Advances in Intelligent Systems Research, volume 147

International Conference on Network, Communication, Computer Engineering (NCCE 2018)

The Realization of Mobile Robot's Dynamic Obstacle Avoidance with Deep Reinforcement Learning Based on Deterministic Strategy Gradient

Yunxiong Wu [a)]

School of computers, Guangdong University of Technology, Guangzhou 510003, China.

[a)] Corresponding author: 2441596205@qq.com

Abstract. When the deep reinforcement learning algorithm based on visual perception is applied to the issue of robot's dynamic obstacle avoidance, the perception ability of convolutional neural network is combined with the decision control ability of reinforcement learning, and the process from raw visual input to decision action output is realized. But the application scope of deep reinforcement learning algorithm based on Q value is still in low-dimensional and discrete action space. If the continuous action space is discretized, the problem of excessively huge motion space and extremely difficult convergence of network model will be caused. Besides, fine adjustment cannot be realized for the network model, and meanwhile, the division of motion space will also result in information loss. Hence, a deep reinforcement learning algorithm based on deterministic strategy gradient was proposed in this paper, and the strategy was parameterized via convolutional neural network through the integration of reinforcement learning algorithms based on strategy and value.

Key words: Deep Reinforcement Learning, Mobile Robot, Intelligent Decision-making, Visual Control, Local path planning.

INTRODUCTION

The present navigation decision system requires us to design various subtle strategies to deal with the complicated environment. A tiny negligence in strategy design can produce disastrous consequences. In addition, the traditional obstacle avoidance decision system of mobile robot sets mobile robot as the sole intelligent agent in the environment and treats other moving objects in the environment as "obstacles". Actually, in the navigation decision process of mobile robot, the robot has an interactive relation with other moving objects in the environment, and it is an issue of multi-agent decision control [1-2].

However, as for the navigation system based on deep reinforcement learning algorithm [3], it is hard to explain the decision-making process, so it is controversial in the dynamic obstacle avoidance issue of navigation [5]. The model sets visual perception as input and decision control action as output. Only rough classification is conducted on the basis of end-to-end neural network [4], and the middle reasoning process cannot be explained. Besides, it is difficult to conduct targeted improvement when failure happens to the system. Therefore, a deep reinforcement learning algorithm based on deterministic strategy gradient [6] was proposed in this paper. This algorithm is separated from rock-bottom perception and drive control, its division of labor is clear, and the interpretability is high. This architecture can avoid the awkwardness that traditional decision systems have to adopt conservative strategies due to the difficulty of strategy design. Meanwhile, the system instantaneity and robustness for dynamic environment are guaranteed.

DEEP REINFORCEMENT LEARNING ALGORITHM BASED ON DETERMINISTIC STRATEGY GRADIENT

Derivation of Deep Reinforcement Learning Algorithm Based on Deterministic Strategy Gradient

In network model training of deep reinforcement learning algorithm based on deterministic strategy gradient, the overall network model can be divided into two sub-models of key network and target network. Actor-Critic framework is adopted for each sub-network, including both strategy network and evaluation network. The training of strategy network model adopts Silver's algorithm, and the strategy gradient is defined as:

$$\frac{\partial J(\theta^{\mu})}{\partial \theta^{\mu}} = E_s[\frac{\partial Q(s,a \mid \theta^{Q})}{\partial \theta^{\mu}}] \tag{1}$$

In the above formula, $J(\theta^{\mu})$ is the objective function of model optimization, and the following can be gained according to the strategy network $a = \pi(s \mid \theta^{\mu})$:

$$\frac{\partial J(\theta^{\mu})}{\partial \theta^{\mu}} = E_s[\frac{\partial Q(s,a \mid \theta^{Q})}{\partial a}\frac{\partial \pi(s \mid \theta^{\mu})}{\partial \theta^{\mu}}] \tag{2}$$

The deterministic strategy network model updates model parameters along the trajectory direction of high cumulative return.

The evaluation network updates network model parameters through deep Q value model, and the corresponding gradient information is defined as:

$$\frac{\partial L(\theta^{Q})}{\partial \theta^{Q}} = E_{s,a,r,s' \sim D}[(T\arg etQ - Q(s,a \mid \theta^{Q}))\frac{\partial Q(s,a \mid \theta^{Q})}{\partial \theta^{Q}}] \tag{3}$$

$$T\arg etQ = r + \gamma Q'(s', \pi(s' \mid \theta^{\mu'}) \mid \theta^{Q'}) \tag{4}$$

In the above formula, $\theta^{\mu'}$ and $\theta^{Q'}$ represent parameters of target strategy network and target value network respectively. The network model draws training samples at random from the experience replay pool via experience replay method, and updates network parameters along the direction of big cumulative sum of trajectory improvement.

Data Pre-Processing and Convolutional Neural Network Model Structure

The deep convolutional neural network based on deterministic strategy gradient adopts Actor-critic framework. Three layers of CNN and two layers of FNN are applied in network architecture design. The network inputs state S only and the output is various continuous actions a. The evaluation network inputs state S and action a, and outputs evaluation values of evaluation network actions. The network model structure is shown in Fig. 1. At the same time, the network model will conduct gray processing for input data, so as to reduce the handling capacity of network training data.

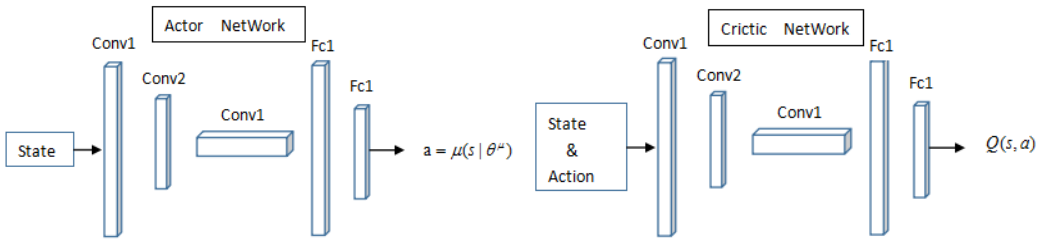

FIG. 1. Convolutional network structure

EXPERIMENTAL RESULTS AND ANALYSIS

Training Visualization of Convolutional Neural Network Model

We conducted model evaluation through calculating loss function and average Q value of the model at regular intervals in the network model training process. Fig. 1 shows the variation trend of loss function, and Fig. 2 presents the change of Q value in the robot training process. According to the figure, after neural network training reaches the 50th stage, the network model tends to converge and become relatively stable.

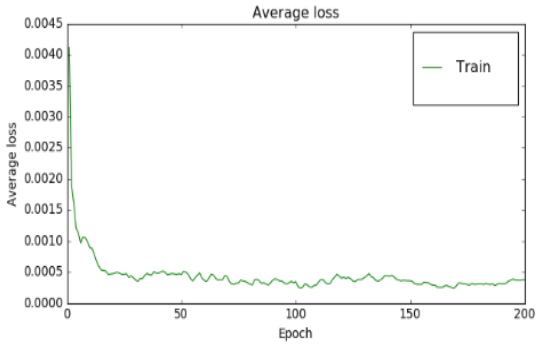

FIG.2. Network training loss function

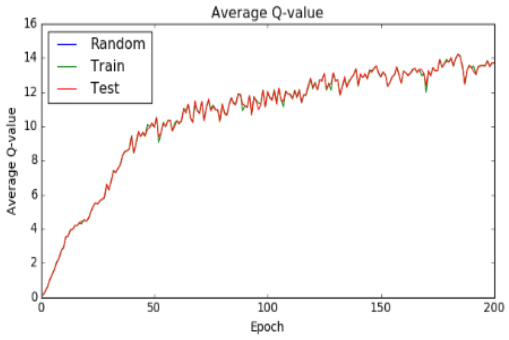

FIG.3. The average state action value

Experimental Results

Fig. 4 shows the experimental results about dynamic obstacle avoidance of mobile robot. The green mobile robot ball moves from the initial position at the left bottom to the target point at the top right corner along the trajectory in the narrow passage, and the blue obstacle ball moves from the top right corner to the left bottom along the trajectory. When the mobile robot encounters the obstacle, the robot can move toward the target point by choosing an optimum action according to the requirement of minimum deviation from trajectory and obstacle avoidance without colliding with the narrow passage.

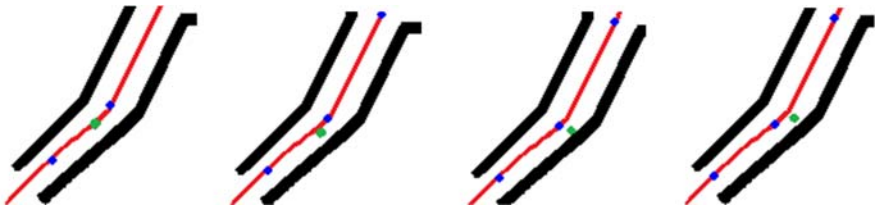

FIG.4. Fragment screenshots of mobile robot trajectory tracking and dynamic obstacle avoidance

SUMMARY

The experiment shows that the deep reinforcement learning algorithm [7-8] based on deterministic strategy gradient possesses good instantaneity and strong robustness in dynamic obstacle avoidance and can effectively realize dynamic obstacle avoidance. Besides, the efficiency of algorithm optimization policy is high, the model training speed is fast, and it has much less time steps than the deep reinforcement learning based on value function. It performs stably in high-dimensional input and task control of continuous action space, so this algorithm can be applied to the scene of continuous action space.

REFERENCES

1. Mnih V, Kavukcuoglu K, Silver D, et al. Playing Atari with deep reinforcement learning//Proceedings of Workshops at the 26th Neural Information Processing Systems 2013. Lake Tahoe, USA, 2013:201-220.
2. Silver D, Lever G, Heess N, et al. Deterministic policy gradient algorithms//Proceedings of the International Conference on Machine Learning. Beijing, China, 2014: 387-395.
3. Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning. Nature, 2015, 518(7540):529-533.
4. Yu Kai, Jia Lei, Chen Yu-Qiang, Xu Wei. Deep learning: yesterday, today, and tomorrow. Journal of Computer Research and Development, 2013, 50(9): 1799-1804 (in Chinese).
5. Kober J, Peters J. Reinforcement learning in robotics: a survey. International Journal of Robotics Research, 2013, 32(11): 1238-1274.
6. Sutton R S, Mcallester D A, Singh S P, et al. Policy gradient methods for reinforcement learning with function approximation//Proceedings of the Advances in Neural Information Processing Systems. Denver, USA, 1999: 1057-1063.
7. Watkins C J C H. Learning from delayed rewards. Robotics & Autonomous Systems, 1989, 15(4): 233-235.
8. Lillicrap T P, Hunt J J, Pritzel A, et al. Continuous control with deep reinforcement learning. Computer Science, 2016, 8(6): A187.