Neural Networks 42 (2013) 1–27

Contents lists available at SciVerse ScienceDirect

Neural Networks

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

Wavelet neural networks: A practical guide

Antonios K. Alexandridis [a,*], Achilleas D. Zapranis [b]

[a] University of Kent, School of Mathematics, Statistics and Actuarial Science, Cornwallis Building, P.O. CT2 7NF, Canterbury, UK

[b] University of Macedonia of Economic and Social Sciences, Department of Accounting and Finance, 156 Egnatia St., P.O. 54006, Thessaloniki, Greece

ARTICLE INFO

Article history:
Received 21 November 2010
Received in revised form 25 May 2012
Accepted 13 January 2013

Keywords:
Wavelet networks
Model identification
Variable selection
Model selection
Confidence intervals
Prediction intervals

ABSTRACT

Wavelet networks (WNs) are a new class of networks which have been used with great success in a wide range of applications. However a general accepted framework for applying WNs is missing from the literature. In this study, we present a complete statistical model identification framework in order to apply WNs in various applications. The following subjects were thoroughly examined: the structure of a WN, training methods, initialization algorithms, variable significance and variable selection algorithms, model selection methods and finally methods to construct confidence and prediction intervals. In addition the complexity of each algorithm is discussed. Our proposed framework was tested in two simulated cases, in one chaotic time series described by the Mackey–Glass equation and in three real datasets described by daily temperatures in Berlin, daily wind speeds in New York and breast cancer classification. Our results have shown that the proposed algorithms produce stable and robust results indicating that our proposed framework can be applied in various applications.

1. Introduction

Wavelet networks are a new class of networks that combine the classic sigmoid neural networks (NNs) and the wavelet analysis (WA). WNs have been used with great success in a wide range of applications. However a general accepted framework for applying WNs is missing from the literature. In this study, we present a complete statistical model identification framework in order to apply WNs in various applications. To our knowledge we are the first to do so. Although a vast literature about WNs exists, to our knowledge this is the first study that presents a step by step guide for model identification for WNs. Model identification can be separated in two parts, model selection and variable significance testing. In this study a framework similar to the one proposed by Zapranis and Refenes (1999) for the classical sigmoid NNs is adapted. More precisely, the following subjects were thoroughly examined: the structure of a WN, training methods, initialization algorithms, variable significance and variable selection algorithms, model selection methods and finally methods to construct confidence and prediction intervals. Only in Iyengar, Cho, and Phoha (2002) some of these issues are studied to some extent.

WA has proved to be a valuable tool for analyzing a wide range of time-series and has already been used with success in image processing, signal de-noising, density estimation, signal and image compression and time-scale decomposition. WA is often regarded as a "microscope" in mathematics, (Cao, Hong, Fang, & He, 1995), and it is a powerful tool for representing nonlinearities, (Fang & Chow, 2006). The major drawback of the WA is that it is limited to applications of small input dimension. The reason is that the construction of a wavelet basis is computationally expensive when the dimensionality of the input vector is relatively high, (Zhang, 1997).

On the other hand NNs have the ability to approximate any deterministic non-linear process, with little knowledge and no assumptions regarding the nature of the process. However the classical sigmoid NNs have a series of drawbacks. Typically, the initial values of the NN's weights are randomly chosen. Random weight initialization is generally accompanied with extended training times. In addition, when the transfer function is of sigmoidal type, there is always significant change that the training algorithm will converge to local minima. Finally, there is no theoretical link between the specific parameterization of a sigmoidal activation function and the optimal network architecture, i.e. model complexity (the opposite holds true for WNs).

In Pati and Krishnaprasad (1993) it has been demonstrated that it is possible to construct a theoretical formulation of a feedforward NN in terms of wavelet decompositions. WNs were proposed by Zhang and Benveniste (1992) as an alternative to feedforward NNs which would alleviate the aforementioned weaknesses associated with each method. The WNs are a generalization of radial basis function networks (RBFN). WNs are one hidden

* Corresponding author.
E-mail addresses: A.Alexandridis@kent.ac.uk (A.K. Alexandridis), zapranis@uom.gr (A.D. Zapranis).

0893-6080/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2013.01.008

layer networks that use a wavelet as an activation function, instead of the classic sigmoidal family. It is important to mention here that the multidimensional wavelets preserve the "universal approximation" property that characterizes NNs. The nodes (or wavelons) of WNs are the wavelet coefficients of the function expansion that have a significant value. In Bernard, Mallat, and Slotine (1998) various reasons were presented in why wavelets should be used instead of other transfer functions. In particular, firstly, wavelets have high compression abilities, and secondly, computing the value at a single point or updating the function estimate from a new local measure, involves only a small subset of coefficients.

WNs have been used in a variety of applications so far, i.e. in short term load forecasting, (Bashir & El-Hawary, 2000; Benaouda, Murtagh, Starck, & Renaud, 2006; Gao & Tsoukalas, 2001; Ulugammai, Venkatesh, Kannan, & Padhy, 2007; Yao, Song, Zhang, & Cheng, 2000), in time series prediction, (Cao et al., 1995; Chen, Yang, & Dong, 2006; Cristea, Tuduce, & Cristea, 2000), signal classification and compression, (Kadambe & Srinivasan, 2006; Pittner, Kamarthi, & Gao, 1998; Subasi, Alkan, Koklukaya, & Kiymik, 2005), signal denoising, (Zhang, 2007), static, dynamic (Allingham, West, & Mees, 1998; Oussar & Dreyfus, 2000; Oussar, Rivals, Presonnaz, & Dreyfus, 1998; Pati & Krishnaprasad, 1993; Postalcioglu & Becerikli, 2007; Zhang & Benveniste, 1992), and nonlinear modeling, (Billings & Wei, 2005), nonlinear static function approximation, (Jiao, Pan, & Fang, 2001; Szu, Telfer, & Kadambe, 1992; Wong & Leung, 1998), to mention the most important. In Khayamian, Ensafi, Tabaraki, and Esteki (2005) WN were even proposed as a multivariate calibration method for simultaneous determination of test samples of copper, iron, and aluminum.

In contrast to classical "sigmoid NNs", WNs allow for constructive procedures that efficiently initialize the parameters of the network. Using wavelet decomposition a "wavelet library" can be constructed. In turn, each wavelon can be constructed using the best wavelet of the wavelet library. The main characteristics of these procedures are: (i) convergence to the global minimum of the cost function, (ii) initial weight vector into close proximity of the global minimum, and as a consequence drastically reduced training times, (Zhang, 1997; Zhang & Benveniste, 1992). In addition, WNs provide information for the relative participation of each wavelon to the function approximation and the estimated dynamics of the generating process. Finally, efficient initialization methods will approximate the same vector of weights that minimize the loss function each time.

In Zapranis and Alexandridis (2008) and Zapranis and Alexandridis (2009) we give a concise treatment of wavelet theory. For a complete theoretical background on wavelets and wavelet analysis refer to (Daubechies, 1992; Mallat, 1999). Here the emphasis is in presenting the theory and mathematics of wavelet neural networks.

The rest of the paper is organized as follows. In Section 2 we present the WN. More precisely in Section 2.1 the structure of a WN is described. In Section 2.2 various initialization methods were described. In Section 2.3 a training method of the WN is presented and in Section 2.4 the stopping conditions of the training are described. In Section 2.5 the various initialization methods are compared and evaluated. A model selection algorithm is described in Section 3 and is evaluated in two cases in Section 3.1. Next, various criteria for selecting significant variables are presented while a variable selection algorithm is analytically presented in Section 4.1. In Section 4.2 the proposed algorithm is evaluated in two cases. In Section 5 methods to estimate the model and variance uncertainty are described. In Section 5.1 a framework for constructing confidence intervals is presented while in Section 5.2 a framework for constructing prediction intervals is presented. In Section 5.3 the proposed framework for constructing confidence and prediction intervals is evaluated in two cases. In Section 6 the proposed framework is applied in real data described by temperature in Berlin. Similarly, our framework is applied in wind speed data in Section 7. In Section 8 a WN is constructed for breast cancer classification while in Section 9 the proposed framework is applied in modeling and predicting the chaotic Mackey–Glass equation. Finally, in Section 10 we conclude.

2. Wavelet neural networks for multivariate process modeling

2.1. Structure of a wavelet network

In this section the structure of a WN is presented and discussed. A WN usually has the form of a three layer network. The lower layer represents the input layer, the middle layer is the hidden layer and the upper layer is the output layer.

In the input layer the explanatory variables are introduced to the WN. The hidden layer consists of the hidden units (HUs). The HUs are often referred as wavelons, similar to neurons in the classical sigmoid NNs. In the hidden layer the input variables are transformed to dilated and translated version of the mother wavelet. Finally, in the output layer the approximation of the target values is estimated.

The idea of a WN is to adapt the wavelet basis to the training data. Hence, the wavelet estimator is expected to be more efficient than a sigmoid NN, (Zhang, 1993). In Billings and Wei (2005), Kadambe and Srinivasan (2006), Mellit, Benghamen, and Kalogirou (2006), Xu and Ho (1999) an adaptive WN was used. In Chen et al. (2006) a local linear WN was proposed. The difference is that the connections weights between the hidden layer and output layer are replaced by a local linear model. In Fang and Chow (2006) and Jiao et al. (2001) a multiwavelet NN is proposed. In this structure, the activation function is a linear combination of wavelet bases instead of the wavelet function. During the training phase, the weights of all wavelets are updated. The multiwavelet NN is also enhanced by the DWT. Their results indicate that the proposed model increases the approximation capability of the network. In Khayamian et al. (2005) a principal component-wavelet NN was introduced. In this context, first principal component analysis (PCA) has been applied to the training data in order to reduce the dimensionality. Then a WN was used for function approximation. In Zhao, Chen, and Shen (1998) a multidimensional wavelet-basis function NN was used. More precisely (Zhao et al., 1998) use a multidimensional wavelet function as the activation function in the hidden layer. Then the sigmoid function was used as an activation function in the output layer. (Becerikli, 2004) proposes a network with unconstrained connectivity and with dynamic elements (lag dynamics) in its wavelet processing units called dynamic WN.

In this study, we implement a multidimensional WN with a linear connection between the wavelons and the output. Moreover, in order for the model to perform well in the presence of linearity, we use direct connections from the input layer to the output layer. Hence, a network with zero HUs is reduced to the linear model.

The structure of a single hidden-layer feedforward WN is given in Fig. 1. The network output is given by the following expression:

$$g_\lambda(\mathbf{x};\mathbf{w}) = \hat{y}(\mathbf{x}) = w_{\lambda+1}^{[2]} + \sum_{j=1}^{\lambda} w_j^{[2]} \cdot \Psi_j(\mathbf{x}) + \sum_{i=1}^{m} w_i^{[0]} \cdot x_i. \qquad (1)$$

In that expression, $\Psi_j(\mathbf{x})$ is a multidimensional wavelet which is constructed by the product of m scalar wavelets, $\mathbf{x}$ is the input vector, m is the number of network inputs, λ is the number of HUs