Knowledge-Driven Deep Reinforcement Learning for Wireless Networks

Thi My Tuyen Nguyen, The Vi Nguyen, Tung Son Do, Dongwook Won, and Sungrae Cho
School of Computer Science and Engineering, Chung-Ang University, Seoul 06974, Republic of Korea
Email: {*tuyen, tvnguyen, tsdo, dwwon*}*@uclab.re.kr, srcho@cau.ac.kr*

***Abstract*—Deep reinforcement learning (DRL) is a promising technique for solving numerous optimization problems in wireless networks, owing to its high adaptability to dynamic and uncertain wireless environments. However, the lack of interpretability, inherent in the black-box nature of the deep neural networks (DNNs) within DRL, limits its deployment in real-world applications. In this paper, we focus on integrating domain knowledge into DRL to enhance robustness and learning efficiency. Specifically, we review a teacher-student framework, where a teacher agent utilizes knowledge from classical model-based methods to guide a student DRL agent toward improved decision-making.**

***Index Terms*—Deep reinforcement learning (DRL), knowledge-driven DRL, Wireless networks**

I. Introduction

Next-generation 6G networks are envisioned to deliver ubiquitous three-dimensional coverage through space-air-ground integration, intelligent-green network operations, and Internet of Everything (IoE) [1]–[5]. The 6G networks are expected to support services with significantly more stringent demands than the previous 5G generation, including higher reliability, lower latency, higher throughput, and greater energy efficiency. To meet these demands, it is important to efficiently optimize multi-dimensional resources, including computing, spectrum, power, and time. However, the inherent large scale, high density, and heterogeneity of 6G networks make the efficient solutions to the optimization problems in 6G systems particularly challenging [6]–[13].

The problem can be reformulated as a mathematical optimization problem and solved by using many optimization methods. Inherent in 6G networks' properties, the optimization problems can be formulated in the complicated form of objective and constraint functions, including non-convex functions due to the diversified QoS requirements, integration with new 6G functions (e.g., joint sensing, communication, and computing). In addition, optimization variables are high-dimensional due to a huge number of devices, large antenna arrays, and large amount of data. Furthermore, optimization problems in 6G networks usually involve complex, dynamical network parameters, such as channel state information and traffic state. Therefore, it is essential to design efficient and real-time controls, incorporating all the characteristics of the upcoming 6G networks. In the following, we introduce common approaches to tackle the optimization problems in wireless networks.

A. Classical Optimization Methods

Classical optimization methods are widely used in solving non-convex optimization problems in wireless networks, such as semidefinite relaxation, successive convex approximation [14], [15]. However, these methods face with critical disadvantages, as follows. Firstly, the iterative procedure in most algorithms requires a long time to converge, along with the high complexity. In addition, whenever the channel condition changes, the algorithms have to be re-executed entirely, limiting their ability to adapt quickly to rapid channel fluctuations. These disadvantages limit the applications of the classical optimization methods in real scenarios, especially in time-sensitive services in 6G networks.

B. Machine Learning-based Methods

Besides the traditional mathematical optimization methods, machine learning techniques are promising approaches. Among them, reinforcement learning techniques are gaining significant attention in solving sequential decision-making problems. The problem can be modeled as Markov decision process (MDP) and solved by reinforcement learning (RL), where the agent learns how to interact with the unknown environments to find the optimal policy (i.e., how to take an action at a state). However, the training process in conventional RL techniques is often slow, especially with the large-scale state space. To address these challenges, deep RL (DRL) has been proposed, where the deep neural networks (DNNs) are combined with RL to approximate value functions. Leveraging the potential of the excellent feature-capturing and fast online inference ability, DRL can successfully improve the learning performance of the RL algorithms. Additionally, by collecting historical training data, DRL can keep track of the dynamics of environments in real time, yielding higher adaptability. Moreover, DRL is well-suited for long-term optimization by learning the long-term policy rather than the instantaneous one.

Although DRL models offer greater flexibility and robustness than optimization-based methods under uncertainty and dynamic environments, their practical implementation remains challenging because of some reasons. First, DRL faces poor interpretability due to the black-box nature of the employed DNNs and slow convergence in online learning. Second, training a globally optimal policy is difficult because the action space is usually too large for exhaustive exploration. Moreover, storing sufficient training experiences is constrained by the limited memory of local devices. Third, although DRL can

979-8-3315-7896-1/26/$31.00 ©2026 IEEE

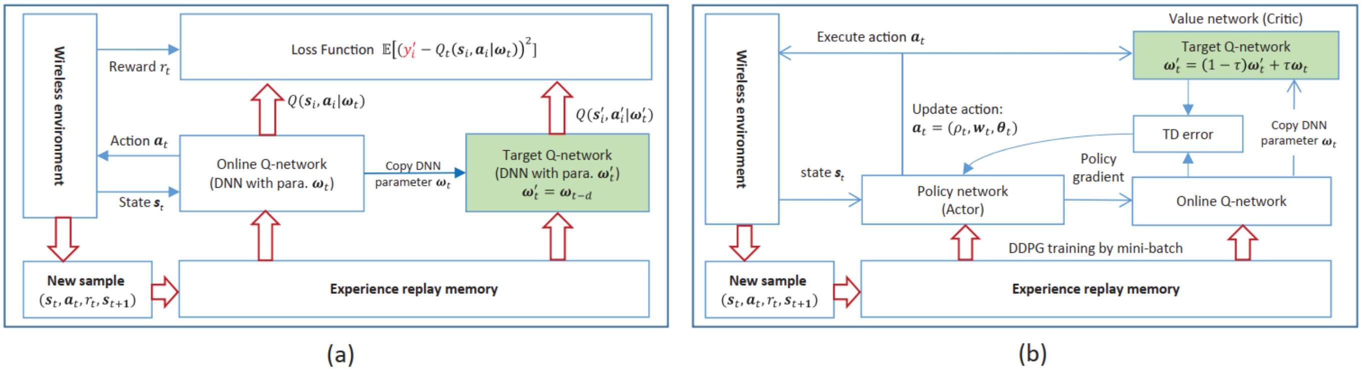

Fig. 1. Conventional DRL algorithms: (a) DQN and (b) DDPG [16].

be implemented with simple architectures, it involves many hyperparameters that require costly manual tuning, where improper settings may significantly degrade performance.

To improve the learning performance of DRL, communications-specific domain knowledge can be embedded into the DRL, including theoretical principles/rules and expert experiences, and insights can be exploited. Accordingly, knowledge-driven DRL in wireless networks refers to approaches that explicitly incorporate communication-specific domain knowledge into DNN models to compensate for limited training data and to guide the design of both network architectures and learning algorithms [17].

Motivated by these observations, this paper provides a comprehensive survey of recent advances in integrating domain knowledge with DRL and emphasizes their applications to key challenges in wireless networks.

II. Knowledge-driven DRL Algorithm

A. Deep Reinforcement Learning: Preliminaries

A general RL problem is formulated through MDP that is defined by the tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma \rangle$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $\mathcal{P}$ is the state transition probability, $\mathcal{R}$ is the reward function, and γ is the discount factor. At each time step t, given the current state $s_t \in \mathcal{S}$, the agent selects an action $a_t \in \mathcal{A}$ according to a policy π that maps states to actions. Following this policy, the agent transitions to the next state $s_{t+1} \sim \mathcal{P}(s'|s_t, a_t)$ and receives a reward $r_t = \mathcal{R}(s_t, a_t)$. The objective of the agent is to learn an optimal policy that maximizes the expected cumulative discounted reward, which is formulated as

$$\pi^* = \arg\max_{\pi} \mathbb{E}_\pi \left\{ \sum_{k=0}^{\infty} \gamma^k r_{t+k} \right\}. \tag{1}$$

In each decision-making step, the *state* function is defined as

$$V_\pi(s) = \mathbb{E}_\pi \left\{ \sum_{k=0}^{\infty} \gamma^k r_{t+k} \mid s_t = s \right\}, \tag{2}$$

which quantifies the expected cumulative discounted reward when starting from a state $s \in \mathcal{S}$ and following policy π thereafter. Similarly, the *state-action* function is defined as

$$Q_\pi(s, a) = \mathbb{E}_\pi \left\{ \sum_{k=0}^{\infty} \gamma^k r_{t+k} \mid s_t = s, a_t = a \right\}, \tag{3}$$

representing the expected cumulative discounted reward when starting from state s, taking action a, and subsequently following policy π.

In DRL frameworks, DNNs are incorporated into RL to improve learning efficiency in large-scale wireless networks characterized by high-dimensional state and action spaces. In general, DRL algorithms can be divided into two major categories: *value-based* and *policy-based* methods. Particularly, *value-based* methods (e.g., Q-learning, DQN) approximate the state-action value function using DNNs, and the corresponding policy is determined by selecting the action that yields the highest value. These methods are particularly suitable for problems with discrete action spaces. On the other hand, the *policy-based* methods learn directly the policy maps states directly to optimal actions or the probability of each action through DNNs. These methods are suitable for both discrete and continuous action spaces. Although DRL demonstrates superior performance, the black-box nature of DNN models limits their transparency and interpretability. This leads to trustworthiness issues, as it is difficult to verify model robustness in real-world deployment and to diagnose erroneous decisions. For instance, the authors in [18], [19] discuss the wrong and risky decisions, where wrong decisions degrade the average performance of a DRL agent, while risky decisions can induce high performance variance.

B. Domain Knowledge and Knowledge-driven DRL in Wireless Networks

To address aforementioned issues in DRL, domain knowledge can be ultilized to guide the decision-making process and detect erronous decisions. Generally, domain knowledge can be categorized into two classes: scientific knowledge and expert knowledge [17]. In particular, scientific knowledge includes theoretical transmission rules/laws/principles (e.g.,

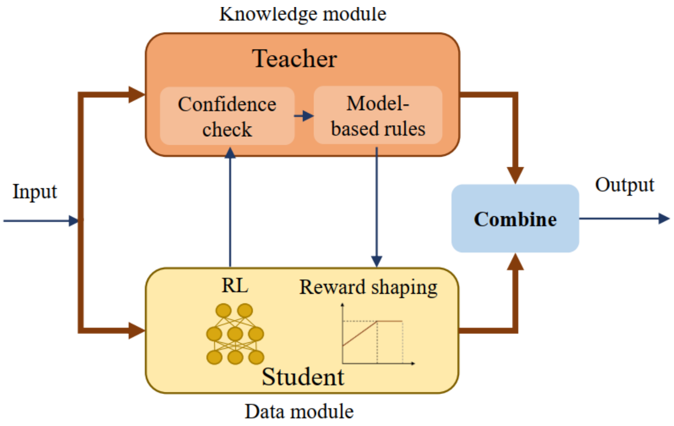

Fig. 2. Teacher-student learning framework [17].

Shannon's capacity formula), network modelling methodologies, and theoretical solutions to wireless network optimization. Expert knowledge encompasses practical experience and domain-specific insights that have been accumulated, refined, and validated by engineers and researchers in the field of wireless communication networks. Knowledge-driven DRL in wireless networks refers to approaches that explicitly integrate domain knowledge into DNNs employed in the DRL framework to compensate for limited training data, poor robustness, and improve learning efficiency.

In the following, we introduce a general framework that integrating domain knowledge into DRL [17]. As illustrated in Fig. 2, a teacher module (knowledge block), implemented using explainable theoretical algorithms, and a student module (data-driven block), implemented using dynamic DRL, operate concurrently to address the target problem. The confidence check module compares the outcomes from both modules, and the student network refines its policy by imitating the teacher's superior guidance through reward shaping. The final decision is obtained by combining the outputs of the two modules, thereby improving the overall reliability and robustness of the solution.

III. Conclusion

In conclusion, this paper has proposed the integration of domain knowledge into DRL to address the critical interpretability and robustness issues posed by its black-box nature. Through the teacher-student framework, the DRL student can be advised by the teacher knowledge. This approach enhances learning efficiency and decision-making, thereby facilitating the deployment of more reliable and trustworthy DRL solutions in real-world wireless networks.

Acknowledgment

This work was supported in part by the IITP (Institute of Information & Communications Technology Planning & Evaluation) - ITRC (Information Technology Research Center) (IITP-2026-RS-2022-00156353, 50%) grant funded by the Korea government (Ministry of Science and ICT) and in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00209125).

References

[1] Y. Shi, L. Lian, Y. Shi, Z. Wang, Y. Zhou, L. Fu, L. Bai, J. Zhang, and W. Zhang, "Machine learning for large-scale optimization in 6g wireless networks," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 4, pp. 2088–2132, 2023.

[2] J. Oh, D. Lee, D. S. Lakew, and S. Cho, "DACODE: Distributed adaptive communication framework for energy efficient industrial iot-based heterogeneous wsn," *ICT Express*, vol. 9, no. 6, pp. 1085–1094, 2023.

[3] T. S. Do, T. P. Truong, Q. T. Do, and S. Cho, "TranGDeepSC: Leveraging ViT knowledge in CNN-based semantic communication system," *ICT Express*, vol. 11, no. 2, pp. 335–340, 2025.

[4] M. C. Ho, A. T. Tran, D. Lee, J. Paek, W. Noh, and S. Cho, "A DDPG-based energy efficient federated learning algorithm with SWIPT and MC-NOMA," *ICT Express*, vol. 10, no. 3, pp. 600–607, 2024.

[5] D.-T. Hua, Q. T. Do, N.-N. Dao, and S. Cho, "On sum-rate maximization in downlink UAV-aided RSMA systems," *ICT Express*, vol. 10, no. 1, pp. 15–21, 2024.

[6] C. Song, D. Lee, Y. Lee, W. Noh, and S. Cho, "Deep learning based energy-efficient transmission control for STAR-RIS aided cell-free massive MIMO networks," *ICT Express*, vol. 11, no. 2, pp. 341–347, 2025.

[7] T. T. H. Pham, W. Noh, and S. Cho, "Multi-agent reinforcement learning based optimal energy sensing threshold control in distributed cognitive radio networks with directional antenna," *ICT Express*, vol. 10, no. 3, pp. 472–478, 2024.

[8] W. J. Yun, S. Park, J. Kim, M. Shin, S. Jung, D. A. Mohaisen, and J.-H. Kim, "Cooperative multiagent deep reinforcement learning for reliable surveillance via autonomous multi-UAV control," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 10, pp. 7086–7096, 2022.

[9] D. Kwon and D. K. Kim, "Channel estimation overhead reduction scheme and its impact in IRS-assisted systems," *ICT Express*, vol. 10, no. 1, pp. 58–64, 2024.

[10] S. H. Gardner, T.-M. Hoang, W. Na, N.-N. Dao, and S. Cho, "Metaverse meets distributed machine learning: A contemporary review on the development with privacy-preserving concerns," *ICT Express*, 2025.

[11] S. Park, H. Baek, and J. Kim, "The matrix: Quantum AI for interacting two worlds in prioritized metaverse spaces," *IEEE Communications Magazine*, 2024.

[12] D. Kwon, J. Jeon, S. Park, J. Kim, and S. Cho, "Multiagent ddpg-based deep learning for smart ocean federated learning IoT networks," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9895–9903, 2020.

[13] S. Park, G. S. Kim, Z. Han, and J. Kim, "Quantum multi-agent reinforcement learning is all you need: Coordinated global access in integrated TN/NTN cube-satellite networks," *IEEE Communications Magazine*, vol. 62, no. 10, pp. 86–92, 2024.

[14] T. V. Nguyen, T. P. Truong, T. M. T. Nguyen, W. Noh, and S. Cho, "Achievable rate analysis of two-hop interference channel with coordinated IRS relay," *IEEE Transactions on Wireless Communications*, vol. 21, no. 9, pp. 7055–7071, 2022.

[15] T. M. T. Nguyen, T. V. Nguyen, W. Noh, and S. Cho, "Statistical delay guarantee for the URLLC in IRS-assisted NOMA networks with finite blocklength coding," *IEEE Transactions on Wireless Communications*, 2024.

[16] S. Gong, J. Lin, B. Ding, D. Niyato, D. I. Kim, and M. Guizani, "When optimization meets machine learning: The case of irs-assisted wireless networks," *IEEE Network*, vol. 36, no. 2, pp. 190–198, 2022.

[17] R. Sun, N. Cheng, C. Li, W. Quan, H. Zhou, Y. Wang, W. Zhang, and X. Shen, "A comprehensive survey of knowledge-driven deep learning for intelligent wireless network optimization in 6g," *IEEE Communications Surveys & Tutorials*, 2025.

[18] Y. Zheng, L. Lin, T. Zhang, H. Chen, Q. Duan, Y. Xu, and X. Wang, "Enabling robust drl-driven networking systems via teacher-student learning," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 1, pp. 376–392, 2021.

[19] J. Hu, L. Chen, S. Shen, and T. Wang, "Explainable multi-agent deep reinforcement learning for joint task offloading and resource allocation in distance and channel-aware noma vehicular edge networks," *IEEE Internet of Things Journal*, 2025.