Jurnal Teknologi

Full paper

Time Series Forecasting using Least Square Support Vector Machine for Canadian Lynx Data

Shuhaida Ismail*, Ani Shabri

Department of Mathematics, Science Faculty, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia
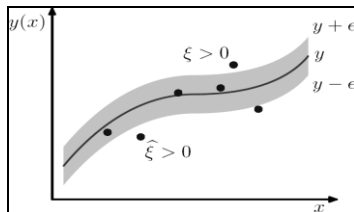
*Corresponding author: ismail.shuhaida@gmail.com

Graphical abstract

Abstract

Time series analysis and forecasting is an active research area over the last few decades. There are various kinds of forecasting models have been developed and researchers have relied on statistical techniques to predict the future. This paper discusses the application of Least Square Support Vector Machine (LSSVM) models for Canadian Lynx forecasting. The objective of this paper is to examine the flexibility of LSSVM in time series forecasting by comparing it with other models in previous research such as Artificial Neural Networks (ANN), Auto-Regressive Integrated Moving Average (ARIMA), Feed-Forward Neural Networks (FNN), Self-Exciting Threshold Auto-Regression (SETAR), Zhang's model, Aladang's hybrid model and Support Vector Regression (SVR) model. The experiment results show that the LSSVM model outperforms the other models based on the criteria of Mean Absolute Error (MAE) and Mean Square Error (MSE). It also indicates that LSSVM provides a promising alternative technique in time series forecasting.

Keywords: Time series forecasting; support vector regression; least square support vector machine; canadian lynx data

Abstrak

Analisis dan ramalan siri masa adalah kawasan penyelidikan yang aktif sejak beberapa dekad yang lalu. Terdapat pelbagai jenis model ramalan yang telah dibangunkan dan penyelidik telah bergantung pada teknik statistik untuk meramalkan masa depan. Kertas ini membincangkan penggunaan model Kurang Persegi Mesin Sokongan Vektor (LSSVM) untuk peramalan Lynx Kanada. Objektif kertas ini adalah untuk mengkaji fleksibiliti LSSVM dalam peramalan siri masa dengan membandingkannya dengan beberapa model lain dalam kajian sebelumnya seperti Rangkaian Neural Buatan (ANN), Auto-Regresif Bersepadu Purata Bergerak (ARIMA), Rangkaian Neural Suap-Kehadapan (FNN), Auto-Regresif Ambang Sendiri-Menarik (SETAR), Model Zhang, model hibrid Aladang dan model Sokongan Vektor Regresi (SVR). Keputusan eksperimen menunjukkan bahawa model LSSVM melebihi performa model lain berdasarkan kriteria Ralat Min Mutlak (MAE) dan Ralat Min Kuasa Dua (MSE). Ia juga menunjukkan bahawa LSSVM menyediakan teknik alternatif yang memberangsangkan dalam peramalan siri masa.

Kata kunci: Peramalan siri masa; sokongan vektor regresi; kurang persegi mesin sokongan vektor; data lynx Kanada

■1.0 INTRODUCTION

Time series forecasting, or time series prediction, takes an existing series of data and forecasts the data values. The goal is to observe or model the existing data series to enable future unknown data values to be forecasted accurately. The accuracy of time series forecasting is fundamental to many decision processes and hence the research for improving the effectiveness of forecasting models has never been stopped[11]. The reason that forecasting is so important is that prediction of future events is a critical input into many types of planning and decision making. In time to time, time series forecasting becomes an active research area over the last few decades. Various kinds of forecasting models have been developed and researchers have relied on statistical techniques to predict time series data.

Artificial neural network (ANN) has found increasing consideration in forecasting theory, leading to successful applications in various forecasting domains including economic and many more. In the last decade, ANN is being used more frequently in the analysis of time series forecasting, pattern classification and pattern recognition capabilities[8,12]. ANN provides an attractive alternative tool for both forecasting researchers and has shown their nonlinear modeling capability

in data time series forecasting[15]. The ANN can learn from examples (pass data), recognize a hidden pattern in historical observations and use them to forecast future values. In addition to that, they are able to deal with incomplete information or noisy data and can be very effective, especially in situations where it is not possible to define the rules or steps that lead to the solution of a problem. However, there are some disadvantages of the ANN. The network structure is hard to determine and it is usually determined by using a trial–and–error approach[7].

Other than ANN, the Support Vector Machine (SVM) model which was first suggested by Vapnik[10], have become state–of–the–art tools for linear and nonlinear input–output knowledge discovery[10,17] and it can be employed for solving pattern recognition, regression estimation problems, data mining, classification, and time series forecasting[9,11]. Several studies showed that SVM is a powerful methodology and has become the most wanted in studies due to ability to solve most nonlinear regression and time series problem. The ability of SVM to solve nonlinear regression estimation problems makes SVM successful in time series forecasting. SVM are very specific types of learning algorithms characterized by the capacity control of the decision function and the use of the kernel function. Established on the unique theory of the structural risk minimization principle to estimate a function by minimizing an upper bound of the generalization error, SVM is shown to very resistant over-fitting problem, eventually achieving high generalization performance in solving various time series forecasting problems[9].

An SVM are divided into two types: SVM for classification or also known as SVC and SVM for regression also known as SVR. SVR are applied to solve the regression problems with the introduction of an alternative loss function[13,17]. Detailed discussions of SVMs and SVRs have been given in several literature[13,16,17,18]. Another key property of SVM is that training SVM is equivalent to solving a linearly constrained quadratic programming problem so that the solution of SVM is always unique and globally optimal, unlike other networks' training, which requires non-linear optimization with the danger of being stuck into local minima[3]. The numerical results indicated that the SVM is superior to the multi-layer back propagation neural network in financial time series forecasting[9]. However, the standard SVM is solved using complicated quadratic programming methods, which are often time consuming and has higher computational burden because of the required constrained optimization programming.

On the other hand, LSSVM is a modification from existing SVM, which has been successfully applied to solve various problems, among others in data mining, classification, regression and time series forecasting[3,11]. The ability of SVM to solve nonlinear regression estimation problems makes SVM successful in time series forecasting. The LSSVM's reformulation greatly simplifies the problem in such a way that the solution is characterized by a linear system, more precisely a Karush-Kuhn-Tucker (KKT) liner system, which takes a similar form as the linear system that one solves in every iteration step by interior point methods for standard SVM.

LSSVM encompasses similar advantages as SVM, but its additional advantages is that it requires solving a set of only linear equations, which is much easier and computationally more simple. The method uses equality constraints instead of inequality constraints and adopts the least squares linear system as its loss function, which is computationally attractive. An LSSVM also has better convergence and high precision. Hence, this method is easier to use than quadratic programming solvers in SVM method. Extensive empirical studies[17] have shown that

LSSVM is comparable to SVM in terms of generalization performance.

In this paper, LSSVM model is proposed in order to improve the accuracy of time series forecasting. With the capability of LSSVM, the proposed model is expected to be useful for time series forecasting. The prediction results of LSSVM model are compared with others forecasting models developed by the previous researcher such as Aladang, Kajitani, Khashei and Zhang[2,5,6,11].

■2.0 FORECASTING MODELS

This section, we present the SVR and LSSVM models used in Canadian lynx forecasting. The choice of these models in this study was because these models have been widely and successfully used in time series forecasting.

2.1 Support Vector Regression (SVR)

SVR is closely related to SVM classifiers in terms of theory and implementation. The loss function must be modified to include a distance measure. The regression can be linear and non-linear. Similar to classification problems, a nonlinear model is usually required to adequately model data. In the same manner as the non-linear SVC approach, a non-linear mapping can be used to map the data into a high dimensional feature space where linear regression is performed.

Vapnik[10,14] introduced the ε-insensitive zone in the error lost function. From a theoretical point of view, this zone represents the degree of precision at which the bounds on the generalization ability apply. Training vectors that lie within the zone are deemed correct, whereas those outside this zone are deemed incorrect and contribute to the error loss function. These incorrect vectors become the support vectors (see Figure 1). Vectors lying on and outside the dotted lines are support vectors, whereas those within the ε-insensitive zone are not important in terms of the regression function. The regression surface then can be determined only by support vectors.
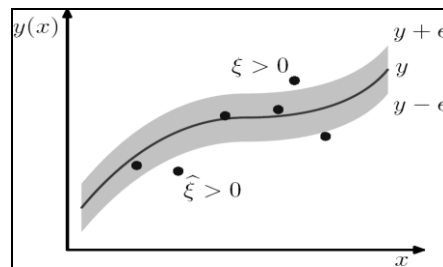

Figure 1 One-dimensional non-linear regression with epsilon intensive band

Fundamentally, SVR is linear regression in the feature space. Although it is simple and not very useful in real-world situations, it forms a building block for understanding complex SVRs. The goal of SVR is to find a function $f(x)$ that deviates not more than ε from the targets y_k for all the training data, and at the same time, is as flat as possible. Let linear function $f(x)$ takes the form:

$$f(x) = w^T x + b$$

The optimization problem in the primal weight space becomes:

$$\min \frac{1}{2} w^T w + C \sum_{k=1}^{N} (\xi_k + \xi_k^*)$$

subject to

$$\begin{cases} y_k - w^T \varphi(x_k) - b \le \varepsilon + \xi_k \\ w^T \varphi(x_k) + b - y_k \le \varepsilon + \xi_k^* \\ \xi_k, \xi_k^* \ge 0 \qquad k = 1,...,N \end{cases}$$

This constrained optimization problem is solved using the primal lagrangian form:

$$\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{k=1}^{N} (\alpha_k - \alpha_k^*) \varphi(x_k)$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow w = \sum_{k=1}^{N} (\alpha_k - \alpha_k^*) = 0$$

$$\frac{\partial L}{\partial \xi_i} = 0 \rightarrow c - \alpha_k - \eta_k = 0$$

$$\frac{\partial L}{\partial \alpha_i} = 0 \rightarrow c - \alpha_k^* - \eta_k^* = 0$$

Here the kernel trick has been applied with $K(x_k, x_l) = \varphi(x_k)^T \varphi(x_l)$ or $k, l = 1,..., N$. Kernal trick is a technique to write a nonlinear operator as a linear one in a space of higher dimension. The dual representation of the model becomes the following equation, where α_k, α_k^* are the solution to the QP problem.

$$f(x) = \sum_{k=1}^{n} (\alpha_k - \alpha_k^*) K(x, x_k) + b$$

2.2 Least Square Support Vector Regression

Least Square Support Vector Machine (LSSVM) is a modification of the standard Support Vector Machine (SVM) was develops by Suykens and Vandewalle[14]. The basic LSSVM is used for the optimal control of non-linear Karush-Kuhn-Tucker systems for classification as well as regression.

Consider a set data $D = \{(x_1, y_1), (x_2, y_2),...,(x_n, y_n)\}$, $x_i \in \Re^p$, $y_i \in \Re$, x is the input vector, y is the expected output and n is the number of data. The LSSVM has been developed to find the optimally non-linear regression function

$$y(x) = w^T \varphi(x) + b \qquad (1)$$

By combining the functional complexity and fitting error, the optimization problem of LSSVM is given as:

$$\text{Min } J(w, \xi) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{i=1}^{n} \xi_i^2 \qquad (2)$$

Such that :

$$y(x) = w^T \varphi(x_i) + b + \xi_i$$
$$i = 1, 2, 3, ..., n \qquad (3)$$

This formulation consists of equality instead of inequality constraints. To solve this optimization problem, Lagrange function is constructed as

$$L(w, b, \xi; \alpha) = J(w, b, \xi) -$$
$$\sum_{i=1}^{l} \alpha_i \{ w^T \varphi(x_i) + b - y_i + \xi_i \}$$

$$(4)$$

where α_i are the Langrange multipliers, which can be positive or negative. The solution of (4) can be obtained by partially differentiating with respect to w, b, ξ_i and α_i

$$\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{l} \alpha_i \varphi(x_i)$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{l} \alpha_i = 0$$

$$\frac{\partial L}{\partial \xi_i} = 0 \rightarrow \alpha_i = \gamma \xi_i$$

$$\frac{\partial L}{\partial \alpha_i} = 0 \rightarrow w^T \varphi(x_i) + b - y_i + \xi_i = 0$$

$$(5)$$

After elimination of the variables w and ξ_i one obtains the following matrix solution.

$$\begin{bmatrix} 0 & 1_v^T \\ 1_v^T & \Omega + \frac{1}{\gamma} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \qquad (6)$$

with $y = [y_1, y_2,..., y_l], 1_v^T = [1, 1,..., 1]$, $\alpha = [\alpha_1, \alpha_2,..., \alpha_l]$ and Mercer's condition is applied within the Ω matrix;

$$\Omega_{ij} = y_i y_j \varphi(x_i)^T \varphi(x_j)$$
$$= y_i y_j K(x_i, x_j) \qquad (7)$$

The fitting function namely the output of LSSVM Regression is,

$$y(x) = \sum_{i=1}^{n} \alpha_i K(x_i, x) + b \qquad (8)$$

where α_i, b are the solutions to the linear system and $K(x_i, x_j)$ is a kernel function. The most popular kernel function is Radial Basis Function (RBF)[20], as shown in Equation (9).

$$K(x_i, x_j) = \exp \left(- \frac{\|x - x_i\|^2}{\delta^2} \right)$$

$$(9)$$

◼3.0 DATASET

The lynx data sets are one of the most frequently used time series. The lynx series contains the number of lynx trapped, per year in the Mackenzie River, Northern Canada. The data show a periodicity of approximately 10 years, corresponding from 1821–1934 containing 114 observations (see Figure 2). For the experiment's purpose, we divided the whole data into a training

set containing 100 data points and the testing set containing the rest of 14 data points. The training set is used for model development while the testing set is used to evaluate and established the model.
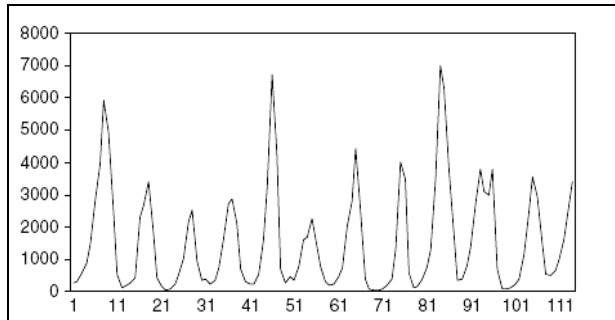

Figure 2 Canadian lynx data series (1821–1934)

■**3.0 PERFORMANCE CRITERIA**

The performances of the each model for both the training data and forecasting data are evaluated and were selected according to the mean absolute error (MAE) and mean square error (MSE), which are widely used for evaluating results of time series forecasting. The MAE and MSE are defined as

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left|y_t - \hat{y}_t\right|$$

$$MSE = \frac{1}{N}\sum_{t=1}^{N}\left(y_t - \hat{y}_t\right)^2$$

where y_t and $\hat{y}_t$ are the observed and the forecasted rice yields at the time *t*. The criterions to judge for the best model are relatively small of MAE and MSE in the modeling and forecasting.

■**4.0 APPLICATION**

In this study, RBF was employed as a kernal for both of the models. Dibike[20] employed some diverse kernel functions for their modeling and demonstrated that the RBF kernel has superior efficiency than other kernel[4]. The advantage of the RBF kernel is that it nonlinear maps the training data into a possibly infinite dimensional space. This can effectively handle situations when predictors and predicted is non-linear and is computationally simple than the polynomial kernel. Therefore, RBF kernel was chosen in this study because it performs well under general smoothness assumptions.

4.1 Testing Using SVR

It is well known that SVR generalization performance (estimation accuracy) depends on a good setting of hyper-parameters C, ε and the kernel parameters. The problem of optimal parameter selection is further complicated by the fact that SVR model complexity (and hence its generalization performance) depends on all three parameters. Existing software implementations of SVM regression usually treat SVM hyper-parameters as user-defined inputs. In order to better evaluate the

performance of the proposed approach, the parameters C, ε, and γ in the range of search was set up to [1, 10] at increment of 1.0 for C, and [0.1, 0.5] at increment of 0.1 for ε, with γ fixed as 0.5. For each hyper-parameter pair (C, ε) in the search space, 10-fold cross validation on the training set is performed to predict the prediction error where it will repeated ten times to increase the reliability of the results.

4.2 Testing Using LSSVM

In order to better evaluate the performance of the proposed approach, we consider a grid search of (γ, σ^2) with γ in the range 10 to 1000 and σ^2 in the range 0.01 to 1.0. For each hyper-parameter pair (γ, σ^2) in the search space, 10-fold cross validation on the training set is performed to predict the prediction error.

■**5.0 RESULT AND COMPARISON**

In this section, the predictive capabilities of the proposed model are compared with artificial neural networks (ANNs), auto-regressive integrated moving average (ARIMA), and Zhang's hybrid ANNs/ARIMA model[11], Khashei's ANN model[6], Aladang's hybrid Elman's Recurrent Neural Networks (ERNN) and ARIMA model[2] and Kajitani's (Self-Exciting Threshold Autoregression) SETAR and Feed- Forward Neural Networks (FNN) model[5] using well-known real data sets: the Canadian lynx data. The mean absolute error (MAE) and Mean Squared Error (MSE), which are widely used for evaluation of time series forecasting's result were used as a performance measurement in this study. The usage of MAE and MSE in this study are suitable as the previous studies were conducted using the same evaluation performances. The MAE and MSE values for the last 14 observations of the models are summarized in Table 1.

By considering these results, our proposed model has yielded a slightly better result, were the MSE and MAE values of the proposed model is the smallest among others. Thus, the SVR model results perform slightly better than using other models. Figure 3 showed the forecasting values for the last 14 observations used in this study. The solid line represents the actual time series data while the dot line represents the forecast values.

Table 1 The comparison of the performance of SVR with other forecasting models of Canadian Lynx Data

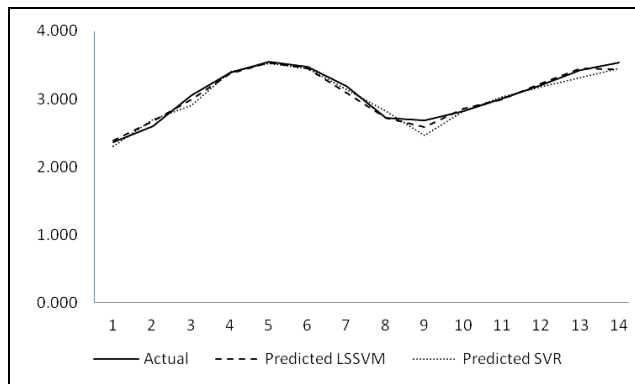| Model log10 (lynx) Max=3.844539, F=14	MSE	MAE
Zhang' ARIMA	0.0205	0.1123
Zhang' ANN	0.0205	0.1121
Zhang' Hybrid	0.0172	0.1040
Khashei & Bijari' ANN	0.0136	0.0896
Kajitani' SETAR	0.0140	-
Kajitani' FNN	0.0090	-
Aladag' Hybrid	0.0090	-
SVR	0.0085	0.0746
LSSVM	0.0030	0.0418

Figure 3 The SVR prediction of Canadian lynx

◼6.0 CONCLUSION

Time series forecasting is very important for future forecasting because it is involved in the decision making process. Previous data and information need in order to forecast the future. In summary, the main steps for forecasting are to analyze the historical or past time series data to identify the patterns that can be used. This pattern was further expanded to provide a forecast.

In this paper, we present a LSSVM model for Canadian lynx forecasting. From the experimental results comparing the performance of SVR, LSSVM and seven other models done by previous researchers, it indicates that LSSVM significantly outperform other models and it can be concluded that LSSVM provides an alternative technique for time series forecasting. In the future work, we hope to increase the forecasting accuracy by employing the forecasting and clustering models together.

Acknowledgement

References

[1] S. Ani. 2001. Comparision of Time Series Forecasting Methods Using Neural Networks and Box-Jenkins Model. *Matematika.* 17: 1–6.
[2] C. H. Aladag, E. Egrioglu and C. Kadilar. 2009. Forecasting Nonlinear Time Series with a Hybrid Methodology. *Applied Mathematics Letters.* 22: 1467–1470.
[3] L. Cao. 2003. Support Vector Machines Experts For Time Series Forecasting. *Neurocomputing.* 51: 321–339.
[4] Y. B. Dibike, S. Velickov, D. P. Solomatine, and M. B. Abbott. 2001. Model Induction with Support Vector Machines: Introduction and Applications. *ASCE Journal of Computing in Civil Engineering.* 15(3): 208–216.
[5] Y. Kajitani, W. H. Keith, and A. I. Mcleod. 2005. Forecasting Nonlinear Time Series With Feed-Forward Neural Networks: A Case Study of Canadian Lynx Data. *Journal of Forecasting.* 24: 105–117.
[6] M. Khashei, and M. Bijari. 2010. An Artificial Neural Network (p,d,q) Model for Timeseries Forecasting. *Expert Systems with Applications.* 37. 479–489.
[7] O. Kisi. 2004. River Flow Modeling Using Artificial Neural Networks. *Journal of Hydrologic Engineering.* 9(1): 60–63.
[8] R. Sharda. 1994. Neural Networks for the MS/OR Analyst: An Application Bibliography. *Interfaces.* 24(2): 116–30.
[9] F. E. H. Tay and L. J. Cao. 2001. Improved Financial Time Series Forecasting by Combining Support Vector Machines with Self-Organizing Feature Map. *Intelligent Data Analysis.* 5: 339–354.
[10] V. Vapnik. 1995. *The Nature of Statistical Learning Theory.* Springer Verlag, Berlin.
[11] G. P. Zhang. 2003. Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. *Neurocomputing.* 50: 159–175.
[12] H. F. Zou, G. P. Xia, F. T. Yang, and H. Y. Wang. 2007. An Investigation and Comparison of Artificial Neural Network and Time Series Models for Chinese Food Grain Price Forecasting. *Neurocomputing.* 70: 2913–2923.
[13] N. Cristianini, and J. Shawe-Taylor. 2000. *An Introduction To Support Vector Machines and Other Kernel Based Learning Methods.* Cambridge, Cambridge University Press.
[14] J. A. K. Suykens, T. V. Gestel. 2005. *Least Square Support Vector Machine.* New Jersey, World Scientific.
[15] C. L. Wu, K. W. Chau, *et al.* 2008. River Stage Prediction Based on a Distributed Support Vector Regression. *Journal of Hydrology.* 358(1–2): 96–111.
[16] H. Wang, and D. Hu. 2005. Comparison Of SVM And LSSVM For Regression. International Conference on Neural Networks and Brain. 1: 279–283
[17] A. J. Smola and B. Scholkopf. 2004. A Tutorial on Support Vector Regression. *Statistics and Computing.* 14: 199–222.
[18] C. J. C. Burges. 1998. A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery. 2: 121–167.
[19] M. T. Gencoglu and M. Uyar. 2009. Prediction of Flashover Voltage of Insulators Using Least Squares Support Vector Machines. *Expert Systems With Applications.* 36: 10789–10798.
[20] Y. B. Dibike, and D. P. Solomatine. 2001. River Flow Forecasting Using Artificial Neural Networks. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere.* 26(1): 1–7.