Automatica 81 (2017) 240–252

Contents lists available at ScienceDirect

Automatica

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

Novel iterative neural dynamic programming for data-based approximate optimal control design[☆]

Chaoxu Mu[a], Ding Wang[b], Haibo He[c,1]

[a] Tianjin Key Laboratory of Process Measurement and Control, School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China
[b] The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[c] Department of Electrical, Computer and Biomedical Engineering, University of Rhode Island, RI, USA

ARTICLE INFO

Article history:
Received 26 April 2016
Received in revised form
15 December 2016
Accepted 21 February 2017

Keywords:
Iterative neural dynamic programming (INDP)
Data-based control
Approximate optimal control
Heuristic dynamic programming (HDP)
Affine and non-affine nonlinear systems

ABSTRACT

As a powerful method of solving the nonlinear optimal control problem, the iterative adaptive dynamic programming (IADP) is usually established on the known controlled system model and is particular for affine nonlinear systems. Since most nonlinear systems are complicated to establish accurate mathematical models, this paper provides a novel data-based approximate optimal control algorithm, named iterative neural dynamic programming (INDP) for affine and non-affine nonlinear systems by using system data rather than accurate system models. The INDP strategy is built within the framework of IADP, where the convergence guarantee of the iteration is provided. The INDP algorithm is implemented based on the model-based heuristic dynamic programming (HDP) structure, where model, action and critic neural networks are employed to approximate the system dynamics, the control law and the iterative cost function, respectively. During the back-propagation of action and critic networks, the approach of directly minimizing the iterative cost function is developed to eliminate the requirement of establishing system models. The neural network implementation of the INDP algorithm is presented in detail and the associated stability is also analyzed. Simulation studies are conducted on affine and non-affine nonlinear systems, and further on the manipulator system, where all results have demonstrated the effectiveness of the proposed data-based approximate optimal control method.

1. Introduction

Since nonlinear systems are widely existed in most industrial fields, the optimal control problem of nonlinear systems has attracted great attention in recent several decades. The nonlinear optimal control problem is usually formulated as coping with the nonlinear Hamilton–Jacobi–Bellman (HJB) equation, which is often difficult to be solved (Bellman, 1957; Lewis & Syrmos, 1995; Si, Barto, Powell, & Wunsch, 2004). As is known to all, when the optimal control problem of linear systems is studied, the linear HJB equation can be evolved as the algebraic Riccati equation. The famous iterative solution strategy was proposed by converting the algebraic Riccati equation to a series of linear Lyapunov equations (Kleinman, 1968). Along this direction, the iterative solution strategy was extended to solve the approximate optimal control of a trainable manipulator in Saridis and Lee (1979). However, this iterative solution method only fits this kind of HJB equations that are linear partial differential equations. Motivated by these success, there has been a great deal of research developed to approximately solve the HJB equation with the great improvement of intelligent computation (Beard, Saridis, & Wen, 1997; Mu, Sun, Song, & Yu, 2016; Si et al., 2004; Wang, Liu, Wei, Zhao, & Jin, 2012; Werbos, 1992). In Werbos (1992), an adaptive/approximate dynamic programming (ADP) algorithm was proposed to approximately solve optimal control problems in forward time by involving neural networks for function approximation. The generalized HJB equation was formulated to solve the optimal control problem from a view of successive approximation (Beard et al., 1997). For continuous-time nonlinear systems, a nearly constrained-optimal state feedback control method using a neural network HJB approach was given in Abu-Khalaf and Lewis (2004), and was extended to synchronous policy iteration in Vamvoudakis

[☆] The material in this paper was not presented at any conference. C. Mu and D. Wang make equal contributions to this work and share the first authorship. This paper was recommended for publication in revised form by Associate Editor James Lam under the direction of Editor Ian R. Petersen.

E-mail addresses: cxmu@tju.edu.edu (C. Mu), ding.wang@ia.ac.cn (D. Wang), he@ele.uri.edu (H. He).

[1] Fax: +1 401 782 6422.

http://dx.doi.org/10.1016/j.automatica.2017.03.022
0005-1098/© 2017 Elsevier Ltd. All rights reserved.

and Lewis (2010). Simultaneously, for discrete-time nonlinear systems, the iterative adaptive dynamic programming (IADP) strategy was improved to obtain the approximate solution of the nonlinear HJB equations by using neural networks (Al-Tamimi, Lewis, & Abu-Khalaf, 2008; Dierks, Thumati, & Jagannathan, 2009; Wang, Jin, Liu, & Wei, 2011; Wang, Mu, & Liu, 2017; Zhang, Luo, & Liu, 2009). The value-iteration-based ADP algorithm was developed with several convergence results of both inner-loop and outer-loop iterations in Heydari (2014). In addition, there are several latest developments related to ADP, including approximation-error-based adaptive optimal control (Heydari, 2016), event-triggered optimal control design (Vamvoudakis, Mojoodi, & Ferraz, 2017; Wang, Mu, He, & Liu, in press; Wang, Mu, Liu, & Ma, in press), ADP-based variable structure or switching control design (Fan & Yang, 2016; Heydari & Balakrishnan, 2014b; Mu, Ni, Sun, & He, 2017), cooperative control of multi-agent systems (Heydari & Balakrishnan, 2014a; Zhang, Liang, Wang, & Feng, 2017), and so on.

In the industrial field, two prominent features are presented with the technological innovation and progress. One is that more and more real systems are facing the difficulty in establishing process models to support the control design due to increasing scales and complex operations. The other is that vast volume of data is stored during the industrial process but does not get used efficiently. Thus, the problem of data-based optimal control for nonlinear systems is significant and challenging. Recently, several data-based approximate optimal control approaches have been reported. For example, an online direct heuristic dynamic programming method was proposed by Si and Wang without requiring the controlled system model (Si & Wang, 2001), or was more specifically called neural dynamic programming (NDP), which was further developed to the tracking control problem of nonlinear systems (Yang, Liu, Wang, & Wei, 2014; Yang, Si, Tsakalis, & Rodriguez, 2009). The data-based online policy iteration approach was proposed to obtain adaptive optimal controllers for continuous-time linear systems with unknown system dynamics (Jiang & Jiang, 2012). A model-free approximate policy iteration method was developed based on a least-square weight updating for affine nonlinear continuous-time optimal control design (Luo, Wu, Huang, & Liu, 2015). Based on the identification of neural networks, a data-driven robust approximate optimal control was designed for the tracking control of continuous-time general nonlinear systems (Zhang, Cui, Zhang, & Luo, 2011). The robust ADP was studies for the robust optimal control design for a class of uncertain nonlinear systems (Jiang & Jiang, 2014). The approach of goal representation adaptive dynamic programming was proposed by adapting reinforcement signal (He, Ni, & Fu, 2012), which has been applied to tracking control problem (Mu, Ni, Sun, & He, 2016), maze navigation (Ni, He, Wen, & Xu, 2013) and power systems (Tang, Mu, & He, 2016).

Compared with the NDP algorithm, this proposed method is an off-line algorithm by integrating the cost function iteration and the control law iteration into the NDP approach, while the NDP algorithm is with the merit of online learning and control. Compared with the iterative adaptive dynamic programming (IADP) method, the proposed method has built the data-based learning control framework by using a model network, while the IADP strategy is usually established on the known controlled system model and is particularly effective for affine nonlinear systems even a model network is utilized in this method (Wang et al., 2012; Wang, Liu, Zhang, & Zhao, 2016). The contribution of this paper is summarized as follows. First, we propose the ε-optimal iterative ADP algorithm based on a prescribed error bound, where the convergence of the iterative algorithm as well as the equivalence of stopping criterion is proved from the view of theoretical analysis. Second, the INDP approach based on a HDP structure is developed to implement the data-based optimal control via estimating both the iterative control law and the iterative cost function. The novel design on the weight updating of the action neural network makes the implementation can be operated by only using system data, which has greatly improved the realization of the algorithm without involving the accurate system model. Third, by using a Lyapunov approach, the uniformly ultimately boundedness (UUB) stability is provided for the INDP controller.

This paper is organized as follows. In Section 2, the optimal control problem is formulated for general discrete-time nonlinear systems. Section 3 presents the ε-optimal INDP algorithm and the iteration convergence analysis. The implementation strategy of INDP algorithm and the associated stability proof are provided in Section 4. In Section 5, three simulation examples are given to demonstrate the effectiveness of the proposed data-based INDP approximate optimal control scheme. Finally, we summarize this paper in Section 6.

2. Problem statement and preliminaries

In this paper, the studied discrete-time nonlinear systems are generally described by

$$x_{t+1} = F(x_t, u_t), \quad t = 0, 1, 2, \dots \tag{1}$$

where $x_t = [x_{1t}, x_{2t}, \dots, x_{nt}]^T \in \mathbb{R}^n$ is the state vector at time step t, and $u_t = [u_{1t}, u_{2t}, \dots, u_{mt}]^T \in \mathbb{R}^m$ is the control vector at time step t. The system function $F(x_t, u_t)$ is Lipschitz continuous on $\Omega_x \subseteq \mathbb{R}^n$ and $F(0, 0) = 0$.

Definition 1 (*Werbos, 1992; Zhang et al., 2009*)**.** A nonlinear dynamical system is said to be stabilizable on a compact set $\Omega_x \subseteq \mathbb{R}^n$, if for any initial condition $x_0 \in \Omega_x$, there exists a control sequence $u_0, u_1, u_2, \dots, u_t \in \mathbb{R}^m$, such that the state $x_t \to 0$ as $t \to \infty$.

For the optimal control of discrete-time nonlinear system (1), it is expected to obtain an optimal control law u_t, which enables all the states of system (1) to stabilize at the origin and minimizes the following cost function $J(x_t)$,

$$J(x_t) = \sum_{k=t}^{\infty} \beta^{k-t} R(x_k, u_k), \tag{2}$$

where $R(x_k, u_k)$ is the utility function, $R(x_k, u_k) \geq 0$ for any x_k and u_k, and $R(0, 0) = 0$. β is the discount factor with $0 < \beta \leq 1$. Generally speaking, the utility function can be chosen as the quadratic form of the states and the control variables, which is as follows:

$$R(x_k, u_k) = x_k^T P x_k + u_k^T Q u_k, \tag{3}$$

where P and Q are symmetric positive definite matrices with appropriate dimensions.

A feedback control is used in this paper, such that $u_t = u(x_t)$. The admissible control is introduced for the optimal control problem, which stabilizes system (1) at the origin and guarantees that the total cost function (2) is finite.

Definition 2 (*Prokhorov, Santiago, & Wunsch, 1995; Si et al., 2004*)**.** A feedback control u_t defined on Ω_x is said to be admissible with respect to (2) if u_t is continuous on a compact set $\Omega_u \subseteq \mathbb{R}^m$, $u(0) = 0$, u_t stabilizes system (1) on Ω_x, and $J(x_0)$ is finite $\forall x_0 \in \Omega_x$.

Note that the infinite-horizon cost function can be rewritten in a recursive form, then Eq. (2) is rewritten as

$$\begin{aligned} J(x_t) &= x_t^T P x_t + u_t^T Q u_t + \beta \sum_{k=t+1}^{\infty} \beta^{k-t-1} R(x_k, u_k) \\ &= x_t^T P x_t + u_t^T Q u_t + \beta J(x_{t+1}). \end{aligned} \tag{4}$$