PREDICTING STOCK PRICE OF CONSTRUCTION CORPORATES USING OPTIMIZED MACHINE LEARNING REGRESSION SLIDING-WINDOW

DỰ ĐOÁN GIÁ CỔ PHIẾU CÔNG TY XÂY DỰNG BẰNG CỬA SỔ TRƯỢT HỒI QUY MÁY HỌC TỐI ƯU

Kha Thi Nguyen[1], Thi Phuong Trang Pham[2]

[1]The University of Danang – Campus in Kontum; nguyenkha130490@gmail.com

[2]College of Technology - The University of Danang; trangpham3112@gmail.com

Abstract - Forecasting changes in stock market prices play a critical role in the development of company. However, predicting stock prices is a difficult problem because of the complexity of the stock market data. Therefore, the objective of this paper is to propose an optimized machine learning regression sliding-window model for forecasting stock prices for construction companies. The proposed model integrates the firefly algorithm (FA) and the least squares support vector regression (LSSVR). The FA automatically fine-tunes the hyper-parameters of the LSSVR to construct an optimized LSSVR model. The developed model is called the FA-LSSVM. This investigation uses stock datasets of construction companies in Taiwan; namely 2545.TW, 2597.TW and 5534.TW to verify the effectiveness of the proposed model. The results reveal that for 2597.TW stock price dataset, the FA-LSSVM model obtains highest performance with a root mean square error (RMSE) of 1.548, a mean absolute error (MAE) of 0.617, a mean absolute percentage error (MAPE) of 1.372%, and a mean square error (MSE) of 2.396. Therefore, the proposed model is a promising predicting technique for non-linear time series.

Key words - Stock market prices; sliding-window; machine learning; firefly algorithm; least squares support vector regression; time series.

Tóm tắt - Dự đoán thay đổi giá cổ phiếu đóng vai trò quan trọng trong sự phát triển của công ty. Nhưng dự đoán giá cổ phiếu rất khó khăn vì số liệu của nó rất phức tạp. Mục đích của bài báo là đề xuất mô hình dự đoán giá cổ phiếu các công ty xây dựng dựa vào cửa sổ trượt và hồi quy máy học tối ưu. Mô hình liên kết giữa thuật toán đom đóm và bình phương vec-tor hỗ trợ hồi quy. FA giúp điều chỉnh hệ số của LSSVR. Mô hình đưa ra gọi là FA-LSSVM. Nghiên cứu sử dụng dữ liệu giá cổ phiếu một số công ty xây dựng Đài Loan; 2545.TW, 2597.TW và 5534.TW; để làm rõ tính hiệu quả của mô hình. Kết quả cho thấy đối với dữ liệu 2597.TW, mô hình FA-LSSVM đạt kết quả tốt nhất với sai số tiêu chuẩn 1,548, sai số bình phương tuyệt đối 0,617, phần trăm sai số bình phương tuyệt đối 1,372%, và sai số toàn phương trung bình 2,396. Vì vậy, mô hình đề xuất là công cụ dự báo đáng mong đợi cho vấn đề chuỗi thời gian phi tuyến tính.

Từ khóa - Giá thị trường chứng khoán; cửa sổ trượt; máy học; thuật toán con đom đóm; bình phương vector hỗ trợ hồi quy; chuỗi thời gian.

1. Introduction

Stock prices are one of the most challenging applications of time series forecasting. Stock prices are predicted to determine the future value of companies' stock or other financial instruments. Therefore, there are more and more numerous models which have been proposed to provide the investors with precise predictions in stock prices. For example, Oh and Kim (2002) [1] presented backpropagation neural networks that incorporated chaotic and piecewise techniques to deal with problems of predicting stock markets. Leigh et al. (2002) [2] combined pattern recognition with neural networks to predict the New York Stock Exchange Composite index. Autoregressive integrated moving average (ARIMA) technique, which is one of a time series-based approaches, have been widely used in forecasting stock price data. However, ARIMA models cannot easily solve nonlinear patterns. Therefore, such linear models do not always satisfactorily approximate complex real-world problems [3, 4].

Machine learning (ML) technique is another method for nonlinear modeling, such as the support vector machines (SVMs) algorithm. Support vector regression (SVR). A variant of the SVM, was developed by Vapnil *et al.* (1995) [5]. However, one of main challenges for the classical SVR is its high computational complexity, because the algorithm involves constrained optimization programming. The least square support vector regression (LSSVR) is a highly enhanced machine-learning technique with many advanced features that support abstract generalization and fast computation [6]. To improve the predictive accuracy of the LSSVR model, the parameters of the LSSVR must be optimized because the performance of the LSSVR depends on the selected regularization parameter (C) and the kernel function parameter (σ), which are called LSSVR hyper-parameters.

Thanks to the development of artificial intelligence (AI) technology, many intelligent algorithms are now used in parameter optimization, including the genetic algorithm (GA) and the particle swarm optimization algorithm (PSO). Notably, firefly algorithm, which is a nature-inspired metaheuristic method, has recently performed extremely well in solving various optimization problems such as stock price forecasting [7] and electricity price prediction [8]. Many studies have also shown that the FA can solve optimization problems more efficiently than can conventional algorithms, including GA and PSO [9, 10].

Thus, this work develops an intelligent time series prediction model using sliding-window metaheuristic optimization. The proposed model integrates firefly algorithm and least squares support vector regression (FA-LSSVR) to forecast the prices of construction corporate stocks, which is an unsolved problem in the field of stock price prediction. Moreover, to validate the effectiveness of proposed model, this paper uses several datasets concerning stock prices in construction companies and compares them to experimental results.

The remainder of this paper is organized as follows. Section 2 elucidates the sliding-window, LSSVR, FA, and the predictive evaluation methods. The collection and preprocess of stock price datasets, and analytical results are mentioned in Section 3. Finally, conclusions are given in Section 4.

2. Methodology

In time series prediction, the time series are typically expanded into three or higher-dimensional space to exploit the information that is implicit in them. Selecting a suitable pairing of embedding dimension m (lag) and time delay τ is very important for phase space reconstruction [11]. In this study, the optimum values of the embedding parameters are determined by performing a sensitivity analysis.

The LSSVR approach proposed by Suykens *et al.* (2002) [12] is a well-developed ML technique. However, the prediction accuracy of the LSSVR is highly dependent on the determination of its hyper-parameters. Therefore, as a part of this study, the enhanced FA algorithm is developed to optimize LSSVR hyper-parameters, i.e., the regularization parameter (C) and the sigma of the RBF kernel (σ).

Firefly algorithm (FA) is a modern heuristic algorithm proposed by the Yang in 2008 [13]. In optimization problems, each firefly represents a potential solution, and absolute light intensity of firefly is associated with fitness function. The core idea of FA is that a firefly with greater light intensity will attract other fireflies with less light intensities. Moreover, the brightest firefly will move randomly in a certain range [14].

In this work, the FA is chosen as the training algorithm to enhance the efficiency and reduce the computational burden on the LSSVR. A novel FA-LSSVR model can be formulated to achieve fast and accurate prediction results.

The performance measures that are used to assess the predictive accuracy of the proposed system include the root mean square error (RMSE), the mean absolute error (MAE), the mean absolute percentage error (MAPE), the synthesis index (SI), and the mean square error (MSE). These indexes are used to measure whether the predicted values are close to the actual values.

3. Data preparation and analytical results

3.1. Data preparation

Historical daily prices are taken from Yahoo! Finance, a publicly accessible website by Xiong et al. (2014) [7]. Daily data in six years (5th October, 2011 to 31st May 2017) of on four stocks - Highwealth Construction Corp. (2542.TW), Huang Hsiang Construction Corporation (2545.TW), Ruentex Engineering & Construction Co., Ltd. (2597.TW) and Chong Hong Construction Co., Ltd. (5534.TW) – were downloaded from Yahoo! Finance. The data was closing stock prices. 2542.TW, 2545.TW, 2597.TW, and 5534.TW stock datasets provide important indicators of the overall performance of the construction market in Taiwan. Table 1 presents details about the stocks and the number of data instances for each stock.

Table 1. *Description of analyzed stocks and datasets*

Dataset	Total No. of data points	Duration
2542.TW	1457	Oct 5, 2011 to May 31, 2017
2545.TW	1457	Oct 5, 2011 to May 31, 2017
2597.TW	1457	Oct 5, 2011 to May 31, 2017
5534.TW	1457	Oct 5, 2011 to May 31, 2017

3.2. Analytical results

Table 2 compares the performances of the proposed model using four stock datasets. For the 2597.TW stock, the RMSE, MAE, MAPE and MSE values are 1.548, 0.617, 1.372% and 2.396, respectively. The one day-ahead prediction of the price of the 2597.TW stock is better than that of the price of the stock of any other construction company Notably, it improves the prediction accuracy of other stocks from 99.080% to 99.212% with the MAPE.

Table 2. *One day-ahead prediction performance comparison using the proposed system (for 4 stock datasets)*

Dataset	Average performance measure in test data				SI	Rank	Improved by proposed system (%)			
	RMSE	MAE	MAPE (%)	MSE			RMSE	MAE	MAPE	MSE
2542.TW	2.308	0.863	1.745	5.239	1.000	4	73.195[a]	39.525[a]	99.212[a]	53.606[a]
2545.TW	1.372	0.558	1.701	1.883	0.220	2	54.903[a]	10.884	99.192[a]	22.529
5534.TW	1.728	0.823	1.494	2.986	0.476	3	64.197[a]	33.089[a]	99.080[a]	22.871
2597.TW	1.548	0.617	1.372	2.396	0.134	1	-	-	-	-

Note: Measurements of performance improvement and hypothesis testing results are calculated using average performance measures
[a] *Significance level at 1%*

Table 3 reviews the key literature that is relevant to the present work and provides details about the choices of datasets and performance measures in those papers. Table 3 lists the max, min and average values of the performance measures that were obtained in previous works [6-12], in which different datasets and learning algorithms were used to predict the closing prices of stock one day ahead. Clearly, the one day-ahead predictions of the values of four stocks achieve good outcomes as previously reported models.

Table 3. *Summary of relevant works*
(ordered by relevance to this paper)

BIST 100 [15, 16]	RMSE	Max	3810.072
		Min	2493.273
		Average	2413.851
	MAE	Max	2951.554
		Min	2597.321
		Average	2833.042

	MAPE (%)	Max	3.863
		Min	3.381
		Average	3.688
	MSE	Max	14516650.000
		Min	11236305.000
		Average	12651969.667
S & P 500 [8, 17, 18]	RMSE	Max	0.082
		Min	0.014
		Average	0.037
	MAPE (%)	Max	46.900
		Min	0.550
		Average	9.436
	MSE	Max	0.007
		Min	0.00019
		Average	0.003
NASDAQ [19, 20]	RMSE	Max	50.167
		Min	21.600
		Average	42.035
	MAPE (%)	Max	0.190
		Min	0.130
		Average	0.160
	MSE	Max	20901.200
		Min	90.320
		Average	5721.558

4. Conclusions

This work presents a novel approach based on a firefly algorithm and least squares support vector regression (FA-LSSVR), to construct a stock price forecasting expert system with the aim of improving forecasting accuracy. To evaluate the proposed approach, the work applies four datasets for stocks in Taiwan. In particular, the one day prediction of 2597.TW stock prices is better than that of any construction company stock prices with an RMSE of 1.548, an MAE of 0.617, an MAPE of 1.372% and an MSE of 2.396. Therefore, the proposed system can be used as a tool to forecast stock prices for short-term investment and will motivate future research. In the future works, the system can be verified by other stock markets such as China, Japan and Hong Kong.

REFERENCES

[1] K.K. Oh KJ, Analyzing stock market tick data using piecewise nonlinear model, *Expert Systems with Application,* 22 (2002) 249-55.

[2] W. Leigh, M. Paz, R. Purvis, An analysis of a hybrid neural network and pattern recognition technique for predicting short-term increases in the NYSE composite index, *Omega 30(2)* (2002) 69-76.

[3] S.-Y. Lin, C.-H. Chen, C.-C. Lo, Currency Exchange Rates Prediction based on Linear Regression Analysis Using Cloud Computing, *International Journal of Grid and Distributed Computing* 6 (2013) 10.

[4] A. Abdoos, M. Hemmati, A.A. Abdoos, Short term load forecasting using a hybrid intelligent method, *Knowledge-Based Systems 76* (2015) 139-147.

[5] V.N. Vapnik, *The nature of statistical learning theory*, Springer-Verlag New York, Inc.1995.

[6] J.-S. Chou, A.-D. Pham, Nature-inspired metaheuristic optimization in least squares support vector regression for obtaining bridge scour information, *Information Sciences 399* (2017) 64-80.

[7] T. Xiong, Y. Bao, Z. Hu, Multiple-output support vector regression with a firefly algorithm for interval-valued stock price index forecasting, *Knowledge-Based Systems 55* (2014) 87-100.

[8] A.K. Rout, R. Bisoi, P.K. Dash, A low complexity evolutionary computationally efficient recurrent Functional link Neural Network for time series forecasting, IEEE Power, Communication and Information Technology Conference, PCITC 2015, *Institute of Electrical and Electronics Engineers Inc.*, Bhubaneswar, India, 2015, pp. 576-582.

[9] S.K. Pal, Rai, C. S., and Singh, A. P., Comparative study of firefly algorithm and particle swarm optimization for noisy non-linear optimization problems., (2012) *Int. J. Intell. Syst. Appl.*, 4(10), 50–57.

[10] J. Olamaei, M. Moradi, T. Kaboodi, A new adaptive modified Firefly Algorithm to solve optimal capacitor placement problem, *Electrical Power Distribution Networks (EPDC)*, 2013 18th Conference on, 2013, pp. 1-6.

[11] N.-D. Hoang, A.-D. Pham, a.M.-T. Cao, A Novel Time Series Prediction Approach Based on a Hybridization of Least Squares Support Vector Regression and Swarm Intelligence, *Applied Computational Intelligence and Soft Computing* (2014) 8.

[12] J.A.K. Suykens, T.V. Gestel, J.D. Brabanter, B.D. Moor, J. Vandewalle, *Least squares support vector machines*, World Scientific, Singapore 2002.

[13] X.-S. Yang, *Nature-Inspired Metaheuristic Algorithms*, Luniver Press 2008.

[14] S. Łukasik, S. Żak, Firefly Algorithm for Continuous Constrained Optimization Tasks, in: N.T. Nguyen, R. Kowalczyk, S.-M. Chen (Eds.), Computational Collective Intelligence. Semantic Web, Social Networks and Multiagent Systems: First International Conference, ICCCI 2009, Wrocław, Poland, October 5-7, 2009. *Proceedings, Springer Berlin Heidelberg, Berlin*, Heidelberg, 2009, pp. 97-106.

[15] E. Hadavandi, H. Shavandi, A. Ghanbari, Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting, *Knowledge-Based Systems* 23(8) (2010) 800-808.

[16] M. Göçken, M. Özçalıcı, A. Boru, A.T. Dosdoğru, Integrating metaheuristics and Artificial Neural Networks for improved stock price prediction, *Expert Systems with Applications* 44 (2016) 320-331.

[17] C.M. Anish, B. Majhi, Hybrid nonlinear adaptive scheme for stock market prediction using feedback FLANN and factor analysis, *Journal of the Korean Statistical Society* 45(1) (2016) 64-76.

[18] R. Dash, P. Dash, Efficient stock price prediction using a Self Evolving Recurrent Neuro-Fuzzy Inference System optimized through a Modified Differential Harmony Search Technique, *Expert Systems with Applications* 52 (2016) 75-90.

[19] D. Bhattacharya, A. Konar, P. Das, Secondary factor induced stock index time-series prediction using Self-Adaptive Interval Type-2 Fuzzy Sets, *Neurocomputing 171* (2016) 551-568.

[20] Shunrong Shen, Haomiao Jiang, T. Zhang, Stock Market Forecasting Using Machine Learning Algorithms, *IEEE Transactions on Neural Networks* 84(4) (2015) 21.

(The Board of Editors received the paper on 31/08/2017, its review was completed on 19/10/2017)