Engineering Applications of Artificial Intelligence 40 (2015) 28–36

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Improved extreme learning machine for multivariate time series online sequential prediction ☆

Xinying Wang, Min Han *

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

ARTICLE INFO

Article history:
Received 11 July 2014
Received in revised form
22 December 2014
Accepted 22 December 2014
Available online 29 January 2015

Keywords:
Online prediction
Multivariate time series
Extreme Learning Machine
LM algorithm

ABSTRACT

Multivariate time series has attracted increasing attention due to its rich dynamic information of the underlying systems. This paper presents an improved extreme learning machine for online sequential prediction of multivariate time series. The multivariate time series is first phase-space reconstructed to form the input and output samples. Extreme learning machine, which has simple structure and good performance, is used as prediction model. On the basis of the specific network function of extreme learning machine, an improved Levenberg–Marquardt algorithm, in which Hessian matrix and gradient vector are calculated iteratively, is developed to implement online sequential prediction. Finally, simulation results of artificial and real-world multivariate time series are provided to substantiate the effectiveness of the proposed method.

1. Introduction

Time series prediction, which is bounded in both scientific researches and engineering applications, has attracted increasing attention for years (De Gooijer and Hyndman, 2006). Due to the complexity of underlying systems, nonlinear or chaotic time series prediction has aroused more and more concerns (Zhao et al., 2009; Li et al., 2012). Furthermore, the time series observed from complex systems generally compromises multiple variables, and there is more dynamic information of the underlying dynamic system contained in multivariate time series than in univariate time series (Cao et al., 1998; Chakraborty et al., 1992). As a consequence, multivariate time series prediction has become an increasingly important research direction (Popescu, 2011; von Bünau et al., 2009).

In existing literature, support vector machines (Sapankevych and Sankar, 2009), neural networks (Shi and Han, 2007; Pino et al., 2008), and other machine learning methods (Bai and Li, 2012) have been researched to predict time series. Moreover, according to Takens' Embedding Theorem (Takens, 1981), the time series can be reconstructed to the phase-space by a time delayed coordinate, which can translate time correlation to spatial correlation in phase-space. Because of the universal approximation capability and distributed computing characteristic, neural networks have become one of the most influential prediction tools (Jaeger and Haas, 2004; Zemouri et al., 2003; Niska et al., 2004).

But the traditional gradient-based learning algorithms of neural networks converge slowly and are easy to be trapped in local optimums, which constrain the prediction performance of neural networks. To deal with the shortcomings of traditional neural networks, extreme learning machine (ELM) has been developed (Huang et al., 2006b). Compared with other neural networks with random weights (Huang, 2014), the input weights and the bias of hidden nodes of ELM are generated randomly before learning, and an optimal output weights can be obtained by a one-shot algorithm. Owing to its simple structure, fast learning speed and good generalization performance, ELM has been successfully applied to function approximation (Rong et al., 2009), time series prediction (Nizar et al., 2008; Lian et al., 2013), pattern classification (Man et al., 2012; Miche et al., 2010; Luo and Zhang, 2014) and other fields (Soria-Olivas et al., 2011; Ye et al., 2013). Although ELM has greatly improved the neural network training speed and accuracy, there are still some shortcomings (Huang et al., 2011).

The ridge regression algorithm is introduced to improve the stability and generalization performance of ELM (Deng et al., 2009; Huang et al., 2012), and second order Newton optimization algorithm is applied in ELM training (Balasundaram, 2013). Besides, online learning variants of ELM are proposed to satisfy real-time and online learning requirements. Online sequential ELM (OS-ELM) (Liang et al., 2006) provides a sequential implementation of the least squares solution of ELM. Successively, ensemble of online sequential extreme learning machine (EOS-ELM) (Lan et al., 2009), online sequential extreme learning machine with forgetting mechanism (FOS-ELM) (Zhao et al., 2012), regularized online sequential learning algorithm (ReOS-ELM) (Huynh and Won, 2011), low complexity adaptive

☆This paper is under the support of the National Natural Science Foundation of China (61374154) and the National Basic Research Program of China (973 Program) (2013CB430403).

* Corresponding author.
E-mail address: minhan@dlut.edu.cn (M. Han).

http://dx.doi.org/10.1016/j.engappai.2014.12.013
0952-1976/

forgetting factor OS-ELM (LAFF-COS-ELM) (Lim et al., 2013), online sequential ELM-TV (Ye et al., 2013) and online sequential extreme learning machine with kernels (OS-ELMK) (Wang and Han, 2014a) have been proposed and their superior performances have been testified.

Considering the advantages of ELM, some variants of ELM have been used to predict multivariate time series. In Wang and Han (2012), a model selection algorithm is applied to determine the optimal structure of ELM, and the resulting model is used to predict multivariate chaotic time series. And in Wang and Han (2014), different kernels are used together to map the multivariate time series and the resulting multiple kernel extreme learning machine (MKELM) is proposed. However, these two methods are all included in the batch or offline prediction framework. In order to solve the problem of online sequential prediction of multivariate time series, an improved ELM prediction model is presented in this paper. The multivariate time series is first reconstructed to the phase-space where ELM is used to approximate the input–output mapping. An improved Levenberg–Marquardt (LM) algorithm is developed to optimize the output weights of the ELM prediction model online sequentially. When new samples are observed, the Hessian matrix and the gradient vector are updated iteratively, and the corresponding output weights are tuned immediately. As a result, the ELM can learn the latest observed time series in real-time. The paper is organized as follows. In the second section, the problem definitions are given. Some preliminary work is briefly reviewed in the third section. Next, an improved online sequential LM algorithm is presented, and it is incorporated in the ELM prediction model. Finally, three experiments of artificial and real-world multivariate time series are conducted to illustrate the effectiveness of the proposed method compared with other existing approaches.

2. Problem definitions

The variables and notations are defined in Table 1.

3. Preliminaries

In this section, we will give a brief review of multivariate time series reconstruction and ELM model.

3.1. Multivariate time series reconstruction

Time series is a sequence of value points, measured at successive times spaced typically at uniform time intervals. Time series prediction is the use of a model to predict future value based on previously observed values. In order to establish a prediction model for the time series data generated from a nonlinear dynamic system, a time delayed phase-space reconstruction is used as preprocessing, generally. According to Takens' Embedding Theorem (Takens, 1981), as enough delayed coordinates are used, scalar time series is sufficient to reconstruct the dynamic of the underlying systems. However, it is not certain whether a given scalar time series is sufficient to reconstruct the dynamic or not (Popescu, 2011). Additionally, multivariate time series contains more dynamic information than scalar time series, using available multivariate time series would improve the prediction performance (Chakraborty et al., 1992).

Considering M dimensional time series: $X_1, X_2, \ldots, X_N$, where $X_i = (x_{1,i}, x_{2,i}, \ldots, x_{M,i})$, $i = 1, 2, \ldots, N$. As in the case of scalar time series (where $M = 1$), a time delayed reconstruction can be made as follows:

$$\begin{aligned} \boldsymbol{V}_n = [&x_{1,n}, x_{1,n-\tau_1}, \ldots, x_{1,n-(d_1-1)\tau_1},\\ &x_{2,n}, x_{2,n-\tau_2}, \ldots, x_{2,n-(d_2-1)\tau_2},\\ &\ldots\\ &x_{M,n}, x_{M,n-\tau_M}, \ldots, x_{M,n-(d_M-1)\tau_M}]^{\mathrm{T}} \end{aligned} \quad (1)$$

where τ_i, d_i, $i = 1, \ldots, M$, are the time delays and the embedding dimensions, respectively. As Takens' Embedding Theorem (Takens, 1981), if d or each d_i is large enough, there exists generally a function $F : \Re^d \to \Re^d \left(d = \sum_{i=1}^{M} d_i\right)$ such that

$$\boldsymbol{V}_{n+1} = F(\boldsymbol{V}_n) \quad (2)$$

The equivalent form of (2) can be written as

$$\begin{aligned} x_{1,n+1} &= F_1(\boldsymbol{V}_n)\\ x_{2,n+1} &= F_2(\boldsymbol{V}_n)\\ &\vdots\\ x_{M,n+1} &= F_M(\boldsymbol{V}_n) \end{aligned} \quad (3)$$

The remaining problems are how to choose the time delays τ_i and embedding dimensions $d_i, i = 1, \ldots, M$, so that (2) or (3) holds. There are several methods for choosing the time delay for scalar time series, such as mutual information and autocorrelation (Sun et al., 2014).

3.2. Extreme learning machine prediction model

ELM has a simple three-layer structure: input layer, output layer, and hidden layer which contains a large number of nonlinear processing nodes. The weights connecting the input layer to the hidden layer, and the bias values within the hidden layer of ELM are randomly generated and maintained throughout the learning process, and only the output weights need to be learned. Generally, ELM has interpolation capability and universal approximation capability (Huang et al., 2006a), so ELM can be a promising time series prediction tool.

Mathematically, ELM can be formulated as a function as follows:

$$\sum_{i=1}^{L} w_i g(\boldsymbol{W}_{in(i)}, b_i, \boldsymbol{x}_j) = \sum_{i=1}^{L} w_i g(\boldsymbol{W}_{in(i)} \cdot \boldsymbol{x}_j + b_i) = y_j, \quad j = 1, \ldots, N. \quad (4)$$

where $\boldsymbol{x}_j \in \Re^n$ is the input vector, $\boldsymbol{W}_{in(i)} \in \Re^n$ is the weight vector connecting the input nodes to the ith hidden node, $\boldsymbol{W}_{in(i)} \cdot \boldsymbol{x}_j$ denotes the inner product of $\boldsymbol{W}_{in(i)}$ and $\boldsymbol{x}_j$, $b_i \in \Re$ is the bias of the ith hidden node, $g(\cdot)$ is the sigmoid activation function , $w_i \in \Re$ is the output weight connecting the ith hidden node to the output node, $y_j \in \Re$ is the output of ELM, L is the number of hidden nodes, and N is the number of training samples. In the ELM learning framework, $\boldsymbol{W}_{in(i)}$ and b_i are randomly chosen beforehand.

The function (4) can be further expressed by the following matrix-vector form:

$$\boldsymbol{A}\boldsymbol{w} = \boldsymbol{Y}. \quad (5)$$

where

$$\boldsymbol{A} = \begin{pmatrix} g(\boldsymbol{W}_{in(1)}, b_1, \boldsymbol{x}_1) & \ldots & g(\boldsymbol{W}_{in(L)}, b_L, \boldsymbol{x}_1)\\ \vdots & \ddots & \vdots\\ g(\boldsymbol{W}_{in(1)}, b_1, \boldsymbol{x}_N) & \cdots & g(\boldsymbol{W}_{in(L)}, b_L, \boldsymbol{x}_N) \end{pmatrix}_{N\times L},$$

$\boldsymbol{Y} = [y_1, \ldots, y_N]^{\mathrm{T}}$, and $\boldsymbol{w} = [w_1, w_2, \ldots, w_L]^{\mathrm{T}}$. Matrix $\boldsymbol{A}$ is called the hidden layer output matrix of ELM in Huang et al. (2011); the ith column of $\boldsymbol{A}$ is the ith hidden node's output vector with respect to inputs $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ and the jth row of $\boldsymbol{A}$, denoted as $\boldsymbol{a}_j$ is the output vector of the hidden layer with respect to input $\boldsymbol{x}_j$.

If the ELM model with L hidden nodes can learn these N training samples with no residuals, there exists $\boldsymbol{w}$, so that

$$\sum_{i=1}^{L} w_i g(\boldsymbol{W}_{in(i)} \cdot \boldsymbol{x}_j + b_i) = t_j, \quad j = 1, \ldots, N. \quad (6)$$