ISSN: 2414-1895 DOI: 10.6919/ICJE.202112_7(12).0093

Stock prediction method based on integrated learning

Dickson Di Sen Wang[1*], Cheng Guo[2,a], Yue Yang[3,b], Yijian Zhang[4,c], Zihan Wang[5,d]

[1]The British school of Guangzhou, Guangzhou, Guangdong, China

[2]Guanghua Cambridge International School, Shanghai, China

[3]Ealing International School, Dlian, Liaoning, China

[4]Beijing No.35 High School, Beijing, China

[5]Hangzhou No.14 Secondary School, Hangzhou, Zhejiang, China

*Corresponding author: dickson_wang@bsg.org.cn

[a]2633573653@qq.com

[b]1795323096@qq.com

[c]390723161@qq.com

[d]2230762189@qq.com

These authors have contributed equally to this work

Abstract

With the rapid development of the stock market, fields related to stock prediction have attracted the attention of more and more researchers. Stock forecasts can help the healthy development of the stock market and are of great significance to society, companies and individual investors. Through experiments, we found that the support vector regression model (SVR) is not stable due to the distribution characteristics of the training data. The specific manifestation is that the stock price forecasts of certain stocks have large deviations. In response to this problem, this paper improves the SVR model from the perspective of integrated learning. Based on the SVR model, we integrated two simple and effective learners, linear regression model (LR) and K nearest neighbor model (KNN) to enhance the generalization ability of the SVR model. Experiments show that the ensemble model proposed in this paper has a significant improvement in stock price prediction accuracy compared to the pure SVR model.

Keywords

Stock price prediction, SVR, Integrated learning.

1. Introduction

A stock market is a place that reflects the development of a national economy, which can be also called a "barometer" or an "early warning device". As early as 1611, some businessmen bought and sold shares of overseas trading companies in Amsterdam, this is the embryonic form of stock exchanges. The stock market developed rapidly, after entering the 20th century. Especially after 1970, the stock scale was expanding and transactions became more and more active. This happened because, in developed countries, the scale and intensification as well as the progress of communication network technology have increased.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202112_7(12).0093

With the continuous improvement of living standards and having the hope of getting higher returns, more and more individuals and institutions put money into the stock market. However, most investors' have weak risk-taking ability, and the dramatic fluctuation in stock prices is likely to lead to great losses towards shareholders. Therefore, the stock market has always needed a more efficient investment theory than in the past to provide better investment methods for investors to improve investment returns and reduce investment risks.

Nowadays, many researchers have begun to study stock price prediction. Some researchers list historical data in the stock market in chronological order and use time-series statistical models based prediction methods, such as ARMA model, GARCH model, Markov model and average moving model (MA) and many more to establish the quantitative relationship between the past and future prices, which can predict the changes of the stock market in the next few days more accurately.

2. Literature review

Since the 1990s, some researchers began to apply machine learning to stock price prediction. In the literature, Gencay et al. established a forward neural network model. The results show that the effect of the neural network model is better than that of the simple model MA, however, the prediction effect is not up to industrial standards. In the literature, Zhang et al. proposed a method to combine ARIMA and artificial neural networks, and proved that the prediction accuracy of this combination method is higher than that of the ARIMA model in the nonlinear data processing. Some scholars use short-term and short-term memory networks (LSTM) and cyclic neural networks (RNN) to predict stocks.

Support vector machine model (Support Vector Machine, SVM) is another machine learning model that is often used in stock price prediction research. SVM method uses the principle of post-festival risk minimization and has good generalization ability. Support vector regression (support vector regression, SVR) is an important branch of SVM. In 2003,

Kim used the SVR model in stock price prediction and compared it with the BP neural network model, it was found that the accuracy of the model was higher. After that, more and more investors start to use the SVR model to predict the stock price. We used the SVR model to predict the stock price and found that the price forecast of individual stocks (such as Amazon stock, stock code AMZN, and Tianchang group, stock code 2182.HK) was very deviated. Aiming at this problem, we suggested an improved algorithm and will be proposed in this paper.

Two are selected as the analysis resultants of the model predication. One is Amazon(AMZN),the other is Ping An Bank(000001). The data of two model are the historical data of the two stocks from June 2016 to June 2020.

Amazon stock(AMZN):It can be found from the data that AMZN mainly shows a increasing trend, rising volatility, while it will fall in a few cases. The range for rising is within 11per cent. However, the range of decline is wide. And in the period of growing, the closing price is higher than opening price. Meanwhile, the closing price is lower than the opening price during the decline.

Ping An Bank(000001):it can be also found from the data that there are great variation for Ping An Bank,up and down. In total, the rise are almost same as the fall. Even ,there are continue drops. The value can increase by up to 13 per cent and decrease by down to 14 per cent. At the same time ,the closing value is higher than opening value in rise progress. At the same time, the closing value is lower than opening value in descending progress.

Short-term securities refer to all kinds of marketable securities that can be realized at any time and held for no more than one year, as well as other investments for no more than one year. I found a few stocks to have a general study of their trend. The first stock is NIKE, whose overall trend over the past year has been upward. From 90 per share to about 120 per share. However, it has a long period of decline in between. It began to decline from about February 20, 2020, and reached the lowest at

ISSN: 2414-1895 DOI: 10.6919/ICJE.202112_7(12).0093

the end of March, about 60 per share. This is probably because of coVID-19, which is causing people not to go shopping, and some other reasons, which is causing the stock price to plummet. It did not begin to return to pre-epidemic levels until early June. After that, the stock began to rise until September of this year. The second stock is Boeing. From September last year to the end of February this year, the stock price has been in a very stable, little volatility, and has remained at 340-380, until the end of February this year, the stock price began to fall rapidly. From the end of February to the middle of March, the share price fell from 340 to 95 in just half a month. The share price has never been back to where it was. Until a small peak in early June, the stock has remained around 170 since then.

SVR (Support Vector Regression) is same as SVC, selects a part of data from the practiced database to support vector more positively. And analysis the predicted object through the value of sample.

Linear regression refers to using linear functions to fit samples in vector Spaces. The model takes the comprehensive distance between the actual positions of all samples and the linear function as the loss, and calculates the parameters of the linear function by minimizing the loss. For linear regression, a sample is calculated as long as it does not fall exactly on the linear function as the model.

The model of SVR is also a linear function 'y= kx + b', but different from linear regression. It is a more tolerant predictive model than linear regression. The SVR creates an "interval band" on both sides of the linear function. For all samples falling into the interval band, the loss is not calculated. Only those outside the interval band are included in the loss function. Then the model is optimized by minimizing the width of the interval band and the total loss.

3. Experiment

3.1 Experimental data

We obtained the stock historical data of many domestic and foreign companies through the yfinance module. In the experiment part, we mainly focused on the price prediction of individual stocks (such as Amazon stock, stock code AMZN and Tianchang Group, stock code 2182.HK), because the SVR model showed a large deviation in the price prediction of these several stocks.

We mainly obtained the stock prices of Amazon and Tianchang Group during 2000-01-01 to 2018-12-01, and used the closing price as the target value for training and prediction.

We will get 70% of the data as training data and 30% as testing data.

3.2 Experimental results

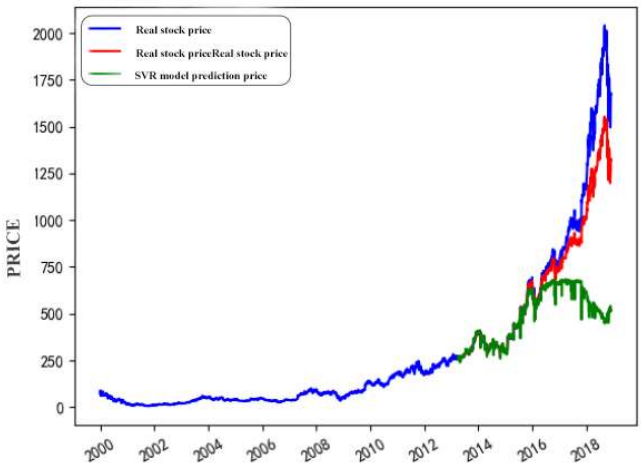

Figure 1: Amazon stock forecast results

ISSN: 2414-1895 DOI: 10.6919/ICJE.202112_7(12).0093

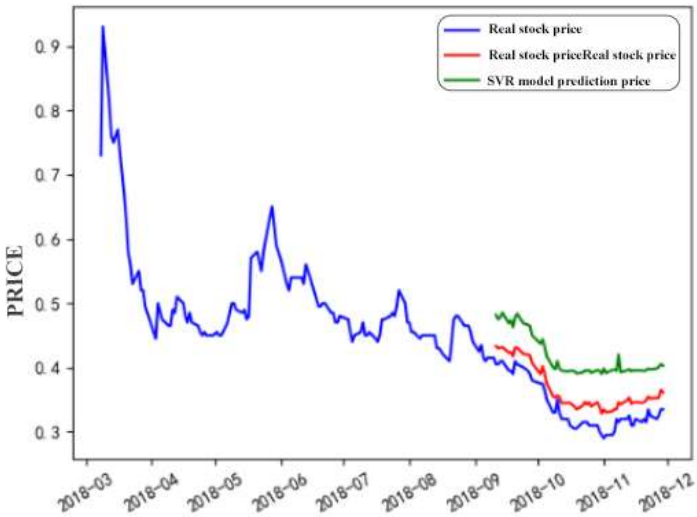

Figure 2: Tianchang Group Stock Forecast

We used the SVR model and the integrated model proposed in the paper to forecast the stock prices of Amazon and Tianchang Group respectively, and the experimental results are shown in Figure 1-2 below.

It is obvious from the figure that the prediction result of the integrated model is better and more accurate than that of the simple SVR model, no matter for the prediction of Amazon's stock or the stock price of Tianchang Group.

Table 1 lists the results obtained after the training of stock price data by the two models. In addition to the root mean square error RMSE and determination coefficient introduced before R2, the evaluation index also has a score of 10% discount cross validation. K folding cross validation (k-fold cross-validation) first divides all the data into K sub-samples, one of which is not repeated as the test set, and the other K-1 samples are used for training. Repeat K times and get a score of K times on average. Here we take K=10.

Table 1 The results obtained after the training of stock price data

	SVR Model	Integration Model
RMSE	0.0674	0.0319
	273.8249	82.44144
R2	0.5339	0.8987
	0.5512	0.9588
Cross_val_score	0.1218	0.8989
	0.5367	0.9587

In Table 1, the shadow table is the data of Tianchang stock, and the blank background is Amazon stock. We can see from the table that the three evaluation indexes of the integrated model are all improved to a certain extent than the single SVR model, which shows that the root mean square error RMSE decreases, the correlation index increases, and the cross-validation score increases on the test set. The evaluation index and cross-validation score have been greatly improved.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202112_7(12).0093

4. Summary

This paper finds out the problem that the generalization ability of SVR model is not strong enough through the prediction experiment of multiple stock prices by SVR model, and puts forward an improved algorithm based on integrated learning. The idea of integrated learning is to combine a variety of weak learners, to promote strengths and circumvent weaknesses, and finally to obtain a stronger learner. the integrated model proposed in this paper integrates simple and effective linear model LR and K nearest neighbor model based on the SVR model. training the prediction results with different stock data proves that the integrated model has better accuracy than the simple SVR model in many evaluation indexes. At the same time, this paper is still further mentioned Ascension space: this paper only considers the influence factors of closing price, if more factors can be combined, better prediction results may be obtained. When using the SVR model, this paper adopts the setting of default parameters. The search for optimal parameters is also a direction to explore.

References

[1] Zhao Zhenquan, Jiang Yingkun, Chen Shoudong, Dai Li. Appropriate scale of China's securities market. China in 2001: Analysis and Forecast of Economic Situation, 2001.

[2] Xu Haoran, Xu Bo, Xu Kewen. Overview of the application of machine learning in stock forecasting [J]. Computer Engineering and Applications, 2020, 56(12): 19-24.

[3] Vapnik, Vladimir. The nature of statistical learning theory. Springer science & business media, 2013.

[4] Kim, Kyoung-jae. "Financial time series forecasting using support vector machines." Neurocomputing 55.1-2 (2003): 307-319.

[5] Mei, Wenjuan, et al. "Stock price prediction based on ARIMA-SVM model." (2018).