Engineering Applications of Artificial Intelligence 25 (2012) 533–543

Contents lists available at SciVerse ScienceDirect

Engineering Applications of Artificial Intelligence

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

Identification of nonlinear systems with outliers using wavelet neural networks based on annealing dynamical learning algorithm

Chia-Nan Ko

Department of Automation Engineering, Nan Kai University of Technology, 568 Chung-Cheng Road, Tsaotun, Nantou 54243, Taiwan, ROC

ARTICLE INFO

Article history:
Received 11 March 2011
Received in revised form
25 August 2011
Accepted 16 September 2011
Available online 1 October 2011

Keywords:
Wavelet support vector regression
Annealing dynamical learning algorithm
Particle swarm optimization
Wavelet neural networks
Outliers

ABSTRACT

This paper presents an annealing dynamical learning algorithm (ADLA) to train wavelet neural networks (WNNs) for identifying nonlinear systems with outliers. In ADLA–WNNs, wavelet-based support vector regression (WSVR) is adopted to determine the initial translation and dilation of a wavelet kernel and the weights of WNNs due to the similarity between WSVR and WNNs. After initialization, ADLA with nonlinear time-varying learning rates is applied to train the WNNs. In the ADLA, the determination of the learning rates would be a key work for the trade-off between stability and speed of convergence. A computationally efficient optimization method, particle swarm optimization (PSO), is adopted to find the optimal learning rates to overcome the stagnation in the training procedure of WNNs. Due to the advantages of WSVR and ADLA (WSVR–ADLA), the WSVR-based ADLA–WNNs (WSVR–ADLA–WNNs) can robust against outliers and achieve the promising efficiency of system identifications. Three examples are simulated to confirm the performance of the proposed algorithm. From the simulated results, the feasibility and superiority of the proposed WSVR–ADLA–WNNs for identifying nonlinear systems with artificial outliers are verified.

1. Introduction

Identification of nonlinear systems can be found in various industries (Elfelly et al., 2010; Fu et al., 2009; Johnson et al., 2009; Lendaris, 2009; Luitel and Venayagamoorthy, 2010; Manel et al., 2006; Rouss and Charon, 2008; Vieira et al., 2005). However, it should point out that structural identification and parameter estimation of nonlinear systems are rather difficult issues in system identification. For scientific and engineering applications, the obtained training data are always subject to outliers. The intuitive definition of an outlier (Hawkins, 1980) is "an observation which deviates so much from other observations as to arouse suspicions that it is generated by a different mechanism". Generally, outliers may occur due to various reasons, such as erroneous measurements or noisy data from the tail of noise distribution functions. Recently, many researchers have endeavored to investigate the issue of identifying nonlinear systems with outliers (Chuang et al., 2000, 2002, 2004; Fu et al., 2010; Jeng et al., 2010; Lee et al., 1999; Swanchez, 1998).

Neural networks (NNs) are extensively used for approximating functions due to its simplicity and faster convergence (Ait Gougam et al., 2008; Azadeh et al., 2007; Chuang et al., 2002; Muzhou and Xuli, 2011; Narendra, 1990; Yang et al., 2011). Since NNs approximate functions without requiring a mathematical description of how outputs functionally depend on inputs, they are often referred to as model-free estimators (Kosko, 1992). The basic modeling philosophy of model-free estimators is that they learn from examples without any knowledge of the model type. When outliers exist in the training data, there still have some problems in the traditional NNs approaches. Hence, robust NNs are proposed to overcome the problems of the traditional NNs while facing outliers. These robust approaches could indeed improve the learning performance when training data contain outliers (Chuang et al., 2000, 2002, 2004; Fu et al., 2010; Lee et al., 1999).

In last decade, some researchers have developed the structure of NNs based on the wavelet functions to construct wavelet neural networks (WNNs) (Billings and Wei, 2005; Subasi et al., 2005; Tzeng, 2010; Wei et al., 2010; Wu and Chan, 2009; Xu and Ho, 2002). The WNNs are constructed based on the wavelet transform theory (Zhang and Benveniste, 1992). Wavelet decomposition is a powerful tool for function approximation using a wavelet function (Chui, 1992). Unlike the functions used in the conventional NNs, wavelet functions are spatially localized, so that the learning capability of the WNNs is more efficient than the conventional NNs. When utilizing WNNs, the number of wavelet layer nodes, the initial parameters of the kernel, and the initial weights of the networks must be determined first. These parameters are usually determined according to the experience of the

E-mail address: t105@nkut.edu.tw

0952-1976/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.engappai.2011.09.019

designer or are just chosen randomly. However, improper initialization usually results in slow convergence speed and poor performance of WNNs.

The support vector machine (SVM) is a new universal learning machine proposed by Vapnik in 1995, which is applied to both regression and pattern recognition (Min and Cheng, 2009; Schölkopf et al., 2000; Trafalis and Gilbert, 2006; Zhang et al., 2010). Due to the excellent performance of SVM and the approximating ability of wavelet kernel function, several researchers have combined wavelet and SVM (WSVM) to apply to investigation (Fernandez, 2007; Wang and Fu, 2010; Zhang et al., 2004, 2005). In WSVM, a regression method (WSVR) with wavelet kernel function is usually adopted to determine the initial structures of WNNs. After initialization, an annealing robust learning algorithm (ARLA) is applied to train the WNNs. In ARLA, a learning rate serves as an important role in the training procedure. In general, the learning rate is selected as a time-invariant constant by trial and error (Chuang et al., 2004; Fu et al., 2009, 2010; Hsieh et al., 2008; Lin, 2006). However, there still exist several problems of unstable or slow convergence. Some researchers have engaged in exploring the learning rate to improve the stability and the speed of convergence (Hsieh et al., 2008; Song et al., 2008; Yoo et al., 2007; Yu, 2004).

In this paper, annealing dynamical learning algorithm (ADLA) is proposed to overcome the stagnation in searching a globally optimal solution or the drawback of slow convergence of training WSVR-based WNNs for identifying nonlinear systems with outliers. That is, first, WSVR method is used to determine the initial translation and dilation of a wavelet kernel and the structures of WNNs (i.e. the proper number of wavelet layer nodes, the parameters of wavelet kernel function, and the synaptic weights). Then, the ADLA based on nonlinear time-varying learning rates is then applied to adjust the parameters of wavelet kernel and the synaptic weights for improving learning performance, in which a popular optimization approach, PSO, is adopted to find optimal learning rates. Finally, three simulation examples are illustrated to show the performances of the WSVR–ADLA–WNNs. From the simulated results, the proposed WNNs have the superiority over the conventional WNNs using fixed learning rates for identifying nonlinear systems with artificial outliers are verified.

2. WNNs for identification of nonlinear systems

The WNNs are constructed based on the wavelet transform theory and are alternatives of feed forward neural networks for identifying arbitrary nonlinear systems.

2.1. Structure of wavelet neural networks model

Using a multi-resolution analysis (MRA), the wavelet transform expands a signal or function onto a set of wavelet basis function (Daubechies, 1992; Mallat, 1989). The basis function $\Psi_{a,b}(\mathbf{x})$ can be derived from a mother wavelet $\Psi(\mathbf{x})$ through translations and dilations as

$$\Psi_{a,b}(\mathbf{x}) = \frac{1}{\sqrt{a}}\Psi\left(\frac{\mathbf{x}-\mathbf{b}}{a}\right) \quad (1)$$

where $a \in R, a > 0$ and $\mathbf{b} \in R^d$, a is the dilation and $\mathbf{b}$ is the translation. The normalization factor $\sqrt{a}$ in (1) ensures that $\Psi_{a,b}(\mathbf{x})$ has a constant norm in the space of square integrable functions. Then, an approximation of $f(\mathbf{x}) \in L^2(R)$ can be regarded as a linear combination of wavelets (Chui, 1992), it is expressed as

$$\tilde{f}(\mathbf{x}) = \sum_{j=1}^{m} w_j \Psi_{a_j,b_j}(x) \quad (2)$$

where $\tilde{f}(\mathbf{x})$ is the approximation of the function $f(\mathbf{x})$ and w_j is the weight of the jth wavelet.

Wavelet functions have efficient time-frequency localization properties. The wavelets have been applied in various research fields due to the capability of decomposing signals (Zhang, et al., 1995; Zhang and Benveniste, 1992). Based on the properties of good learning ability of NNs, combining wavelets with neural networks (NNs), wavelet neural networks (WNNs) have been dramatically developed (Muzhou and Xuli, 2011; Subasi et al., 2005; Wei et al., 2010; Wu and Chan, 2009; Yilmaz and Oysal, 2010; Zhang, 1997).

Generally, WNNs has a three-layered network structure that consists of input, wavelet, and output layers depicted in Fig. 1. The semantic meaning and operation function of the nodes in each layer are described as follows:

Layer 1 (input layer): In this layer, the input data $\mathbf{x} = [x_1, x_2, \ldots, x_n]$ are directly transmitted into the wavelet layer, wher n is the number of nodes; namely, the number of input variables.

Layer 2 (wavelet layer): In this layer, Morlet wavelet function (Zhang et al., 2004) is adopted as the activation function of the wavelet nodes connected with the input data is expressed as

$$\mu_j(x_i) = \cos\left(1.75 \times \frac{x_i - b_j}{a_j}\right)\exp\left(-\frac{(x_i-b_j)^2}{2a_j^2}\right) \text{ for } i = 1,2,\ldots,n,\ j = 1,2,\ldots,r \quad (3)$$

where a_j and b_j are the dilation and the translation of the jth wavelet function for the ith input variable x_i, respectively. The product of the jth multi-dimensional wavelet with n input dimensions of x_i is defined as

$$\psi_j(\mathbf{x}) = \prod_{i=1}^{n}\mu_j(x_i) = \prod_{i=1}^{n}\cos\left(1.75 \times \frac{x_i-b_j}{a_j}\right)\exp\left(-\frac{(x_i-b_j)^2}{2a_j^2}\right) \text{ for } j = 1,2,\ldots,r. \quad (4)$$

Layer 3 (output layer): According to the theory of MRA (Daubechies, 1992), the kth output of the WNNs using a linear combination of wavelets at different resolution levels is represented as

$$\hat{y}_k = \sum_{j=1}^{r} y_{jk} = \sum_{j=1}^{r} w_{jk}\psi_j = \sum_{j=1}^{r} w_{jk}\prod_{i=1}^{n}\cos\left(1.75 \times \frac{x_i-b_j}{a_j}\right)\exp\left(-\frac{(x_i-b_j)^2}{2a_j^2}\right) \quad k = 1,2,\ldots,p. \quad (5)$$

2.2. WNNs-based identification of nonlinear systems

In general, an unknown nonlinear system can be expressed as

$$\mathbf{y}(t+1) = f(\mathbf{y}(t), \mathbf{y}(t-1), \cdots, \mathbf{y}(t-n_y), \mathbf{u}(t), \mathbf{u}(t-1), \cdots, \mathbf{u}(t-n_u)) \quad (6)$$

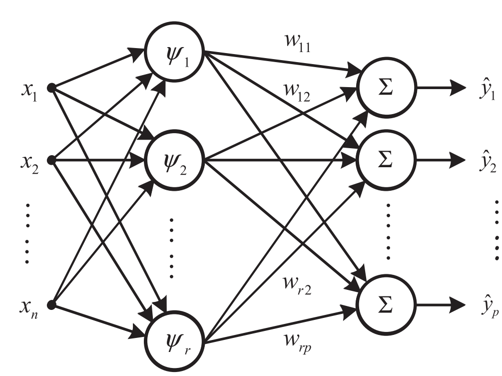

Fig. 1. Structure of WNNs that consists of input, wavelet, and output layers.