TEXAS
The University of Texas at Austin

WHAT STARTS HERE CHANGES THE WORLD

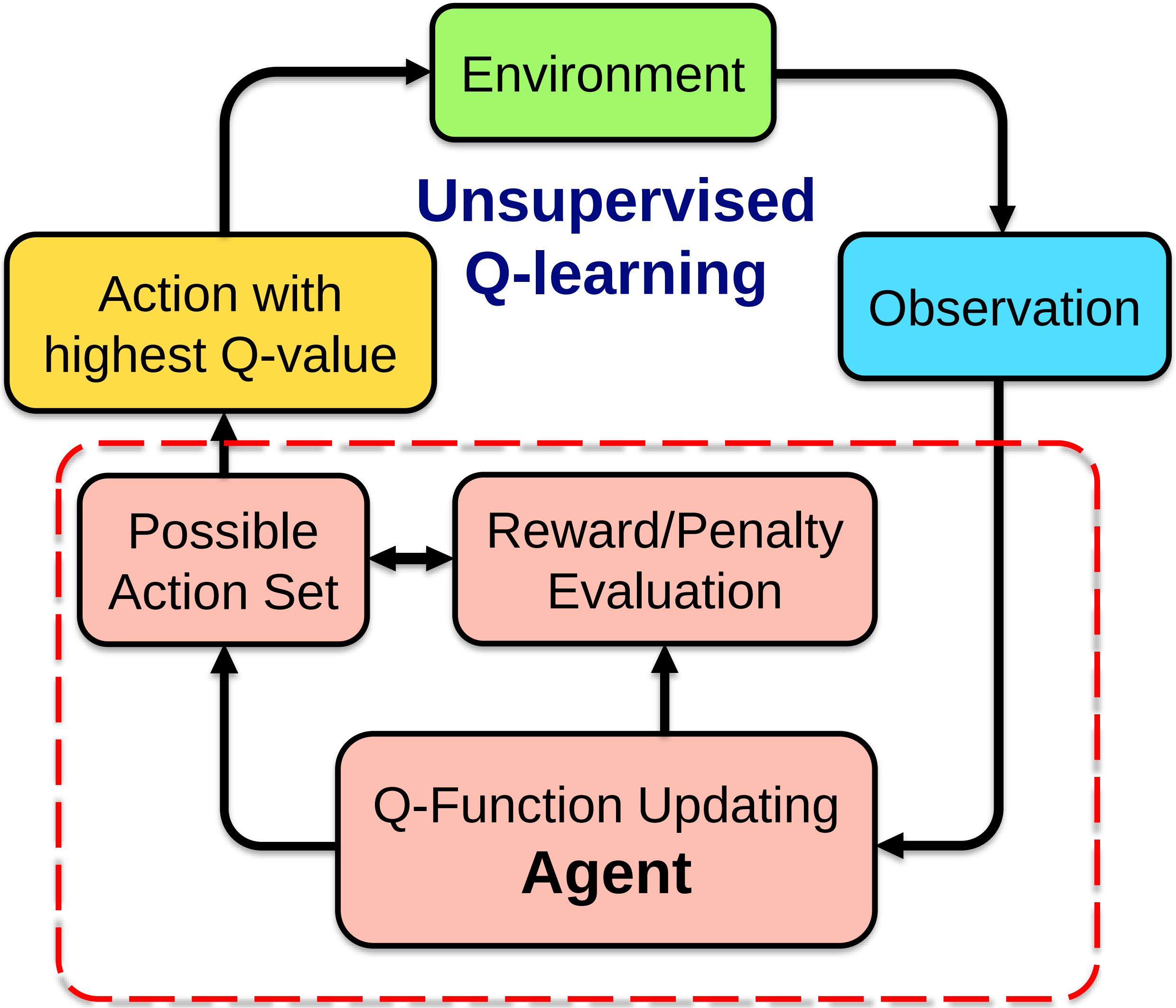
Grover Method for Quantum Reinforcement Learning

Zhanzhi Jiang and Qiang Gao

Outline

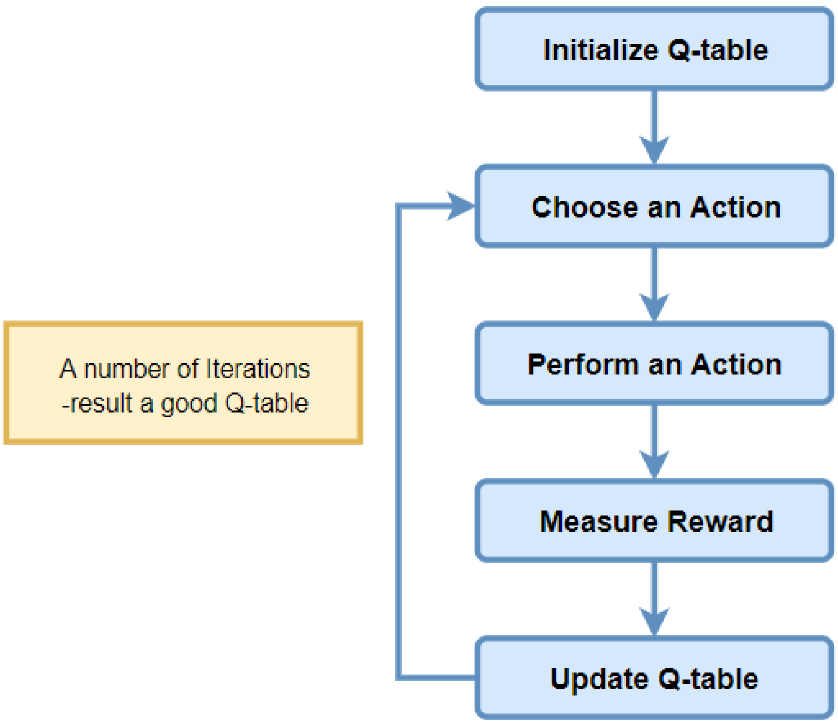
- Introduction to reinforcement learning (RL)
- Quantum RL based on grove algorithm
- Tests with simple environments
- Summary and future directions

Reinforcement learning



Q-learning

$Q(s,a)$: expectation of total reward for taking action a at state s



Update Q-table:

$$Q(s,a) \leftarrow (1-\alpha)Q(s,a) + \alpha\left(R(s,a) + \gamma \max_{a'} Q(s',a')\right)$$

α: learning rate
$R(s,a)$: reward for taking action a at state s
γ: discount factor for convergence

Method to implement Q-learning agents:

- Direct searching the Q-table
- Decision Tree and Deep Neural Networks (Alpha Go)
- Quantum agent: Grover amplitude amplification

Quantum Q-Learning Agent

Actions of every state s is encoded in a quantum state $|a_s\rangle = \sum_i c_i |a_i\rangle$

Take an action: the Grover iteration

Measure $|a_s\rangle$
Get $|a_i\rangle$

→

Take a_i
Reach s'

→

Update $|a_s\rangle$
$\rightarrow U^L_{Grover}|a_s\rangle$

→

Update $Q(s, a)$

U^L_{Grover}: Grover oracle for performing the Grover quantum circuits

reward

$$L = \mathrm{int}\left[\min\left\{k\left(r + \boxed{V(s)}\right), \frac{\pi}{4\theta} - \frac{1}{2}\right\}\right]$$: Grover length

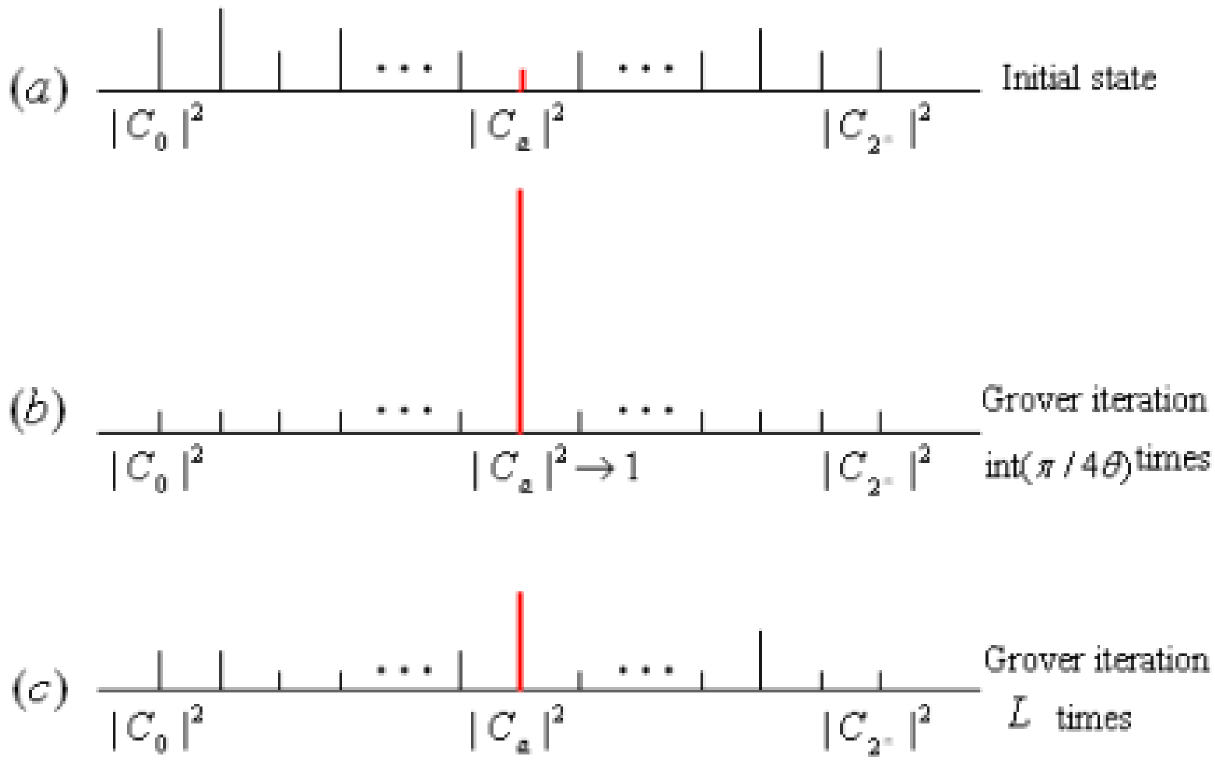
a parameter related to the state space dimension

a tuning parameter

$$V(s) = \max_a Q(s, a)$$

Ganger, M. and Hu, W. *International Journal of Intelligence Science*, 9, 1-22 (2019)

How Grover algorithm works



- Grover Searching Algorithm: amplify the amplitude of target state to 1
- Grover Q-learning agent: amplify the amplitude of the action according to $Q(s', a)$

arXiv:0810.3828

Quantum Q-learning algorithm

Algorithm 3 Quantum Q-learning (QQRL)

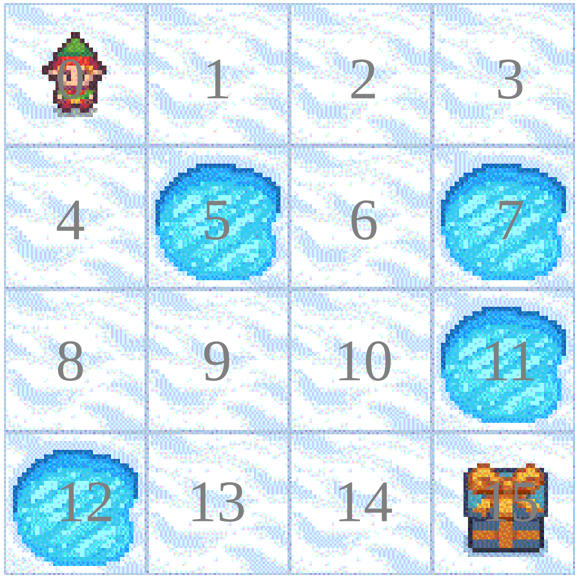
for all episodes **do**
  for all $s \in S$ **do**
    Observe a from $|a_s\rangle$
    Take action a, observe next state s' and reward r
    $V(s') = \max_{a'} Q(s', a')$
    $L = \left\lfloor \min\left\{k(r + V(s')), \frac{\pi}{4\theta} - \frac{1}{2}\right\}\right\rfloor$
    $|a_s\rangle \leftarrow \hat{U}_g^L |a_s\rangle$ ▷ Grover's Algorithm
    $Q(s,a) \leftarrow Q(s,a) + \alpha(r + \gamma V(s') - Q(s,a))$
  end for
end for

Modifications:
The update of the V-values requires a search for maximal values in Q-table. To avoid that, we directly update $V(s)$ using $Q(s,a)$ with $|a\rangle$ being the measured results when given state s. The quantum mechanics ensure us with high possibilities to get $\max_a Q(s,a)$.

Thus, we achieve a full-quantum Q-learning algorithm which might overperform its classical counterpart when the state space is large enough.

Ganger, M. and Hu, W. *International Journal of Intelligence Science*, 9, 1-22 (2019)

Frozen Lake



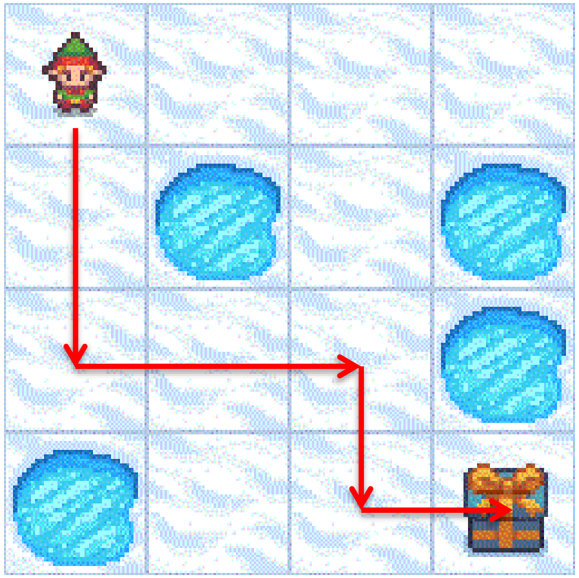
State: site location 0-15 (global env)
Action: move up, down, left, right

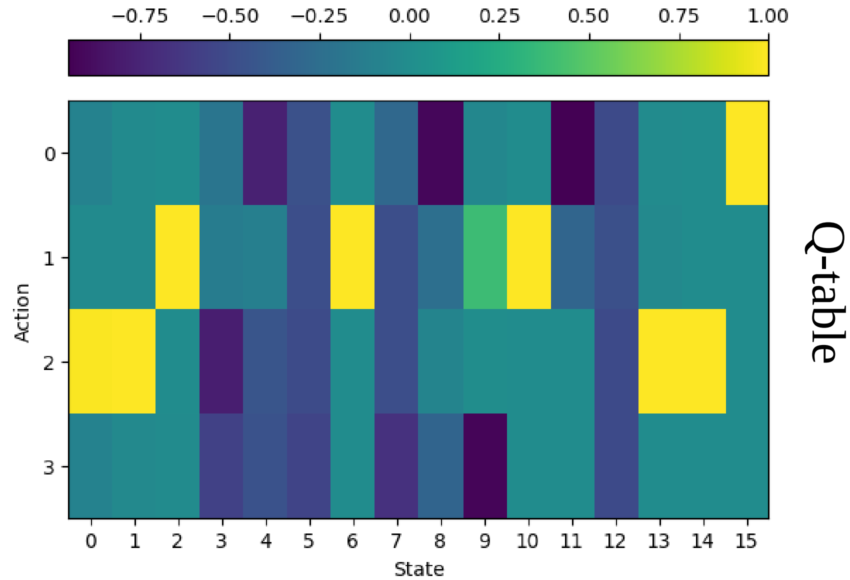
Next State	Reward
Goal	99, Done
Same site	-10
Ice	-1
Hole	0, Done

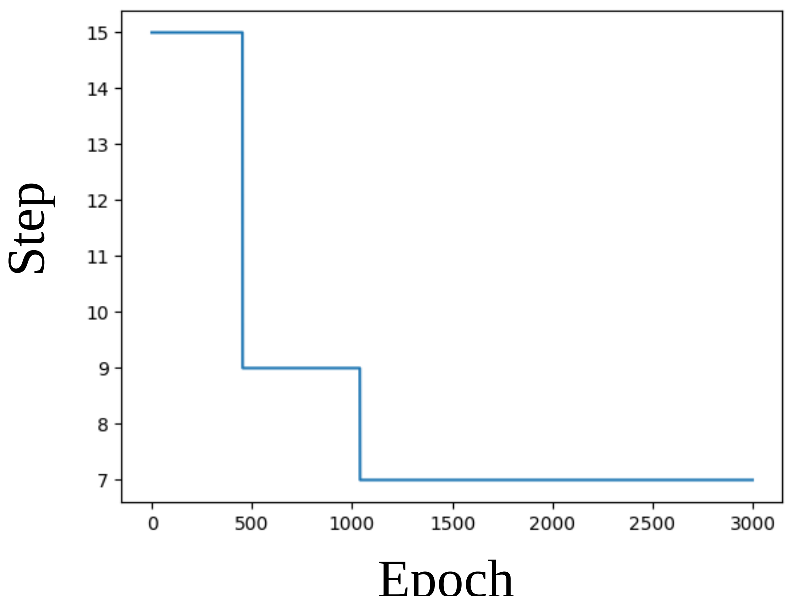
Task: finding the optimal path from the start point (site 0) to the goal (site 15) without falling down in the hole

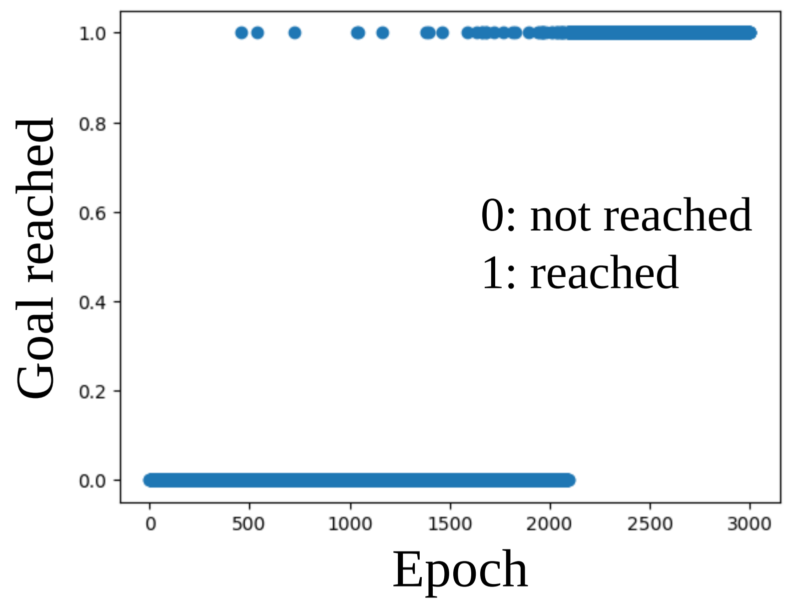
https://www.gymlibrary.dev/environments/toy_text/frozen_lake/

Result

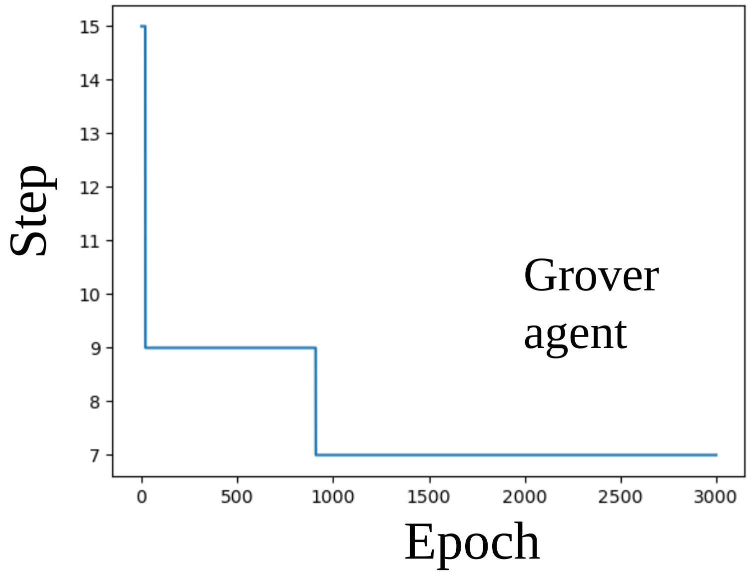


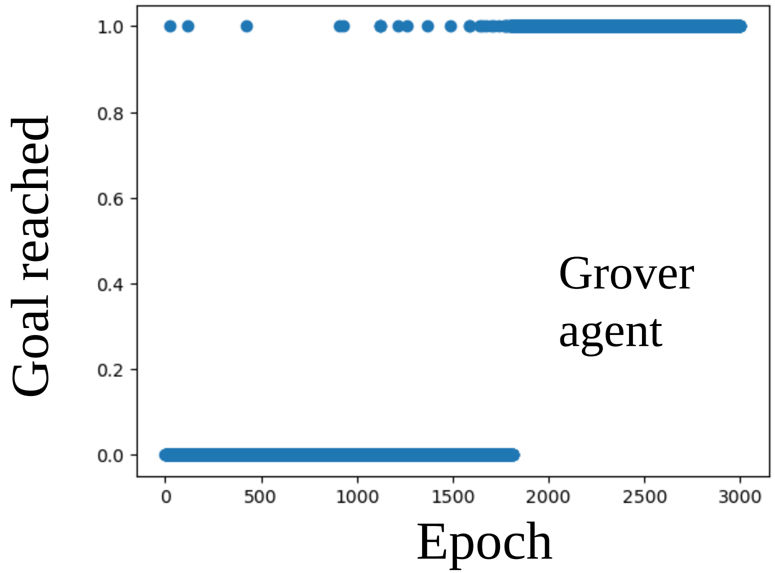

−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
Action
State
0
1
2
3
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Q-table

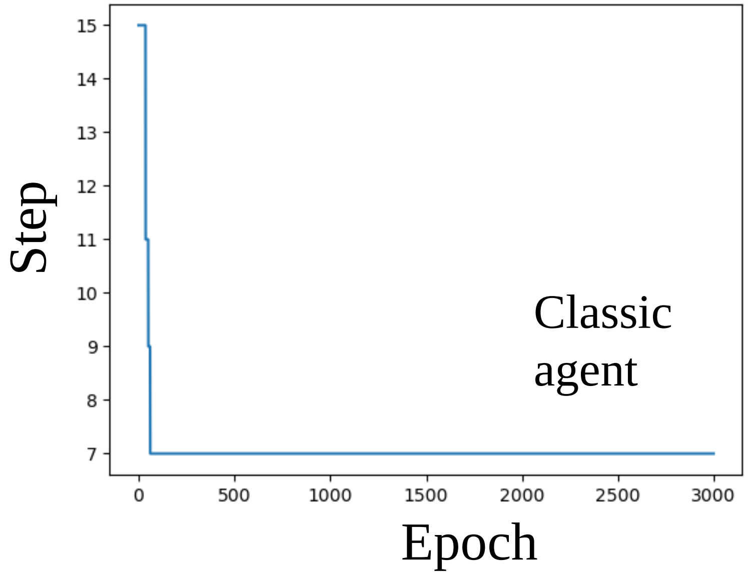

Step
15
14
13
12
11
10
9
8
7
0
500
1000
1500
2000
2500
3000
Epoch

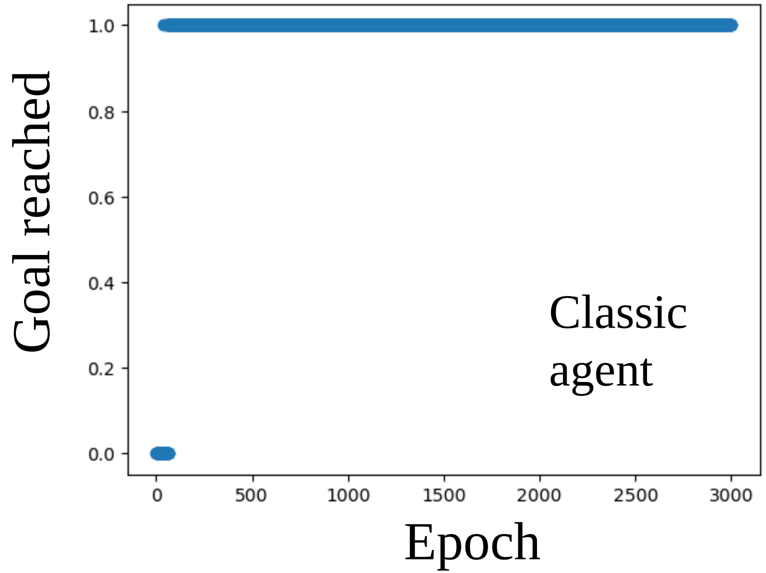

Goal reached
1.0
0.8
0.6
0.4
0.2
0.0
0: not reached
1: reached
0
500
1000
1500
2000
2500
3000
Epoch

Compare with classic agent

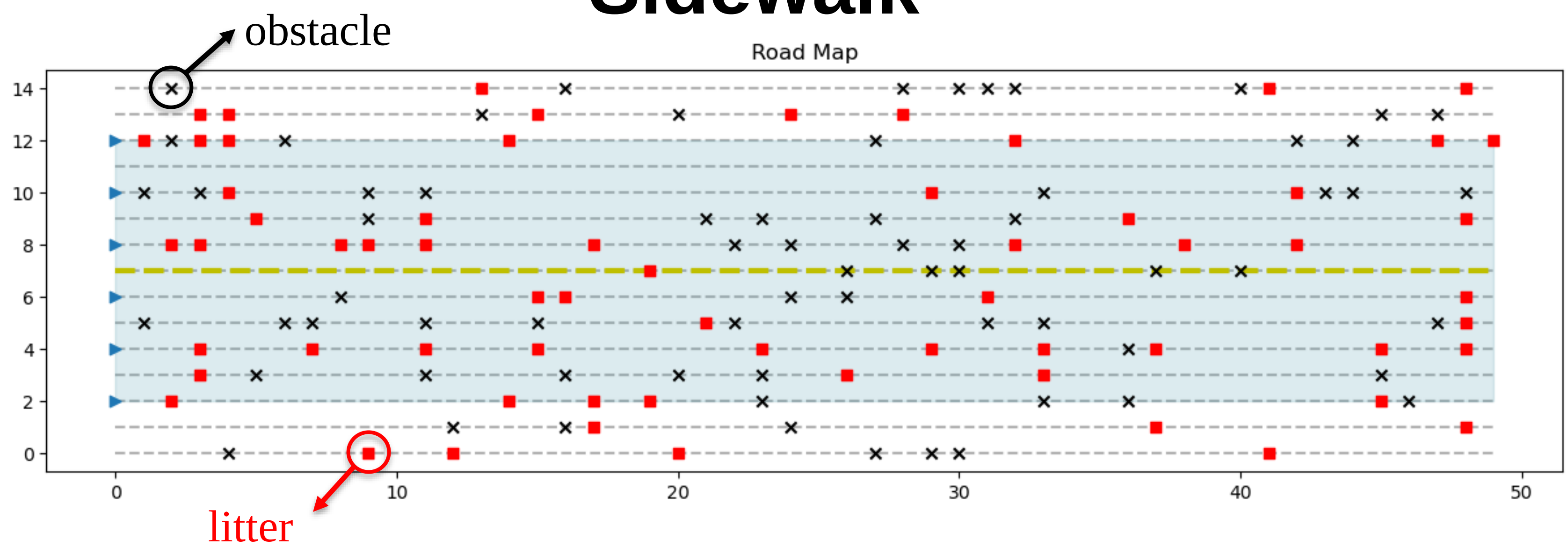








Sidewalk



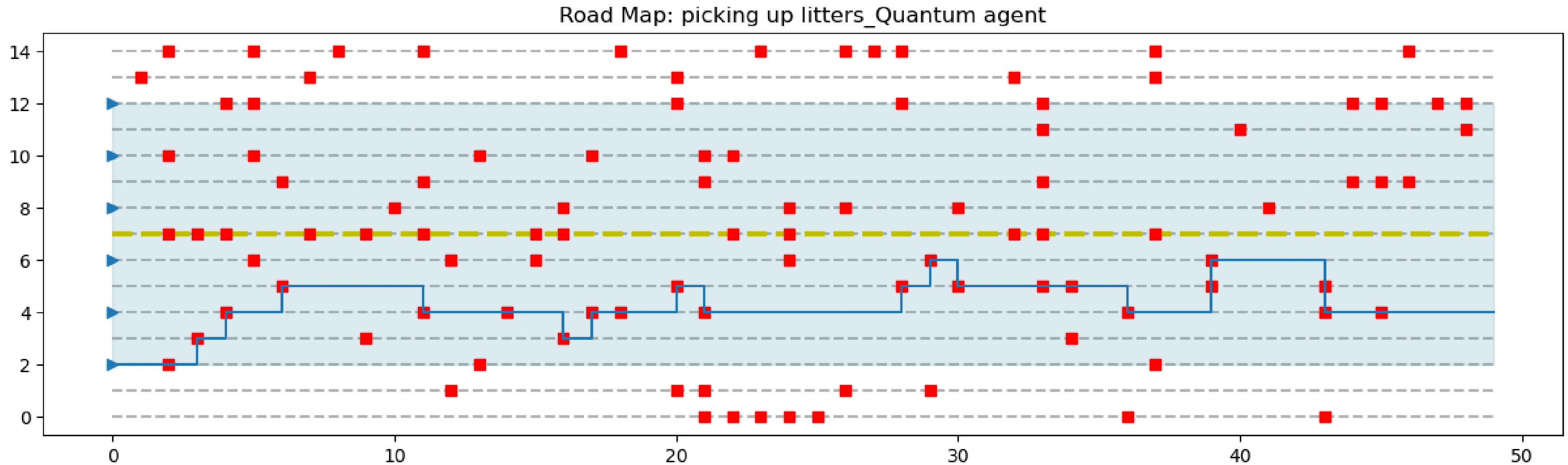
Task: picking up litters and avoiding obstacles when going through the sidewalk

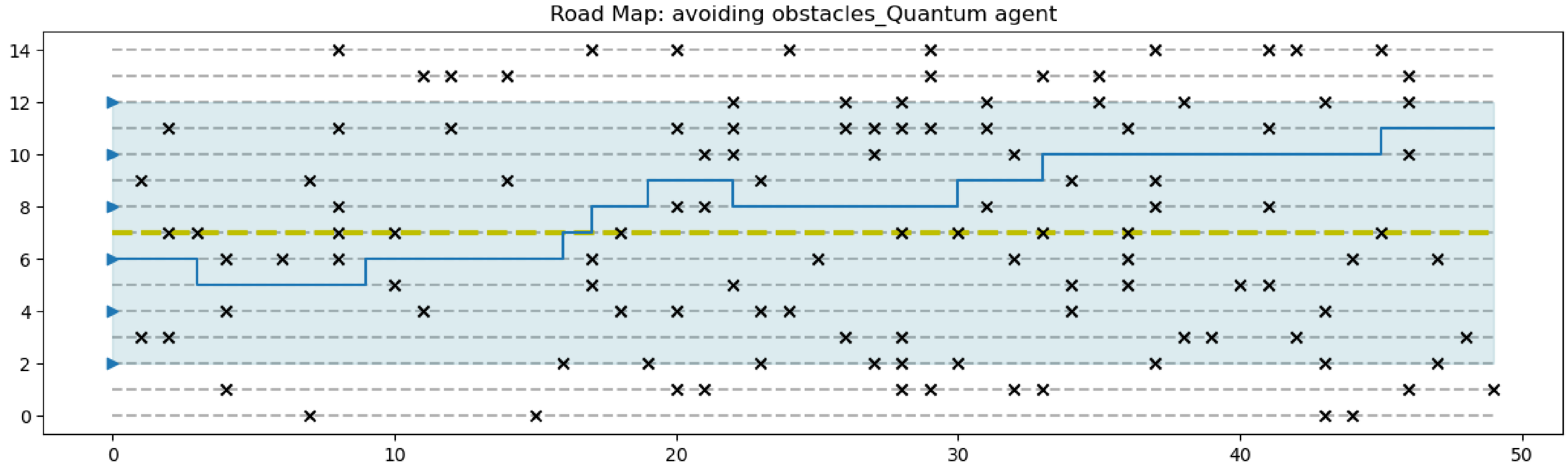
State: four neighbors having litters (obstacles) or not, labeled by 0-15 (local env)
Action: move up, down, left, right

Next position	Reward
Litter (moving →, ↑, ↓)	15
No litter (moving →)	5
No litter (moving ↑, ↓)	0
Moving ←	-5

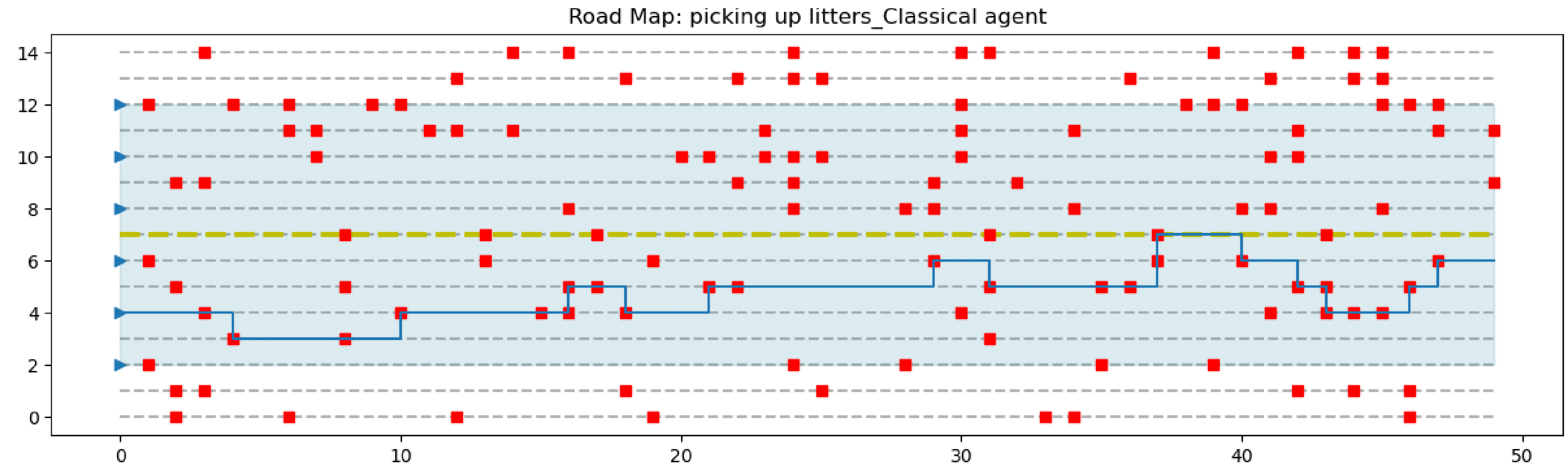
Next position	Reward
Obstacle	-3
No obstacle (moving →)	8
No obstacle (moving ↑, ↓)	3
No obstacle (moving ←)	1

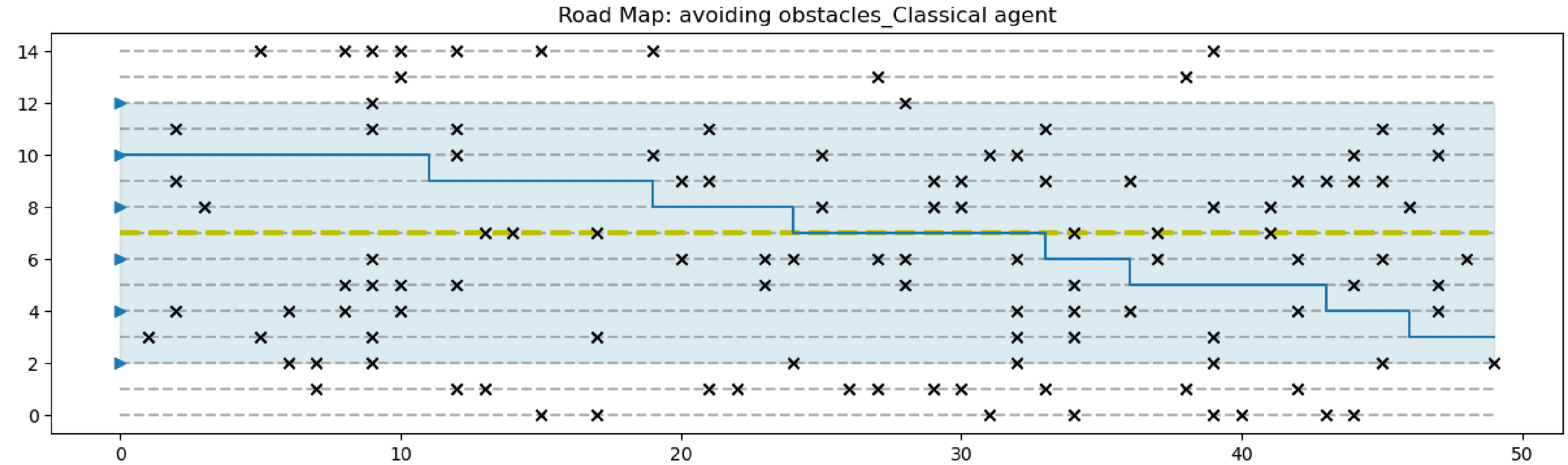
Quantum Agent





Classical Agent





Summary and future directions

- We have demonstrated the feasibility of using quantum algorithm to do reinforcement learning for both global (Frozen-lake) and local (Sidewalk) environments.

- For systems with small state space, the quantum algorithm cannot outperform its classical counterpart. More powerful and efficient quantum algorithm is needed to achieve the quantum advantage.

- The next step will be implementing this algorithm on a real quantum computer.