European Journal of Operational Research 207 (2010) 807–816

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Stochastics and Statistics

Adaptive neural network model for time-series forecasting

W.K. Wong [a,*], Min Xia [a,b], W.C. Chu [a]

[a] Institute of Textiles and Clothing, The Hong Kong Polytechnic University, Hong Kong
[b] College of Information Science and Technology, Donghua University, Shanghai, China

ARTICLE INFO

Article history:
Received 27 January 2010
Accepted 14 May 2010
Available online 1 June 2010

Keywords:
Time-series
Forecasting
Adaptive metrics
Neural networks

ABSTRACT

In this study, a novel adaptive neural network (*ADNN*) with the adaptive metrics of inputs and a new mechanism for admixture of outputs is proposed for time-series prediction. The adaptive metrics of inputs can solve the problems of amplitude changing and trend determination, and avoid the over-fitting of networks. The new mechanism for admixture of outputs can adjust forecasting results by the relative error and make them more accurate. The proposed *ADNN* method can predict periodical time-series with a complicated structure. The experimental results show that the proposed model outperforms the auto-regression (*AR*), artificial neural network (*ANN*), and *adaptive k-nearest neighbors* (*AKN*) models. The *ADNN* model is proved to benefit from the merits of the *ANN* and the *AKN* through its' novel structure with high robustness particularly for both chaotic and real time-series predictions.

1. Introduction

Many planning activities require prediction of the behavior of variables (e.g. economic, financial, traffic and physical). The predictions support the strategic decisions of organizations (Makridakis, 1996), which in turn sustain a practical interest in forecasting methods. Time-series methods are generally used to model forecasting systems when there is not much information about the generation process of the underlying variable and when other variables provide no clear explanation about the studied variable (Zhang, 2003).

Time-series forecasting is used to forecast the future based on historical observations (Makridakis et al., 1998). There have been many approaches to modeling time-series dependent on the theory or assumption about the relationship in the data (Huarng and Yu, 2006; Chen and Hwang, 2000; Taylor and Buizza, 2002; Kim and Kim, 1997; Zhang et al., 1998; Wang and Chien, 2006; Singh and Deo, 2007). Traditional methods, such as time-series regression, exponential smoothing and autoregressive integrated moving average (Brooks, 2002) (*ARIMA*), are based on linear models. All these methods assume linear relationships among the past values of the forecast variable and therefore non-linear patterns cannot be captured by these models. One problem that makes developing and implementing this type of time-series model difficult is that the model must be specified and a probability distribution for data must be assumed (Hansen et al., 2002). Approximation of linear models to complex real-world problems is not always satisfactory.

Recently, artificial neural networks (*ANN*) have been proposed as a promising alternative to time-series forecasting. A large number of successful applications have shown that neural networks can be a very useful tool for time-series modeling and forecasting (Adya and Collopy, 1998; Zhang et al., 1998; Celik and Karatepe, 2007; Wang and Chien, 2006; Sahoo and Ray, 2006; Singh and Deo, 2007; Barbounis and Teocharis, 2007; Bodyanskiy and Popov, 2006; Freitas and Rodrigues, 2006). The reason is that the *ANN* is a universal function approximator which is capable of mapping any linear or non-linear functions (Cybenko, 1989; Funahashi, 1989). Neural networks are basically a data-driven method with few priori assumptions about underlying models. Instead they let data speak for themselves and have the capability to identify the underlying functional relationship among the data. In addition, the *ANN* is capable of tolerating the presence of chaotic components and thus is better than most methods (Masters, 1995). This capacity is particularly important, as many relevant time-series possess significant chaotic components.

However, since the neural network lacks a systematic procedure for model-building, the forecasting result is not always accurate when the input data is very different from the training data. Like other flexible non-linear estimation methods such as kernel regression and smoothing splines, the *ANN* may suffer either under-fitting or over-fitting (Moody, 1992; Geman et al., 1992; Bartlett, 1997). A network that is not sufficiently complex can fail to fully detect the signal in a complicated data set and lead to under-fitting. A network that is too complex may fit not only the signal but also the noise and lead to over-fitting. Over-fitting is especially misleading because it can easily lead to wild prediction far beyond the range of the training data even with the noise-free data. In order to solve this problem, a novel *ANN* model is proposed

* Corresponding author. Tel.: +00852 64300917.
E-mail address: tcwongca@inet.polyu.edu.hk (W.K. Wong).

0377-2217/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2010.05.022

in this study with the adaptive metrics of inputs, and the output data is evolved by a mechanism for admixture. The adaptive metrics of inputs of the model can adapt to local variations of trends and amplitudes. Most inputs of the network are close to the historical data in order to avoid a dramatic increase in the forecasting error due to the big difference between training data and input data. In using the proposed mechanism for admixture of outputs, the forecasting result can be adjusted by the relative error, making the forecasting result more accurate.

The forecasting results generated by the proposed model are compared with those obtained by the traditional statistical *AR* model, traditional *ANN* architectures (*BP* network), and *adaptive k-nearest neighbors* (*AKN*) method (Kulesh et al., 2008) in the related literature. The experimental results indicate that the proposed model outperforms the other models, especially in chaotic and real data time-series predictions.

This paper is organized as follows. In the next section, the fundamental principle of the proposed method is introduced. The experimental results are presented in Section 3. The last section concludes this study.

2. Methodology

We focus on one-step-ahead point forecasting in this work. Let $y_1, y_2, y_3, \ldots, y_t$ be a time-series. At time t for $t \geqslant 1$, the next value y_{t+1} will be predicted based on the observed realizations of $y_t, y_{t-1}, y_{t-2}, \ldots, y_1$.

2.1. The ANN approach to time-series modeling

The *ANN* is a flexible computing framework for a broad range of non-linear problems (Wong et al., 2000). The network model is greatly determined by data characteristics. A single hidden-layer feed-forward network is the most widely used model for time-series modeling and forecasting (Zhang and Qi, 2005). The model is characterized by a network of three layers of simple processing units connected by acyclic links. The hidden layers can capture the non-linear relationship among variables. Each layer consists of multiple neurons that are connected to neurons in adjacent layers. The relationship between the output y_{t+1} and the inputs $y_t, y_{t-1}, y_{t-2}, \ldots, y_{t-p+1}$ has the following mathematical representation:

$$y_{t+1} = \alpha_0 + \sum_{j=1}^{q} \alpha_j g\left(\beta_{0j} + \sum_{i=1}^{p} \beta_{ij} y_{t-i+1}\right) + \varepsilon, \tag{1}$$

where α_j ($j = 0,1,2,\ldots,q$) and β_{ij} ($i = 0,1,2,\ldots,p$; $j = 1,2,3,\ldots,q$) are the model parameters called connection weights, p is the number of input nodes and q is the number of hidden nodes. The logistic function is often used as the hidden-layer transfer function, which is,

$$g(x) = \frac{1}{1 + e^{-x}}. \tag{2}$$

A neural network can be trained by the historical data of a time-series in order to capture the characteristics of this time-series. The model parameters (connection weights and node biases) can be adjusted iteratively by the process of minimizing the forecasting errors (Liu et al., 1995).

2.2. Adaptive neural network model for forecasting (ADNN)

It is well known that the *ANN* may suffer either under-fitting or over-fitting (Moody, 1992; Geman et al., 1992; Bartlett, 1997). A network that is not sufficiently complex can fail to fully detect the signal leads to under-fitting. Over-fitting generally occurs when a model is excessively complex. A model which has been under-fitting or over-fitting will generally have poor predictive performance, as it can exaggerate minor fluctuations in the data. For these two problems, the over-fitting is more important when the signal data is sufficient and the network is sufficiently complex. Thus, in this paper we emphasize on the problem of over-fitting for the *ANN*. Generally, the *ANN* algorithm is said to over-fitting relative to a simpler one if it is more accurate in fitting known data (hindsight) but less accurate in predicting new data (foresight). In order to avoid over-fitting, the adaptive neural network model is proposed. In this model, the hindsight data is used to modify the inputs of the *ANN* in the prediction processing making the inputs approach to the learning data. Thus, this algorithm can reduce the chance of over-fitting. Based on the current *ANN*, an extension is done to develop the adaptive neural network (*ADNN*) model for time-series forecasting. Firstly, a strategy is used to initialize the input data $y_t, y_{t-1}, y_{t-2}, \ldots, y_{t-m+1}$, where m is the number of input nodes. The strategy adopts the adaptive metrics which are similar to the *adaptive k-nearest neighbor* method. The data set $y_t, y_{t-1}, y_{t-2}, \ldots, y_{t-m+1}$ is compared with the other parts of this time-series, which have the same length. The determination of the closeness measure is the major factor in prediction accuracy. Closeness is usually defined in terms of metric distance on the Euclidean space. The most common choices are the Minkowski metrics:

$$L_M(Y_t, Y_r) = \left(|y_t - y_r|^d + |y_{t-1} - y_{r-1}|^d + \cdots + |y_{t-m+1} - y_{r-m+1}|^d\right)^{\frac{1}{d}}. \tag{3}$$

This equation gives the value difference between Y_t and Y_r, but the differences of trends and amplitudes are not presented. In time-series forecasting, the information on trends and amplitudes is the crucial factor. In this study, adaptive metrics are introduced to solve this problem and the arithmetic is presented as:

$$L_A(Y_t, Y_r) = min_{\lambda_r, u_r} f_r(\lambda_r, u_r), \tag{4}$$

$$f_r(\lambda_r, u_r) = \left(|y_t - \lambda_r y_r - u_r|^d + |y_{t-1} - \lambda_r y_{r-1} - u_r|^d + \cdots + |y_{t-m+1} - \lambda_r y_{r-m+1} - u_r|^d\right)^{\frac{1}{d}}, \tag{5}$$

where h_r and l_r are the largest and smallest elements of vector correspondingly, $\lambda_r \in \left[1, \frac{h_r}{l_r}\right]$, $u_r \in [0, h_r - l_r]$. The parameter of minimization λ_r equilibrates the amplitude difference between Y_t and Y_r. The parameter u_r is responsible for the trend of time-series.

The optimization problem (4) can be solved by the algorithm of Levenberg–Marquardt (Press et al., 1992) optimization or other gradient methods for $d \geqslant 1$. In this study, d is assumed to be 2 and gives the widely used Euclidean metrics.

$$f_r(\lambda_r, u_r) = \left(|y_t - \lambda_r y_r - u_r|^2 + |y_{t-1} - \lambda_r y_{r-1} - u_r|^2 + \cdots + |y_{t-m+1} - \lambda_r y_{r-m+1} - u_r|^2\right)^{\frac{1}{2}}. \tag{6}$$

For $d = 2$, two equations are considered:

$$\begin{cases} \frac{\partial f_r(\lambda_r, u_r)}{\partial \lambda_r} = 0, \\ \frac{\partial f_r(\lambda_r, u_r)}{\partial u_r} = 0. \end{cases} \tag{7}$$

When the corresponding linear system is solved, the solution of the minimization problem can be obtained analytically:

$$u_r = \frac{z_1 z_2 - z_3 z_4}{m z_2 - z_3^2}, \quad \lambda_r = \frac{m z_4 - z_1 z_3}{m z_2 - z_3^2},$$

where