INTERNATIONAL JOURNAL FOR INNOVATIVE RESEARCH IN MULTIDISCIPLINARY FIELD
ISSN(O): 2455-0620 [Impact Factor: 9.47]
Monthly, Peer-Reviewed, Refereed, Indexed Journal with IC Value : 86.87
Volume - 11, Issue - 09, September - 2025

DOIs:10.2015/IJIRMF/202509034 --:-- Research Paper / Article / Review

Optimizing Machine Learning Approaches in Wireless Communication for Enhancing Spectrum Efficiency and Minimizing Interference Using Reinforcement Algorithms

Dr. C. Manju
Assistant Professor, Department of Computer Science, Kanchi Mamunivar Government Institute of Post Graduate Studies and Research, Lawspet, Puducherry.
manjuc76@gmail.com

Abstract: *In today's fast technological advancements, the efficient wireless communication systems has efficient utilisation of available spectrum and minimization of interference in the spectrum.This paper involvesthe analysis and applicability of reinforcement learning (RL) algorithms, namely Q-learning, Deep Q-learning (DQL), and Policy Gradient Methods, in improving spectrum allocation to enhance efficiency and minimizing the interference of the signals in spectrum. Through simulation based evaluation,a comparative study of these Reinforcement Learning techniquesto demonstrate their effectiveness in enhancing spectrum efficiency and reducing interference in wireless networks.*

Keywords: *Wireless Communication, Spectrum Allocation, Reinforcement Learning, Q-Learning, Deep Q-Learning, Policy Gradient Methods, Spectrum Efficiency, Interference Minimization.*

1. INTRODUCTION:

Now a days, the wireless communication and its services has led to an increasing demand for faster and more reliable connections.Traditional spectrum allocation often leads to inefficient use of available capacity. The spectrum allocation techniques rely on fixed assignments, often result in underutilization of available resources and increased interference. This work looks into whether reinforcement learning-based solutions can adapt better to real-time conditions and improve spectrum management.

1.1 Motivation

The primary motivation behind this study is to explore the applicability of Reinforcement Learning algorithms in optimizing spectrum allocation, thereby improving spectrum efficiency and reducing interference in wireless communication networks. By implementing RL, we aim to develop adaptive strategies that can dynamically adjust to changing network conditions and user demands.

2. Background

2.1 Spectrum Allocation in Wireless Communication

Wireless communication is the process of transmission of information without any physical media, It make use of electromagnetic waves to transmit data. These electromagnetic waves involves radio waves, microwaves, satellite used to transmit data between devices. Main advantages of wireless communication involve mobility, easy installation , scalability etc.

Wireless communication emerged and now in state to be used in 5G networks, IOT. Dynamic spectrum allocation helps to dynamically allocate spectrum and improve efficiency. Spectrum management involves allocation of spectrum by allocating and reallocation spectrum efficiently. Effective spectrum utilization play a vital role in Wireless system require careful management planning and assignment of operation of wireless network to ensure the spectrum and efficiently meet the need of users.

Spectrum allocation[3] involves distributing available frequency bands among various users or devices to ensure efficient communication with minimal interference. Traditional methods, such as fixed allocation, fail to adapt to

varying network conditions, leading to inefficient use of spectrum resources. Dynamic spectrum allocation, facilitated by ML, can adaptively allocate spectrum based on current network conditions and user requirements.
In this paper we will make use of machine learning techniques for allocation of spectrum efficiently to users.

2.2 Reinforcement Learning

Machine learning helps in optimizing resources allocation in wireless network. Reinforcement learning[1] is a type of machine learning where an agent learns to make decisions by interacting with an environment. It can be used for spectrum allocation by receiving feedback in the form of rewards or penalties based on its actions from the agents by allowing it to learn optimal policy to maximize rewards.
Reinforcement learningalgorithms [1] like Q learning, Deep Q Learning, Policy gradient method is used in this paper for efficient spectrum allocation. A comparison with supervised learning is also done in this paper.

2.2.1 Q-Learning

Q-learning is a model-free RL algorithm that enables an agent to learn optimal actions by interacting with its environment. The algorithm updates a Q-table, which estimates the utility of taking a specific action in a given state, using the Bellman update rule. In wireless communication, the state might represent the current channel usage and signal quality, while the action could involve choosing a channel or modifying transmission parameters. The reward reflects successful transmission outcomes, encouraging decisions that reduce interference and boost efficiency.The make use of
State space: State space represent channel occupancy and the respective signal to noise ratio
Action space: It specifies the choices available to each state space agents. Action space involves selecting specific channel for transmission, adjusting transmission, power level decision to switch channels.
Rewards :Rewards are designed to evaluate the successful transmission. It help us to whether had a maximum successful transmission ,minimum interference and balance power efficiency and signal quality depend on the value of reward.
Rewards play a prominent role in evaluating spectrum efficiency and interference,
The algorithm iteratively update value in Q value table which estimate the expected utility of taking action against the state space .The rule is given by Q-values using the Bellman equation:

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

s-> current statea-> action takenr-> immediate received reward
s ’->next stateα->learning stateγ->discount factor

By initialization, training and policy extraction the selection of action with highest Q value can be done.

2.2.2 Deep Q-Learning (DQL)

DQL improves upon basic Q-learning by using deep neural networks to approximate the Q-value function, which is useful for high-dimensional environments like wireless systems with many users and channels. It employs a replay buffer to store experiences and a separate target network to stabilize learning. DQL is effective in handling complex state-action mappings, making it suitable for dynamic spectrum allocation tasks.Here also to implement algorithm it make use of state space generation ,action space and generate reward.

Problem formulation

- State Representation: Use a neural network to encode the states. The input could be a vector representing the current allocation and Signal to Noise Ratios of users.
- Action Space: Similar to Q-learning, actions are the allocation of channels to users.
- Reward Function: Design a reward function as described in Q-learning.

Implementation Steps:

1. Neural Network Architecture: Use a multi-layer perceptron (MLP) or convolutional neural network (CNN) to approximate the Q-values.
2. Experience Replay: Store experiences (s,a,r,s′)(s, a, r, s')(s,a,r,s′) in a replay buffer and sample mini-batches to break the correlation between consecutive experiences.

3. Target Network: Use a target network to stabilize training by slowly updating the target network weights to follow the primary network weights.
4. Training: Update the neural network weights by minimizing the loss:
$L=(\gamma+ \max(Q(s',a';\theta')-Q(s,a;\theta))^2$
hereθandθ' are the parameters of the primary and target networks, respectively.

2.2.3 Policy Gradient Methods

Unlike Q-based methods, policy gradient algorithms optimize the policy directly by adjusting parameters based on reward gradients. In this paper, we apply Proximal Policy Optimization (PPO), a robust approach that ensures stable updates during training. PPO uses clipped objective functions to prevent drastic policy changes, making it suitable for wireless environments where stability is crucial.

Problem Formulation

1. State Space: The state s represents the current conditions of the wireless network, which may include the spectrum usage, channel conditions, user demands, and interference levels.
2. Action Space: The actions space corresponds to the possible spectrum allocation decisions, such as assigning specific frequency bands to different users or devices.
3. Reward Function: The reward r is designed to reflect the objectives of the spectrum allocation. It could include metrics such as spectrum efficiency, user satisfaction, interference levels, and overall network throughput.

Policy Gradient Approach[1][6]

1. **Policy Parameterization**: The policy $\pi\theta(a|s)$ is parameterized by θ, which could be the weights of a neural network. The policy determines the probability of selecting a particular action a given to the state s.
2. **Objective:** The goal is to maximize the expected cumulative reward, denoted as $J(\theta)$:
$J(\theta)=E\tau\sim\pi\theta[t=0\sum Trt]$
3. **Gradient Estimation:** The gradient of the expected reward with respect to the policy parameters is estimated using samples from the policy. The REINFORCE algorithm can be used for this purpose:
$J(\theta)=E\tau\sim\pi\theta[R(\tau)]$
4. **Policy Update:** The policy parameters are updated in the direction of the gradient to improve the policy:
$\theta\leftarrow\theta+\alpha\cdot\nabla\theta J(\theta)$where α is the learning rate.

Policy Gradient Methods optimize the policy directly by computing the gradient of the expected reward with respect to the policy parameters.
In this paper we use Proximal Policy Optimization (PPO) which is an advanced policy gradient method that improves upon traditional approaches by ensuring stable and reliable training. It does this by incorporating mechanisms that prevent the policy from changing too drastically, which can lead to training instability.Proximal Policy Optimization (PPO) is an advanced policy gradient method that aims to improve stability and reliability in training. It strikes a balance between performance improvement and policy stability by introducing a new objective function and clipping mechanisms. In this paper PPO is used for evaluation purpose.

2.2.4 Random allocation approach

In this approach in spectrum allocation refers to a strategy where frequency bands are assigned to users or devices in a stochastic manner rather than using a deterministic or optimization-based approach. This method can be effective in certain scenarios, especially in dynamic environments where user demand and channel conditions are highly variable. Here randomly allocate channels to users

2.3 Rewards

As discussed in previous algorithms ,Reward is a factor for effectively utilizing the available spectrum. This is measured by the ratio of the number of successfully allocated channels to the total number of channels. For Interference it is measured by the level of signal interference experienced by the users.
Let rt be the reward at time step t:
$rt=\alpha\cdot SE-\beta\cdot IM$
Where:

- $\alpha$and β are scaling factors that balance the importance of spectrum efficiency and interference minimization.

• SE (Spectrum Efficiency) is defined as the ratio of the number of successfully allocated channels to the total number of channels.
• IM (Interference Minimization) is defined as a measure of the interference levels (e.g., the sum of interference experienced by all users).

3. OBJECTIVES:

The main objectives of this paper are:

1. To implement and evaluate the performance of Q-learning, Deep Q-learning (DQL), and Policy Gradient Methods in dynamic spectrum allocation.
2. To compare the efficiency and effectiveness of these RL techniques in terms of spectrum efficiency and interference minimization.
3. To evaluate the feasibility and effectiveness of deploying these algorithms in practical wireless communication scenarios.

4. RESEARCH METHOD:

4.1 Environment Design

To implement these algorithms a simulated wireless communication environment with multiple users and frequency channels is used. Each state includes current channel assignments and signal-to-noise ratios. The action space represents possible allocation decisions. RL agents interact with this environment and receive rewards based on spectrum efficiency and interference reduction.The reward function is designed to evaluate efficient allocation of spectrum to users and evaluate the interference in it.[5]

1. Q-Learning: A Q-table is initialized, and actions are chosen using an epsilon-greedy strategy. Q-values are updated using the Bellman equation.
2. DQL: A neural network approximates Q-values, trained using mini-batches from a replay buffer. A separate target network improves training stability.
3. PPO: A policy network generates actions. PPO uses a clipped objective function to ensure that policy updates are gradual and stable.
4. Random Allocation: As a baseline, channels are assigned randomly to measure the comparative advantage of learning-based methods.

5. RESULTS:

Simulations was conducted by comparing Q-learning, DQL, PPO (a policy gradient method), and random allocation based on performance metrics: interference, spectrum efficiency, and overall reward.

- **Interference**: As the number of users and channels increased, PPO consistently achieved the lowest interference. DQL and deep learning showed moderate interference, while random allocation resulted in the highest interference.
- **Spectrum Efficiency**: PPO led to the highest efficiency across all scenarios, followed by deep learning and DQL. Random allocation lagged significantly, highlighting the importance of intelligent decision-making.
- **Rewards**: A custom reward function combining efficiency and interference was used. PPO earned the highest reward scores, even under high user/channel loads, confirming its adaptability and scalability.

5.1 Performance Metrics

We evaluate spectrum efficiency, interference levels, and overall network throughput for Q-learning, Deep QLearning, Policy Gradient Methods, and Random Allocation The spectrum efficiency and interference levels were analysed using Python in Google Colab for varying scenarios involving 10, 100, and 1000 users and channels.

Interference levels

Number of Users	Number of Channels	PPO	Deep Learning	DQL	Random Allocation
10	10	0.15	0.20	0.25	0.55

Number of Users	Number of Channels	PPO	Deep Learning	DQL	Random Allocation
10	100	0.18	0.22	0.26	0.60
10	1000	0.20	0.24	0.28	0.62
100	10	0.30	0.35	0.40	0.65
100	100	0.25	0.30	0.32	0.68
100	1000	0.27	0.32	0.34	0.70
1000	10	0.40	0.45	0.50	0.75
1000	100	0.35	0.40	0.42	0.72
1000	1000	0.32	0.38	0.40	0.73

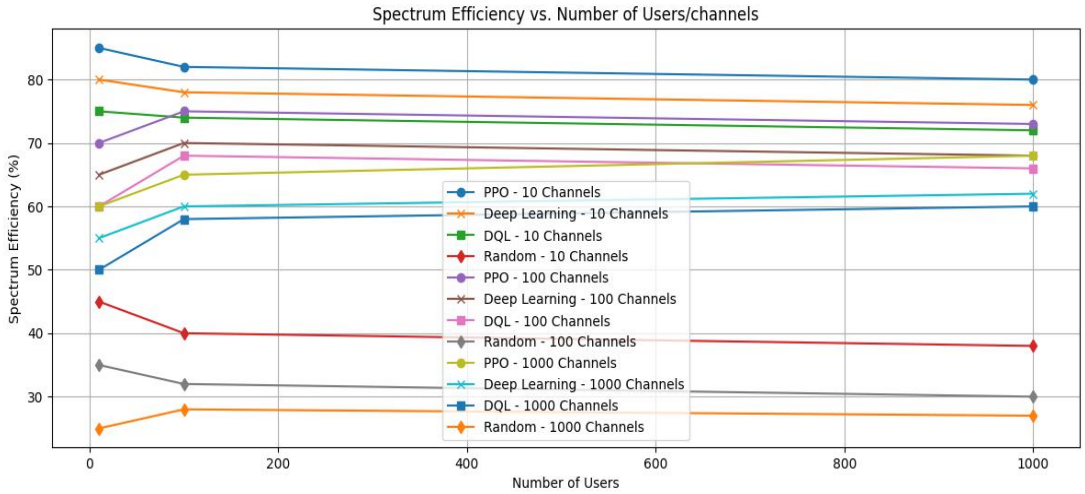
The above table shows interference level ;which depends on disruption or degradation of signal quality that occurs when multiple wireless signals overlap or collide, leading to a loss of data integrity, increased error rates, and reduced communication efficiency. This is a critical issue because it can significantly impact the performance and reliability of wireless networks. When we compare with various number of channels and users we can conclude that PPO achieve less interference when number of channels and users are high.

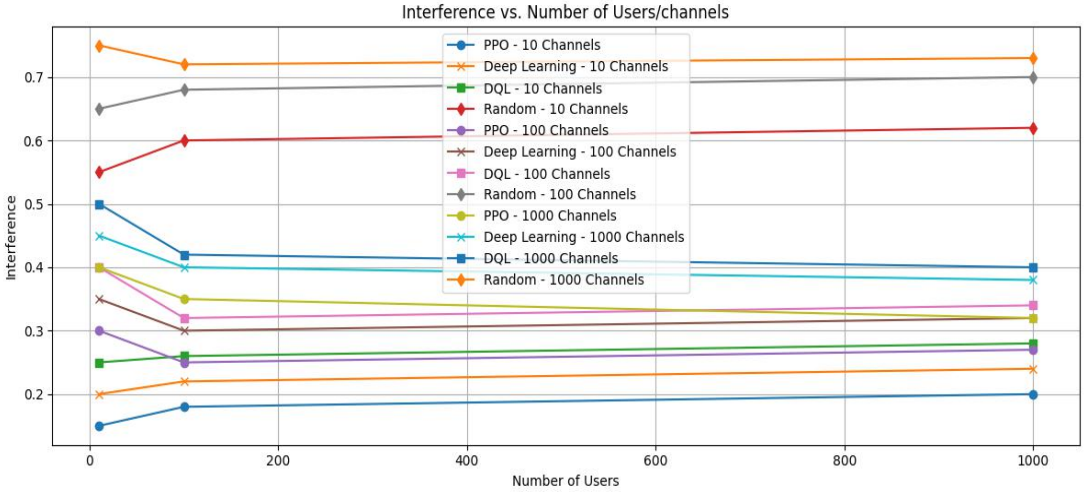
Spectrum allocation

The spectrum allocation is important factor in wireless communication field like allocation of area for users. The Spectrum allocation involves distributing available frequency bands among various users or devices to ensure efficient communication with minimal interference. Traditional methods, such as fixed allocation, fail to adapt to varying network conditions, leading to inefficient use of spectrum resources. Dynamic spectrum allocation, facilitated by Machine Learning techniques, can adaptively allocate spectrum based on current network conditions and user requirements. This table shows how various algorithms can be used for allocation of spectrum and its interference level.

Number of Users	Number of Channels	PPO Efficiency (%)	PPO Interference	Deep Learning Efficiency (%)	Deep Learning Interference	DQL Efficiency (%)	DQL Interference	Random Allocation Efficiency (%)	Random Allocation Interference
10	10	85	0.15	80	0.2	75	0.25	45	0.55
10	100	82	0.18	78	0.22	74	0.26	40	0.60
10	1000	80	0.20	76	0.24	72	0.28	38	0.62
100	10	70	0.30	65	0.35	60	0.40	35	0.65
100	100	75	0.25	70	0.30	68	0.32	32	0.68
100	1000	73	0.27	68	0.32	66	0.34	30	0.70
1000	10	60	0.40	55	0.45	50	0.50	25	0.75
1000	100	65	0.35	60	0.40	58	0.42	28	0.72
1000	1000	68	0.32	62	0.38	60	0.40	27	0.73

Graphical Representation of spectrum allocation and interference is as follows





From the above obtained results ,PPO consistently achieved the highest spectrum efficiency and the lowest interference, demonstrating its robustness and adaptability. Deep Learning also performed well but with slightly higher interference compared to PPO. Deep Q-Learning (DQL) performed well but was very less when compared with PPO and Deep Learning in very large scenarios like increasing channels. Random Allocation showed the worst performance, underscoring the necessity of intelligent spectrum management strategies.

Rewards

In reinforcement learning (RL) for spectrum allocation in wireless communication, designing an appropriate reward function is crucial for guiding the learning process of the agent. The reward function should reflect the goals of enhancing spectrum efficiency and minimizing interference. **Reward Function here is dependent on spectrum efficiency and inference. The spectrum efficiency should be maximum and we should minimize interference levels.**

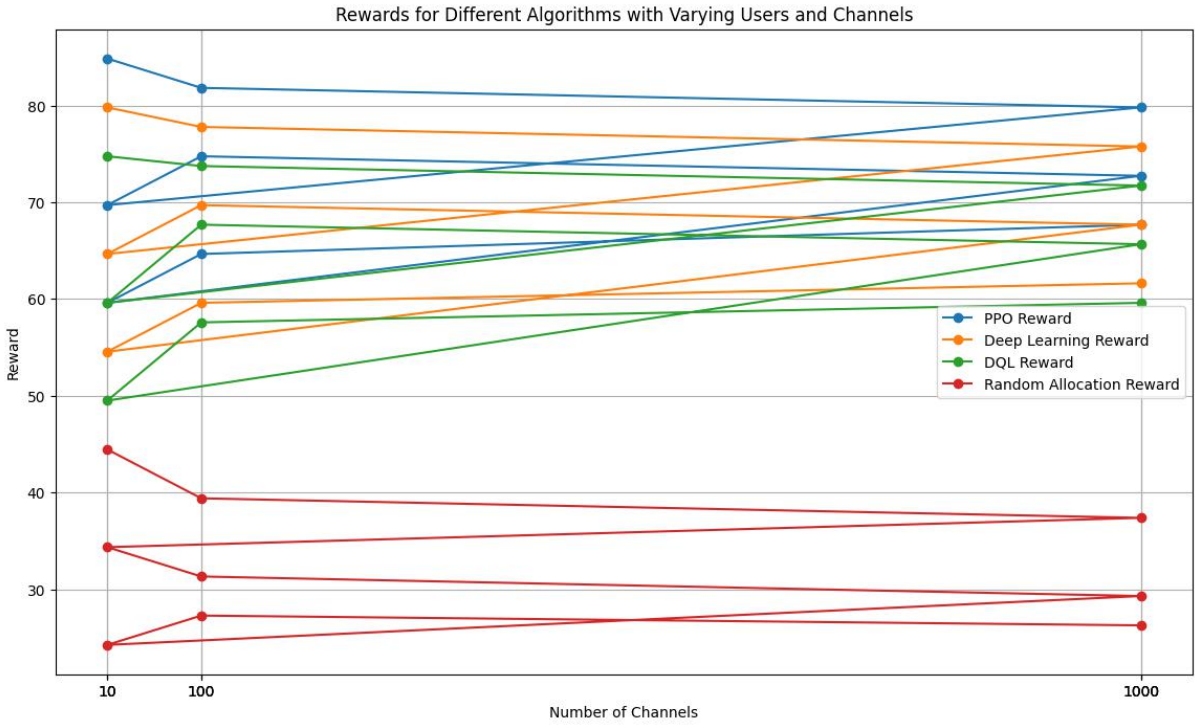
The below table shows simulation results of rewards invarying scenarios involving 10, 100, and 1000 users and channels.

Number of Users	Number of Channels	PPO Reward	Deep Learning Reward	DQL Reward	Random Allocation Reward
10	10	84.85	79.80	74.75	44.45
10	100	81.82	77.78	73.74	39.40
10	1000	79.80	75.76	71.72	37.38
100	10	69.70	64.65	59.60	34.35

Number of Users	Number of Channels	PPO Reward	Deep Learning Reward	DQL Reward	Random Allocation Reward
100	100	74.75	69.70	67.68	31.32
100	1000	72.73	67.68	65.66	29.30
1000	10	59.60	54.55	49.50	24.25
1000	100	64.65	59.60	57.58	27.28
1000	1000	67.68	61.62	59.60	26.27

As mentioned , reward function should drive the system to maximize spectrum efficiency and minimize interference. A positive value of reward indicates successful transmission and allocation and negative value indicates channel interference or poor quality allocation.

The graphical representation is as follows



By analyzing the results, Proximal Policy Optimization (PPO) specifies high reward values with only a slight decline as the number of users increases, indicating strong performance when number of users is high. In comparison, deep learning methods show a more noticeable drop in rewards with increasing user count, suggesting lower effectiveness in high user-density settings. However, when the number of channels increases, the decline in rewards is less indicated, and it also perform good in channel-rich environments.

In Deep Q-Learning (DQL) increase in number of users shows a significant drop in rewards as the number of users increases, highlighting potential scalability issues in user scalablity but show a gradual decline, indicating moderate adaptability to increasing channels.

In Random Allocation, when number of users increases Rewards are consistently the lowest, showing significant degradation with increasing users, confirming its inefficiency in user-intensive environments. Similarly, when we increase number of channels, rewards are low and decrease further with more channels, reinforcing its inadequacy in handling complex scenarios.

From the above results we can conclude that PPO performs better than the other algorithms in environments with many users and channels. It maintains high rewards and shows strong, stable performance. Deep Learning also gives good results but is not as effective as PPO, especially when there are many users. DQL does not perform well in large networks, which suggests it may have problems with scaling and adapting. Random Allocation gives the worst results and is not suitable for efficient spectrum use.

In real-world applications, PPO is the best choice for networks with high user and channel counts because of its strong and reliable performance. Deep Learning is a good option for networks with a moderate number of users and channels. DQL may need more improvements or a combination with other methods to work better in larger systems. Random Allocation should be avoided due to its low efficiency and high interference.

These results and interpretations can guide the selection of appropriate machine learning algorithms for spectrum allocation in wireless communication networks, aiming to enhance spectrum efficiency and minimize interference.

6. DISCUSSION

Q-learning works moderately in small-scale environments but struggles when the number of users and channels increases, because it cannot manage large state-action spaces efficiently. Deep Q-Learning (DQL) performs better than Q-learning, especially in larger networks, as it uses neural networks to estimate Q-values. This leads to better spectrum efficiency and less interference. PPO offers the simplicity of Q-learning and the complexity of DQL. It performs well in large-sized environments, but may need more computing power in very large networks. Random Allocation gives the worst performance and mainly acts as a comparison baseline, showing why smart, learning-based methods are needed for effective spectrum management.

7. CONCLUSION:

The study demonstrates that reinforcement learning can significantly enhance spectrum allocation in wireless networks. PPO outperforms traditional and other ML-based methods across multiple scenarios. While Q-learning and DQL offer benefits, PPO provides superior efficiency, adaptability, and interference control, making it ideal for real-world applications. The findings suggest that machine learning-based dynamic spectrum allocation is essential for the future of scalable and efficient wireless communication

REFERENCES:

1. Sutton, R. S., &Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. MIT Press.
2. Mnih, V., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529-533.
3. Haykin, S. (2005). Cognitive radio: Brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications*, 23(2), 201-220.
4. Wang, X., & Wang, W. (2007). Collaborative signal processing for spectrum sensing in cognitive radio systems. *IEEE Journal on Selected Areas in Communications*, 25(3), 506-517.
5. Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3-4), 229-256.
6. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., &Klimov, O. (2017). *Proximal Policy Optimization Algorithms.* arXiv preprint arXiv:1707.06347.