Information–Theoretic Foundations of Quantum Reinforcement Learning (RL) Algorithms

Bridging Quantum Information Measures and RL Performance

Oluwaseyi Giwa

African Institute for Mathematical Sciences
South Africa
Supervised by Prof. Joan Soler

April 19, 2025

Motivation

- The rapid development of quantum computing opens new frontiers in reinforcement learning (RL).
- Key Question: How can information–theoretic tools guide the design and analysis of quantum RL algorithms?
- Goals: Derive performance bounds, analyze convergence, and leverage quantum measures (e.g., QFI, mutual information) for efficient policy design and speedup.

Outline

Brief description of RL

- Reinforcement learning problems involve learning what to do—how to map situations to actions—so as to maximize a numerical reward signal.
- The main idea behind RL is for an agent to learn good policy through interaction with a dynamic environment by trial and error. The agent in each state performs an action and based on the reward obtained, the agent decides to either keep performing the specific action or explore.
- It is then appropriate to say RL is a sequential decision making process. This could be interpreted in mathematical form as the Markov Decision Processes (MDPs).

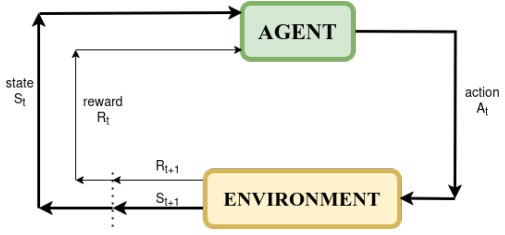

Figure: The agent-environment interaction in reinforcement learning

Brief description of RL contd.

A Markov Decision Process is a tuple $(\mathcal{S},\ \mathcal{A},\ p,\ r,\ \gamma)$, where

- $\mathcal{S}$ denote the set of states in an environment,
- $\mathcal{A}$ a set of actions,
- $p:\ \mathcal{S} \times \mathcal{R} \times \mathcal{S} \times \mathcal{A} \longrightarrow\ [0,1]$ is the dynamic function.

$$p(s', r|s, a) = Pr S_{t+1} = s',\ R_{t+1} = r|S_t = s, A_t = a,\ \text{where } \mathcal{R} \subset \mathbb{R}.$$

- $r : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \longrightarrow \mathcal{R}$ is a reward function.

$$r(s, a, s') = \mathbb{E}[R_{t+1}|S_t = s,\ A_t = a,\ S_{t+1} = s'].$$

- γ is a discount factor with $\gamma \in [0, 1]$.

Policy Gradient in RL

- Policy, π, is a mapping from each state, $s \in \mathcal{S}$ and action, $a \in \mathcal{A}(s)$, to the probability $\pi(a|s)$ of taking action a when in state s.
- The goal in RL is to find a policy (a strategy for choosing actions) that maximizes the expected cumulative reward (also known as the expected return).
- Policy gradient methods maximize the expected total reward by repeatedly estimating the gradient.

$$\nabla_\theta J(\theta) = \mathbb{E}_{\pi\theta} \left[\nabla_\theta \log \pi\theta(a|s) Q^\pi(s, a) \right]$$

Quantum RL Overview

- Quantum RL: Agent and/or environment leverage quantum information processing.
- Advantages: Potential quadratic speedup, parallel exploration via superposition.
- Challenges: Noise (NISQ era), decoherence, and state preparation.

Quantum Information Measures in RL

- Quantum Mutual Information: is a measure of the correlation between subsystems of a quantum state.

$$I(X;Y)_\rho = S(\rho_X) + S(\rho_Y) - S(\rho_{XY})$$

where $S(\rho)$ is the von Neumann entropy and ρ is the density matrix.

$$S(\rho) = -\text{tr}(\rho \log \rho)$$

- In quantum reinforcement learning, it makes sense to say that Quantum Mutual Information quantifies correlations between agent and environment.

Quantum Fisher Information (QFI)

- QFI is used to qualify the utility of an input state. In RL, this could be useful for preconditioning policy gradient updates.
- For a parametrized state,

$$F_Q(\theta) = \text{tr}[\rho(\theta)L(\theta)^2]$$

$L(\theta)$ is called the symmetric logarithmic derivative defined via

$$\frac{\partial \rho(\theta)}{\partial \theta} = \frac{1}{2}\left(\rho(\theta)L(\theta) + L(\theta)\rho(\theta)\right)$$

Directed Information in Quantum Settings

- Measures causal flow from the agent's actions X^n to the environment's responses Y^n.
- This is useful for optimizing feedback in quantum RL where quantum correlations (e.g. entanglement) play a role.
- We can then define directed information in mathematical terms as;

$$I(X^n \longrightarrow Y^n) = \sum_{i=1}^{n} I\left(X^i; Y_i | Y^{i-1}\right)$$

Quantum Natural Policy Gradients

- Classical update:

$$\Delta\theta = \eta \nabla_\theta J(\theta)$$

- Quantum Natural Policy Gradient incorporates QFI as a metric tensor

$$\Delta\theta = \eta F_Q^{-1}(\theta) \nabla_\theta J(\theta)$$

- This helps to adjust learning rate based on the quantum state.

Mathematical Framework for Quantum RL

- To analyze RL policies in quantum settings, we consider the parameterized quantum circuits, state evolution, and cost function in variational quantum algorithms.
- Parameterized quantum circuits: $U(\theta) = U_L(\theta_L) \dots U_1(\theta_1)$
- State evolution: $\rho' = U(\theta)\rho U^\dagger(\theta)$
- Cost function in variational quantum algorithms:

$$C(\theta) = \text{tr}[\rho^2] - \text{tr}[(U(\theta)\rho U^\dagger(\theta))^2]$$

Feedback and Optimization via Directed Information

- An agent updates its policy based on observed rewards.
- Directed information can capture how much information flows from past actions to future rewards, given by

$$I(\theta; r) = \sum_i I(\theta_i; r_i | r^{i-1})$$

- The goal is to maximize this flow to ensure efficient learning in quantum environments.

Conclusion

- Information–theoretic measures provide deep insights into the learning dynamics of quantum RL algorithms.
- Incorporating quantum measures (QFI, directed information) into policy updates leads to improved convergence and efficiency under certain conditions.

Challenges and Questions

- How robust are these methods under realistic noise and decoherence?
- What is the role of entanglement in enhancing or impeding the information flow in feedback loops?
- When does quantum preconditioning (via QFI) outperform classical methods?

Future Research Directions

- Algorithm Design: Develop hybrid quantum–classical RL architectures with provable performance bounds.
- Experimental Validation: Implement on near-term quantum devices to test convergence and robustness.
- Extended Information Measures: Explore other quantum information quantities (e.g., quantum Rényi entropies) in RL contexts.

References

Oluwaseyi Giwa Oluwaseyi Giwa notes on Reinforcement Learning, 2025.

Oluwaseyi Giwa
`https://github.com/OluwaseyiWater/quantumRL`

Richard Sutton and Andrew Barto
Reinforcement Learning: An Introduction, MIT Press, 1992.

Arnu Proteius
AIMS 2024-2025 Course on Reinforcement Learning, AIMS, 2024.

References

Hamann, A., Dunjko, V. & Wölk, S. Quantum-Accessible Reinforcement Learning beyond Strictly Epochal Environments, Quantum Mach. Intell. 3, 22 (2021), https://doi.org/10.1007/s42484-021-00049-7

Mark Wilde
Quantum Information Theory, Cambridge Press, 2013.

Dong Daoyi and Chen Chunlin and Li Hanxiong and Tarn Tzyh-Jong
Quantum Reinforcement Learning, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics).
2008, 38, 5, 1207-1220, doi=10.1109/TSMCB.2008.925743.

References

Emma Brunskill CS234 Reinforcement Learning, Stanford University Spring 2024, Stanford University, 2024.

Andre Sequeira and Luis Paulo Santos and Luis Soares Barbosa
On Quantum Natural Policy Gradients, arXiv preprint, 2024,
https://doi.org/10.48550/arXiv.2401.08307

The End