Neurocomputing 000 (2017) 1–11

Contents lists available at ScienceDirect

Neurocomputing

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

Concise deep reinforcement learning obstacle avoidance for underactuated unmanned marine vessels

Yin Cheng, Weidong Zhang*

Department of Automation, Shanghai Jiaotong University, Shanghai 200240, PR China

ARTICLE INFO

Article history:
Received 8 April 2017
Revised 19 June 2017
Accepted 26 June 2017
Available online xxx

Communicated by Hongli Dong

Keywords:
Deep learning
Reinforcement learning
Obstacle avoidance
Nonlinear dynamics
Underactuated unmanned marine vessel

ABSTRACT

This research is concerned with the problem of obstacle avoidance for the underactuated unmanned marine vessel under unknown environmental disturbance. A concise deep reinforcement learning obstacle avoidance (CDRLOA) algorithm is proposed with the powerful deep Q-networks architecture to overcome the usability issue caused by the complicated control law in the traditional analytic approach. Furthermore, a comprehensive reward function is specifically designed for obstacle avoidance, target approaching, speed modification, and attitude correction. Compared to the analytic methods, the proposed algorithm based on reinforcement learning shows notable advantages in utility and extendibility. With the same CDRLOA system, the targets and the constraints are highly customizable for various of special requirements. Extensive experiments conducted have demonstrated the effectiveness and conciseness of the proposed algorithm.

1. Introduction

The application of the autonomous maritime system is becoming more and more prevalent due to its flexibility and versatility both in the civil and military field. For all kinds of application scenarios, it is of extreme importance to avoid obstacles such as rocks, floaters, debris and other ships. For the autonomous maritime vessels, the obstacle detection, information fusion, avoidance algorithm and the control strategy must be located onboard vessels. Consequently, the major challenge is the realization of real-time obstacle avoidance control strategy. Therefore, the applicable decision-making operator has an essential role in autonomous navigation and obstacle avoidance. Many positive results on this topic have been reported in the literature and readers are referred to the papers [1,2]. Lisowski and Smierzchalski [3] first applied the mathematical algorithm on the ship's dynamic mathematical model (static, kinetic, dynamic and matrix models) by generating a sequence of maneuvers.

However, mathematical algorithms have their particular limitations, and the avoidance performance intrinsically depends on the fine-grained models of obstacles and the dynamics of vessels. The slight changes of obstacles and the disturbance of the environment may lead to model's failure. Moreover, as the maritime system complexity increase, the mathematical algorithms become harder to design and deploy. In all the study as mentioned earlier of obstacles avoidance, there exist three main issues to be resolved: (1) The algorithms' ability to deal with the complex dynamic system is limited. Traditional mathematical algorithms are feeble to the changes and uncertainty of systems, while the weak representation capacity is the major flaw of traditional reinforcement learning approaches. (2) The control law obtained by most mathematical algorithms is formed as complicated formulas, while they are often too complicated to be deployed in practical applications. (3) The previous architectures are designed to accommodate some specified situations. Consequently, these approaches are not interoperable and portable for diverse and complex navigation requirements.

On the other side, the recent development of artificial intelligence area [4,5] has profound effects on the industrial world, which brings researchers powerful algorithms to characterize and control the extremely complex system under the changing environment. The ancient game of Go has long been viewed as the most difficult and challenging classic game, while the strategy of Go players can also be considered as the output of a controlled system with high complexity. David Silver and Aja Huang [6] managed to design a group of deep neural networks known as AlphaGo that are trained by the deep reinforcement learning (DRL) from games of self-play to beat human Go champions. Comparing with prior knowledge based traditional algorithms, DRL is with greater capacity to adapt complex system environment while it is capable of self-learning. Positive results in [7] have demonstrated that the successful control policies can be learned directly from DRL on

* Corresponding author.
E-mail addresses: wdzhang@sjtu.edu.cn, ycheng_mit@sjtu.edu.cn (W. Zhang).

http://dx.doi.org/10.1016/j.neucom.2017.06.066
0925-2312/© 2017 Published by Elsevier B.V.

challenging classic games under various scenarios. Besides, traditional reinforcement learning approaches have been applied in the autonomous movement of the four-wheeled robot [8]. After lots of self-learning processes, the robot car had succeeded in navigating in the environment with multiple obstacles. Fathinezhad and Derhami [9] proposed supervised fuzzy sarsa method for robot navigation by utilizing the advantages of both supervised and reinforcement learning algorithms.

Motivated by all of these theories and realistic reasons, this paper focuses on the field of advanced artificial intelligence approaches, e.g. deep learning architectures, reinforcement learning algorithms. A concise autonomous obstacles avoidance system, which is implemented by avoidance reward algorithm and deep reinforcement learning approach, is proposed for the complex unmanned marine vessels with unknown dynamics, taking autonomous surface vessels (ASVs) as the cases. The main contributions of this work can be summarized as follows:

(1) A concise deep reinforcement learning obstacles avoidance (CDRLOA) system is developed to deal with the complex navigation situations and the unknown dynamics of the environment. Combining the proposed avoidance reward algorithm and deep reinforcement learning (DRL) approach, CDRLOA system improves its controller with the self-play process. By applying the concise control strategies learned from the system, vessels are able to reach the target through highly simplified control instructions while avoiding the collision.

(2) Corroborating the effectiveness of the proposed algorithm with the obstacles avoidance navigation tasks, which consist of obstacles scattered within the domain, the destination to be reached, and the standard ASV with unknown environmental dynamics. The experiments have achieved satisfying results and shown the conciseness of the control strategies learned from CDRLOA system.

2. System description and preliminaries

2.1. Preliminaries

Throughout the paper, $|\cdot|$ denotes the absolute value for a scalar variable or the members of a specific set. $\mathbb{E}$ denotes the expectation operator, and $\mathbb{V}$ denotes the variation operator. $\hat{(\cdot)}$ is the estimation of $(\cdot)$ and $\delta(\cdot)$ is the Dirac function [10]. The case study of obstacles avoidance takes ASVs as the example. The horizontal motion of a surface vessel is unusually described by the motion components in surge, sway, and yaw [11]. Based on this, $\boldsymbol{\upsilon} = [u, v, r]^T \in \mathbb{R}^3$ and $\boldsymbol{\eta} = [x, y, \psi]^T \in \mathbb{R}^3$ are chosen as the velocity vector and position vector. Among these, (ψ) is the heading of the vessel and $(x,\ y)$ is the position in the earth-fixed inertial frame. The linear velocities $\boldsymbol{\upsilon} = [u, v, r]^T$ correspond to surge and sway, and for yaw in the body-fixed frame of vessel. r_0 is the obstacle detection radius. Fig. 1 illustrates the major concepts of the movement process in this case. The nonlinear dynamic equations of motion [12] can be expressed as:

$$\begin{aligned} &\dot{\boldsymbol{\eta}} = \boldsymbol{R}(\psi)\boldsymbol{\upsilon} \\ &\boldsymbol{M}\dot{\boldsymbol{\upsilon}} = \boldsymbol{\tau} - \boldsymbol{C}(\boldsymbol{\upsilon})\boldsymbol{\upsilon} - \boldsymbol{D}(\boldsymbol{\upsilon})\boldsymbol{\upsilon} - \boldsymbol{g}(\boldsymbol{\upsilon}) + \boldsymbol{\tau_w} \end{aligned} \tag{1}$$

where $\boldsymbol{R}(\cdot)$ is the 3 DOF rotation matrix for the horizontal motion of ASV. This matrix has the properties that $\boldsymbol{R}(\psi)^T\boldsymbol{R}(\psi) = \boldsymbol{I}$ and $\|\boldsymbol{R}(\psi)\| = 1$ for all ψ. Generally, $\frac{d}{dt}\{\boldsymbol{R}(\psi)\} = \dot{\psi}\boldsymbol{R}(\psi)\boldsymbol{S}$, where

$$\boldsymbol{J}(\eta) \overset{\text{3DOF}}{=} \boldsymbol{R}(\psi) = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}, \boldsymbol{S} \overset{\text{3DOF}}{=} \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{2}$$

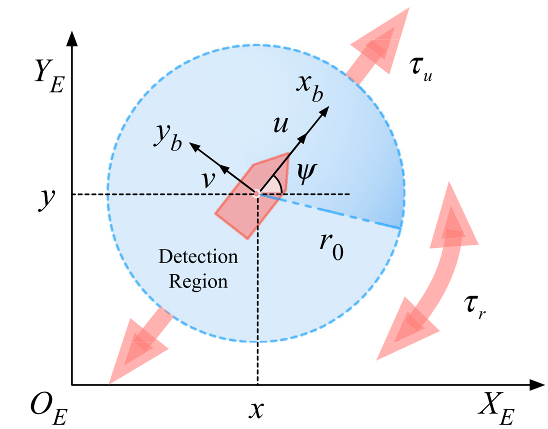

Fig. 1. Major components of the autonomous underactuated vessel.

The system inertia matrix $M = M^T = M_A + M_{RB} > 0$ is the combination of added mass matrix M_A and rigid-body matrix M_{RB}:

$$M_A = \begin{bmatrix} -X_{\dot{u}} & 0 & 0 \\ 0 & -Y_{\dot{v}} & -Y_{\dot{r}} \\ 0 & -Y_{\dot{r}} & -N_{\dot{r}} \end{bmatrix}, M_{RB} = \begin{bmatrix} m & 0 & 0 \\ 0 & m & mx_g \\ 0 & mx_g & I_z \end{bmatrix} \tag{3}$$

Similarly, the skew-symmetric matrix $C(\upsilon) = -C(\upsilon)^T$ of Coriolis and centripetal terms are also consisted of two parts:

$$C(\upsilon) = \begin{bmatrix} 0 & 0 & c_{13}(\upsilon) \\ 0 & 0 & c_{23}(\upsilon) \\ -c_{13}(\upsilon) & -c_{23}(\upsilon) & 0 \end{bmatrix} = C_A(\upsilon) + C_{RB}(\upsilon) \tag{4}$$

$$C_A(\upsilon) = \begin{bmatrix} 0 & 0 & Y_{\dot{v}}v + Y_{\dot{r}}r \\ 0 & 0 & -X_{\dot{u}}u \\ -Y_{\dot{v}}v & X_{\dot{u}}u & 0 \end{bmatrix} \tag{5}$$

$$C_{RB}(\upsilon) = \begin{bmatrix} 0 & 0 & -m(x_g r + v) \\ 0 & 0 & mu \\ m(x_g r + v) & -mu & 0 \end{bmatrix} \tag{6}$$

$D(\upsilon)$ is the nonlinear damping matrix for the system inertia

$$D(\upsilon) = \begin{bmatrix} d_{11}(\upsilon) & 0 & 0 \\ 0 & d_{22}(\upsilon) & d_{23}(\upsilon) \\ 0 & d_{32}(\upsilon) & d_{33}(\upsilon) \end{bmatrix} \tag{7}$$

with

$$\begin{aligned} d_{11}(\upsilon) &= -X_u - X_{|u|u}|u| - X_{uuu}u^2 \\ d_{22}(\upsilon) &= -Y_v - Y_{|v|v}|v| - Y_{|r|v}|r| \\ d_{23}(\upsilon) &= -Y_r - Y_{|v|r}|v| - Y_{|r|r}|r| \\ d_{32}(\upsilon) &= -N_v - N_{|v|v}|v| - N_{|r|v}|r| \\ d_{33}(\upsilon) &= -N_r - N_{|v|r}|v| - N_{|r|r}|r| \end{aligned} \tag{8}$$

The coefficients $[\{X_{(\cdot)}, Y_{(\cdot)}, N_{(\cdot)}\}$ are the so-called hydrodynamic derivatives that represent the hydrodynamic forces and moments acting on the vessel. $g(\upsilon) = [g_u, g_v, g_r]^T \in \mathbb{R}^3$ indicates the unmodeled dynamics.

The control input vector τ denotes the propulsion surge force and the yaw moment, which is given by

$$\tau = \begin{bmatrix} \tau_u & 0 & \tau_r \end{bmatrix}^T \in \mathbb{R}^3 \tag{9}$$

To further complicate the requirements, an underactuated vessel is under consideration. As can be observed in the propulsion force and moment vector τ, the independent actuators for the sway control are unnecessary. Accordingly, this vessel has more extensive applicability and fewer requirements for hardware. The disturbance from the environment can be represented by the vector

Please cite this article as: Y. Cheng, W. Zhang, Concise deep reinforcement learning obstacle avoidance for underactuated unmanned marine vessels, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.06.066