Neural Networks 50 (2014) 175–182

Contents lists available at ScienceDirect

Neural Networks

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

Finite time convergent learning law for continuous neural networks

Isaac Chairez *

Professional Interdisciplinary Unit of Biotechnology at the Instituto Politecnico Nacional, Av. Acueducto de Guadalupe sn, Col. Barrio La Laguna, Del. Gustavo A. Madero, Mexico, D.F., Mexico

ARTICLE INFO

Article history:
Received 2 November 2012
Received in revised form 17 October 2013
Accepted 18 October 2013

Keywords:
Neural networks
Learning algorithms
Super-twisting
Lyapunov method
Discontinuous systems

ABSTRACT

This paper addresses the design of a discontinuous finite time convergent learning law for neural networks with continuous dynamics. The neural network was used here to obtain a non-parametric model for uncertain systems described by a set of ordinary differential equations. The source of uncertainties was the presence of some external perturbations and poor knowledge of the nonlinear function describing the system dynamics. A new adaptive algorithm based on discontinuous algorithms was used to adjust the weights of the neural network. The adaptive algorithm was derived by means of a non-standard Lyapunov function that is lower semi-continuous and differentiable in almost the whole space. A compensator term was included in the identifier to reject some specific perturbations using a nonlinear robust algorithm. Two numerical examples demonstrated the improvements achieved by the learning algorithm introduced in this paper compared to classical schemes with continuous learning methods. The first one dealt with a benchmark problem used in the paper to explain how the discontinuous learning law works. The second one used the methane production model to show the benefits in engineering applications of the learning law proposed in this paper.

1. Introduction

Neural networks (NN) have been exploited since thirty years ago to solve many problems in control theory (Hopfield, 1982), pattern recognition and some others scientific and technological fields (Bishop, 2008; Han, Wang, & Qiao, 2013; Noorgard, Ravn, Poulsen, & Hansen, 2000; Siegelmann & Sontag, 1994). In particular, the several so-called weight adjustment methods or learning laws were developed using different methodologies (Haykin, 1994; Poznyak, Sanchez, & Yu, 2001) such as backpropagation and momentum just to mention a few (Fukushima, 2013). For the well known static NN, several schemes based on static optimisation methods were developed with relative success in different areas (Xingjian, 2012). Later on, recursive methods helped to introduce the learning laws for recurrent NN using different least mean square methods (LMS) (Chen, Cowan, & Grant, 1991). Recently, differential NN used continuous versions of the same LMS or adaptive schemes to produce the learning laws (Lewis, Yesildirek, & Liu, 2001). Nevertheless, most of the previous algorithms just provided asymptotic convergence for the parameters involved in the NN description (Mathias & Rech, 2012).

The emergence of new tools for analyzing and designing discontinuous algorithms using Lyapunov functions has opened new horizons in different areas (Moreno, 2010; Polyakov & Poznyak, 2011). In particular, a special class of generalised super-twisting algorithm was developed to prove finite time convergence for affine systems with respect to parameters (Moreno & Guzman, 2011).

In this paper, a new method is presented to produce a novel type of learning law for a class of continuous NN with access to the whole state of the uncertain system. The method is based on the particular application of the so-called generalised super-twisting algorithm and a new type of Lyapunov-like method to analyse the convergence of discontinuous algorithms.

The paper is organised as follows: Section 2 defines the approximation scheme based on NN with continuous dynamics. Section 3 describes the identifier introduced in this paper as well as the finite-time convergent learning law. Section 4 defines the problem statement considered here. Section 5 gives the main contribution regarding the learning law associated with the NN with continuous dynamics. Section 6 describes the numerical results used to demonstrate the benefits achieved by the method introduced in this paper. Section 8 concludes the paper with some remarks.

2. The uncertain system and its approximation based on the NN

The class of nonlinear dynamics with the real-valued state considered in this paper is characterised by the following mathematical model:

$$\frac{d}{dt}x(t) = f(x(t), u(t)) + \xi(t) \tag{1}$$
$$x(0) = x_0, \quad \|x_0\| < \infty$$

* Tel.: +52 5514019307.
E-mail address: jchairez@ctrl.cinvestav.mx.

0893-6080/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.neunet.2013.10.004

where $x \in \Re^n$; $x := [x_1, \ldots, x_n]^\top$ defines the system state. The function $u \in \Re^m$, $m < n$ represents an exogenous input or a feedback control action that fulfils

$$\|u(t)\| \leq u^+ \quad \forall t \geq 0 \quad u^+ \in \Re^+.$$

Anyway, this signal is assumed to be measurable. In this paper, the state is also assumed to be measurable. The system (1) is assumed to be stable; therefore $\|x(t)\| < \infty \; \forall t \geq 0$. Particularly, the following condition is required to solve the uncertain system and to obtain the convergence regimen of the identifier based on the continuous NN.

Condition 1: The function $f(x, u) : \Re^{n+m} \to \Re^n$ is Lipschitz with respect to its first argument, that is

$$\|f(x, u) - f(y, u)\|^2 \leq L\|x - y\|^2$$

for $x, y \in \Re^n$ and $L \in \Re^+$ is a positive constant. The source of uncertainty is not coming only from the low level of knowledge associated with the mathematical structure of $f(x, u)$, but also from the presence of external noises or/and perturbations in the system dynamics x. This source of perturbation is formally represented by $\xi \in \Re^n \; \forall t \geq 0$ and fulfils the following inclusion:

$$\|\xi(t)\|^2 \leq \xi^+ \quad \forall t \geq 0 \quad \xi^+ \in \Re^+.$$

The non-parametric mathematical model will be obtained using a particular type of the least mean square algorithm. Using this method demands a very important assumption: the nonlinear function $f(x, u)$ should admit a numerical reconstruction based on NN. This NN is represented by $f_0(x, u)$. This assumption is paramount to admit the existence of solution for the adaptive modelling problem. A number of possible approximations may be used here. Among others, the classical least mean square based on basis functions such as polynomials, sinusoids functions, wavelet functions and NN is a suitable option. No matter what selection is made to obtain the approximation, the following construction to represent the uncertain system should be considered:

$$\frac{d}{dt}x(t) = Ax(t) + f_0(x(t), u(t)) + \eta(x(t)) + \xi(t) \tag{2}$$

where $\eta(x) : \Re^n \to \Re^n$ is used to include the modelling errors generated by the approximation used in this paper. The matrix $A \in \Re^{n\times n}$ is introduced to represent a feasible linear part of the uncertain system (1). The class of systems analysed in this study and the assumption on the existence of a solution lead us to consider that the approximation error fulfils the following sector restriction:

$$\|\eta(x)\|^2 \leq \eta_0 + \eta_1\|x\|^2 \tag{3}$$

with η_0 and η_1 are positive scalars. The Stone–Weierstrass theorem claims that if the number of basis functions is tending to infinity, the approximation of the uncertain function will be exact. Nevertheless, it is practically impossible to construct a suitable numerical algorithm with such a characteristic. In this paper, a finite number of basis functions are used. That is why we consider the modelling error included in $\eta(x)$ with the characteristics described in (3). Here one must note that $\eta(x)$ is not including the control function u explicitly. However, this is not needed because the control action can be either bounded or locally linear with respect to the state, that is

$$\|u\| \leq \begin{cases} u^+ \\ l^+\|x\| \end{cases}$$

with u^+; l^+ positive scalars. These two conditions are included in the condition given in (3) and do not fail to fulfil the constrain given in the last inequality. The first order bound of u with respect to x is valid because the approximation achieved by the NN is also local.

Usually, the nominal part $f_0(x, u)$ is represented by linear combinations of continuous functions $\Gamma(x, u)$ such as it was explained in different Ref. (Chairez, 2009) regarding the approximation capability showed by different NN. Therefore, the so-called nominal section is approximated using the classical linear regression form, that is

$$f_0(x, u) := W\Gamma(x, u). \tag{4}$$

Here $\Gamma(x, u)$ represents the set of activation functions used to reconstruct the uncertain system described in (1). Generally, this function is defined as

$$\Gamma(x, u) := \begin{cases} \psi_1(x) \\ \psi_2(x)u. \end{cases}$$

The functions $\psi_1(x)$ and $\psi_2(x)$ are the basis functions of the same Hilbert space described above. Additionally, W are parameters used to adjust the contribution of each basis function required to obtain the approximation result. In previous results (Poznyak et al., 2001), the approximation parameters have been proposed as a combination of linear and nonlinear terms. Particularly, in the NN, one of the widely used activation functions is

$$\psi_{s,r}(x) := \tanh(c_r^\top x) \quad s = 1, 2; \; r = 1, \ldots, N; \; c_r \in \Re^n.$$

Here $\tanh(\cdot)$ is a real valued hyperbolic tangent that is continuous and bounded, that is

$$\|\psi_j(x) - \psi_j(\bar{x})\|^2 \leq L_{\psi_j}\|x - \bar{x}\|^2$$
$$\|\psi_j(x)\|^2 \leq \bar{L}_{\psi_j}^2$$
$$x, \bar{x} \in \Re^n \quad L_{\psi_j}, \bar{L}_{\psi_j} \in \Re^+.$$

3. The identifier based on the NN with continuous dynamics

The identifier based on the NN is proposed to follow the classical strategy used within the adaptive parameter identification framework. This construction used a structural copy of the approximation for an uncertain system based on the NN defined in (2). In particular, this identifier uses what is called a series structure because x is used in the differential equation for the identifier trajectories, namely $\hat{x}$. Therefore, the identifier based on the NN has the following structure (Lewis et al., 2001):

$$\begin{aligned} \frac{d}{dt}\hat{x}(t) &= Ax(t) - K_1\Phi_1(\Delta(t)) + \hat{W}(t)\Gamma(x(t), u(t)) + r(x(t)) \\ \frac{d}{dt}\hat{W}(t) &= -K_2\Phi_2(\Delta(t))\Gamma^\top(x(t), u(t)). \end{aligned} \tag{5}$$

In this expression, $\hat{W}(t)$ is the estimation for the matrix W. The couple of matrices $K_1 \in \Re^{n\times n}$ and $K_2 \in \Re^{n\times n}$ are gains that must be selected to achieve an accurate simultaneous estimation of both the state vector and parameters. The functions Φ_1 and Φ_2 are defined following the structure proposed in Cruz-Zavala, Moreno, and Fridman (2011). Therefore, these function are

$$\begin{aligned} \Phi_1(\Delta) &:= \mu_1\mathrm{diag}(|\Delta|^{1/2})S(\Delta) + \mu_2\Delta \\ \Phi_2(\Delta) &:= \mu_1^2 S(\Delta) + 1.5\mu_1\mu_2\mathrm{diag}(|\Delta|^{1/2})S(\Delta) + \mu_2^2\Delta. \end{aligned} \tag{6}$$

Here $\mu_1 \in \Re^{n\times n}$ and $\mu_2 \in \Re^{n\times n}$ are additional adjustable matrices that must be selected to ensure the convergence of the identification error defined by $\Delta := \hat{x} - x$ and the parameter error defined by $\tilde{W} = \hat{W} - W \in \Re^{2n\times n}$. The function $|\Delta(t)|^{1/2} \in \Re^n$ stands for the vector formed with the absolute values of the components of Δ, that is

$$|\Delta|^{1/2} = \left[|\Delta_1|^{1/2}, |\Delta_2|^{1/2}, \ldots, |\Delta_n|^{1/2}\right]^\top$$