Neurocomputing 149 (2015) 1461–1468

Contents lists available at ScienceDirect

Neurocomputing

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

Stability analysis of heuristic dynamic programming algorithm for nonlinear systems

Tao Feng [a], Huaguang Zhang [a,*], Yanhong Luo [a], Jilie Zhang [b]

[a] Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110004, PR China
[b] College of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, PR China

ARTICLE INFO

Article history:
Received 17 March 2014
Received in revised form
15 August 2014
Accepted 20 August 2014
Communicated by Hongyi Li
Available online 6 September 2014

Keywords:
Convergence
Stability
Heuristic dynamic programming (HDP)
Optimal control
Value-iteration

ABSTRACT

In this paper, a value-iteration based heuristic dynamic programming (HDP) algorithm is developed to solve the optimal control for the continuous time affine nonlinear systems. First, a rigorous convergence proof of the HDP algorithm is given. Second, stability issues of the HDP algorithm for nonlinear systems are investigated. It is commonly believed that the main drawback of the HDP algorithm is that only the limit function of the iterative control sequence is proved to be stabilized, thus infinite iterations are executed. To confront this problem, we present a novel stability result for the HDP algorithm, which indicates that the resulting iterative control laws after finite iterations can guarantee the closed-loop stability. A similar stability result is also obtained for the discrete time nonlinear systems. Therefore, the practicality of the HDP algorithm is greatly improved. Single neural network (NN) structure is employed to implement the algorithm. It should be pointed that the algorithm can be implemented without knowing the internal dynamics of the systems. Finally, two numerical examples are given to demonstrate the effectiveness of the developed methods.

1. Introduction

One of the most basic design principles in the feedback control design is to guarantee the closed-loop stability of the nonlinear systems. Optimal control aims to design a feedback control law, which not only guarantees the system closed-loop stability, but also follows the optimal manner according to an overall performance index. In the past decades, a mountain of work has been done for the optimal control of nonlinear systems. Dynamic programming [1], which is proved to be a powerful method, has been extensively applied to generate the optimal control for nonlinear systems. However, one notable drawback of this method is the computing cost with the increasing dimension of the nonlinear systems, which is referred to as the "curse of dimensionality". Approximate dynamic programming (ADP) [2] methods have been proposed to circumvent this difficulty. Different from the DP methods, ADP solves the optimal control problems forward-in-time [2,3].

The optimal control for linear systems with respect to a quadric performance index can be achieved by solving the algebra Riccati equation (ARE). However, for nonlinear systems, the optimal feedback control depends on obtaining the solution to the Hamilton–Jacobi–Bellman (HJB) equation, which is challenging to solve directly due to its inherently nonlinear nature. To confront this difficulty, iterative methods have been proposed to obtain the solution of the HJB equation indirectly which can be roughly sorted into two classes [4]: policy-iteration and value-iteration. For the policy-iteration algorithm [5–11], all the iterative control laws stabilize the system, however, an initial stabilized control law is required, which is often difficult to obtain in practical applications.

For the value-iteration algorithm, an initial stabilized control law is not required. Zhang et al. [12] studied the near-optimal control for a class of discrete-time affine nonlinear systems with control constraints by the iterative DHP method. Al-Tamimi et al. [13] derived a value-iteration based HDP algorithm to solve the optimal control problems and provided a full rigorous convergence proof. In [14], the HDP algorithm has been used to solve the non-affine nonlinear systems with respect to a discounted performance index. An iterative value-iteration based ADP method has been proposed in [15] to solve a class of nonlinear zero-sum differential games. An iterative DHP algorithm has been proposed in [16] for optimally controlling a large class of nonlinear discrete-time systems affected by an unknown time variant delay and system uncertainties. In [17], Huang et al. proposed an optimal tracking control scheme based on HDP algorithm by transforming the original tracking problem into a regulation problem with respect to the state tracking error. A SN-DHP based technique has been developed in [18] to find the near optimal controller for

* Corresponding author.
E-mail addresses: sunnyfengtao@163.com (T. Feng), hgzhang@ieee.org (H. Zhang), neuluo@neu.edu.cn (Y. Luo), jilie0226@home.swjtu.edu.cn (J. Zhang).

http://dx.doi.org/10.1016/j.neucom.2014.08.046
0925-2312/

unknown affine nonlinear discrete-time systems. An on-line learning and control approach based on ADP for wind farm control and integration with the grid has been investigated in [19]. In [20], a greedy HDP algorithm has been developed to solve the zero-sum game problems for affine discrete-time systems, which can be used to solve the Hamilton–Jacobi–Isaacs equation associated with H_∞ optimal regulation control problems. The data-driven ADP methods have also received considerable attention recently. A model-free optimal control scheme for a class of linear discrete-time systems with multiple delays in state, control and output vectors has been developed in [21], where the optimal control can be obtained using only measured input/output data from systems by ADP technology. For more details, see [22–24] and references therein. However, most of the research on value-iteration focuses on the discrete-time nonlinear systems, the value-iteration based HDP algorithm for the continuous-time nonlinear systems remains unstudied. This motivates our work.

The main drawback of the value-iteration algorithm is that only the limit function of the iterative control sequence has been proved to be stabilized while the iterative control laws may be not [25]. This greatly limits the applications of the value-iteration algorithm. Li et al. [26] proposed general value-iteration (GVI) algorithm and Wei et al. [28] proposed a stable θ-ADP scheme, but the initial values of both are difficult to be obtained. Convergence of the ADP algorithm does not mean that the iterative control laws provide the closed-loop stability of the considered nonlinear systems. The closed-loop stability of the nonlinear systems must be guaranteed when the optimality is achieved. However, it is worthy noting that in the existing references, say [12–14,16–18], the optimal iterative control laws obtained by the value-iteration algorithm are indeed stabilized, rather than just the limit function of the iterative control sequence. A theoretical explanation for this phenomenon has not yet been given, to our best knowledge. In this paper, novel stability results for iterative control laws are proposed. It is proved that for the infinite horizon problem, the resulting iterative control laws after finite iterations can guarantee the closed-loop stability of the nonlinear systems, which greatly increases the practicability of the value-iteration based HDP algorithm.

The rest of the paper is organized as follows. In Section 2, the value-iteration based HDP algorithm for the continuous-time affine nonlinear systems is developed and a rigorous convergence proof is given. Novel stability results of the HDP algorithm for the continuous-time nonlinear systems are proposed. In Section 3, stability issues of the HDP algorithm for discrete-time nonlinear systems are investigated. NN implementations of the HDP algorithm are given in Section 4. Two simulation examples are employed in Section 5 to demonstrate the effectiveness of the developed methods.

2. HDP algorithm for continuous-time nonlinear systems

Consider the affine continuous-time nonlinear system of form

$$\dot{x}(t) = f(x(t)) + g(x(t))u(x(t)), \quad x(0) = x_0, \tag{1}$$

where $x(t) \in \mathbb{R}^n$ is the state vector and $u(t) \in \mathbb{R}^m$ is the input vector, $f(x(t)) \in \mathbb{R}^n$ and $g(x(t)) \in \mathbb{R}^{n\times m}$. It is assumed that $f(x(t)) + g(x(t))u(x(t))$ is Lipschitz continuous on a set $\Omega \subseteq \mathbb{R}^n$ which contains the origin, and that the dynamical system is stabilizable on Ω, which means that there exists a continuous control function $u(x(t)) \in \mathbb{R}^m$ such that the system is asymptotically stable on Ω.

We consider the following quadric performance index:

$$J(x(t)) = \int_t^\infty x(\tau)^T Q x(\tau) + u(x(\tau))^T R u(x(\tau)) \, d\tau, \tag{2}$$

where the state weighting matrix $Q \in \mathbb{R}^{n\times n}$ is nonnegative definite and the inputs weighting matrix $R \in \mathbb{R}^{m\times m}$ is positive definite. The objective is to find the control law $u(x(t))$ which minimizes the infinite-horizon cost function (2). Note that the control law $u(x(t))$ needs to be stabilized and guarantees that (2) is finite, i.e., the control law must be admissible [5].

2.1. Value-iteration based HDP algorithm for continuous-time nonlinear systems

In this subsection, we propose the value-iteration based HDP algorithm for continuous-time nonlinear systems and give the convergence proof. Note that the key difference between the HDP algorithm and the general policy-iteration algorithm with $k=1$ (which is in fact a variant of the value-iteration algorithm) in [10] is that the initial control law is not necessary stabilized.

Defining the Hamiltonian of the problem as

$$H(x(t), u(t), \partial V/\partial x) = x(t)^T Q x(t) + u(x(t))^T R u(x(t)) + \left(\frac{\partial V}{\partial x}\right)^T (f(x(t)) + g(x(t))u(x(t))), \tag{3}$$

then we can start with an initial value $V_0(x(t)) \geq 0$, and then solves for u_0 as

$$u_0(x(t)) = \arg\min_{v(x(t))} H(x(t), v(x(t)), \partial V_0/\partial x), \tag{4}$$

then we update the cost function as

$$V_1(x(t)) = \int_t^{t+h} x(\tau)^T Q x(\tau) + u_0(x(\tau))^T R u_0(x(\tau)) \, d\tau + V_0(x(t+h)), \tag{5}$$

where $h > 0$ is the sampling period.

The value-iteration based HDP algorithm iterates between the following two steps:

- *Value update step*: update the value using
$$V_{i+1}(x(t)) = \int_t^{t+h} x(\tau)^T Q x(\tau) + u_i(x(\tau))^T R u_i(x(\tau)) \, d\tau + V_i(x(t+h)). \tag{6}$$
- *Policy improvement step*: determine the improved policy using
$$u_{i+1}(x(t)) = \arg\min_{v(x(t))} H(x(t), v(x(t)), \partial V_{i+1}(x(t))/\partial(x(t))). \tag{7}$$

In the above recurrent iteration, i is the iteration index. The cost function and control law are updated until they converge to the optimal values. The following convergence theorem is inspired by the innovative work of [26,27].

Theorem 1. *Suppose the condition*

$$0 \leq J^*(x(t+h)) \leq \theta \int_t^{t+h} x(\tau)^T Q x(\tau) + u(x(\tau))^T R u(x(\tau)) \, d\tau \tag{8}$$

holds uniformly for some $0 < \theta < \infty$ *and that* $0 \leq \delta J^* \leq V_0 \leq \omega J^*$, $0 \leq \delta \leq 1$, $1 \leq \omega \leq \infty$. *The control law sequence* $\{u_i\}$ *and value function sequence* $\{V_i\}$ *are iteratively updated by (6) and (7). Then the value function* V_i *approaches the optimal value function* $J^*(x(t))$ *according to the inequalities*

$$\left[1 + \frac{\delta - 1}{(1+\theta^{-1})^i}\right] J^*(x(t)) \leq V_i(x(t)) \leq \left[1 + \frac{\omega - 1}{(1+\theta^{-1})^i}\right] J^*(x(t)). \tag{9}$$

Define $V_\infty(x(t)) = \lim_{i\to\infty} V_i(x(t))$, *then* $V_\infty(x(t)) = J^*(x(t))$.

Proof. The proof is given in the Appendix.