Applied Mathematical Sciences, Vol. 7, 2013, no. 57, 2833 – 2840
HIKARI Ltd, www.m-hikari.com

Hybrid ARIMA-Support Vector Machine Model for Agricultural Production Planning

Thoranin Sujjaviriyasup

Department of Industrial Engineering, Faculty of Engineering
Khon Kaen University, Thailand 40002
Thoraninbki@gmail.com

Komkrit Pitiruek

Department of Production Technology, Faculty of Technology
Khon Kaen University, Thailand 40002
komkri@kku.ac.th

Abstract

In this study we develop the hybrid models for forecasting in agricultural production planning. Real data of Thailand's orchid export and Thailand's pork product are used to validate candidate models. Autoregressive Integrate Moving Average (ARIMA) is also selected as a benchmarking to compare other developed models. The main concept of building the models is to combine different forecasting techniques in order to overcome the time-series forecasting errors. The combined models of Support Vector Machine (SVM) and ARIMA are considered as they can be represented both nonlinear and linear values. We perform many experiments on the combination of SVM and ARIMA and select the most precision model, which is the SVM (10) and ARIMA hybrid model, by using statistical criteria. For orchid export case, comparing to ARIMA, the error reduction from MAE, RMSE, and MAPE is 2.46%, 1.96%, and 4.63%, respectively. Moreover, the error reduction from MAE, RMSE, and MAPE is 8.08%, 6.24%, and 6.88%, respectively, for the case of pork product.

Keywords: Hybrid forecasting model, ARIMA, Support Vector Machine

1. Introduction

Nowadays, an international trading on agricultural produces is very competitive, especially agricultural countries that they play both producer and exporter roles. Therefore, they have to level up agricultural process and management in order to obtain sustainable achievement. Unfortunately, a harvesting is sometimes unpredictable because of geography and climate. Moreover, a demand in each period is uncertainty so it is difficult to manage and plan for the products. Consequently, future information is a useful and very challenging in agricultural production planning because more accurate information can reduce any risk. However, due to the uncertainty, a forecasting model [1, 15] plays an important role to capture the future information. A forecasting model [3, 9, 13] is not only a model that employs statistics [11] to formulate the model such as ARIMA and Holt-Winters but it is also a model that employs machine learning algorithm [6, 12, 10, 5] to formulate the model such as Artificial Neural Networks (ANNs) or Support Vector Machine (SVM). Moreover, there are many interests to combine more than one forecasting model [14, 7, 4, and 8] in order to overcome time series forecasting error. The combined model such as ARIMA, it has good performance in linear forecasting, and SVM that has good performance in nonlinear forecasting. In this research, we develop hybrid model of ARIMA and SVM and propose traditional model, machine learning model, and hybrid model to compare their performance using cross – validation check by MSE, RMSE, and MAPE criteria. The real data of Thailand's orchid export and Thailand's pork product are used to study in this research.

2. Numerical data

The Thailand's orchid export and Thailand's pork product export [2] from January 2007 to December 2012 is used to study in this research as show in Figure 1 – 2.

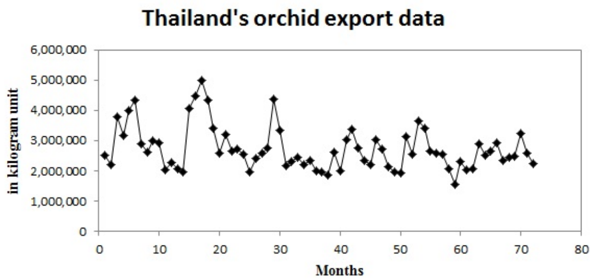

Figure 1: Thailand's orchid export data

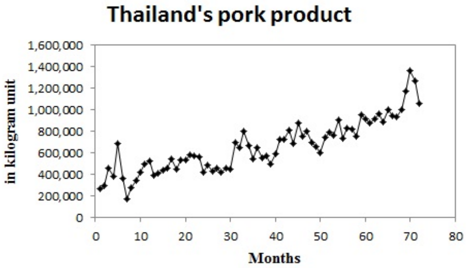

Figure 2: Thailand's pork product export data

3. Data preparation process

An algorithm of data preparation in traditional model and machine learning model is shown in Figure 3.

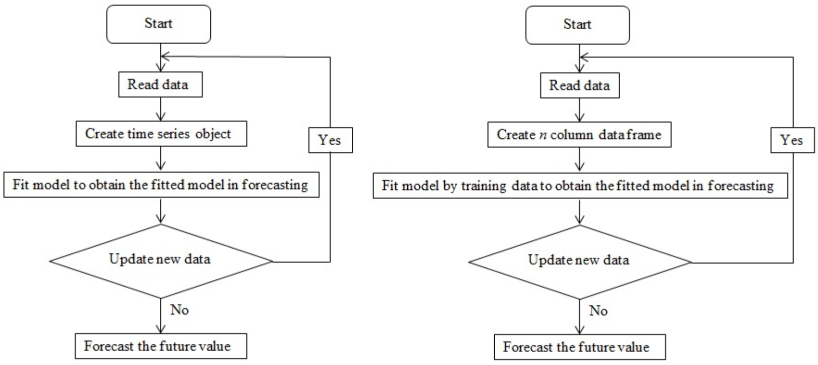

Figure 3: An algorithm of data preparation in traditional model and machine learning model

An algorithm of data preparation in hybrid model is shown in Figure 4.

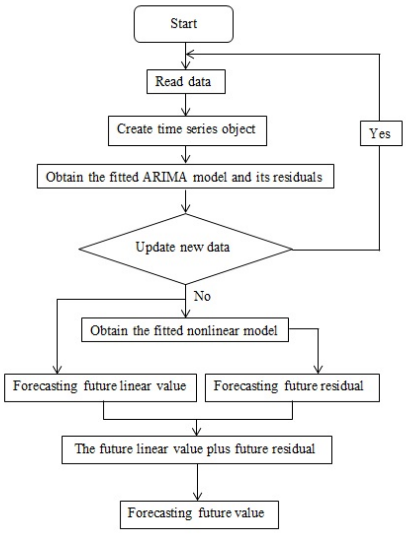

Figure 4: An algorithm of data preparation in hybrid model

4. Methodology

ARIMA model is composed of moving average (MA) and Autoregressive (AR) as equation (1)

$$\left(1-\sum_{i=1}^{p}\varphi_i L^i\right)(1-L)^d X_t = \left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t \tag{1}$$

Where L^i is Lag operator

φ_i is a parameter of autoregressive part

θ_i is a parameter of moving average part

ARIMA model used in this paper that is appropriately selected by AIC (Akaike's Information Criteria), which gives the lowest value. When updated data are realized, the data will be used to forecast for the following information and this procedure will be repeated.

Holt–Winters model or triple exponential smoothing is a mathematical formula that is the combination of the smoothed values of the constant part for time, the sequence of best estimates of the linear trend and the sequence of seasonal correction factor as equation (2).

The additive Holt-Winters model is

$$\begin{aligned} s_0 &= x_0 \\ s_t &= \alpha(x_t - c_{t-L}) + (1-\alpha)(s_{t-1} + b_{t-1}) \\ b_t &= \beta(s_t - s_{t-1}) + (1-\beta)b_{t-1} \\ c_t &= \gamma(x_t - s_t) + (1-\gamma)c_{t-L} \\ F_{t+m} &= (s_t + mb_t)c_{t-L+((m-1)(\mathrm{mod}\, L))} \end{aligned} \tag{2}$$

Where x_0 is beginning value at time $t=0$

s_t is the smoothed value of the constant part for time

b_t is the sequence of best estimates of the linear trend

c_t is the sequence of seasonal correction factor

F_{t+m} is the forecasting value of *m* ahead

Holt-Winters method used in this study, the fitted model is decided using Sum Squared Error (SSE). In the model, we use the present data to predict the future data. The updated data will also be used for forecasting the next data repeatedly.

SVM model is machine learning as supervised learning model to find fitted coefficient to set up a line to classify dataset in training process. Not only SVM model can formulate linear model but also SVM model can formulate nonlinear model to classify the nonlinear dataset. In this research, we slot 10 historical data and also use 1-9 historical data to input into the SVM model use kernel function as polynomial with degree and coefficient is 9 and 0.2, respectively, to predict the future numbers of 10th data.

For hybrid model, we propose two models in this research and they are made up of linear model and nonlinear functions and used to forecast the linear and nonlinear demands. The equation is as equation (3)

$$y = \hat{L} + \hat{N} + \varepsilon \tag{3}$$

Where y is actual value
$\hat{L}$ is linear value
$\hat{N}$ is nonlinear value
ε is white noise

In this part, the ARIMA is used to predict the linear values of future value. Then, the residuals obtained from the ARIMA are entered to SVM as the dataset of 10. The SVM model in this part, we use kernel function as polynomial with degree and coefficient is 9 and 0.2, respectively. However, dataset of 1-9 is used to train the selected model and predict the data of 10th. However, these obtained data are nonlinear data represented in Equation (3) and algorithm as shown in Figure 3.

5. Cross – validation process

We develop hybrid forecasting models in order to obtain the most accurate model for predicting the uncertain demand. The criteria used in the experiment are mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) in cross-validation check process. All criteria equations are shown in Equation (4) – (6). The ARIMA is selected and used as the benchmarking of univariate time-series forecasting method in order to compare with candidate model.

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} \left| \hat{\mathrm{y}}_\mathrm{i} - \mathrm{y}_\mathrm{i} \right|}{\mathrm{n}} \tag{4}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} \left(\hat{\mathrm{y}}_\mathrm{i} - \mathrm{y}_\mathrm{i} \right)^2}{\mathrm{n}}} \tag{5}$$

$$\mathrm{MAPE} = \frac{\sum_{i=1}^{n} \left| \hat{\mathrm{y}}_\mathrm{i} - \mathrm{y}_\mathrm{i} \right| / \mathrm{y}_\mathrm{i}}{\mathrm{n}} \times 100\,\% \tag{6}$$

We separate an export volume of Thailand's orchid and Thailand's pork product into two datasets as a formulation dataset and a validation dataset. The formulation dataset is 70% of an export data from January 2007 to March 2011. The rest is used to test the validation of forecasting model as shown in Table 1–2.

Table 1: MAE, RMSE, and MAPE of forecasted data of Thailand's orchid export

Model	Criteria (metric ton unit)		
	MAE	RMSE	MAPE
ARIMA	256.4365	365.1041	10.59%
Holt-Winters(additive)	436.0040	553.7446	18.14%
SVM(10)	342.6776	477.8260	14.08%
Hybrid ARIMA and SVM(10)	**250.1385**	**357.9637**	**10.10%**

Table 2: MAE, RMSE, and MAPE of forecasted data of Thailand's pork product export

Model	Criteria (metric ton unit)		
	MAE	RMSE	MAPE
ARIMA	1,113.9550	1,389.8392	11.49%
Holt-Winters(additive)	1,126.5990	1,459.5095	11.60%
SVM(10)	1,326.0020	1,728.3100	13.12%
Hybrid ARIMA and SVM(10)	**1,023.9810**	**1,303.1289**	**10.70%**

The results of hybrid ARIMA-SVM model with actual value in metric ton unit from April 2011 to December 2012. For the orchid and the pork product are shown in Figure 5 to 6.

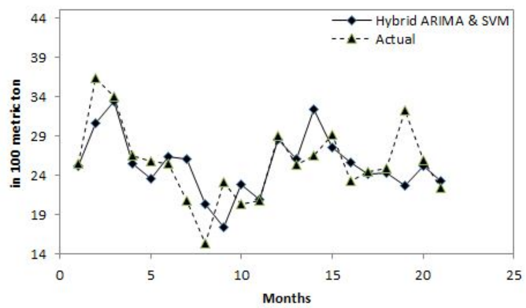

Figure 5: The hybrid ARIMA-SVM for Thailand's orchid dataset

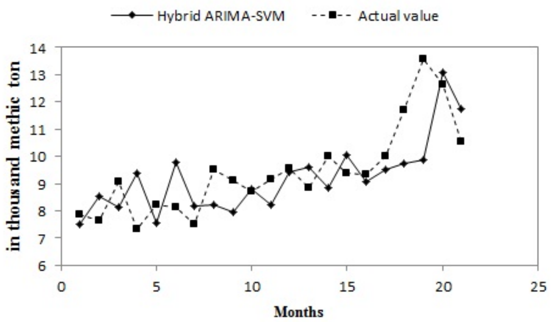

Figure 6: The hybrid ARIMA-SVM for Thailand's pork product dataset

6. Conclusion

The results indicate that the developed hybrid model of ARIMA and SVM with the cross-validation check is the most accurate forecasting tool among the others. The developed hybrid model is flexible and can reduce error of forecasting in the case of MAE, RMSE, and MAPE. For the orchid dataset are 2.46% MAE, 1.96% RMSE, and 4.63% MAPE, respectively. For the pork product dataset are 8.08% MAE, 6.24% RMSE, and 6.88% MAPE, respectively. In both dataset, the hybrid ARIMA-SVM is compared to ARIMA model. Additionally, the hybrid model provides more accuracy than single forecasting model although each single model does not provide more accuracy. In additional, the hybrid model can validate in improving performance in forecasting by using more than one dataset. In this way, the hybrid model may apply to another dataset in beyond. Moreover, the hybrid model of ARIMA and SVM is most appropriate to forecast Thailand's orchid export and Thailand's pork product.

References

[1] E. Suryani, ***et al.***, "Air passenger demand forecasting and passenger terminal capacity expansion: A system dynamics framework", Expert Systems with Applications 37 (2010) 2324 – 2339.

[2] http://www.oae.go.th/oae_report/export_import/export.php

[3] J. G. D. Gooijer and R. J. Hyndman, "25 year of time series forecasting", International Journal of Forecasting 22 (2006) 443 – 473.

[4] L. Shijin, ***et al.***, "A Hybrid Forecasting Model of Discharges based on Support Vector Machine", Procedia Engineering 28 (2012) 136 – 141.

[5] M. A. D. Oliveira, "An Application of Neural Networks Trained with Kalman Filter Variants (EKF and UKF) to Heteroscedatic Time Series Forecasting", Applied Mathematical Sciences, 6 (2012) 3675 – 3686.

[6] M. Khashi and M. Bijari, "An artificial neural network (*p*, *d*, *q*) model for timeseries forecasting", Expert Systems with Applications 37 (2010) 479 – 489.

[7] M. Khashi and M. Bijari, "A new class of hybrid models for time series forecasting", Expert Systems with Applications 39 (2012) 4344 – 4357

[8] M. Khashi and M. Bijari, "A novel hybridization of artificial neural networks and ARIMA models for time series forecasting", Applied Soft Computing 11 (2011) 2664 – 2675.

[9] M. Paliwal and U. A. Kumar, "Neural networks and statistical techniques: A review of applications", Expert Systems with Applications 36 (2009) 2 – 17.

[10] M. Shaverdi, ***et al.***, "Prediction of Stock Price of Iranian Pertochemical Industry Using GMDH-Type Neural Network and Genetic algorithm", Applied Mathematical Sciences, 7 (2012) 319 – 332.

[11] M. Shukla and S. Jharkharia, "Applicability of ARIMA Models in Wholesale Vegetable Market: An Investigation", Proceedings of the 2011 International Conference on Industrial Engineering and Operations Management Kuala Lumpur, Malaysia, January 22 – 24, 2011.

[12] N. Mohamed, ***et al.***, "Forecasting Short Term Load Demand Using Multilayer Feed-forward (MLFF) Neural network Model", Applied Mathematical Sciences, 6 (2012) 5359 – 5368.

[13] S.N. Hassan, ***et al.***, "A Comparison of the Forecast Performance of Double Seasonal ARIMA and Double Seasonal ARFIMA Model of Electricity Load Demand", Applied Mathematical Sciences, 6 (2012) 6705 – 6715.

[14] T. M. Choi, ***et al.***, "A hybrid SARIMA wavelet transform method for sales forecasting", Decision Support Systems 51 (2011) 130 – 140.

[15] Z. Hongbo and Z. Nian, "Forecast of Energy Demand in the Next Decade", Energy Prodedia 5 (2011) 2536 – 2539.

Received: March 20, 2013