Apply RF-LSTM to Predicting Future Share Price

Haoxian Nan*

The college of communication engineering, Jilin University, 130012, China

Abstract: In recent years, share has become a more and more heated topic in the whole world. As one of the most significant items, the data is featured by great complexity, especially in price prediction. According to the previous relevant study, its prediction has high requirements for the model, which indicates using a single model cannot acquire relatively accurate prediction results. With regard to this problem, integrating random forest and long short-term memory is illustrated to solve that. The first step is the normalization of related share data, which is to reduce the influence caused by the discrepancy of different data. And then, random forest is used for choosing relatively optimal feature rally. In contrast to single decision tree, the application of random forest has simplified the complexity of training. After that, long short-term memory is used for forecasting the price and optimize plentiful important parameters in the model. According to test consequence, the error rate of the integrated model is decreased obviously.

1. Introduction

The fluctuations in the share market are of great importance to the prediction of share prices. The role of accurate predicting share price in the share market with severe fluctuations and risks is increasingly prominent. Both financial institutions and regulatory authorities have paid enough attention to this. Thus, studying share price variety is major to national economic development and preventing financial risks. There are many factors for changes of share price in the ticket market, such as company news, industry performance, investors’ emotions, social media public opinion and economy factors and so on. High volatility and uncertainty of the share price make it become a major problem in the research of financial fields. shares are risky asset production and prediction of share prices can help investors prevent risky insurance, improving the security and profitability of share investment. Therefore, researching share price prediction method to maximize the investment style insurance and increasing investment income have important practical significance.

In recent years, some combined models have achieved better results in the forecast of share price tendencies. Such as A. M. Erher. Combined [1] with a self-regulating mobile average model, index smooth model and RNN to form a hybrid model. The verification consequence indicate that the hybrid model is greater than simply using RNN; Li Haiyan [2] adopts the main component analysis method [2]. The method of combining PCA, genetic algorithms, and BPNN is used for the price prediction of the share marker, and the results prove that the preciseness of the prediction is more accurate; Y. J. BAEK [3] put forward a classical model named LSTM that prevents overfitting and the share market index prediction framework as the foundation of the LSTM model. The consequences indicate that the model has good predictive preciseness; H. H. Y. Kim [4] integrated the this model with different broad-regression conditions models to propose a new hybrid LSTM model to improve predictive performance; TAN Z [5] based on random forests Prediction of share prices, the importance of using RF analysis features, and set appropriate parameters based on the importance of features to formulate reasonable strategies. Because the share market is unstable, nonlinear, mutable and non-parameter, the RF method can better study these characteristics compared to traditional data analysis methods. RF is capable of analyzing complex characteristics and possesses a fast learning speed. It could be utilized as a feature selection tool for multi-dimensional data. In recent times, it has been extensively applied in various predictions, classification, and characteristics [6-8].

2. Relevant index

The technical index is the specific result acquired by the processing of the original data in accordance with a certain indicator algorithm. The processing result is a data sort. The share price forecast based on the RF-LSTM combination model forecast 137 share market prediction analysis, which has the characteristics of intuitive and specific application when judging the share market. various technical index have individual scope and restriction. When the share characteristics represent, a single technical index cannot ensure the synthesis and preciseness of the characteristic representation. Thus, the selection of multiple representative and easy-to-quantifying technical indexes could improve the
preciseness of the complementary and characteristics of data. This article builds forecasting features based on the 11 technical indexes commonly applied in the share market. As shown in Fig.1, these indexes comprehensively reflect the tendency change information of the share price, and can include most influencing factors that can include share price prediction.

3. Main methods

3.1 Random forest

Random forest refers to a method that uses multiple decision-making trees to train, classify, and predict the sample data. It can also give the data to the importance of the data while classifying the data. The role in classification, the random forest itself can be used as a feature selection method. Use the importance of the variables of random forest algorithms to sort the features, and then use the sequence to search the search method. Each time from the feature collection, remove the less important (minimum importance score) feature, iterate one by one, and calculate the classification correctly. In the end, the feature collection with the least number of variables and the highest classification rate of variables is used as a feature selection result. Optimizing the input variable combination with the importance of the variable importance of random forests can significantly improve the predictive performance of this forest model, simplify the complexity of the model, and improve the prediction preciseness.

3.2 Long short-term memory

Long-term memory network LSTM (Long Short-Term Memory) is a variant of RNN. The core concept is the cell state and the “door” structure. The cell state is equivalent to the path of information transmission, allowing information to pass in the sequence connection. It solves the issue of gradient explosion and gradient disappearance of RNN during the long sequence training process. When the number of network layers increases, the subsequent nodes weaken the perception of previous nodes. Time push will cause the partial loss of the former information. On the foundation of ordinary RNN, LSTM appends memory units in the neutral units within the hidden layer, so the storage information on the time sequence is controlled. Thus, it can deepen the potential laws among data and make the prediction more accurate and reliable.
4. Process analysis

Flow chat of RF-LSTM is shown as Fig.2.

Firstly, Obtain share data, build technical index for predicts as features, and treat feature sets.
Secondly, train a random forest and use the Bootstrap method to resample.
Thirdly, as for every decision tree, choose the corresponding out-of-bag data to calculate the data error and record it as ERROR1.
Fourthly, randomly additional noise disturbance to all samples of OOB data, and then calculate OOB data errors again, record it as ERROR2. The feature importance equation is \( \sum (\text{ERROR2-ERROR1})/\text{AMOUNT} \).
Fifthly, sort all features according to their significance.
Finally, LSTM is used to make prediction and get the final conclusion.

5. Experimental feasibility analysis

5.1 Experimental data and environment

The original data studied in this article comes from akshare financial package. Akshare is a free, convenient Python financial data interface package. The whole data will be split into training rally and test rally. In this project, the last 30 days are selected as test set. The selected features are tested on the independent test set, and the classification predictions that have not been selected as a comparison experiment will be selected.

The introduction of this experimental environment is shown as Fig.3.

5.2 Experimental index

The validity of the evaluation index evaluation model was evaluated with average absolute error (MAE), average square error (MSE), and average square root error (RMSE).

5.3 Experimental result

The experimental result is illustrated in Fig.4 and Fig.5 below:
6. Final conclusion

According to two resulting graphs, the fitting effect of RF-LSTM is relatively good. There is a suitable fit in the model, and the error is relatively low. On the one hand, the use of random forest algorithms to select optimal features can reduce the dimensions of data and the complexity of training. On the other hand, using long-term memory networks to build prediction models and adjust model parameters can improve the preciseness of share price prediction.

In the future, artificial intelligence will be applied much more widely than now in a variety of fields, shown as Fig.6. The aspect of data prediction mentioned in this paper is just a tip of an iceberg, which indicates the great significance of artificial intelligence in modern society. There are some cites of its application in other fields below.

Firstly, computer vision, the basic of various biometric authentication, is one of the main fields which is based on a great amount of artificial intelligence key techniques.

Secondly, the ability to decide in the automatic drive is the representative of deep learning. Nowadays, this technique is becoming more and more mature with the development of related algorithms.

Lastly, in the wireless communication, the researchers also integrate plenty of classical artificial intelligence with the signal processing. For example, Q-algorithm used to be utilized into the distribution of communication nodes.

Reference
2. Li H Y. Research on share Price Prediction Based on PCA-GA-BPNN Model[D]. Xi’an: Xi’an University of Technology, 2018.