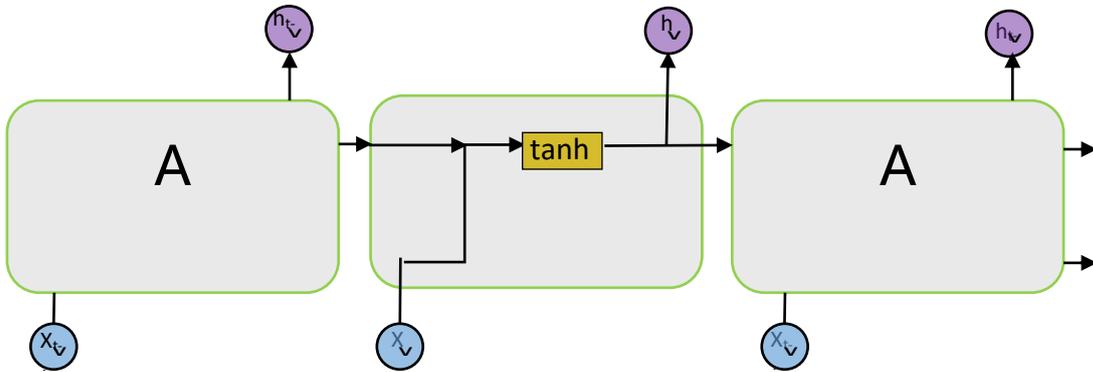


Figure 1 . RNN

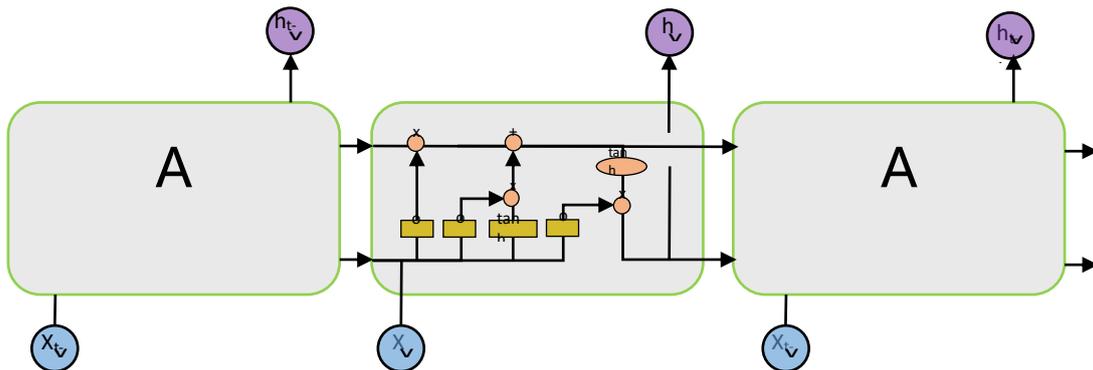


Figure2. Lstm

1.1.2. Structure of LSTM

1.1.2.1. Forget gate

Target: cell state

Function: selectively forgetting the information in cell state

Let us return to an example of a language model that predicts the following word on account of what we see. In the circumstances, the unit status possibly conclude the category of the latest topic, so you can choose the correct pronoun. We always want to forget the previous one because we meet a new subject.

Such as, my grandma was not at home yesterday, so I . . . When mentioning me, people may choose to forget about my grandma [3].

1.1.2.2. Input layer gate

Target: cell state

Function: selectively record new information into cell state

In our language model example, in order to replace the old topic that needs to be forgotten, we need to add a new topic category to the cell state.

He has something to do today, so I... When the word "I" is processed, the subject I is updated into the cell.

1.1.2.3. Output layer gate:

Target : Hidden layer HT

In the case of this language model, because he sees pronouns, he may need to output information related to verbs. If we want to know the singular form of a verb, then we should also know that if the verb is singular, we can do the same.

For example: in the above example, when the word "I" is processed, it can be predicted that the next word is a verb, which is more likely to be the first person.

It will save the previous information to the hidden layer.

1.2. Concepts related to stocks

1.2.1. Influencing factors of stock price

1.2.1.1. Basic factors

Basic factors usually include economic factors, political factors, human factors, etc. that do not include the stock market. The basic factors are very important, because a little change in it will cause drastic changes in the stock market price. [4].

For investors, they invest for profit, so their most concern is whether the stocks they buy in the market can be sold at a higher price. Therefore, the purpose of investors and professional scholars and analysts who study the stock market is often to study the stock and predict its price trend through various methods. Some people have evolved from the relationship between supply and demand in the market to the technical analysis method of stock price, while others have formed stocks by analyzing the relationship between the factors that affect the stock price.

The study of its internal value mainly refers to the analysis of its internal value. The stock price is not static, and many factors such as politics, economy, and culture can affect its trend.

We can see that the stock price on the market is not necessarily the same as its real price, which is also one of the beauty of stock trading. The price of stocks may continue to rise to an outrageous price due to false advertising and popular with the masses. The stock price may also be affected by other factors, such as false reports, the price will deviate from the value, and even the listed company will go bankrupt due to the crazy dumping of stocks by investors, which will inevitably lead to stock market chaos and even crisis. If a stock is found to be overvalued in the market, it must be cashed out; if the valuation of another stock is too low, it will definitely cause crowds to buy it. These three points must be decisive for the value of stocks: First, whether the national economic environment is good. Regardless of prosperity, the status of various economic sectors such as industry and agriculture, commerce, transportation, public utilities and finance. Third, the

operating conditions of the issuing enterprise. If it is managed and profitable, the stock price will be high and the actual value will be high. Otherwise, the market price will be very low.

1.2.1.2. Political factors

We can usually know that the price of stocks will fluctuate with government activities and regulations. This includes national economic policies and regional development plans, as well as related laws and regulations and future administrative measures. It is also more and more sensitive to changes in stock prices

(1) The impact of war. The war caused global turmoil and consumed a lot of military supplies. In people's eyes, war means stock prices fall, but the impact of war on stock prices in different industries is different. For example, based on the effects of wars in the Middle East before, local wars will bring about large-scale fluctuations in commodity prices, that is, oil and gold prices will rise, and the US dollar will also rise to a certain extent. Therefore, investors should promptly buy stocks in oil and related industries, and sell stocks that are vulnerable to war. The rising stage of commodities such as crude oil is generally concentrated before the outbreak of the war and the early stage of the war, which means that war expectations have a greater leading role in commodity prices. In the middle and late stages of the war, the risk of price decline will become greater and greater, and speculative funds are very high. It is possible to hit the market with all its strength at some point in the later period, triggering a sharp drop in prices.

(2) Changes in the international situation. For example, improving diplomatic relations will raise the share prices of the multinational companies concerned. When diplomatic relations improve, investors should seize the opportunity to buy shares of relevant multinational companies.

(3) Major domestic political events, such as political turmoil, will also have a significant impact on the stock market. In other words, it will affect the psychology of stock investors, thus indirectly affecting the level of stock prices.

(4) Main national economic policies, such as industrial policy, taxation policy, monetary policy, etc. The establishment of a stock market in my country is itself an attempt by the country to deepen reform and opening up and actively explore. Every step of the development of the stock market is carried out under the guidance of national policies. The state's attitude towards the stock market is the fundamental factor in determining the development prospects of the stock market. The stock prices of industries supported and developed by the state will be pushed up, and the stock prices of industries restricted by the state will be

adversely affected. For example, politics limits the prices of social and public goods and services, including transportation, natural gas, water, and electricity. This will directly affect the profitability of public utilities and cause the stock price of public utility companies to fall. These political factors will have an impact on the stock market itself, that is, through the company's profits and market interest rates, the stock price changes. Policy changes aimed at the stock market generally bring about substantial changes in the stock market. If there are signs that the country may introduce a bad policy, it should be avoided as soon as possible. Short positions are waiting for opportunities and never take a fluke. With rapid fluctuations and with the increase in volume and energy, you can quickly move in and out to buy leading stocks to maximize returns.

1.2.1.3. The impact of the nature of the company's industry on the stock price

The growth of the company is subject to the restriction of the industry and the rise and fall of the industry. For example, if the enterprise belongs to the electronics, fine chemical industry and growth industry, it has a good development prospect and is attractive to investors; on the contrary, enterprises belonging to coal, cotton textile and sunset industries have poor development prospects and relatively low investment income. Therefore, the nature of the company's industry has a great impact on the stock price, so we must analyze the nature of the company's industry. The concrete analysis should be carried out from the following three aspects

(1) To know whether a company's product is a production resource or a consumption resource, we need to observe and study the characteristics of the product. Production resources are to meet the needs of the original products required by the production of other products, and the latter directly meet people's consumer needs. The two types are interdependent and mutually restrictive, and require proportional and coordinated development; there should be a balance between the supply and demand of production materials and the supply and demand of consumption materials. The oversupply of some means of production and consumer goods affects the balance of proportions within and between the two major sectors, and affects the smooth progress of reproduction in the entire society. Therefore, when the economic environment changes, the two will be affected differently. Generally speaking, the impact of economic changes on production resources is greater than the impact on consumption resources, that is, during economic development, the growth rate of production resources is faster than the growth rate of consumption resources. On the contrary, the production of production resources shrinks faster. In terms of consumption resources, we should also analyze whether

the company's products are necessities or luxury goods, because the price elasticity of different product attributes is different. People have little flexibility for necessities, but very much for luxury goods.

(2) Analyze the sales target and sales scope of the company's products from the demand table. If the company's products are mainly sold at home and abroad, domestic sales are vulnerable to domestic political and economic factors, and export sales are vulnerable to international economic and trade climate. Different demand objects have different requirements for the performance and quality of goods. The company determines the output according to the demand. Otherwise, it is bound to affect the company's product sales, thus affecting the level of profit, reducing investment income and stock price.

(3) From the production form, the company is labor-intensive, capital intensive or knowledge-based technology intensive. Technology is a labor-intensive factor, while technological capital is a labor-intensive factor. In underdeveloped countries or regions, labor-intensive enterprises account for a large proportion, while in developed countries or regions, capital intensive enterprises tend to dominate. In addition, different types of companies, labor productivity and competitiveness are different, will also affect the level of product sales and profitability, resulting in differences in investment returns.

1.2.2. Stock related technical indicators

1.2.2.1. KDJ Index

KDJ index, also known as random index, is a new practical technical analysis index. It is first used to analyze the futures market, and then widely used to analyze the short-term trend of the stock market. It is the most commonly used technical analysis tool in futures and stock markets [5].

In the statistical system of stock analysis, the random index KDJ is the most commonly used. According to the principle of statistics, according to the highest price and lowest price of a certain period (usually 9 days, 9 weeks, etc.) and the closing price of the last calculation cycle, the immature random value RSV of the last calculation cycle is calculated, and then the square of smooth moving average K value is used to calculate D and draw a curve chart to study and judge the stock price trend.

Through the highest price, the lowest price, the closing price at the end of the cycle, and the proportion between them. Then, the immature random value (RSV) of the last day of the cycle is calculated and calculated according to the moving average method

Calculate the value of K, D and J:

$$RSV = (C_n - L_n) / (H_n - L_n) * 100 \quad (1)$$

The k-value is the n-day moving average of RSV. The K-line connected by k-value is also called fast line. The velocity of change in the curve is moderate.

D value is the moving average value of K value in n days, and the change speed of D line is the most among the three lines Slow, known as the slow line.

J value changes the fastest, as an auxiliary observation K line and D line issued by the sale signal, known as an express line or confirmation line. Three lines show that KDJ can reflect the price fluctuation trend in the same coordinate indicators, according to the curve can be more intuitive to buy and sell the timing of research and judgment.

$$K = \frac{2}{3}K_p + \frac{1}{3}RSV$$

$$D = \frac{2}{3}D_p + \frac{1}{3}K$$

$$J = 3 * K - 2 * D \quad (2)$$

Where C_n is the closing price on the nth day, L_n is the lowest price within n days, and H_n is the highest

price within n days. The values of days cycle n are generally 5, 9, 19, 36, 45, 73, etc. In the short and medium-term stock trend analysis, in order to make the curve response is more rapid and accurate, and the parameter n is 9. K_p, D_p. Is the value of K and D of the previous day, if there is no previous value a day is replaced by 50.

KDJ index measures the overbought and oversold phenomenon through the value and mutual state of three lines, and then shows the market Status and send a buy sell signal. In the technical analysis of the stock market, overbooking and oversold is a kind of analysis of the general trend of rise and fall Powerful tools. The phenomenon of overbought and oversold is that investors overreact to some news and the stock price is excessive.

Fluctuation, in essence, is the embodiment of inertial factors of price trend. After inertia, the effect is one.

The price will fall after overbooking and rebound after oversold. Grasp overbought.

The movement law of oversold can effectively help to analyze the stock law.



Figure 3. An example of KDJ indicator

1.2.2.2 MACD index (Moving average convergence and divergence)

MACD should first calculate the fast moving average (generally 12 days) and the slow moving average (generally 26 days). Take these two values as measurement 2

The basis of "difference" between fast line and slow line. The so-called "difference" (DIF) is the 12 day moving average minus the 26 day moving average [5]. So, in a sustained rise,

The 12-day moving average is at the 26 day moving average. The positive difference (+ DIF) will increase.

On the contrary, in a downward trend, the difference may become negative (- DIF), i.e. absolute value

The bigger you come. As for the turning of the market, the positive and negative deviation should be reduced to a certain extent, which is the real signal of market reversal. The reversal signal of MACD is defined as "differential"

9-day moving average (9-day DIF).

$$EMA_{(n)} = \frac{n-1}{n+1} * PEMA_{(n)} + \frac{2}{n+1} * C$$

$$DIF = EMA_{(12)} - EMA_{(26)}$$

$$DEA = \frac{n-1}{n+1} * PDEA_n + DIF \quad (3)$$

$$MACD = (DIF - DEA) * 2 \quad (4)$$

MACD index is mainly through the moving average (EMA), deviation (DIF) and the 9-day shift of dispersion

The relationship between the dynamic average (DEA) is dominated by the deviation, and the 9-day moving average of the deviation is 0.

Auxiliary analysis and judgment. The difference is equal to the short-term EMA value minus the long-term EMA value, which lasts for a short period of time. The EMA on the 12th is higher than that on the 26th, and the positive difference between them will become larger.

On the contrary, it is falling. In the potential, the difference will decrease or even be negative. After the trend of -- wave rising or falling, when the deviation is absolute. When the value is reduced to a certain extent, it is the signal of market reversal. In practical application, DIF line protrudes upward. The crossing point of breaking DEA line is "golden cross", which is the buying signal. The intersection of dif line and DEA line is

"Dead cross" is a sell signal. In addition, the red and green histogram drawn according to MACD value can also be intuitively displayed.



Figure 4. The red and green histogram

1.3. Prediction Process

1.3.1. Data acquisition

Before training the neural network model and forecasting, we need to collect data first. Stock data could be obtained through

Tiger finance and economics, sina finance and economics, Dongfang fortune net, tonghuashun and other large financial websites export. Number of training to prepare the model

According to the set, the closing price of the stock will be used for prediction and model evaluation [6].

1.3.2. Data normalization processing

For the convenience of calculation, the maximum (max) and minimum (min) of the opening price obtained in this paper are in the whole range

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (5)$$

From the sample. The real opening price and normalized opening price of Yanghe shares (002304).

Describe. Yanghe shares (002304) real daily opening price chart is as follows, in which the abscissa time span.

From 20150101 to 20181231, a total of 974 trading days, the vertical axis is the opening price, Yanghe can be observed

The opening price of the shares reached the maximum value in 2018, exceeding 140 yuan [7].

1.3.3. Data categories

Table 1. Classification of Data name

Input variables	Variable name	Variable symbol
X (t-1 trading day)	Opening price	open
	Closing price	close
	Opening price	open
	Minimum price	low
	Up and down range	change
	Amplitude	amplitude
	Turnover rate	turnover rate
	Search volume	search volume
Y (t trading day)	Opening price	open

1.3.4. Predication

The training set of the opening price of Yanghe shares from 2015 to 2017, and the long-term and long-term memory model can be used to predict the opening price of Yanghe shares in 2018, that is, the prediction calling the predict function will be returned in the format provided by the output layer of the network. Because in the process of data processing, we normalize the data, so we need to invert the scaling ratio when we combine the prediction with the test data set. The model converts the test data set and the expected stock market data to the original data. The error score of the model can be calculated by the predicted value and the actual value. The error of the opening price of Yanghe shares is 0.0159%. In this case, the root mean square error (RMSE) of the error given by the same unit as the variable itself can be obtained, and the final RMSE of the model on the test data set can be obtained. After

calculation, the final root mean square error (RMSE) of the prediction series of the model is 1.766. In order to see the degree of model fitting and prediction, the training test fit diagram is drawn. As shown in the figure below, the horizontal axis is still 20150101-20181231. The trading days of these four years are shown in the form of data, and the vertical axis is the daily opening price of Yanghe shares. We predict the opening price of t day by taking 8 characteristics of T-1 as variables. The blue line throughout is the actual daily opening price, the recording line is the opening price predicted by training set from 2015 to 2017, and the red line is the predicted opening price of test set of Yanghe shares in 2018. It can be seen from the chart that the predicted opening price is consistent with the real opening price trend, indicating that our model is basically successful. However, there is still a large deviation between the predicted value and the real opening price

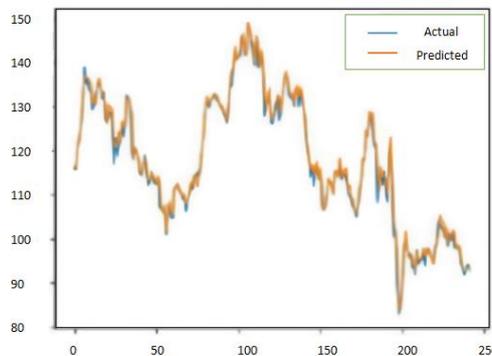


Figure 5. Prediction chart of opening price in 2018

After sifting out the forecast chart in 2018 for further observation, the predicted opening price of Yanghe shares is basically less than the actual opening price in terms of maximum and minimum value, which shows that there are obvious prominent parts at the tip of the blue line, and there are 45 times. This shows that the open price forecasting model based on the long-term and short-term memory model LSTM is conservative, and the real stock market has more ups and downs than the predicted value.

(the prediction of opening price based on the model)

2. CONCLUSION

Stock trend prediction is actually a complex project integrating finance and computer science. However, there are still many problems in the existing research content and algorithm model can affect the stock trend of all aspects. What's more, the problem of stock trend prediction is very serious. The method is also different. Based on the comparison of a large number of experiments, a new algorithm model is proposed and its effectiveness is verified. However, there are still some deficiencies in the research level There are two reasons. First of all, there are a lot of random emergencies in the stock market (no matter which country or nature), and the impact of emergencies on the stock market is highly random and unpredictable, that is, the so-called noise is unimaginable. Secondly, continuous variables as a prediction target is a bad design, because it will make the prediction space too large, resulting in infinite search space. This view comes from reinforcement learning. One of the technical points of reinforcement learning is to limit the prediction space. Even if the objective world is continuous and infinite, we need to use the technology similar to tile coding to make it discrete and limited.

ACKNOWLEDGMENT

When the essay was about to end, I looked at the paper full of words and talked. The dreams at the beginning, the tempering experience and the happiness that I had are all in the end with a light smile, becoming a passage in the long journey of life. In this experience, the most profound influence on me is the teachers who have verified my growth and the brothers and sisters who accompanied me when I grew up. Special thanks to my mentor for his selfless help in my study and life. The mentor's profound knowledge, dialectical thinking, innovative consciousness and strict self-discipline character set me a successful example and also become my goal in the future. Here, I sincerely wish that my teacher will stay here, and peaches and plums will be full of fragrance. Alma mater is getting better and better.

REFERENCES

- [1] J. Zhang(2019)Introduction to LSTM model. <https://blog.csdn.net/yu444/article/details/86764352>
- [2] Comparison of LSTM and RNN, lanmengyiyu, 2018, <https://blog.csdn.net/lanmengyiyu/article/details/79941486>
- [3] Detailed explanation of LSTM working principle and mechanism, HadoopDev, 2019, <https://blog.csdn.net/hadoopdevelop/article/details/88964175>
- [4] Y. Jiang, The main factors affecting the stock price, 2018 <https://www.zhihu.com/question/29362860>
- [5] X. Wang. Research on stock prediction based on improved LSTM model[D]. Chongqing University of Technology, 2020.
- [6] R. Liu. Research on intelligent optimization algorithm and its application in the financial field[D]. Jilin University, 2020..
- [7] M. Cui. Quantitative analysis of stock market based on long and short-term memory network[D]. Yunnan University of Finance and Economics, 2019.