BUILDING A MULTI-OUTPUT HYBRID MODEL FOR INTERVAL-VALUED TIME SERIES FORECASTING

XÂY DỰNG MỘT MÔ HÌNH LAI ĐA ĐẦU RA ĐỂ DỰ BÁO CHUỖI THỜI GIAN CÓ GIÁ TRỊ TRONG KHOẢNG THỜI GIAN

Thuy-Linh Le[1], Duc Sy Nguyen[1], Le Nhat Hoang Tran[1], Vy-Tang Thi Khanh[2]

[1]The University of Danang, University of Technology and Education; lttlinh@ute.udn.vn

[2]Engineering of Technical Division, Danang Power Company Limited

Abstract - Time series analysis and forecasting is an attractive research area over the last few years, especialy in interval-valued time series. Financial market is an example; a useful model may be of great interest to home brokers who do not possess sufficient knowledge to invest in such companies. In this paper, multi-input multi-output least square support vector regression (MIMO-LSSVR) is an improved algorithm based on support vector machine (SVM), with the combination of the sliding-window algorithm is proposed for interval-valued time series forecasting, a new branch in time series analysis field. The experiment shows MIMO-S-LSSVR positive outcomes than previous results. A retest using twotime series data sets in three years demonstrates that the proposed model is a promising alternative for interval-valued time series forecasting.

Key words - Time series; interval-valued time series; sliding-window algorithm; multi-input multi-output model

Tóm tắt - Phân tích và dự báo chuỗi thời gian là một lĩnh vực nghiên cứu hấp dẫn trong những năm gần đây đặc biệt là đối với dữ liệu dạng khoảng thời gian. Thị trường tài chính là một ví dụ, một mô hình hữu ích là cần thiết đối với các nhà đầu tư không chuyên. Trong bài báo này, thuật toán hồi quy vectơ hỗ trợ vuông góc tối thiểu đa đầu ra đa đầu vào (MIMO-LSSVR) là một thuật toán cải tiến dựa trên máy vectơ hỗ trợ (SVM), cùng với sự kết hợp của thuật toán cửa sổ trượt được đề xuất cho dự báo chuỗi thời gian có giá trị trong khoảng thời gian, một nhánh mới trong lĩnh vực phân tích chuỗi thời gian. Thí nghiệm cho thấy mô hình MIMO-S-LSSVR cho kết quả tốt hơn so với nghiên cứu trước đó. Mô hình được kiểm chứng lại bởi hai chuỗi dữ liệu trong ba năm chứng minh đây một kỹ thuật thay thế đầy hứa hẹn cho dự báo chuỗi thời gian có giá trị trong khoảng thời gian.

Từ khóa - Chuỗi thời gian; chuỗi thời gian giá trị khoảng; thuật toán cửa sổ trượt; mô hình đa đầu vào đa đầu ra

1. Introduction

Machine learning technique, a strongly useful tool in Artificial Intelligence (AI) field, has been successfully applied for a wide variety of problems, especially in time-series forecasting. Support vector machines (SVMs) are specific successful models of ML. Notably, machine learning techniques are increasingly used to solve the desired problems [1], the design of engineering systems [2] and financial risk prediction [3]. Recently, support vector machines have been successfully utilized to solve classification and regression problems [4].

The crucial part of a hybrid model is setting the tuning parameters. In fact, identifying the best values of parameters for a model is an optimization problem. The least squares support vector regression (LSSVR) algorithm is a further development of SVM by Suykens in 2002 [5]. The LSSVR approach considerably reduces computational complexity and increases efficiency compared to SVM. However, the standard LSSVR only resolves single variable problems [6].

Multi-output regression is based on learning a mapping from multivariate input space to multivariate output space. Xu et al. developed a multi-output least squares support vector regression (MLSSVR) for input-output mapping in multivariate space [6]. In contrast, the development of LSSVR for multi-output cases has drawn less attention, despite the fact that a diverse range of engineering problems involve many variables and so are best solved using multivariate input-output mapping.

This research develops a multi-input multi-output hybrid forecasting model based on the LSSVR method combining with the sliding window technique and is optimized by PSO algorithm for ITS data forecasting to predict stock prices. This model to predict one-step- and multi-step-ahead forecasting as a cornerstone to develops a standalone system. The proposed model is compared with previous research and retested by two datasets which are S&P500 and VN index 100 from 2016 to 2018. The rest of this paper is organized as follows. The next section describes the research methodology. Section 3 presents analytical results and a discussion. The final section provides concluding remarks and an outline for future work.

2. Methodology

2.1. Construction of an interval-valued financial data series

An interval-valued set A has variable x, where $x_i=[x_{iL}, x_{iH}]$ and $x_{iL} \leqslant x_{iH}$. The particular value of [X] for the i^{th} element in set A can be denoted by either of two methods. The interval is ranged by the highest and lowest bound or mid-point and half-range for each element point in value chain. Table 1 shows the range of interval-valued data and Figure 1 illustrates the structure of an interval from 01/04/2016 to 22/01/2016.

Table 1. *The range of interval-valued data*

Time	Upper bound	Lower bound
04/01/2016	573.14	567.60
05/01/2016	568.90	562.91
06/01/2016	569.18	562.94
07/01/2016	568.59	557.12
08/01/2016	559.88	552.88
11/01/2016	558.82	553.10
12/01/2016	561.95	553.99
13/01/2016	563.49	557.64

14/01/2016	554.35	547.79
15/01/2016	554.02	539.96
18/01/2016	529.52	517.77
19/01/2016	535.27	526.83
20/01/2016	536.50	529.50

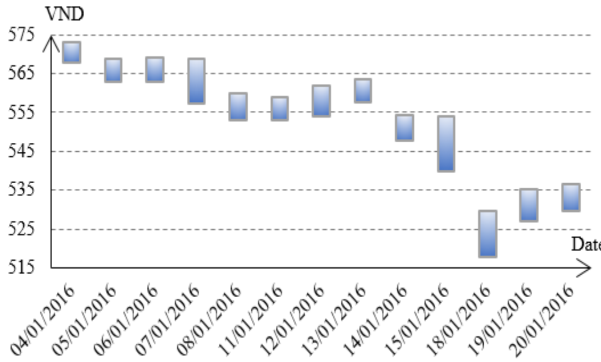

***Figure 1.** Interval-valued of daily VN100 stock price*

2.2. Phase space reconstruction using sliding-window

In time series prediction, the time series are typically expanded into three or higher-dimensional spaces to exploit the implicit information. LSSVR model is inherently limited to handle time series data while it is an advanced tool to solve problems in regression data series and classification data series. The sliding-window algorithm reconstructs data set phase space and optimizes data series by feeding back the output history data matrix as the next input matrix and removing the oldest history data in the chain [7]. Thus, contemporaneously, the main challenge of time series analysis could be solved by the sliding window method.

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \xrightarrow{m} X = \begin{bmatrix} x_1 & x_2 & \cdots & x_{m-1} & x_m \\ x_2 & x_3 & \cdots & x_m & x_{m+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{N-m-\tau+1} & x_{N-m-\tau+2} & \cdots & x_{N-\tau-1} & x_{N-\tau} \end{bmatrix}, \quad Y = \begin{bmatrix} x_{m+\tau} \\ x_{m+1+\tau} \\ \vdots \\ x_N \end{bmatrix} \tag{1}$$

2.3. Multi-input multi-output least square support vector regression model

In LSSVR for function estimation, given a training dataset $\{x_k, y_k\}_{k=1}^{N}$, the optimization problem is formulated as:

$$\min_{\omega,b,e} J(\omega,e) = \frac{1}{2}\|\omega\|^2 + \frac{1}{2}C\sum_{k=1}^{N} e_k^2;$$
$$y_k = \langle \omega, \varphi(x_k) \rangle + b + e_k, k = 1,...N \tag{2}$$

The resulting LSSVR model for function estimation is:

$$f(x) = \sum_{k=1}^{N} \alpha_k K(x, x_k) + b \tag{3}$$

Let R be the set of real numbers, for every x∈ R, x= (x1, x2,…, xi) represents a row vector with i dimensions, and x= (x1, x2,…, xj) represents a column vector with j dimensions, with Rij denotes a matrix with i rows and j columns. Multi-output regression aims to predict an output vector Y∈ Rij from a given input vector X∈ Rmn. More explicitly, a multi-input multi-output regression problem can be constructed as the learning of a mapping from Rij to Rmn. This relationship in multi-output least squares support vector regression is depicted in Figure 2.

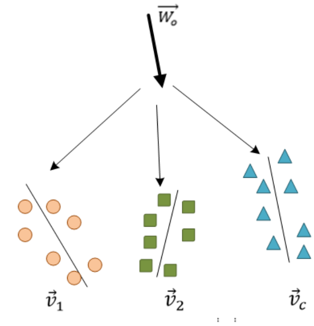

***Figure 2.** Illustration of the Underlying MIMO.*

All $w_i \in R^{rh}$ can be written as $w_i = w_o + v_i$, i ∈N, the mean vector $w_o \in R^{rh}$ is“small” [6]. Thus, the learning objective of multi-output least squares support vector regression can be transformed into the optimization problem as follows.

Where R^n ~> R^{rh} is a mapping to some higher-dimensional Hilbert space H with rh dimensions. Hence, the objective function with constraints can be minimized.

$$\min_{w_o \epsilon Rrh, V \epsilon Rrh, b\, \epsilon Rc} (w_o, V, \Xi) = \frac{1}{2}(w_o^T w_o) + \frac{1}{2}\frac{\lambda}{c}\,\mathrm{trace}(V^T V) + \gamma\frac{1}{2}\,\mathrm{trace}(\Xi^T \Xi), \tag{4}$$

$$\text{s.t. } Y = Z^T W + \mathrm{repmat}\,(b^T, r, 1) + \Xi, \tag{5}$$

where

$\Xi = (\xi_1, \xi_2, ..., \xi_c) \epsilon R^{r\,x\,c}$, as a slack variable

$W = (w_0 + v_1, w_0 + v_2, ..., w_0 + v_c) \in R^{rh\,x\,c}$

$Z = (\varphi(x_1), \varphi(x_2), ..., \varphi(x_l)) \in R^{n_h\,x\,l}, \lambda, \gamma \in R$ as two positive, real regularized parameters.

The Lagrangian function for the problem (3), (4) is formulated as follows.

$$L\,(w_0, V, b, \Xi, A) = J\,(w_0, V, \Xi) - \mathrm{trace}\left(A^T\left(Z^T W + \mathrm{repmat}\,(b^T, r, 1) + \Xi - Y\right)\right), \tag{6}$$

where $A = (\alpha_1, \alpha_2, ..., \alpha_c) \in R^{r\,x\,c}$ is a matrix that consists of Lagrange multipliers.

After the optimization, the corresponding decision function for the multiple outputs becomes,

$$F\,(x) = \varphi\,(x)^T W^* + b^{*T} = \varphi\,(x)^T \mathrm{repmat}\,(w_0^*, 1, c) + \varphi\,(x)^T V^* + b^{*T} \tag{7}$$

2.4. MIMO-S-LSSVR model

A feature combine process can be used to remove historical data point and apply slide-step predictive result to optimize learning data set, thereby improving both efficiency and accuracy. Setting 80% data point for training and 20% for test model. Particle swarm optimization algorithm (PSO) [8], a population-based stochastic global optimization technique, is used to optimize the parameters of MIMO-SLLVR model. C and σ setting range are [10^{-3}; 10^{12}] and [10^{-3}; 10^{3}], respectively. The population size is 50, max generation is 25 and PSO learning parameters c_1 and c_2 are 2.05.

3. Performance validation with the previous results

In this section, the performance of the proposed model is compared with the models in the previous studies.

In 2008, by presenting the effective evidence in evaluating time series forecasts in interval-valued data, Maia [9] marked a scientific forecasting revolution in time series prediction. The proposed model was used to forecast a financial interval time series of the same data set and horizon as [9]. The experiment result in Table 2 shows that the proposed model gets the best in the test with the daily stock prices of the Brazilian Petroleum Company SA which is used from 2005 to 2006 year with 484 intervals that represent the highest and lowest daily stock prices.

Table 2. *MIMO-S-LSSVR evaluation vs. previous models based on MAE and MSE value*

Model	MAE (USD)		MSE (USD)	
	X_U	X_L	X_U	X_L
Autoregressive	0.9944	1.0760	1.7791	2.1185
ARIMA [9]	0.9758	1.0610	1.7248	2.0941
ANN [9]	0.9486	1.0004	1.5439	1.7922
Hybrid [9]	0.9210	0.8964	1.3566	1.6998
MIMO-S-LSSVR	**0.4276**	**0.5017**	**0.3305**	**0.7241**

4. Case study results and discussion

Setting an embedding dimension or lag is the most important in the sliding-window algorithm. In this paper, the optimal lag in the model is determined by a sensitivity analysis. The result shows that the optimal lag of S&P500 and VN100 dataset is 4 days, respectively. Total data points is 754 from January 2016 to December 2018. The performance of predictive models is verified by the root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and correlation coefficient (R). Table 3 indicates two experimental data sets test results.

Table 3. *Performance evaluates the predictive model of S&P500 and VN100 daily stock prices data sets*

	Stock	Interval	MIMO-S-LSSVR			
			RMSE (VND)	MAE (VND)	MAPE	R
1-day ahead	S&P500	Upper bound	33.44	24.16	0.89%	0.90
		Lower bound	45.08	32.44	1.21%	0.87
	VN100	Upper bound	15.51	11.63	1.30%	0.83
		Lower bound	18.39	13.56	1.55%	0.77
3-days ahead	S&P500	Upper bound	54.11	41.79	1.51%	0.76
		Lower bound	65.12	52.04	1.92%	0.71
	VN100	Upper bound	17.53	12.99	1.45%	0.78
		Lower bound	20.52	15.23	1.73%	0.72

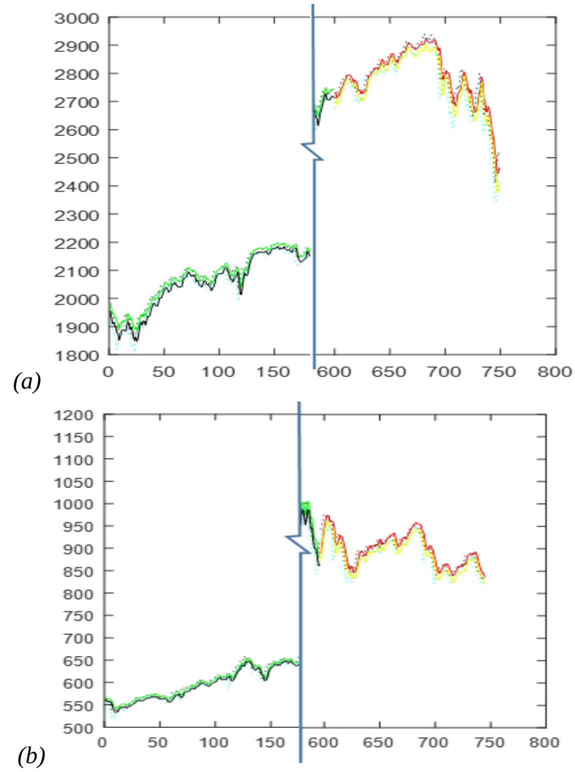

Figure 3. *Historical and predicted values 1-day ahead validation. (a) S&P500 stock price. (b) VN100 stock price*

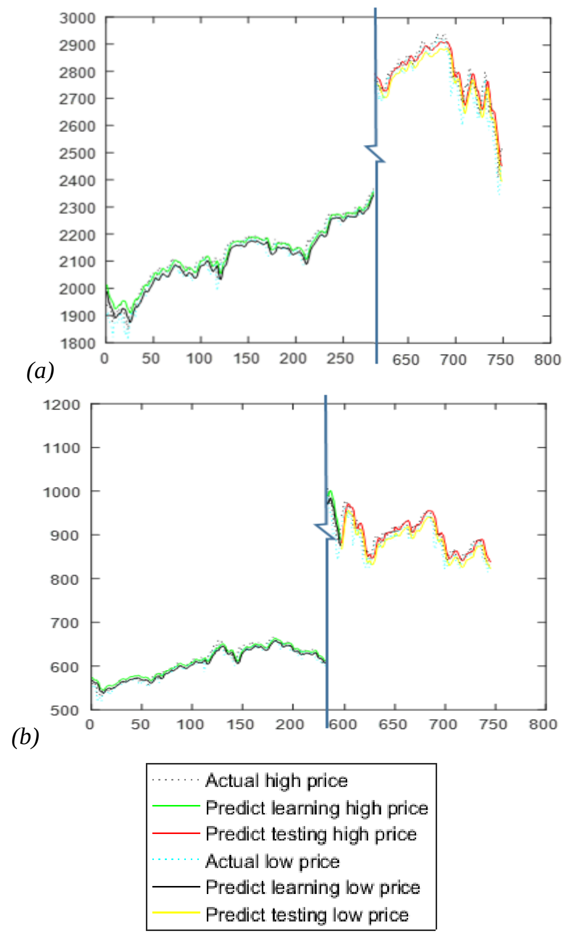

Figure 4. *Historical and predicted values 3-day ahead validation. (a) S&P500 stock price. (b) VN100 stock price*

After being verified in the low horizon with the multivariable model of Maia, the proposed model is tested in the same horizon prediction, H=1. The proposed model shows high performance in 1-day ahead forecast with upper bound and lower bound MAPE of S&P500 and VN100 index is 0.89%, 1.21% and 1.30%, 1.55% respectively.

In another case, the model is used to predict in higher horizon, H=3, The proposed model shows high performance in 1-day ahead forecast with upper bound and lower bound MAPE of S&P500 and VN100 index is 1.51%, 1.92% and 1.45%, 1.73% respectively.

The positive results show that MIMO-S-LSSVR is the highly expect model in time series forecasting, especially in interval-valued time series analysis.

5. Conclusion and future work

The interval-valued time series attracts attention in various fields because the information in ITS includes demands fields, economic, political and investment trends. This work has developed an expert multi-output technique for evaluating and predicting the interval-valued time series. Moreover, the proposed sliding-window method is useful for handling time series to optimize data size.

Additionally, phase space reconstruction by sliding-window metaheuristic optimization improves input performance for the nonlinear time series. PSO is used to optimize the parameters of the proposed model, which is aimed to improve the prediction accuracy.

This study is conducted only on the financial market data sets. The remaining problem comes from the necessary self-tuning setting of these parameters of the model. In the future, creating a self-tuning system is needed to train model on the different varieties of time series data sets.

Aknowledgement: This research is funded by University of Technology and Education – The University of Danang under project number T2019-06-128.

REFERENCES

[1] J. S. Chou and A. D. Pham, "Smart artificial firefly colony algorithm-based support vector regression for enhanced forecasting in civil engineering", *Computer-Aided Civil and Infrastructure Engineering*, vol. 30, pp. 715-732, 2015.

[2] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, and S. Mishra, "Decision tree and SVM-based data analytics for theft detection in smart grid", *IEEE Transactions on Industrial Informatics*, vol. 12, pp. 1005-1016, 2016.

[3] A. Oztekin, R. Kizilaslan, S. Freund, and A. Iseri, "A data analytic approach to forecasting daily stock returns in an emerging market", *European Journal of Operational Research*, vol. 253, pp. 697-710, 2016.

[4] C. Cortes and V. Vapnik, "Support-vector networks", *Machine learning*, vol. 20, pp. 273-297, 1995.

[5] S. J. AK, *Least squares support vector machines*: World Scientific, 2002.

[6] S. Xu, X. An, X. Qiao, L. Zhu, and L. Li, "Multi-output least-squares support vector regression machines", *Pattern Recognition Letters*, vol. 34, pp. 1078-1084, 2013.

[7] R.-P. Mundani, J. Frisch, V. Varduhn, and E. Rank, "A sliding window technique for interactive high-performance computing scenarios", *Advances in Engineering Software*, vol. 84, pp. 21-30, 2015.

[8] X.-S. Yang, *Engineering optimization: an introduction with metaheuristic applications*: John Wiley & Sons, 2010.

[9] A. L. S. Maia, F. d. A. de Carvalho, and T. B. Ludermir, "Forecasting models for interval-valued time series", *Neurocomputing*, vol. 71, pp. 3344-3352, 2008.

(The Board of Editors received the paper on 25/9/2019, its review was completed on 09/03/2020)