Analysis and Development Trends of Network Resource Allocation in Wireless Communications

Zongyi Yan *

Electronic Information School, Wuhan University, Wuhan, China

* Corresponding Author Email: 2023302121012@whu.edu.cn

Abstract. Mobile communication technology has evolved significantly from 1G analog systems to 5G intelligent networks, focusing on improvements in spectral efficiency, network capacity, and adaptability. Currently, 5G encounters challenges such as limited spectrum resources and increased bandwidth pressure, especially in high-density user environments. While traditional Non-Orthogonal Multiple Access (NOMA) technologies, like SCMA, and Device-to-Device (D2D) communication enhance spectrum reuse, they still face issues with dynamic resource allocation and cross-layer collaboration. This paper introduces an intelligent resource allocation framework leveraging Deep Reinforcement Learning (DRL) for dynamic power control and interference coordination through multi-agent collaboration. A Markov decision process model is developed, and a distributed DRL algorithm is created to optimize local and global performance in cellular networks. Experiments show that DRL-driven SCMA codebook scheduling can improve spectral efficiency by 20% while enabling distributed interference management and network slicing optimization in D2D scenarios. Nonetheless, challenges remain in practical DRL deployment, such as online training costs and policy interpretability. Future advancements will involve integrating DRL with sixth-generation (6G) technologies like intelligent reflecting surfaces (RIS) and terahertz beamforming, fostering a shift towards cognitive communication systems with autonomous perception and global optimization.

Keywords: Mobile communication technology, non-orthogonal multiple access, sparse code multiple access, direct device communication, dynamic resource allocation.

1. Introduction

The evolution of wireless communication technology is driven by the need for efficient resource allocation. Since Bell Labs introduced the first-generation cellular system (AMPS) in the 1970s, mobile communication has experienced transformative changes—transitioning from analog to digital, narrowband to broadband, and shifting focus from voice-centric services to the integration of multiple services. The 1G era was pivotal in establishing the groundwork for wireless communication through dynamic spectrum selection and cellular architecture, while the subsequent 2G (GSM) and 3G (WCDMA) technologies facilitated the widespread adoption of voice and low-speed data through digitization and multiple access techniques. The advent of 4G and 5G brought about further innovations, including Orthogonal Frequency Division Multiple Access (OFDMA) and Non-Orthogonal Multiple Access (NOMA), which overcame spectral efficiency challenges and spurred the explosive growth of mobile internet and the Internet of Things (IoT). However, as 5G commercialization accelerates and research into 6G intensifies, the exponential increase in network data traffic has exacerbated conflicts arising from high-density user scenarios, diverse service demands, and limited spectrum resources, revealing significant shortcomings in conventional resource allocation strategies.

The primary challenge in wireless resource allocation lies in achieving a balance between spectrum efficiency and interference management. Power-domain multiplexing techniques like NOMA, which utilize superposition coding and successive interference cancellation (SIC), greatly enhance spectral usage. Nevertheless, their dynamic optimization necessitates complex channel state information (CSI) and user pairing strategies, leading to substantial computational demands. Likewise, Device-to-Device (D2D) communication mitigates core network congestion through decentralized architectures, but its spectrum-sharing mechanisms require coordinated solutions to address cross-layer interference and energy efficiency. Concurrently, the synergistic utilization of high-frequency bands (such as

millimeter-wave and terahertz) alongside low-frequency bands has become a significant focus for 6G. For example, terahertz beamforming can facilitate ultra-high-speed data transmission across a 30 GHz operational bandwidth, but its limited ability to penetrate obstacles and provide coverage necessitates integration with emerging technologies like Reconfigurable Intelligent Surfaces (RIS).

In this landscape, artificial intelligence (AI) and Deep Reinforcement Learning (DRL) present transformative solutions for resource allocation challenges. DRL employs end-to-end learning frameworks to adaptively optimize power allocation, codebook scheduling, and interference management, significantly reducing computational costs in contrast to traditional game-theoretic methods. Research findings indicate that DRL can achieve a 20% enhancement in spectral efficiency for Sparse Code Multiple Access (SCMA) through dynamic power adjustments. In D2D scenarios, multi-agent collaboration facilitates distributed interference management, ultimately extending the battery life of IoT devices. However, deploying DRL is not without its challenges, including the acquisition of online training data, policy interpretability, and protocol standardization, all of which necessitate breakthroughs through digital twin technologies and explainable AI methods.

This study aims to comprehensively integrate cutting-edge multiple access technologies, intelligent algorithms, and emerging trends in 6G to propose a multi-layer resource allocation framework. By investigating the synergies among NOMA, D2D, and DRL, as well as exploring the dynamic regulatory potential of terahertz bands and RIS, it provides theoretical foundations for developing high-density, low-latency 6G networks. Recent advancements, such as terahertz polarization state modulation and RIS-enabled reconfiguration of wireless environments, illustrate innovative pathways for efficient utilization of high-frequency resources. These innovations not only advance wireless networks.

2. Evolution of wireless communication and resource allocation challenges

2.1. Network development history and resource conflicts

The current landscape of wireless communication technology is marked by unprecedented demands on resource allocation, driven by the widespread deployment of 5G networks and emerging applications such as the Internet of Things (IoT), ultra-high-definition video streaming, and massive machine-type communications. These advancements have led to exponential growth in network data traffic, exacerbating systemic challenges in spectrum scarcity, bandwidth limitations, and interference management. Traditional orthogonal resource allocation models, designed for legacy systems, struggle to address the dual pressures of coverage and capacity. Millimeter-wave bands, while offering high throughput, suffer from limited coverage and poor penetration, whereas low-frequency bands face congestion due to overcrowded device deployments. This imbalance is particularly acute in urban environments and IoT-dense scenarios, where high user density and centralized device clusters create fierce competition for limited resources, resulting in latency spikes, energy inefficiency, and degraded quality of service (QoS). Furthermore, the lack of coordination across the industry ecosystem—spanning spectrum allocation, chip design, and infrastructure deployment—has intensified resource fragmentation and operational costs, hindering the scalable adoption of next-generation technologies.

To overcome these challenges, innovative approaches such as non-orthogonal multiple access (NOMA), sparse code multiple access (SCMA), and device-to-device (D2D) communication have emerged as transformative solutions. NOMA addresses spectrum scarcity by enabling non-orthogonal resource sharing through power-domain multiplexing, allowing multiple users to coexist on the same frequency band with differentiated power levels. By superimposing user signals and leveraging successive interference cancellation (SIC), it significantly enhances spectral efficiency while supporting massive connectivity in high-density environments. SCMA complements this by introducing sparse codebook-based modulation, where data streams are mapped to multi-dimensional codewords with low collision probability. This reduces decoding complexity and improves connectivity density, making it ideal for scenarios requiring simultaneous access from numerous low-

power IoT devices. Meanwhile, D2D communication alleviates core network congestion by enabling direct data exchange between proximate devices, bypassing traditional base station routing. This not only reduces latency and spectrum contention but also enhances energy efficiency, particularly in localized applications like smart factories or vehicular networks. Together, these technologies form a synergistic framework for dynamic resource optimization: NOMA maximizes spectral utilization, SCMA enhances access scalability, and D2D offloads traffic pressure, collectively enabling adaptive interference management and heterogeneous service coordination. Their integration addresses the critical gaps in current systems, offering a pathway to balance coverage-capacity trade-offs, mitigate resource contention, and support the ultra-reliable, low-latency requirements of future 6G networks. By redefining resource allocation paradigms, they pave the way for sustainable growth in an era defined by hyper-connectivity and diverse application demands.

2.2. Existing resource allocation methodology

Non-Orthogonal Multiple Access (NOMA) is a technique to improve the efficiency of wireless communication through non-orthogonal resource allocation, the core of which lies in the power domain multiplexing and interference co-management. NOMA allows users to transmit with different power stacks at the same time and frequency resources, and the receiving end decodes the signals layer by layer by means of the Successive Interference Cancellation (SIC) technique: the user with the poorer channel conditions is assigned higher power, and its signal is decoded and canceled first. The signals of users with poorer channel conditions are assigned higher power and their signals are prioritized for decoding and cancellation, followed by decoding the signals of low-power users. This mechanism significantly improves spectral efficiency, especially for scenarios with large differences in user channels or high density (e.g., IoT, dense urban areas). NOMA is a key technology in fifth-generation (5G) systems, has emerged as a promising alternative to traditional Orthogonal Multiple Access (OMA) by enabling efficient spectrum utilization and balancing user fairness with spectral efficiency [1]. Unlike OMA, which relies on orthogonal resource allocation, NOMA allows simultaneous transmission to multiple users over shared time, frequency, or code resources through power-domain or code-domain multiplexing [2]. In power-domain NOMA, users are differentiated by distinct power levels at the transmitter, with receivers employing successive interference cancellation (SIC) to decode signals. Code-domain NOMA, meanwhile, optimizes resource sharing through advanced coding schemes. For downlink scenarios, NOMA prioritizes fairness by allocating higher power to users with weaker channel conditions while leveraging superposition coding and SIC at base stations. Critical challenges involve optimizing power allocation and user pairing strategies, as exhaustive searches for ideal resource combinations remain computationally intensive. Research indicates that NOMA significantly enhances system capacity and cell-edge user throughput compared to OMA, with further performance gains achievable through integration with technologies like Bioresearch by Liang Xiaolin (2024) indicates that NOMA can increase achievable rate by up to 30% compared to OMA. Effective algorithms for dynamic resource management continue to be a focus for maximizing NOMA’s potential in real-world deployments [3].

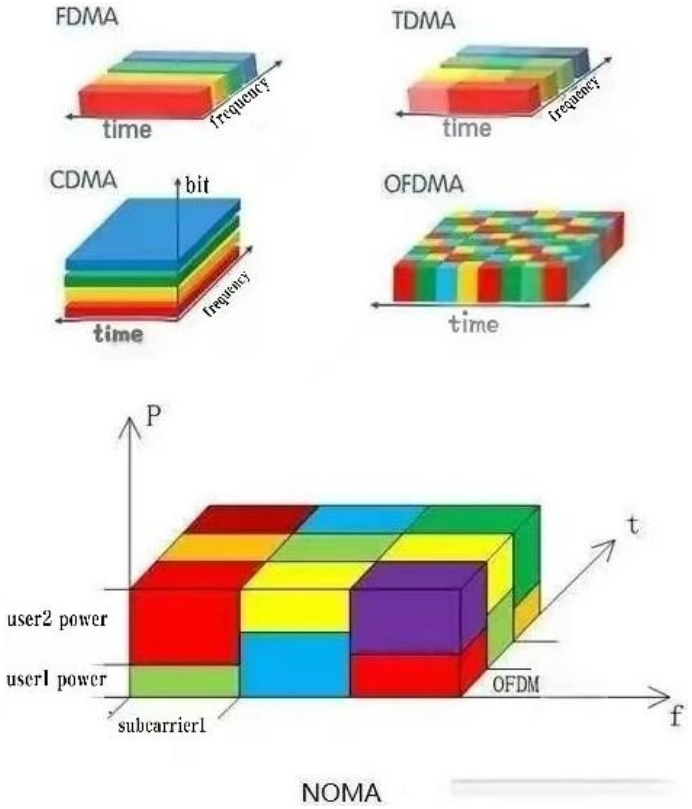

Figure 1. Schematic diagram of NOMA principle

As a critical 5G innovation, Sparse Code Multiple Access (SCMA)—a non-orthogonal multiple access (NOMA) variant developed by Huawei—integrates principles of OFDMA and CDMA to enable multi-user transmission in the frequency domain [4] [5]. By leveraging multi-dimensional codebooks, SCMA enhances spectral efficiency and user capacity through dynamic power allocation tailored to individual nodes, ensuring optimal throughput and service quality. Its downlink implementation employs a three-tier power allocation strategy that dynamically adjusts transmission power based on channel state information (CSI) and Quality of Service (QoS) requirements, maintaining communication reliability.

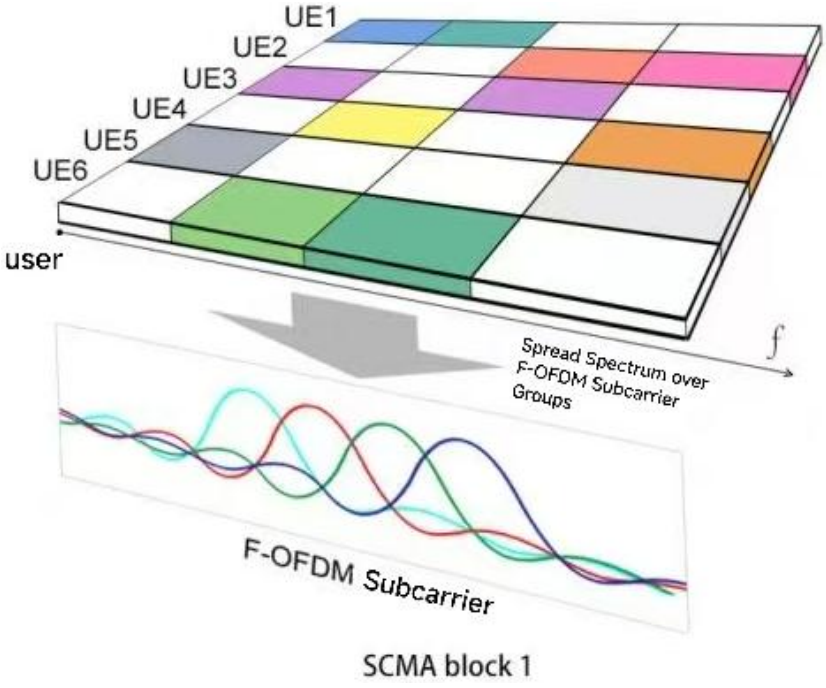

Figure 2. Schematic diagram of SCMA principle

While SCMA optimizes multi-user transmission through advanced coding and power control, 5G networks further enhance connectivity scalability by integrating Device-to-Device (D2D) communication—a paradigm that shifts traffic away from centralized infrastructure. Complementing this, Device-to-Device (D2D) communication enables direct data exchange between proximate devices, bypassing cellular infrastructure to reduce network congestion and improve QoS [6] [7]. Standardized as Proximity Services (ProSe) by 3GPP, D2D operates via in-band (licensed spectrum) or out-band (unlicensed spectrum) modes [8]. In-band D2D further divides into underlay (shared spectrum with cellular users) and overlay (dedicated spectrum) approaches, while out-band supports autonomous (base station-independent) or controlled (base station-managed) operations. Key resource allocation challenges, such as power distribution and spectrum management, focus on balancing energy efficiency, spectral utilization, and system throughput, positioning D2D as a cornerstone for 5G's high-density connectivity demands.

In recent years, Deep Reinforcement Learning (DRL) has demonstrated unique advantages in resource allocation. Addressing challenges posed by multi-dimensional resource competition and dynamic demand variations in complex environments, traditional methods often struggle with high-dimensional state modeling and real-time decision efficiency. DRL establishes an end-to-end mapping from environmental states to allocation strategies by integrating the perception capabilities of Deep Neural Networks (DNNs) with the sequential decision-making mechanisms of Reinforcement Learning (RL). Its core lies in agents progressively learning optimal strategies to maximize long-term rewards through iterative trial-and-error interactions, particularly suited for scenarios with uncertainty and time-varying resource dynamics. Research indicates that DRL not only resolves Multi-Agent Collaboration (MAC) optimization problems challenging for traditional approaches but also enables adaptive policy adjustments in unknown environments, achieving dynamic allocation of computing resources, communication bandwidth, and spectrum resources. Leveraging model-free learning capabilities, systems can execute efficient decisions without prior knowledge of environmental dynamics, offering a novel pathway for autonomous resource management in intelligent networks.

3. Intelligent resource management in wireless networks using DRL

DRL [9] has become a key technology in the field of Dynamic Spectrum Access by virtue of its powerful ability to process high-dimensional state and action spaces, and the ability to efficiently map the environment states to optimal actions to maximize the Q-value by combining a deep neural network with a Q-learning framework to form a Deep Q Network (DQN). Studies have shown that the DRL algorithm developed by Talaat et al [10]. Successfully optimizes multi-player utility sharing in multi-intelligence collaborative scenarios (e.g., MNIST image classification games and channel switching puzzles), while Li et al. design a DRL-based spectrum sensing strategy for environments in which a single user dynamically interacts with an unknown joint Markov model, overcoming the problems of unknown system dynamics and the study by Chang et al. further focuses on the dynamic access decision of a single cognitive agent in N channels, and theoretically demonstrates that the optimal policy can be resolved when the channel state transfer probability is known, while the deep Q-learning is used to train the state-action value function when the model is unknown, and experimentally verifies that the learning policy is as close as possible to the theoretical optimal solution [11] [12]. Approximation. Aiming at the challenges of IoT devices with limited size and insufficient spectrum sensing capability, Tan et al. propose a spectrum sensor assistance system based on reinforcement learning, which collaboratively realizes dynamic spectrum access through external sensors and effectively improves the spectrum utilization of existing networks [13]. Together, these results highlight the core advantages of DRL in dynamic spectrum allocation - circumventing complex modeling and adaptively optimizing spectrum resource allocation through an end-to-end learning mechanism, which provides an important technological path for autonomous decision-making and efficient resource management in smart wireless networks.

Deep Reinforcement Learning (DRL) provides breakthrough solutions for the optimization of 5G core technologies (e.g., SCMA and D2D) by integrating deep neural networks and reinforcement learning mechanisms: in SCMA scenarios, DRL dynamically adjusts the power allocation and multi-dimensional codebook scheduling, and optimizes the spectral efficiency based on the real-time channel state and service demand, which significantly reduces computational complexity and improves user capacity compared to the traditional game-theoretic approach that relies on fixed models. Compared with traditional game theory methods relying on fixed models, its end-to-end learning capability significantly reduces computational complexity and improves user capacity; for the spectrum sharing and interference coordination problems of D2D communication, DRL realizes distributed power control and dynamic spectrum access through multi-intelligence collaboration, which extends the endurance of IoT devices while reducing co-frequency interference, while traditional centralized scheduling is difficult to match the demand of high-density scenarios due to signaling overhead and response delay; furthermore, DRL supports SCMA and multi-dimensional codebook scheduling. DRL supports cross-layer co-optimization of SCMA and D2D, realizes dynamic allocation of network slice resources (e.g., prioritizing ultra-low latency of URLLC) through a layered reinforcement learning framework, and builds a self-healing network to cope with sudden interferences or link interruptions by means of a continuous learning mechanism, which is more adaptable to 5G heterogeneous environments than the traditional layered optimization strategy; however, the actual deployment of DRL still needs to overcome the challenges of obtaining online training data, model interpretation, and protocol standardization, as well as the need of online training data acquisition, model interpretation, and protocol standardization. However, the actual deployment of DRL still needs to overcome the challenges of online training data acquisition, model interpretability, and protocol standardization and integration.

The primary challenge in obtaining online training data is the data distribution shift in dynamic environments. In wireless communication scenarios, user mobility and channel variability lead to highly unstable real-time data distributions. For example, in densely populated urban areas, rapid changes in user locations and channel states can quickly render training data obsolete, making it difficult for DRL models to capture long-term dynamics. Privacy and security constraints further complicate data acquisition. To comply with regulations like GDPR, user-level data (such as location and traffic patterns) must be anonymized, potentially undermining data utility—anonymized channel state information (CSI) collected by base stations may not accurately reflect user behavior, reducing the predictive accuracy of DRL models. Additionally, labeling costs and noise issues cannot be overlooked. Automatic labels based on instantaneous signal-to-noise ratios may be impacted by multipath interference, particularly in D2D communication, where dynamic interference can lead to estimate discrepancies and affect policy convergence.

Lack of model interpretability also limits the trustworthiness of DRL deployments. The use of complex neural networks results in opacity in decision-making logic. For instance, in Sparse Code Multiple Access (SCMA) systems, DRL-driven codebook scheduling may allocate certain users to low-priority levels without clear reasons for the selection of specific power levels. This is particularly crucial in regulatory contexts, where operators need to verify algorithms' fairness, and regulatory bodies require compliance audits. Stringent requirements from organizations like the European Telecommunications Standards Institute (ETSI) regarding algorithm transparency could restrict DRL's application in critical tasks. Additionally, when DRL policies result in network performance declines, tracing decisions using historical reward functions becomes difficult, extending operational response times.

Protocol standardization lags further impede DRL's large-scale deployment. Current communication protocols (such as 5G NR) do not define interaction interfaces for DRL, complicating cross-vendor collaboration. For instance, Huawei's SCMA may be incompatible with Ericsson's Dynamic Spectrum Sharing (DSS), hindering distributed control. Real-time constraints in ultra-reliable low-latency communication (URLLC) scenarios require DRL to respond within milliseconds, but traditional signaling (such as CSI feedback) can introduce delays. If base station cooperation

relies on X2 interfaces and protocols lack low-latency optimizations for DRL, real-time decision-making efficiency will be significantly constrained. Furthermore, the absence of a unified framework can cause parameter mapping issues between different DRL algorithms (like DQN and PPO) and communication protocols, resulting in global policy conflicts if independently trained models yield inconsistent action formats.

DRL method may achieve deep coupling with 6G intelligent hypersurface (RIS), terahertz beam fouling and other technologies in the future will promote the evolution of wireless networks to the"cognitive intelligence"paradigm of environment sensing and autonomous decision making.

An example will be given to show how DRL work. The network consists of N cells, each of which contains a centrally located Base Station (BS) and M User Equipment (UE) randomly distributed in the cell coverage area. All BSs share the flat fading spectrum resources to provide services to M UEs. All base stations and users are equipped with a single antenna, and all BSs multiplex a single spectrum resource in the same time slot. The set of BSs and UEs can be denoted as $\mathrm{N} = \{1,2,3,\ldots,\mathrm{N}\}$ and $\mathrm{M} = \{1,2,3,\ldots,\mathrm{M}\}$, respectively. A server acting as a central trainer is located in the cloud, which is responsible for the training and distribution of the BS power allocation policies. The whole network adopts a fully synchronized time system: at the beginning of each time slot, each BS will associate multiple UEs within its own cell for power allocation, and each UE is associated with the BS of the cell it is currently in by default. in addition, the UEs will move randomly within the cell at the beginning of each time slot. The above proposed power allocation problem is modeled as a Markov Decision Process (MDP) and a Multi-Agent Deep Reinforcement Learning (DRL) approach is used to achieve dynamic power allocation. Specifically, each base station (BS) is considered as an independent intelligence and individualized State, Action and Reward functions are designed for it. The core elements of the MDP-based power allocation model are described below.

(1) State S: The state is the basic information required by the intelligent body to perceive the environment, but if the state information is too large and redundant, it will be difficult for the intelligent body to quickly and accurately extract effective features. Therefore, the state designed for intelligent body n in this paper is defined as:

$$\mathrm{s}(t) = \{G_n(t), P_n(t-1), C_n(t-1), G_{i\in D}(t-1), C_{i\in D}(t-1)\} \tag{1}$$

In addition to the underlying channel states, two new features are introduced in this paper to enhance the decision-making capability of the intelligences. Specifically, $\mathrm{G_n(t)}$ denotes the normalized channel gain of the current cell n, and $\mathrm{P_n(t-1)}$ and $\mathrm{C_n(t-1)}$ record the transmit power and transmission rate of the cell in the previous time slot, respectively. By fusing the historical state information of the neighborhood, the intelligent body needs to weigh its own channel conditions and the dynamic interference level of the neighboring networks during action selection, thus avoiding the global performance degradation triggered by the local greedy strategy.

(2) Action A: The transmit power is discretized into |A| levels, from which the intelligent body needs to select the optimal action. Under the maximum power constraint $\mathrm{P_{max}}$, the set of available discrete powers is defined as:

$$a_t = \left\{0, \frac{P\max}{|A|-1}, \frac{2P\max}{|A|-1}, \ldots, P\max\right\}, \tag{2}$$

(3) Reward R: The environment provides feedback to the intelligent body on the merits of its actions through a reward mechanism to drive the policy update. The reward function designed here is not only affected by the intelligent body's own actions, but also needs to consider the interference effect of neighboring cell base stations (BSs). Specifically, when the neighboring cell BSs use high-power communication, the interference in the current cell will be significantly enhanced. Therefore, the reward function needs to take into account the correlation performance between its own transmission rate and that of neighboring cells. Based on this, this paper defines the reward rt for time slot t as:

$$\sum_{i\in D}\sum_{m=1}^{M}\left(\left(\log_2\left(1+\frac{g_{i,i,m}(t)p_{i,m}(t)}{\sum_{j\neq i,j\neq n}^{D} g_{j,i,m}(t)\sum_{m=1}^{M} p_{j,m}(t)+\sum_{k\neq m}^{M} g_{i,i,m}(t)p_{i,k}(t)+\sigma^2}\right)\right)-C_{i,m}(t)\right) \quad (3)$$

The first term in curly brackets in Eq. represents the sum of the theoretical rates when ignoring the interference of cell n to the neighboring cells, and the second term is the sum of the actual rates of the neighboring cells after the interference. The difference between the two quantifies the interference loss caused by cell n to the neighbor. The penalty term is introduced to constrain the intelligence from over-optimizing the local rate-avoiding the global performance degradation triggered by the decrease in the neighbor's rate due to its over-boosting power. The mechanism provides dynamic feedback on the interference cost and guides the intelligent body to adjust its power strategy in subsequent time slots to achieve a balanced improvement between local optimization and the overall rate of the system.

Similar examples are also given. Paper [14] gives a model-free distributed execution DQN algorithm is proposed to solve the power allocation problem of power output in wireless communication networks. In the paper, each transmitter is modeled as an intelligent body, which collects instantaneous cross cell CSI and QoS at the beginning of each time slot and adjusts its own transmission power. Another method is in paper [15].

4. Conclusion

Deep Reinforcement Learning (DRL) provides breakthrough solutions for dynamic spectrum access and 5G core technology optimization by fusing the high-dimensional state-action space processing capability and end-to-end decision-making mechanism of deep neural networks and Q-learning framework. Its core innovations are reflected in three aspects: first, avoiding complex modeling in dynamic spectrum allocation, realizing utility optimization in multi-intelligence collaboration scenarios through the DQN framework, and solving adaptive sensing and multi-channel dynamic decision-making in unknown Markov environments to approach the theoretical optimum; second, upgrading intelligent architecture in 5G heterogeneous networks-SCMA improves spectral efficiency through real-time channel state-driven codebook scheduling and cross-layer resource allocation, and DQN improves the state-action-space processing capability and end-to-end decision-making mechanism to improve spectrum efficiency. resource allocation to improve spectrum efficiency, D2D constructs a distributed interference coordination mechanism with the help of multi-intelligence collaboration to break through the response delay limitation of traditional centralized scheduling in high-density scenarios, and supports the dynamic optimization of network slicing through a hierarchical reinforcement learning framework; thirdly, designing intelligent architectures with global equilibrium through quantized interference loss and neighborhood historical state fusion, and designing intelligent architectures with global equilibrium and dynamic decision-making intelligent architecture to approach the theoretical optimum. Third, the reward function with global equilibrium characteristics is designed by quantizing the interference loss and neighbor history state fusion, which solves the performance degradation problem caused by the local greedy strategy and provides a scalable autonomous decision-making paradigm for multi-base station power allocation. Its practical value has been verified in terms of capacity enhancement in dense scenarios, range extension of IoT devices, and network self-healing enhancement. Despite the issues of online training data acquisition, model interpretability, protocol standardization and integration, digital twin simulation and interpretable AI can accelerate its engineering landing. In the future, DRL will be deeply integrated with 6G Reflective Surface Intelligence (RIS), terahertz beam fusion, and other technologies to promote the evolution of wireless networks to the cognitive intelligence paradigm of “environment perception-autonomous decision-making-global optimization”.

References

[1] Panda, S. (2020). Joint user patterning and power control optimization of MIMO–NOMA systems. Wireless Personal Communications, 112, 1 - 17.

[2] S. M. R., Avazov, N., Dobre, O. A., et al. (2016). Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges. IEEE Communications Surveys and Tuto.

[3] Liang, X., Liu, Q., Cao, W., et al. (2020). Fairness optimization and power allocation in cognitive NOMA/OMA V2V network with imperfect SIC.

[4] Miuccio, L, Panno, D, Riolo, S, Joint Control of Random Access and Dynamic Uplink Resource Dimensioning for Massive MTC in 5G NR Based on SCMA, IEEE INTERNET OF THINGS JOURNAL.

[5] Han, S., Huang, Y., Meng, W., et al. (2018). Optimal power allocation for SCMA downlink systems based on maximum capacity. IEEE Transactions on Communications, 67 (2), 1480 – 1489.

[6] Yu, S., Ejaz, W., Guan, L., et al. (2017). Resource allocation schemes in D2D communications: overview, classification, and challenges. Wireless Personal Communications, 96 (1), 303 – 322.

[7] Mishra, P. K., Kumar, A., Pandey, S., et al. (2018). Hybrid resource allocation scheme in multi-hop device-to-device communication for 5G networks. Wireless Personal Communications, 103 (3), 2553 – 2573.

[8] Dhilipkumar, S., Arunachala Perumal, C., & Thanigaivelu, K. (2019). A comparative study of resource allocation schemes for D2D networks underlay cellular networks. Wireless Personal Communications, 106 (3), 1075 – 1087.

[9] Tian Di, An Intelligent Optimization Method for Wireless Communication Network Resources Based on Reinforcement Learning

[10] F. M. Talaat, “Effective deep Q-networks (EDQN) strategy for resource allocation based on optimized reinforcement learning algorithm,” Multimedia Tools and Applications, vol. 81, no. 28, pp. 39945 - 39961, 2022.

[11] Feriani, A; Hossain, E, “Single and Multi-Agent Deep Reinforcement Learning for AI-Enabled Wireless Networks: A Tutorial”.

[12] Chang, H. H., Song, H., Yi, Y., et al. (2018). Distributive dynamic spectrum access through deep reinforcement learning: A reservoir computing-based approach. IEEE Internet of Things Journal, 6 (2), 1938 - 1948.

[13] Tan, X., Zhou, L., Wang, H., et al. (2022). Cooperative Multi-Agent Reinforcement-Learning-Based Distributed Dynamic Spectrum Access in Cognitive Radio Networks. IEEE Internet of Things Journal, 9 (19), 19477 - 19488.

[14] A. Nzaldo, A. G. Andrade, "Deep Reinforcement Learning for Power Control in Multi-Tasks Wireless Cellular Networks," in *2022 IEEE International Mediterranean Conference on Communications and Networking (MeditCom)*. IEEE, 2022: 61 - 65.

[15] Y. S. Nasir and D. Guo, "Multi-Agent Deep Reinforcement Learning for Dynamic Power Allocation in Wireless Networks," IEEE Journal on Selected Areas in Communications, vol. 37, no. 10, pp. 2239 - 2250, 2019.