Parameterized Quantum Policies for Reinforcement Learning

Shiven Tripathi
Department of Electrical Engineering
IIT Kanpur
shiven@iitk.ac.in

Somya Lohani
Department of Computer Science
IIT Kanpur
somyaloh@iitk.ac.in

1 Background

1.1 Machine Learning

Machine learning aims to solve the problem of finding patterns based on large sets of data inputs which might be labelled or through other information signals which are used to train a model to solve a task like classification, generation etc. Supervised learning uses data samples for which ground truth label values are available, whereas, in unsupervised learning, unlabelled data is leveraged. Recently a lot of progress in ML has been achieved by leveraging deep neural networks, which can be trained to perform well on complex tasks by learning informative latent representations over input samples. A growing problem of such networks is the vast amount of computing which is required in their training and inference.

1.2 Reinforcement Learning

A standard formulation for RL is based on Markov decision processes with a defined state space, action space defined for each state, state transition probabilities corresponding to the action taken on a state, a reward function corresponding to taking action on a state and an objective function which signals whether the desired goal of the actions on the MDP has been achieved. RL algorithms used to solve these MDPs can be categorized into value-based or policy-based, on the basis of which function the optimization or learning is done. The goal of any RL algorithm is to learn a policy which is mapping to actions for each state, which correlates most strongly to the completion of the attainment of the MDP. Deep neural network-based models have shown tremendous progress recently, helping achieve superhuman capability in games like chess and Go, which have state spaces exceeding the number of atoms in the universe.

1.3 Parameterised Quantum Circuits

Near-term quantum computers consist of few (less than a hundred) qubits which behave noisily, resulting in large gate circuits being infeasible. To solve this hardware problem, hybrid quantum computing is evolving, which leverages some classical processing steps along with quantum computation. Recent attempts to combine quantum computing principles with machine learning have resulted in the development of parameterized quantum circuits. These consist of quantum algorithms which depend on certain free parameters, which can be tuned to optimize specific loss or cost metrics associated with the output function that the circuit computes. The key intuition behind this approach is that by designing certain sub-parts of a problem classically, the need for quantum resources is drastically reduced and can be leveraged solely for the classically intractable part of the problem.

Conference on Neural Information Processing Systems (NeurIPS).

2 Contributions

The use of policies based on PQCs is shown to solve classical RL tasks. By building environments, the solution to which requires solving the discrete logarithm problem, the quantum advantage is demonstrated in RL. Previous work has explored PQCs as value function approximators rather than direct policy functions. Also, this work solves the MDP for a classical state, showing more applicability to practical problems. The extreme amount of computational resources which classical RL algorithm leverage to solve even standard benchmark tasks makes it one of the most promising fields to show a quantum advantage. This work shows a quantum RL formulation using PQCs, which is able to handle standard RL benchmarks well, along with theoretically proven and empirically verified tasks which would be very tough for classical agents.

3 Results

3.1 Benchmark Environments

On cart pole and frozen lake standard benchmark environments provided by OpenAI, solutions are provided by the QPC agent. We find that satisfactory performance can be achieved even with a low number of qubits, further validating the hypothesis that near-term quantum approaches can also be leveraged for practical solutions. This performance is achieved with interaction with the classical environment, and quantum processing is only used for policy determination.

3.2 Quantum Advantage

Discrete logarithm problem is classically intractable but has efficient solutions using quantum approaches. By designing environments which are modifications of standard benchmarks but have supervised learning based on solutions of DLP, it can be theoretically proven to be very tough for classical systems. It is demonstrated that using a quantum parameterized circuit; such problems can yield solutions for such an environment. For example, for the cliff walk environment adding a temporal structure, an agent, if it assigns the correct temporal value, moves to the intended fixed state and receives a reward. Otherwise, the episode terminates.

4 Limitations

The results are shown by implementing a quantum processing unit on a simulator. It is highly likely that some of the learning of any such model would be hindered in a real quantum system which would be noisy. The discrete logarithm problem embedded environments largely were just theoretical constructs useful to demonstrate some quantum advantage, but it would remain unclear if this progress directly leads to the practical application of quantum RL.

5 Discussion

Somewhat due to the lack of informativeness of the training reward signals, or the sheer large state space, we find that classical RL algorithms tend to be very resource intensive. A step towards the demonstration of quantum advantage in RL through certain constructed task environments could potentially reduce the resource demands while enabling solutions to even more complex problems, which can be formulated as an MDP.

6 References

Jerbi, S., Gyurik, C., Marshall, S. C., Briegel, H. J., Dunjko, V. (2021). Parametrized Quantum Policies for Reinforcement Learning, Advances in Neural Information Processing Systems