Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 48 (2015) 14 – 21

Procedia Computer Science

International Conference on Intelligent Computing, Communication & Convergence

(ICCC-2015)

Conference Organized by Interscience Institute of Management and Technology,

Bhubaneswar, Odisha, India

A Model Ranking Based Selective Ensemble Approach for Time Series Forecasting

Ratnadip Adhikari*, Ghanshyam Verma, Ina Khandelwal

Department of Computer Science and Engineering, the LNM Institute of Information Technology, Jaipur-302031, India

Abstract

Time series analysis is a highly active research topic that encompasses various domains of science, engineering, and finance. A major challenge in this field is to obtain reasonably accurate forecasts of future data from analyzing the past records. A fruitful alternative to using a single forecasting technique is to combine the forecasts from several conceptually different models. Numerous research studies in literature strongly recommend this approach, due to the fact that a combination of multiple forecasts almost always substantially reduces the overall forecasting errors as well as outperforms the component models. In this paper, we propose an ensemble method that selectively combines some of the constituent forecasting models, instead of combining all of them. On each time series, the component models are successively ranked as per their past forecasting accuracies and then we combine the forecasts of a group of high ranked models. Empirical analysis is conducted with nine individual models and four real-world time series datasets. Results clearly show that our proposed ensemble mechanism achieves consistently better accuracies than all component models and other conventional forecasts combination schemes.

Peer-review under responsibility of scientific committee of International Conference on Computer, Communication and Convergence (ICCC 2015)

* Corresponding author. Tel.: +91-966-093-3276;
E-mail address: adhikari.ratan@gmail.com

Peer-review under responsibility of scientific committee of International Conference on Computer, Communication and Convergence (ICCC 2015)
doi:10.1016/j.procs.2015.04.104

Keywords: Time series; Combination of forecasts; Forecasting accuracy; Box-Jenkins models; Neural networks; Support vector machines.

1. Introduction

Time series analysis is a highly important and dynamic research domain, having numerous practical applications. Its primary objective is to develop a mathematical model that estimates the underlying data generation process, retaining the statistical properties of the series and then to forecast desired number of future observations through this model. Appropriate modeling and forecasting of a time series is a considerably difficult task, mainly due to several unintended characteristics, often associated with the series. These include *nonstationarity*, irregular fluctuations, seasonal and cyclical variations, deviations from the standard statistical specifications, and severe *multicollinearity* among the observations [1]. The most appropriate alternative is to combine the forecasts from several structurally different models, instead of adopting only one model. Forecasts combination is based on the rational ideology that no specific model alone can consistently achieve best forecasts for a class of time series, but multiple models in unison can provide a very close estimation of the actual data generation process [2]. A number of renowned research works in this domain have demonstrated that a combination of forecasts generally comes up with much better forecasting accuracy than each component model. Moreover, this approach also substantially reduces the risk associated with selecting a single individual forecasting technique [3, 4].

Throughout the past two decades, there has been an overwhelming amount of research on combining forecasts, mainly due to its outstanding potency of accuracy enhancement. As a result, a variety of combination techniques have been developed in literature [2, 3]. Most of them form a weighted linear combination of the component forecasts, the weights being determined from the past forecasting records of the participating models. Their range varies from the simple statistical techniques, e.g. simple average, trimmed mean, winsorized mean, median, etc. [5] to more advanced methods, e.g. the outperformance and optimal linear combination of forecasts [2, 6]. Recently, Adhikari and Agrawal [7] have comprehensively reviewed the performances of several linear forecasts combination techniques on nine real time series datasets. An important finding from the past as well as recent research is that the simple techniques of combining generally achieves considerably better accuracies than more complex schemes. We further notice that there has been little work on selecting the suitable models in the ensemble and as such, the existing works combines all component forecasts. However, it is obvious that not all models will produce good forecasts for the particular time series and so tactically discarding some of them can potentially improve the overall accuracy to a large extent. This observation is the primary motivation behind the present work.

In this paper, we propose an ensemble methodology that combines the forecasts from some selected component models. The appropriate subset of forecasts to combine is selected through a ranking mechanism. At first, the models are successively ranked between one and the total number of models, so that a model with a comparatively smaller in-sample forecasting error gets a smaller, i.e. in fact a better rank and vice versa. Then, starting with the first rank, we consecutively select a predefined number of models and form a weighted linear combination of their forecasts. The weight to each model in this group is assigned to be inversely proportional to its in-sample forecasting error. In this manner, the proposed approach selectively combines the forecasts from a group of better performing models and discards the others. In order to check the precision and effectiveness of our approach, empirical analysis is carried out with nine individual forecasting models on four real time series datasets. The forecasting performance of the proposed ensemble is compared with those of the individual models as well as a number of other traditional linear combination techniques, through two popular error measures.

The remainder of the paper is organized as follows. Section 2 describes various well-known linear forecasts combination techniques and Sect. 3 presents the proposed ensemble mechanism. Section 4 reports the empirical analysis and finally Sec. 5 concludes this paper.

2. The ensemble forecasting paradigm

The most popular and widely used ensemble method is to form a linear combination of the constituent forecasts. Let us consider that $\mathbf{Y} = \left[y_1, y_2, \ldots, y_N \right]^{\mathrm{T}}$ be the actual out-of-sample testing dataset of a time series and