Expert Systems with Applications 37 (2010) 974–984

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Development of a PSO–SA hybrid metaheuristic for a new comprehensive regression model to time-series forecasting

J. Behnamian, S.M.T. Fatemi Ghomi *

Department of Industrial Engineering, Amirkabir University of Technology, 424 Hafez Avenue, Tehran, Iran

ARTICLE INFO

Keywords:
Curve fit
Non-linear regression
Forecasting
Hybrid metaheuristic
Particle swarm optimization
Simulated annealing
Time-series
Fitness efficiency index

ABSTRACT

Forecasting has always been a crucial challenge for organizations as they play an important role in making many critical decisions. Much effort has been devoted over the past several decades to develop and improve the time-series forecasting models. In these models most researchers assumed linear relationship among the past values of the forecast variable. Although the linear assumption makes it easier to manipulate the models mathematically, it can lead to inappropriate representation of many real-world patterns in which non-linear relationship is prevalent. This paper introduces a new time-series forecasting model based on non linear regression which has high flexibility to fit any number of data without preassumptions about real patterns of data and its fitness function. To estimate the model parameters, we have used hybrid metaheuristic which has the ability of estimating the optimal value of model parameters. The proposed hybrid approach is simply structured, and comprises two components: a particle swarm optimization (PSO) and a simulated annealing (SA). The hybridization of a PSO with SA, combining the advantages of these two individual components, is the key innovative aspect of the approach. The performance of the proposed method is evaluated using standard test problems and compared with those of related methods in literature, ARIMA and SARIMA models. The results in solving on 11 problems with different structure reveal that the proposed model yields lower errors for these data sets.

1. Introduction

Determination of a suitable model and best fitted parameters has a widespread application in various aspects of engineering and business. Time-series analysis is an important tool for forecasting the future in terms of past history. Time-series methods are generally used when there is not much information about the generation process of the underlying variable and when other variables provide no clear explanation of the studied variable (Zhang, 2003). By studying many related variables together, a better understanding is often obtained. Robust forecasting must rely on how well the time-series is designed. Many techniques for time-series analysis have been developed assuming linear relationships among the series variables. Gooijer and Hyndman (2006) have published a review of the literature for the methods used to model time-series over the last 25 years. The most popular methods are the moving average, exponential smoothing and ARIMA (autoregressive integrated moving average) methods. These methods all assume linear relationship among the past values of the forecast variable. Unfortunately, many real-world applications involve non-linearities between environmental variables. In this case assumption of simple relationship among time-series variables can produce poor results regarding the ability to predict the future. In many cases, such inaccuracies can produce major problems. For example, forecasting river flows or business and economic trends needs to be handled very carefully to produce accurate results. Tong (1983, 1990) introduced a description of some of the drawbacks of linear modeling for time-series analysis. This includes for example, their inability to explain the sudden bursts of very large amplitudes at irregular time intervals. It seems necessary, therefore, that non-linear models be used for the analysis of the real-world temporal data.

The usefulness of alternative solution methods, approximate in nature but less restrictive, for general non-linearities has been noted by several authors. For instance, Kim and Kim (1977) used a genetic fuzzy predictor ensemble; Salustowicz and Schmidhuber (1997) used probability instructions combined with genetic programming; Sheta and De Jong (2001) applied genetic algorithms to tune radial basis functions; Zhang (2003) combined the ARIMA structure with neural networks; Chen, Yang, Dong, and Abraham (2005) proposed a neural model; Hwarng and Hui-Kuang Yu (2006), Singh (2007) presented fuzzy time-series models; Zhang and Kline (2007) used neural networks for quarterly time-series forecasting; (Valenzuelaa et al., 2008) also proposed the use of hybrid ARIMA and artificial neural networks models; Zhang (2007) proposed a neural ensemble model that incorporates noise into

* Corresponding author. Tel.: +98 21 66413034; fax: +98 21 66413025.
E-mail address: fatemi@aut.ac.ir (S.M.T. Fatemi Ghomi).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.05.079

the data used to build different training sets; and Gómez-Ramírez, Najim, and Ikonen (2007) applied an enhanced polynomial artificial neural networks. Koop and Potter (2000) suggest that conditional–mean non-linearities may be a feature of the data generating process (DGP), but may not be large enough to yield much of an improvement to forecasting, as well as the explanation that they are present and important, but that the wrong types of non-linear models have been used to try. In addition, Taio and Tsay (1989) addressed the modeling for the multi-variable time-series using statistical non-linear models.

Nevertheless, incorporating non-linearities into models can lead to very difficult mathematical problems, in which the optimal set of parameters may be difficult to know. This difficulty could explain why there have been fewer studies devoted to non-linear time-series than to linear time-series. Time-series analysis and design has attracted a considerable number of interest from the evolutionary computation community to study, model and forecast time-series data behavior. Iba, Hitoshi, Hung, and Taisuke (1993) developed a GA-based approach for system identification and time-series prediction problems. This approach is called STROGANOFF (structured representation on genetic algorithms for non-linear function fitting). Mulloy, Riolo, and Savit (1996) applied the generating process to the task of chaotic and time-series prediction problem. Recently Da Silva (2008) presented a non-linear model that combined radial basis functions and the ARMA(p,q) structure. He used a scatter search metaheuristic to find optimal set of parameters.

This paper develops a comprehensive regression model for forecasting time-series data. A hybrid metaheuristic (HMH) approach which integrates several features from particle swarm optimization (PSO) and simulated annealing (SA) are used to tune the model parameters. An advantage of this approach is that no assumptions are made about the data pattern in contrast to more traditional methods (Hlavackova & Neruda, 1993).

The rest of paper is organized as follows. Section 2 proposes a comprehensive regression model. Section 3 deals with hybrid metaheuristic. Especially, it outlines the configuration of hybrid metaheuristic to estimate parameters and to initially estimate parameters. Section 4 gives computational experiments using benchmark data to indicate that the proposed approach is effective when compared to existing algorithms for the problem. Finally, section 5 is devoted to conclusions and suggestions for future studies.

2. Proposed comprehensive regression model

Regression has been used widespread to fit the most suitable for data (Ragsdale & Plane, 2000). The main purpose in all kinds of regression is to form $y = f(x) + \xi_i$ for a series of data in which Y is the dependent variable measured by experiment and X is independent variable which is changed during the experiment. F is a function to describe the relationship of X and Y, consisting of one or more parameters. ξ_i is the ith observation error from mean level of data having a normal distribution with zero mean (Brown, 2001). The general scheme of our proposed model is as follows:

$$f(X) = \frac{f_1(x)}{c_1 + c_2\left(f_1(x) + c_3 \times \ln\left(b_0 + \sum_{i=1}^{5} b_i X^{a_i}\right)\right)} \quad (1)$$

in which

$$f_1(x) = b_0 + \sum_{i=1}^{5} b_i X^{a_i} + b_6^{a_6 X} + b_7 \mathrm{Sin}(a_7 X) + b_8 \mathrm{Cos}(a_8 X) \quad (2)$$

and model parameters are $b_0, b_1, b_2, b_3, b_4 \ldots b_8, a_1, a_2, a_3 \ldots a_8, c_1, c_2$ and c_3. Note that, to establish the model we have used the general form of logistic regression equation while covering all the possible cases in data patterns (Mitchell, 2006).

The point to consider is the large number of parameters and the way of estimating these parameters, while downsizing the parameters in estimates leads to the reduction of regression degree for fitness function resulting in the case of linear regression with a low efficacy (Ragsdale & Plane, 2000). There is a direct relationship between the number of parameters in non-linear fitness model and its efficacy, based on Eye and Schuster (1998). According to their book, the best fitness is for an equation with the most number of parameters while the authors have tried to fit a regression line with different number of parameters for data. As is shown in Fig. 1, the fourth degree model has the highest optimality to reduce the error in fitness model and it has been reasoned that the higher number of parameters has reduced the fitness error while it has reduced the capability of model.

This feature is well presented for fewer numbers of data in Figs. 2 and 3. The accuracy of passed curve from points is nearly 100%, i.e., the model has the most accurate case of fitness but in Fig. 2, due to the lack of enough observations the model has not been able to identify the general pattern of data for future predictions, therefore the value of prediction has gradually decreased due to the last declining data. But in Fig. 3, this number of data with enough observations is given to model and the 100% accuracy was obtained. The prediction of future was carried out accurately while there was the descending trend of last real data.

Our proposed model has performed the fitness with a high efficiency and without any pre-assumption in both cases while the higher degree of freedom leads to the higher accuracy of prediction. Therefore, it is suggested that the more number of data is used to estimate parameters while it works well if it is used to fit curve to data. It is worth mentioning that the optimal value of some parameters is zero which must be taken into account in calculating the freedom degree of model.

3. The hybrid metaheuristic approach

Generally, the curve must describe data more accurately. A linear regression is a simple trend to establish a linear relationship between X and Y used in patterned data of first degree that can be carried out with a few simple points (Brown, 2001; Carman & Chasteen, 1996). Describing the data by non-linear function is a more complex issue which has been simplified by the advent of new techniques. There are some methods suggested for non-linear fitness but some drawbacks have confused the users. One of the suitable tools to do non-linear regression is to use hybrid metaheuristics.

The problem of finding such a set of parameters for (1) is complex, and alternative methods for deriving these parameters is of significant interest in time-series studies. When dealing with non-linear models, two approaches are usually followed. The first seeks to develop exact methods to derive the optimal parameters for the model. The second tries to find a good "near-optimal" set of parameters using metaheuristic procedures (Da Silva, 2008). The metaheuristic methods attempt to lead the parameter search

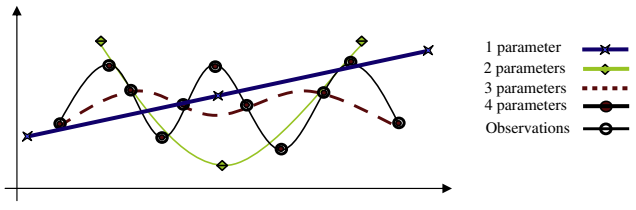

Fig. 1. Curve fitting for different number of parameters (Eye & Schuster, 1998).