Dqn-Based Resource Allocation for Power Management in Cellular Network

Xiaoyu Li

College of Information Engineering, Shanghai Maritime University, Shanghai, China

Abstract. In recent years, power allocation in cellular networks has become increasingly crucial with the rapid development of mobile communications. This report presents a comparison of different power allocation approaches, including Deep Q-Network (DQN), Policy Gradient, Fractional Programming (FP), and Weighted Minimum Mean Square Error (WMMSE). The study first constructs a grid-based multi-cell downlink cellular network environment that incorporates practical features such as path loss, channel fading, and interference models. Through experiments, the study evaluates these methods in terms of performance, computational efficiency, and stability. The results show that DQN achieves a balanced performance with an average reward of 1.52 and runtime of 316-349 seconds, while WMMSE achieves slightly higher performance (1.58) but requires significantly more computational resources (850 seconds). The findings show that deep reinforcement learning approaches, particularly DQN, offer promising solutions for power allocation in cellular networks, combining strong performance with reasonable computational efficiency. This study aims to explore the impact of different methods in Power Management in Cellular Networks.

1 Introduction

In recent years, with the rapid development of mobile communication technology, power allocation in cellular networks has attracted increasing attention from researchers. Due to the complexity and dynamics of networks, traditional power allocation methods such as Fractional Programming (FP) and Weighted Minimum Mean Square Error (WMMSE) often struggle to adapt to rapidly changing network environments. Therefore, designing intelligent and adaptive power allocation strategies has become an important research direction. This trend is supported by Meng et al., who showed that deep reinforcement learning approaches for power allocation can outperform traditional methods in cellular networks while adapting to changing environments [1]. Reinforcement Learning (RL), as an important branch of machine learning, has been widely applied in automatic control, games, robotics, and other fields. Through the interaction between agents and the environment, it learns optimal strategies through trial and error, enabling autonomous decision-making in complex environments. As demonstrated by Fang et al. [2] in network resource allocation contexts, RL approaches can effectively address complex optimization problems. In particular,

202210310193@stu.shmtu.edu.cn

algorithms such as Deep Q-Network (DQN) and Policy Gradient have shown great potential in handling decision-making problems in high-dimensional state spaces. Xiong et al. [3] further demonstrated that DQN-based strategies can effectively optimize resource allocation in dynamic wireless networks, achieving near-optimal performance. Therefore, applying reinforcement learning to power allocation optimization in cellular networks can make the allocation system more intelligent, reduce human intervention, and improve system performance.

This paper aims to study the application of DQN and Policy Gradient in cellular network power allocation and compare them with traditional FP and WMMSE methods. The study first constructed a reinforcement learning environment with historical data and set reasonable allocation rules. Then, DQN and Policy Gradient were used to train intelligent agents to learn optimal allocation strategies. Through experimental analysis, paper evaluated the performance of the reinforcement learning models and compared them with traditional strategies to verify their effectiveness. Finally, paper summarize the advantages and challenges of various methods in power allocation and explore future research directions.

The remainder of this paper is organized as follows: Section 2 provides a comprehensive literature review covering traditional power allocation methods and reinforcement learning applications in network resource management. Section 3 details the methodology, presenting the mathematical foundations and implementation specifics of DQN, Policy Gradient, FP, and WMMSE approaches. Section 4 describes the experimental setup, including the system model, network configuration, and presents the comparative results in terms of performance metrics, computational efficiency, and stability analysis. Finally, Section 5 concludes the paper with a summary of findings and suggestions for future research directions in power allocation for cellular networks.

2 Literature Review

2.1 Traditional Power Allocation Methods

In the field of cellular network power allocation, traditional algorithms mainly include Fractional Programming (FP) and Weighted Minimum Mean Square Error (WMMSE) methods. These methods are based on mathematical modeling and optimization theory, attempting to find optimal power allocation schemes under given constraints. The FP method seeks optimal solutions through iterative updates, featuring good theoretical convergence and high stability. The WMMSE method optimizes system performance by introducing weights, with the advantages of relatively simple implementation and near-optimal performance. Chowdhury et al. further developed the WMMSE approach by unfolding the algorithm into a neural network architecture parameterized by graph neural networks, achieving a balance between computational efficiency and optimal performance [4]. Qi and Guo proposed an energy-saving strategy for 4G/5G networks using an improved FP-Tree algorithm, addressing the significant power consumption challenges in 5G deployment. However, these traditional methods often have high computational complexity and struggle to adapt to scenarios with high real-time requirements [5].

2.2 Application of Reinforcement Learning in Power Allocation

Recent research has shown that reinforcement learning-based methods have significant advantages when dealing with dynamic, non-stationary network environments. Meng et al. conducted a comprehensive comparison of different deep reinforcement learning approaches for power allocation in multi-user cellular networks [1]. Their work demonstrated that deep

reinforcement learning techniques, particularly Deep Deterministic Policy Gradient (DDPG), can outperform traditional model-based methods in terms of sum-rate performance while maintaining lower computational complexity.

Reinforcement learning, especially deep reinforcement learning, has shown promising application prospects in power allocation problems. Compared with traditional methods, reinforcement learning methods are data-driven and do not require precise system modeling. Through continuous interaction with the environment, reinforcement learning algorithms can adaptively learn optimal strategies. Similarly, Shi et al. [6] applied hierarchical DRL for resource allocation in mobile networks, confirming the advantages of these methods in dynamic environments.

2.3 DQN and Policy Gradient Methods

DQN is a reinforcement learning algorithm that combines deep learning with Q-learning. It uses deep neural networks to approximate action value functions and can handle decision-making problems in high-dimensional state spaces. DQN improves training stability and efficiency through techniques such as experience replay and target networks, as demonstrated by Pang et al. in their work on remote state estimation systems [7]. Nasir and Guo further showed that DQN approaches can be effectively applied to dynamic power allocation in wireless networks, achieving near-optimal performance [8]. Policy Gradient is another important class of reinforcement learning methods that directly learns policy functions and is suitable for problems with continuous action spaces. Sharara et al. demonstrated the effectiveness of Policy Gradient-based reinforcement learning for computing resource allocation in Open Radio Access Networks (O-RAN), showing that such approaches can achieve near-optimal performance with much lower computational complexity compared to traditional optimization methods [9]. These two methods have their own characteristics in power allocation problems and are suitable for different application scenarios.

3 Methodology

3.1 DQN

DQN, as a deep reinforcement learning method, combines deep neural networks with Q-learning to handle the power allocation problem. The DQN implementation consists of several key components. The state space design incorporates Channel State Information (CSI), normalized interference levels, historical power allocation decisions, and previous transmission rates. This comprehensive state information enables the agent to thoroughly understand the current network environment.

For the action space design, the study discretize the continuous power range [5dBm, 38dBm] into power_num discrete levels. Each action corresponds to a specific transmission power value, and the agent selects one power level as its action at each time step. As suggested by Meng et al., this discretization approach provides a balance between implementation simplicity and performance optimization [1]. The network architecture consists of an input layer that receives the state vector, hidden layers with fully connected layers using ReLU activation, and an output layer that produces Q-values corresponding to each possible action.

The learning process employs an experience replay buffer to store transition samples (st, at, rt, st+1), utilizes an ε-greedy strategy for exploration, updates network parameters by minimizing TD error, and implements a target network to maintain training stability. The study's DQN implementation draws from the experience replay and target network design

principles introduced by Cheng et al. to enhance training stability and convergence speed [10].

The Q-learning update rule used in DQN:

$$Q(s,a) = Q(s,a) + \alpha[r + \gamma \cdot \max'_a Q(s',a') - Q(s,a)] \quad (1)$$

The loss function for training the neural network:

$$L(\theta) = E\left[\left(r + \gamma \cdot max'_a Q(s',a';\theta^-) - Q(s,a;\theta)\right)^2\right] \quad (2)$$

3.2 Policy Gradient

The Policy Gradient method directly learns the policy function by optimizing the policy network parameters to maximize expected returns. The main update formula is as follows:

Update Rule:

$$\nabla_\theta J(\theta) = E_\pi[\nabla_\theta \log \pi(a|s;\theta)\, Q(s,a)] \quad (3)$$

The policy network parameters θ define the policy function π(a|s;θ), which gives the probability of selecting action a in states. Q(s,a) evaluates the expected return of state-action pairs, while ∇θ represents the gradient used to update the parameters during optimization.

Output:

$$\pi(a|s;\theta) = softmax\big(f(s;\theta)\big) \quad (4)$$

Where $f(s;\theta)$ is the output layer of the policy network.

The expected reward objective function is:

$$J(\theta) = E_\pi[\sum r_t] \quad (5)$$

The REINFORCE algorithm specifically optimizes:

$$\nabla_\theta J(\theta) = E_\pi[\nabla_\theta \log \pi(a|s;\theta) \cdot G_t] \quad (6)$$

where G_t represents the cumulative reward from time step t.

Meng et al. [1] demonstrated that policy gradient methods like REINFORCE can achieve strong performance in power allocation tasks, though they typically exhibit higher variance and less stable convergence compared to DQN.

3.3 FP And WMMSE

Fractional Programming (FP) is an important method for solving power control and beamforming problems. In multi-cell network power allocation, the objective is to maximize the sum-rate:

$$\mathrm{C}(\mathrm{g}^{\mathrm{t}}, \mathrm{p}^{\mathrm{t}}) \coloneqq \sum_{\{\mathrm{n,k}\}} \mathrm{C}^{\mathrm{t}}_{\{\mathrm{n,k}\}} \quad (7)$$

the rate of each user is:

$$\mathrm{C}^{\mathrm{t}}_{\{\mathrm{n,k}\}} = \log_2\big(1 + \gamma^{\mathrm{t}}_{\{\mathrm{n,k}\}}\big) \quad (8)$$

This is a non-convex NP-hard optimization problem. FP solves it by transforming it into a series of convex subproblems.

This is a non-convex NP-hard optimization problem. FP solves it by transforming it into a series of convex subproblems.

WMMSE takes a different approach by iteratively minimizing the weighted mean square error to solve the power allocation problem. It introduces receiver weights and equalizers as auxiliary variables and obtains local optimal solutions through alternating optimization. Chowdhury et al. simplified WMMSE's computational demands by transforming the iterative algorithm into a neural network using graph neural networks to capture wireless network topology [4].

Each iteration consists of three steps where first the power is fixed to optimize equalizers, then both power and equalizers are fixed to optimize weights, and finally the weights and equalizers are fixed to optimize power.

Mathematically, these steps can be represented as:

$$w_k = (1 + SINR_k)^{-1} \tag{9}$$

$$u_k = h_k * \frac{p_k}{\sum_j h_j * p_j + \sigma^2} \tag{10}$$

$$p_k = \left[\lambda^{-1} \cdot w_k \cdot u_k \cdot \frac{h_k}{\sum_j w_j \cdot u_j^2 \cdot h_j}\right]^+ \tag{11}$$

Where w_k is the weight, u_k is the equalizer, and p_k is the power allocation for user k.

The main challenges of these two algorithms include their requirement for accurate mathematical models despite hardware and channel imperfections in practical systems, their high computational complexity which limits practical deployment, and their difficulty in accommodating heterogeneous service requirements and dynamically changing environments.

4 Experimental

4.1 Experimental procedure

4.1.1 System Model and Problem Formulation

This study construct a grid-based multi-cell downlink cellular network system to compare and evaluate the performance of different power allocation algorithms. This system model incorporates key features from practical scenarios, including multi-user interference, channel fading, and power constraints.

As noted by Qi and Guo, 5G NR networks consume 3-5 times more power than 4G networks, with higher base station density, making energy efficiency an important consideration alongside performance optimization [5].

4.1.2 Network Model: Research consider a grid cellular network consisting of $n_x \times n_y$ cells. Each cell is equipped with a base station (BS) that can serve up to maxM users. The total

number of users in the system is M = $n_x \times n_y \times$ maxM. Within each cell, users are randomly distributed at distances ranging from $\min_{dis}$ to $\max_{dis}$ kilometers from their serving BS, with angular positions uniformly distributed between -π and π radians. This setup simulates typical base station and user distributions in practical cellular networks.

4.1.3 Channel Model: To accurately simulate the wireless propagation environment, the channel model incorporates three main components. First, the standard path loss model is adopted to represent signal attenuation over distance, defined as:

$$PL(d) = 120.9 + 37.6 \log_{10}(d)\, dB \tag{12}$$

Where d represents the distance between the transmitter and receiver in kilometers.

Second, the Jakes model is used to describe time-varying Rayleigh fading, which captures the small-scale fading effects. The channel temporal correlation is characterized by the coefficient:

$$\rho = J_0(2\pi f_d T_s) \tag{13}$$

Third, log-normal shadowing with 8dB standard deviation is considered to model large-scale fading, accounting for signal attenuation caused by terrain and buildings.

4.1.4 Interference Model

The interference model adopts a set-based approach to evaluate system performance with both accuracy and efficiency. The paper defines parameter L to establish the interference region range for each cell. This results in c = 3L(L+1) + 1 adjacent interfering base station and a total of K = maxM × c interfering users. This comprehensive model accounts for both intra-cell and inter-cell interference effects, effectively capturing the primary interference sources while maintaining computational tractability.

4.1.5 System Parameters amd Constraints

The system operates under specific parameters and constraints. Transmission power ranges from 5dBm to 38dBm, while noise power is set at -114dBm (AWGN). The SINR (Signal-to-Interference-plus-Noise Ratio) is calculated as the ratio of desired signal power to the sum of interference power and noise power. For numerical stability, the paper establishes an upper limit of maxC for SINR values. The system utilizes a bandwidth of 10MHz for operation.

4.1.6 Optimization Objective

The system's optimization objective is to maximize the sum rate while satisfying power constraints:

$$maximize: R = E\left[\sum\left(log_2(1 + SINR)\right)\right] \tag{14}$$

Subject to:

$$0 \le P \le P_{max} \tag{15}$$

where each user's achievable rate is calculated using the Shannon formula

$$R = W \times \log_2(1 + SINR) \tag{16}$$

With W being the bandwidth weight factor.

Table 1 shows the system model parameters and configuration.

Table 1: System Model Parameters and Configuration.

Category	Parameter	Value
Network Configuration	Grid Size	5 × 5 cells
	Users per Cell (maxM)	4
	Total Users (M)	100
Channel Parameters	Path Loss Model	120.9 + 37.6log10(d) dB
	Doppler Frequency (fd)	10 Hz
	Time Slot Duration (Ts)	20ms
	Shadowing	8Db std
	Distance Range	0.01-1 km
Power Settings	Maximum Power	38 dBm
	Minimum Power	5 dBm
	Noise Power	-114 dBm
	Power Levels	10
Training Parameters	Maximum Episodes	5000
	Batch Size	500
	Buffer Size	50000
	State Dimension	3C + 2 (C=16)
	Initial ε	0.2
	Final ε	0.0001
	Learning Rate	1e-3
Interference Model	Region Parameter (L)	2
	Adjacent BSs (c)	3L(L+1) + 1
	Max Interfering Users (K)	maxM × c

4.2 Performance Comparison Between DQN and Policy Gradient

The comparison between DQN and Policy Gradient approaches revealed distinct learning characteristics and performance trends. Over 5000 training episodes, DQN demonstrated more stable learning behavior but with a relatively slower initial learning rate. In contrast, Policy Gradient exhibited faster initial learning but showed higher variance in performance. The final average rewards achieved by DQN ranged from 1.374 to 1.516, while Policy Gradient reached slightly higher values between 1.572 and 1.597. Figure 1 shows the performance comparison of DQN and policy gradient.

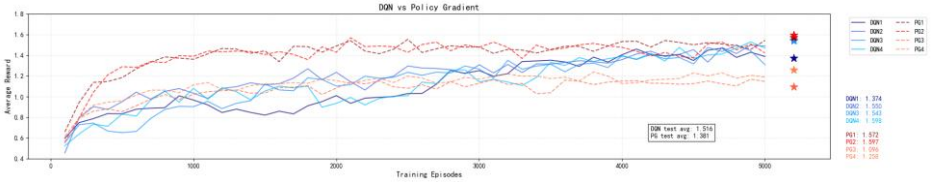

Figure 1: Performance Comparison of DQN and Policy Gradient. (Picture credit: Original)

4.3 Comprehensive Method Comparison

Through comprehensive evaluation of DQN, Policy Gradient, FP (Fractional Programming), and WMMSE methods, the paper observed significant performance differences. In terms of performance metrics, WMMSE achieved the highest average reward of approximately 1.58, followed by DQN with a strong performance of about 1.52. Policy Gradient achieved a moderate performance level of around 1.38, while FP, although showing relatively lower performance (approximately 1.32), demonstrated the fastest execution speed.

Regarding computational efficiency, the experiment showed that WMMSE required the longest computation time, taking about 850 seconds to complete a full run. In stark contrast, the FP method needed only about 200 seconds to complete calculations. DQN and Policy Gradient maintained a balance between these extremes, exhibiting moderate computational requirements. Figure 2 shows the average reward comparison.

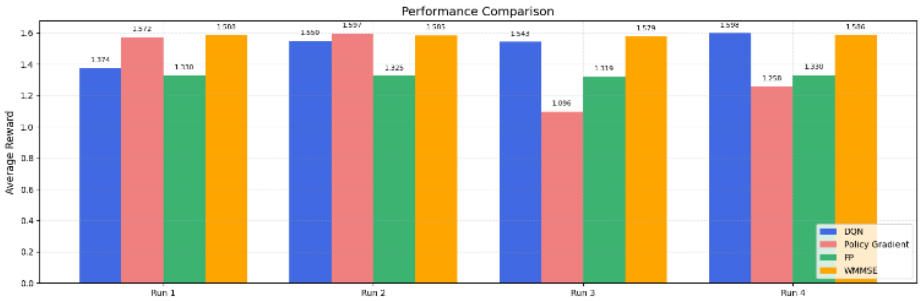

Figure 2: Average Reward Comparison. (Picture credit: Original)

4.4 Runtime

In four experimental runs, the paper conducted an in-depth analysis of computational efficiency across different methods. The results showed that the Policy Gradient method consistently required longer execution times, ranging from 473 to 511 seconds. In contrast, DQN demonstrated significantly higher processing efficiency with runtime between 316 and 349 seconds. The average time difference of approximately 150-170 seconds between these methods clearly demonstrates DQN's advantage in computational efficiency. Figure 3 shows the runtime comparison of all methods.

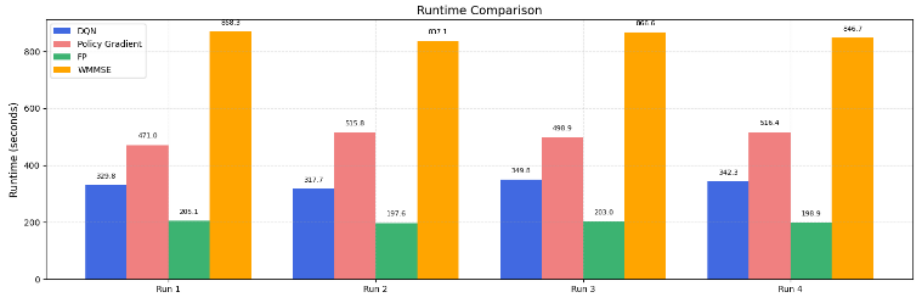

Figure 3: Runtime Comparison of All Methods. (Picture credit: Original)

4.5 Comprehensive Method Comparison

Table 2: Performance-Efficiency Trade-off Analysis.

Evaluation Metric	DQN	Policy Gradient	FP	WMMSE
Stability	4/5	3/5	5/5	5/5
Performance	4/5 Average~1.52	3/5 Average~1.4	3/5 Average~1.32	5/5 Average~1.58
Runtime	4/5 ~330 seconds avg	3/5 ~330 seconds avg	5/5 ~200 seconds Avg	1/5 ~850 seconds avg
Progress	4/5	4/5	1/5	1/5
Total Score	16/20	13/20	14/20	12/20

Table 2 shows the performance-efficiency trade-off analysis. Through experimental analysis, the paper observed significant trade-offs between performance and computational efficiency across different methods. DQN found the optimal balance in this trade-off, delivering good performance at a reasonable computational cost. While WMMSE achieved optimal performance, its significant computational overhead limited its practical applications. The FP method excelled in rapid execution but showed relatively weak performance. Policy Gradient demonstrated competitive performance, despite requiring longer training times.

Notably, FP and WMMSE, as traditional model-based methods, have reached their theoretical performance limits and are unlikely to achieve further improvements through optimization. In contrast, deep reinforcement learning-based methods like DQN and Policy Gradient still have significant room for improvement. Through enhanced network structures, optimized training strategies, or adoption of more advanced learning algorithms, these methods have the potential to achieve better performance. This scalability and potential for improvement give deep reinforcement learning methods greater advantages in practical applications.

5 Conclusion

Experiments comparing DQN, Policy Gradient, FP, and WMMSE methods revealed trade-offs between performance and computational efficiency.

DQN emerged as a solution that this study identifies as promising, achieving strong performance (average reward ~1.52) with reasonable computational costs (316-349 seconds). While WMMSE achieved slightly higher performance metrics (~1.58), its substantial computational requirements (~850 seconds) limit practical applications, according to this research. Traditional model-based approaches like FP and WMMSE have reached their theoretical limits, while this study suggests that DRL-based methods (DQN and Policy Gradient) show potential for further improvement. The research indicates that the adaptability of DQN to dynamic network conditions represents an advantage over traditional methods, particularly in scenarios where network parameters fluctuate frequently due to user mobility and varying traffic patterns.

The evaluation across multiple metrics demonstrated that this paper finds DQN's overall capability to be strong, suggesting it offers a practical balance for real-world cellular network applications. This research indicates that DRL approaches represent an advancement in power allocation for cellular networks, combining performance with efficiency and future enhancement potential. Future research could explore hybrid approaches that combine the

theoretical guarantees of traditional methods with the adaptability and efficiency of reinforcement learning techniques, potentially leading to more robust power allocation strategies for next-generation cellular networks. Additionally, the integration of these methods with other network optimization techniques could further enhance overall system performance and resource utilization in complex heterogeneous network environments.

References

1. Meng, F., Chen, P., Wu, L., & Cheng, J.: 'Power allocation in multi-user cellular networks: Deep reinforcement learning approaches', IEEE Trans. Wireless Commun., 2020, 19, (10), pp. 6255–6267.
2. Fang, C., Zhang, T., Huang, J., Xu, H., Hu, Z., Yang, Y., ... & Luo, X.: 'A DRL-driven intelligent optimization strategy for resource allocation in cloud-edge-end cooperation environments', Symmetry, 2022, 14, (10), 2120.
3. Xiong, K.: 'Research on resource allocation issues in wireless virtual networks based on deep reinforcement learning' (Master's thesis, University of Electronic Science and Technology of China), 2019.
4. Chowdhury, A., Verma, G., Rao, C., Swami, A., & Segarra, S.: 'Unfolding WMMSE using graph neural networks for efficient power allocation', IEEE Trans. Wireless Commun., 2021, 20, (9), pp. 6004–6017.
5. Qi, R., & Guo, X.: 'Analysis of intelligent energy saving strategy of 4G/5G network based on FP-tree', Procedia Comput. Sci., 2022, 198, pp. 486–492.
6. Shi, W., Li, J., Wu, H., Zhou, C., Cheng, N., & Shen, X.: 'Drone-cell trajectory planning and resource allocation for highly mobile networks: A hierarchical DRL approach', IEEE Internet Things J., 2020, 8, (12), pp. 9800–9813.
7. Pang, G., Liu, W., Li, Y., & Vucetic, B.: 'DRL-based resource allocation in remote state estimation', IEEE Trans. Wireless Commun., 2022, 22, (7), pp. 4434–4448.
8. Nasir, Y. S., & Guo, D.: 'Multi-agent deep reinforcement learning for dynamic power allocation in wireless networks', IEEE J. Sel. Areas Commun., 2019, 37, (10), pp. 2239–2250.
9. Sharara, M., Pamuklu, T., Hoteit, S., Vèque, V., & Erol-Kantarci, M.: 'Policy-gradient-based reinforcement learning for computing resources allocation in O-RAN', in: 2022 IEEE 11th Int. Conf. Cloud Networking (CloudNet), IEEE, 2022, pp. 229–236.
10. Cheng, M., Li, J., & Nazarian, S.: 'DRL-cloud: Deep reinforcement learning-based resource provisioning and task scheduling for cloud service providers', in: 2018 23rd Asia South Pacific Design Autom. Conf. (ASP-DAC), IEEE, 2018, pp. 129–134.