Stock price prediction based on particle swarm algorithm optimised SVM univariate time series algorithm

Yuhang Li*

*Corresponding author email: lyh15204130550@qq.com

College of business administration, Northeastern University, Liaoning, Shenyang, 110000, China

Abstract. In this paper, using the Coca-Cola stock price data from the UCL public dataset, the support vector machine univariate time series algorithm optimised by particle swarm algorithm is used to conduct a prediction study on the stock opening price. By observing the X-Y scatter plot of the stock price and the actual stock price, it is found that the distribution of data points in the training set and the test set mainly focuses on the Y=X straight line, indicating that the predicted value is very close to the actual value and the model has good prediction effect. Further analysis of the line graphs of predicted and actual stock prices shows that the root mean square error (RMSE) of the training set is 0.78195 and the RMSE of the test set is 0.56134, which verifies the high accuracy of the model. Overall, the support vector machine univariate time series algorithm optimised based on particle swarm algorithm can not only accurately predict the stock price values, but also better capture the stock price trend.

Keywords: SVM, Stock price prediction, Time series.

1. Introduction

Stock price forecasting is an important research topic in finance, and its background can be traced back to the 1970s. In today's information explosion era, the stock market fluctuates dramatically, and investors need accurate forecasts to guide their decisions, and time series algorithms play a key role in stock price forecasting [1,2].

Firstly, the research background of stock price forecasting stems from the concern about uncertainty and risk in the financial market. The stock market is affected by a variety of factors, such as economic policies, company performance, international situation, etc. Price fluctuations are not only affected by intrinsic fundamentals, but also by external market sentiment and other factors [3].

Secondly, time series algorithms play an important role in stock price forecasting. Time series analysis is a method that uses historical data to predict future trends. By modelling and analysing stock prices, turnover and other data, the hidden patterns and trends can be revealed, so as to achieve accurate forecasts of future price movements. Common time series algorithms include ARIMA model [4], GARCH model [5], LSTM neural network [6] and so on.

ECIT 2024, March 15-17, Changsha, People's Republic of China

DOI 10.4108/eai.15-3-2024.2346191

ARIMA (Autoregressive Integrated Moving Average) model is a classical time series analysis method for linear data and captures trends and seasonal variations in the data [7].GARCH (Generalised Autoregressive Conditional Heteroskedasticity) model is specifically designed to deal with the characteristics of data with Heteroskedasticity in the financial sector, which is more effective when stock prices are more volatile [8]. And LSTM (Long Short-Term Memory) neural network performs well in dealing with nonlinear and non-stationary time series data due to its powerful memory and nonlinear modelling capabilities [9].

In summary, stock price forecasting is not only one of the important topics in the field of finance, but also an important reference for investors to make decisions. Time series algorithms play a crucial role in stock price forecasting, and by using these algorithms scientifically and reasonably, and analysing and applying them in combination with real-time market intelligence, they are expected to improve the accuracy and effectiveness of investment decisions. In this paper, the support vector machine univariate time series algorithm optimised based on particle swarm algorithm is used to forecast the opening price of stocks, which provides some reference for the subsequent researchers.

2. Data set sources and data analysis

The Coca-Cola Company, a well-known beverage company, produces, markets and sells a wide range of non-alcoholic beverages across several categories worldwide. The company's products include sparkling soft drinks, sparkling flavoured drinks, water, sports drinks, coffee and tea, fruit juices, value-added dairy products, plant-based beverages, and other types of beverages. Additionally, Coca-Cola provides concentrated beverages, syrups, and fountain syrups to a variety of fountain retailers to meet the needs of different customer segments.

The dataset consists of the daily stock prices of Coca-Cola Company and includes information on the opening, closing, high and low prices for each day from 2 January 1962 to 5 December 2023. In this paper, the daily opening price from 1 January 2021 to 5 December 2023 is selected as the data set for model training and prediction [10].

By analysing the stock price changes, we can better understand the fluctuation of Coca-Cola Company's stock price in the past years, which can help investors to make corresponding investment strategies and decisions. The stock price chart can visualise the changes in stock price over time, helping people to observe the market trend and make predictions and judgements accordingly. Based on this data and information, investors can gain a deeper understanding of Coca-Cola's performance in the financial market and make a more informed investment choice by analysing it in conjunction with other factors. At the same time, the company's internal management can also take this opportunity to assess business performance and formulate future development strategies. The change in the stock price curve is shown in Figure 1.

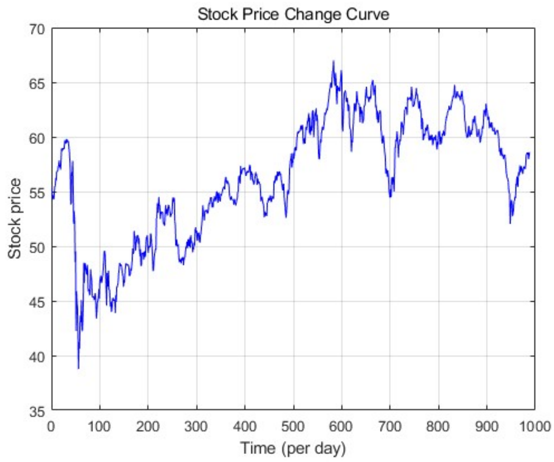

Figure 1. Partial data. (Photo credit : Original)

3. Method

3.1. Particle swarm algorithm

The particle swarm algorithm is a heuristic optimisation algorithm inspired by collaborative behaviour in groups of organisms, such as the collective intelligence of a flock of birds or a school of fish when searching for food. In the particle swarm algorithm, each individual is considered as a particle, and these particles move in the solution space and search for the optimal solution by continuously adjusting their position and speed. Each particle updates its position based on its own historical experience and information from neighbouring particles, thus gradually converging to the global optimal solution. By simulating social behaviour and information sharing, the particle swarm algorithm can be effectively applied to a variety of optimization problems with good convergence and global search capability. The schematic diagram of the particle swarm algorithm is shown in Figure 2.

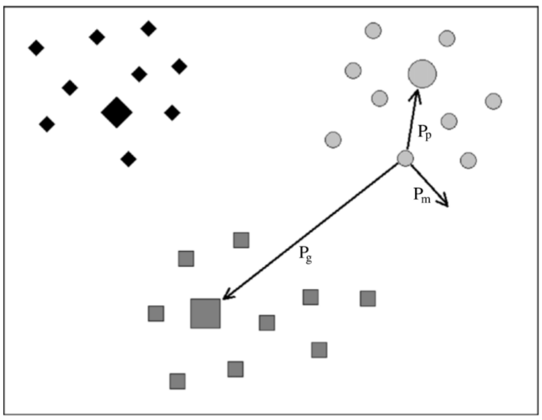

Figure 2. Partial data. (Photo credit: Original)

In particle swarm algorithms, the fitness function is used to evaluate the quality of the candidate solutions represented by each particle in the problem space. The fitness function is usually associated with the characteristics of a specific problem, in addition, the particle swarm algorithm involves the setting of some key parameters, such as the learning factor and the inertia weights, in order to better guide the search process of the particles. The learning factor influences the speed and magnitude of the position adjustment of the particles based on individual and social experiences, while the inertia weight affects the extent to which the particles maintain the inertia of their motion in the search space, and these parameters have an important impact on the convergence speed and global search capability of the algorithm. By appropriately defining the fitness function and adjusting the key parameters, the particle swarm algorithm can effectively solve various complex optimisation problems and achieve satisfactory optimisation results.

3.2. Support vector machine

Support vector machine is a supervised learning algorithm for classification and regression analysis. The basic idea is to find an optimal hyperplane that effectively separates different classes of data points. In two-dimensional space, this hyperplane is a straight line; in multidimensional space, this hyperplane is a hyperplane in a higher-dimensional space. The schematic diagram of the support vector machine is shown in Figure 3.

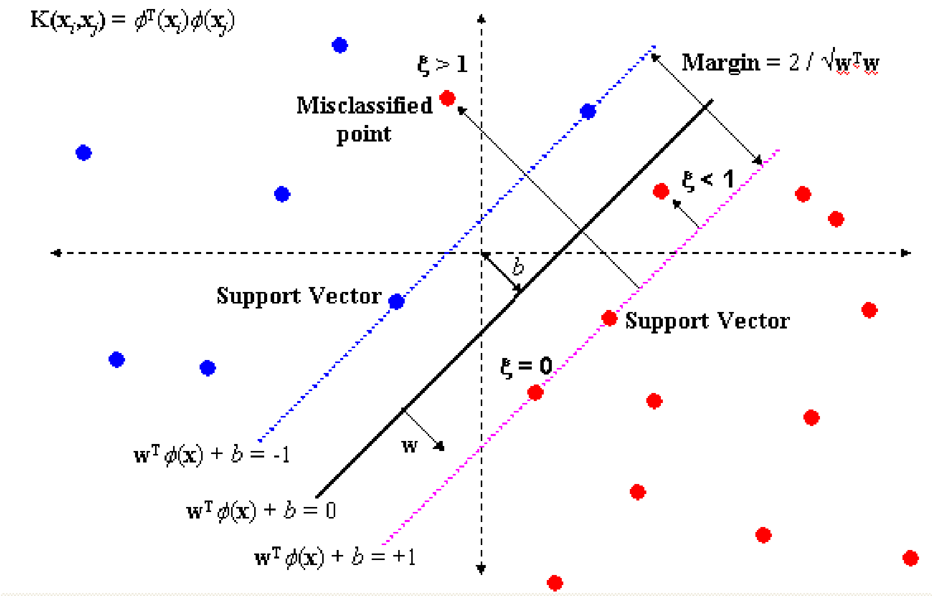

Figure 3. Partial data. (Photo credit: Original)

Support vector machines are powerful supervised learning algorithms whose key concepts include support vectors, intervals, and kernel functions. Support vectors are the sample points in the training data that are closest to the decision boundary (hyperplane), and they play a key role in defining the decision boundary. The interval is the distance from the support vectors to the hyperplane, and the optimisation goal of the SVM is to maximise this interval to improve the generalisation and robustness of the model. Kernel functions, on the other hand, allow SVMs to solve complex problems by performing nonlinear divisions in high-dimensional spaces and mapping raw features to spaces more suitable for linear classification. Through the understanding and application of these key concepts, Support Vector Machines have been widely used and achieved good results in the fields of classification, regression and anomaly detection.

3.3. Univariate time series algorithm for support vector machines based on particle swarm algorithm optimisation

Support Vector Machine (SVM), a powerful supervised learning algorithm, is known for its ability to find the optimal hyperplane in a high-dimensional space for data classification or regression. In univariate time series forecasting, it is a common approach to use SVM for modelling and prediction after converting time series data into feature vectors. However, the performance of SVM in practical applications depends greatly on the selected hyperparameters, such as the penalty coefficient C and kernel function parameters. The proper tuning of these hyperparameters is crucial for the generalisation ability and prediction performance of the model, so effective tuning strategies are needed to optimise the SVM model.

The univariate time series prediction based on particle swarm optimisation (PSO) algorithm combined with SVM is a novel idea that combines heuristic optimisation and supervised learning methods. This approach not only improves the performance of the model, but also promises better results in practical applications. Optimising the hyperparameter selection process of SVM by PSO algorithm enables the model to capture the patterns and trends in time series data more accurately, which brings new possibilities and opportunities to the field of time series analysis. The univariate time series forecasting algorithm of SVM based on PSO optimisation has a broad application prospect and is expected to become one of the effective tools and methods for solving practical time series forecasting problems.

4. Result

For the support vector machine part, the parameters are set as n estimators=100, max depth=10, min samples split=2, min samples leaf=1; for the particle swarm algorithm part, the number of particles is 50, the number of iterations is 100, and the learning factors c1 and c2 are both set to 2.0. In terms of the device parameters, a CPU Intel i7-10700K, 32GB of RAM, and GPU Nvidia RTX 3080 for training, and the Python programming language combined with the Scikit-learn library to implement the Support Vector Machines, and the PySwarm library to implement the Particle Swarm algorithm. Finally, for dataset allocation, the dataset is divided in the ratio of 70% for training set, 15% for validation set as well as 15% for testing set.

The X-Y scatter plots of predicted and actual stock prices for the training set are output, while the X-Y scatter plots of predicted and actual stock prices for the test set are output, as shown in Figure 4.

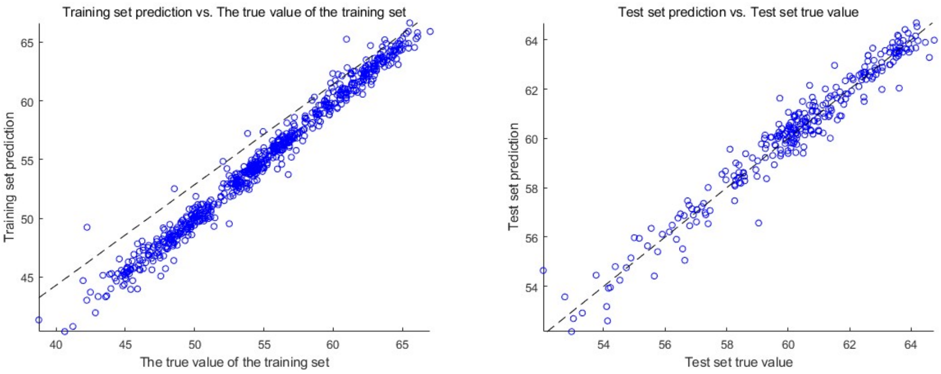

Figure 4. Loss and Accuracy change curves. (Photo credit: Original)

As can be seen from Fig. 2, the X-Y scatter distributions of the training set and test set are concentrated on the straight line of Y = X. It can be proved that the predicted values are very close to the actual values, and the model's prediction effect is very good. Output the line graphs of predicted stock prices and actual stock prices for the training set, and also output the line graphs of stock prices and actual stock prices for the test set, as shown in Fig. 5.

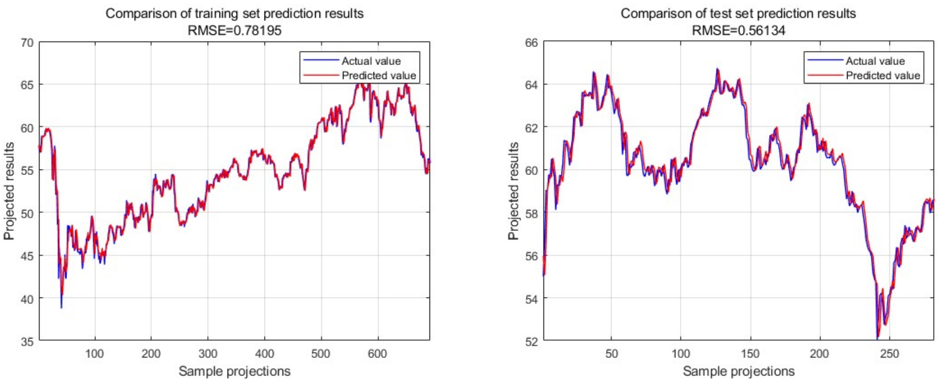

Figure 5. Partial data. (Photo credit: Original)

Based on the analysis of the line graphs of the predicted and actual stock prices, we can find that the root mean square error of the model is 0.78195 in the training set and 0.56134 in the testing set. These data indicate that the predicted values of the model are very close to the actual values and have high accuracy. In addition, when observing the line graphs, it can also be found that the univariate time series algorithm for support vector machine optimised based on particle swarm algorithm is not only able to accurately predict the value of the stock price, but also able to capture the trend change of the stock price.

By training and testing the model, and comparing and analysing the prediction results with the actual situation, we have come to the above conclusions. This research result provides useful reference information for the financial market, and investors can use this algorithm to better formulate trading strategies and risk management plans. The particle swarm algorithm optimised univariate time series algorithm for support vector machines demonstrates its potential and superiority in stock price forecasting, providing investors with a reliable and effective decision-making tool.

5. Conclusion

The univariate time series algorithm for Support Vector Machines (SVM) optimised based on Particle Swarm Algorithm shows satisfactory performance in predicting the opening price of Coca-Cola stock. By observing the X-Y scatter distributions of the training set and test set as well as the line graphs of the predicted stock price against the actual stock price, it is clear that there is a high degree of agreement between the predicted values and the actual values, which indicates that the model has excellent accuracy. More strikingly, the algorithm not only accurately predicts stock price values, but also captures stock price trend changes, which provides an important reference for investors.

In evaluating the model performance, we focused on the root mean square error (RMSE) of the training and test sets, and the results show that the RMSE of the training set is 0.78195, and the RMSE of the test set is 0.56134. These low error values further validate the model's excellent performance in the stock price prediction task. By comparing the graphical presentation of the actual data and the prediction results, we can intuitively feel the accuracy and stability brought by the SVM univariate time series algorithm optimised based on the particle swarm algorithm.

This algorithm not only performs well in forecasting stock prices, but also demonstrates a good grasp of market volatility and trend changes. Investors can make more informed decisions and better plan their investment strategies with this powerful tool.

References

[1] Dang, Bo, et al. "Enhancing Kitchen Independence: Deep Learning-Based Object Detection for Visually Impaired Assistance." Academic Journal of Science and Technology 9.2 (2024): 180-184.

[2] Dong, **nqi, et al. "The prediction trend of enterprise financial risk based on machine learning arima model." Journal of Theory and Practice of Engineering Science 4.01 (2024): 65-71.

[3] Zhang, Cheng, Nilam Nur Amir Sjarif, and Roslina Ibrahim. "Deep learning models for price forecasting of financial time series: A review of recent advancements: 2020–2022." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 14.1 (2024): e1519.

[4] Vuong, Pham Hoang, et al. "A bibliometric literature review of stock price forecasting: From statistical model to deep learning approach." Science Progress 107.1 (2024): 00368504241236557.

[5] An, Zhiyong, et al. "A novel hierarchical feature selection with local shuffling and models reweighting for stock price forecasting." Expert Systems with Applications (2024): 123482.

[6] Chadidjah, Anna, I. Jaya, and Farah Kristiani. "The comparison stateless and stateful LSTM architectures for short-term stock price forecasting." International Journal of Data and Network Science 8.2 (2024): 689-698.

[7] Jia, Yuanzhe, Ali Anaissi, and Basem Suleiman. "ResNLS: An improved model for stock price forecasting." Computational Intelligence 40.1 (2024): e12608.

[8] Zheng, Li, Yuying Sun, and Shouyang Wang. "A novel interval-based hybrid framework for crude oil price forecasting and trading." Energy Economics 130 (2024): 107266.
[9] Harshith, N., and Prity Kumari. "Memory based neural network for cumin price forecasting in Gujarat, India." Journal of Agriculture and Food Research 15 (2024): 101020.
[10] Beniwal, Mohit, Archana Singh, and Nand Kumar. "Forecasting long-term stock prices of global indices: A forward-validating Genetic Algorithm optimization approach for Support Vector Regression." Applied Soft Computing 145 (2023): 110566.