Neural Networks 50 (2014) 1–11

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Correcting and combining time series forecasters

Paulo Renato A. Firmino [a], Paulo S.G. de Mattos Neto [b], Tiago A.E. Ferreira [a,*]

[a] Department of Statistics and Informatics, Federal Rural University of Pernambuco, 52171-900, Recife, Pernambuco, Brazil

[b] Department of Computing, University of Pernambuco, Garanhuns, Pernambuco, Brazil

ARTICLE INFO

Article history:
Received 6 March 2013
Received in revised form 22 October 2013
Accepted 25 October 2013

Keywords:
Time series forecasters
Unbiased forecasters
Maximum likelihood estimation
Linear combination of forecasts
Artificial neural networks hybrid systems

ABSTRACT

Combined forecasters have been in the vanguard of stochastic time series modeling. In this way it has been usual to suppose that each single model generates a residual or prediction error like a white noise. However, mostly because of disturbances not captured by each model, it is yet possible that such supposition is violated. The present paper introduces a two-step method for correcting and combining forecasting models. Firstly, the stochastic process underlying the bias of each predictive model is built according to a recursive ARIMA algorithm in order to achieve a white noise behavior. At each iteration of the algorithm the best ARIMA adjustment is determined according to a given information criterion (*e.g.* Akaike). Then, in the light of the corrected predictions, it is considered a maximum likelihood combined estimator. Applications involving single ARIMA and artificial neural networks models for Dow Jones Industrial Average Index, S&P500 Index, Google Stock Value, and Nasdaq Index series illustrate the usefulness of the proposed framework.

1. Introduction

All phenomena sampled over time generate a time series. Due to the time ordering and usual stochastic nature underlying these series, their statistical modeling has been paramount to analyzing and forecasting the respective phenomena (Box, Jenkins, & Reinsel, 1994; Kantz & Schreiber, 2003). The well-known linear autoregressive integrated moving average (ARIMA) (Box et al., 1994), non-linear autoregressive conditionally heteroscedastic (ARCH) (Engle, 1982), and generalized ARCH (GARCH) (Bollerslev, 1986) models are examples of statistical formalisms for time series analysis and forecasting.

Regarding non-linear models, approaches based on artificial neural networks (ANN) for time series forecasting have produced convincing results in recent decades (Ferreira, Vasconcelos, & Adeodato, 2008; Gerald & Dimitri, 2007; Hippert & Taylor, 2010; Morabito & Versaci, 2003; Xiao Niu, feng Shi, & Wu, 2012; Zhang & Eddy Patuwo, 1998, 2001). The main challenge of these ANN models has been to adjust their attributes, like connection weights, architecture (*e.g.* structure, transfer function) and learning algorithms. In this way, it has been usual to spend considerable computational resources in order to select the best ANN model (Amjady & Keynia, 2010; Ferreira et al., 2008; Gerald & Dimitri, 2007; Hippert & Taylor, 2010; Morabito & Versaci, 2003; Xiao Niu et al., 2012). These strategies commonly use some evolutionary computation technique based on population dynamics – for example, genetic algorithms (Michalewicz, 1996), evolution strategy (Beyer & Schwefel, 2002), swarm intelligence techniques, like ant colony optimization (Dorigo, Manirezzo, & Colomi, 1996) and particle swarm optimization (de M. Neto, Petry, Rodrigues, & Ferreira, 2009), and so on (Morabito & Versaci, 2003; Xiao Niu et al., 2012).

Anyway, regardless of dealing with linear (*e.g.* ARIMA) or non-linear patterns (*e.g.* ANN), much of the time series literature implicitly assume that there is a true model for the series and that this model is known before it is fitted to the data (Chatfield, 2000), only resting to infer its parameters. In other terms, this modeling perspective only deals with parameters uncertainties and neglects the intrinsic structural or model uncertainty. Corroborating with the fact that it is very difficult to find a unique and true model for a given time series, some authors such as Neuman (2003) also have pointed out that adopting just one model may lead to statistical bias and underestimation of uncertainty. With those arguments in mind, model uncertainty seems to be present in any time series analysis.

Currently, model uncertainty research has been in the vanguard of time series analysis. For the sake of illustration, Taskaya-Temizel and Casey (2005) have studied the performance of autoregressive neural network hybrids, warning for the danger of underestimating the relationship between the models' linear and non-linear components. In turn, some authors (Amendola & Storti, 2008; Dell'Aquila & Ronchetti, 2006; Lean, Shouyang, & L, 2005; Lux & Morales-Arias, 2010) have rejected the hypothesis that

* Corresponding author. Tel.: +55 8133206490.
E-mail addresses: taef.first@gmail.com, tiago@deinfo.ufrpe.br (T.A.E. Ferreira).

0893-6080/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.neunet.2013.10.008

a unique and true model is achievable. Instead, these researchers have been challenged by the attempt of combining diverse models in order to present aggregate forecasts. In the specific case of ANN models, there are basically two methodologies: mixtures of experts methods (Hansen, 1999) and ensemble methods (Hansen, 1999; Shi, Xu, & Liu, 1999). The former consists of a system that is composed of several single networks, where each one learns to handle a subset of the complete data set of training. This procedure can be viewed as a modular version of an multilayer ANN. The latter works on multiple predictors and uses (non-)linear combinations of them, resulting in an aggregate forecaster.

Actually, whether one considers ANN or other modeling approaches, model uncertainty literature reviews (Clemen, 1989; Jeong, 2009; Wallis, 2011) have emphasized linear combination of forecasters (LCF). This class of estimators is among the most effective and simplest ways to deal with model uncertainty. In LCF the combined estimator is given by a weighted average of single models, where the weight of each model is usually a function of the variance of its residuals and its correlation with other models. This is the case with minimum variance (MV) combined estimators – one of the most widely known LCF approaches (Menezes, Bunn, & Taylor, 2000). Specifically, the MV LCF are commonly used by both the statistics and the neural network communities (Hashem, 1997).

Generally, LCFs are based on the assumption that the residuals of each single model are caused by random shocks, characterizing white noises, or unpredictable, independent and unbiased terms. However, mostly because of the heterogeneity of the phenomenon under study or even due to disturbances not captured by the models, it is yet possible that such suppositions are violated when applying ARIMA and ANN models (de M. Neto, Lima, & Ferreira, 2010; Sitte & Sitte, 2002), for instance. In fact, many ANN models are relatively ineffective in making estimates for state spaces they have not been trained on (Chen & Leung, 2004), eventually leading to biased noises.

The present paper illustrates cases involving ARIMA and ANN models where the white noise supposition is violated and suggests a model uncertainty approach to overcome the problem. Specifically, a two-step LCF model is presented. Firstly, the time trend of the residuals of each single model is fitted via a recursive ARIMA-based algorithm. At each iteration of the algorithm an ARIMA model is adjusted to the current remaining residuals. This algorithm is repeated until a white noise residual is achieved, *i.e.* until an ARIMA(0, 0, 0) is reached. Only after this step, a LCF model is introduced in order to combine the resulting unbiased forecasters.

The paper is organized as follows. The next section presents the proposed framework and suggests an algorithm to envelop the main ideas of the paper. Section 3 illustrates the usefulness of the proposed approach by means of four case studies involving ARIMA and ANN models and financial time series, namely the Dow Jones Industrial Average Index, S&P500 Index, Google Stock Value, and Nasdaq Index series. Section 4 brings some concluding remarks.

2. The proposed maximum likelihood (ML) approach

The proposed framework can be divided into three parts (see Fig. 1). In the first part, namely the Classical Approach, the residuals of each model are supposed to be white noises. This step reflects the elaboration of the single forecasters and it is not the object of study of the present paper. Anyway, one important step to be highlighted in the elaboration of the single models is the correct establishment of the relevant lags, *i.e.* the past points of the series that should be considered to perform predictions with reasonable accuracy. Thus, the single forecasts must be seen as an input for the proposed approach and specific details regarding their functional form are neglected. Therefore, the single models are considered black-box like models and their parameters uncertainties are not considered.

In turn, the Correction Procedure is dedicated to model the residuals of each forecaster when exposed to application. The purpose of this step is to capture any tendency of the phenomena of interest not enveloped by the original forecasting models. Then, if necessary, the prediction of the error series is used to correct the forecaster estimates for the future values of the series (Forecast_k^c in Fig. 1).

Finally, in the Aggregation Procedure the forecasts of the corrected models are combined via ML estimation. Specifically, the adopted LCF penalizes forecasters which involve greater error variance and correlations than the remaining forecasters. Thus, in the adopted aggregation method, the weight of each model is proportional to its statistical efficiency and independence with regard to the remaining predictors. This step is based on the supposition that the input forecasters are unbiased, reinforcing the importance of the correction phase. The output of this process is the combination of the corrected single models forecasts (Forecast^a in Fig. 1).

Such a framework is formally presented as follows. Following general literature practice, random variables and constants will be represented by upper and lower case letters respectively. Subindexes: t, i, and j to correspond time, the forecaster, and the recursive-ARIMA model, in this order. Respectively, u, $\widehat{U}$, and E are the unknown to be predicted, an estimator (forecaster) for the unknown u, and its respective random error (bias). In order to deal with the model uncertainty problem in the cases where the estimates are performed by experts, Mosleh and Apostolakis (1986) consider two error structures:

$$E = \widehat{U} - u \tag{1}$$

namely additive error model and

$$E = \frac{\widehat{U}}{u} \tag{2}$$

so-called multiplicative error model.

For modeling these two error structures, some authors (Droguett & Mosleh, 2008; Shirazi & Mosleh, 2009) suggest respectively normal and lognormal distributions. Specifically, Droguett and Mosleh (2008) use independent and identically distributed (*iid*) additive (or multiplicative) errors for handling predictions performed by mathematical models. Thus, regarding a time series context, the set of errors resulting from the predictor, $\mathbf{E} = (E_1, E_2, \dots, E_t, \dots, E_n)$, is seen as a sample of n *iid* normal (or lognormal) distributed random variables reflecting the model performance, in such a way that the closer the additive (multiplicative) E_t to zero (one) the better the forecaster is. Actually, this reasoning is in accordance with the basic framework of time series analysis. The assumption of *iid* normal (for additive) or lognormal (for multiplicative) distributed errors plays the role in the usual formalisms, such as ARIMA models.

One can see that for the additive error, $\widehat{U}_t = u_t + E_t$, and for the multiplicative error, $\widehat{U}_t = u_t \cdot E_t$. Basically, the idea of the frameworks originated from Mosleh and Apostolakis (1986) is to measure the error model in the light of observed values of $\mathbf{E}$, here related by the performance data set $\mathbf{e} = (e_1, e_2, \dots, e_t, \dots, e_n)$ (where a lowercase letter denotes the observed value of the random variable denoted by the same letter capitalized), and then to use the resulting error probability distribution when performing new predictions, by means of the Bayes theorem. Among the arguments for studying each model bias, Droguett and Mosleh (2008) emphasize the practical possibility of adopting models only partially applicable to the problem at hand. Anyway, mostly because