Neurocomputing 134 (2014) 289–295

Contents lists available at ScienceDirect

Neurocomputing

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

Innovative NARX recurrent neural network model for ultra-thin shape memory alloy wire

Han Wang, Gangbing Song *

Department of Mechanical Engineering, University of Houston, 4800 Calhoun Rd., Houston, TX 77204, USA

ARTICLE INFO

Article history:
Received 20 May 2013
Received in revised form
5 August 2013
Accepted 16 September 2013
Available online 15 January 2014

Keywords:
NARX
RNN
Jordan network
Elman network
SMA wire

ABSTRACT

The Nonlinear Autoregressive model with Exogenous inputs (NARX) has been utilized in many dynamic systems with complicated nonlinearities. Since NARX models can employ past values of time series as inputs, it is possible to estimate the property of hysteresis to Shape Memory Alloys (SMAs). The innovation of this paper lies in the development of a creative Jordan-plus-Elman NARX recurrent neural network (Jordan–Elman network) model as well as its training procedure. In this paper, the proposed model is applied to an ultra-thin SMA wire with a diameter of 0.001 in., which can be actuated/heated by an electric current. Experimental results demonstrate that the Jordan–Elman network dramatically improves the modeling error (mean squared error) in comparison with a Jordan NARX neural network. In addition, with good generalization results, the proposed model successfully identifies and estimates the hysteretic behavior of the ultra-thin SMA wire including major loops and minor loops at various frequencies.

1. Introduction

Shape Memory Alloys (SMAs), such as NiTi-based alloys, have been used as one of the most promising smart actuators in many engineering applications including civil [1], mechanical [2], biomedical [3], aerospace [4], etc. SMAs can exhibit two major properties, the shape memory effect and pseudoelasticity. The shape memory effect is based on the transition between two solid phases: the martensite phase for relative low temperatures, and the austenite phase for relative high temperatures. This property can be utilized to generate force or motion by electrically heating the material. Pseudoelasticity refers to the ability of the material in a relatively high temperature to accommodate strains of this magnitude during loading, and then recover upon unloading without permanent deformations.

SMAs in wire form have been commonly used in many applications, since they are generally the least expensive and most readily available from [5,6]. The distinctly non-uniform strain and temperature fields have great implications on the performance, reliability and controllability of practical devices. Due to the difficulty in accurately controlling the martensite–austenite proportion, SMAs have traditionally been used as "on–off" electromechanical actuators. The main difficulty of controlling SMAs is caused by the hysteretic behavior. Without a proper pre-defined memory or some similar algorithm, the controller should not be sufficient to compensate the hysteresis. Researchers have developed some advanced control methods by using inverse models of SMAs based on the corresponding mathematical model for the thermo-mechanical behavior covering hysteresis [7], as it has been done for some other hysteretic materials [8]. However, good models for phase transitions are very difficult to obtain precisely.

The dynamic behavior of SMAs has been of great interest for actuator applications. Thus, the simulation of the SMA-actuator response (e.g. time, frequency) is highly reasonable as it facilitates and supports the design layout process. Due to the complexity of the material behavior and the limited experimental basis, the development of constitutive models for SMAs has been hampered for many years. Although researchers have been studying the constitutive models for the characterization of the SMA [9–12], these methods are mostly simplified in order to enable their solvability. Therefore, Artificial Neural Networks (ANNs), as a nonlinear approach, can be appropriately trained to learn the SMAs′ hysteretic behavior from experimental data, and to provide a forward control algorithm through its inverse model [13].

Since an Artificial Neural Network (ANN) with at least one hidden layer has been successfully proven to universally approximate arbitrary bounded non-constant functions [14], it has been successfully applied to many applications of prediction and modeling in areas such as financial [15], biomedical [16], engineering [17], communication [18], etc. Normally, a regular feedforward neural network only contains time series of inputs in the input layer, and predicts an output from the output layer. Some nonlinear behaviors, especially non-mapping, such as hysteresis, are

* Corresponding author.
E-mail address: gsong@uh.edu (G. Song).

0925-2312/$ - see front matter © 2014 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.neucom.2013.09.050

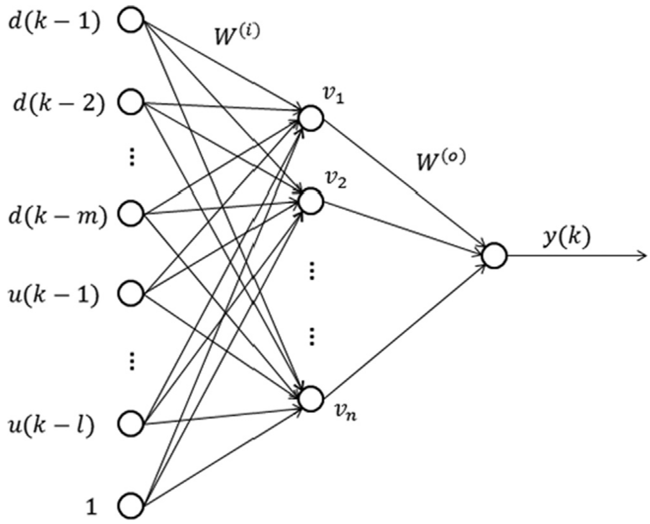

Fig. 1. Network structure of series–parallel mode Jordan NARX network.

normally difficult to estimate. By using advanced neural networks, researchers have developed many approaches to model hysteretic behaviors for various smart materials including piezoelectric [19] and SMAs [20]. Also, recurrent neural networks (RNNs) with hysteretic transfer functions have been introduced to model piezoelectric actuators [21]. RNNs have proven to be universal approximators in state space model forms [22]. Jordan networks and Elman networks are two major types of recurrent neural networks for modeling dynamic systems. The Jordan networks have information fed back from the output layer whereas the Elman networks use feedback from the hidden layer. Their structures make them suitable for representing various types of nonlinear dynamic systems, and thus many applications have been utilizing these two types of networks. Researchers have presented a Jordan network based NARX neural network model to capture the unknown dynamics of brain activities [23]. An innovative approach to model nonlinear complex systems based on internal RNNs as well as their modified backpropagation algorithms has been proposed [24]. Since Recurrent Neural Networks (RNNs) are very complex, normally the training and learning process may take a long period.

In this paper, two experimental modeling methods are introduced to estimate the time response of an ultra-thin SMA wire in the presence of hysteresis: one is a Jordan NARX recurrent neural network (Jordan network) model, and the other is an innovative Jordan-plus-Elman NARX RNN (Jordan–Elman network) model. A creative learning procedure is proposed to effectively compose a convergent RNN model and efficiently reduce the training time.

This paper is organized as follows. Section 2 describes the modeling methodology of three types of RNN NARX models: series–parallel model Jordan NARX network, parallel mode Jordan NARX network and Jordan-plus-Elman NARX network as well as the associated training procedures. Section 3 presents the experimental setup in Smart Material and Structure Laboratory, University of Houston to verify the effectiveness of the proposed modeling methodology. Section 4 compares the experimental results with modeling simulation results, and discusses the results for various modeling methodology. Section 5 provides a conclusion and potential application for future work.

2. Modeling methodology

In this section, three recurrent network structures will be introduced: the series–parallel mode Jordan NARX networks, the parallel mode Jordan NARX networks and the Jordan-plus-Elman NARX networks. Each network structure will be presented and discussed in the following subsection, and the experimental results can be found in Section 4.

The Nonlinear Autoregressive model with Exogenous inputs (NARX) is a commonly used discrete nonlinear system that can be mathematically represented as

$$y(k) = f\{u(k-1),\ u(k-2),\ \ldots,\ u(k-l),\ y(k-1),\ y(k-2),\ \ldots,\ y(k-m)\}, \tag{1}$$

where $u(k) \in \mathbb{R}$ and $y(k) \in \mathbb{R}$ denote the inputs and the outputs of the NARX model at the discrete time step, k, respectively. $l \geq 0$ and $m \geq 0$ are the input memory and the output memory used in the NARX model. The unknown function $f(\cdot)$, that is generally nonlinear, can be approximated, for instance, by a regular multilayer feedforward network. The resulting model architecture is called a NARX network or Jordan NARX network. A powerful class of dynamical models has been shown to be computationally equivalent to Turing machines [25]. Due to its complex structure, concerned in training the Jordan NARX network, two modes are introduced: the series–parallel mode and the parallel mode.

2.1. Series–parallel mode Jordan NARX network (Jordan-SP)

The network structure of the series–parallel mode Jordan NARX network (Jordan-SP) is presented in Fig. 1. $v_1, v_2, \ldots, v_n$ are the neurons in the hidden layer; $W^{(i)}$ is the weight matrix from the input layer to the hidden layer and $W^{(o)}$ is the weight vector from the hidden layer to the output layer.

The structure is a regular Feedforward Neural Network (FNN) structure, in which the output's regressor is formed only by the actual data of the system's output:

$$\hat{y}(k) = f\{u(k-1),\ u(k-2),\ \ldots,\ u(k-l),\ d(k-1),\ d(k-2),\ \ldots,\ d(k-m)\}, \tag{2}$$

where $d(k)$ is the desired or actual output data and $\hat{y}(k)$ is the network's estimated output at the time step, k. The state of the p-th neuron and its output are defined as

$$x_p(k) = \sum_{q=1}^{l+m+1} W_{pq}^{(i)} u_q(k), \tag{3}$$

$$v_p(k) = f(x_p(k)) \tag{4}$$

where the subscripts stand for the index of the element in a vector or a matrix, $W_{pq}^{(i)}$ is the weight connecting the q-th input and the p-th neuron in the hidden layer. Then the output of the network is

$$y(k) = \sum_{p=1}^{n} W_p^{(o)} v_p(k) \tag{5}$$

In this paper, the hyperbolic tangent sigmoid function, as shown in (6), is applied to the hidden layer and the linear function is applied to the output layer.

$$f(x) = \frac{2}{1+e^{-2x}} - 1 \tag{6}$$

Since the motivation is to let the neural network model track the desired outputs, the cost function, in which the Mean Squared Error (MSE) method is utilized in this paper, should be minimized between the desired outputs and the network's estimated output. Thus, the cost function for N steps is

$$J = \frac{1}{N} \sum_{k=1}^{N} (\hat{y}(k) - d(k))^2, \tag{7}$$

which is a function of $W^{(i)}$ and $W^{(o)}$.

In the training process, for example, by using the Levenberg–Marquardt backpropagation algorithm [26,27], error functions in