Least Square Support Vector Machines as an Alternative Method in Seasonal Time Series Forecasting

Ani Shabri

Department of Mathematical Sciences, Faculty of Science
Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia

Abstract

The least square support vector machines (LSSSVM) model is a novel forecasting approach and has been successfully used to solve time series problems. However, the applications of LSSVM model in a seasonal time series forecasting has not been widely investigated. This study aims at developing a LSSVM model to forecast seasonal time series data. To assess the effectiveness of this model, the airline passenger series exhibits nonlinear behaviour and shows multiplicative seasonal behaviour was applied. In order to obtain the optimal model parameters of the LSSVM, a grid search algorithm and cross-validation method were employed. In this study, seasonal autoregressive integrated moving average (SARIMA) and artificial neural network (ANN) models are employed for forecasting the same data sets. Empirical results indicate that the LSSVM yields well forecasting performances. Thus, the LSSVM model provides a promising alternative for seasonal time series forecasting.

Keywords: ARIMA, LSSVM, forecasting, time series, seasonal

1 Introduction

Seasonality is especially observable in the economic and business time series data. Most time series consist of trends and seasonal effects. However, the seasonal time series is a complex and nonlinear problem. Improving accuracy in time series forecasting is an important yet often difficult task. Much effort has
been devoted over the past several decades to the development and improvement of time series forecasting models.

One of the best-known approaches in the development of time series model is the seasonal auto-regressive integrated moving average (SARIMA) model. The SARIMA model is one of the most popular approaches in seasonal time series forecasting owing to its statistical properties, as well as the well-known Box–Jenkins methodology used for constructing the model. The SARIMA model has been successfully utilized in many fields of forecasting such as in economic, engineering, foreign exchange, stock market and social [8]. Although the SARIMA model has been highly successful in both academic research and industrial application during the past three decades, it suffers from a major limitation owing to its pre-assumed linear form of the model.

Recently, artificial neural network (ANN) model has been extensively studied and also used as an alternative in forecasting seasonal data pattern [17]. Some literatures indicated that ANN can obtain desirable results in seasonal and trend forecasting [4, 11, 15]. While some researchers claim that ANN is unsuccessful in finding out seasonal effect in the data structure [6, 9, 17].

Suykens et al. [8] proposed a modified version of support vector machines (SVM) called least squares support vector machines (LSSVM). In recent years, LSSVM extended to cope with forecasting problems, and has been used successfully in various areas of pattern recognition and time series forecasting problems [5, 7, 13, 16]. However, applications of the LSSVM models in seasonal time series data have not been widely studied. Therefore, this study attempts to develop a LSSVM model in the seasonal time series forecasting problems.

2 Methodology

2.1 The Autoregressive Integrated Moving Average

The ARIMA models introduced by Box and Jenkins [1], has been one of the most popular approaches to the analysis of the time series forecasting. The general components of an ARIMA models consist of seasonal and non-seasonal parts, which are commonly known as SARIMA models. The form of SARIMA($p,d,q)(P,D,Q)_s$ model can be written as

\[ \phi_p(B)\Phi_p(B^s)(1 - B)^d(1 - B^s)^D x_i = \theta_q(B)\Theta_q(B^s)a_i \]  

where $\phi(B)$ and $\theta(B)$ are polynomials of order $p$ and $q$, respectively; $\Phi(B^s)$ and $\Theta(B^s)$ are polynomials in $B^s$ of degrees $P$ and $Q$, respectively; $p$ order of non-seasonal auto regression; $d$ number of regular differencing; $q$ order of the non-seasonal moving average; $P$ order of seasonal auto regression; $D$ number of seasonal differencing; $Q$ order of seasonal moving average; and $s$ length of season. Random errors, $a_i$ are assumed to be independently and identically distributed with a mean of zero and a constant variance $\sigma^2$. A more detailed description of
the ARIMA models is published by Box and Jenkins [1] and thus, descriptions of these methods are not given here.

2.2 The Least Square Vector Machines Model

The LSSVM is a new version of SVM modified by Suykens et al. [14]. LSSVM involves the solution of a quadratic optimization problem with a least squares loss function and equality constraints instead of inequality constraints. Consider a training sample set \((x_i, y_i)\) with input \(x_i \in \mathbb{R}^n\) and output \(y_i \in \mathbb{R}\). In feature space SVM models take the form

\[
y(x) = w^T \phi(x) + b
\]  

(2)

where the nonlinear mapping \(\phi(x)\) maps the input data into a higher dimensional feature space. LSSVM introduces a least square version to SVM regression by formulating the regression problem as

\[
\min R(w, e) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{i=1}^{n} e_i^2
\]  

subject to the equality constraints

\[
y(x) = w^T \phi(x_i) + b + e_i, \quad i = 1, 2, ..., n
\]  

(4)

To solve this optimization problem, Lagrange function is constructed as

\[
L = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{i=1}^{n} e_i^2 - \sum_{i=1}^{n} \alpha_i \{w^T \phi(x_i) + b + e_i - y_i \}
\]  

(5)

where \(\alpha_i\) is Lagrange multipliers. The solution of (5) can be obtained by partially differentiating with respect to \(w, b, e_i\) and \(\alpha_i\)

\[
\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{n} \alpha_i \phi(x_i)
\]  

(6)

\[
\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{n} \alpha_i = 0
\]  

(7)

\[
\frac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i = \gamma e_i
\]  

(8)
\[
\frac{\partial L}{\partial \alpha_i} = 0 \rightarrow w^T \phi(x_i) + b + e_i - y_i = 0, \quad i = 1, 2, ..., n \tag{9}
\]

then the weight \( w \) can be written as combination of the Lagrange multipliers with the corresponding data training \( x_i \).

\[
w = \sum_{i=1}^{n} \alpha_i \phi(x_i) = \sum_{i=1}^{n} \gamma e_i \phi(x_i) \tag{10}
\]

Putting the result of (10) into (2), then the following result is obtained:

\[
y(x) = \sum_{i=1}^{n} \alpha_i \phi(x_i)^T \phi(x_i) + b = \sum_{i=1}^{n} \alpha_i K(x_i, x) + b \tag{11}
\]

where a positive definite kernel is defined as follows:

\[
K(x_i, x) = \phi(x_i)^T \phi(x_i)
\]

The \( \alpha \) vector and \( b \) can be found by solving a set of linear equations:

\[
\begin{bmatrix}
0 & 1^T \\
1 & \phi(x_j)^T \phi(x_j) + \gamma^{-1} I
\end{bmatrix}
\begin{bmatrix}
b \\
\alpha
\end{bmatrix}
= \begin{bmatrix}
0 \\
y
\end{bmatrix}
\]

where \( y = [y_i; \ldots; y_n] \), \( 1 = [1; \ldots; 1] \), \( \alpha = [\alpha_1; \ldots; \alpha_n] \). This finally leads to the following LSSVM model for function estimation:

\[
y(x) = \sum_{i=1}^{n} \alpha_i K(x_i, x) + b \tag{12}
\]

where \( \alpha_i \), \( b \) are the solution to the linear system. Kernel function, \( K(x_i, x) \) represents the high dimensional feature space that is nonlinearly mapped from input space \( x \). Many works have demonstrated favorable performances of the Radial basis function (RBF) [10]. Therefore, RBF is used as the kernel function in this study. The RBF is given by \( K(x_i, x) = \exp(-\gamma \|x_i - x\|^2) \), where \( \gamma \) is the kernel parameters.

### 3 Indices of Performance Evaluation

In this study, one-step ahead forecasting is considered. In practice, short-term forecasting results are more useful as they provide timely information for the correction of forecasting value. In this study, three performance criteria such as
Least square support vector machines

means square error (MSE), mean absolute error (MAE) and correlation \((R)\) were used to compare LSSVM performance with ARIMA model or traditional methods. These criteria are given below:

\[
MSE = \frac{1}{n} \sum_{t=1}^{n} (x_t - \hat{x}_t)^2, \quad MAE = \frac{1}{n} \sum_{t=1}^{n} |x_t - \hat{x}_t|
\]

\[
R = \frac{\frac{1}{n} \sum_{t=1}^{n} (x_t - \bar{x})(\hat{x}_t - \bar{x})}{\sqrt{\frac{1}{n} \sum_{t=1}^{n} (x_t - \bar{x})^2} \sqrt{\frac{1}{n} \sum_{t=1}^{n} (\hat{x}_t - \bar{x})^2}}
\]

where \(x_t\) is the actual and \(\hat{x}_t\) is the forecasted value of period \(t\), and \(n\) is the number of total observation.

4 Case Study

In order to examine whether LSSVM gives better results or not, the data set of airline passengers from the literature was used. The airline passenger data set was first used by Brown [2] and then by Box and Jenkins [1]. The airline passenger data set consists of the total number (in thousands) of passengers on international airlines from January 1949 to December 1960. Fig. 1 shows that the data have an upward trend together with seasonal variation whose size is roughly proportional to the local mean level called multiplicative seasonality. The airline passenger series in its original form exhibits nonlinear behaviour and shows multiplicative seasonal behaviour. As in many other studies involving this time series, the data from the first 11 years (132 observations) are used for modelling and the 12 last observations are used for testing.

![Airline passenger data from 1949 to 1960](image)

The series involves taking natural logarithms of the data followed by seasonal \((D=1)\) and non-seasonal \((d=1)\) differencing to make the series stationary. The sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF) of stationary this series is plotted in Fig. 2.
Fig. 2. ACF and PACF plots used for the selection of ARIMA model for air passenger series

The ACF is damping out in exponential waves with significant spikes near lags 1 and 12. The PACF also has significance values at lag 1 and 12. This indicates a possible SARIMA($p, 1, q)(P, 1, Q)_s$ model with $p, q, P$ and $Q$ equal to 1 with $s = 12$. All combinations are tested to determine the best model out of these candidate models. The identification of best model for this series based on minimum AIC is shown in Table 1.

Table 1. Comparison of AIC for selected candidate models

<table>
<thead>
<tr>
<th>Model</th>
<th>$AIC$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SARIMA(1,1,1)(1,1,1)</td>
<td>-372.306</td>
</tr>
<tr>
<td>SARIMA(1,1,0)(1,1,0)</td>
<td>-370.904</td>
</tr>
<tr>
<td>SARIMA(1,1,0)(0,1,1)</td>
<td>-374.686</td>
</tr>
<tr>
<td>SARIMA(0,1,1)(1,1,0)</td>
<td>-371.454</td>
</tr>
<tr>
<td>SARIMA(0,1,1)(0,1,1)</td>
<td><strong>-376.732</strong></td>
</tr>
</tbody>
</table>

Bold values indicate the best AIC

The model finally selected was SARIMA(0,1,1)(0,1,1)$_{12}$. The residual ACF (RACF) and the PACF (RACF) of the best model are demonstrated in Fig. 3. The RACF and RPACF lie within the confidence limits, which clearly supports the fact that the residuals from the best model are white noise.

Fig. 3. The residual ACF (RACF) and the PACF (RACF) of the best model
One of the most important steps in developing a satisfactory LSSVM model is the selection of the input variables. A classical method in the selection of the input variables from a time series by using a set of time delayed samples of series \([1, 2, \ldots, 3, 2, \ldots, k] \). Using this method, the number of nodes in the input layer is equal to the number of delays or lagged variables \(x_{t-\tau}, x_{t-2\tau}, \ldots, x_{t-k\tau} \), where \(\tau \) is the time delay, and \(k \) is the number of chosen delays. The output, \(x_{t+p} \), is the predicted value of a time series defined as

\[
x_{t+p} = f(x_t, x_{t-\tau}, x_{t-2\tau}, \ldots, x_{t-k\tau})
\]

where \(P \) is a prediction time into the future. In this case, we gradually increased the number of lagged variables from 1 to 12 through exhaustive numerical simulation runs.

A new method to determine the number of nodes in the input layer is based on the Box-Jenkins analysis. Since one of the main goals of this work is to construct a LSSVM model for the airline passengers, therefore, the Box-Jenkins approach is used to discover which lagged or delayed variables are necessary to present as inputs to the LSSVM. Using the Box-Jenkins technique on the airline passengers time series presented in Fig.2, the best model found is SARIMA model of order \((0,1,1)\times(0,1,1)_{12}\) given by

\[
(1-B)(1-B^{12})x_t = (1-0.3405B)(1-0.6305B^{12})a_t
\]

The model can be written as

\[
x_t = x_{t-1} + x_{t-12} - x_{t-13} - 0.3405a_{t-1} - 0.6305a_{t-12} + 0.2147a_{t-13} + a_t
\]

For the airlines data, the output \(x_t\) can be defined as

\[
x_t = f(x_{t-1}, x_{t-12}, x_{t-13})
\]

All combinations of number of nodes are tested to determine the best input model for the LSSVM model. In the training and testing of LSSVM model, the same input structures of the data set from Box-Jenkins model were used. In order to obtain the optimal model parameters of the LSSVM, a grid search algorithm and cross-validation method were employed. In this study, a grid search of \((\gamma, \sigma^2)\) with \(\gamma \) in the range 10 to 1000 and \(\sigma^2 \) in the range 0.01 to 1.0 was conducted to find the optimal parameters. For each hyper parameter pair \((\gamma, \sigma^2)\) in the search space, 10-fold cross validation on the training set was performed to predict the prediction error.

Table 2 shows the performance results obtained in the testing period of the LSSVM approach. The best values of the MSE, MAE and R were obtained using
lag 1 and 12. For further analysis, the best performance of the SARIMA and LSSVM were compared with the best results studied by Faraway and Chatfield [3]. In Table 3, it shows that LSSVM has good performance and this model outperform SARIMA and ANN in terms of all the standard statistical measures for airlines series. The performance of LSSVM in predicting the air passenger is superior to classical SARIMA model. The results show that LSSVM model can be successfully applied to establish accurate and reliable in forecasting of seasonal time series.

Table 2. The MSE, MAE and R statistics of LSSVM model in test period

<table>
<thead>
<tr>
<th>Lags</th>
<th>MSE</th>
<th>MAE</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,12,13</td>
<td>380.804</td>
<td>15.123</td>
<td>0.9655</td>
</tr>
<tr>
<td>1,12</td>
<td><strong>206.393</strong></td>
<td><strong>12.800</strong></td>
<td><strong>0.9834</strong></td>
</tr>
<tr>
<td>1,13</td>
<td>3335.259</td>
<td>46.045</td>
<td>0.7569</td>
</tr>
<tr>
<td>12,13</td>
<td>294.723</td>
<td>15.356</td>
<td>0.9789</td>
</tr>
</tbody>
</table>

Table 3. Comparison of performance of the LSSVM model with ARIMA, ANNs and ARIMA models studied by Faraway and Chatfield [3]

<table>
<thead>
<tr>
<th>Model</th>
<th>Indices of Performance Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SSE</td>
</tr>
<tr>
<td>Faraway’ SARIMA [3]</td>
<td>4328</td>
</tr>
<tr>
<td>Faraway’ ANN[3]</td>
<td>2900</td>
</tr>
<tr>
<td>SARIMA(0,1,1)x(0,1,1)_12</td>
<td>6810</td>
</tr>
<tr>
<td>LSSVM</td>
<td>2477</td>
</tr>
</tbody>
</table>

5 Conclusion

There are plenty of models used to predict time series data. In this paper, the LSSVM model was proposed for improving forecasting performance of seasonal time series. To illustrate the capability of the LSSVM model, airlines data was chosen as a case study. One of the most important factors in developing a satisfactory LSSVM model is the selection of the input variables. Empirical results on air passenger data sets using two different models clearly reveal the efficiency of the LSSVM model in terms of MSE and MAE values. These results show that the LSSVM model provides a robust modelling capable of capturing the seasonal nature of the time series data and thus producing more accurate forecasts.

Acknowledgements. The authors thankfully acknowledged the financial support that afforded by MOE, UTM and GUP Grant (VOT 4F681).
References


**Received: September 1, 2015; Published: October 12, 2015**