Neurocomputing 000 (2018) 1–10

Contents lists available at ScienceDirect

Neurocomputing

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

Control Design of a Marine Vessel System Using Reinforcement Learning[☆]

Zhao Yin[a], Wei He[a,*], Chenguang Yang[b], Changyin Sun[c]

[a] School of Automation and Electrical Engineering and Key Laboratory of Knowledge Automation for Industrial Processes, Ministry of Education, University of Science and Technology Beijing, Beijing 100083, China
[b] Zienkiewicz Centre for Computational Engineering, Swansea University, SA1 8EN, UK
[c] School of Automation, Southeast University, Nanjing 210096, China

ARTICLE INFO

Article history:
Received 25 January 2018
Revised 19 April 2018
Accepted 22 May 2018
Available online xxx

Communicated by Mou Chen

Keywords:
Reinforcement Learning
Critic Neural Networks
Actor neural networks
Lyapunov method
Marine Vessel

ABSTRACT

In this paper, our main goal is to solve optimal control problem by using reinforcement learning (RL) algorithm for marine surface vessel system with known dynamic. And this algorithm is an optimal control algorithm based on policy iteration (PI), and it can obtain the suitable approximations of cost function and the optimized control policy. There are two neural networks (NNs), where critic NN aims to estimate the cost-to-go and actor NN is utilized to design suitable input controller and minimize the tracking error. A novel tuning method is given for critic NN and actor NN. The stability and convergence are proven by Lyapunov's direct method. Finally, the numerical simulations are conducted to demonstrate the feasibility and superiority of presented algorithm.

1. Introduction

Recently, marine vessels have been used in various fields, for example, ocean exploration, marine transportation, etc. [1–7]. With the continuous development of society, the traditional control methods are unable to satisfy the growth in the marine transportation and the needs for modern navigation safety. In order to increase tracking accuracy, there are a lot of studies have been proposed with different control methods of marine surface vessels [2,8–13].

For marine surface vessel system, it is a difficult problem to ensure the stability in the brutal environment. Therefore, there have been many researches presented in the last couple of years. For example, an adaptive robust tracking control law with finite-time for a fully actuated marine vessel with unknown interference is proposed in [8]. In [14], a control law for trajectory tracking is proposed for the marine vessels system with state constraints and dynamics uncertainties. The authors present a control method of tracking the desired trajectory for a fully actuated marine vessel in [11]. And a control problem of a variable length crane system is investigated in [15]. In [10,16], the authors propose the sliding-mode control method for a surface vessels system.

In the mathematical view, the optimal control problem is equal to solve Hamilton–Jacobi–Bellman (HJB) equation. Because of the difficulty of nonlinear nature of the HJB equation, more and more researchers put effort into this field in order to solve this puzzle. More achievements have presented the reasonable methods to cope with the discrete-time HJB equation. In [17,18], many useful points about this problem have been given.

Reinforcement learning is an approach to deal with the aforesaid problem [18–23]. For a typical structure of reinforcement learning, there includes two neural networks, and the actor neural network updates its output value based on the value of Critic Neural Network. These two neural networks must execute coordinately, and the ultimate target is to reach the global optimum of cost function. The authors provide an adaptive neural network control by using RL algorithm for a robot manipulator systems with unknown functions and input dead-zone in [24]. In this paper, we propose a surface marine vessel by using reinforcement learning and prove its availability.

In recent years, PI has been discussed in [25–31]. This method belongs to optimal learning for dealing with optimal control problems. For the linear time-invariant system, it can reduce the prob-

☆ This work was supported in part by the National Natural Science Foundation of China under Grants 61522302, 61761130080, U1713209, Grant NA160436 and International Exchanges Grant IE 170247 from the Royal Society, UK, and the Fundamental Research Funds for the China Central Universities of USTB under Grant FRF-BD-17-002A.

* Corresponding author.
E-mail address: weihe@ieee.org (W. He).

https://doi.org/10.1016/j.neucom.2018.05.061
0925-2312/

lem of Kleinman algorithm to solve the Riccati function problem. It is the same as other reinforcement learning algorithms, PI is applied on critic/actor neural networks which are used to approximate the unknown parameters. In this paper, a method about synchronous policy iteration is investigated and it is inspired by PI [32]. This method is one of the generalized PI proposed in [33].

For the past few years, adaptive neural network has been applied for the nonlinear systems broadly, and it can be learned to approximate solution of any nonlinear equations as long as the hidden layer with enough nodes [34–45]. In [46,47], authors use NN to approximate the unknown system parameters. Two NNs are utilized to approximate the input deadzone and unknown system dynamics in [48]. In [49], a novel Critic Neural Network controller is presented for nonlinear feedback systems, and the control design is based on the predictor model. An adaptive neural network controller is presented to cope with the problem of system uncertainties [12,41,50–58]. An adaptive NN control method based on radial basis function for nonlinear multiagent systems is investigated in [59]. In [60], the authors employ an adaptive NN method for an underactuated wheeled inverted pendulum model. In [61], a trajectory tracking control for marine vessel with full-state constraints and system unknown is designed. In the controller, an adaptive neural networks are used to compensate the dynamics uncertainties. To sum up, the NN is a more and more important technique and can be applied to many fields.

In this paper, there are several main contributions. (i) The critic NN is designed to approach the optimal cost function of the marine vessel system, and we tune the critic NN weights when an adoptable policy is specified. (ii) And an extra NN actor neural is proposed, and in standard policy iteration we adjust both NN synchronous in real time. (iii) RL is applied to control the position of a three degrees of freedom multiple-input–multiple-output (MIMO) marine vessel system, which has a good control effect.

In what follows, Section 2 covers problem formulation that contains system modeling and some necessary lemmas, assumptions and properties. The two neural networks control design and stability analysis are shown in Section 3. Next, the simulation is given to show the feasibility and effectiveness of our controller. At last, Section 5 concludes this paper.

2. Problem formulation

Some notations are proposed as follows, and we will use some symbols: $\mathbb{R}^+$ denotes a positive real number, $\mathbb{R}^n$ is the n-dimensional Euclidean space, $\|\cdot\|$ is the norm of Euclidean vector, $|\varpi|$ is the absolute value of a scalar ϖ, $\|\varpi\|$ is the norm of vector ϖ, that is $\|\varpi\| = \sqrt{\varpi^T\varpi}$, and $\|\cdot\|_2$ represents the matrix 2-norm.

2.1. System modeling

In this paper, the dynamic of a marine surface vessel [1] is described as

$$\dot{\eta} = J(\eta)\upsilon$$
$$M\dot{\upsilon} + C(\upsilon)\upsilon + D(\upsilon)\upsilon + g(\eta) = u \tag{1}$$

where $\eta = [\eta_x, \eta_y, \eta_\psi]^T \in \mathbb{R}^3$ denotes the earth-frame positions and heading, $u \in \mathbb{R}^3$ presents the control input of the systems, $\upsilon = [\upsilon_x, \upsilon_y.\upsilon_\psi]^T \in \mathbb{R}^3$ presents the velocities of vessel in the vessel-frame. $M \in \mathbb{R}^{3\times3}$ is a symmetric positive definite inertia matrix, $C(\upsilon) \in \mathbb{R}^{3\times3}$ denotes centripetal and Coriolis torques, $D(\upsilon) \in \mathbb{R}^{3\times3}$ is the damping matrix, and $g(\eta)$ presents the restoring force, and $J(\eta)$ is the transformation matrix which is defined as

$$J(\eta) = \begin{bmatrix} \cos\eta_\psi & -\sin\eta_\psi & 0 \\ \sin\eta_\psi & \cos\eta_\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

We can let $x_1 = \eta$, $x_2 = \upsilon$, then we are able to get following description of our system:

$$\dot{x}_1 = J(x_1)x_2$$
$$\dot{x}_2 = M^{-1}[u - C(x_2)x_2 - D(x_2)x_2 - g(x_1)] \tag{3}$$

Then the vessel dynamical system is given by

$$\dot{x}(t) = A(x(t)) + B(x(t))u(x(t)); \quad x(0) = x_0 \tag{4}$$

where

$$\dot{x}(t) = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix},$$
$$A(x(t)) = \begin{bmatrix} J(x_1)x_2 \\ M^{-1}[-C(x_2)x_2 - D(x_2)x_2 - g(x_1)] \end{bmatrix},$$
$$B(x(t)) = \begin{bmatrix} \mathbf{0}_{3\times3} \\ M^{-1} \end{bmatrix} \tag{5}$$

with $\mathbf{0}_{3\times3}$ denoting 3×3 zero matrices.

Assumption 1. [62] According to (4), we can assume $B(x)$ is bounded, and matrix $B(x)$ has full column rank for all $x \in \mathbb{R}^n$, and we need to define $B^+ = (B^TB)^{-1}B^T$ is bounded and locally Lipschitz.

Assumption 2. [63] Let $x_d(t)$ be the bounded desired trajectory, and we can assume that there exists a Lipschitz continuous equation $f_d(\cdot) \in \mathbb{R}^n$ with $f_d(0) = 0$ such that

$$\dot{x}_d(t) = f_d(x_d(t)) \tag{6}$$

Then denoting the tracking error as,

$$e = x(t) - x_d(t) \tag{7}$$

From (3), (6) and (7), we can obtain the tracking error dynamics

$$\dot{e}(t) = A(x(t)) + B(x(t))u(x(t)) - f_d(x_d(t)) \tag{8}$$

The input controller u_d corresponding to the desired trajectory x_d is

$$u_d(x_d) = B^+(x_d)\dot{x}_d - A(x_d) \tag{9}$$

Therefore, we need to define a new state $\varpi \in \mathbb{R}^{12}$ as

$$\varpi = [e^T, x_d^T]^T \tag{10}$$

According to (8) and Assumption 1, we can obtain the derivative of (10)

$$\dot{\varpi} = E(\varpi) + F(\varpi)\nu \tag{11}$$

where the functions $E \in \mathbb{R}^{12}, G \in \mathbb{R}^{12\times3}$, and controller $\nu \in \mathbb{R}^3$, we have

$$E(\varpi) = \begin{bmatrix} A(e + x_d) - f_d(x_d) + B(e + x_d)u_d \\ f_d(x_d) \end{bmatrix}, \tag{12}$$

$$F(\varpi) = \begin{bmatrix} B(e + x_d) \\ 0^{6\times3} \end{bmatrix}, \quad \nu = u - u_d \tag{13}$$

Assumption 3. [18] We can assume that, $A(0) = 0$, $A(x) + B(x)u$ is Lipschitz continuous on a set $\Omega \subseteq \mathbb{R}^6$ which contains the origin, and the dynamics system achieves stability on Ω. That is, there exists a continuous control torque $\nu(t) \in \mathbb{U}$ so that the system is asymptotically stable on Ω. On the other hand, we assume that the system parameters M, $C(\upsilon)$, $D(\upsilon)$, and $g(\eta)$ are all known.

Please cite this article as: Z. Yin et al., Control Design of a Marine Vessel System Using Reinforcement Learning, Neurocomputing (2018), https://doi.org/10.1016/j.neucom.2018.05.061