Neurocomputing 165 (2015) 163–170

Contents lists available at ScienceDirect

Neurocomputing

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

Online optimal control of unknown discrete-time nonlinear systems by using time-based adaptive dynamic programming

Geyang Xiao [a,b], Huaguang Zhang [a,b,*], Yanhong Luo [a,b]

[a] College of Information Science and Engineering, Northeastern University, Box 134, 110819 Shenyang, PR China
[b] The Key Laboratory of Integrated Automation of Process Industry (Northeastern University) of the National Education Ministry, 110004 Shenyang, PR China

ARTICLE INFO

Article history:
Received 11 June 2014
Received in revised form
22 January 2015
Accepted 2 March 2015
Communicated by D. Liu
Available online 23 March 2015

Keywords:
Adaptive dynamic programming
Online optimal control
Reinforcement learning
Discrete-time systems

ABSTRACT

In this paper, an online optimal control scheme for a class of unknown discrete-time (DT) nonlinear systems is developed. The proposed algorithm using current and recorded data to obtain the optimal controller without the knowledge of system dynamics. In order to carry out the algorithm, a neural network (NN) is constructed to identify the unknown system. Then, based on the estimated system model, a novel time-based ADP algorithm without using system dynamics is implemented on an actor–critic structure. Two NNs are used in the structure to generate the optimal cost and the optimal control policy, and both of them are updated once at the sampling instant and thus the algorithm can be regarded as time-based. The persistence of excitation condition, which is generally required in adaptive control, is ensured by a new criterion while using current and recorded data in the update of the critic neural network. Lyapunov techniques are used to show that system states, cost function and control signals are all uniformly ultimately bounded (UUB) with small bounded errors while explicitly considering the approximation errors caused by the three NNs. Finally, simulation results are provided to verify the effectiveness of the proposed approach.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The theory of optimal control is concerned with finding a control law for a given system and user defined optimality criterion. Traditional optimal control design methods are generally offline and require complete knowledge of the system dynamics [1]. Adaptive control techniques on the other side are designed for online use of uncertain systems. However, classical adaptive control methods are generally far from optimal.

During the last few decades, reinforcement learning (RL) [2–4] has successfully provided a way to bring together the advantages of adaptive and optimal control. A class of RL-based adaptive optimal controllers, called approximate/adaptive dynamic programming (ADP), was first developed by Werbos [5,6]. Extensions of the RL-based controllers to DT systems have been considered by many researchers [7–20]. In [7], the authors attempted to solve the DT nonlinear optimal control problem offline using ADP approaches and neural networks by assuming that there are no NN reconstruction errors. Based on the results of [7], other researchers developed offline ADP approaches in some complicated situations, such as optimal tracking problem [8,14,21], optimal control with control constraints [10], optimal control with time delays [11,12], optimal control with finite approximation errors [15]. However, the above works are all required the knowledge of system dynamics and using offline tuning law.

Since the mathematical models of real-world system dynamics are often difficult to build, it has become one of the main foci of control practitioners to design the optimal controller for nonlinear systems with unknown dynamics. The work of [9] analyzed the convergence of unknown DT nonlinear systems using offline-trained neural networks, but this method introduced the Lebesgue integral [7], which required data of a subset of the plant, in the tuning law and thus spent too much time on off-line training. In [20], the authors developed one way to control the unknown DT nonlinear systems using globalized dual heuristic programming, and others employed the single network dual heuristic dynamic programming (SN-DHP) technique in the ADP algorithm in [19]. Both of them introduced the gradient-based adaptation tuning law instead of the way in [9]. However, without using recorded system data, iterations were needed in the tuning law [19,20] and the critic NN and actor NN could not be updated with respect to time at each sampling interval. Moreover, although [9,20,21] constructed a NN to identify the unknown system dynamics, they assumed that the NN identification error approached to zero, and thus the effects of the estimation error on the convergence of the actor–critic algorithms were not considered.

* Corresponding author at: The Key Laboratory of Integrated Automation of Process Industry (Northeastern University) of the National Education Ministry, 110004 Shenyang, PR China.
E-mail addresses: xgyalan@gmail.com (G. Xiao), hgzhang@ieee.org (H. Zhang), neuluo@gmail.com (Y. Luo).

http://dx.doi.org/10.1016/j.neucom.2015.03.006
0925-2312/© 2015 Elsevier B.V. All rights reserved.

On the other hand, online adaptive-optimal controller designs were presented in [17,18,22–24] to overcome the iterative offline training methodology. The central theme of the approaches in [24,23] as well as several works in [22] is that the cost function and optimal control signal are approximated by online parametric structures, such as NN. Although the proposed methods in [22–24] are verified via numerical simulations, the approximation errors are not considered and proofs of convergence are not demonstrated. The work of [17] presented a novel approach that relied on current and recorded system data for adaptation and proved the convergence while the approximation errors are considered, and recently the authors in [18] improved this method in the presence of unknown internal dynamics and called it time-based ADP algorithm. However, since the requirement of the knowledge of control coefficient matrix in the tuning law, the general time-based ADP algorithm [17,18] becomes invalid while dealing with the unknown DT nonlinear system. Meanwhile, most of the online adaptive optimal control algorithms with ADP require a persistence of excitation (PE) condition [25–27] that is important in NN identification. Refs. [17,18] proposed a similar condition to ensure the PE requirement, but they did not give the lower bound in the proof.

The contributions of this paper lie in the development of an online adaptive learning algorithm to solve an infinite horizon optimal control problem for unknown DT nonlinear systems. By performing identification process, the time-based ADP algorithm, which makes use of current and recorded system data, is applicable to deal with the optimal control problem of unknown nonlinear systems. However, the general time-based ADP technique requires knowing the system dynamics. By using current and recorded system information, the PE condition is ensured by a new criterion with explicit lower bound and the unknown nonlinear DT system can be controlled once at the sampling instant. Convergence of the system states and NN implementation is demonstrated while explicitly considering all the NN reconstruction errors in contrast to previous works [9,20,21].

2. Background

Consider the affine DT nonlinear system described by

$$x_{k+1}=f(x_k)+g(x_k)u(x_k) \tag{1}$$

where $x_k\in\mathbb{R}^n$, $f(x_k)\in\mathbb{R}^n$, $g(x_k)\in\mathbb{R}^{n\times m}$ and $u(x_k)\in\mathbb{R}^m$. Without loss of generality, assume that the system is controllable, sufficiently smooth, drift free, and that $x=0$ is a unique equilibrium point on a compact set Ω while the states are considered measurable. In the following part, $u(x_k)$ is denoted by u_k for simplicity.

Define the infinite horizon cost function

$$\begin{aligned}J(x_k)&=\sum_{n=k}^{\infty}Q(x_n)+u_n^{\mathrm{T}}Ru_n\\&=Q(x_k)+u_k^{\mathrm{T}}Ru_k+J(x_{k+1})\\&=\rho(x_k,u_k)+J(x_{k+1})\end{aligned} \tag{2}$$

where $Q(x_k)$ is a positive definite, $R\in\mathbb{R}^{m\times m}$ is a symmetric positive definite matrix, and $\rho(x_k,u_k)=Q(x_k)+u_k^{\mathrm{T}}Ru_k$ is the utility function.

In order to control (1) in an optimal manner, it is desired to select the control sequence u_k to minimize the cost function (2) for all x_k. Further, it is required that the control sequence u_k is admissible and $J(x_k=0)=0$ so that the cost function serves as a Lyapunov function.

According to Bellmans optimality principle, the infinite horizon optimal cost function $J^*(x_k)$ satisfies the DTHJB equation

$$J^*(x_k)=\min_{u_k}(\rho(x_k,u_k)+J^*(x_{k+1})) \tag{3}$$

where x_{k+1} can be derived from Eq. (1).

The optimal control u_k is found by solving $\partial J^*(x_k)/\partial u_k=0$, and then it is given by

$$u_k^*=-\frac{1}{2}R^{-1}g(x_k)^{\mathrm{T}}\frac{\partial J^*(x_{k+1})}{\partial x_{x+1}}. \tag{4}$$

The optimal control (4) for unknown DT nonlinear systems is generally unavailable due to its dependence on $g(x_k)$ and x_{k+1}. To circumvent these deficiencies, a NN identification scheme for unknown systems is presented next.

3. NN identification of the unknown nonlinear system

To begin the NN identifier construction, the system dynamics (1) are rewritten as

$$x_{k+1}=f(x_k)+g(x_k)u_k=H(x_k,u_k). \tag{5}$$

The function $H(x_k,u_k)$ has a NN representation on a compact set S according to the universal approximation property of NN, which can be written as

$$H(x_k,u_k)=W_{\mathrm{s}}^{\mathrm{T}}\theta(Y_{\mathrm{s}}^{\mathrm{T}}z_{\mathrm{s}}(k))+\varepsilon_{sk}=W_{\mathrm{s}}^{\mathrm{T}}\theta(\bar{z}_{\mathrm{s}}(k))+\varepsilon_{sk} \tag{6}$$

where $W_{\mathrm{s}}\in\mathbb{R}^{l\times n}$ and $Y_{\mathrm{s}}\in\mathbb{R}^{(n+m)\times l}$ are the constant ideal weight matrices. l is the number of hidden layer neurons. $\theta(\cdot)$ is the NN activation function, $z_{\mathrm{s}}(k)=[x_k^{\mathrm{T}}\ u_k^{\mathrm{T}}]^{\mathrm{T}}$ is the NN input and let $\bar{z}_{\mathrm{s}}(k)=Y_{\mathrm{s}}^{\mathrm{T}}z_{\mathrm{s}}(k)$, ε_{sk} is the bounded NN functional approximation error and satisfies $\|\varepsilon_{sk}\|\le\varepsilon_{\mathrm{sM}}$ and $\|\varepsilon'_{sk}\|\le\varepsilon'_{\mathrm{sM}}$ for constants $\varepsilon_{\mathrm{sM}}$ and $\varepsilon'_{\mathrm{sM}}$ respectively. Additionally, the NN activation functions and their gradients are bounded such that $\|\theta(\cdot)\|\le\theta_{\mathrm{M}}$ and $\|\theta'(\cdot)\|\le\theta'_{\mathrm{M}}$ for constants θ_{M} and θ'_{M} respectively.

During the system identification process, keep Y_{s} constant while only tune W_{s}, the identification scheme is then defined as

$$\hat{x}_{k+1}=\hat{W}_{\mathrm{s}}^{\mathrm{T}}(k)\theta(\bar{z}_{\mathrm{s}}(k)) \tag{7}$$

where $\hat{x}_k$ is the estimated system state vector, and $\hat{W}_{\mathrm{s}}$ is the estimation of the ideal constant weight matrix W_{s}. The parameter estimation error is defined as $\tilde{W}_{\mathrm{s}}(k)=W_{\mathrm{s}}-\hat{W}_{\mathrm{s}}(k)$.

Define the error performance as $E_{\mathrm{s}}(k+1)=\tilde{x}_{k+1}^{\mathrm{T}}\tilde{x}_{k+1}/2$, where $\tilde{x}_k=x_k-\hat{x}_k$. To minimize the error performance $E_{\mathrm{s}}(k+1)$, the weights tuning law are proposed as

$$\begin{aligned}\hat{W}_{\mathrm{s}}(k+1)&=\hat{W}_{\mathrm{s}}(k)-\alpha_{\mathrm{s}}\left[\frac{\partial E_{\mathrm{s}}(k+1)}{\partial\hat{W}_{\mathrm{s}}(k)}\right]\\&=\hat{W}_{\mathrm{s}}(k)-\alpha_{\mathrm{s}}\theta(\bar{z}_{\mathrm{s}}(k))\tilde{x}_{k+1}^{\mathrm{T}}\end{aligned} \tag{8}$$

where $\alpha_{\mathrm{s}}>0$ is the learning rate.

Theorem 1 (*Liu et al. [20]*). *Let the identification scheme (7) be used to identify the nonlinear system (1), and let the parameter update law (8) be used to tune the NN weights. Then, the state estimation error dynamics $\tilde{x}_k$ is asymptotically stable while the parameter estimation error $\tilde{W}_{\mathrm{s}}(k)$ is bounded.*

After a sufficient learning session, the estimation error can be denoted as

$$\tilde{x}_{k+1}=W_{\mathrm{s}}^{\mathrm{T}}\theta(\bar{z}_{\mathrm{s}}(k))-\hat{W}_{\mathrm{s}}^{\mathrm{T}}\theta(\bar{z}_{\mathrm{s}}(k))+\varepsilon_{sk}=\tilde{W}_{\mathrm{s}}^{\mathrm{T}}\theta(\bar{z}_{\mathrm{s}}(k))+\varepsilon_{sk}. \tag{9}$$

According to Theorem 1, we can assume $\|\tilde{W}_{\mathrm{s}}\|\le\tilde{W}_{\mathrm{sM}}$, where $\tilde{W}_{\mathrm{sM}}$ is a small positive constant. Then we can conclude that $\|\tilde{x}_{k+1}\|\le\tilde{x}_{\mathrm{M}}$ where $\tilde{x}_{\mathrm{M}}$ is a small bounded positive constant.

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
✓ امکان دانلود نسخه ترجمه شده مقالات
✓ پذیرش سفارش ترجمه تخصصی
✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
✓ دانلود فوری مقاله پس از پرداخت آنلاین
✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات