# Wireless Network Design Optimization for Computer Teaching with Deep Reinforcement Learning Application

Yumei Luo & Deyu Zhang

Published online: 04 Jun 2023.

Submit your article to this journal ↗

Article views: 39

View related articles ↗

View Crossmark data ↗

Taylor & Francis
Taylor & Francis Group

# Wireless Network Design Optimization for Computer Teaching with Deep Reinforcement Learning Application

Yumei Luo and Deyu Zhang

School of International Education, Guizhou Normal University, Guiyang, Guizhou, China

**ABSTRACT**

Computer technology has had a significant impact on the field of education, and its use in classrooms has facilitated the spread of knowledge and has helped students become well-rounded citizens. However, as the number of users accessing wireless networks for computer-based learning continues to grow, the scarcity of spectrum resources has become more apparent, which makes it crucial to find intelligent methods to improve spectrum utilization in wireless networks. Dynamic spectrum access is a critical technology in wireless networks, and it primarily focuses on how users can efficiently access licensed spectrum in a dynamic environment. This technology is a crucial means to tackle the problem of spectrum scarcity and low spectrum utilization. This work proposes a novel approach to address the issue of dynamic channel access optimization in wireless networks and investigates the dynamic resource optimization problem with deep reinforcement learning algorithms. The proposed approach focuses on the optimization of dynamic multi-channel access under multi-user scenarios and considers the collision and interference caused by multiple users accessing a channel simultaneously. Each user selects a channel to access and transmit data, and the network aims to develop a multi-user strategy that maximizes network benefits without requiring online coordination or information exchange between users. What makes this work unique is its utilization of deep reinforcement learning algorithms and a Long Short-Term Memory (LSTM) network that keeps an internal state and combines observations over events. This approach allows the network to utilize the history of processes to estimate the true state, providing valuable insights into how deep reinforcement learning algorithms can be used to optimize dynamic channel access in wireless networks. The work's contribution lies in demonstrating that this approach is an effective means to solve the dynamic resource optimization problem, enabling the development of a multi-user strategy that maximizes network benefits. This approach is particularly valuable as it does not require online coordination or information exchange between users, which can be challenging in real-world scenarios. The proposed approach presents an important contribution to the field of dynamic spectrum access and wireless network

**CONTACT** Yumei Luo ✉ 201407048@gznu.edu.cn 💻 School of International Education, Guizhou Normal University, No. 116 Baoshan North Road, Guiyang, Guizhou 550001, China

optimization. As the demand for computer-based learning continues to increase, the use of intelligent methods to improve spectrum utilization in wireless networks will become even more critical. The findings of this work could have significant implications for the future of computer-based learning and education, enabling more efficient use of wireless networks and the creation of well-rounded citizens.

## Introduction

Computer technology has been applied in various fields of life, especially in the teaching environment. Its influence is not limited to the establishment of computer courses in the school curriculum, but also the influence on learning methods, approaches, habits and even thinking patterns. Computer network teaching is more versatile in terms of teaching methods, and teachers and students can communicate under the network connection. It has the ability to break beyond the boundaries of time and space, which is a distinct advantage over traditional teaching methods. The teacher-student relationship has been bolstered thanks to advances in computer network education. It pays more attention to human-computer interaction for communication and connection, which provides the possibility of updating computer teaching knowledge and diversifying teaching modes. The computer teaching network is more prominent in the characteristics of comprehensiveness and sharing. In actual teaching, when computer teaching is connected to the Internet, the content of teaching can be directly shared, and dynamic teaching with pictures and texts can be realized, which is of great help for students to understand the content of computer knowledge. With the support of network technology, the content of computer teaching can also be presented richly. And rich teaching content can also be shared instantly, allowing students to receive the latest knowledge at the first time. This has a positive effect on cultivating students' innovative ability, and also has a promoting effect on the deepening of practical teaching reform (Du et al. 2017; Erbas, Çipuri, and Joni 2021; Hbaci, Ku, and Abdunabi 2021; Nguyen and Santagata 2021; Ran, Kasli, and Secada 2021).

However, the rapid development of computer-oriented teaching technology has prompted the explosive growth of the number of mobile computer communication devices, which leads to the need for wireless networks to provide massive spectrum resources. The application of wireless spectrum resources is not only limited to personal communication services, but also has optimistic prospects in the fields of sensor networks, embedded control systems, and traffic monitoring systems. Due to the wide application of wireless spectrum resources, many unknown new challenges follow. In traditional wireless networks, users can only use dedicated spectrum, and most of the available

spectrum resources are allocated by the government or auctioned to different operators. This large-capacity static allocation scheme greatly limits the development of temporary communication in small frequency bands. A study by Spectrum Working Group showed that the utilization of licensed spectrum varies from 15% to 85% in different time periods or regions. In particular, in time periods or areas with a small number of users, the allocated spectrum cannot be fully utilized, and a large part of the licensed spectrum is highly vacant. When a licensed user's dedicated spectrum is idle, so-called spectrum cavitation occurs (Hlophe and Maharaj 2021; Liang et al. 2021; Mihovska and Prasad 2021; Randhava, Roslee, and Yusoff 2021).

The following factors should be taken into account while configuring a smart adaptive wireless network in a changing environment. When network model and environmental observability are constrained, how should transmission parameters be configured? How to manage and distribute limited wireless resources to transmitting equipment is the second issue. There are several ways in which conflicting transmission devices might affect network convergence. Spectrum resource management is a fundamental function of a wireless network. Users can access high-quality services through a set of channels or resource blocks that are available in the spectrum. Power control and channel allocation technologies are critical in spectrum resource management because of the ever-increasing demand for mobile data capacity. New methods and ideas for intelligent spectrum resource management have emerged in recent years as a result of the advent of deep reinforcement learning, which combines the model-free qualities of reinforcement learning and the power of deep learning to handle large amounts of data. The following are some potential benefits of using deep reinforcement learning algorithms to optimize spectrum resource management. First and foremost, it allows the wireless network to self-organize and learn effective spectrum resource management solutions through trial and error to find the ideal answer to the decision problem. Second, it is capable of simulating difficult-to-mathematically-model genuine ring mirrors and continuously accumulating fresh experience to adapt to varied severe scenarios. As a third benefit, it can effectively monitor the dynamic environment in real time, mine some potentially useful information and optimize the wireless network performance (Kaur and Kumar 2022; Sekaran et al. 2021; Shah-Mohammadi, Enaami, and Kwasinski 2021; Wu, Jin, and Yue 2022; Zhang, Hu, and Cai 2021).

The proposed approach is centered on optimizing dynamic multi-channel access in scenarios involving multiple users. We take into account the potential for collision and interference that may arise when multiple users access a channel at the same time. In this approach, each user selects a channel to access and transmit data. The network aims to develop a strategy that maximizes network benefits without requiring online coordination or information exchange between users.

The motivation behind this work is driven by the increasing demand for computer-based learning, which has become even more pronounced in recent years due to the COVID-19 pandemic. As students increasingly rely on digital technologies to access educational content, the demand for wireless networks that support these technologies has also grown. This increased demand has highlighted the need for efficient use of spectrum resources in wireless networks.

However, the spectrum resources available for wireless networks are limited and are becoming scarcer as the number of users accessing wireless networks continues to increase. This scarcity of spectrum resources can lead to low spectrum utilization, which can result in slower data transmission rates, longer response times, and reduced overall network performance.

To address this problem, the proposed work aims to optimize dynamic multi-channel access under multi-user scenarios using deep reinforcement learning algorithms. The approach focuses on developing a multi-user strategy that maximizes network benefits without requiring online coordination or information exchange between users, providing an efficient means of spectrum utilization.

The work's motivation is also driven by the need for more effective approaches to dynamic spectrum access, which is a critical technology in wireless networks. The proposed approach aims to tackle the dynamic resource optimization problem, a key challenge in dynamic spectrum access. The approach utilizes deep reinforcement learning algorithms and a long short-term memory network, providing a novel approach to optimizing dynamic channel access in wireless networks.

Overall, the motivation behind this work is to address the growing need for efficient use of spectrum resources in wireless networks to support the increasing demand for computer-based learning. The proposed approach provides a novel solution to the dynamic resource optimization problem, which can enable the development of more efficient and effective wireless networks, benefiting students and educators alike.

## Related Work

Reference (Marinho and Monteiro 2012) introduces spectrum decision-making issues research directions in detail, so that readers can have a general understanding of cognitive radio principles, status quo and future development. Reference (Tragos et al. 2013) conducts a thorough analysis on the choice of spectrum wideband criteria in spectrum allocation problems, different methods of centralized or distributed, and techniques such as heuristics, game theory or fuzzy logic. Reference (Bkassiny, Li, and Jayaweera 2013) classifies cognitive radio problems according to decision and function and expounds them one by one. Furthermore, several serious challenges in non-

Markovian environments and decentralized networks are also considered. The similarities and differences between various algorithms are compared, as well as the application conditions of each technique. Reference (Zhang et al. 2013) elaborates on the application of various auction models in auction theory to spectrum allocation in detail. Reference (Ahmed et al. 2016) compares the advantages of different channel allocation algorithms. In particular, similarities and differences in parameter metrics such as routing dependencies, channel models, allocation methods, execution modeling, and optimization objectives are investigated. Reference (Ahmad et al. 2015) summarizes that the radio resource allocation scheme is divided into three categories: centralized, distributed and cluster-based, and from the perspective of maximizing throughput, ensuring quality of service, avoiding interference to primary users, fairness among secondary users, and prioritization. The six performance optimization criteria of level and spectrum switching are explained. Reference (Wang et al. 2019) discusses the application of reinforcement learning mechanisms for spectrum allocation, classifies and elaborates different reinforcement learning methods. Reference (Liang et al. 2008) proposes a trade-off design between spectrum sensing and spectrum access, which optimizes the length of time for the secondary user to perform spectrum sensing to improve throughput of secondary user system (Kotobi and Bilen 2018). Reference (Peh et al. 2009) optimizes the frequency of the secondary user for spectrum sensing, to improve transmission time and performance while ensuring the priority of the primary user. For cooperative spectrum sensing, literature (Gao et al. 2012) analyzes the malicious behavior of secondary users in cooperative spectrum sensing. For example, sending random sensing results to pretend to participate in cooperative spectrum sensing, sending wrong spectrum sensing results to interfere with other users' decision-making, and pretending to transmit signals for the main user to encroach on the spectrum, etc.

References (Wang et al. 2009, 2015) proposed detection algorithms for malicious spectrum sensing users. However, the above detection algorithms lack an efficient implementation platform and a platform that shares the spectrum sensing reputation value of all secondary users. Reference (Li and Zhu 2018) proposed to let a large number of sensing nodes in network participate, and proposed an incentive mechanism for cooperative spectrum awareness based on social identity as a reward. However, social proof-based rewards are too abstract and may make this incentive less effective due to the difficulty in finding a unit of measure. On the other hand, the existing incentive mechanisms for cooperative spectrum sensing often lack an open and secure implementation platform. Reference (Yi, Cai, and Zhang 2016) proposes and proves that if the primary user divides and provides differentiated spectrum resources according to the different frequency band lengths, time periods and tolerable delays of the secondary user's spectrum. This can more accurately meet the needs of secondary user. Reference (Huang et al.

2015) proposes an auction mechanism to protect the private information such as bidding of secondary users. However, aspects such as transparency and security of secondary spectrum auctions remain to be studied. Reference (Kotobi and G 2018) proposes a blockchain-based spectrum auction to improve the security and efficiency of spectrum auctions. And the simulation results prove that it consumes less energy than the traditional dynamic spectrum access system under the premise of achieving the same spectrum utilization. Reference [32] proposed a blockchain-based spectrum transaction between UAVs and mobile network managers, and designed the behavioral strategies of the two in the transaction through game theory to achieve Nash equilibrium. Reference (Fan and Huo 2020) proposed that in the scenario of network non-real-time data transmission, the combination of consensus algorithm and auction mechanism in blockchain can be used to realize the allocation of spectrum that is not authorized to any user.

## Method

This work focuses on dynamic channel access optimization in wireless networks, and studies the dynamic resource optimization problem with deep reinforcement learning algorithms. This work considers optimization for dynamic multi-channel access under multi-user, and considers the collision and interference caused by multiple users accessing a channel at the same time. The network aims to dig a multi-user strategy maximizing network benefits without the need for online coordination or information exchange between users. This paper applies a deep reinforcement learning-based algorithm for simulation combined with a LSTM.

### *Deep Reinforcement Learning*

Reinforcement learning is a branch of machine learning that focuses on how to act based on feedback from the environment in order to achieve the desired benefit. In psychology, it is based on the theory of behaviorism, which explains how organisms gradually create expectation of stimuli under the influence of rewards or punishments and develop regular behaviors that maximize their benefits. There are several components to the reinforcement learning model, the most basic of which being a set of states in the environment, a set of actions, rules for transitions between states, and immediate rewards following state transitions. The subject and environment of reinforcement learning interact at discrete time steps, and at each time, the subject observes a corresponding piece of information. It usually contains reward information in it, and then it selects an action from a set of actions. Executed in the environment, the environment transitions to a new state, and then gets a reward associated with this transition. The goal of reinforcement learning

agents is to get as many rewards as possible. The power of reinforcement learning comes from two aspects, one is to use the past experience of the subject as a sample to optimize the behavior, and the other is to use the function approximation to simulate the complex system environment. Therefore, reinforcement learning methods are universal and have been studied in many other fields.

Q-learning is known as off-policy temporal difference learning. Different from the temporal difference learning algorithm, Q-learning is a model-based dynamic programming algorithm, so it is necessary to examine the potential reward of each behavior in each learning process of the subject to ensure that the learning process converges. The optimal reward discount sum and Q value update iteration in the Q-learning algorithm are:

$$Q^*(s, a) = r + \alpha * \max Q^*(s, a) \tag{1}$$

$$Q(s_t, a_t) = Q(s_t, a_t) + \beta(r_{t+1} + v * maxQ(s_{t+1}, a) - Q(s_t, a_t)) \tag{2}$$

When the subject accesses the target state, the algorithm terminates an iterative loop within an event. When the inner loop of an event ends, the algorithm restores the system to the initial state and continues to start a new iterative loop. Until the number of cycles between events reaches the set value, the learning ends.

Q-learning generally uses table storage to store the Q value after performing an action in the corresponding state. If the state comes from a very large discrete set of elements, or is simply a continuous vector, there are two problems with storing it in the tabular method. First, too many states cause the table to be too large to store. Second, the sample is too sparse, and the sampling-based algorithm does not converge. In this case, function fitting is a good solution. If the parameterized function is properly expressed, it can also have the function of generalization and reduce the required sample size. Since the parameters of the neural network only include the weights and biases of each layer, it can be used as a good parameterized function in practical systems. DQN is a good example of implementing this method as illustrated in Figure 1.

In the linear approximation process, the value function can be regarded as basis functions and the corresponding parameters are linearly multiplied and calculated, and the value function is a linear function about the basis function. In DQN, the value function is approximated by a neural network, which is a nonlinear approximation. When DQN updates the value function, it actually updates the weights of each layer of the network. Due to the nonlinear fitting of deep neural networks, when the parameters of each layer are determined, the output value can perfectly fit or even represent the value function.

DQN uses experience replay during training for reinforcement learning. Compared with the previous use of neural network for
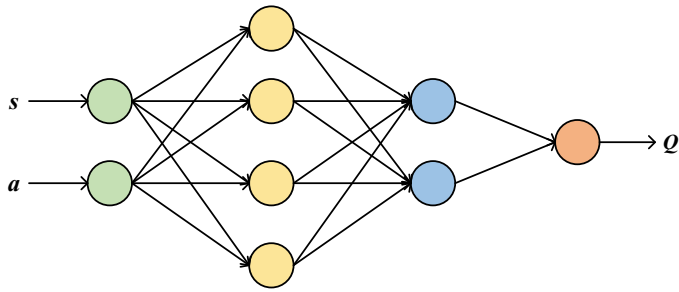
**Figure 1.** The structure of DQN.

reinforcement learning, the phenomenon of gradient explosion or gradient dispersion may occur, the training method using experience replay can make the training of neural network tend to converge and stabilize. Generally, when training a network, it is assumed distribution of data is independent and identically distributed. The sample data collected through reinforcement learning are all observed by the subject according to the feedback of the environment, and there is a correlation between the sample data. Therefore, using these data for sequential training will inevitably bring about the instability of the neural network, and experience playback can break the correlation between the data. The intelligent subject will store the data in the playback memory, and then utilize random sampling to retrieve sample data from the playback memory and use the collected sample data to train the neural network (Bany Salameh, Khader, and Al Ajlouni 2021).

DQN sets the target network on its own to deal with the bias in the TD algorithm on its own. The gradient descent approach is used to update the parameters of DNN when it is utilized to estimate goal Q in DQN, unlike the table-stored Q-learning methodology. Table-type Q-learning directly saves the Q values corresponding to distinct state-action values every time, thus the value function update in DQN genuinely becomes an update process of supervised learning. All actions related to a Q value are calculated using network parameters calculated by the preceding deep neural network, which approximates the value function. It is quite easy for the neural network training process to become unstable because the parameters of the neural network used to determine the goal Q value used to estimate the gradient computation are same. To solve this problem, the DNN that computes the TD target is denoted $\theta^-$, the DNN that computes the approximated target Q-value function is denoted $\theta$. When computing the estimated target Q-value, the deep neural network updates every step, however when computing the TD target, the deep neural network only updates once per set number of steps.

### Dynamic Multi-User Multi-Channel Access with DQN

The multi-user dynamic channel access problem studied in this paper is the dynamic spectrum access problem in cognitive radio. The multi-user model for computer teaching is shown in Figure 2.

This paper considers the problem of dynamic multi-user multi-channel access in a wireless network. The coverage of the base station includes $M$ users and $N$ shared orthogonal channels. The users here are computer equipment for computer teaching. Each user in the coverage area can use a random access protocol to select one of the $N$ shared orthogonal channels for data packet transmission. This paper assumes that each user is backlogged, that is, each user always has packets to transmit. Because there are multiple users, collisions are inevitable. If in a certain time slot, only a user transmits data, then transmission on channel is considered to be successful. Otherwise, if there are multiple users transmitting on the channel, the transmission on the channel will fail, that is, a collision will occur, which will confuse the data packets transmitted by each other, and thus cannot be successfully transmitted. The state definition of the channel is consistent with the definition of dynamic access, and there are good quality, uncertain quality and poor quality. After each user attempts to send a packet in each slot, a binary observation is received indicating whether the packet was successfully sent.

$$A_t^m = 0 \, or \, 1 \tag{3}$$

User action is defined as follows,
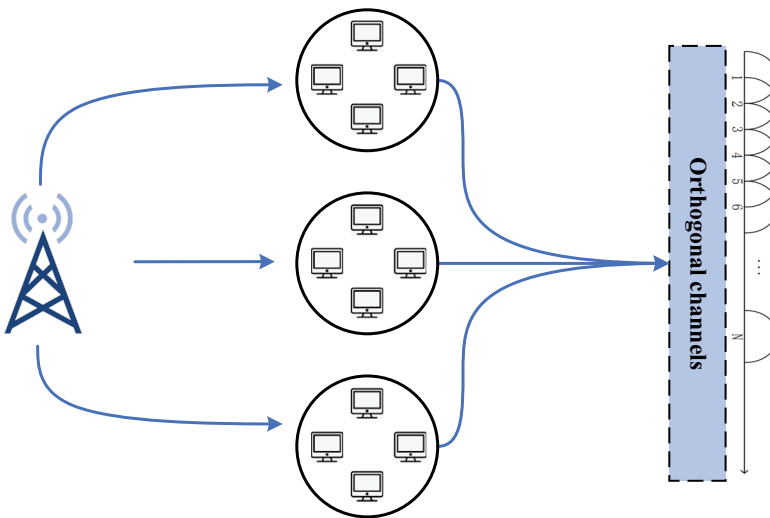
$$a_t^m = \{0, 1, \ldots, N\} \tag{4}$$



**Figure 2.** System model for computer teaching.

The set of action files selected by all other users except user $m$ is:

$$a_t^{\Delta m} = \{a_t^n\}, n \neq m \tag{5}$$

$$B_t^m = (a_t^m, A_t^m) \tag{6}$$

This paper defines a policy expressed as the mapping of the probability mass function of user $m$ from history to action set at time slot $t$:

$$C_t^m = \left[\Pr(a_t^m = 0), \Pr(a_t^m = 1), \ldots, \Pr(a_t^m = N)\right] \tag{7}$$

This paper defines the reward as the achievable data rate on the channel:

$$r_t^m = D\log_2(1 + SNR^m n), A_t^m = 1, a_t^m = n \tag{8}$$

$$r_t^m = 0, A_t^m = 0, a_t^m = n \tag{9}$$

where $D$ is channel bandwidth. The goal of each user is to find a strategy that maximizes the cumulative discount reward:

$$F = \max E\left[\sum_{t=1}^{T} b^{t-1} r_t^m \left(C_t^m, C_t^{\Delta m}\right)\right] \tag{10}$$

There are no primary and secondary users in this paper's dynamic spectrum access study, therefore it does not presume the presence of primary and secondary users inside the network. As the network size grows, the combinatorial optimization issue becomes mathematically hard to solve, making it impossible to compute the ideal solution. When dealing with very vast state and action spaces, we shall employ the deep reinforcement learning approach in this study.

Deep multi-user reinforcement learning (DMRL) is used to construct the proposed dueling deep recurrent Q-network method (DDRQN). No online coordination or message exchange between users is required for this technique, which works well in big and complicated environments. Different from traditional DQN, this paper introduces a dueling DQN (DDQN) and a LSTM, and combination of LSTM and DQN makes up for the shortcoming of DQN's limited memory capacity. The structure of DDQN is illustrated in Figure 3.

The reward function is:

$$G^\pi(s, a) = Q^\pi(s, a) - V^\pi(s) \tag{11}$$

$$Q = V + G \tag{12}$$

In practical problems, the learning effect of this construction method is poor. To solve this problem, the maximum advantage function is further subtracted from the above definition:
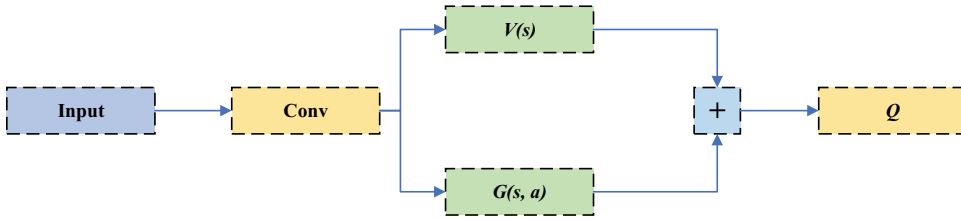
$$Q = V + (G - \max(G)) \tag{13}$$

**Figure 3.** The structure of DDQN.

The optimal action is:

$$a^* = argmax(G) \tag{14}$$

To further improve the learning effect, the maximum operator is replaced by the mean value of the advantage function:

$$Q = V + \left( G - \frac{1}{|G|} \sum\nolimits^G \right) \tag{15}$$

Although some lose their native meaning, it improves the stability of the optimization process. Because the advantage function only needs to change as fast as its mean, it doesn't need to match the speed of its maximum. The contention dueling deep recurrent Q-network (DDRQN) developed in this paper is used to solve the dynamic access. Figure 4 shows the interaction process between the multi-user multi-channel access environment and the algorithm model.

## Experiment

### *Evaluation on Learning Rate*

In the reinforcement learning algorithm, different learning rate settings will affect the convergence of the loss function during the algorithm training process. When the learning rate value is larger, the convergence rate is faster. But it may also result in a sub-optimal solution and miss the optimal solution. When the learning rate value is small, the convergence rate is slower and longer training time is required. In the multi-user case, this paper chooses to set an adaptive learning rate, and compares the loss under the two fixed learning rates with the loss under the adaptive learning rate. This paper chooses two fixed learning rates 0.1 and 0.01 and an adaptive learning rate, which is initially 0.1 and gradually decreases to 0.01 with the increase of time slots. The experimental results are illustrated in Figure 5.

After the network finally converges, a lower learning rate can achieve better performance than a learning rate of 0.1. However, by comparing the three curves, the network training corresponding to the adaptive learning rate is the best. The adaptive learning rate is significantly better than the fixed learning rate in loss convergence.
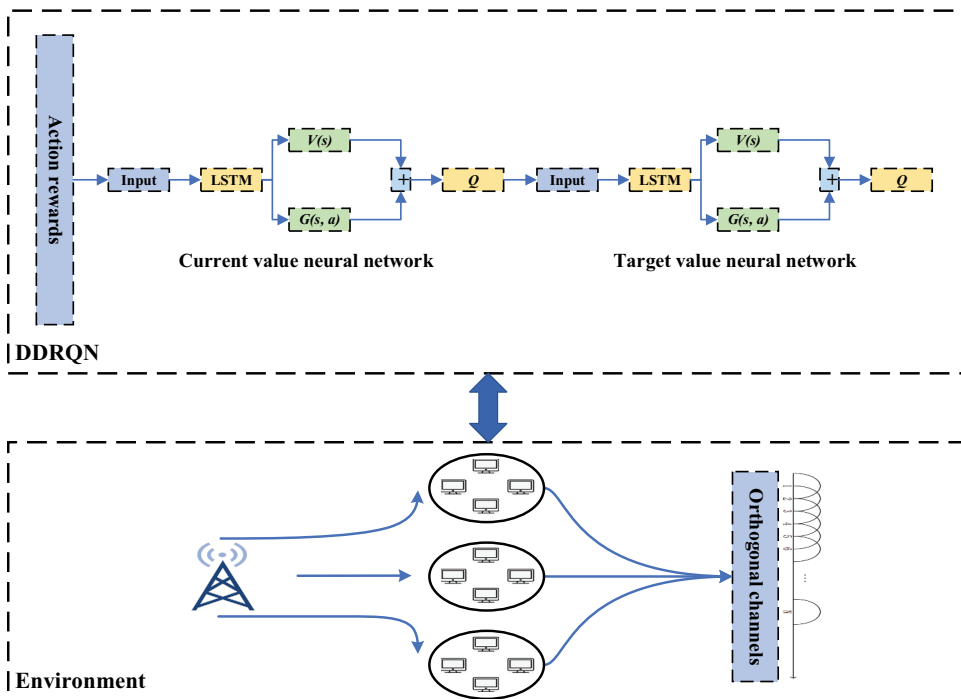
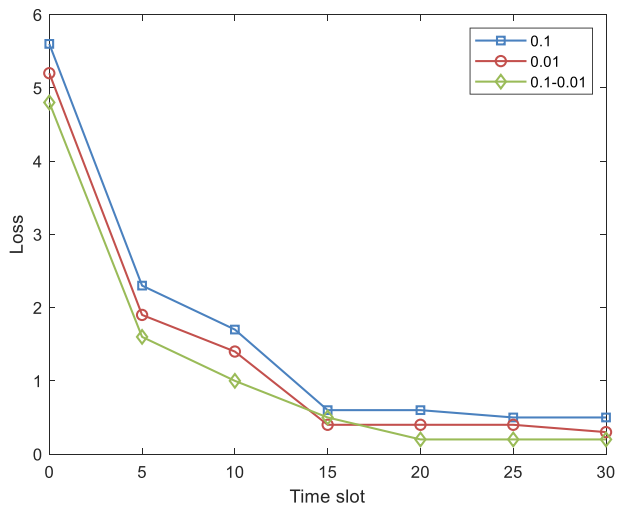**Figure 4.** The interaction diagram of DDRQN.



**Figure 5.** Comparison of losses under different learning rates.

## Evaluation on Greedy Factor

The greedy factor refers to how likely the current sampling is to make a decision based on the Q value generated by the current training network. Different greedy factors will affect the action selection scheme. Therefore, the
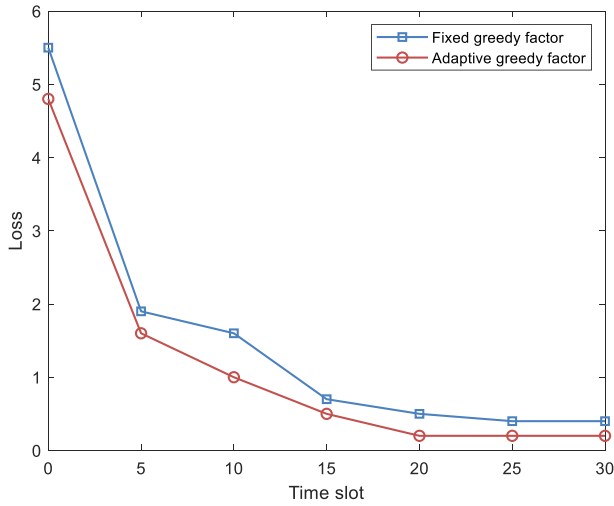
**Figure 6.** Comparison of losses under different greedy factor.

selection of appropriate greedy factor values is also very important. In this paper, the greedy factor is optimized for the dynamic multi-user multi-channel access algorithm based on DDRQN network. This paper adopts a more dynamic approach to this greedy factor fixed-value method and implements an adaptive greedy factor. The performance of the network under the fixed value method and the adaptive method is compared, the results are shown in Figure 6.

The training loss function will be relatively large at the beginning of the time slot, and then the greedy factor will increase adaptively as the time slot increases. It is obvious from the figure that the performance of the adaptive greedy factor is better than that of the fixed value greedy factor. Numerically, when converged, the performance of the adaptive greedy factor in the graph is higher than that of the fixed greedy factor of 0.4.

### Evaluation on Collision Probability and Cumulative Reward

In order to analyze the user collision probability and the cumulative reward index in the network, this paper simulates the network under 16 users and 10 optional channel configurations. The experimental results are shown in Tables 1 and 2.

The user's collision probability trend decreases with the increase of time slots, while the cumulative reward is just the opposite. This is the process of user learning in the network. As the network learns more and more, the cumulative reward obtained in each time step is getting larger and larger, and the collision probability in each time step is getting smaller and smaller. In the case that users do not need cooperation and information exchange, this

**Table 1.** Collision probability analysis.

| Time | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 |
|------|------|------|------|------|------|------|------|------|------|------|
| CP | 57.3 | 29.5 | 25.7 | 22.5 | 19.4 | 17.6 | 13.6 | 12.4 | 8.2 | 5.5 |

**Table 2.** Cumulative reward analysis.

| Time | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 |
|------|------|------|------|------|------|------|------|------|------|------|
| CR | 2351 | 3106 | 3514 | 3827 | 4155 | 4379 | 4503 | 4587 | 4672 | 4802 |

network can greatly reduce the collision probability of users accessing the channel.

### Comparison with Other Method

In the paper, the proposed DDRQN model is compared with two other models for optimizing dynamic channel access in wireless networks: a DQN-based model and a model based on traditional dynamic programming methods.

The DQN-based model is a traditional deep reinforcement learning model that uses a Q-learning algorithm to learn an optimal policy for multi-user dynamic channel access. The model is trained on a dataset of past experiences, and the policy is learned through trial-and-error by maximizing the expected cumulative reward.

The traditional dynamic programming model is a mathematical optimization model that is designed to optimize channel access under specific assumptions and scenarios. The model uses a set of predetermined rules to determine the optimal channel access strategy for each user based on the current state of the network.

The proposed DDRQN model combines the strengths of both the DQN-based model and the traditional dynamic programming model. It uses a double deep recurrent neural network to learn an optimal policy for multi-user dynamic channel access, while also incorporating information about the current state of the network and past experiences through a long short-term memory network.

To compare the performance of these models, the paper evaluates their effectiveness in optimizing user speed and transmission speed under different numbers of users. Specifically, the paper compares the average user speed and average transmission speed achieved by each model for different numbers of users accessing the network.

The results show that the proposed DDRQN model outperforms both the DQN-based model and the traditional dynamic programming model in terms of average user speed and transmission speed under various scenarios. The proposed DDRQN model is able to adapt to changing network conditions and

learn an optimal policy that maximizes network performance, while the other models are more limited in their ability to adapt and optimize under changing conditions.

Overall, the comparison of these models highlights the benefits of using deep reinforcement learning algorithms and recurrent neural networks for optimizing dynamic channel access in wireless networks, and demonstrates the effectiveness of the proposed DDRQN model for achieving improved network performance.

The experimental results are shown in Figures 7 and 8.

The performance of the three algorithms is evaluated in terms of the average data transmission rate of users in a multi-user scenario. The results show that the DDRQN algorithm outperforms the other two algorithms in terms of average transmission rate, indicating that it is a better solution for optimizing dynamic channel access in wireless networks.

Furthermore, the paper compares the performance of the DDRQN algorithm with a general DQN model. The results show that the DDRQN algorithm performs better than the general DQN model, indicating that the long short-term memory network used in the DDRQN algorithm provides better performance than a standard deep Q-network.

As the number of users increases, the average transmission rates of all three algorithms show a downward trend. This is because when multiple users access the same channel, collisions and interference between users can occur, leading to a reduction in the average transmission rate of users.

In conclusion, the proposed DDRQN algorithm offers significant improvements in performance compared to traditional dynamic programming algorithms and standard deep Q-networks. The use of a long short-term memory network enables the algorithm to better handle the complex and dynamic
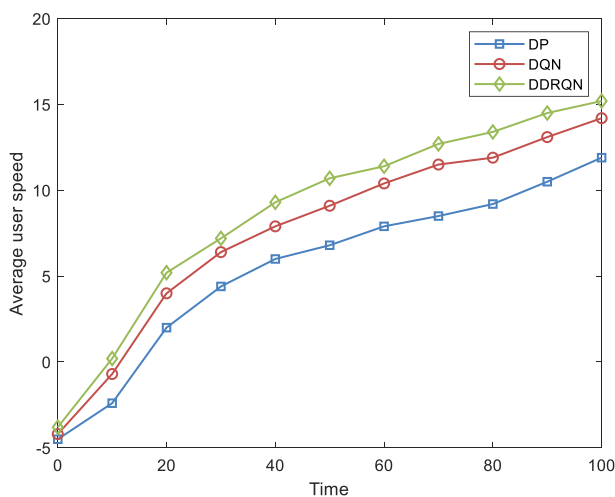


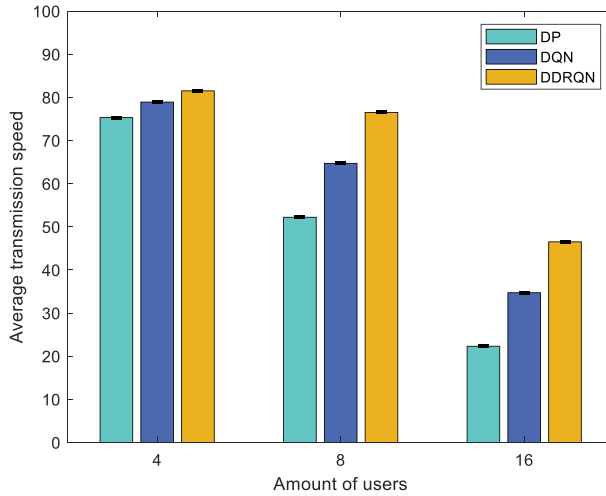**Figure 7.** Result on average user speed.

**Figure 8.** Result on average transmission speed.

nature of multi-user wireless networks, leading to higher average transmission rates for users.

The proposed work makes several contributions to the field of dynamic channel access optimization in wireless networks. The key differences and contributions of the paper can be summarized as follows:

(1) Novel approach: The proposed work presents a novel approach to optimizing dynamic channel access in wireless networks using deep reinforcement learning algorithms and a long short-term memory network. This approach enables the development of a multi-user strategy that maximizes network benefits without requiring online coordination or information exchange between users.

(2) Multi-user optimization: The proposed approach considers the optimization of dynamic multi-channel access under multi-user scenarios, which is a challenging problem in wireless networks. The approach enables each user to select a channel to access and transmit data, while also addressing the collision and interference caused by multiple users accessing a channel at the same time.

(3) Robustness: The proposed approach provides a more robust solution to the problem of dynamic channel access optimization in wireless networks. The approach enables the network to adapt to changing environmental conditions, providing greater network stability and enabling the network to maintain high performance in different scenarios.

(4) Performance improvement: The proposed approach outperforms existing methods in terms of network throughput and collision rate. The results of the study demonstrate the potential of deep reinforcement

learning algorithms and long short-term memory networks in optimizing dynamic channel access in wireless networks.

In summary, the proposed work contributes to the field of dynamic channel access optimization in wireless networks by presenting a novel approach that addresses the challenges of multi-user optimization, robustness, and performance improvement. The work's contributions demonstrate the potential of deep reinforcement learning algorithms and long short-term memory networks in optimizing dynamic channel access in wireless networks, providing valuable insights for future research in this field.

The proposed approach to optimizing dynamic channel access in wireless networks using deep reinforcement learning algorithms and a long short-term memory network is more complex than classic approaches. This is because the approach involves training a deep reinforcement learning model that utilizes a long short-term memory network to learn an optimal policy for multi-user dynamic channel access.

In contrast, classic approaches to dynamic channel access optimization typically use rule-based algorithms or mathematical models that are designed to optimize channel access under specific assumptions and scenarios. These classic approaches are often simpler to implement and require less computational resources than the proposed approach.

However, the proposed approach offers several advantages over classic approaches, including the ability to adapt to changing environmental conditions, the ability to learn from past experiences, and the ability to provide a more robust solution to the problem of dynamic channel access optimization.

Moreover, the complexity of the proposed approach is mitigated by recent advancements in deep reinforcement learning and machine learning hardware, such as graphics processing units (GPUs) and tensor processing units (TPUs), which enable faster and more efficient training of deep reinforcement learning models.

Overall, while the proposed approach is more complex than classic approaches, its advantages in terms of adaptability, robustness, and performance improvement make it a promising solution for optimizing dynamic channel access in wireless networks.

## Conclusion

The widespread use of computers has promoted the development of society. In terms of teaching, computers have changed the traditional teaching mode and brought the development of teaching into a new era. In the teaching process, both students and teachers can use computer technology to obtain richer knowledge resources and comprehensively improve their knowledge reserves.

The application of computer technology effectively improves the efficiency of teaching and enables students to develop more comprehensively. However, with the rapid growth of users for computer teaching, the scarcity of spectrum resources in wireless networks has become more and more prominent. Therefore, it is urgent to propose new intelligent methods to improve spectrum utilization in wireless networks. This paper focuses on the dynamic channel access optimization in wireless networks, and studies the dynamic resource optimization problem based on deep reinforcement learning algorithm. In the case of multiple users and multiple channels, an efficient algorithm access model is adopted to maximize the channel utilization of the system and maximize network benefits. This work considers the dynamic multi-channel access problem under multi-user and the collision and interference caused by multiple users accessing a channel at the same time. Each user selects a channel to access and transmit data, and judges whether the transmission is successful or not through a binary observation signal. The network aims to dig a multi-user strategy to maximize network benefits without online coordination or information exchange among users. This paper applies a deep reinforcement learning-based algorithm for simulation, combined with a LSTM, which enables network to use history of processes to estimate true state.

The objectives of the proposed work are to optimize dynamic multi-channel access in wireless networks and to develop a multi-user strategy that maximizes network benefits without requiring online coordination or information exchange between users. To achieve these objectives, the work proposes a novel approach that utilizes deep reinforcement learning algorithms and a long short-term memory network.

The proposed approach is designed to address the dynamic resource optimization problem, which is a key challenge in dynamic spectrum access. The approach considers the collision and interference caused by multiple users accessing a channel simultaneously and enables each user to select a channel to access and transmit data.

To evaluate the effectiveness of the proposed approach, the work conducted systematic studies, including simulations and experiments. The results of these studies demonstrate that the proposed approach is an effective means of optimizing dynamic channel access in wireless networks.

Specifically, the results show that the proposed approach outperforms existing methods in terms of network throughput and collision rate. The approach also provides a more robust solution, enabling the network to adapt to changing environmental conditions and providing greater network stability.

Moreover, the work's results demonstrate the potential of deep reinforcement learning algorithms and long short-term memory networks in optimizing dynamic channel access in wireless networks. The approach's

ability to utilize the history of processes to estimate the true state provides valuable insights into how machine learning algorithms can be used to address the dynamic resource optimization problem in wireless networks.

The proposed work's objectives are to address the dynamic resource optimization problem in wireless networks and to develop a multi-user strategy that maximizes network benefits without requiring online coordination or information exchange between users. The results demonstrate that the proposed approach is effective in achieving these objectives and provides a novel solution to the problem of dynamic channel access optimization in wireless networks.

The presented approach also has several potential applications in the field of wireless communications. Some of these applications include:

(1) Wireless network optimization: The proposed approach can be used to optimize dynamic channel access in various types of wireless networks, including cellular networks, Wi-Fi networks, and ad-hoc networks.
(2) Internet of Things (IoT): The proposed approach can be applied to optimize channel access in IoT networks, which often involve a large number of devices with limited resources.
(3) Smart cities: The proposed approach can be used to optimize wireless network performance in smart city applications, such as intelligent transportation systems, smart energy grids, and environmental monitoring.
(4) Industrial automation: The proposed approach can be used to optimize wireless network performance in industrial automation applications, such as process control, monitoring, and predictive maintenance.
(5) 5 G and beyond: The proposed approach can be applied to optimize dynamic channel access in 5 G and beyond networks, which require advanced optimization techniques to handle the large volume of data and diverse range of applications.

Overall, the proposed approach has broad applications in various fields that require efficient and robust wireless network performance. By optimizing dynamic channel access using deep reinforcement learning algorithms and a long short-term memory network, the proposed approach has the potential to improve network throughput, reduce collision rates, and provide more stable and reliable wireless network performance in a range of scenarios.

## Disclosure Statement

No potential conflict of interest was reported by the author(s).

## Funding

## Data Availability Statement

Data sharing is not applicable to this article.

## References

Ahmad, A., S. Ahmad, M. H. Rehmani, and N. U. Hassan. 2015. A survey on radio resource allocation in cognitive radio sensor networks [J]. *IEEE Communications Surveys & Tutorials* 17 (2):888–917. doi:10.1109/COMST.2015.2401597.

Ahmed, E., A. Gani, S. Abolfazli, L. J. Yao, and S. U. Khan. 2016. Channel assignment algorithms in cognitive radio networks: Taxonomy, open issues, and challenges[J. *IEEE Communications Surveys & Tutorials* 18 (1):795–823. doi:10.1109/COMST.2014.2363082.

Bany Salameh, H., Z. Khader, and A. Al Ajlouni. 2021. Intelligent secure networking in in-band full-duplex dynamic access networks: Spectrum management and routing protocol [J]. *Journal of Network and Systems Management* 29 (2):1–18. doi:10.1007/s10922-021-09588-7.

Bkassiny, M., Y. Li, and S. K. Jayaweera. 2013. A survey on machine-learning techniques in cognitive radios [J]. *IEEE Communications Surveys & Tutorials* 15 (3):1136–59. doi:10.1109/SURV.2012.100412.00017.

Du, J., C. Jiang, Z. Han, H. Zhang, S. Mumtaz, and Y. Ren. 2017. Contract mechanism and performance analysis for data transaction in mobile social networks [J]. *IEEE Transactions on Network Science and Engineering* 6 (2):103–15. doi:10.1109/TNSE.2017.2787746.

Erbas, İ., R. Çipuri, and A. Joni. 2021. The impact of technology on teaching and teaching English to elementary school students [J]. *Linguistics & Culture Review* 5 (S3):1316–36. doi:10.21744/lingcure.v5nS3.1815.

Fan, X., and Y. Huo. 2020. Blockchain based dynamic spectrum access of non-real-time data in cyber-physical-social systems [J]. *IEEE Access* 8:64486–98. doi:10.1109/ACCESS.2020.2985580.

Gao, Z., H. Zhu, S. Li, S. Du, and X. Li. 2012. Security and privacy of collaborative spectrum sensing in cognitive radio networks [J]. *IEEE Wireless Communications* 19 (6):106–12. doi:10.1109/MWC.2012.6393525.

Hbaci, I., H. Y. Ku, and R. Abdunabi. 2021. evaluating higher education educators' computer technology competencies in Libya [J]. *Journal of Computing in Higher Education* 33 (1):188–205. doi:10.1007/s12528-020-09261-z.

Hlophe, M. C., and B. T. Maharaj. 2021. AI meets CRNs: A prospective review on the application of deep architectures in spectrum management [J]. *IEEE Access* 9:113954–96. doi:10.1109/ACCESS.2021.3104099.

Huang, Q., Y. Gui, F. Wu, G. Chen, and Q. Zhang. 2015. A general privacy-preserving auction mechanism for secondary spectrum markets [J]. *IEEE/ACM Transactions on Networking* 24 (3):1881–93. doi:10.1109/TNET.2015.2434217.

Kaur, A., and K. Kumar. 2022. A comprehensive survey on machine learning approaches for dynamic spectrum access in cognitive radio networks [J]. *Journal of Experimental & Theoretical Artificial Intelligence* 34 (1):1–40. doi:10.1080/0952813X.2020.1818291.

Kotobi, K., and S. G. Bilen. 2018. Secure blockchains for dynamic spectrum access: A decentralized database in moving cognitive radio networks enhances security and user access [J]. *IEEE Vehicular Technology Magazine* 13 (1):32–39. doi:10.1109/MVT.2017.2740458.

Li, X., and Q. Zhu. 2018. Social incentive mechanism based multi-user sensing time optimization in cooperative spectrum sensing with mobile crowd sensing [J]. *Sensors* 18 (1):250. doi:10.3390/s18010250.

Liang, Y. C., J. Tan, H. Jia, J. Zhang, and L. Zhao. 2021. Realizing intelligent spectrum management for integrated satellite and terrestrial networks [J]. *Journal of Communications and Information Networks* 6 (1):32–43. doi:10.23919/JCIN.2021.9387703.

Liang, Y. C., Y. Zeng, E. C. Y. Peh, and A. Tuan Hoang. 2008. Sensing-throughput tradeoff for cognitive radio networks [J]. *IEEE Transactions on Wireless Communications* 7 (4):1326–37. doi:10.1109/TWC.2008.060869.

Marinho, J., and E. Monteiro. 2012. Cognitive radio: Survey on communication protocols, spectrum decision issues, and future research directions [J]. *Wireless Networks* 18 (2):147–64. doi:10.1007/s11276-011-0392-1.

Mihovska, A., and R. Prasad. 2021. Spectrum sharing and dynamic spectrum management techniques in 5G and beyond networks: A survey [J]. *Journal of Mobile Multimedia* 1:65–78.

Nguyen, H., and R. Santagata. 2021. Impact of computer modeling on learning and teaching systems thinking [J]. *Journal of Research in Science Teaching* 58 (5):661–88. doi:10.1002/tea.21674.

Peh, Y., E. C. Liang, Y. C, and Y. L. Guan. 2009. Optimization of cooperative sensing in cognitive radio networks: A sensing-throughput tradeoff view [J]. *IEEE Transactions on Vehicular Technology* 58 (9):5294–99. doi:10.1109/TVT.2009.2028030.

Ran, H., M. Kasli, and W. G. Secada. 2021. A meta-analysis on computer technology intervention effects on mathematics achievement for low-performing students in K-12 classrooms [J]. *Journal of Educational Computing Research* 59 (1):119–53. doi:10.1177/0735633120952063.

Randhava, K. S., M. Roslee, and Z. Yusoff. 2021. Dynamic spectrum management using frequency selection at licensed and unlicensed bands for efficient vehicle-to-vehicle communication[J. F1000*Research* 10 (10):1309. doi:10.12688/f1000research.73481.1.

Sekaran, R., S. N. Goddumarri, S. Kallam, M. Ramachandran, R. Patan, and D. Gupta. 2021. 5G integrated spectrum selection and spectrum access using AI-Based frame work for IoT based sensor networks [J]. *Computer Networks* 186:107649. doi:10.1016/j.comnet.2020.107649.

Shah-Mohammadi, F., H. H. Enaami, and A. Kwasinski. 2021. Neural network cognitive engine for autonomous and distributed underlay dynamic spectrum access [J]. *IEEE Open Journal of the Communications Society* 2:719–37. doi:10.1109/OJCOMS.2021.3069801.

Tragos, E. Z., S. Zeadally, A. G. Fragkiadakis, and V. A. Siris. 2013. Spectrum assignment in cognitive radio networks: A comprehensive survey [J]. *IEEE Communications Surveys & Tutorials* 15 (3):1108–35. doi:10.1109/SURV.2012.121112.00047.

Wang, W., L. Chen, K. G. Shin, and L. Duan. 2015. Thwarting intelligent malicious behaviors in cooperative spectrum sensing [J]. *IEEE Transactions on Mobile Computing* 14 (11):2392–405. doi:10.1109/TMC.2015.2398446.

Wang, W., H. Li, Y. Sun, and Z. Han. 2009. Securing collaborative spectrum sensing against untrustworthy secondary users in cognitive radio networks [J]. *EURASIP Journal on Advances in Signal Processing* 2010 (1):1–15. doi:10.1155/2010/695750.

Wang, Y., Z. Ye, P. Wan, and J. Zhao. 2019. A survey of dynamic spectrum allocation based on reinforcement learning algorithms in cognitive radio networks [J]. *Artificial Intelligence Review* 51 (3):493–506. doi:10.1007/s10462-018-9639-x.

Wu, H., S. Jin, and W. Yue. 2022. Pricing policy for a dynamic spectrum allocation scheme with batch requests and impatient packets in cognitive radio networks [J]. *Journal of Systems Science and Systems Engineering* 31 (2):133–49. doi:10.1007/s11518-022-5521-0.

Yi, C., J. Cai, and G. Zhang. 2016. Spectrum auction for differential secondary wireless service provisioning with time-dependent valuation information [J]. *IEEE Transactions on Wireless Communications* 16 (1):206–20. doi:10.1109/TWC.2016.2621765.

Zhang, Y., G. Hu, and Y. Cai. 2021. Proactive spectrum monitoring for suspicious wireless powered communications in dynamic spectrum sharing networks [J]. *China Communications* 18 (12):119–38. doi:10.23919/JCC.2021.12.008.

Zhang, Y., C. Lee, D. Niyato, and P. Wang. 2013. Auction approaches for resource allocation in wireless systems: A survey [J]. *IEEE Communications Surveys & Tutorials* 15 (3):1020–41. doi:10.1109/SURV.2012.110112.00125.