2020 7th International Conference on Information Science and Control Engineering (ICISCE) | 978-1-7281-6406-9/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICISCE50968.2020.00026

A new stock price prediction model based on improved KNN

Qian Yunneng
Northwest University
Xi'an,China
Qianyn20000414@163.com

***Abstract*—With the rapid development of economy, how to accurately analyze and forecast the stock price has become a key problem. There are many methods and models about stock price prediction, such as KNN algorithm. However, the traditional KNN algorithm only uses the data of latest one day to predict the change trend of the next day, which is of little significance for reference. Therefore, this paper proposes an improved KNN algorithm, that is, the share price information group of the first N days is synthesized into a sample, which is input to the KNN model for learning. Experiments show that the improved KNN algorithm has better predictive performance than the traditional KNN algorithm.**

Keywords-KNN;stock price prediction;machine learning

I. INTRODUCTION

In the world economy, the stock market plays an important role. Through the issuance of stocks, a large amount of capital flows into the stock market, which promotes the concentration of capital, improves the organic composition of enterprise capital, and greatly accelerates the development of commodity economy. On the other hand, through the circulation of stocks, small amounts of funds are collected, which accelerates the concentration and accumulation of capital. In order to better regulate the development direction of the stock market, it is also very important to timely and accurately predict the trend of stock market based on historical data and current status, which will help the governments and investors to make correct investment decisions.

To correctly predict the trend of stock market, many algorithms and models about stock price forecasting have been proposed. Guo Jianfeng, Li Yu and An dong [1] proposed an improve BP neural network with LM algorithm, and then introduced traditional algorithm to optimize the initial weight and threshold of BP neural network. In [1], lM-BP and GA-LM-BP models are used to predict the transaction price of CSI 300 futures. The experimental results show that the LM-BP model optimized by genetic algorithm is superior to the LM-BP model in iteration time and number. The convergence speed of GA-LM-BP is accelerated, which overcomes the shortcoming of easily falling into the local optimal solution, and the feasibility of GA-LM-BP is higher in the specific stock price forecast analysis.

Sun Chen, Li Yang, Li Xiaoge and Yu Jiaoyan [2] proposed an optimized neural network (CS-BP) to predict the stock market, aiming at the low accuracy of the current intelligent algorithm for the stock market prediction. The results are compared with those of neural network model (PSO-BP) optimizd by Particle swarm and neural network model(GA-BP) optimized by genetic algorithm. Based on the back-test and analysis of the closing price data of SZ300091 (Jintongling) day line, it can be seen that the optimized neural network model proposed in [2] is obviously superior to these two algorithms, and can effectively predict the stock market, with a prediction accuracy of about 98.633% for 30 days.

Xu chenmeng and Fang hua [3] selected the daily opening price of stock SH601398(Gongshangyinhang) (February 14, 2018 solstice February 14, 2019), conducted first-order difference to stabilize the data, and then used ARMA model to predict the future prices (February 15, 2019 solstice February 19, 2019).After comparing the forecast result with the real value, it is found that the forecast result is more accurate and the error is smaller, indicating that ARMA model is suitable for short-term stock price forecast, which further confirms the effectiveness of time series model in finance.

The BCC-ESN model [4] constructed by Zhang Ningzhi, Zhou Jiali and Sun Wujun uses the colony drug algorithm of bacteria (BCC) to optimize the weight structure of echo state network(ESN), which inherits the excellent properties of ESN and has a higher model prediction ability. Experiments show that BCC-ESN model has better learning and prediction ability than feedforward neural network. The BCC-ESN model is more accurate than the other three networks by comparing the short-term price forecast of Shanghai Stock Exchange index with BP network, Elman network and ESN network. At the same time, in terms of operational efficiency, BCC-ESN model inherits the operational advantages of ESN and is obviously superior to other neural network prediction models. It is a feasible and efficient prediction algorithm, and especially has extensive practical value in stock time series prediction. Aiming at the problems encountered in the training prediction of BCC-ESN model, such as the problems of too long time consuming and over fitting, this paper also provides simple and feasible ideas and methods.

978-1-7281-6406-9/20/$31.00 ©2020 IEEE
DOI 10.1109/ICISCE50968.2020.00026

For the purpose of information fusion, Rao Dongning, Deng Fudong and Jiang Zhihua [5] tried to use three information sources, namely basic economic characteristics, technical indicators and online public opinions, to forecast stock prices simultaneously. In [5], the scheme is as following: First, different types of information source data are processed specifically to form a unified data set, and then the SVM classifier is used to establish a prediction model. The experimental results show that when linear kernel function is selected and non-trading day data is considered, the prediction effect of the three sources combination is better than that of single source or pairwise combination. In addition, when collecting data, it was found that although there was no trading on non-trading days (such as weekends or suspension periods), online public opinions surged. Therefore, the non-trading day public sentiment and emotional data were added to the experimental data to improve the classification accuracy. The results show that although it is difficult to predict the stock price based on multi-source fusion, it has a better prediction effect after proper selection of features and targeted data preprocessing.

Yang Zhenjian, Xia kewen [6] presented an ipos price forecast model based on gaussian process regression to improve the prediction accuracy of stock price, through the effect of the extract of ipos price index, with its training the historical data of the NASDAQ (NASDAQ) price of ipos, with particle swarm algorithm to optimize of gaussian process parameters to predict the price of ipos. Taking the listed stocks of 8 companies as examples, the prediction results show that the Gaussian process regression method improves the accuracy of stock price prediction and can be effectively applied to the new stock price prediction.

Additionally, mechine learning and deep learning algorithms are all adopted in stock price prediction [7], for example, ARMA model [8], neural network [9], Support Vector Regression [10], multivariate linear regression [11], deep neural network, etc. Among them, the most commonly used one is to learn and model the historical data of stocks based on the deep neural network [12] (represented by LSTM), and then to predict the change trend of stock prices. While this kind of method requires massive historical data to model the stock price, which requires a large amount of computation and cannot meet the need of real-time stock trading decision.

In addition, the stock market trading is a complex game process. Each participant in the stock market can obtain historical data of the stock and make predictions about the future price trend of the stock. The stock market is a typical zero-sum game, and there is no win-win situation for all stock holders. In view of this characteristic, each participant will adopt the corresponding trading strategy to influence the stock price, so as to make the stock price rise and fall to the most beneficial direction. In other words, when using long-term historical data to predict the stock price, the stock price has become unpredictable due to the influence of game factors.

In order to solve the time-consuming problem of conventional methods, this paper proposed an improved KNN model for stock price prediction. KNN [13~16] algorithm has many advantages, such as being easy to implement and good generalization ability. Therefore, KNN algorithm can meet the real-time and lightweight requirement of stock trading decision through proper optimization.

II. THE IMPROVED KNN

Traditional KNN algorithm takes the characteristic of the single data object as input, namely forecast the change trend of the day after simply by day data, and when the elected sample data is very close (that is, the test data and sample data are very similar), the second day of the change may have different trends such as up, down, and the same, when the result of voting of K most similar samples may have lost its meaning.

Since the traditional KNN method only predicts the change trend of the following day with the data of latest one day as shown in Figure 1, and it has no general reference significance.

1	Characteristic attribute 1
2	Characteristic attribute 2
3	Characteristic attribute 3
4	Characteristic attribute 4
5	Characteristic attribute 5
6	Characteristic attribute 6
7	Characteristic attribute 7
8	Characteristic attribute 8
...	.....

Figure 1. Traditional KNN input.

To solve the shortcoming of traditional KNN, an improved KNN was proposed. The proposed scheme is based on the assumption that the stock price of the following day is not only determined by the stock price of previous one day, but also depends on the stock price of the previous N days, and the overall trend of the change of the stock price of the first N days can better reflect the trend of the next day.

Based on this assumption, the stock price information group of the first N days is synthesized into a sample and input into KNN model for learning.

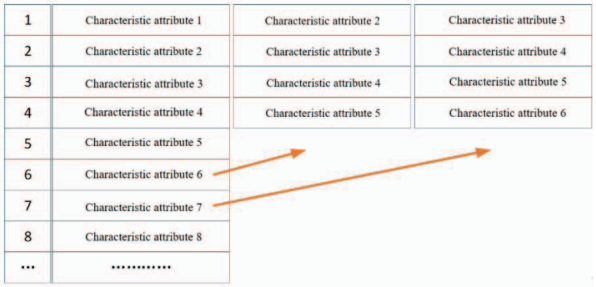

Figure 2. Improved KNN input.

As shown in the Figure 2, in the improved KNN-based stock price forecast, for each day's stock price information, the stock price information of the preceding N days is supplemented as the stock price information of the extended day, and the extended information is applied to the stock price forecast of the following day.

This method takes into account the stock price trend over a previous period (N days) rather than an isolated point and should theoretically have better predictive performance. The differece between traditional KNN and the proposed improved KNN is shown in Figure 3 and Figure 4.

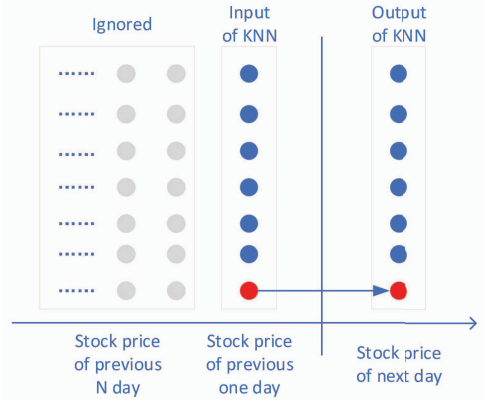

Figure 3. The prediction mechanism of traditional KNN.

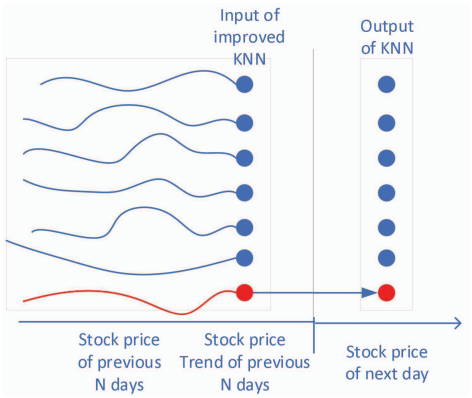

Figure 4. The prediction mechanism of improved KNN.

III. EXPERIMENT

To verify the performance of the improved KNN, an experiment was performed on improved KNN and traditional KNN. The historical price data of stock Neimengyiji from March 2004 to June 2020 (stock number is SH600967) is used in the experiment. The data includes the closing price, opening price, the highest price and the lowest price of. The historical price data of stock Neimengyiji is as shown in Figure 5.

date	code	name	closing price	highest price	lowest price	opening price
2020/6/3	'600967	Neimengyiji	10.02	10.1	9.93	10.06
2020/6/2	'600967	Neimengyiji	10.03	10.24	10.01	10.13
2020/6/1	'600967	Neimengyiji	10.1	10.12	9.92	9.92
2020/5/8	'600967	Neimengyiji	9.85	9.94	9.8	9.81
2020/5/7	'600967	Neimengyiji	9.8	9.88	9.72	9.85
2020/5/6	'600967	Neimengyiji	9.85	9.85	9.49	9.49
2020/4/9	'600967	Neimengyiji	9.56	9.65	9.5	9.56
2020/4/8	'600967	Neimengyiji	9.55	9.64	9.22	9.29
2020/4/7	'600967	Neimengyiji	9.32	9.37	9.18	9.18

Figure 5. The data example used in the experiment.

A. *The analysis process of traditional KNN*

The traditional KNN method used in this paper divides the existing data set into training set and test set. Among them, the first 3,000 pieces of data are the data before 2020, and the following data are the data in 2020.The data before 2020 is used as the training set and the data in 2020 as the test set. The stock trend is modeled based on KNN algorithm and gridsearch method is used to find the best parameters.

B. *The analysis process of the improved KNN*

The improved KNN method used in this paper converts time series into a supervised learning problem. The existing data set is divided into training set and test set, among which, the first 3000 data are the data before 2020, and the following data are the data in 2020.The data before 2020 is used as the training set and the data in 2020 as the test set. The stock trend is modeled based on KNN algorithm and gridsearch method is used to find the best parameters.

C. *Experiment result analysis*

After preprocessing data for the tradition KNN, we fit the model and make the prediction thereby. The result reports that the standard error of stock price prediction based on traditional KNN algorithm is 3.9710598792900704. Then, the training data is used for parameter estimation and regression prediction, and the standard error of stock price prediction based on regression prediction is 4.21897630556746. Comparably, after preprocessing data for the improved KNN, the model is fitted and the prediction is made, and the standard error of the stock price prediction based on the improved KNN algorithm is 3.6600746663610986. The details are shown in the following table, which demonstrates that the standard error of stock price prediction is lower when utilizing our improved KNN method, leading to the superiority of our proposed method.

TABLE I. COMPARISONS AMONG VARIOUS PREDICTION MODELS

Algorithms	standard error of stock price prediction
Regression prediction	4.21897630556746
Tradition KNN	3.9710598792900704
Improved KNN	3.6600746663610986

IV. CONCLUSION AND DISCUSSION

Based on the traditional KNN algorithm, the characteristics of a single data object are taken as the input, that is, the change trend of the following day is predicted only through the data of one day. For reference, this paper proposes an improved KNN algorithm, that is, the stock price information of the first N days is combined into a sample, which is then input to the KNN model for learning. The experimental results illustrate that the prediction accuracy of the improved KNN algorithm can be greatly enhanced compared with that of the original KNN method or the regression prediction method, verifying the effectiveness of our proposed improved KNN method.

ACKNOWLEDGMENT

First and foremost, I would like to express my sincere gratitude to my supervisor, Prof. Huang Jianhui. During the long-lasting period that I spent writing this paper, Prof. Huang kept teaching me the programming method and essay writing method, and encouraged me a lot as well. I have been extremely fortunate to have such a nice supervisor who was so patient and responsible. I consider myself blessed to have worked under the expert hand of this great advisor and mentor. His dedication and passion had made it possible for me to gain the experience to effectively write research papers.

Second, my deepest love and gratitude are devoted to me family members. Their love will be an endless motivation for me to strive forward and help me overcome the difficult times in my life.

REFERENCES

[1] [1]GUO Jian-feng,LI Yu,AN Dong.Prediction of Short-term Stock Price Based on LM-GA-BP Neural Network[J]. Computer Technology and Development,2017,27(01):152-155+159.

[2] [2]Sun Chen,Li Yang,Li Xiaoge,Yu Jiaoyan.STOCK FORECASTING MODEL BASED ON OPTIMISING BP NEURAL NETWORK WITH CUCKOO SEARCH[J]. Computer Applications and Software,2016,33(02):276-279.

[3] [3]XU Chen-meng,FANG Hua.Empirical Study on Stock Price Forecast by ARMA Model[J]. Economic Research Guide,2019(31):77-82.

[4] [4]ZHANG Ning-zhi,ZHOU Jia-li,SUN Wu-jun.Stock Index Prediction Based on Optimized Echo State Network[J]. Journal of Industrial Engineering and Engineering Management,2014,28(01):94-101.

[5] [5]RAO Dong-ning,DENG Fu-dong,JIANG Zhi-hua.Stock Price Movements Prediction Based on Multisources[J]. Computer Science,2017,44(10):193-202.

[6] [6]YANG Zhen-jian,XIA Ke-wen.Price Forecasting Method for IPO Based on Gaussian Process Regression[J]. Computer Simulation,2013,30(01):293-296+304.

[7] [7]Lin Sheng,Qi Ke,Wei Kaicong,Zhang Wei.Generality of Machine Learning in Stock Price Forecasting[J].China Economist,2019(03):71-73+78.

[8] [8]ZHANG Gui-Sheng, ZHANG Xin-Dong. A differential-information based ARMAD-GARCH stock price forecasting model[J].Systems Engineering-Theory&Practice,2016,36(05):1136-1145.

[9] [9]JIANG Yi,LIN Yong-Peng.The Application of RBF Neural Networks to Stock Price Forecasting[J]. Mind and Computation,2007(04):413-419.

[10] [10]SUN Qiu-Yun,LIU Jin-Qing,LIU Yin,WU Qing-Xiang. Stock Price Prediction Model Based on SVR with Parameters Optimized by Improved GA [J]. Computer Systems & Applications,2015,24(09):29-34.

[11] [11]CHEN Lu-lu. Based on Multivariate Linear Regression Analysis——Forecasting Stock Prices in China Citic Bank [J]. Economic Research Guide,2016(19):75-76.

[12] [12]LUO Xianglong,LI Danyang,YANG Yu,ZHANG Shengrui. Short-term Traffic Flow Prediction Based on KNN-LSTM [J]. Journal of Beijing University of Technology,2018,44(12):1521-1527.

[13] [13]Liu Xiaoying. Research on the Prediction of Stock Prices Trend Based on K Nearest Neighbor[D].Northeast Agricultural University,2015.

[14] [14]Yang Mao,Jia Yunpeng,Mu Gang,Yan Gangui,Liu Jia. Wind power real- time prediction research based on the improved KNN algorithm[J]. Electrical Measurement & Instrumentation,2014,51(24):38-43.

[15] [15]LIN Fang,LIN Yan,LV Xianlong,CHENG Xingong,ZHANG Huiyu,CHEN Bojian. Short-term Parallel Power Load Forecasting Based on Balanced KNN[J]. Electric Power,2018,51(10):88-94+102.

[16] [16]LU Xuanmin,Yuan Wenle,Qiu Yang,Yang Fan. Improved dynamic prediction fingerprint localization algorithm based on KNN [J]. Application Research of Computers,2017,34(07):2016-2018.