PERFORMANCE ANALYSIS OF A QUANTUM-CLASSICAL HYBRID REINFORCEMENT LEARNING APPROACH

Evan Mitchell
Trinity College Dublin

Pabitra Mitra
Indian Institute of Technology, Kharagpur

Biswajit Basu
Trinity College Dublin

ABSTRACT

Quantum Machine Learning (QML) is a nascent field of technology that is yet to be fully explored. While previous QML implementations have demonstrated performance efficiency gains over classical benchmarks, it has not been studied in detail whether shallow unentangled quantum circuits can provide the same benefits to reinforcement learning algorithms. Towards this goal, we present a shallow Deep Q-Network (DQN) hybrid quantum-classical Variational Quantum Circuit (VQC) model in the Cartpole-v0 environment that provides an increase in training stability and average reward for any given training run with a simpler unentangled quantum circuit than what is proposed in prior literature.

1 INTRODUCTION AND BACKGROUND

Quantum reinforcement learning (QRL) is an emerging field that combines the principles of quantum computing and reinforcement learning (RL) algorithms. QML methods show the capacity for higher learning accuracies (Liu et al., 2019), training run time reduction through efficient data sampling (Phillipson, 2020) and resilience to the barren plateau problem (Abbas et al., 2021). Previous studies have shown how QRL has theoretical superiority over classical RL for learning problems with huge state spaces in unknown probabilistic environments (Dong et al., 2008; Dunjko et al., 2017). Many previous implementations utilise VQCs, which are hybrid algorithms that use a classical optimizer to train a parameterized quantum circuit in an iterative manner, thereby approximating a neural network (Phillipson, 2020).It has been demonstrated that hybrid VQC-based Proximal Policy Opitimization models can achieve reward increases in the CartPole-V0 environment of over 80% more than classical models (Kwak et al., 2021), with other LSTM models (Chen, 2022) and VQC DQN models (Hsiao et al., 2022) displaying increased performance with a coupled parameter reduction. However, despite some QRL implementations in the literature displaying high performance, no consensus on the best performing architecture is to be reached.

2 METHODOLOGY

The QRL architecture proposed in this paper uses a classical DQN algorithm with an epsilon-greedy policy as a base structure with an integrated quantum circuit block within the classical neural network framework at the data input level (A.2). Modelled using the Qiskit library, the hybrid component is a simple VQC consisting of a layer of Hadamard gates and R_y gates with a trainable parameter, θ, followed by a series of measurement gates as seen in Fig. 3 (A.2). This is achieved by simulating the quantum circuit 100 times for each forward pass, averaging the results to achieve the measured output, with the parameter θ acting as the 'weight' in the VQC. The output of these measurement gates / simulations generates a length-4 bit string, the output of which is then passed into a fully connected feed-forward neural network, modelled in PyTorch, which interprets the results of the measurement. The opposing classical model just omits the quantum circuit and acts as a classical DQN with identical hyperparameters. Hyperparameter tuning was completed by taking a parameter (learning rate, batch size, γ, ϵ, etc.) and iteratively cycling over a series of values for said parameter until the highest performance was achieved. The model was trained on the CartPole-v0 environment (Brockman et al., 2016), which is often used as a benchmark in QRL implementations due to it's low state-space size (4 qubits, length-4 state space, length-2 action space), helping qubit limitations. During forward passes for each episode of the CartPole-v0 environment, the model received a length-4 state tensor from the environment, which is passed into the quantum layer through to the classical layer, which chooses an action for the environment. CartPole-V0 then returns a reward which acts as the maximising value for the DQN.

3 Results

To determine the performance difference between the Hybrid and the Classical model, both models were trained 15 times each utilising the same hyperparamers, with the only difference being the addition of the VQC to the hybrid model.

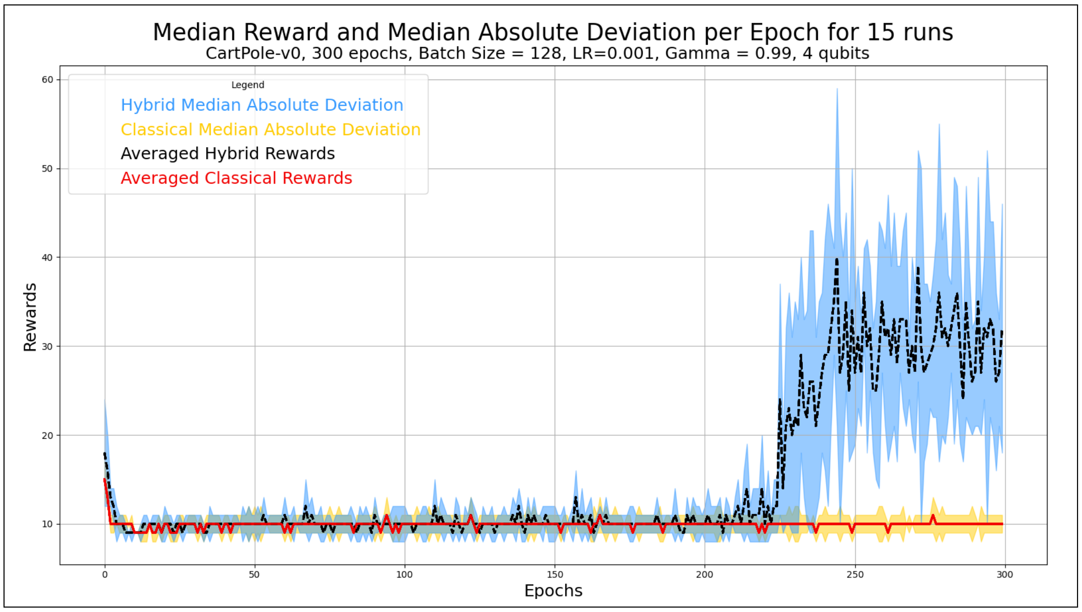

Figure 1: Median Reward and Median Absolute Deviation for 15 runs of both the hybrid and classical models, with the MAD showing the limited deviation from the baseline with the classical model compared to the hybrid model.

The classical models quickly learned the optimal policy in a small subset (20%) of training runs, with the remaining 80% remaining at the baseline (10 reward). In comparison, the hybrid model was able to learn a partial policy above the baseline in 60% of training runs (A.2), showcased by the diverging median line on the graph in Fig. 1. As observed, the hybrid model displays increased stability in comparison to the classical model, with a lower divergence between training runs as the majority of curves reached above the baseline (A.2). These results were achieved with fewer trainable parameters than previous QRL implementations (Kwak et al., 2021) with a shallower circuit design than other Cartpole-v0 methods (Chen, 2022), while also maintaining equivalent performance. These curves suggest that the classical models are more sensitive to the initial conditions and the epsilon greedy policy, with the hybrid models being more robust in comparison (possibly due to the quantum noise inherent in VQCs) or due to learning a better policy than the classical models on average.

4 Conclusion

In conclusion, the hybrid models outperform the classical models on average with a higher median reward and a more stable average reward than the classical models, indicating a more robust tolerance to the initial conditions and the exploration-exploitation trade-off, possibly due to quantum noise. This was achieved using a shallow quantum circuit with minimal modifications to standard classical DQN algorithms, with possible applications in environments which have variable starting conditions. This architecture could also be scaled up to a larger number of qubits in a wider variety of environments by using tensor trained networks in future works (Ballarin et al., 2023). These experimental observations can provide useful guidance in designing VQC RL models in future studies.

URM Statement

The authors acknowledge that at least one key author of this work meets the URM criteria of ICLR 2024 Tiny Papers Track.

References

Amira Abbas, David Sutter, Christa Zoufal, Aurelien Lucchi, Alessio Figalli, and Stefan Woerner. The power of quantum neural networks. *Nature Computational Science*, 1(6):403–409, June 2021. ISSN 2662-8457. doi: 10.1038/s43588-021-00084-1. URL `http://dx.doi.org/10.1038/s43588-021-00084-1`.

Marco Ballarin, Stefano Mangini, Simone Montangero, Chiara Macchiavello, and Riccardo Mengoni. Entanglement entropy production in quantum neural networks. *Quantum*, 7:1023, May 2023. ISSN 2521-327X. doi: 10.22331/q-2023-05-31-1023. URL `http://dx.doi.org/10.22331/q-2023-05-31-1023`.

Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. Openai gym, 2016.

Samuel Yen-Chi Chen. Quantum deep recurrent reinforcement learning, 2022.

Daoyi Dong, Chunlin Chen, Hanxiong Li, and Tzyh-Jong Tarn. Quantum reinforcement learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 38(5):1207–1220, October 2008. ISSN 1083-4419. doi: 10.1109/tsmcb.2008.925743. URL `http://dx.doi.org/10.1109/TSMCB.2008.925743`.

Vedran Dunjko, Jacob M. Taylor, and Hans J. Briegel. Advances in quantum reinforcement learning. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, October 2017. doi: 10.1109/smc.2017.8122616. URL `http://dx.doi.org/10.1109/SMC.2017.8122616`.

Jen-Yueh Hsiao, Yuxuan Du, Wei-Yin Chiang, Min-Hsiu Hsieh, and Hsi-Sheng Goan. Unentangled quantum reinforcement learning agents in the OpenAI Gym. 3 2022.

IBM. Explore gates and circuits with the quantum composer. Technical report, IBM Quantum Learning, 2023a. URL `https://learning.quantum.ibm.com/tutorial/explore-gates-and-circuits-with-the-quantum-composer#non-unitary-operators-and-modifiers`.

IBM. Qiskit is the open-source toolkit for useful quantum. Technical report, IBM Quantum Learning, 2023b. URL `https://www.ibm.com/quantum/qiskit`.

Yunseok Kwak, Won Joon Yun, Soyi Jung, Jong-Kook Kim, and Joongheon Kim. Introduction to quantum reinforcement learning: Theory and pennylane-based implementation, 2021.

Junhua Liu, Kwan Hui Lim, Kristin L. Wood, Wei Huang, Chu Guo, and Heliang Huang. Hybrid quantum-classical convolutional neural networks. *Science China Physics, Mechanics & Astronomy*, 64, 2019. URL `https://api.semanticscholar.org/CorpusID:207847492`.

Ashley Montanaro. Quantum algorithms: an overview. *npj Quantum Information*, 2(1), January 2016. ISSN 2056-6387. doi: 10.1038/npjqi.2015.23. URL `http://dx.doi.org/10.1038/npjqi.2015.23`.

Frank Phillipson. Quantum machine learning: Benefits and practical examples. In *QANSWER*, 2020. URL `https://api.semanticscholar.org/CorpusID:211574012`.

A APPENDIX

A.1 QUANTUM COMPUTING CONCEPTS

Quantum Computing - Quantum computing is designed and operated on the probabalistic principles of quantum mechanics, primarily the superposition and entanglement of qubits, to augment classical computations or to solve problems that classical computers would not be able to solve within finite time frames (Montanaro, 2016).

Qubit - The inherent unit of quantum computing; a two-state quantum-mechanical system which can exist in a superposition and can be entangled together.

Quantum Machine Learning - Quantum Machine Learning (QML) augments or replaces classical machine learning methods with quantum circuitry / algorithms in order to enhance efficiency or to solve complex architectural problems inherent to classical computing.

Variational Quantum Circuit - A Variational Quantum Circuit (VQC) utilises a hybrid quantum-classical algorithm that uses a classical optimizer to train a parameterized quantum circuit in an iterative manner, with the weights in any given artificial neural network being approximated by these parameters (Phillipson, 2020).

Hadamard Gates - Rotates qubit states $|0\rangle$ and $|1\rangle$ to $|+\rangle$ and $|-\rangle$, respectively. It is useful for making superpositions of qubits (IBM, 2023a).

RY Gates - The RY gate implements $exp(-i(\theta/2)Y)$. On the Bloch sphere, this gate corresponds to rotating the qubit state around the y axis by the given angle and does not introduce complex amplitudes (IBM, 2023a).

Qiskit - Qiskit is IBM's open-source quantum computing software, facilitating programming, simulation, and access to quantum processors (IBM, 2023b)

A.2 ADDITIONAL GRAPHS

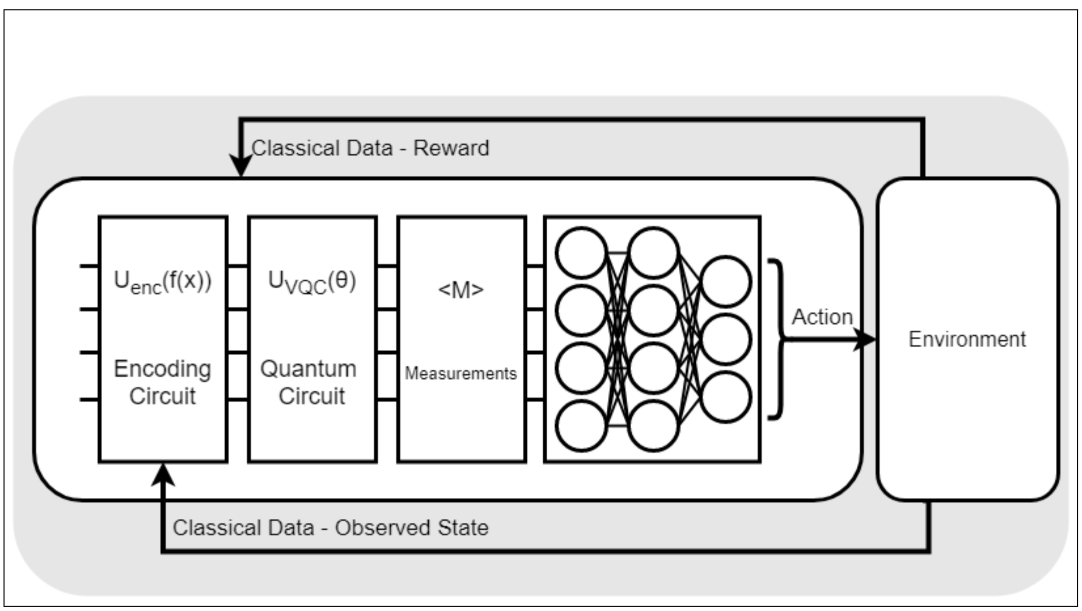

Figure 2: A diagram of the Hybrid Quantum-Classical RL architecture, with the environment inputs, VQC and FCN graphed.

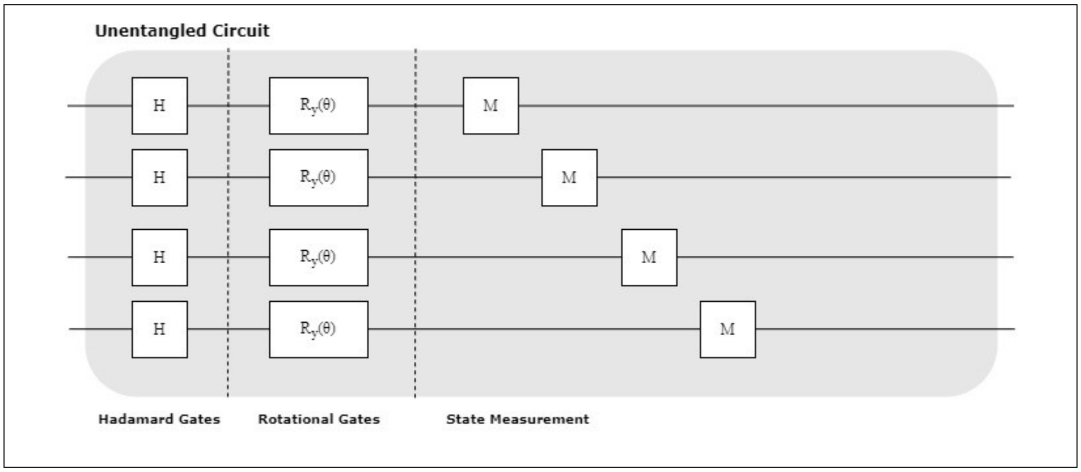

Figure 3: A diagram of the simple Variational Quantum circuit utilised in the quantum computational block.

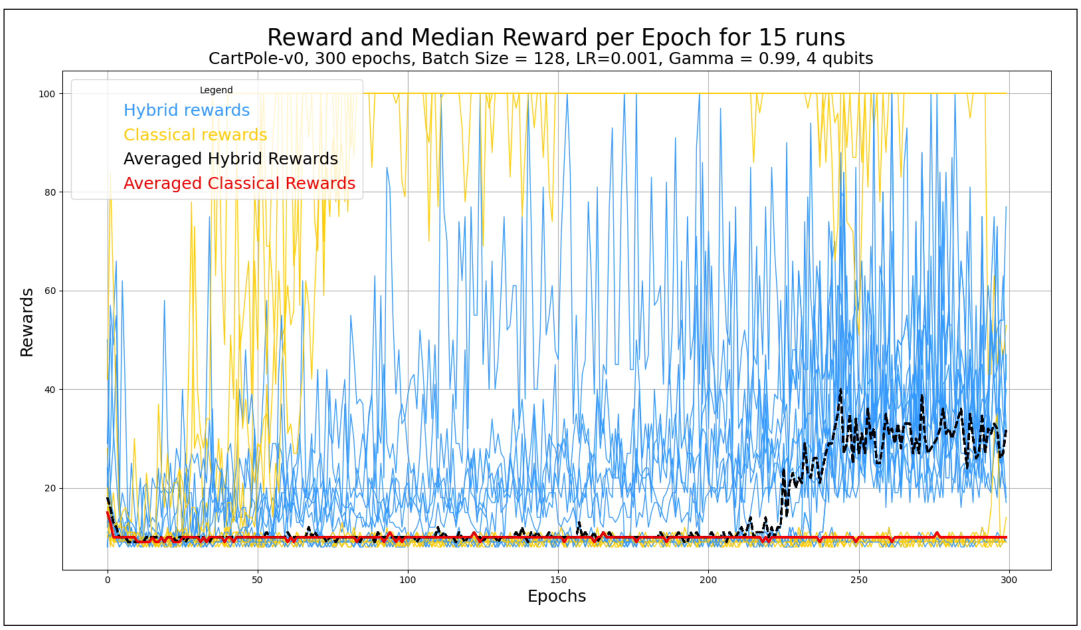

Figure 4: Reward per epoch and median reward for 15 runs of both the hybrid and classical models.

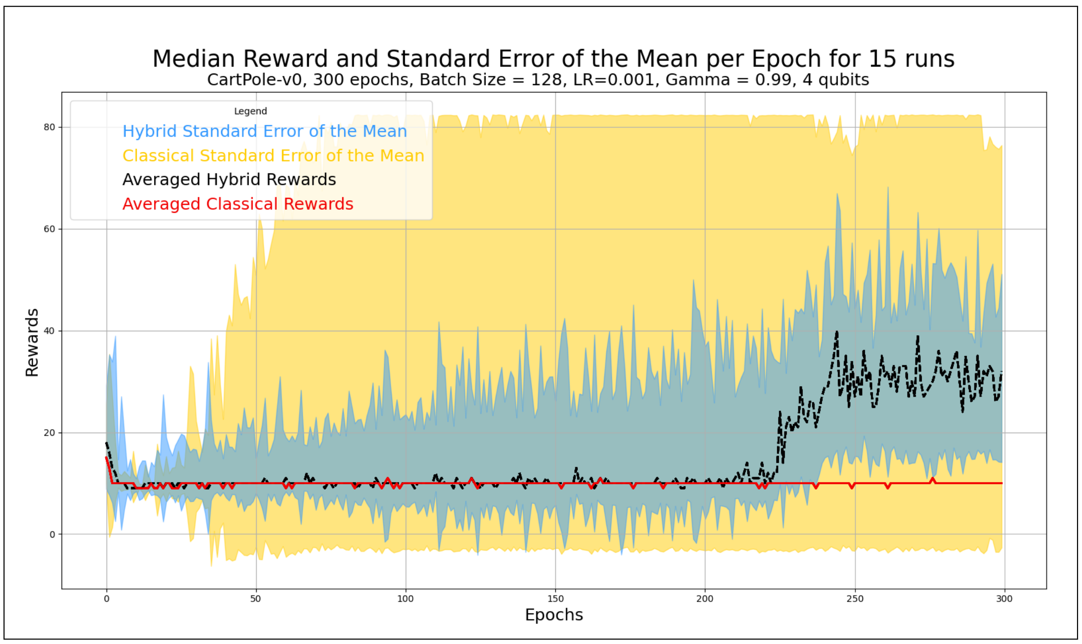

Figure 5: Median Reward and Standard Error of the Mean for 15 runs of both the hybrid and classical models, with the standard error showing the high variability of the classical model compared to the hybrid model.