Quantum algorithms for reinforcement learning with a generative model[1]

Daochen Wang (University of Maryland)
QISE-NET June Meeting 2021

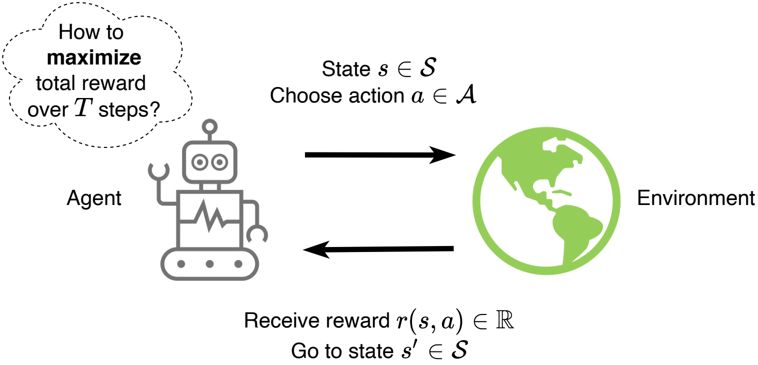
Aarthi Sundaram   Robin Kothari   Ashish Kapoor   Martin Roetteler

(Microsoft Research)

[1] Full paper to appear at ICML 2021.

Reinforcement Learning (RL)

Main question of RL: how should an agent interact with its environment to maximize its total reward?



Note: s' is random and follows some probability distribution $p(\cdot \mid s, a)$. In the *generative model* we assume quantum sample access to these distributions for any (s, a) of our choice, i.e., access to the oracle

$$\mathcal{O} : |s\rangle |a\rangle |0\rangle |0\rangle \mapsto |s\rangle |a\rangle \sum_{s' \in \mathcal{S}} \sqrt{p(s' \mid s, a)}\, |s'\rangle |\psi_{s'}\rangle .$$

RL has many applications in engineering, finance, gaming, natural language processing, robotics, and so on

	Playing Go	Trading stocks
States ($\mathcal{S}$)	Board positions	Market positions
Actions ($\mathcal{A}$)	Valid moves	Buy or sell a stock
Rewards	win: $+1$ draw: 0 lose: -1	Net profit

Image credit: Nature

Quantum algorithms for RL

Summary of quantum speedups

Notation: $S = |\mathcal{S}|$, $A = |\mathcal{A}|$, $T =$ number of steps, $\epsilon =$ error;
q^*, v^*, $\pi^* =$ optimal (Q-value function, value function, policy).

Goal: output an ϵ-accurate estimate of	Classical sample complexity	Quantum sample complexity	
	Upper and lower bound	Upper bound	Lower bound
q^*	$\frac{SAT^3}{\epsilon^2}$	$\frac{SAT^{1.5}}{\epsilon}$	$\frac{SAT^{1.5}}{\epsilon}$
v^*, π^*	$\frac{SAT^3}{\epsilon^2}$	$\min\{\frac{SAT^{1.5}}{\epsilon}, \frac{S\sqrt{A}T^3}{\epsilon}\}$	$\frac{S\sqrt{A}T^{1.5}}{\epsilon}$

Quantum speedups from applying quantum mean estimation and maximum finding to value iteration

- Quantum mean estimation (Montanaro, 2015) estimates $\mathbb{E}[X]$ to error ϵ using $\tilde{O}(\sqrt{\mathrm{Var}[X]}/\epsilon)$ quantum samples of X.
- Quantum maximum finding (Dürr and Høyer, 1996) finds the maximum of a size-n list using $\tilde{O}(\sqrt{n})$ quantum queries to it.
- They can speed up, e.g., the value iteration algorithm for v^*:

  $v \leftarrow \mathbf{0} \in \mathbb{R}^A$, $\gamma \leftarrow 1 - 1/T$
  for $\ell = 1, 2, \ldots, L = \tilde{O}(T)$ **do**
      for $s \in \mathcal{S}$ **do**
          $v(s) \leftarrow \max_{a \in \mathcal{A}}\{r(s,a) + \gamma \mathbb{E}[v(s') \mid s' \sim p(\cdot|s,a)]\}$
      end
  end

- But this value iteration turns out to be highly sub-optimal so we speed up a modern variant (Sidford et. al., 2018) instead which gives us the (near-)optimal bounds in the summary.

Thank you for your attention!

Open problems

Thank you for your attention! Here are some of our open problems:

1. Can we close the gap between the upper and lower bounds for computing v^* and π^*?
2. Can we quantize model-based classical algorithms? As a first step, can we get a tight bound for quantum distribution learning in ℓ_1-norm?
3. Can we circumvent our quantum lower bounds? We note that exponential speed-ups exist for finding $v^*(s_0)$ and $\pi^*(s_0)$, for a fixed $s_0 \in \mathcal{S}$, in convoluted cases based on the glued-trees construction (Childs et. al., 2002).