Research on manipulator control based on deep reinforcement learning

Haoyu Chi

1 Electrical and control engineering, Xi'an University of science and technology, Xi'an, Shaanxi, China
email:19206204087@stu.xust.edu.cn

Abstract. With the gradual improvement of the influence of intelligent robots in production and life, it has greatly facilitated people's production and life. Therefore, people's requirements for intelligent robots are also increasing, and are developing towards more humanization and intelligence. However, at present, there are still many imperfections in the field of intelligent robot technology in China. In order to solve the problems in work, we must further strengthen the research on artificial intelligence theory and robot technology. Only in this way can we realize the all-round development of intelligent robot system. So this paper will discuss the deep reinforcement learning in the theory of artificial intelligence, and explain its basic theory, research status, existing problems and future development direction. Moreover, under the background of the overall improvement of the current industrial development level, this paper will also talk about the manipulator widely used in the industrial field and the research status of manipulator control based on deep reinforcement learning, hoping to provide effective help for the development of related fields.

1. Introduction

With the progress of science and technology and the development of society, the development of artificial intelligence and robot has become more and more mature. Artificial intelligence refers to the discipline that simulates people's intelligent behavior and thinking process by exploring some applications of computers. It mainly includes manufacturing intelligent computers equivalent to human brain and realizing the intelligent principles of computers, so that computers can be used in various fields. The application of artificial intelligence in robot includes recognition, direction recognition, perception, proof and other functions. Intelligent robot belongs to the experimental product of artificial intelligence, which can well reflect its technical development requirements. If the two are effectively combined, it will inevitably have a positive impact on the construction and development of industrial machinery.

Machine learning is an interdisciplinary subject in many fields. According to the common classification, it can be divided into supervised learning, unsupervised learning, deep learning and reinforcement learning[1]. Reinforcement learning theory is proposed based on the mechanism of "learning and feedback", that is, learning from failure. It focuses on learning in continuous interaction with the environment and obtaining positive or negative feedback, so as to optimize the decision-making required for action. Because the process of reinforcement learning does not need prior knowledge and supervision signals, the application of reinforcement learning in some complex and unknown environments has broad prospects. However, reinforcement learning focuses on decision-making and lacks the perception of the environment. Therefore, deep learning, which is widely and
deeply used in the field of image recognition, can be well combined with reinforcement learning to form deep reinforcement learning.

Deep reinforcement learning is one of the most concerned directions in the field of artificial intelligence in recent years. It uses the image recognition technology of deep learning to optimize the state elements in reinforcement learning, and then combined with the decision-making ability of reinforcement learning to directly control the action of agent. This idea can solve some decision-making ability in complex environment. For example, in 2013, Google proposed the deep Q-network algorithm (dqn), which combines the deep convolution neural network with the Q-learning algorithm, uses the deep convolution neural network to extract image feature information, and uses the Q-learning algorithm to make decisions. In 2015, the industrial robot FANUC Corporation held the world robot exhibition in Tokyo. The newly developed manipulator is displayed. Through the deep learning algorithm, the manipulator learns to select objects in random positions by trial and error. After 8 hours of training, the manipulator can pick out parts from a pile of sundries with an accuracy of 90%.

In the field of robotics, especially in the development of industrial robots, the manipulator is an important research object, because it replaces human beings to engage in production in some harsh fields, such as flow operation. Working on the ground with the manipulator can save human physical strength. At the same time, due to the extensive use of manipulator in life, human productivity has been greatly improved, our working environment has also been improved, and human life has become more and more intelligent. Therefore, this paper mainly discusses the development and application of deep reinforcement learning in manipulator trajectory planning.

2. Research status and theory of deep reinforcement learning

Deep reinforcement learning is a new algorithm combining deep learning and reinforcement learning, which realizes end-to-end learning from perception to action. Input images, text, audio, video, etc. through the processing of the deep neural network constructed by deep reinforcement learning, the direct output action can be realized without manual intervention.

2.1 Deep reinforcement learning theory

Deep reinforcement learning algorithm has a wide range of applications, which can be learned from its learning process. Firstly, at any time, the agent will interact with the environment and obtain the state quantity, and then optimize the state quantity through deep learning to obtain the corresponding accurate state quantity. Then evaluate the value of each action and take corresponding decisions to execute the action. Finally, after interacting with the environment through action, the next state and feedback will be obtained. After repeating the above process, we can achieve the ultimate goal. Figure 1 is a schematic diagram of deep reinforcement learning theory.

![Figure 1. Schematic diagram of deep reinforcement learning theory](image)

2.2 Research status of deep reinforcement learning

Deep reinforcement learning was first applied to games, and deep mind, Google's artificial intelligence team, is a pioneer in this field. Deepmind has published many research results in the field of deep
reinforcement learning, breaking through the limitations of traditional algorithms on state input, and greatly promoting the application of deep reinforcement learning in various fields, such as robot control, game decision-making and so on. In 2015, based on the original dqn algorithm, Google team deepmind improved and added the target network, improved the stability of the algorithm, and its feature extraction does not rely on manual work, so as to give better play to its self-learning ability.

With the continuous development of deep reinforcement learning in recent years, many algorithms have been proposed, such as ddpg, A3C, sac and so on. Figure 2 shows the classification of existing deep reinforcement learning algorithms, so that readers can better understand them.

Figure 2. Classification of deep reinforcement learning

3 Overview of manipulator kinematics and trajectory planning
With the advent of intelligent and unmanned era, the wide application of industrial robot shows that it has played a great role in many fields. Whether for enterprises or individuals, industrial robot has driven labor productivity, economic benefits and improved workers' working conditions by improving the level of production automation. In recent years, the global intelligent manufacturing industry has developed rapidly, which further promotes the development of industrial robots. At present, industrial robots have been widely and efficiently used in welding, spraying, handling, assembly, machining and other fields.

3.1 Overview of Manipulator Kinematics
The kinematics of manipulator studies the motion of manipulator without considering the force producing motion. Kinematics studies the position, velocity and acceleration of the manipulator. The research of manipulator kinematics involves geometry and time-based content, especially the relationship between each joint and its variation with time. The movement credits of the manipulator are forward kinematics and inverse kinematics: forward kinematics is to find the corresponding end position and attitude of the manipulator when the joint angle of the manipulator is known; Inverse kinematics is to know the end position and attitude of the manipulator and solve the joint angle of the corresponding manipulator. As shown in Figure 3, it is the configuration diagram between adjacent joints in the manipulator. Therefore, analyzing the kinematics of the manipulator is of great significance for the control of the manipulator.
3.2 Manipulator trajectory planning

Trajectory planning can be divided into two categories: Cartesian space trajectory planning and joint space trajectory planning. Cartesian space trajectory planning is to use mathematical expressions to describe the expected trajectory of the robot end in Cartesian space, and express the position, attitude, velocity and acceleration of the robot end as a function of time\(^{[10]}\). Joint space trajectory planning is to convert the end motion trajectory path points in Cartesian space coordinate system into corresponding joint angles by solving the inverse kinematics solution, construct the function of joint variables in time, and solve its first-order and second-order derivatives.

4 Research status of deep reinforcement learning in manipulator trajectory planning

At present, the use of deep reinforcement learning algorithm to train manipulator to control grasping motion is mainly in the experimental research stage. In recent years, a variety of related technologies and algorithms have been proposed, and new research results continue to emerge.

Du Zhijiang et al\(^{[10]}\). A variable admittance human-computer interaction model based on fuzzy reinforcement learning is proposed. It learns the human operation characteristics in the continuous human-computer interaction process, and achieves the control purpose of the operator by adjusting the admittance control model. The variable admittance control model adds the fuzzy sarsa algorithm to realize the damping change conditions in each stage of the interaction process, so as to make the control of the manipulator more flexible.

Zhang Fangyi et al\(^{[11]}\). It is proved that the manipulator can learn from the image independently without any prior knowledge. The depth Q network algorithm (dqn) used in it reaches the target position after simulation training. The input image elements include the ground position and target position of the manipulator. However, in the simulation, it is the image obtained from the third perspective. In practice, if the image is obtained by the camera, it can not effectively control the manipulator to reach the target position. Fig. 4 is an algorithm flowchart of dqn.
5 Existing problems

Manipulator trajectory planning based on deep reinforcement learning is a promising research direction, but there are also many problems at this stage, especially for the research of deep reinforcement learning algorithms, there are still many immature places, which often makes the control of manipulator unable to achieve the expected effect quickly and accurately. Let's discuss such issues below.

5.1 Algorithm sample efficiency problem

Reinforcement learning is a main branch of machine learning, which mainly studies how to make agents infer optimal control decisions from their interaction with the environment. At present, reinforcement learning algorithms often need a large amount of interactive data to achieve good learning effect, which limits the application of existing technologies in practical problems with expensive interactive data. In order to reduce the high dependence of reinforcement learning on the amount of data, we need to have a deeper understanding of the sample efficiency of related algorithms. Although the existing theoretical analysis can describe the relationship between algorithm, problem instance and sample efficiency to a certain extent, its analysis results are too targeted at the most difficult problem instance, and cannot give a sufficiently accurate prediction of sample efficiency on the problem of general difficulty. This makes the existing theoretical results difficult to help users and researchers compare, select, set and improve algorithms.

5.2 Learning method under abnormal reward function

The feedback of many tasks is sparse. For example, the task of walking the maze in the game can get a positive feedback only when walking out of the maze. The other actions will not get any positive feedback, so they will be rewarded only when they successfully complete the task. In this way, if the agent is allowed to explore at random, it will be difficult to get any positive feedback and effective strategy evaluation, resulting in the inability to learn useful experience. Another case is that the reward function is difficult to define accurately. In these cases, it is difficult to determine its form even using manual methods.

In order to solve this kind of sparse reward problem, the reward function can be defined in the grasping process of the manipulator. For example, the distance between the manipulator and the target will produce the corresponding reward function, or the opening of the manipulator will produce the reward function.

6 Conclusion

For the future development direction, there are many problems to be solved. In typical tasks with sparse reward or difficult to define, the internal reward will be affected by the internal randomness of the environment. For example, in alpha go, you can only judge the result from the beginning to the end of the game. At this time, the agent can be rewarded, which is difficult to evaluate in the middle of the game; In the navigation task, the agent can be rewarded only when it reaches the specified position.
within the specified time step, and there is no reward in each step of the intermediate process; In the grasping task of manipulator, the manipulator can be rewarded only after completing a series of complex attitude control and successfully grasping the target. The failure of any step in the middle will lead to the failure of winning the reward. Therefore, if the sparse reward problem can be solved, the number of interactions can be reduced, the learning speed can be accelerated, and the sample utilization can be improved.

To solve the above problems, researchers have also proposed some solutions. For example, in terms of optimizing the reward function, Finn et al. [3] proposed the cost function in inverse reinforcement learning under the reward function represented by deep neural network. Hadfield et al. [4] proposed a method to approximately solve the reward function, which can avoid the negative impact of the reward function. Christiano et al. [5] proposed to learn the reward function according to human preferences, obtain human preferences by selecting trajectories, and use supervised learning to approximate human preferences. Zhao Kaifeng et al. [6] summarized the early inverse reinforcement learning methods.

The above methods have achieved preliminary results in solving the sparse reward problem, but there are still deficiencies. In short, in the development and application of manipulator, researchers need to constantly solve these problems to improve the intelligence and stability of manipulator, so as to better serve human beings and do what human beings can't do. Therefore, the application of deep reinforcement learning in the field of manipulator is a promising development direction. Through the discussion of manipulator control based on deep reinforcement learning, this paper provides readers with the development status and existing problems in this field, makes readers have a preliminary understanding of manipulator control, and hopes to let more people join the field of deep reinforcement learning and manipulator.

References
[1] Zhao Zilong. On the development trend of artificial intelligence and robot [J]. China new communication, 2018, 20 (24): 156
[2] Zheng Wei. Application of artificial intelligence in the field of intelligent robots [J]. Electronic testing, 2021 (10): 127-128
[3] FINN C, LEVINE S, ABBEEL P. Guided cost learning: Deep inverse optimal control via policy optimization[C]//International Conference on Machine Learning. 2016: 49-58.
[4] HADFIELD-MENELL D, MILLI S, ABBEEL P, et al. Inverse reward design[C]//Advances in neural information processing systems. 2017: 6765-6774.
[5] CHRISTIANO P F, LEIKE J, BROWN T, et al. Deep reinforcement learning from human preferences[C]//Advances in Neural Information Processing Systems. 2017: 4299-4307.
[6] ZHANG KAIFENG, YU YANG.Methodologies for Imitation Learning via Inverse Reinforcement Learning: A Review[J]. Journal of Computer Research and Development, 2019, 56(2): 254-261.
[7] Zhang Haizhu, Fu Xiuhui. Research and Simulation of trajectory planning of six axis industrial manipulator [J]. Science and technology wind, 2020 (32): 172-173 + 184
[8] Guo Yong, Lai Guang. Summary of research on joint space trajectory planning and optimization of industrial robots [J]. Mechanical transmission, 2020, 44 (02): 154-165
[9] Li Li, Shang Junyun, Feng Yanli, Huai Yawen. Review on trajectory planning of articulated industrial robots [J]. Computer engineering and applications, 2018, 54 (05): 36-50
[10] Du Zhijiang, Wang Wei, Yan Zhiyuan, et al. Human computer interaction method of minimally invasive surgical manipulator based on fuzzy reinforcement learning [J]. Robot. 2017, 39 (03): 363-370
[11] Arel I. Deep reinforcement learning as foundation for artificial general intelligence[M]. Theoretical Foundations of Artificial General Intelligence. Atlantis Press, Paris, 2012: 89-102.