Performance Analysis of Quantum Reinforcement Learning for Different Quantum Encoding Schemes: UAV Trajectory Optimization Application

Silvirianti*, Lia Suci Waliani†, and Soo Young Shin‡
e-mail: {*silvirianti93,†liasuciwaliani, ‡wdragon}@kumoh.ac.kr

***Abstract*—This paper investigates the performance of quantum reinforcement learning (QRL) under different quantum encoding schemes. In QRL, quantum encoding undertakes a significant role in transforming input data i.e., states and actions, which are in the form of classical values into quantum states. Quantum encoding has various schemes, it can be challenging to decide the most appropriate scheme which results in the highest prediction accuracy for the QRL. Thus, in this paper, some of the well-known quantum encoding schemes are employed on the QRL to optimize unmanned aerial vehicle (UAV) trajectory for maximum sum-rate as a study case. The performances of QRL under different quantum encoding schemes are compared and analyzed based on the achievable sum-rate as a performance metric. The results show that QRL with angle encoding outperforms QRL with amplitude encoding by achieving higher sum-rate as cumulative rewards.**

***Index Terms*—Accuracy, quantum encoding, quantum reinforcement learning, sum-rate, UAV trajectory.**

I. Introduction

Quantum reinforcement learning (QRL) has gained popularity in research interests because of its ability to solve various time-sequential optimization problems with lower complexity compared to classical methods [1], [2]. Quantum encoding, one of the vital processes in QRL, transforms input data into quantum states. Several works employed one of the existing quantum encoding schemes i.e., amplitude encoding, angle encoding, etc., in their proposed quantum algorithms [3], [4]. However, to the best of authors' knowledge, limited works have investigated the comparison performance of each quantum encoding scheme that is beneficial to acknowledge the most appropriate scheme that returns the highest prediction accuracy for the QRL. This paper provides a comparison and performance analysis of QRL for different quantum encoding schemes. As a study case, the QRL with several quantum encoding schemes are employed to optimize UAV trajectory for achieving maximum sum-rate in UAV communication networks scenario.

II. Quantum Encoding Schemes for Quantum Reinforcement Learning

In general, the QRL is akin to classical RL, where given the current states condition $\mathcal{S}^{(t)}$, a learning agent aims to choose the best action $\mathcal{A}^{(t)}$ and obtain maximum cumulative rewards $\mathcal{R}^{(t)}$ over time τ by learning the optimal action policy, denoted by $\Gamma_{\mathcal{A}}$, that is predicted employing neural networks. The difference is highlighted in the classical RL which utilizes classical neural networks, whereas QRL employs quantum neural networks for predicting the optimal action policy. In QRL, the states as learning inputs, which originally formed in classical values are required to be transformed into quantum states. The process of transforming classical values to quantum states, in quantum machine learning, is well-known as quantum encoding. Quantum encoding has various schemes which can be presented as follows: **(i) Amplitude encoding**, where the classical input data is encoded into the amplitudes of a quantum state. Let us consider the states information $\mathcal{S}^{(t)}$ as classical input data of QRL that consists of $\mathcal{K}$ samples, with $\mathcal{D}$ features each, which can be expressed as $\mathcal{S}^{(t)} = s_1^{(t),(1)}, \ldots, s_{\mathcal{D}}^{(t),(1)}, s_1^{(t),(2)}, \ldots, s_{\mathcal{D}}^{(t),(2)}, \ldots, s_1^{(t),(\mathcal{K})}, \ldots, s_{\mathcal{D}}^{(t),(\mathcal{K})}$. The amplitude vector of $\mathcal{S}^{(t)}$ can be defined as $\lambda = \bar{C}\Big(s_1^{(t),(1)}, \ldots, s_{\mathcal{D}}^{(t),(1)}, s_1^{(t),(2)}, \ldots, s_{\mathcal{D}}^{(t),(2)}, \ldots, s_1^{(t),(\mathcal{K})}, \ldots, s_{\mathcal{D}}^{(t),(\mathcal{K})}\Big)$, where $\bar{C}$ denotes the normalization constant. It is worth noting that the amplitude vector has to be normalized $|\lambda|^2 = 1$. Finally, the states of QRL can be represented by the amplitudes of a m-qubit[1] of a quantum state, which can be expressed as $\left|\mathcal{S}^{(t)}\right\rangle = \sum_i^{2^m} \lambda_i \left|i\right\rangle$, where λ_i denotes the element of the amplitude vector λ and $|i\rangle$ is the computational basis state. The amplitude encoding operation can be presented as $E^{\text{atd}}(\mathcal{S}^{(t)}) = \mathcal{S}^{(t)} \rightarrow \left|\mathcal{S}^{(t)}\right\rangle = \bigotimes_{b=1}^{\mathcal{B}} \mathbf{RY}(\tanh{(s_b)})\mathbf{H}$, where $\mathbf{RY}(\cdot)$ denotes the rotation on Y-axis on $\left|s_b^{(t)}\right\rangle$ by ψ phase, $\mathbf{H}$ denotes Hadamard gate and s_b represents b-th data of $\mathcal{S}^{(t)}$. **(ii) Angle encoding**, allows to encode classical input data as rotation angles. Let us consider the states of QRL consisting of $\mathcal{B}$ data, the operation of angle encoding can be presented as $E^{\text{angle}}(\mathcal{S}^{(t)}) = \mathcal{S}^{(t)} \rightarrow \left|\mathcal{S}^{(t)}\right\rangle = \bigotimes_{b=1}^{\mathcal{B}} \mathbf{RX}(\tanh{(s_b)})$, where $\mathbf{RX}(\cdot)$ denotes the rotation on X-axis on $\left|s_b^{(t)}\right\rangle$ by ψ phase and s_b represents b-th data of $\mathcal{S}^{(t)}$. In QRL, the encoding operation is also applied to network weights Θ, which can be expressed as $E^{\{\text{atd,angle}\}}(\Theta^{(t)})$, where index $\{\text{atd}, \text{angle}\}$ denotes amplitude encoding and angle encoding, respectively. Based on the quantum encoding scheme that is utilized, the feed-forward operation of QRL can be defined as $\mathcal{F}_{\text{QRL}} \triangleq \Big(E^{\{\text{atd,angle}\}}(\Theta^{(t)})\big(\prod_{m=1}^{M} \mathbf{CZ}(q_2|q_1) \otimes \ldots \otimes \mathbf{CZ}(q_{\text{M}}|q_{\text{M-1}})\big)E^{\{\text{atd,angle}\}}(\mathcal{S}^{(t)})\Big)$, where M denotes the number of qubits. At the end of $\mathcal{F}_{\text{QRL}}$ operation, quantum

[1]In quantum computing, a qubit is a term for quantum bits.

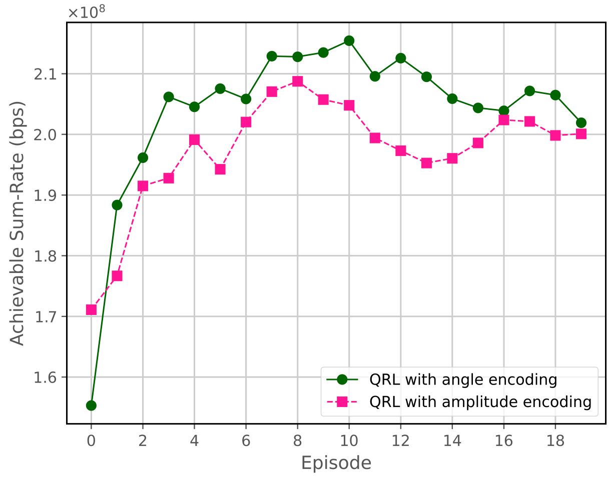

Figure 1. Achievable sum-rate under QRL with different quantum encoding schemes.

measurements, denoted by $\mathbb{Z}(|\cdot\rangle)$, are conducted to attain outputs in classical values. Subsequently, learning loss can be calculated as $\mathcal{L}(\Theta) = {}^{1}/\kappa \sum_k \left(y_k - \mathcal{F}_{\text{QRL}}(s_k, a_k; \Theta)\right)$, where $y_k = r_k + \gamma\Big(\mathcal{F}_{\text{QRL}}\big(s_{k+1}, \mathcal{F}_{\text{QRL}}(s_{k+1}; \Theta); \Theta\big)\Big)$, where $\mathcal{K}$ is the number of sample data. The gradient of QRL can be calculated employing parameter-shift rules [5], which can be expressed as follows $\nabla_\theta \mathcal{L}(\Theta) = \frac{1}{2\sinh(\phi)}\Big(\mathcal{L}(\Theta + \phi) - \mathcal{L}(\Theta - \phi)\big)\Big)$, where ϕ is the shifting parameter value. Finally, the networks weights can be updated using $\Theta = \Theta - \vartheta \nabla_\theta \mathcal{L}(\Theta)$, where ϑ denotes a learning step.

III. Study Case: UAV Trajectory Optimization

As a particular study case, the QRL with different quantum encoding schemes are employed to optimize UAV trajectory for achieving the maximum achievable sum-rate. The details of the wireless system model are as follows. A UAV is considerably employed as an aerial base station with limited battery energy. The UAV flies over a targeted service area, denoted by $\mathcal{G} \in \mathbb{R}^3$, and serves N terrestrial users. The UAV position at the t-th time can be defined as $g^{(t)} = \{x^{(t)}, y^{(t)}, a^{(t)}\}$, whilst the position of the n-th user at the t-th time can be defined as $g_n^{(t)} = \{x_n^{(t)}, y_n^{(t)}\}$. Herein, non-orthogonal multiple access (NOMA) is employed to enhance spectral efficiency, where multiple users are grouped into several groups. Let us assume $\mathcal{J}$ NOMA groups, where each j-th group accommodates N_j users. The distance between UAV and the n-th user in the j-th group can be defined as $V_{n,j}^{(t)} = \Big((x^{(t)} - x_{n,j}^{(t)})^2 + (y^{(t)} - y_{n,j}^{(t)})^2 + a^{(t)^2}\Big)^{1/2}$. Moreover, the transmitted power allocation for the n-th user in the j-th group at t-th time can be defined as $p_{n,j}^{(t)} = \mu_{n,j} P_{\text{Tx}}$, where $\mu_{n,j} \in (0, 1]$ is NOMA power coefficient for the n-th user in the j-th group and P_{Tx} denotes the UAV transmit power. The achievable rate of the n-th user in the j-th group at the t-th time can be defined as $R_{n,j}^{(t)} = \frac{B}{\mathcal{J}} \log_2\left(1 + \frac{p_{n,j}|h_{n,j}^{(t)}|^2}{\sum_{i \neq n}^{N_j - 1} p_{i,j}|h_{n,j}^{(t)}|^2 + \sigma^2}\right)$, where $|h_{n,j}^{(t)}|^2$ denotes the channel gain between the UAV and the n-th user in the j-th group. The objective is to optimize UAV trajectory that results in the maximum achievable sum-rate, which can be defined as follows

$$\max_{g^{(t)}} SR^{(t)} = \sum_{n=1}^{N_j} R_{n,j}^{(t)}, \tag{1a}$$

$$\text{s.t. } C_1 : \sum_{n=1}^{N_j} p_{n,j}^{(t)} \leq P_{\text{Tx}}, C_2 : R_{n,j}^{(t)} \geq \kappa, \tag{1b}$$

where κ is the minimum required achievable rate.

Based on the objective problem, the state space for QRL can be defined as $\mathcal{S}^{(t)} = \{h_{n,j}^{(t)}, \forall n \in N_j, j \in \mathcal{J}\}$, where $h_{n,j}^{(t)}, \forall n \in N_j, j \in \mathcal{J}$ is the channel condition between UAV and all the terrestrial users at t-th time. Moreover, the action space can be formulated as $\mathcal{A}^{(t)} = \{\varphi_\alpha = \chi_{\varphi_\alpha}.\pi, \ \varphi_\beta = \chi_{\varphi_\beta}.\pi\}$, where φ_α and φ_β denote angular and polar velocities of UAV, respectively. Finally, the reward can be defined as follows

$$\mathcal{R}^{(t)} = \begin{cases} SR^{(t)} = \sum_{n=1}^{N_j} R_{n,j}^{(t)}, & \text{if } C_1, C_2 \text{ are satisfied,} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

IV. Performance Analysis and Conclusion

The simulation parameters were presented as follows: $\mathcal{G} = 100 \times 100 \times 100 \ m^3$, $a^{(0)} = 50$ m, $v = 5.56$ m/s, $N_j = 2$, $B = 1$ MHz, $P_{Tx} = 30$ dBm, $\vartheta = 0.001$, $\gamma = 0.99$, $M = 6$, Episode $= 20$, Step $= 20$. The quantum operation $\mathcal{F}_{\text{QRL}}$ was performed employing IBM Qiskit [6]. Figure 1 shows the achievable sum-rate that represents a reward of QRL which was designed in Section III. As can be seen in Fig. 1, QRL with angle encoding achieved a higher sum-rate compared to QRL with amplitude encoding for the case of UAV trajectory optimization. The results indicate that the angle encoding scheme can improve prediction accuracy than the amplitude encoding. This is due to the angle encoding allows for a more flexible representation by adjusting the phase angles of the quantum state.

Acknowledgment

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2023-RS-2023-00259061) supervised by the IITP (Institute for Information Communications Technology Planning Evaluation). This work was supported by Institute of Information communications Technology Planning Evaluation (IITP) grant funded by the Korea government(MSIT) (No. 2021-0-02120, Research on Integration of Federated and Transfer learning between 6G base stations exploiting Quantum Neural Networks).

References

[1] W. J. Yun, J. P. Kim, S. Jung, J. -H. Kim and J. Kim, "Quantum Multiagent Actor–Critic Neural Networks for Internet-Connected Multirobot Coordination in Smart Factory Management," *IEEE Internet of Things J.*, vol. 10, no. 11, pp. 9942-9952, 1 June1, 2023.

[2] Silvirianti and S. Y. Shin, "Quantum Reinforcement Learning with Stochastic Parameter-Shift Rules for Energy-Efficiency in UAV-NOMA," *Proc. Symp. Korean Inst. Commun. Info. Sci.*, 2023, pp. 176-178.

[3] B. Narottama, D. K. Hendraningrat, and S. Y. Shin, "Quantum-inspired evolutionary algorithms for NOMA user pairing," *ICT Express*, vol. 8, pp. 11-17, Feb. 2022.

[4] B. Narottama and S. Y. Shin, "Quantum Federated Learning for Wireless Communications," *Proc. Symp. Korean Inst. Commun. Info. Sci.*, 2020, pp. 208-209.

[5] D. Wierichs, J. Izaac, C. Wang, and C.Y. Lin, "General parameter-shift rules for quantum gradients," *Quantum*, vol. 6, pp. 677, Mar. 2022.

[6] Abraham, H., Akhalwaya, I.Y., Aleksandrowicz, G., Alexander, T., Alexandrowics, G., Arbel, E., Asfaw, A., et al, "Qiskit: An Open-source Framework for Quantum Computing (Version 0.7.2). Zenodo.," 2019.