Research on automatic control of end position of flexible manipulator based on deep reinforcement learning

Guoqiang Luan, Renjun Chen, Junqi Zhang, Yutong Han

Department of Mechanical Engineering, College of Arts and Information Engineering, Dalian Polytechnic University, Dalian, 116400, China

Keywords: Deep reinforcement learning; Flexible manipulator; End position; Automatic control

Abstract: Due to the fact that traditional industrial flexible robotic arms often use rigid materials with high self weight in the connecting rods and joint components to ensure repetitive accuracy at the end, in order to minimize the displacement of the end position caused by the elastic deformation of the material. This paper investigates the automatic control problem of the end position of flexible robotic arms based on the DRL (Deep reinforcement learning) algorithm. Aiming at the problem that DDPG (Deep Deterministic Policy Gradient) algorithm takes a long time in the training process, a DDPG algorithm based on transfer learning is proposed to plan the end position of flexible manipulator. The motion planning problem of flexible manipulator is decomposed into the superposition of the action output of initial motion planning strategy and the action output of DRL strategy to solve the problem of automatic control of the end position of flexible manipulator. The experimental results show that the proposed residual DRL method is basically stable above 90% at the end of 5000 rounds of training. However, the success rate of DDPG algorithm is rising slowly with the training, and the success rate is close to 60% in 5000 rounds. The experimental results show that DDPG algorithm based on transfer learning can effectively improve the efficiency of DRL agent training.

1. Introduction

Flexible manipulator has the advantages of light weight, small inertia, high efficiency, low energy consumption, large load-mass ratio and compact construction design, and has been widely used in mechanical manufacturing, aerospace and other fields [1]. Flexible manipulator has been widely used in solving repetitive and high-risk tasks, but in the face of complex situations, especially when the flexible manipulator itself has faults, the control strategy becomes particularly complicated.

In order to ensure the repeatability of the end of the traditional industrial flexible manipulator, rigid materials with large self-weight are often used in the connecting rod and joint parts, so as to minimize the end position deviation caused by the elastic deformation of the materials. Therefore, the weight of the chassis and connecting rod of the flexible manipulator becomes very heavy, which obviously increases the driving energy consumption, increases the requirements of the use environment, and brings a lot of inconvenience to the installation and transportation of the flexible manipulator [2-3]. In this paper, the automatic control of the end position of flexible manipulator based on DRL (Deep reinforcement learning) algorithm is studied. Aiming at the problem that DDPG (Deep Deterministic Policy Gradient) algorithm takes a long time in the training process, a DDPG algorithm based on transfer learning is proposed to plan the end position of flexible manipulator.

2. Research method

2.1. Problem description

In recent years, with the development of robotics, the application of flexible manipulator with light weight, high speed and high load-to-weight ratio has become a trend in industrial

  DOI: 10.25236/icceme.2023.004

manufacturing and aerospace fields [4-5]. Compared with rigid flexible manipulator, flexible manipulator has incomparable advantages. Due to the flexibility of the joint or connecting rod structure, the flexible manipulator produces elastic vibration during the movement, which leads to the decrease of the positioning accuracy of the end. However, the actual industrial operation often requires high positioning accuracy, which makes the application of flexible arm a difficult problem.

The application of DRL algorithm to the motion planning of flexible manipulator is a research hotspot in the field of robot control in recent years. However, compared with the application of DRL in other fields, the control of flexible manipulator has its particularity, and not all DRL algorithms can be directly applied to the control of flexible manipulator. The structure of the flexible manipulator considered in this paper is shown in Figure 1. The flexible manipulator consists of a base, a main arm, a flexible joint long arm and a working load.

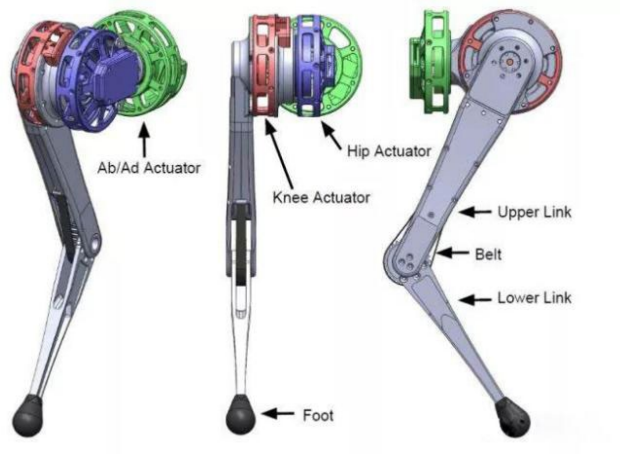

Figure 1 Structure diagram of flexible manipulator

Among them, the main arm usually plays the role of main bearing, and its size is short, its cross section is thick and its stiffness is large, so its flexibility effect can be ignored. Flexible long arms are usually long in size, thin in section and flexible in order to reduce weight. In order to suppress the residual vibration of the flexible arm, two pairs of piezoelectric plates are symmetrically attached to the root of the flexible arm, one of which is used as a sensor and the other as an actuator.

The connecting rod of flexible manipulator is made of light and flexible materials, which has high load-to-weight ratio, small inertia and rapid action response. From the light flexible manipulator mounted on spacecraft to the hard disk drive device, it can be regarded as the practical application of flexible arm [6]. In this paper, the centralized mass method is chosen as the modeling method of flexible connecting rod; At the same time, Lagrange's energy analysis method is chosen to establish the dynamic equation of the flexible manipulator. According to the actual situation, it can be assumed that the connecting rod of the flexible manipulator only produces transverse vibration perpendicular to the rod direction, and does not produce longitudinal vibration along the rod direction, and the connecting rod does not undergo torsional deformation [7-8].

In fig. 1, we define the terminal variable ξ_i of the flexible arm connecting rod, which represents the included angle formed by the connecting line between the terminal and the origin of the deformed connecting rod i and the horizontal axis $\widetilde{O}_{i-1}\widetilde{X}_{i-1}$. According to the geometric relationship, it is easy to get the satisfaction of $\theta_i, \xi_i, y_i(l_i)$:

$$\xi_i = \theta_i + arctg[y_i(l_i)/l_i] \tag{1}$$

Then, we define the terminal vector $\xi = [\xi_1, \xi_2]^T$ of the flexible arm, and according to the above formula, we can obtain the vector relationship among θ, ξ, δ:

$$\xi = \theta + \Gamma\delta \tag{2}$$

The standard model is obtained through the mechanism modeling of the spatial two-link three-degree-of-freedom flexible manipulator, so it should have the inherent vibration characteristics of the flexible manipulator. Besides, the flexible manipulator system is one of the classic examples of non-minimum phase system, so its theoretical model should be underactuated under the condition of closed-loop feedback control.

2.2. Automatic control of end position of flexible manipulator

With the rapid development and popularization of robot technology, the motion planning function of flexible manipulator plays an increasingly important role. There are many defects in the traditional teaching motion planning method of flexible manipulator. Using DRL algorithm to solve the motion planning problem is a new research trend in recent years. DRL makes use of the ability of deep learning to map from environment perception to action decision-making. DRL can act on the environment according to actions and gain the ability of feedback self-learning. DRL begins with deep Q learning. Another important feature of DRL is the need to compromise between "exploration" and "utilization" and weigh their respective proportions in the training process [9]. "Utilization" refers to the behavior that the agent has learned through DRL algorithm in a certain state.

In order to complete the grasping task of multi-degree-of-freedom flexible manipulator in 3D space, it is necessary not only to plan the path according to the position of the target, but also to consider the posture of the end of the flexible manipulator when grasping the target. Aiming at the problems of long training time and large training samples in the training process of DDPG algorithm, this paper improves DDPG algorithm by using transfer learning algorithm, and puts forward DDPG algorithm based on transfer learning. Experiments show that the proposed algorithm has faster training speed than DDPG algorithm.

Transfer learning can reduce the cost of learning, reuse the previous work, do not have to start learning from scratch, and can be quickly transferred in the face of new tasks, which is more timely. The application of transfer learning has been very extensive, and intelligent customer service and voice assistant all have the technical figure of transfer learning [10]. In this paper, the motion planning problem of flexible manipulator is decomposed into the superposition of the action output of initial motion planning strategy and DRL strategy, which is called residual DRL motion planning. The idea of superposition output of initial strategy and DRL strategy is regarded as the core idea of the method proposed in this paper to solve the problem of automatic control of the end position of flexible manipulator.

If the initial policy is π_0, for the input state s, the output action of the initial policy is:

$$a_0 = \pi_0(s) \tag{3}$$

Substitute τ_0 as the system input of the initial operation into the actual flexible manipulator operation system, and repeat the operation continuously. After each repeated operation, perform an iterative operation, and correct the control input of the last time. For using the same method, the output error signal of the last operation is transformed into the same basis function space expression, that is, the control input function after the i iteration can be obtained:

$$\tau_i(t) = \Phi(t)\Psi_i + w(t) \tag{4}$$

After migrating the agent in the source domain, the agent in the target domain needs to continue learning in the face of the environment in the target domain under a new reward function different from that in the source domain. Formula (5) is the overall state value function after policy migration.

$$Q(s,a) = Q_{source}\left(\delta(s_{target}), a_{source}\right) + Q_{target}\left(s_{target}, v(a_{source})\right) \tag{5}$$

In order to reduce the terminal error through iterative learning, it is necessary to correct the next system input by the following formula (6) according to the position error of the terminal point of the flexible manipulator in the current operation after each operation. That is, the iterative learning process is carried out in the basis function space, and the system input τ_{i+1} of the $i+1$ th time is adjusted according to the following formula after the i th operation:

$$\tau_{i+1} = \tau_i + L_p \varepsilon_i \tag{6}$$

Among them, ε_i is the position error of the end point of the flexible manipulator running for the i time, and L_p is the design of iterative learning law in the basis function space.

In this way, the neural network parameters in the source task can be completely preserved without modification. When learning the strategy of a new task, on the basis of the strategy of the source task, the strategy of the new task is studied according to the changed reward function, that is, the reward function in the target task, which greatly reduces the previous learning and training time.

In this paper, firstly, training is carried out on a 2D flexible manipulator simulation model without physical properties. The main purpose of this process is to find a reasonable setting mode of state vector and reward function. Then, based on this setting mode, training is moved to a 3D simulation environment with physical properties. The physical properties of the flexible manipulator in the 3D simulation environment are very close to those of the real flexible manipulator in the real world.

3. Simulation experiment

In order to verify the effectiveness of the automatic control method of the end position of flexible manipulator proposed in this paper, the simulation experiment of the automatic control of the end position of flexible manipulator is carried out. The flexible manipulator is made of aluminum alloy. The DDPG algorithm based on transfer learning is used to train the position planning of robot left arm in an obstacle environment.

Fig. 2 is a control error variation curve of the left arm position planning of the training robot based on the DDPG algorithm of transfer learning in the presence of obstacles.

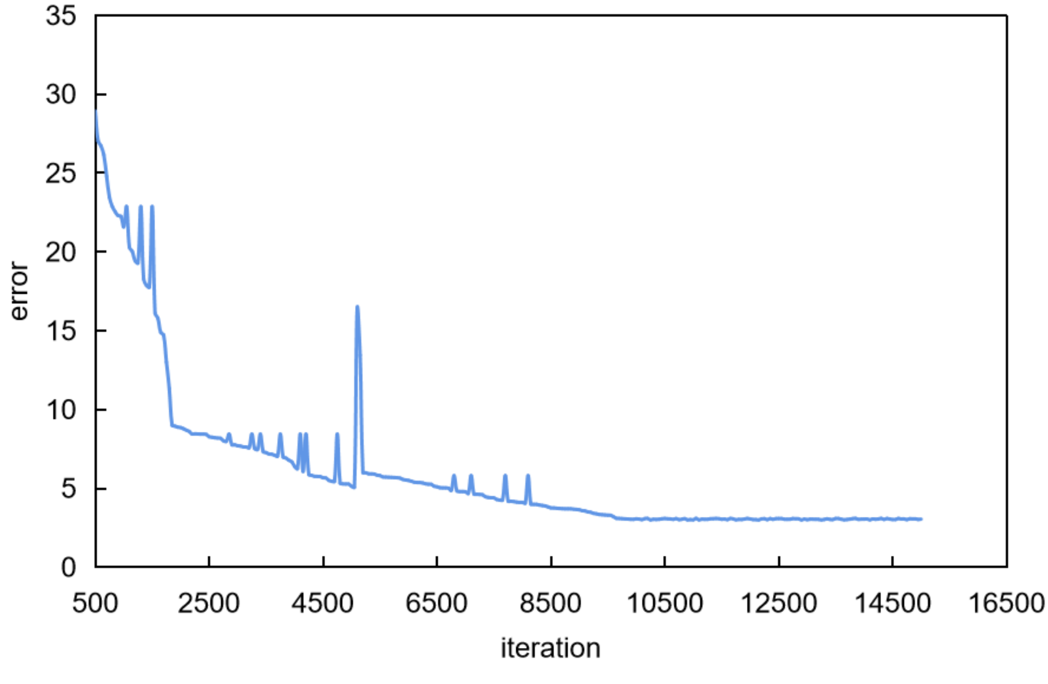

Figure 2 Variation diagram of training process control error

The control error was large at the beginning. After the strategic gradient descent training, the control error decreased obviously, and reached a stable state after about 10,000 steps. The control error decreased slightly, the gradient of neural network update became smaller, and the training process approached to a stable state.

The success rates of DDPG algorithm based on migration learning and DRL training with DDPG algorithm alone are counted. The success rate of the currently trained agent strategy in the experimental environment is evaluated and tested 50 times every 10 rounds, and the current success rate is calculated. Figure 3 shows the evaluation results of success rate.

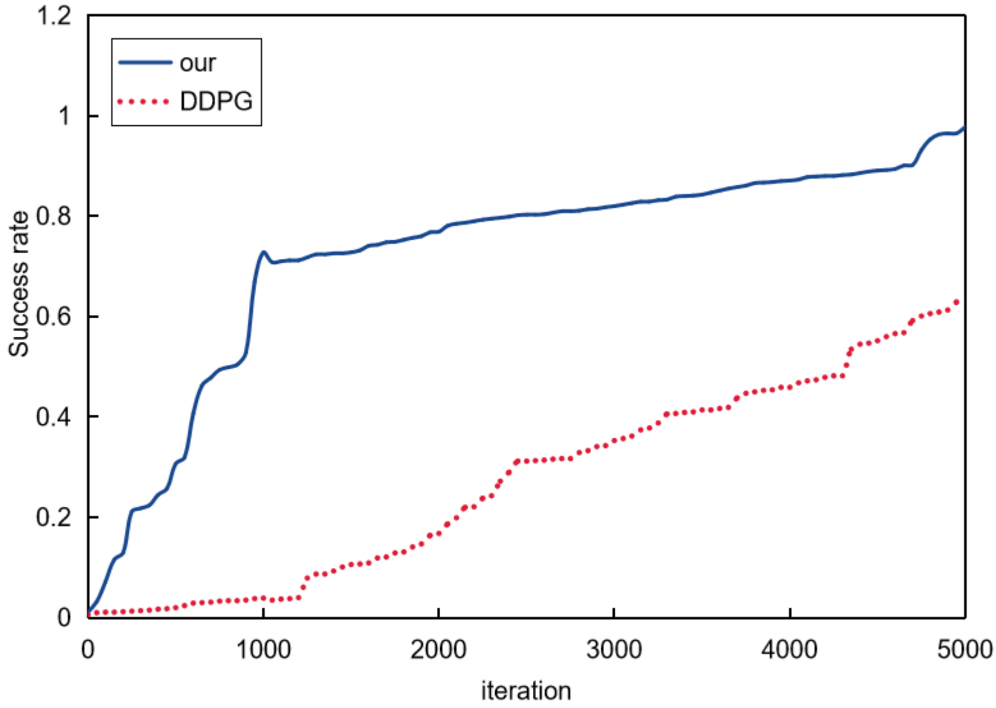

Figure 3 Comparison of success rate

The experimental results show that the success rate of the proposed residual DRL method has reached more than 60% after 1000 steps of training, and the success rate of DDPG algorithm at this time is close to zero. At the end of 5000 rounds of training, it is basically stable at more than 90%. However, the success rate of DDPG algorithm is rising slowly with the training, and the success rate is close to 60% in 5000 rounds.

Through this experiment, it can be found that the proposed DDPG algorithm based on transfer learning is much faster than the single DDPG algorithm for DRL training, and the effect is obvious. The experimental results show that DDPG algorithm based on transfer learning can effectively improve the efficiency of DRL agent training.

4. Conclusions

Flexible manipulator has been widely used in solving repetitive and high-risk tasks, but in the face of complex situations, especially when the flexible manipulator itself has faults, the control strategy becomes particularly complicated. In this paper, the automatic control of the end position of flexible manipulator based on DRL algorithm is studied. Aiming at the time-consuming problem of DDPG algorithm in the training process, a DDPG algorithm based on transfer learning is proposed to plan the end position of flexible manipulator. The experimental results show that DDPG algorithm based on transfer learning can effectively improve the efficiency of DRL agent training.

References

[1] Shitole C, Sumathi P. Sliding DFT-Based Vibration Mode Estimator for Single-Link Flexible Manipulator[J]. IEEE/ASME Transactions on Mechatronics, 2015, 20(6):1-8.

[2] Fareh R, Al-Shabi M, Bettayeb M, et al. Robust Active Disturbance Rejection Control For Flexible Link Manipulator[J]. Robotica, 2019, 38(1):1-18.

[3] Zhang H. Vibration characteristics of flexible manipulator based on EVO trajectory effect by the constraints and parameters variation[J]. Journal of Vibroengineering, 2022, 24(3):20.

[4] Yao W, Zhang D, Guo Y, et al. Improved hierarchical sliding mode control for flexible-joint manipulator[J]. Dongnan Daxue Xuebao (Ziran Kexue Ban)/Journal of Southeast University (Natural ence Edition), 2018, 48(2):201-206.

[5] Gao H, He W, Zhou C, et al. Neural Network Control of a Two-Link Flexible Robotic Manipulator Using Assumed Mode Method[J]. IEEE transactions on industrial informatics, 2019, 15(2):755-765.

[6] Rong J L, Xiang Y, Xin P F, et al. Vibration Control and Experiment Research of Two-Link Flexible Manipulator Considering Flexibility of the Joint[J]. Yuhang Xuebao/Journal of Astronautics, 2017, 38(10):1024-1031.

[7] Veysi M, Soltanpour M R, Khooban M H. A novel self-adaptive modified bat fuzzy sliding mode control of robot manipulator in presence of uncertainties in task space[J]. Robotica, 2015, 33(10):2045-2064.

[8] Carmichael M G, Liu D, Waldron K J. A framework for singularity-robust manipulator control during physical human-robot interaction[J]. The International Journal of Robotics Research, 2017, 36(5-7):874.

[9] Song K T, Jiang S Y, Lin M H. Interactive Teleoperation of a Mobile Manipulator Using a Shared-Control Approach[J]. IEEE Transactions on Human-Machine Systems, 2016, 46(6):834-845.

[10] Li L, Hu Y. Autonomous positioning control of manipulator and fast surface fitting based on particle filter and point cloud library technology[J]. International Journal of Advanced Robotic Systems, 2017, 14(5):82.