Resource Allocation in Multi-Cell Networks: A Deep Reinforcement Learning Approach

Harun Ur Rashid
Dept. of Information and Communications Engineering
Hankuk University of Foreign Studies (HUFS)
Seoul, Korea
Email: harun@hufs.ac.kr

Seong Ho Jeong
Dept. of Information and Communications Engineering
Hankuk University of Foreign Studies (HUFS)
Seoul, Korea
Email: shjeong@hufs.ac.kr

***Abstract*— The conventional resource distribution methodologies rely on numerical methodologies to enhance diverse performance metrics. Most of these endeavors can be classified as immediate, given that the optimization determinations stem from the present network condition without regard for historical network states. Although utility theory has the capacity to integrate long-term optimization consequences into these optimization actions, the escalating diversity and intricacy of network settings have made the resource allocation challenges insurmountable. The optimization of resources at an optimum level stand as a foundational hurdle for densely populated and mixed wireless environments with an extensive array of wireless connections. Owing to the intricate and non-linear nature of the optimization conundrum, the quest for the best resource allocation is a resource-intensive undertaking. Among the prospective solutions, reinforcement learning (RL) emerges as a viable candidate to resolve resource allocation dilemmas optimally across fluctuating network scenarios. This paper presents an innovative, centralized RL-based resource allocation method tailored for a multi-cell network, aiming to optimize connection stability and data rate by improving the quality of experience (QoE). Specifically, a deep Q-network (DQN) approach is employed to realize this objective. Empirical findings underscore that the proposed deep reinforcement learning (DRL) based resource allocation strategy delivers better performance within a multi-cell scenario.**

Keywords*— *Resource allocation, reinforcement learning, Deep Q-network, AI, Beyond 5G/6G cellular.

I. Introduction

In recent years, the field of cellular mobile communications has undergone significant advancements. It becomes evident that telecommunication operators must factor in the potential challenges arising from commoditization and the quality of service (QoS) for mobile users during the initial phase of 6G deployment [1]. In the conventional radio access network deployment, individual base stations (BSs) are physically equipped with a fixed number of antennas, facilitating radio functionalities within limited coverage areas. However, achieving higher transmission rates necessitates the installation of a vast number of physical BSs. This introduces complexities in substantial investments, wireless channel interference, different resource allocation, and diverse QoS requirements for different user equipment's (UEs) [2]. The optimal allocation of resources becomes a pivotal concern due to the enormous number of connections and the ultra-dense deployment of base stations on a significant scale. Historically, addressing this challenge involved heuristic techniques as the non-convex nature of the optimization problem presented obstacles. Yet, these approaches are computationally intensive, rendering them impractical for the demands of large-scale cellular networks. In contemporary times, machine learning (ML) techniques have been employed to derive pragmatic solutions for resource allocation challenges within expansive cellular networks [2]. These studies utilize datasets generated through diverse heuristic methodologies. However, these approaches entail considerable computational costs and time consumption. Consequently, the adoption of a supervised deep learning (DL) approach proves to be unsuitable for large-scale network systems [3].

In recent times, the fusion of DL with RL has given rise to Deep Reinforcement Learning (DRL) [4]. By harnessing this combination, DRL exhibits promise in effectively handling complex control problems. The integration of DL and RL empowers DRL to distill valuable insights from vast and high-dimensional datasets, enabling the acquisition of optimal action policies in such complex scenarios.

In the aforementioned context, we put forth a centralized downlink resource allocation strategy rooted in DRL, mainly employing the DQN algorithm. This scheme is tailored for multi-cell networks, aiming to optimize the twin objectives of connection stability and data rates. Our approach makes use of the mobile-env [5] environment, wherein they define the state space, action space, and reward function to guide the DRL agent's decisions. Through an array of simulation experiments encompassing various training parameters, we showcase the scalability and robustness of our proposition, demonstrating its efficacy within a comprehensive network scenario.

II. Related Studies

DRL has emerged as a promising contender for optimizing the long-term utility of resource allocation [6]. In a distinct effort [7], researchers employed the DRL framework to execute joint user association and resource allocation within the heterogeneous network. The overarching objective was to enhance the network's long-term utility while upholding QoS prerequisites. Likewise, in other studies, a multi-agent DQN [8] and a centralized DQN approach [9] were harnessed for power allocation in wireless networks, aiming to maximize the system's weighted sum-rate. Furthermore, in [10], a two-pronged approach involving centralized and multi-agent DRL techniques was adopted for resource allocation within multi-cell scenarios. Predominantly, extant research on distributed strategies utilize multi-agent DRL and actor-critic algorithm [10, 11]. Some investigations, albeit fewer, delve into centralized DRL approaches but often within constrained settings involving smaller base stations and network configurations. Notably, while Actor-Critic methods offer a synthesis of policy-based and value-based learning, their intricate architecture and susceptibility to convergence issues can deter those in search of straightforward solutions. In addition, the accrued benefits may not consistently outweigh the augmented implementation intricacies, particularly when more streamlined alternatives like DQN prove adequate.

III. Resource Allocation in Multi-cell Networks: A DRL approach

A. DQN approach

DQN constitutes a fusion of a deep neural network (DNN) and Q-learning reinforcement algorithm. At any given time, *t,* the DQN agent obtains a state s_t from the encompassing state space S and proceeds to execute an action a_t from the action space $\mathcal{A}$. This action is determined by the agent's adherence to a policy, denoted as $\pi(a_t|s_t)$, which signifies the mapping from state s_t to action a_t. Subsequent to the action a_t being performed, the agent receives a reward r_t and transitions to a new state s_{t+1}. This sequential process continues until the terminal state is reached, upon which the cycle restarts anew. The agent's primary aim is to maximize the cumulative reward, which is characterized as the discounted accumulated reward and denoted as $\mathcal{R}_t$. This cumulative reward is calculated as the summation, over an infinite horizon, of rewards and the equation denoted as below where γ^X is discount factor and x is base stations.

$$\mathcal{R}_t = \sum_{X=0}^{\infty} \gamma^X r_{t+X} \tag{1}$$

In this context, the discount factor γ, existing within the interval (0,1], governs the significance assigned to future rewards relative to present rewards. The action-value function $Q_\pi(s,a)$ and expressed as

$$Q_\pi(s,a) = \mathbb{E}[\mathcal{R}_t | s_t = s, a_t = a] \tag{2}$$

represents the anticipated return upon selecting action a within state s while adhering to policy π. This function encapsulates the expected cumulative reward of following a certain policy. The optimal action-value function, denoted as $Q^*(s,a) = max_\pi Q_\pi(s,a)$, reflects the highest achievable action value attainable by adhering to any policy within state s and for action a. This optimal function is characterized by the Bellman equation:

$$Q^*(s,a) = \mathbb{E}_{s'}[r + \gamma \max_{a} Q'(s', a')|s,a] \tag{3}$$

In the realm of DQN, neural network is leveraged to approximate the optimal action-value function, as $Q(s,a;\ \theta) \approx Q^*(s,a)$. Here, $Q(s,a;\ \theta)$ denotes the DQN, with θ representing the neural network's parameter. Through iterative updates, the Q-network is trained, leading to a reduction in the mean-squared error associated with the Bellman equation.

B. Environment description

Within this study, we employed mobile-env [5], an accessible and uncluttered platform devised to facilitate the training, assessment, and comparison of coordination methodologies, with a particular focus on their applicability in wireless mobile networks.

Given the intricacies inherent to mobile scenarios, achieving a comprehensive depiction of the environment state is impractical, even when considering a centralized agent [10]. Thus, the utilization of a partially observable Markov decision process (MDP) becomes pertinent. This MDP is characterized by a tuple (S, $\mathcal{A}$, $\mathcal{R}$), encompassing states S, actions $\mathcal{A}$, and the reward function $\mathcal{R}$, as expounded upon subsequently.

- States, S: At each time step, the DQN algorithm solely acquires insight into the present connections (C_j) of (UEs), their SINR ($SINR_j$) and utility (U_j) with reference to each cell c_j and UE u_j. Mobile-env standardizes the values of all states to fall within the interval of [−1, 1] (or [0, 1]).
- Actions, $\mathcal{A}$: The protocol streamlines overhead by enabling each UE to connect or disconnect from a single cell at a time, efficiently reducing the action space compared to arbitrary cell subsets. For a_j=i (i{1,...,n}), the action toggles u_j's connection status with cell c_j, creating or ending a connection. Conversely, a_j=0 is a no-op, leaving u_j's connections unchanged.
- Reward, $\mathcal{R}$: As per the specification outlined in mobile-env [5], the primary objective centers on elevating the average QoE for UEs. The reward attributed to DQN at a specific time step t is computed as the mean of current utilities across all UEs. The reward function is expressed

$$\mathcal{R} = \frac{1}{N} \sum j \in \{1, \dots, N\} \tag{4}$$

Internally, the DRL agent endeavors to optimize long-term utility by maximizing cumulative rewards that have been discounted over time.

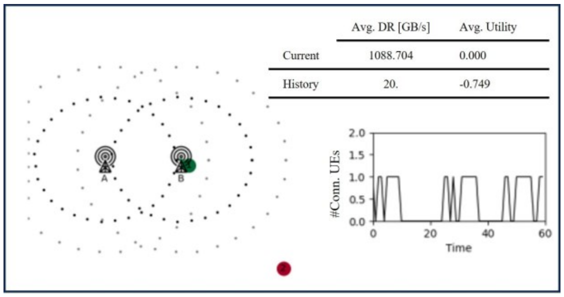
IV. Implementation and Results

The simulation configuration for our proposed resource allocation scheme employs DQN to tailor a multi-cell network approach. This paper explores the approach in a customized mobile-env[5] simulation scenario. The customized mobile-env scenario configuration has two base stations, and UEs is set at two, and their movement velocity is 10m/s, giving rise to a controlled environment for experimentation. To initiate training for our DQN model, it's imperative first to define the neural network-based Q network. For this purpose, we adopt a deep neural network comprising two hidden layers, leveraging the 'tanh' activation function for these layers. In terms of the Q-network's input layer size, it encompasses a state size of 12, accounting for the state space and the number of users. Moreover, the output layer of the DQN, tailored to our specific mobile-env scenario, embodies a total of three actions, representing the number of cells plus an additional option.

The experimental framework employed in this study utilized a computing system with an Intel® Core™ i7-8700 CPU @3.20GHz ×12 Processor, 16GB RAM, and NVIDIA GeForce RTX20270 GPU and operated on the Windows 10 platform. The computing provisions ensure a robust foundation for conducting our investigation.

A visual representation in Fig. 1 provides the simulation scenario under examination. The figure includes two base stations, each with its respective coverage range. Additionally, the movement patterns of users are captured through snapshots taken at different time slots. The lines drawn between UEs, and BSs symbolize the connections, further denoting the QoE through a color spectrum. Green hues represent favorable QoE, while red hues indicate suboptimal performance. This visualization provides an essential context for understanding the dynamic nature of user mobility and connectivity.

The presented outcomes are an aggregation of results obtained from an average of 50 to 70 runs, each spanning 50,000 steps. The evaluation of the proposed DQN scheme hinges on two critical metrics: the sum of average data rate and rewards. These metrics encapsulate the scheme's performance, which is centered on maximizing QoE through the reward function.



	Avg. DR [GB/s]	Avg. Utility
Current	1088.704	0.000
History	20.	-0.749

Fig. 1. Resource allocation in mobile-env without any model deployment

As depicted in Fig. 2 and Fig. 3, the results indicate the successful convergence of the DQN model. The architecture of the DQN, specifically the number of hidden layers, assumes a pivotal role in the scheme's efficacy. This is due to the DQN's function of approximating the action value function (Q). More hidden layers enable the DQN to capture additional features from the state, thereby influencing its learning capacity.

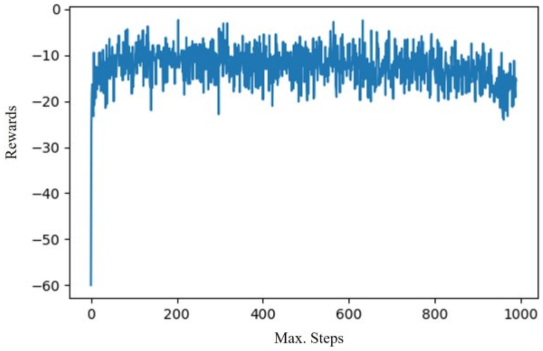
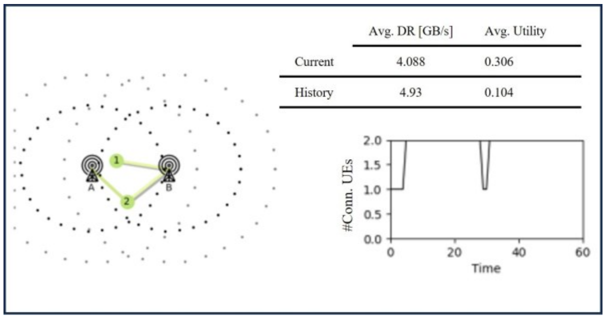


Fig. 2. DQN algorithms average normalized reward over period.

In light of this, we conduct experiments by varying the DQN's hidden layer size and activation function. The outcomes, as depicted in Fig. 2, reveal that increasing the hidden layer size may lead to a slight degradation in the performance of the DRL model. This phenomenon can be attributed to the potential overfitting arising from the DQN learning extraneous features or noise due to excessive hidden layers.



	Avg. DR [GB/s]	Avg. Utility
Current	4.088	0.306
History	4.93	0.104

Fig. 3. Resource allocation in mobile-env with DQN algorithm

In Fig. 3, we delve into the intricacies of the multi-cell selection scenario. Notably, this setup showcases the consistent connectivity of UEs to cells, alongside an examination of data rates and average utilities as crucial indicators of rewarding performance. A discernible pattern emerges, where current utility consistently outperforms previous utility. This trend underscores an improved QoE as a consequence of the approach.

V. Conclusions

We have presented a novel DRL-based scheme for resource allocation by maintaining QoE in multi-cell networks. Specifically, we have used DQN with experience replay for the proposed scheme. Simulation results encompass that a comprehensive exploration of the proposed DQN with two hidden layers is enough to approximate the action-value function for this case. The insightful visualizations shed light on the scheme's intricacies, convergence, and the factors impacting its efficacy. These findings contribute to the broader discourse surrounding resource allocation optimization in multi-cell scenarios, enhancing our understanding of its practical implications.

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-2022-00156353) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation). This work was supported by the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2021-0-00484).

References

[1] M. Zangooei, N. Saha, M. Golkarifard, and R. Boutaba, “Reinforcement Learning for Radio Resource Management in RAN Slicing: A Survey,” *IEEE Commun. Mag.*, vol. 61, no. 2, pp. 118–124, Feb. 2023, doi: 10.1109/MCOM.004.2200532.

[2] K. I. Ahmed, H. Tabassum, and E. Hossain, “Deep Learning for Radio Resource Allocation in Multi-Cell Networks.” arXiv, Aug. 02, 2018. doi: 10.48550/arXiv.1808.00667.

[3] F. Hussain, S. A. Hassan, R. Hussain, and E. Hossain, “Machine Learning for Resource Management in Cellular and IoT Networks: Potentials, Current Solutions, and Open Challenges,” *IEEE Commun. Surv. Tutor.*, vol. 22, no. 2, pp. 1251–1275, 2020, doi: 10.1109/COMST.2020.2964534.

[4] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, “Deep Reinforcement Learning: A Brief Survey,” *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 26–38, Nov. 2017, doi: 10.1109/MSP.2017.2743240.

[5] S. Schneider, S. Werner, R. Khalili, A. Hecker, and H. Karl, “mobile-env: An Open Platform for Reinforcement Learning in Wireless Mobile Networks,” in *NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium*, Apr. 2022, pp. 1–3. doi: 10.1109/NOMS54207.2022.9789886.

[6] M.-L. Tham, A. Iqbal, and Y. C. Chang, “Deep Reinforcement Learning for Resource Allocation in 5G Communications,” in *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Nov. 2019, pp. 1852–1855. doi: 10.1109/APSIPAASC47483.2019.9023112.

[7] N. Zhao, Y.-C. Liang, D. Niyato, Y. Pei, M. Wu, and Y. Jiang, “Deep Reinforcement Learning for User Association and Resource Allocation in Heterogeneous Cellular Networks,” *IEEE Trans. Wirel. Commun.*, vol. 18, no. 11, pp. 5141–5152, Nov. 2019, doi: 10.1109/TWC.2019.2933417.

[8] Y. S. Nasir and D. Guo, “Multi-Agent Deep Reinforcement Learning for Dynamic Power Allocation in Wireless Networks,” *IEEE J. Sel. Areas Commun.*, vol. 37, no. 10, pp. 2239–2250, Oct. 2019, doi: 10.1109/JSAC.2019.2933973.

[9] K. I. Ahmed and E. Hossain, “A Deep Q-Learning Method for Downlink Power Allocation in Multi-Cell Networks.” arXiv, Apr. 29, 2019. doi: 10.48550/arXiv.1904.13032.

[10] S. Schneider, H. Karl, R. Khalili, and A. Hecker, “DeepCoMP: Coordinated Multipoint Using Multi-Agent Deep Reinforcement Learning”.

[11] M. Kouchaki and V. Marojevic, “Actor-Critic Network for O-RAN Resource Allocation: xApp Design, Deployment, and Analysis,” in *2022 IEEE Globecom Workshops (GC Wkshps)*, Dec. 2022, pp. 968–973. doi: 10.1109/GCWkshps56602.2022.10008713.