Abstract

Time series forecasting is an active research area that plays important role in planning and decision making in several practical applications. The main task of this research area is to improve the prediction accuracy.

This thesis proposes three novel hybrid forecasting models which are significantly extended from the Zhang's model and the Khashei and Bijari's model by involving the clustering algorithm (i.e. *k*-means) and the discrete wavelet transform (DWT) for inputs pre-processing. Additionally, instead of including only the lagged values of time series as the input variables, additional variables such as moving averages and annual seasonal index are included into the proposed model. The experiments are conducted comprehensively with several hybridization scenarios in term of structures and variables to find the most suitable forecasting model for Thailand's cassava export.

The first proposed hybrid model (so called ARIMA+ARIMA/ANN+*k*-means/ARIMA/ANN) is the hybrid forecasting model involving autoregressive integrated moving average (ARIMA), artificial neural network (ANN) and the *k*-means clustering. These single models and the *k*-means clustering are used to build the forecasting models in different level of complexity (i.e. ARIMA; hybrid model of ARIMA and ANN; and hybrid model of *k*-means, ARIMA, and ANN). To obtain the final forecasting value, the forecasted values of these three models are combined with the weights generated from discount mean square forecast error (DMSFE) method.

The second proposed hybrid model (so called DWT/ARIMA/ANN) is the hybrid forecasting model of the DWT, the ARIMA, and the ANN without linear or nonlinear assumption on the approximation and the detail. The proposed model starts with decomposing the time series by the DWT to get the approximation and the detail. Then, the approximation and the detail are separately analyzed by the Zhang's model involving the ARIMA and the ANN in order to capture both linear and nonlinear components of the approximation and the detail. Finally, the linear and nonlinear components are additively combined for the final forecasting value.

These two novel hybrid models are applied to three well-known data sets: Wolf's sunspot, Canadian lynx, and exchange rate (British pound to US dollar) to evaluate the prediction capability in three measures (i.e. MSE, MAE, and MAPE). The prediction performance of the proposed models is compared to both the traditional single and hybrid models. The results imply that the proposed models give the best performance in MSE, MAE, and MAPE for all three data sets.

Then, the proposed hybrid models are implemented to Thailand's cassava export as the case study. In addition, we also propose the third novel hybrid model (so called ARIMA/ANN with pre-processed variables), which is the hybrid model of the ARIMA and the ANN with pre-processed variables for Thailand's cassava export forecasting. The experimental results indicate that the DWT/ARIMA/ANN model is the best model for the native starch and the sago. On the other hand, the ARIMA/ANN with pre-processed variables model is the best model for the modified starch.

In conclusion, all three proposed hybrid models have shown their forecasting capability over both the traditional single and hybrid models. Therefore, they can be used as the alternative models for time series prediction. Moreover, the stakeholders involves in the cassava supply chain can apply the proposed models specified for each type of the cassava export to obtain more accurate prediction results of Thailand's cassava export. The proposed models for cassava forecasting can be applied to other commodity products sharing the similar characteristic with the cassava as well.

Keywords: hybrid time series forecasting model, autoregressive integrated moving average (ARIMA), artificial neural network (ANN), *k*-means, discrete wavelet transform (DWT)