

Chapter 76

A Novel Hybridization of ARIMA, ANN, and K–Means for Time Series Forecasting

Warut Pannakkong

School of Knowledge Science, Japan Advanced Institute of Science and Technology, Nomi, Japan

Van-Hai Pham

Pacific Ocean University, Nha Trang, Vietnam

Van-Nam Huynh

School of Knowledge Science, Japan Advanced Institute of Science and Technology, Nomi, Japan

ABSTRACT

This article aims to propose a novel hybrid forecasting model involving autoregressive integrated moving average (ARIMA), artificial neural networks (ANNs) and k-means clustering. The single models and k-means clustering are used to build the hybrid forecasting models in different levels of complexity (i.e. ARIMA; hybrid model of ARIMA and ANNs; 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 the discount mean square forecast error (DMSFE) method. The proposed model is applied to three well-known data sets: Wolf's sunspot, Canadian lynx and the exchange rate (British pound to US dollar) to evaluate the prediction capability in three measures (i.e. MSE, MAE, and MAPE). In addition, the prediction performance of the proposed model is compared to ARIMA; ANNs; Khashei and Bijari's model; and the hybrid model of k-means, ARIMA, and ANN. The obtained results show that the proposed model gives the best performance in MSE, MAE, and MAPE for all three data sets.

INTRODUCTION

Time series forecasting is an active research area that continuously improve effectiveness of forecasting techniques over several decades (De Gooijer & Hyndman, 2006). This research area has contributed to various practical applications: finance (Wei, 2016; Adhikari & Agrawal, 2014), agriculture (Ezzine, Bouziane, & Ouazar, 2014; Garrett, et al., 2013), energy (Sadaci, Enayatifar, Abdulla, & Gani, 2014; Bahrami, Hooshmand, & Parastegari, 2014), transportation (Gosasang, Chandraprakaikul, & Kiattisin, 2011; Xiao, Xiao, & Wang, 2012), environment (Deng, Wang, & Zhang, 2015; Feng, et al., 2015), etc.

Traditionally, one of the most popular forecasting model is autoregressive integrated moving average (ARIMA). The ARIMA usually outperforms other forecasting approaches due to its capability in dealing with non-stationary time series as well as stationary time series. Nevertheless, the ARIMA is a kind of linear model, for this reason, it makes a prior assumption on relationship between historical and future time series as a linear function, which is very difficult to be satisfied in practical situations (Box, Jenkins, & Reinsel, 2008).

Artificial neural network (ANN), an artificial intelligent mimicking biological neurons mechanism, is widely used because of its usage flexibility over the ARIMA. The ANN can fit the relationship between inputs (e.g. historical time series) and outputs (e.g. predicted time series) without pre-assuming their relationship. Moreover, the ANN with only one hidden layer can be used as a universal approximator for continuous functions (Zhang, Patuwo, & Hu, 1998; Hornik, Stinchcombe, & White, 1989). The prediction performances of the ARIMA and the ANN were compared in several studies, and the results indicated that the ANN usually gave better accuracy than the ARIMA (Zou, Xia, Yang, & Wang, 2007; Co & Boosarawongse, 2007; Prybutok, Yi, & Mitchell, 2000; Kohzadi, Boyd, Kermanshahi, & Kaastra, 1996; Ho, Xie, & Goh, 2002; Alon, Qi, & Sadowski, 2001). Recently, in order to improve the effectiveness of the ANN which is a nonlinear model, both the results and the residuals of the ARIMA have been included as the inputs to gain unique capability in linear and nonlinear modeling (so called ARIMA/ANN) (Khashei & Bijari, 2011).

Currently, the prediction performance of the ANN can be improved by applying the ANN to clusters formed by clustering techniques (e.g. *k*-means and self-organizing map (SOM)) instead of applying the ANN directly to whole time series. (Benmouiza & Cheknane, 2013; Ruiz-Aguilar, Turias, & Jiménez-Come, 2015; Amin-Naseri & Gharacheh, 2007). This method can enhance the prediction performance because the observations assigned to the same cluster share similar characteristics that make their pattern to be easier fitted by the ANN. Nevertheless, it can cause the overfitting problem,

In addition, even though, the time series is well separated into the clusters, but we cannot actually know the cluster of each future value. In this situation, the approach to select the suitable cluster for prediction is an interesting issue. A recent work proposed to use summation of the prediction values from every cluster (Ruiz-Aguilar, Turias, & Jiménez-Come, 2015), but in fact, it would be more logical if the future values are produced from the ANN dedicated to their cluster. For this reason, the ANN should be selected based on the prediction of cluster of the future value. In order to do so, the clustering technique that can provide the straight clear-cut boundary between the clusters such as the *k*-means clustering is required. Therefore, a hybrid model of the ARIMA and the ANN with the *k*-means clustering is developed.

Normally, hybridization of the forecasting models has been constructed by combination of single models. Although the hybrid models can outperform the single models, but only in the average of prediction performance. The single models can still give the better results than hybrid models in some forecasting period. The forecasting models having difference of complexity (single, hybrid, and hybrid

25 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/a-novel-hybridization-of-arima-ann-and-k-means-for-time-series-forecasting/289028

Related Content

Netlab Training

(2014). *Medical Diagnosis Using Artificial Neural Networks* (pp. 213-232).

www.irma-international.org/chapter/netlab-training/111012

Comprehensive Modelling of ANN

Meghna Babubhai Patel, Jagruti N. Patel and Upasana M. Bhilota (2022). *Research Anthology on Artificial Neural Network Applications* (pp. 31-40).

www.irma-international.org/chapter/comprehensive-modelling-of-ann/288949

An ANN Model for Predicting the Quantity of Lead and Cadmium Ions in Industrial Wastewater

E. A. Olajubu, Gbemisola Ajayi, Isaiiah Adesola Oke and Franklin Oladiipo Asahiah (2020). *Deep Learning and Neural Networks: Concepts, Methodologies, Tools, and Applications* (pp. 159-173).

www.irma-international.org/chapter/an-ann-model-for-predicting-the-quantity-of-lead-and-cadmium-ions-in-industrial-wastewater/237871

A Model Augmenting Credit Risk Management in the Banking Industry

Okuthe Paul Kogeda and Nicknolt N. Vumane (2020). *Deep Learning and Neural Networks: Concepts, Methodologies, Tools, and Applications* (pp. 123-143).

www.irma-international.org/chapter/a-model-augmenting-credit-risk-management-in-the-banking-industry/237869

Methods and Applications of Graph Neural Networks for Fake News Detection Using AI-Inspired Algorithms

Arpit Jain, Ishta Rani, Tarun Singhal, Parveen Kumar, Vinay Bhatia and Ankur Singhal (2023). *Concepts and Techniques of Graph Neural Networks* (pp. 186-201).

www.irma-international.org/chapter/methods-and-applications-of-graph-neural-networks-for-fake-news-detection-using-ai-inspired-algorithms/323829