Combining Forecasts: A Genetic Programming Approach

Adriano S. Koshiyama, Department of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil

Tatiana Escovedo, Department of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil

Douglas M. Dias, Department of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil

Marley M. B. R. Vellasco, Department of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil

Marco A. C. Pacheco, Department of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil

ABSTRACT

Combining forecasts is a common practice in time series analysis. This technique involves weighing each estimate of different models in order to minimize the error between the resulting output and the target. This work presents a novel methodology, aiming to combine forecasts using genetic programming, a metaheuristic that searches for a nonlinear combination and selection of forecasters simultaneously. To present the method, the authors made three different tests comparing with the linear forecasting combination, evaluating both in terms of RMSE and MAPE. The statistical analysis shows that the genetic programming combination outperforms the linear combination in two of the three tests evaluated.

Keywords: Forecasters Combination, Forecasting Methods, Genetic Programming, Selection, Weighing

1. INTRODUCTION

Daily we face situations of uncertainty about the occurrence or fulfillment of a future project. Such goals as a family trip or a purchase of an asset portfolio are surround by externalities, such as the possibility of rain, or a substantial appreciation of this investment in the next day.

One of the strategies used to anticipate, mitigate or hedge in these environments involves the use of heuristics and procedures that bind a set of subjective and objective information in order to enable the generation of future scenarios, discovery of stylized facts, understanding the dynamics and the exogenous variables that influence this underlying process. This area of science is known as time series
analysis (Box et al., 1994; Shumway & Stoffer, 2000; Brockwell & Davis, 2002; Morettin & Toló, 2004).

The most common way to build a forecasting model lies in the achievement of four steps similar to those established by Box et al. (1994): (i) identification: from the dataset, check which lags or exogenous variables to compose the chosen model; (ii) estimation: given the dataset and the functional form of the model, fit the model parameters using some method (least squares, maximum likelihood, etc.); (iii) diagnosis: stage in which the existence of any violation on the models assumptions are checked, also the quality provided, and finally (iv) forecasting.

Although this relatively straightforward methodology, one of the main challenges takes place in the period before the four steps: the selection of an appropriate forecasting model. Therefore, the choice of a single model, besides being a complex and adverse selection, incurs the neglect of others potential models that provide local forecasts more accurate than the selected model. Bates and Granger (1969) suggested the use of a combination of forecasters, as a way to guarantee the use of the local qualities of each adjusted model, and thus allow a consensual and general better forecasting.

One of the proposals from Granger and Ramanathan (1984), results of a linear combination of the fitted values of each forecasters so that the choice of weights given to each one minimizes the mean squared error (MSE) between the output generated and the expected value (target). However, in a view of the forecasting system the possibility of a nonlinear relationship between such estimates and the fact that some of these models produce poor performance forecasts, adding unwanted noise are also subjects addressed to the linear combination.

Genetic programming (Koza, 1992), a metaheuristic used for selection and combination of terminals (variables) by an evolutionary process of recombination and mutation, can enable the implementation of this nonlinear (or linear) combination of the most relevant models. The aim of this work is to use the Genetic Programming method for combining forecasts in time series analysis.

We exhibit in the next section some related works about methods used for forecasting and model weighing. Section 3 displays each model used in the forecast combination. Section 4 presents the linear and genetic programming forecasting combination, also describing the proposed tests for evaluating each method. Section 5 we analyze the results obtained and, finally, section 6 presents the final considerations.

2. RELATED WORKS

In time series analysis, aiming in univariate forecasting, we can find several approaches traditionally focusing on mathematical and statistical methods (Atsalakis & Valavanis, 2009; Stepnicka et al., 2013). In this way, forecasting models like naive, linear and nonlinear models and Holt-Winters (Hyndman et al., 2008; Taylor, 2012) are mathematical methods that use extracted components (trend, seasonality, etc.) or past samples of the underlying process as a way to produce forecasts. Although, a time series can also be understood as a stochastic process that follows a specific probability distribution. By this principle, regression analysis and autoregressive integrated with moving average (ARIMA) models (Box et al., 1994; Valenzuela et al., 2008) are adequate tools that provide good performance if the time series follows all the statistical assumptions, for example, the time series follows a Gaussian distribution.

In many practical applications, the time series cannot fulfill all assumptions demanded by these tools, and so, a challenge is imposed to achieve good forecasts. Based on this, Computational Intelligence methods, like neural networks (Zhang et al., 1998; Fonseca et al., 2001) and genetic programming (Wagner et al., 2007; Jara, 2011) has been extensively used by forecasters as a path to achieve better solutions.
Application of Biologically Inspired Techniques for Industrial and Environmental Research via Air Quality Monitoring Network
www.igi-global.com/chapter/application-of-biologically-inspired-techniques-for-industrial-and-environmental-research-via-air-quality-monitoring-network/161069?camid=4v1a

Classification of Faults in Power Transmission Systems Using Modern Techniques: An Overview
www.igi-global.com/chapter/classification-of-faults-in-power-transmission-systems-using-modern-techniques/177987?camid=4v1a