Application of Deep Reinforcement Learning in Quantum Control

Yiming Cheng[1,3] [0000-0002-3679-9550], Yixi Yin[1] [0000-0003-2535-596X], Jiaying Lin[2,*][0000-0001-7036-6815], Yaobin Wang[3]

[1] Officers College of PAP, Chengdu, Sichuan, 610213, China

[2] School of Science, Beijing University of Posts and Telecommunications, Beijing, 100876, China

[3] Southwest University of Science and Technology, Sichuan, 621010, China

[*]Corresponding author's e-mail: LinJY@bupt.edu.cn

Abstract. Machine learning technology based on artificial neural network has been successfully applied to solve many scientific problems. One of the most interesting areas of machine learning is reinforcement learning, which has natural applicability to optimization problems in physics. In the quantum control task, it is necessary to find a set of optimal control functions to transfer a quantum system from the initial state to the target state with the highest fidelity possible, which is essentially an optimization task. In this paper, we use Deep Deterministic Policy Gradient algorithm (DDPG) to study the quantum control tasks. We use the algorithm to control the transfer of several quantum systems from one state to another. The results show that DDPG algorithm can find a control strategy to make the fidelity of the final state and the target state of the quantum system be maximum value 1. The results show the potential of DDPG in quantum control.

1. INTRODUCTION

In many areas of physics, such as NMR experiments, ultracold atom systems and quantum computing [1], we need the ability to prepare a physical system to an ideal state. So the development of reliable quantum control technology is essential. In the quantum control tasks, the time evolution of a controlled quantum system satisfies the Schrödinger equation

$$\frac{d}{dt}\psi_t = H_0\psi_t + \sum_{i=1}^{r} u_i(t)\, H_i\psi_t \tag{1}$$

where $H_0$is freedom of system evolution of Hamiltonian,$H_i$is the control Hamiltonian, $u_i(t)$ is the corresponding time-dependent control function. The goal of quantum control is to find a set of control functions $u_i(t)$ to transfer the controlled system from the initial state $|\psi_0\rangle$to the target state $|\psi_f\rangle$. The quality of the state transfer at time t is evaluated using the fidelity, defined as

$$F = \left|\langle\psi_f|\psi_t\rangle\right|^2 \tag{2}$$

Therefore, the key problem is how to find a set of control functions to make the final state of the system close to the target state. With the development of machine learning [2,7,8,9] in recent years, it has been used to solve various complex optimization control problems in science and engineering.

In 2019, Wang Xin [3] et al compared the performance of stochastic gradient Descent (SGD), Krotov optimization algorithms and deep reinforcement learning algorithms such as DQN and PG in the task of controlling qubit evolution. The results show that the DQN and PG algorithms based on deep reinforcement learning are superior to the other two algorithms. However, the DQN and PG used in their work selected control functions only in a discrete action space, so the control sequence that can be found by reinforcement learning will be limited, and it may be impossible to control the evolution trajectory of a quantum system throughout the whole Hilbert space. In this paper, we use Deep Deterministic Policy Gradient (DDPG) algorithm [4] suitable for continuous action space to study its performance in quantum control tasks.

2. THE MODEL SYSTEM

In this section, we introduce several quantum systems to be controlled.First we control the evolution of single qubit and multiple qubits whichonlycontains a single control parameter.We then control a more complex quantum system with two control parameters.

2.1 Single qubit and multiple qubitssystem

We first consider a simple model that control the evolution of a single qubit， which Hamiltonian is

$$H = u(t)\sigma_z + \sigma_x \tag{3}$$

where σ_z and σ_x are pauli operator, $u(t)$ is control parameter. Here the deep reinforcement learning task is to find the best $u(t)$, makes the qubit from initial state $|\psi_0\rangle$ transferred to the final state $|\psi_f\rangle$

$$|\psi_f\rangle = U(t)|\psi_0\rangle \tag{4}$$

where$U(t) = e^{\int_0^T -iH(t)dt}$. We make the initial state $|\psi_0\rangle = [1,0]$ and the final state $|\psi_f\rangle = [0,1]$.

We've looked at single qubit, and now we're looking at multiple qubits. Consider the following Hamiltonian

$$H[u_x(t)] = -\sum_{j=1}^{L} S_{j+1}^z S_j^z + gS_j^z + u(t)S_j^x \tag{5}$$

Set g to 1 in the calculation and consider the case of L=2.There is no restriction on the selection of the initial state and the final state. We set the initial state as the ground state of $u(t) = -2$ and the final state as the ground state of $u(t) = 2$.

2.2 Quantum system with three-mode interaction

we consider a quantum system with three-mode interaction, in which two modes are coupled together by an intermediate mode with two coupling coefficients [5]. Each mode can be regarded as a simple harmonic oscillator. The Hamiltonian of this system is

$$H = g_1(t)\hat{a}_1^+\hat{a}_3 + g_2(t)\hat{a}_2^+\hat{a}_3 + H.c. \tag{6}$$

Where $\hat{a}_i^+(\hat{a}_i)(i = 1,2,3)$ is the creation(annihilation) operator of every mode, $g_1(t) and g_2(t)$is the coupling strength between the corresponding two modes. This hamiltonian can be written as

$$H(t) = \begin{bmatrix} 0 & g_1(t) & 0 \\ g_1(t) & 0 & g_2(t) \\ 0 & g_2(t) & 0 \end{bmatrix} \tag{7}$$

The evolution of this quantum system

$$\psi_t = [\hat{a}_1(t), \hat{a}_3(t), \hat{a}_2(t)]^T \tag{8}$$

is determined by Heisenberg equation

$$\frac{id\psi_t}{dt} = H(t)\psi_t \tag{9}$$

We consider the initial state of this system in mode a_1

$$|\psi_0\rangle = [1,0,0]^T \tag{10}$$

We use the DDPG algorithm to find a suitable set of time-varying coupling coefficients $[g_1(t), g_2(t)]$ to transfer the state to the mode a_2. So the target state $|\psi_f\rangle = [0,0,1]^T$.

3. IMPLEMENTION AND RESULTS

In this section, we will introduce how to use Deep Deterministic Policy Gradient algorithm (DDPG) for quantum control, and give the calculation results.

3.1 Deep reinforcement learning algorithm

In reinforcement learning [6], the agent needs to observe the environment to get the state of the environment, and then choose the next action according to the observed state. After the environment receives the action of the agent, the state changes and returns the changed state to the agent, at the same time giving the agent a reward. Thenthe agent decides the next action according to the feedback information.

Deep reinforcement learning is the combination of deep learning and reinforcement learning.It uses neural network to act as an agent. The DDPG algorithm we used is based on the work [4]. It adopts an actor-critic structure. A network called Critic is used to approximate the value function, its input is the action of the agent and the state of the environment and its output is the value function about the state. The other is called Actor network, its role is to fit a policy function, its input is the state ofthe environment and the output is the next action of the agent.

In the calculations, we use feedforward neuralnetworkto act the Critic and Actor. Each neural network consists of one input layer, two hidden layers where every layer contains 30 neurons and one output layer.

3.2 Results

In our control task, we divided the evolution time T into N segments on average

$$dt = \frac{T}{N} \tag{11}$$

During the evolution time of each segment dt, the coupling coefficientsg_1, g_2 and the control parameter$u(t)$ are fixed value andthey are selected in the interval $[0, E]$ as the action of the agent. We use the state of the quantum system ψ_t as the state of the environment. The reward function defined as

$$R = \begin{cases} 10F(t), & F \in (0, 0.5) \\ 100F(t), & F \in (0.5, 0.9) \\ 1000F(t), & F \in (0.9, 1) \end{cases} \tag{12}$$

where$F(t)$is the fidelity of each segment.

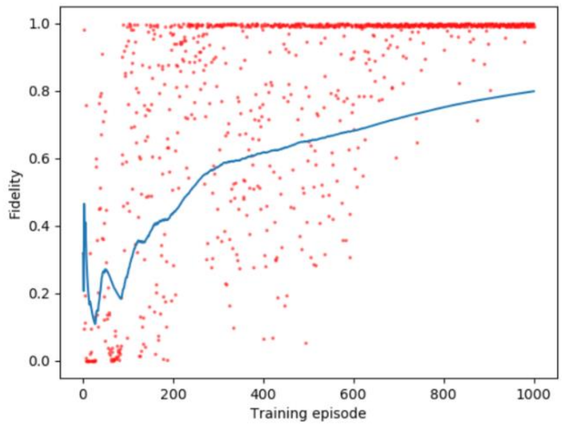

Figure 1. single qubit

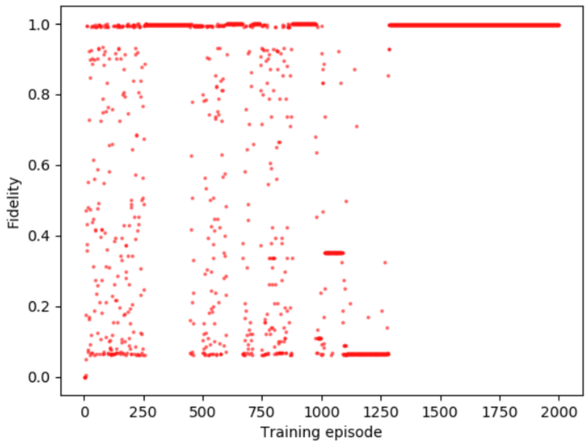

Figure 2. multiple qubits

Fig. 1 shows the training result of single qubit evolution and Fig. 2 shows the training result of multiple qubits evolution. In our calculations, we choose $T = 2\pi$, $N = 20$, $E = 20$. The red dot represents the fidelity at the end of each period, and the blue line is the cumulative average of fidelity for all periods. It can be seen that with the increase of training periods, DDPG algorithm can find the best control strategy and make the fidelity equal to 1.

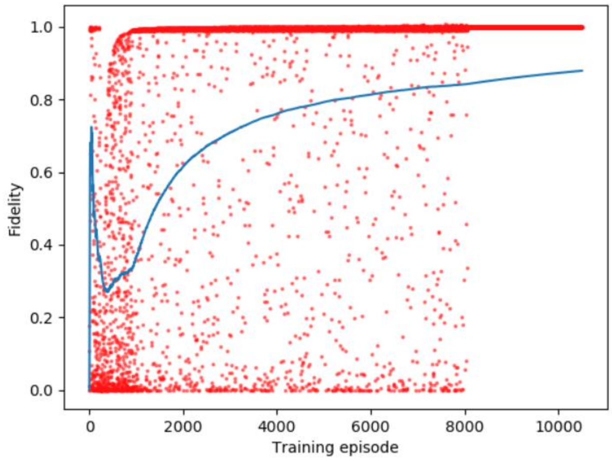

Figure 3.quantum system with three-mode interaction

The Fig. 3 shows the result of quantum system with three-mode interaction. In our calculations,we choose $T = 5$, $N = 50$, $E = 20$. The first 8000 periods is training phase. To explore strategies, the algorithm randomly choosea coupling coefficient in a very small probability ε, and with $1-\varepsilon$ probability choose the optimal value. With the increase of training periods, the ε gradually narrowed to 0. It can be seen that the fidelity of most training periods can reach the maximum value 1, while the fidelity of a few periods will fluctuate due to random actions. Within 8000-10000 periods we stop exploring the strategy, so the coupling coefficient selected by the algorithm for each period is the optimal value. It can be seen that in this case, the fidelity of each period can be maximum value 1.

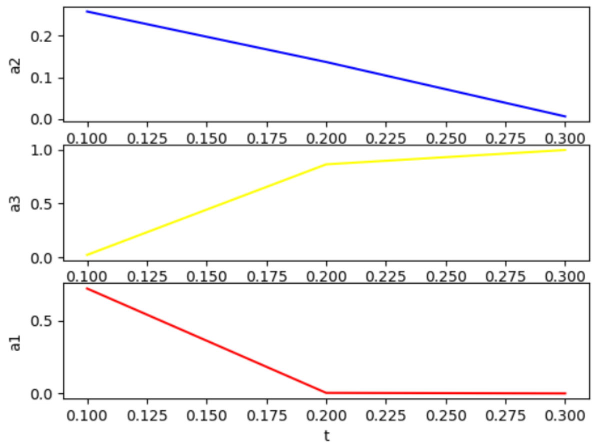

Figure 4.populations of the three modes

Fig. 4 shows that under the optimal strategy, the populations of the three modes change over time. when the control time $t = 0.3$, we can see the evolution of the quantum system is end. Andthepopulationof mode $a_3$goes up to 1, the population of mode $a_1$goes down to 0.

4. CONCLUSION

We have successfully controlled several quantum systems with Deep Deterministic Policy Gradient algorithm. First, we study the control task of single qubit. The results show that the algorithm can control the evolution of qubit with fidelity close to 1. Then the research object is extended to the case of multi-qubits and the result shows that DDPG also can accomplish the task well. Second, we control a more complex quantum system with three-mode interaction, in which two modes are coupled together by an intermediate mode with two coupling coefficients. We use DDPG algorithm to control the transfer of the quantum system from one mode to another by finding the best set of coupling coefficients so that the fidelity close to maximum value 1 and give the population changes with time.

Acknowledgments
This work is supported financially by the Natural Science Foundation of Sichuan (2023NSFSC1966) and Smart Education Research Fund of Southwest University of Science and Technology (22ZHJYZD02, 22SXB004).

References
[1] Nielsen M A., Chuang I. Quantum Computation and Quantum Information: 10th Anniversary Edition[M]. 2011.
[2] Russell S.J., Norvig P. Artificial Intelligence—A Modern Approach[M].2003.
[3] Zhang X M, Wei Z, Wang X. When does reinforcement learning stand out in quantum control? A comparative study on state preparation[J]. npj Quantum Information, 2019,5:85.
[4] Lillicrap T P, Hunt J J, Pritzel A, Heess N, Erez T, Tassa Y, Silver D, Wierstra D. Continuous control with deep reinforcement learning[J]. arXiv preprint arXiv:1509.02971, 2015.
[5] Zhang H, Song X, Ai Q, Wang H, Yang G, and Deng F. Fast and robust quantum control for multimode interactions using shortcuts to adiabaticity[J]. Optics express, 2019,27(5):354520.
[6] Sutton R.S., Barto A. G. Reinforcement Learning: An Introduction [M]. MIT Press, 1998.
[7] Guo M, Wang Y, Huang J, Wang Q, Mo Z, Zhang Y and Shen C. Rgs-SpMM:Accelerate sparse matrix-matrix multiplication by row group splitting strategy on the GPU, The 19th Annual IFIP International Conference on Network and Parallel Computing (NPC2022) , 2022.09.
[8] Mo Z, Wang Y, Zhang Q, Zhang G, Guo M, Zhang Y and Shen C. The Parallelization and Optimization of K-means Algorithm Based on MGPUSim, 31st International Conference on Artificial Neural Networks (ICANN2022), Bristol, UK, 2022.09.
[9] Wang Y, An H, Liang B, Wang L, Cong M, Ren Y. Balancing Thread Partition for Efficiently Exploiting Speculative Thread-Level Parallelism, In： Proceeding of the 7th International Symposium on Advanced Parallel Processing Technology(APPT2007), pp. 40-49, 2007.