Deep reinforcement learning-based controller for obstacle avoidance of deep sea mining vehicles

Qihang Chen\textsuperscript{1,2}, Jianmin Yang\textsuperscript{1, 2, *}, Peng Wang\textsuperscript{1, 2}, Zhixuan Liang\textsuperscript{1}, Changyu Lu\textsuperscript{1, 2}

\textsuperscript{1} State Key Laboratory of Ocean Engineering, Shanghai Jiao Tong University, Shanghai, China
\textsuperscript{2} Yazhou Bay Institute of Deepsea SCI-TECH, Shanghai Jiao Tong University, Sanya, Hainan, China

\textsuperscript{*} Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China

ABSTRACT

Research on deep-sea mining vehicles (DSMVs) is being actively conducted for deep-sea resource development. Developing the obstacle avoidance performance of DSMV will effectively improve the mobility and safety of the DSMV operation process. In this paper, the Modified Twin Delayed Deep Deterministic Policy Gradient (MTD3) algorithm is adapted to train a controller for obstacle avoidance. A Markov decision process model including state, action, and reward functions is designed, and a suitable reinforcement-learning training environment is designed. The obstacle-avoidance ability of the controller is verified by the simulation tests of avoiding the randomly generated one to five obstacles in each episode.

KEY WORDS: Deep-sea mining; deep reinforcement learning; twin delayed deep deterministic policy gradient; obstacle avoidance; Markov decision process model.

INTRODUCTION

The ocean has the largest untapped mineral resources on the earth. In general, there are three main types of deposits of deep-sea mineral resources on the sea floor (Leng et al., 2021): polymetallic nodules, massive Sulfides(SMS), and cobalt-rich crusts. Once these rich mineral resources can be exploited effectively and economically, it will alleviate the shortage of mineral resources to a great extent. Deep-sea mining can be defined as the utilization of hydrodynamic or mechanical methods to transport mineral ores from the seabed to the ocean surface and then transport ores to land-based processing plants by ships (Ma et al., 2019; Ma et al., 2022). As the key equipment of deep-sea ore mining in the future, deep-sea mining vehicle (DSMV) is the carrier of mining equipment. DSMV carries mining devices to collect deep-sea ore as much as possible. However, the complexity and unpredictability of the seabed topography pose great challenges to the movements of DSMV. Therefore, it is necessary to develop an optimal controller for collision avoidance of DSMV.

Traditional approaches to obstacle avoidance of DSMV rely on model-based planning (Dai and Liu, 2013; Liang et al., 2018). For example, Li and Zou (2012) proposed a fuzzy PID approach to control the speed of the left and right tracks to simulate the unilateral obstacle-crossing condition when the vehicle drives in a straight line. Numerical simulation results preliminarily show the feasibility of the proposed methods. In addition to using fuzzy logic, Wu et al. (2021) proposed a model predictive control (MPC) method to handle dead zones and obstacles during trajectory tracking. An obstacle avoidance strategy is used that utilizes the tri-circular arc obstacle-avoidance trajectory with an equal curvature for path re-planning. However, these works are subject to model mismatches, nonlinearity, and external disturbances in dynamic environments.

Recently, the development of deep reinforcement learning (DRL) provides a way to learn optimal control strategies in a changing environment. In reinforcement learning, the controller can be considered an agent, which can receive the state and take action at each time step. After choosing the action, the agent can receive the feedback signal as the reward. Thus, the goal of the agent is to learn optimal policy to maximize the accumulative reward during the training. Compared with the traditional algorithm based on prior knowledge, reinforcement learning not only has self-learning ability but also has a stronger ability to adapt to complex environments. Currently, there is an increasing amount of research that appeals to deep reinforcement learning to solve robotic obstacle-avoidance problems. For example, Muller et al. (2005) proposed a vision-based obstacle avoidance system for a mobile robot by training a six-layer convolutional network that maps raw input images to steering angles. Similarly, Levine et al. (2016) proposed an end-to-end framework that learns control policies mapping raw observations to torques at the robot motors. Liang et al. (2022a); Liang et al. (2022b) presented a hierarchical deep reinforcement learning approach for multi-vehicle cooperation and evaluate it in a real-world testbed. Wang et al. (2023) proposed the DRL-based SAROA approach which shows excellent practicability and robustness for the environmentally-driven USVs in large-scale and uncertain environments.

In this work, the obstacle-avoidance problem of DSMV belongs to the continuous deterministic action problem in continuous space. To solve the continuous-space control problem, Fujimoto et al. (2018) proposed the Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm based on the original Deep Deterministic Policy Gradient algorithm (DDPG). We modified TD3 (MTD3) and optimized the DSMV model so that the DSMV obstacle-avoidance problem can be solved. The MTD3-based controller is trained and validated by one-to-