Noname manuscript No.
(will be inserted by the editor)

A Quantum System Control Method Based on Enhanced Reinforcement Learning

Wenjie Liu[1,2] · Bosi Wang[2] · Jihao Fan[3] · Yebo Ge[2] · Mohammed Zidan[4]

Received: date / Accepted: date

Abstract The design of quantum system control is a key task to a powerful quantum information technology. In practical, traditional quantum system control methods often face different constraints, and are easy to cause both leakage and stochastic control errors under the condition of limited resources. Reinforcement learning has been proved as an efficient way to complete the quantum system control task. So a quantum system control method based on enhanced reinforcement learning (QSC-ERL) is proposed. A satisfactory control strategy is obtained through enhanced reinforcement learning so that the quantum system can be evolved accurately from the initial state to the target state. According to the number of candidate unitary operations, the three-switch control is used for simulation experiments. Compared with other methods, the QSC-ERL can achieve high fidelity learning control of quantum systems and improve the efficiency of quantum system control.

✉ Wenjie Liu
wenjiel@163.com

Bosi Wang
bosi@nuist.edu.cn

[1] Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing 210044, China

[2] School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China

[3] School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

[4] Hurghada Faculty of Computers and Artificial Intelligence, South Valley University, Egypt

Keywords quantum system control · reinforcement learning · quantum computing · machine learning · neural networks

Mathematics Subject Classification (2020) 81Q93 · 81P68 · 68T07

1 Introduction

Quantum system control is considered as an important task of a quantum information technology, aiming for realizing active manipulation or control. For example, the laser technology is used to achieve the desired energy eigenstates of atoms, and the control field is applied to control the quantum gate to reach the specified superposition state (Chu 2002). It can be seen that a good control method is needed to control quantum state from initial state to target state.

Traditional learning algorithms (such as gradient algorithms (Chakrabarti and Rabitz 2007; Roslund and Rabitz 2009), genetic algorithms (Rabitz et al. 2000; Tsubouchi and Momose 2008) have been used in quantum system control, and all of them have shown excellent control effects under specific experimental environment. But in practical, the quantum system to be manipulated usually has different restrictions. There is a class of quantum system control problem with limited control resources. In this case, the gradient-based algorithms are not suitable for solving the above problems. The genetic algorithms need a lot of experimental data to optimize the control performance that complicates the resolution of the problem.

Reinforcement learning (Fang et al. 2020) interacts with the environment in the form of rewards and punishments. With the advent of quantum information technology and the upsurge of machine learning, many researchers combine the two fields to carry out studies (Dong et al. 2008; Chunlin et al. 2012; Chen et al. 2013; Palittapongarnpim et al. 2017). In recent years, studies on quantum system control based on reinforcement learning have been increasing gradually. Vedaie et al. (2018) applied reinforcement learning to realize multi-photon interference measurement. Cardenas-Lopez et al. (2018) proposed a protocol for quantum reinforcement learning, which does not require coherent feedback during the learning process and can be implemented in a variety of quantum systems. Fosel et al. (2018) showed how a network-based "agent" can discover a complete quantum error correction method to protect qubits from noise. In addition, Bukov et al. (2018) used reinforcement learning to prepare the desired quantum states. They also successfully used Q-learning (Watkins et al. 1992) to control quantum systems (Bukov 2018). Yu et al. (2019) used quantum reinforcement learning to make a qubit "agent" adapt to the unknown quantum system "environment" to achieve maximum overlap. Niu et al. (2019) used deep reinforcement learning and proposed a quantum control framework for fast and high-fidelity quantum gate control optimization. Zhang et al. (2019) successfully used reinforcement learning algorithm to solve a class of quantum state control problems, and made a theoretical analysis.

The main contributions of this paper are: (1) In order to validate the effectiveness and generality of quantum system control methods based on reinforcement learning, various reinforcement learning algorithms were used for solving the control problem of quantum systems with limited control resources. (2) A quantum system control method based on enhanced reinforcement learning (QSC-ERL) is proposed to improve the fidelity of quantum state evolution.

The rest of this paper is structured as follows. In Sec. II, we briefly overview the preliminaries about quantum system control and reinforcement learning. In Sec. III, we model the quantum system control problem and present our novel method. In Sec. IV and V, we show the results of simulation experiments and draw our conclusions.

2 Preliminaries

2.1 Learning control of quantum systems

Learning control methods are powerful for solving quantum system control problems (Ma and Chen 2020). The learning methods are often optimized by multiple iterations to realize the evolution of qubits from an initial state to the desired target state. In this paper, the task of quantum system control is set as the quantum pure state transition control problem of n-order quantum system. For the free Hamiltonian H_0 of n-order quantum system, its eigenstate can be defined as $D = \{|\phi_i\rangle\}_{i=1}^{N}$. The quantum state to be evolved $\left|\psi_{(t)}\right\rangle$ of a controlled system can be extended according to the eigenstates in set D:

$$\left|\psi_{(t)}\right\rangle = \sum_{i=1}^{N} c_i(t) \left|\psi_i\right\rangle , \tag{1}$$

where the complex number $c_i(t)$ satisfies $\sum_{i=1}^{N} |c_i(t)|^2 = 1$.

In order to achieve the active control of the quantum system, the control Hamiltonian H_c is introduced into the control $u(t) \in L^2(\mathbf{R})$, which is independent of time and interacts with the quantum system. The $\left|\psi_{(t=0)}\right\rangle$ can be redefined as $|\psi_0\rangle$. The $C(t) = (C_i(t))_{i=1}^{N}$ evolves according to the Schrödinger equation:

$$\begin{cases} \imath\hbar\dot{C}(t) = [A + u(t)B]C(t) \\ C(t=0) = C_0 \end{cases} , \tag{2}$$

where $\imath = \sqrt{-1}$, $C_0 = (c_{0i})_{i=1}^{N}$, $c_{0i} = \langle \varphi_i \mid \psi_0\rangle$, $\sum_{i=1}^{N} |c_{0i}|^2 = 1$, $\hbar$ is the reduced Planck constant, and the matrices A and B correspond to the free Hamiltonian H_0 and the controlled Hamiltonian H_c of the quantum system respectively. $U_{(t_1\to t_2)}$ represents an unitary operation for any state $\left|\psi_{(t_1)}\right\rangle$ of the quantum system. The $\left|\psi_{(t_2)}\right\rangle = U_{(t_1\to t_2)} \left|\psi_{(t_1)}\right\rangle$ of the quantum system is that the quantum state $\left|\psi_{(t_1)}\right\rangle$ evolves from time $t = t_1$ to time $t = t_2$. In addition, $U_{(t_1\to t_2)}$ can also be defined as $U_{(t)}$, $t \in [t_1, t_2]$.

In fact, if the quantum systems evolve freely without control resources limited, it can also arrive at the target state from an initial state. However, there are two unfavorable problems in this way of free evolution control: One is that it is difficult to satisfy the conditions in practice, and will waste a lot of control resources to evolve from an initial state to the desired target state. The other is that free evolutionary control has no certain control law, and is unable to be determined when the quantum system reaches the target state. Our study mainly aims at solving a class of control resource-limited quantum system control problem.

2.2 Quantum control landscapes

The quantum control landscapes (Chakrabarti and Rabitz 2007) has provided a theoretical basis for analyzing the learning control problem of quantum systems, which can be defined as the mapping between the control Hamiltonian and the correlation value of the control performance function. The task of quantum system control can be defined as a problem of maximizing the target performance function. In other words, it can be transformed into a problem of maximizing the state transition probability from the initial state to the desired target state. For the state transition control problem, the quantum control transition can be defined as

$$\begin{aligned} J(u) =& tr(U_{(\varepsilon,T)}|\psi_{initial}\rangle \\ & \langle\psi_{initial}|U^{\dagger}_{(\varepsilon,T)}|\psi_{target}\rangle\langle\psi_{target}|), \end{aligned} \tag{3}$$

where $tr(\cdot)$ is the trace operation, $U^{\dagger}$ is the ad-joint of U, $|\psi_{initial}\rangle$ is the initial quantum state, $|\psi_{target}\rangle$ is the desired target quantum state.

In this paper, it is assumed that the control set $\{u_j, j = 1, 2, \ldots, m\}$ allowed to operate in a controlled quantum system can be given in advance, where each control u_j corresponds to an unitary operation U_j. The goal of learning control is to evolve control from the initial state $|\psi_{initial}\rangle$ to the desired target state $|\psi_{target}\rangle$, and learn a global optimal control sequence u^*:

$$u^* = \arg\max_{u} J(u). \tag{4}$$

2.3 Reinforcement learning

Reinforcement learning (Fang et al. 2020) is described by Markov Decision Process (MDP), which is usually defined by the quadruple $\langle S, A, P, R\rangle$. The S is the set of states, A is the set of actions, and the state $s \in S$, the action $a \in A$. The state transition function $P(s, a, s')$ represents the probability of state transition. The $R(s, a, s')$ represents the reward value function. $P(s, a, s')$ and $R(s, a, s')$ only depend on the current state s and action a that have nothing to do with other historical states and actions. The MDP which adopts the discount criterion is denoted as $M = (S, A, P, \gamma, R)$, where γ is the discount factor.

Reinforcement learning agents learn by interacting with external environment. Specifically, the agent observes the state $s_t \in S$ at each discrete time step $t \in [0, T]$, where T is the end time, and selects an action $a_t \in A$ used for transitioning the state $s_t \in S$ to the next state $s_{t+1} \in S$ with the probability p. After performing an action, the agent is usually given a scalar reward signal r_{t+1}, which reflects how good or bad the action was. The learning process mentioned above is repeated continuously until the agent can learn an optimal strategy, which is a mapping from the state space S to the action set A.

Algorithm 1 The Q-Table learning algorithm

```
Randomly initialize the Q table;
for $episode = 1, M$ do
    Randomly initialize the $s$ state;
    for $step = 1, T$ do
        Select an action $a$ according to the Q table;
        Execute action $a$, receive reward $r$, enter state $s'$;
        $Q(s, a) =$
            $Q(s, a) + \alpha(r + \gamma \max_{a' \in A} Q(s', a') - Q(s, a))$;
        $s \leftarrow s'$;
    end for
end for
```

Q-learning proposed by Watkins et al. (1992) is an offline reinforcement learning algorithm, and is described in **Algorithm 1**. The iteration of the Q-value function and the strategy selection are independent of each other. The approximation goal of Q-learning can be defined as $r + \gamma \max_{a'} Q(s', a')$. The agent can choose actions according to the greedy algorithm or other non-optimal strategies.

3 Methods

3.1 Problem modeling

The two-level quantum system (D'Alessandro and Dahleh 2001) is representative in filed of quantum system control. The spin 1/2 system is one of the typical two-level quantum systems for theoretical and practical research. The state $|\psi\rangle$ of the spin 1/2 system can be defined as:

$$|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{t\phi}\sin\frac{\theta}{2}|1\rangle, \tag{5}$$

where $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi]$ represent the polar and phase angles respectively. A point $\vec{a}$ on the unit sphere can be defined as

$$\vec{a} = (x, y, z) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta). \tag{6}$$

The aim is to design the control of two-level quantum system based on reinforcement learning. In the following, the problem of quantum system control based on reinforcement learning is modeled and described.

The agent in reinforcement learning learns through continuous interaction with the environment. Specific to the quantum system environment, our method divides the state space of the quantum system into a finite discrete set of states S. If the Bloch sphere discretized into multiple "longitude lines" and "latitude lines" is used for representing the state space, each intersection of "longitude" and "latitude" can be defined as each discrete state under the quantum system.

In this paper, Set $A = \{u_j, j = 1, 2, \ldots, m\}$ is defined as a limited set of executable actions (unitary operations) in a quantum environment. Specifically, for the three-switch control, the m is set to 3. Whenever the agent performs action a and the state is transformed from s to s', it will receive the feedback value, and using the fidelity as the reward:

$$r = \begin{cases} 10, \ fidelity \leq 0.5 \\ 100, \ 0.5 < fidelity \leq 0.7 . \\ 10000, \ fidelity > 0.7 \end{cases} \tag{7}$$

The goal of reinforcement learning is to obtain an optimal method π^* and the global optimal control sequence u^* as Eq. (4).

For quantum systems, the agent of reinforcement learning obtains the optimal method by maximizing the long-term cumulative reward in the process of interacting with the environment of quantum systems. Therefore, the agent also needs to constantly interact with the external environment and learns through trial and error. Specifically, the permitted controls at each control step for any quantum state are U_1 (no control), U_2 (positive impulse control), and U_3 (negative impulse control), which is defined as follows:

$$\begin{aligned} &U_1 = e^{-\iota I_z \frac{\pi}{15}}, \\ &U_2 = e^{-\iota (I_z + 0.5 I_x) \frac{\pi}{15}}, \\ &U_3 = e^{-\iota (I_z - 0.5 I_x) \frac{\pi}{15}}, \end{aligned} \tag{8}$$

where $I_z = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, I_x = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. The state of the quantum system in evolutionary control will be limited by the three-switch control. The agent of reinforcement learning will learn under the norms of the three-switch control in interactive learning with the environment of the quantum system. It is mainly embodied in the action selection of the agent in any quantum system state. Under the three-switch control, each action can be performed by the agent is U_1, U_2 and U_3.

Under the above control conditions, a global optimal control method is obtained by using reinforcement learning algorithm to minimize the number of control sequences, so that the spin 1/2 system can reach the target state from the initial state.

3.2 Enhanced reinforcement learning

In order to improve the learning efficiency of Q learning algorithm (Watkins and Dayan 1992) without prior knowledge, it is important to improve the foresight ability of the learning agent. But it brings the following two problems in practice: 1) The state space increases, causing the "dimensionality disaster", which greatly reduces the learning efficiency; 2) The visible space of the learning agent is reduced, making the agent's search process more blind.

Algorithm 2 The enhanced reinforcement learning algorithm

```
Initialize the Q table randomly;
Initialize the neural network as a qualitative layer;
for episode = 1,M do
    Initialize parameters $s, \varepsilon_0, \gamma, \lambda, \alpha, \beta, e = 0$;
    for step = 1,T do
        Generate $\varepsilon \in [0, 1)$ randomly;
        if $\varepsilon \leq 1 - \varepsilon_0$ then
            Select action $a = \arg\max Q(s, a), a \in A$;
            $e = \gamma\lambda e + \frac{\partial}{\partial w} V_{\mathrm{NN}}(s)$
        else
            Select action $a$ randomly;
            if $a == \arg\max_{b \in A} Q(s, b)$ then
                $e = \gamma\lambda e + \frac{\partial}{\partial w} V_{\mathrm{NN}}(s)$
            else
                $e = 0$;
            end
        end
        Execute action $a$, and get $V_{\mathrm{NN}}(s)$ and $V_{\mathrm{NN}}(s')$
        from enhanced neural network;
        Update network:
        $w = w + \beta(r(s, a) + \gamma V_{\mathrm{NN}}(s') - V_{\mathrm{NN}}(s))e$;
        $F(s, a, s') = \gamma V_{\mathrm{NN}}(s') - V_{\mathrm{NN}}(s)$;
        $Q(s, a) = Q(s, a) + \alpha(r(s, a) +$
        $F(s, a, s') + \gamma \max Q(s', a') - Q(s, a))$
        $s \leftarrow s'$;
    end for
end for
```

The enhanced reinforcement learning shown in Fig. 1 consists of a quantitative Q table and a qualitative V value heuristic function obtained by enhanced neural network. And the description of the algorithm is shown in **Algorithm 2**. The agent adopts enhanced neural network in the learning for the table space, which can gradually form a heuristic function. Using the enhanced neural network which has the generalization and foresight ability can discover the evolution trend of the value function in learning, and play a guiding role to avoid the blind behavior of the agent.

3.2.1 Heuristic function based on enhanced neural network

In the enhanced reinforcement learning, the Q-table is updated after each action is executed. At the same time, the enhanced neural network is trained, and a V value fitting surface is gradually formed. Using it as an heuristic function to guide the update of QSC-ERL. And the enhanced neural network shown in Fig. 2 is inspired by common convolutional neural network (Gu et al. 2018) and residual neural network (He et al. 2016) which can make full use of the extracted features. The state s is

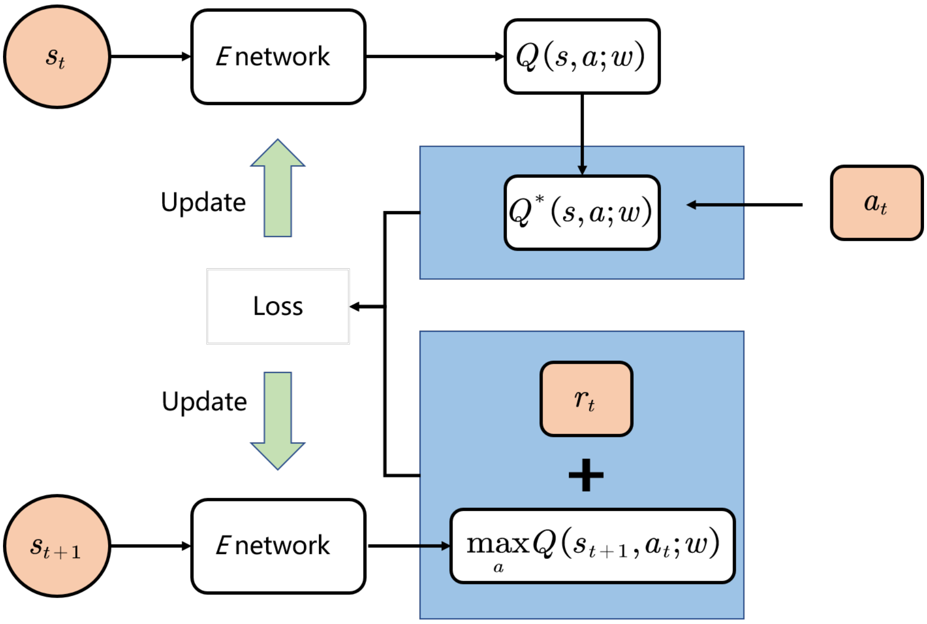

Fig. 1 An overview of enhanced reinforcement learning: the orange rectangle is given by the environment. S_t and S_{t+1} are input into the enhanced neural network which is abbreviated to E network. The algorithm selects Q_t^* according to a_t from $Q_{S_t}$and $maxQ_{t+1}$ from $Q_{S_{t+1}}$ respectively. Then calculating the loss for updating the enhanced neural network between "the blue rectangles".

the input of the neural network, and the Q values got by the probability of actions is the output, where N is the number of actions. In order to obtain the nonlinear characteristics more comprehensively, the LeakyReLU is selected as activation function to give all negative values a non-zero slope.

The heuristic function $F(s,a,s')$ participating in the update of the Q table takes s and s' as input and gets the V value output which is defined as $V_{NN}(s)$ and $V_{NN}(s')$ in state s and s' respectively. And the heuristic function is defined as

$$F(s,a,s') = \gamma V_{NN}(s') - V_{NN}(s). \tag{9}$$

3.2.2 The parameters updating method

The eligibility trace (Singh and Suttun 1996) is an effective method to accelerate the training speed of the reinforcement learning. Updating once can pass the error back several steps to speed up the learning of the enhanced neural network and provide more effective inspiration for the whole algorithm.

The goal of the learning of Q table can be defined as

$$\begin{aligned} Q(s,a) =& Q(s,a) + \alpha[r(s,a,s') + F(s,a,s') \\ &+ \gamma \max_{a'} Q(s',a') - Q(s,a)], \end{aligned} \tag{10}$$

and the updating of V values can be defined as

$$V(s) = \max_a Q(s,a). \tag{11}$$

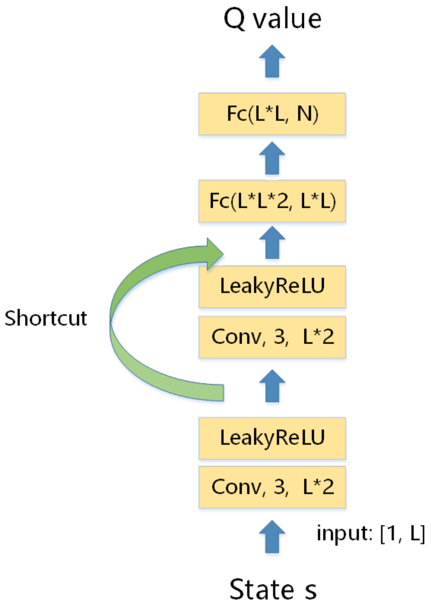

Fig. 2 The enhanced neural network architecture: For each state s fed into the network, the network extracts the features and outputs the Q values.

When the learning agent performs a non-greedy action, its next state often does not obtain the largest Q value. And the QSC-ERL will update the current state-action paired Q value according to the V value of the next state obtained by the greedy strategy. For updating the enhanced neural network, when the agent requires the V value according to the greedy strategy, the eligibility trace is also updated. When the agent performs a non-greedy strategy, the eligibility trace is set as 0, preventing the error from propagating backward.

Updating the weights of the enhanced neural network by the gradient descent method can be defined as

$$\begin{aligned}\varDelta w_t =&\beta(r(s_t)+\gamma V_{\mathrm{NN}}(s_{t+1})-V_{\mathrm{NN}}(s_t))\times \\ &\sum_{k=0}^{t}(\gamma\lambda)^{tk}\frac{\partial}{\partial w}V_{\mathrm{NN}}(s_k),\end{aligned} \tag{12}$$

where β is the learning rate, $0<\beta<1$, λ is the eligibility trace coefficient, $0<\lambda<1$.

The agent updates the weight of the neural network through the difference value $r\left(s_t\right)+\gamma V_{\mathrm{NN}}\left(s_{t+1}\right)-V_{\mathrm{NN}}\left(s_t\right)$ between the next predicted V value of state s and the current target V value. A difference value can update the V value in other state. If the eligibility trace is defined as

$$e_t=\sum_{k=0}^{t}\gamma\lambda\frac{\partial}{\partial w}V_{\mathrm{NN}}(s_k)=\gamma\lambda e_{t-1}+\frac{\partial}{\partial w}V_{\mathrm{NN}}(s_t), \tag{13}$$

Eq. (12) can be rewritten as

$$\varDelta w_t=\beta(r(s_t)+\gamma V_{\mathrm{NN}}(s_{t+1})-V_{\mathrm{NN}}(s_t))e_t. \tag{14}$$

It is easy for modifying the weights from the hidden layer of the neural network to the output layer, and then modify the weights from the input layer to the hidden layer through the back propagation.

The QSC-ERL is carried out synchronously in the learning of Q-Table and the enhanced neural network. The Q-Table based reinforcement learning can obtain more accurate results, but the speed of learning is slow. The enhanced neural network is not accurate enough, but it has better generalization performance. In the initial stage of learning, the effect is not obvious. But with continuous learning, the enhanced neural network is gradually established the trend information, and the convergence speed can be greatly improved.

4 Simulation experiments

4.1 Settings

For simulation experiments, full training for a given scenario can be achieved on a single CPU+GPU workstation (CPU: Intel Xeon Gold 5218, GPU: GeForce RTX 2080 Ti 11G). The state space of the quantum system will be reconstructed from the initial state $s_{initial}=|\psi_{initial}\rangle$ to the target state $s_{target}=|\psi_{target}\rangle$ by the form of the bloch sphere. Specifically, the bloch sphere is discretized by 30 “longitude lines” and 30 “latitude lines”. The state set is $S=\{s_i=|\psi_i\rangle, i=1,2,\ldots,n\}$, and the executable action set is $A=\{a_j=U_j, j=1,2,\ldots,m_\circ\}$. For the spin 1/2 system, the initial state is set as $|\psi_{initial}\rangle(\theta=(\pi/60),\phi=(\pi/30))$, and the target state is $|\psi_{target}\rangle(\theta=(41\pi/60),\phi=(29\pi/30))$.

4.2 Evaluation index

Fidelity is a evaluation index to measure the distance between density operators. It allows us to compare how the state of the system at any given moment is different from the initial state, or how the state of a system is different from a reference state. It allows us to measure quantitatively how different two states really are. For two density matrices ρ, σ it is generalized as the largest fidelity between any two purifications of the given states. And the fidelity function can be defined as

$$F(\rho,\sigma)=(tr\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}})^2, \tag{15}$$

where ρ and σ are the density matrix of source information and target information respectively.

4.3 Results and analysis

The simulation experiments is carried out under the three switch control paradigm. The goal of the experiment is to control the spin 1/2 system from the initial state $|\psi_{initial}\rangle$ to the target state $|\psi_{target}\rangle$. The main purpose is to explore the effectiveness of reinforcement learning algorithm for solving quantum control problem.

Therefore, the simulation experiments is divided for two parts: one is that the tabular Q-learning (TQL) (Sutton and Barto 2018), deep Q-learning (DQL) (Mnih et al. 2015) and policy gradient (PG) (Sutton et al. 2000) are applied to explore the effectiveness of reinforcement learning algorithm for solving quantum control problem. The other is that the NN-QSC (Fosel et al. 2018) and the DRL-QSC (An and Zhou 2019) are compared for verifying that the proposed QSC-ERL performed better than its peers. The parameters of the reinforcement learning algorithms involved in the experiment are set as follows: For all state-action paired, the Q value is initialized to 0, the discount factor is $\gamma=0.99$, the learning rate is $\alpha=0.1$, and the action selection probability is initialized to 1/3.

Fig. 3 shows the comparison of fidelity between algorithms, where the X-axis is the number of episode, and the Y-axis is fidelity. It can be seen from Fig. 3

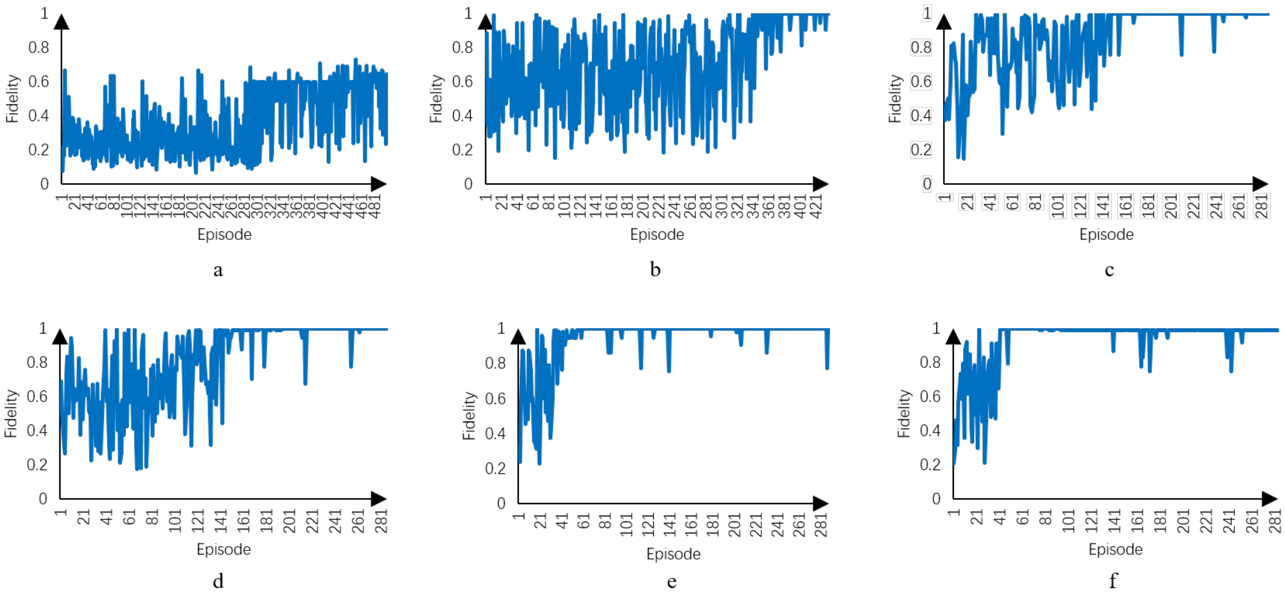

Fig. 3 The comparison of fidelity between algorithms. (a)Fidelity of the TQL algorithm. (b) Fidelity of the PG algorithm. (c) Fidelity of the DQL algorithm. (d) Fidelity of the NN-QSC algorithm. (e) Fidelity of the DRL-QSC algorithm. (f) Fidelity of the QSC-ERL algorithm.

Table 1 The comparison of the number of episodes between algorithms

Name	Episodes	Fidelity
TQL	452	0.73
PG	311	0.99
DQL	135	0.99
NN-QSC	171	0.99
DRL-QSC	60	0.99
QSC-ERL	42	0.99

that reinforcement learning has certain effects on solving the quantum system control problems. However, the TQL algorithm and the PG algorithm cannot get a better fidelity. The DQL algorithm adopted convolutional neural network to guide the learning of the Q learning algorithm. But if want to get a high fidelity between the final state and the target state, it needs to train with more episodes. Table 1 shows the number of episodes when the fidelity can get the maximum, and total number of episodes is set to 500. The data is taken from the average value of 100 experiments. It represents that the ability of algorithms can make the quantum system from the initial state to the desired target state. The experimental results show that most methods converge after training and make the quantum system from the initial state to the desired target state. Specifically, for the TQL, the maximum of the fidelity is 0.73, and others can reach 0.99 after total training. The PG requires about 311 episodes and the DQL requires about 135 episodes. It means that the reinforcement learning based on neural network has the better performance in some degree than the common RL algorithm. The NN-QSC requires about 171 episodes to control the evolution of the quantum system from the initial state to the target state while the DRL-QSC requires about 60 episodes, and the QSC-ERL requires the 42 episodes. So our proposed QSC-ERL algorithm is faster than the NN-QSC and the DRL-QSC for controlling the evolution of the quantum system from the initial state to the target state.

5 Conclusion

In this paper, a quantum system control method based on enhanced reinforcement learning (QSC-ERL) is proposed to achieve the learning control of the spin 1/2 system. A satisfactory control strategy is obtained through enhanced reinforcement learning so that the quantum system can be evolved accurately from the initial state to the target state. Compared with other methods, our method can achieve the quantum system control with high fidelity, and improve the control efficiency of quantum systems.

It should be noted that our method is sufficient for the evolution of quantum state in spin 1/2 system. Other difficult quantum control problems include quantum error correction based on bosonic codes (Michael et al. 2016) and quantum state preparation in the single-photon manifold (Vrajitoarea et al. 2020). And it is a valuable work to conduct a study on providing solutions by using learning theories (Li et al. 2018; Zhang and Wang 2020) and neural network (Xu et al. 2019; Hu et al. 2020), which is also one of our next research.

Acknowledgements The authors would like to thank the anonymous reviewers and editors for their comments that improved the quality of this paper. This work is supported by the National Natural Science Foundation of China (62071240, 61802175), the Natural Science Foundation of Jiangsu Province (BK20171458), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.
Ethical statement Articles do not rely on clinical trials.
Human and animal participants All submitted manuscripts containing research which does not involve human participants and/or animal experimentation.

References

An Z, Zhou D (2019) Deep reinforcement learning for quantum gate control. EPL (Europhysics Letters) 126(6):60002

Bukov M (2018) Reinforcement learning for autonomous preparation of floquet-engineered states: Inverting the quantum kapitza oscillator. Physical Review B 98(22):224305

Bukov M, Day AG, Sels D, Weinberg P, Polkovnikov A, Mehta P (2018) Reinforcement learning in different phases of quantum control. Physical Review X 8(3):031086

Cárdenas-López FA, Lamata L, Retamal JC, Solano E (2018) Multiqubit and multilevel quantum reinforcement learning with quantum technologies. PloS one 13(7):e0200455

Chakrabarti R, Rabitz H (2007) Quantum control landscapes. International Reviews in Physical Chemistry 26(4):671–735

Chen C, Dong D, Li HX, Chu J, Tarn TJ (2013) Fidelity-based probabilistic q-learning for control of quantum systems. IEEE transactions on neural networks and learning systems 25(5):920–933

Chu S (2002) Cold atoms and quantum control. Nature 416(6877):206–210

Chunlin C, Frank J, Daoyi D (2012) Hybrid control of uncertain quantum systems via fuzzy estimation and quantum reinforcement learning. In: Proceedings of the 31st Chinese Control Conference, IEEE, pp 7177–7182

D'Alessandro D, Dahleh M (2001) Optimal control of two-level quantum systems. IEEE Transactions on Automatic Control 46(6):866–876

Dong D, Chen C, Li H, Tarn TJ (2008) Quantum reinforcement learning. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 38(5):1207–1220

Fang W, Pang L, Yi W (2020) Survey on the application of deep reinforcement learning in image processing. Journal on Artificial Intelligence 2(1):39–58

Fösel T, Tighineanu P, Weiss T, Marquardt F (2018) Reinforcement learning with neural networks for quantum feedback. Physical Review X 8(3):031084

Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, Liu T, Wang X, Wang G, Cai J, et al. (2018) Recent advances in convolutional neural networks. Pattern Recognition 77:354–377

He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778

Hu B, Zhao H, Yang Y, Zhou B, Raj ANJ (2020) Multiple faces tracking using feature fusion and neural network in video. INTELLIGENT AUTOMATION AND SOFT COMPUTING 26(6):1549–1560

Li Z, Zhang J, Zhang K, Li Z (2018) Visual tracking with weighted adaptive local sparse appearance model via spatio-temporal context learning. IEEE Transactions on Image Processing 27(9):4478–4489

Ma H, Chen C (2020) Several developments in learning control of quantum systems. In: 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, pp 4165–4172

Michael MH, Silveri M, Brierley R, Albert VV, Salmilehto J, Jiang L, Girvin SM (2016) New class of quantum error-correcting codes for a bosonic mode. Physical Review X 6(3):031006

Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G, et al. (2015) Human-level control through deep reinforcement learning. nature 518(7540):529–533

Niu MY, Boixo S, Smelyanskiy VN, Neven H (2019) Universal quantum control through deep reinforcement learning. npj Quantum Information 5(1):1–8

Palittapongarnpim P, Wittek P, Sanders BC (2017) Robustness of learning-assisted adaptive quantum-enhanced metrology in the presence of noise. In: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, pp 294–299

Rabitz H, de Vivie-Riedle R, Motzkus M, Kompa K (2000) Whither the future of controlling quantum phenomena? Science 288(5467):824–828

Roslund J, Rabitz H (2009) Gradient algorithm applied to laboratory quantum control. Physical Review A 79(5):053417

Singh SP, Sutton RS (1996) Reinforcement learning with replacing eligibility traces. Machine learning 22(1):123–158

Sutton RS, Barto AG (2018) Reinforcement learning: An introduction. MIT press

Sutton RS, McAllester DA, Singh SP, Mansour Y (2000) Policy gradient methods for reinforcement learning with function approximation. In: Advances in neural information processing systems, pp 1057–1063

Tsubouchi M, Momose T (2008) Rovibrational wavepacket manipulation using shaped midinfrared femtosecond pulses toward quantum computation: Optimization of pulse shape by a genetic algorithm. Physical Review A 77(5):052326

Vedaie SS, Palittapongarnpim P, Sanders BC (2018) Reinforcement learning for quantum metrology via quantum control. In: 2018 IEEE Photonics Society Summer Topical Meeting Series (SUM), IEEE, pp 163–164

Vrajitoarea A, Huang Z, Groszkowski P, Koch J, Houck AA (2020) Quantum control of an oscillator using a stimulated josephson nonlinearity. Nature Physics 16(2):211–217

Watkins CJ, Dayan P (1992) Q-learning. Machine learning 8(3-4):279–292

Xu F, Zhang X, Xin Z, Yang A (2019) Investigation on the chinese text sentiment analysis based on convolutional neural networks in deep learning. Comput Mater Contin 58(3):697–709

Yu S, Albarrán-Arriagada F, Retamal JC, Wang YT, Liu W, Ke ZJ, Meng Y, Li ZP, Tang JS, Solano E, et al. (2019) Reconstruction of a photonic qubit state with reinforcement learning. Advanced Quantum Technologies 2(7-8):1800074

Zhang XM, Wei Z, Asad R, Yang XC, Wang X (2019) When does reinforcement learning stand out in quantum control? a comparative study on state preparation. npj Quantum Information 5(1):1–7

Zhang Y, Wang Z (2020) Hybrid malware detection approach with feedback-directed machine learning. Information Sciences 63(139103):1–139103