Neurocomputing 147 (2015) 315–323

Contents lists available at ScienceDirect

Neurocomputing

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

Online multivariate time series prediction using SCKF-γESN model

Min Han*, Meiling Xu, Xiaoxin Liu, Xinying Wang

Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, Liaoning 116023, China

ARTICLE INFO

Article history:
Received 24 February 2013
Received in revised form
13 March 2014
Accepted 16 June 2014
Communicated by H. Zhang
Available online 30 June 2014

Keywords:
Echo state network
Multivariate time series
Online prediction
Square root cubature
Kalman filter

ABSTRACT

In this research, for online modeling and prediction of multivariate time series, we propose a novel approach termed squared root cubature Kalman filter-γ echo state network (SCKF-γESN). First, multivariate time series are modeled by using γ echo state network (γESN). Then, by using squared root cubature Kalman filter (SCKF), we update parameters of γESN and predict future observations online. Furthermore, we add a robust outlier detection algorithm to SCKF to protect SCKF-γESN from divergence caused by outliers. Finally, two numerical examples, by using a multivariate benchmark dataset and a real-world dataset, are conducted to substantiate the effectiveness and characteristics of the proposed SCKF-γESN.

1. Introduction

Recently, there have been many techniques developed in the field of nonlinear multivariate time series prediction. For example, autoregressive regression model [1], wavelet transform [2], support vector machine [3], fuzzy theory [4], self-organizing approaches [5], and neural networks [6]. Among these methods, neural networks have been adopted increasingly for nonlinear time series prediction [7]. The salient feature of neural networks is their appealing ability to learn input–output mappings from data with no necessary for prior knowledge. There are two major types: feed forward neural networks (FNNs) and recurrent neural networks (RNNs). Compared with FNNs, RNNs have the lateral connections or cyclic connections, which have been proven to be computationally effective for time series prediction [8]. Additionally, RNNs can provide a fairly accurate universal approximation to an arbitrary linear function [9]. Therefore they are widely used to predict the future values dependent on the past values [10]. Because RNN training algorithms are on the basis of direct optimization of the network weights, they usually exhibit slow convergence speed associated with high computational requirements, and often yield local optima of the optimized objective functions [11].

As a significant development in RNNs, echo state networks (ESNs) facilitate the application of RNNs and outperform the conventional fully trained RNNs [12–14]. Conventional ESNs consist of a nontrainable sparse recurrent part, i.e., reservoir, and a linear readout part. Input weights and connection weights inside the reservoir are generated randomly. Only the synaptic connections from the recurrent reservoir to output readout neurons are adaptable via supervised learning [15]. Similar to ESNs, extreme learning machines (ELMs) have the property that hidden nodes are generated randomly, and only output weights are calculated analytically [16,17]. But ELMs are single-hidden-layer feed forward networks. They do not exhibit dynamic features. While the reservoir in ESNs can maintain active even in the absence of inputs, which exhibits the ability of dynamic memory. Like RNNs, ESNs can learn to mimic a target system with arbitrary accuracy [12]. Thanks to the merits stated above, ESNs have been successfully applied in time series prediction, positioning, reinforcement learning, inverse modeling, nonlinear control and so on [12,18,19].

In recent years, researchers have found that neurons in conventional ESNs are usually set as hyperbolic tangent function [14], whose jump and slope play a key role in nonlinear approximation [20]. In [21], the authors proposed an input scaling parameter γ to control the activation potential of hyperbolic tangent function. Similarly, in this paper, we also take the input scaling parameter γ into consideration. Both the input scaling parameter γ and the output weights are trained in γESN. With respect to γ, training γESN becomes a nonlinear problem.

There are two major types of training γESN [22]. One is off-line method, in which adjustments to unknown parameters are performed after the presentation of all the samples in the training set. It is rather demanding in terms of storage requirements. The other one is on-line method, in which adjustments to unknown parameters are performed one-by-one. The data which have already been used in the training process can be discarded, so this method can save memory space and ease computational load [16]. Kalman filter (KF) can produce the optimal solution to estimating the

* Corresponding author. Tel.: +86 411 84708719; fax: +86 411 84707847.
E-mail address: minhan@dlut.edu.cn (M. Han).

http://dx.doi.org/10.1016/j.neucom.2014.06.057
0925-2312/

unknown state in a linear dynamic system [23]. However real-world systems are plagued by nonlinearities generally. In this case, it is hard to get a closed-form solution to the state estimation, so we have to make some approximations. The widely known nonlinear filter is extended Kalman filter (EKF), which has been used in many practical applications for decades [24]. But EKF could work well only in a mild nonlinear environment owing to the first order Taylor series approximation for nonlinear functions. Afterwards, deriving a more accurate nonlinear filter that could be applied for a wide range of nonlinear functions has been the main subject of intensive research in the field of Kalman filters. The conventional approaches to solving the problems are unscented Kalman filter (UKF) [25] and cubature Kalman filter (CKF) [26,27]. They share a common property of using a weighted set of symmetric points, which are called sigma-points in UKF and cubature-points in CKF respectively. However, they have the operation of square-rooting, which is making them numerically sensitive and easy to be ill-conditioned. The square-root version of the CKF, termed SCKF, can avoid matrix square-rooting operations and improve the numerical accuracy. We may also seek square root solution to UKF. Unfortunately it is unavailable to get a square-root solution that enjoys the numerical advantage similar to SCKF [26].

In online time series prediction, another main problem is high noise levels, or the influences of outliers. Outliers are much different from the rest of the data based on some measure. Learning observations containing outliers without awareness may lead to fitting those unwanted data and may corrupt the approximation function. Therefore, detecting and removing the outliers are very important. Many algorithms have been proposed in recent years for outlier detection [28,29]. Specially, in [30], the authors proposed an approach to evaluate residual with causality constraint, which operated at a fast speed for prompt fault detection.

In this paper, an online echo state network based on square root cubature Kalman filters, referred to SCKF-γESN, is proposed to deal with the multivariate time series prediction. It can learn the training data one-by-one or chunk-by-chunk. We also combine outlier detection algorithm with SCKF-γESN, to improve the forecasting accuracy and robustness.

The rest of the paper is organized as follows. In Section 2, we give an introduction of γESN. In Section 3, we concisely introduce the proposed SCKF-γESN model. Afterwards, we analyze the merits of the proposed model. In Section 4, simulations are conducted. Furthermore, the performance of the SCKF-γESN model is compared with other models. Finally, in Section 5, discussions and conclusions are given.

2. γESN

In this section, we will describe the essence of γESN from the perspective of input scaling parameter γ beyond conventional ESNs.

Conventional ESNs are novel recurrent discrete-time neural networks. Fig. 1 shows the architecture of conventional ESNs. It consists of a feed-forward input layer with L units, a recurrently connected hidden layer with M units, and a feed-forward output layer with D units [12]. Without loss of generality, we shall address output layer with one readout neuron i.e. D is equal to 1. The center of ESNs is the recurrently interconnected hidden layer, usually called reservoir, because the echo state property (ESP, i.e., remembering a number of previous inputs and asymptotically washing out the initial information) is ensured by a sparse interconnectivity of 1%–5% within the reservoir [12,31].

Denote $\mathbf{u}(k)=[u_k(1),u_k(2),\cdots,u_k(L)]^{\mathrm{T}}\in\mathbb{R}^{L\times 1}$, and $z_k\in\mathbb{R}$ as the activation of the input and output units at time step k, respectively. The echo state $\mathbf{x}_k\in\mathbb{R}^{M\times 1}$ of the reservoir at time step k is generated from the temporal input $\mathbf{u}_k$ and the prior echo states $\mathbf{x}_{k-1}$, and it is defined as follows:

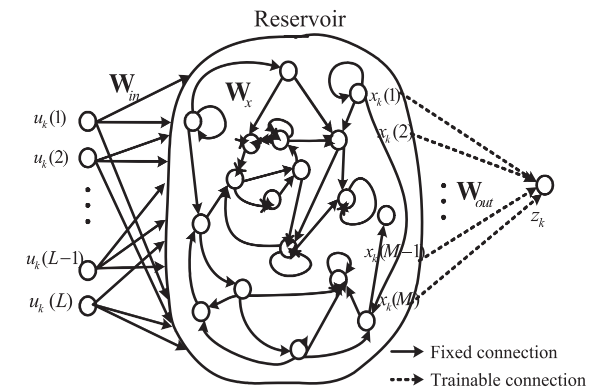

Fig. 1. Conventional ESNs structure. Solid arrows indicate fixed connections and dashed arrows trainable connections.

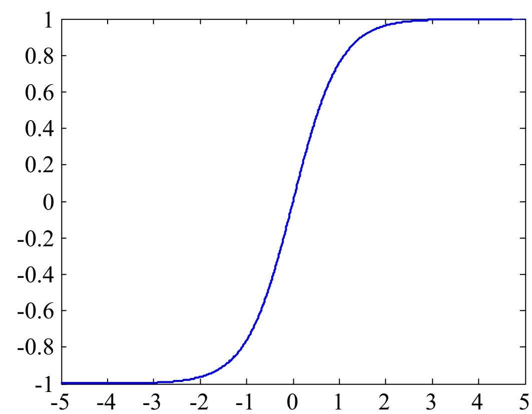

Fig. 2. Hyperbolic tangent sigmoid function.

$$\mathbf{x}_k=f(\mathbf{W}_{in}\mathbf{u}_k+\mathbf{W}_x\mathbf{x}_{k-1}) \tag{1}$$

where $f(\cdot)$ denotes the reservoir activation function. In this paper, we consider reservoirs comprising analog neurons, with $\tan h(\cdot)$ transfer function [15], and set the initial state of the reservoir as zero, i.e., $\mathbf{x}_0=\mathbf{0}$.

The linear output z_k is computed as

$$z_k=\mathbf{W}_{out}\mathbf{x}_k+\omega_k \tag{2}$$

The input-to-reservoir weights matrix $\mathbf{W}_{in}\in\mathbb{R}^{M\times L}$ and the interconnection of reservoir weights matrix $\mathbf{W}_x\in\mathbb{R}^{M\times M}$ are time-invariant and known. Only the reservoir-to-output weights matrix $\mathbf{W}_{out}\in\mathbb{R}^{1\times M}$ is modified via supervised learning. The spectral radius of $\mathbf{W}_x$, denoted by $\rho(\mathbf{W}_x)$, is less than 1. It ensures that the network converges to a stable equilibrium point without oscillation. $\{\omega_k\}$ is an independent Gaussian noise sequence.

In classical MLP models, there are local scaling parameters for input weights [32]. Similarly, in ESNs, add input scaling parameter γ into Eqs. (1) and (2), yielding

$$z_k=\mathbf{W}_{out}f(\gamma\mathbf{W}_{in}\mathbf{u}_k+\mathbf{W}_x\mathbf{x}_{k-1})+\omega_k \tag{3}$$

If the input scaling parameter γ is close to zero, $\gamma\mathbf{W}_{in}\mathbf{u}_k+\mathbf{W}_x\mathbf{x}_k$ will lie in a small interval. ESNs works in the linear region as shown in Fig. 2, i.e., it is suitable for an approximate linear regression task or weak nonlinear regression task. On the other hand, if γ is large, the ESN indicates strong nonlinearity and fits well for a nonlinear task [20]. Accordingly, input scaling parameter γ has a significant effect on the performance of ESNs. Eq. (3) is the