Single-qubit quantum agent in quantum reinforcement learning

variational quantum circuit for proximal policy optimization

Jen-Yueh Hsiao 蕭仁岳(NTU)

2021.8.24

Outline

Motivation

Introduction to the Quantum reinforcement learning

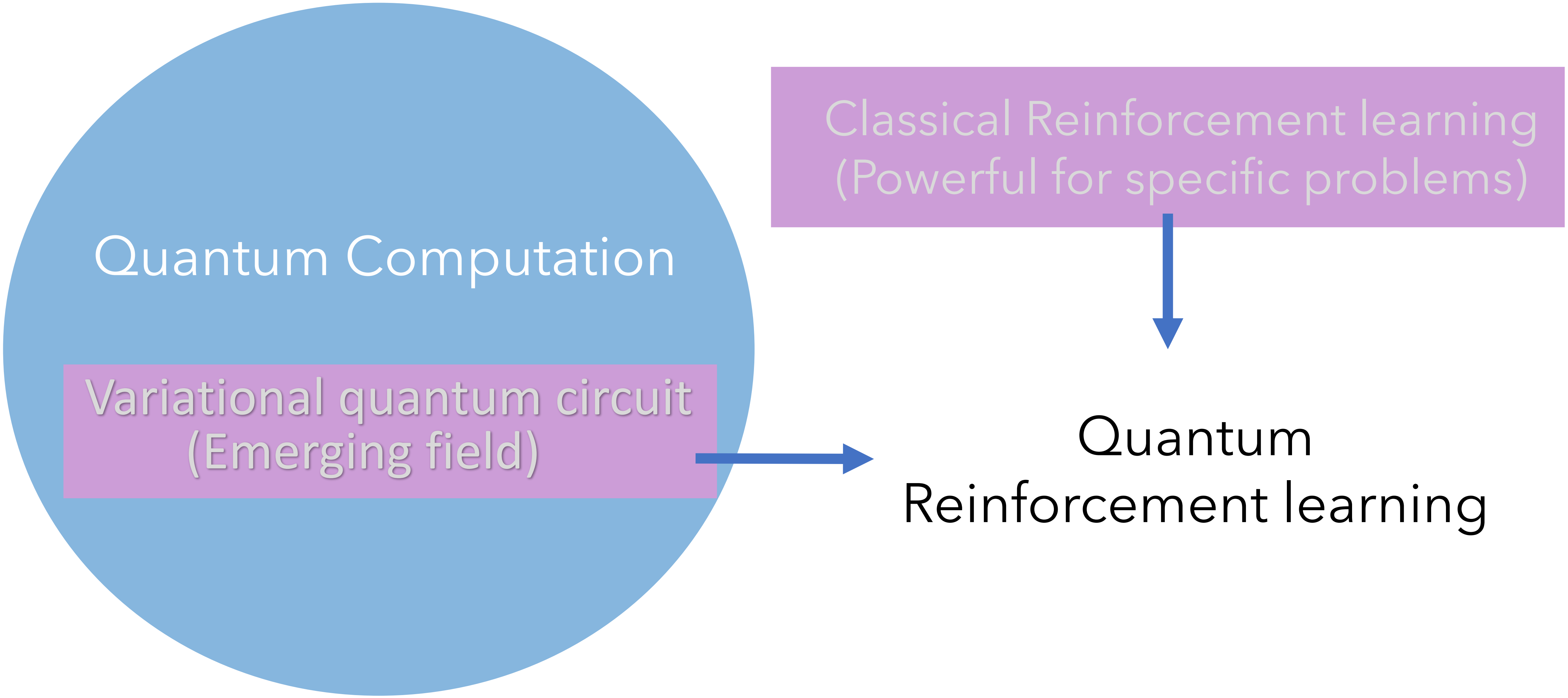
Related works

Research results

Conclusion and open issues

Motivation

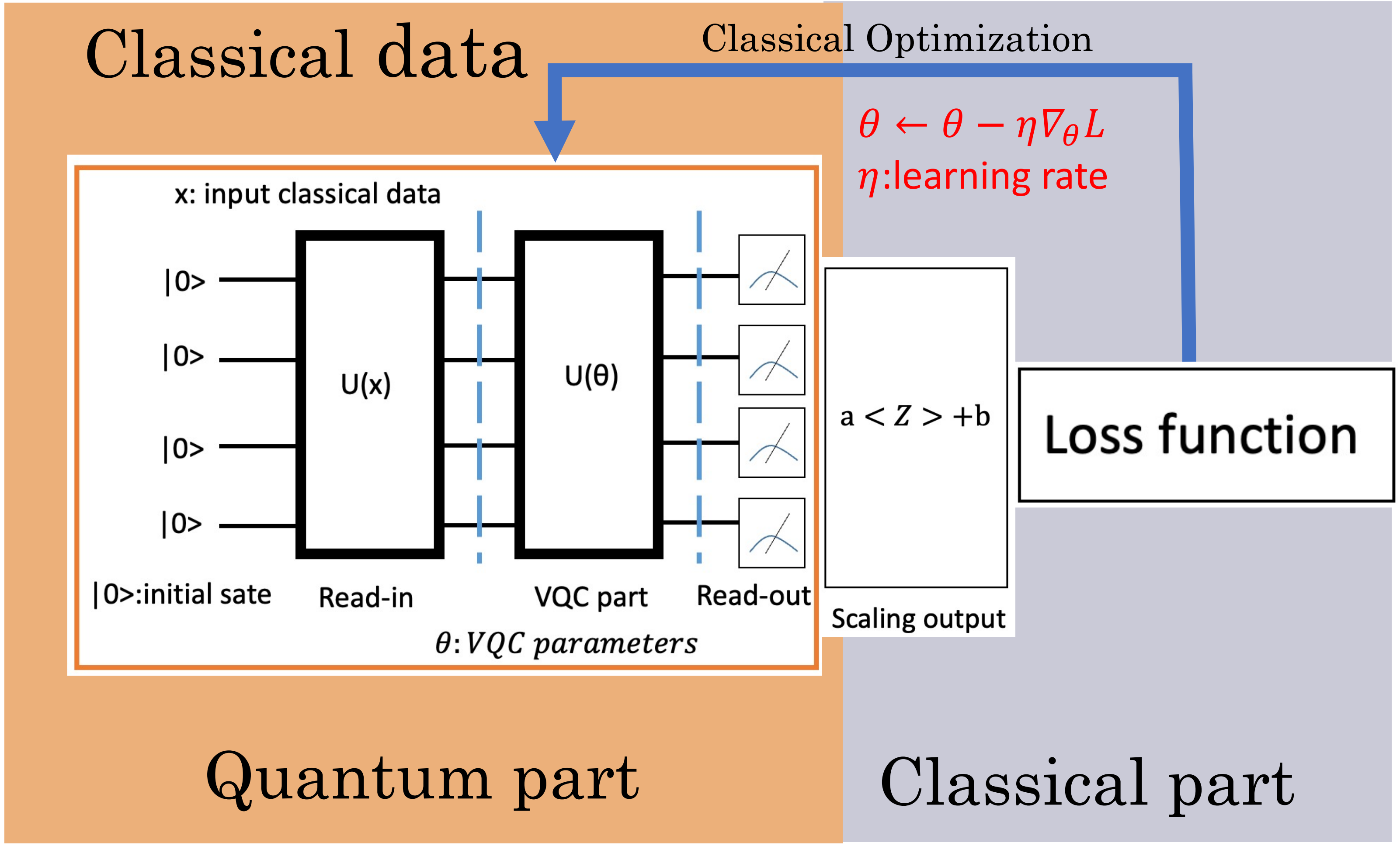
Why investigate Quantum Reinforcement Learning (QRL) ?



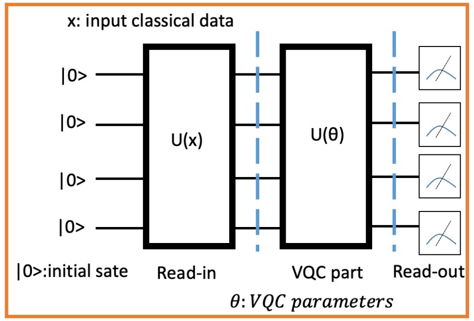
Introduction to the Variational quantum circuit (VQC)

Elements of hybrid quantum machine learning

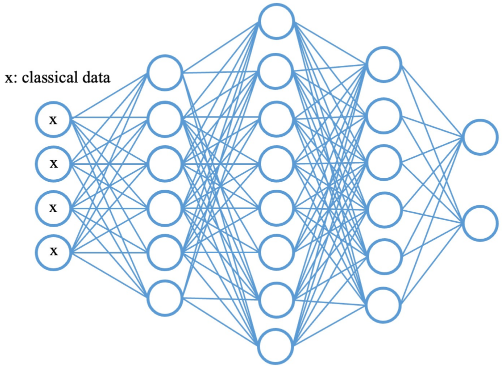
Data   Model   Loss function



Quantum model : VQC

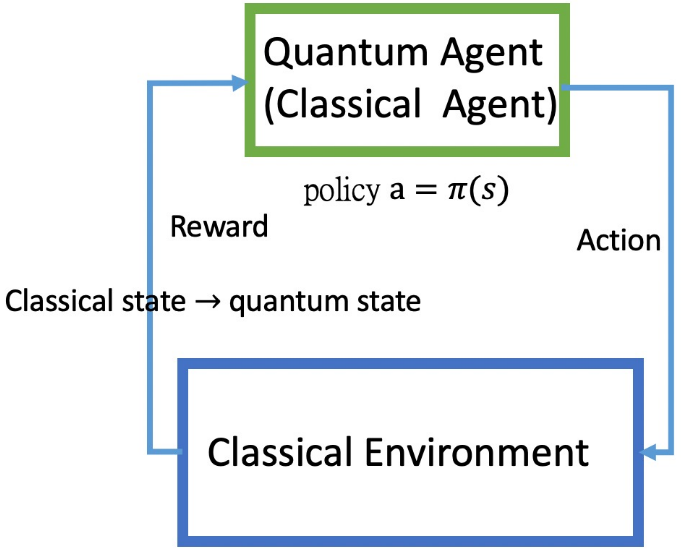


Classical model : NN



Elements of hybrid quantum reinforcement learning

Data



$P_a = P[s_{t+1} = s' \mid s_t = s, a_t = a]$

Model

Classical agent : $\mathrm{f_c}(x, \theta)$
neural network(NN)

Quantum agent : $\mathrm{f_Q}(x, \theta)$
Variational quantum circuit(VQC)

$\theta_a \leftarrow \theta_a + \eta \nabla_{\theta_a} \overline{R_{\theta_a}}$ η:learning rate

$\theta_c \leftarrow \theta_c - \eta \nabla_{\theta_c} L_{\theta_c}$ η:learning rate

Loss function

$$\overline{R_\theta} = \sum_\tau R(\tau) p_\theta(\tau) = \mathbb{E}_{\tau \sim p_\theta(\tau)}[R(\tau)]$$

Train it!!!

Trajectory:$\tau = \{s_1, a_1, s_2, a_2, \ldots, s_T, a_T\}$ *in one episode*

$$\nabla \overline{R_\theta} \approx \frac{1}{N} \sum_{n=1}^{N} \sum_{t=1}^{T} R(\tau^n) \nabla \log p_\theta(a_t^n | s_t^n)$$

t=1~T # of steps in one episode
n=1~N # of episode

Policy gradient : $L = -p_\theta(a_t^n|s_t^n) * R(\tau^n)$

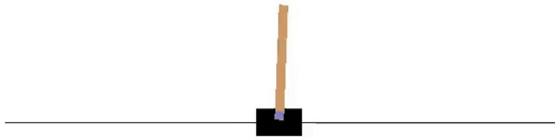
Actor-Critic: Actor L $= - p_\theta(a_t^n|s_t^n) * A(s)$. Critic L $= \frac{1}{N}\left(R(\tau^n) - V_{\theta'}(s)\right)^2$, $A(s) = R(\tau^n) - V_{\theta'}(s)$.

PPO actor L= $-min(clip * A(s)$, ratio* $A(s)$)
PPO critic L $= \frac{1}{N}\left(R(\tau^n) - V_{\theta'}(s)\right)^2 (MSE)$

Clip =(ratio=$\frac{p_\theta(a_t^n|s_t^n)}{p_{\theta old}(a_t^n|s_t^n)}$, $1+\epsilon, 1-\epsilon$) Ratio=$\frac{p_\theta(a_t^n|s_t^n)}{p_{\theta old}(a_t^n|s_t^n)}$

Related works

OpenAI gym environment: CartPole-v0 (v1)



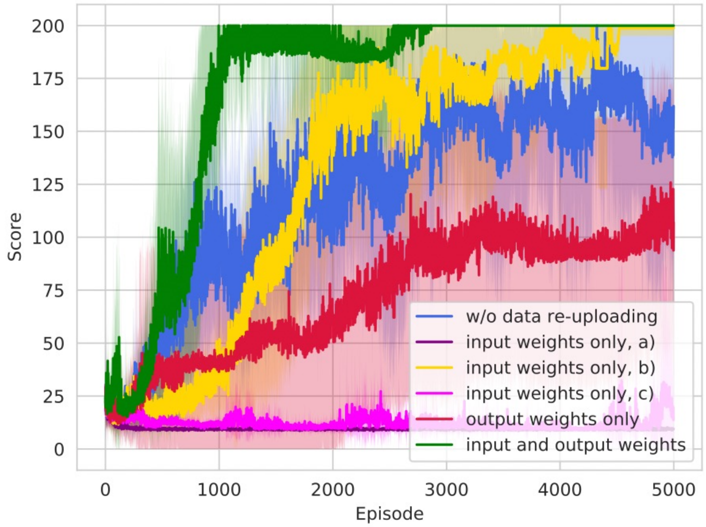
https://gym.openai.com/

- Agent：Cart control
- State：[Cart Position x , Cart Velocity v, Pole Angle θ, Pole Angular Velocity ω]
- Action：left or right
- Reward：Reward is 1 for every step taken, including the termination step

Episode Termination:

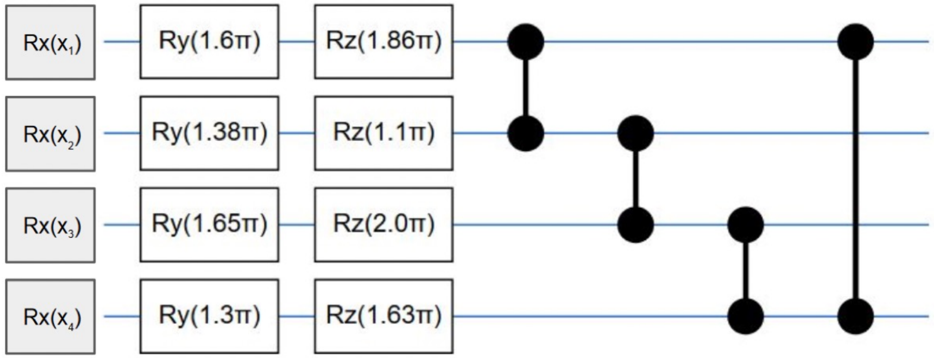
1. Pole Angle is more than 12 degrees(0.209).
2. Cart Position is more than 2.4 or -2.4
3. episode length is greater **than 200 or 500.**
4. Solved Requirements: Considered solved when the average return is greater than or equal to 195.0(475.0) over 100 consecutive trials.

Related works : VQC method for RL's problems

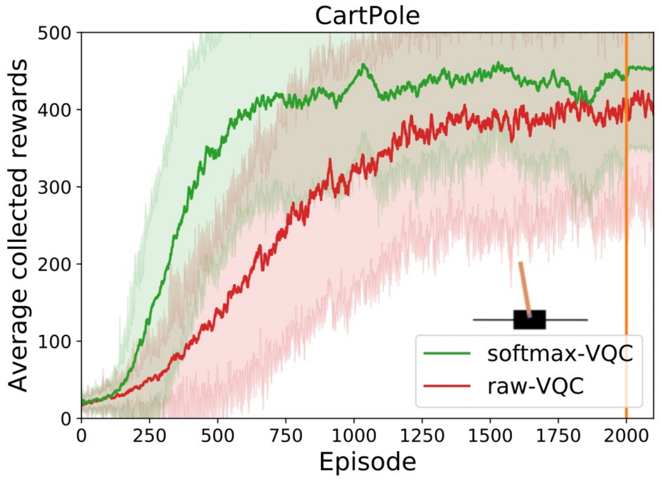


(a) average scores with varying trainable weights

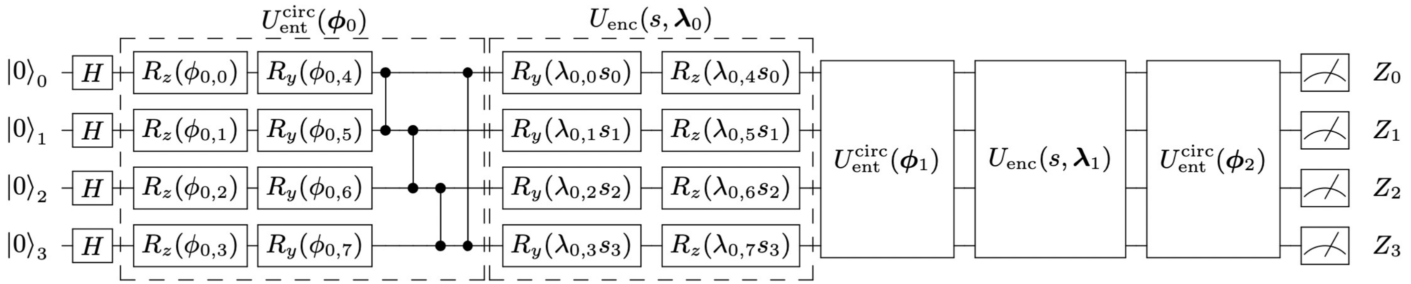
A. Skolik, S. Jerbi, and V. Dunjko
,Quantum agents in the Gym: a variational quantum algorithm for deep Q-learning



(Dated: March 30, 2021) arXiv:2103.15084v1



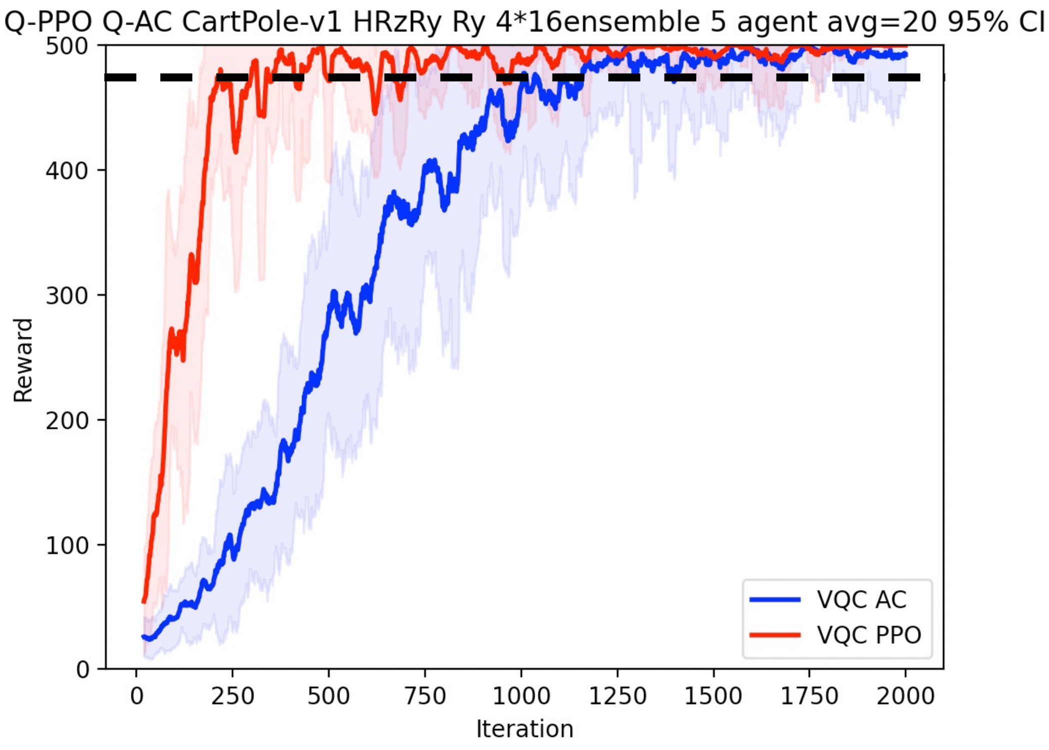
S. Jerbi, C. Gyurik, S. Marshall, H. J. Briegel, and V. Dunjko, Variational quantum policies for reinforcement learning

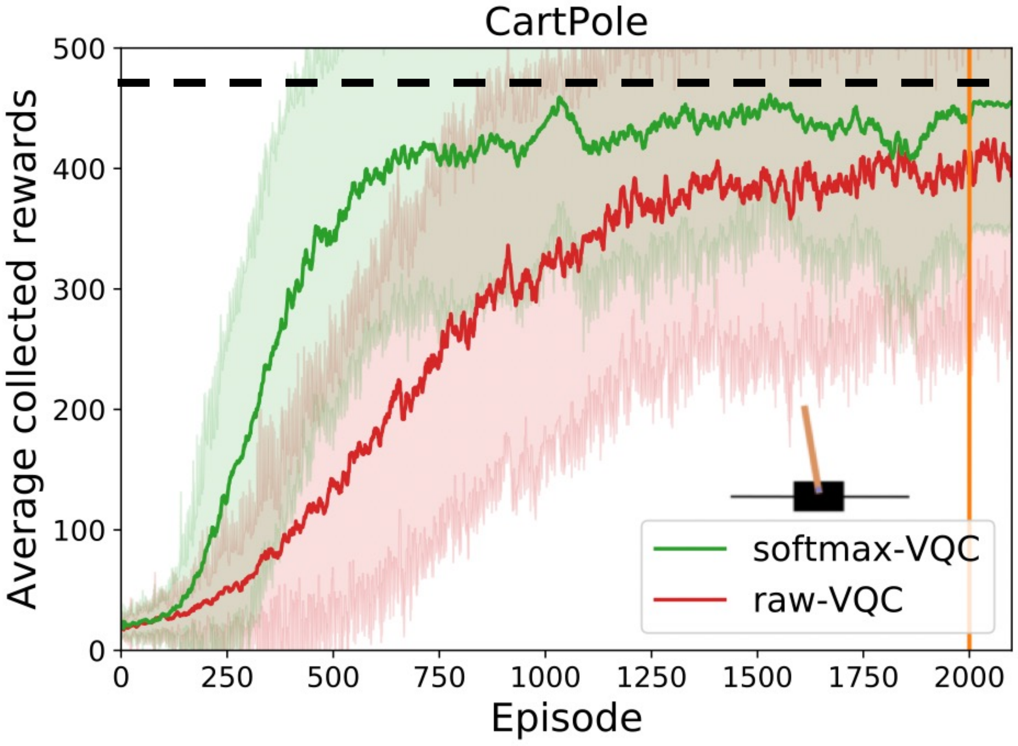


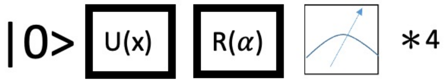
(Dated: March 10, 2021) arXiv:2103.05577v1

Research results

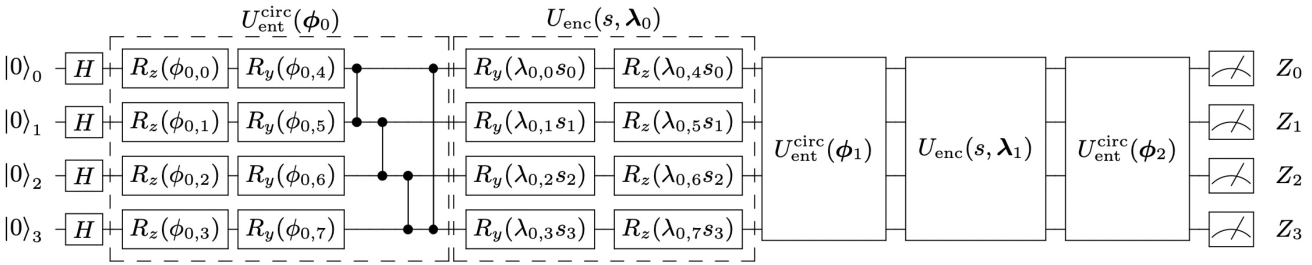
Implementation on classical computer (Training process)





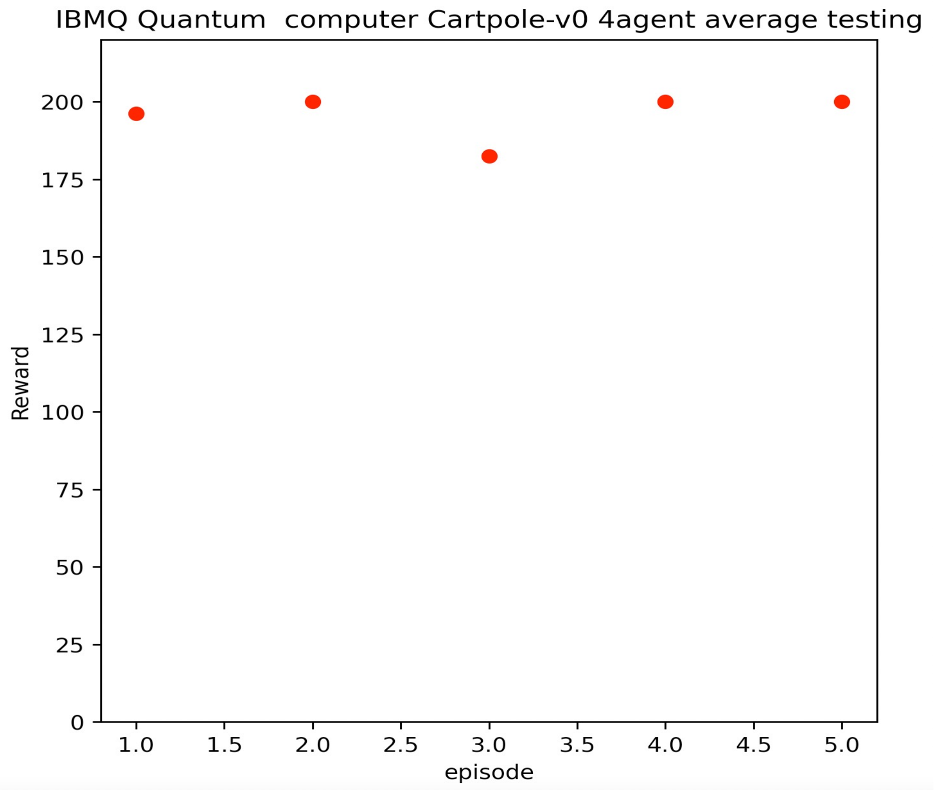


Single-qubit circuit



Implementation on IBM Quantum Computer (Testing Process)

CartPole-v0



First successful implementation on the real quantum device for complex RL's problem in gym with the VQC model

Conclusion and open issues

1. Single-qubit quantum agent has good performance on specific problems.

2. How would we use the entanglement's power with the VQC methods in quantum reinforcement learning？Or quantum machine learning？

3. How would we implement more complex problems by the VQC model？

Thank you for your attention