[1]M. Narender Pavan

[1]Sushil Kumar

[1]Gajendra Nayak

Deep Reinforcement Learning based channel allocation (DRLCA) in Cognitive Radio Networks

Abstract: The Spectrum is a scarce resource in wireless networks, hence these resources need to be perfectly channelized for their better utilization. By the study of the previous decades, it is found that the spectrum is colossally underutilized and the main reason for the underutilization found out to be the policies that are fixed and not dynamic. The dynamic spectrum allocation of frequency bands may overcome this problem. Cognitive radio provides an important concept that can used to solve the problem of underutilization of spectrum. Reinforcement learning is a key technique that is widely used to learn the spectrum allocation behaviour and maximize the system's efficiency. Therefore, in this work, we have designed and developed a Reinforcement learning-based model to allocate the channels among secondary users. Also, a Deep Reinforcement learning-based channel allocation algorithm (DRLCA) has been proposed. The proposed DRLCA is compared with existing JPCRL [1]. In our algorithm, the Python libraries were used for simulation. From the simulation results and analysis, it is found that the DRLCA outperforms the JPCRL in terms of channel utilization by 5%.

Keywords: Cognitive radio, Spectrum, Reinforcement learning, Q-learning, Reward function

I. INTRODUCTION

As the world has hit with the COVID-19 pandemic, a new buzzword has lingered in everybody, especially techie-oriented people is WFH i.e Work from home. WFH has become a new normal and many of the multinational organisations have adopted and implemented the same. Suddenly, everything is happening sitting at home, and a lot of it through online mode. As going out is not an option owing to the lockdown, people have shifted to online classes, video calls, conducting online meetings and conferences or making telemedicine connections to talk to Doctors. This has generated an unprecedented jump in demand for bandwidth. To meet the new and abrupt changes in demand, the existing meager bandwidth needs to be utilized more adroitly and effectively.

As spectrum is a limited resource, this needs to be utilized in a most appropriate manner so that every bit of it counts for its better usage. The need of the hour is to shift the focus from underutilizing these limited bandwidth resources to appropriate utilization. Cognitive radio (CR) technology is widely used in today's bandwidth-centric network scenarios for better utilization of the spectrum, mainly due to its potential to recognize the spectrum holes (i.e unused portions at a particular time).

Cognitive radio offers a better intelligent radio that learns and makes instantaneous decisions based on the surrounding environment. As of now, the spectrum allocation has been handled by Govt agencies in a static way. The majority of the allocated spectrum is unused, according to the various literature [2] [3] [4]. The idea of cognitive radio has been put out to address the issue of underutilization of the spectrum. Joseph Mitola suggested the idea of cognitive radio in his thesis from 1999 [5] [6] [7] [8]. Additionally, Simon Haykin introduced cognitive radio's signal management idea in 2002 [9].

Through the use of the unused spectrum (white space), cognitive radio (CR) enables secondary users (unlicensed users) to increase the bandwidth available for themselves. Cognitive radio consists of two primary parts namely

1 *Corresponding author: M.Narender Pavan School Of Computer and Systems Sciences, Jawaharlal Nehru University email: mnpavan@mail.jnu.ac.in

2 Sushil Kumar School of Computer and Systems Sciences, Jawaharlal Nehru University

3 Gajendra Nayak School of Computer and Systems Sciences, Jawaharlal Nehru University

Cognitive capability and Reconfigurability. A cognitive cycle process manages these elements. The cognitive cycle is divided into several stages, including spectrum sensing, spectrum analysis, sharing, spectrum decision making, spectrum, and spectrum sharing.

Using the spectrum-sharing phase, the unlicensed (secondary user) and licensed (primary user) users will share the vacant piece of the spectrum. Spectrum sharing is governed by three standards: Overlay, Underlay, and Interweave. In the overlay, primary data is prioritised for transmission before secondary data, while the unused piece of the spectrum is shared between the two users. Whereas the interweave approach shares the unused portion of the spectrum with secondary users only when the primary is not in use, the underlay approach defines an interference temperature limit below which both primary and secondary users can send their data simultaneously. Spectrum sharing is carried out using a number of factors, including Interference, Signal to Interference noise ratio (SINR), fairness etc.

Innovations and technological developments are happening so rapidly, recent developments in the technology of Artificial intelligence (AI) [10][11][12] and its applicability in real-time scenarios have shown to bring the brighter side of this technology. The use of this AI technology is booming in the telecommunications sector and the merger of AI and cognitive radio technology has already made tremendous progress and opened a new door for further improvements in technology.

Machine learning for Cognitive wireless communication: Machine learning is a subsidiary of Artificial intelligence and this machine learning-based cognitive technology will solve the problem of shortage of spectrum which is a major concern in today's telecommunication technology [13]. The utilization of machine learning techniques is an important aspect of improving spectrum utilization. Awareness learning is possible through machine learning technology. Rapidly adapting to the wireless environment and self-learning process by observing are the key features of Machine learning [14][15][16]. The performance of the wireless devices improves by using these features in wireless environments. Machine learning aims to learn from the previous data and make accurate predictions without any computational intervention in order to produce the new data set [17][18]. As today's world is exposed to new data so rapidly, it is very difficult to predict and analyse the data in an ever-changing wireless environment. To overcome this problem, Machine learning has been introduced to analyse the ever-growing wireless data [19][20][21][22].

Supervised, Unsupervised, and Reinforcement learning are the three categories of machine learning [23]. The system knows the patterns of the output data when it uses the labelled data. The outcome of the supervised learning is more accurate and reliable. The computer is not trained with labelled data in Unsupervised learning but is presented with data and challenged to find the relationship in it. In Reinforcement learning, machines learn and try to adapt to an ideal behaviour in the given environment. In Cognitive radio networks, a Reinforcement learning-based model can be used for channel allocation. For this, a reward function is generated and an optimum reward value is also generated based on the value of the optimum reward value, the channels have been allocated to secondary users.

Determining the high-quality white spaces and allocation of these temporal white spaces to the secondary users is a major task in spectrum sharing and this can be solved by using the technique of Reinforcement learning. The environment was observed and some kind of feedback (reward) was taken. Through the feedback, the process of learning takes place to control the system and also paves the way to maximize the system performance [24][25][26]. In the proposed work, the utilization of channels has been enhanced using the reinforcement learning technique [27][28][29]. Having known the importance of the interference ratio [30] [31] in sharing the spectrum, we have designed an algorithm using the technique of Reinforcement Learning [32].

2. RELATED WORK AND MOTIVATION

There are several papers related to channel allocation in cognitive radio networks. In this section, some of them that are related to the proposed work are reviewed and presented below:

In a survey paper by [33], the authors discussed in depth about the role of learning in cognitive radio and presented a taxonomy of machine learning techniques and approaches. They have also presented a comparison of a learning algorithm in a tabular format along with their advantages and disadvantages.

In [34], the authors presented the taxonomy of spectrum allocation problem and techniques and also presented a summary of the Reinforcement learning-based spectrum allocation algorithm in a tabular format. The techniques for solving spectrum allocation problems using reinforcement learning were given and discussed extensively.

The authors in [35], proposed a machine learning based opportunistic spectrum access. The authors defined an occurrence aware opportunistic spectrum aware (OA-OSA) which was based on the UCB algorithm and it was considering only local information, further, the authors in this paper, extended the OA-OSA to multiple scenarios. Simulation results reveals that proposed algorithm achieved superior performance in network throughput and less deviation from optimal performance with that of global information. The authors doesn't discuss about the performance parameters like latency and fairness and this can be considered as drawback of the proposed work and the same can be taken up in future work of this paper.

In [36], proposed an algorithm that will maximize spectrum usage without causing any interference to other nodes in the network. For this, the authors used a simple 4-node network and simulated an environment where there were 4 channels available for transmission. Under this environment space, the authors proposed three possible actions that can be taken by an agent at a particular time t. Action-I will not acquire or drop the channel, Action-II will acquire a channel for transmission and Action-III will drop a channel that is already has in use. Accordingly, a function and policy have been designed and simulated, the simulation results prove that the spectrum utilization is maximized without causing any interference to the other nodes.

In [37], Proposed an algorithm that can reduce/prevent excessive spectrum switching and increase the throughput. To achieve the task, the authors have developed an Epsilon-Greedy exploration strategy. Simulation results using NS-3 shows that the proposed model outperformed in terms of channel interchange thereby improving the performance of the secondary user. However, some other system performance parameters may also would have been chosen for better comparison and results analysis.

In [38], A reinforcement learning approach to enhance QoS in LTE-A networks was proposed by the authors. They adopted a finite number of states Markov decision process and a reward value was assigned for average throughput, delay, and packet loss. Simulation results reveal that a good performance was noted especially running the real-time video applications in terms of above parameters. A better traffic classification procedure would have been adopted to further improve the performance and this is the limitation of the above work.

In [39], The issue of channel assignment and power optimization was addressed by the authors and also proposed a centralized reinforcement learning-based autonomic solution. For this, the service arrival of secondary users is taken and secondary users are served on each channel in a dynamic way. The simulation result prevails that the performance of throughput is increased and channel and power are allocated in a centralized way. However, the authors could not consider the interference caused by the incoming user and this is the limitation of the proposed work.

In [40], The concept of Q-learning is used in a distributed manner to select a recognition task for each cognitive radio user. The proposed arrangement works without considering channel state information or priority traffic estimation. The maximum reward is used to select the identification task. The simulation results show that an improvement in the sensing efficiency.

In [41], the authors presented a technique for dynamic spectrum usage by unlicensed users in an IoT Industry-based scenario by a cognitive self-learning approach. In the first step, they developed a cached MAC protocol for single-channel access, and in the second step, a Q-learning-based access algorithm for distributed multiple-channel DSA, and the simulation results show that they have achieved good results. At the same time, the authors could not consider the cases where the nodes of the mesh change their place, this is a shortcoming of the proposed work.

In [42], the authors presented a comprehensive review of reinforcement learning techniques in the concern of cognitive radio networks and their use in dynamic channel selection and channel detection. The authors provided

a detailed discussion of reinforcement learning models where the reinforcement approach was directly applied to different cognitive radio network systems. Performance improvements in terms of higher throughput, lower end-to-end delay/link delay, lower interference level for primary users, lower number of handovers, number of channels detected as idle and higher accumulated fee were also discussed and demonstrated. Finally, the authors raised open questions and challenges for a system based on improving cognitive radio platforms.

In [43],the authors investigate the search of free and busy channels based on channel properties for clustering in the environment of cognitive radio networks. A free and busy channel pool is examined for a random distribution of free and busy channels for every user. To do this, a signal sequence of free and busy states of one channel with a statistical relationship between two consecutive time intervals was applied using the Markov model. They implemented a support vector machine learning algorithm. The simulation results show that the authors can decrease the channel search time by 10 percent. The proposed algorithm can be further improved with deeper channel reserve training.

In [44],the researchers presented a comprehensive overview of general techniques for solving the issues of resource allocation in cognitive radio networks. A critical review of the main approaches applied to resource allocation in cognitive radio networks was performed. The review identifies the biggest challenges in wireless resource allocation. The authors presented a detailed discussion on the formulation of resource allocation problems for cognitive radio networks and also provided a summary of methods for solving resource allocation problems. Finally, the authors presented open problems and challenges. In addition, they offered ideas for possible future research, which can indeed be a good basis for directing research in the field of cognitive radio networks.

From the study of the above work/papers [33-44], it is analysed that none of them considered the interference ratio as the main parameter for designing the reward function in the reinforcement learning technique. This is the reason that motivates us to design and develop the Reinforcement learning-based scheme by using the interference ratio as the main parameter.

3. SYSTEM MODEL AND PROPOSED ALGORITHM

Let us assume that there exist N primary and M secondary users in the system or network. The channels for N primary users is n; out of n channels m channels are available for secondary users; $m = m_1, m_2, m_3, ..., m_n$. There exists a cognitive base station to handle the activity of cognitive users. In our work, the cognitive base station decides each cognitive user. Our system model is divided into three subsections, namely:

1. Markov decision process
2. Q-learning-based reward function
3. Q-learning-based Channel allocation algorithm/proposed scheme

After that, the algorithm for optimized channel allocation has been designed and developed.

a) ***Markov Decision process****:* In our Model, we have n channels and it is considered that $n_1, n_2, n_3, ..., n_n$ states for channels. Let $S = n_1, n_2, n_3, ..., n_n$ is a finite set of these n states. It is assumed that $a_1, a_2, a_3, ..., a_k$ is a set of actions. The transition probabilities upon taking the actions "a_i" in state S is denoted by $P_{ai}s$. In this work $r \in [0, 1]$ is the discount factor. Now the reinforcement function is defined as

$$R : S \rightarrow r$$

Such that

$$R \leq |R_max|$$

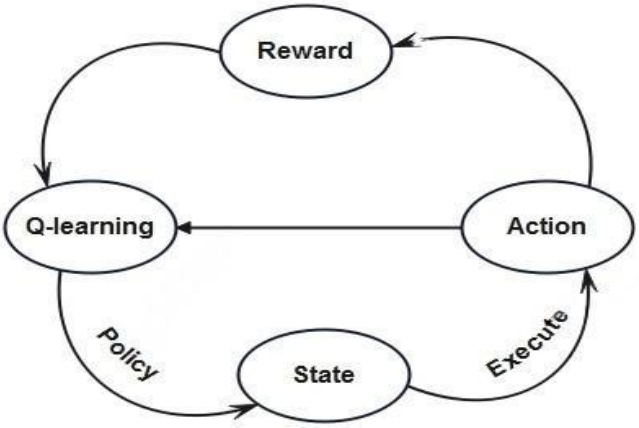

Fig. 1. Q-learning-based channel selection

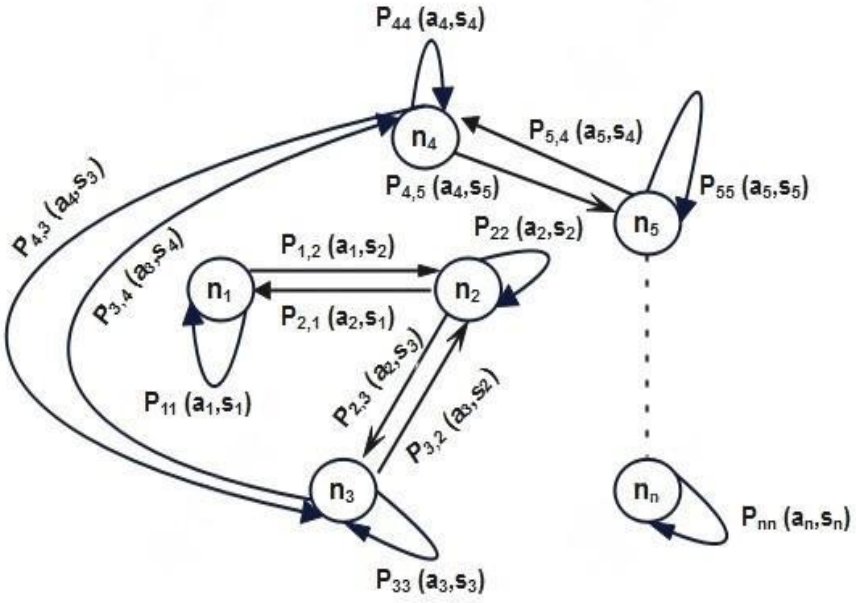

Fig. 2. Markov Decision process

Let SINR be the signal-to-interference plus noise ratio is calculated as:

$$SINR = \frac{\sum P_{ii} G_{ii}}{\sum_{i \neq j} P_{ii} G_{ii} + N} \quad (1)$$

Where

P_{ii}=Power Gain

G_{ii}=Channel Gain

N= Noise factor

The SNR is defined as

$$SNR = \frac{\sum P_{ii} G_{ii}}{\sum_{i \neq j} P_{ii} G_{ii}} \quad (2)$$

This work is the extension of our previous work as [45] and the interference ratio is determined as:

$$Interference\ Ratio\ (IR) = \frac{SINR}{SNR} \quad (3)$$

In this work, the IR ratio is considered as the main parameter to decide the actions taken by cogni- tive/secondary users. The Markov chain-based scenario is already discussed in [45]. The Markov decision process (MDP) can be defined as [46][47][48][49]:

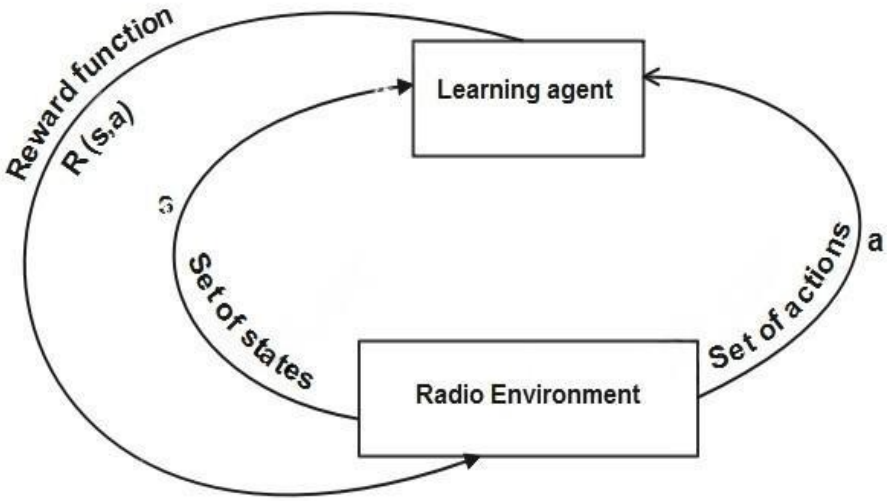

Fig. 3. Learning environment

A finite MDP can be a tuple and is defined as

$$(S, A, P_{sa}, r, R)$$

Where,

S = Finite set of states

A = Set of actions

Psa = the state transition probability

r = discount factor

R = Reinforcement function

b) ***Q-learning and Reward Function:*** The reinforcement-based model consists, a set of states' *S* $(n_1, n_2, n_3, ..., n_n)$, A set of actions $A = (a_1, a_2, a_3, ..., a_k)$, A discount factor *r,* and a policy for cognitive users as defined as $\pi : S \rightarrow A$ as depicted in fig-3.

At any particular time, instance, the cognitive users consider the state $n_n \in S$ of the environment and decide the action a ∈ A and pay attention to its current policy π; In our case, the actions are decided based on interference ratio and reward value. Similarly, the radio environment establishes a transition to the next state n2 ∈ S and generates a reinforcement function R(s, a) known as an immediate reward to the cognitive users. Afterward, the learning cognitive users update their policy and go for the next loop of iteration. Cognitive users intend to develop an optimal policy for every state given by

$$V^{\pi}(S_n) = E[R(S_2) + rR(S_2) + r^2R(S_2) + \cdots]$$

$$V^{\pi}(S_n) = E[\sum_{i=0}^{\infty} r^{i-1}R(S_1)|(S_0)$$

$$V^{\pi}(S_n) = E[\sum_{i=0}^{\infty} r^{i-1} R(S_1, \pi(S_1)|(S_0)=S) \quad (4)$$

Using Bellman's optimality criteria [50], the optimal policy π^* must satisfied the following condition:

$$V^*(S) = V^{\pi^*}(S) = Max[R(s,a) + \gamma \sum_{n_i \epsilon S} P_{S,S'}(a) V^*(S')] \quad (5)$$

where *S,S′* denotes the current and next state, *R(S, a)* is the reward and $P_{S,S'(a)}$ is the state transition probability from *S* to *S′* by action *a*.

The endeavour of q-learning is to find a π^* without the prior knowledge of reward and transition probabilities; the optimized policy may be learned by using the simple *Q*-value iterations and defined as:

$$Q_{\pi}(s,a) = R(s,a) + \gamma \sum_{n_i \epsilon S} P_{S,S'}(a) V^*(S') \quad (6)$$

The optimal policy and optimal state value can be calculated using Equation 4 and Equation 5 and can be defined as:

$$V^*(S) = Max_{a \epsilon A}(Q^*(S,a)) \quad (7)$$

$$\pi^*(S) = argmax(Q^*(S,a)) \quad (8)$$

or can be written as:

$$V^*(S) = Sup_{\pi}(V^{\pi}(S) \quad (9)$$

$$Q^*(S,a) = Sup_{\pi}(Q^{\pi}(S,a) \quad (10)$$

The Q-learning uses

$$Q_{t+1}(S,a) = (1-\alpha)Q(S,a) + \alpha R_i + \gamma Max_{a^t}(Q_t(S,a) \quad (11)$$

where α is the learning rate and $0 \leq \alpha \leq 1$.It is observed that as $t \rightarrow \infty$ and $\alpha \rightarrow 0$ $Q_t(S,a)$ converges to $Q_t{}^*(S',a)$.

The reward plays a key role in navigating the secondary users in the system and is used to take action to tackle the requests of other users. The reward function may be defined in various ways. In our work, the reward function is calculated by the following equation:

$$R(S,a) = \sum_{i=1}^{n} IR_i$$

$$R(S,a) = \sum_{i=1}^{n} \frac{SINR}{SNR}$$

$$R(S,a) = \sum_{i=1}^{n} \frac{\frac{\sum P_{ii} G_{ii}}{\sum_{i \neq j} P_{ii} G_{ii} + N}}{\frac{\sum P_{ii} G_{ii}}{\sum_{i \neq j} P_{ii} G_{ii}}}$$

In our rewards, the interference ratio is considered as the main parameter. The main idea of assigning higher rewards is to give the maximum throughput.

c) Q-learning-based Channel allocation algorithm/proposed scheme:

Algorithm 1: Deep Reinforcement Learning based channel allocation (DRLCA)

Input: Primary users, Secondary users, channels, γ, r, R_max, SINR, SNR, Q_Th

Output: Optimized channel allocations among secondary users

for *each State* $\in A, i = 1, 2, 3 \ldots n$ **do**

calculate the SINR using the following equation:

$$SINR = \frac{\Sigma p_{ii} G_{ii}}{\Sigma_{i \neq j} P_{ij} G_{ij} + N}$$

Calculate the SNR using the equation as below:

$$SNR = \frac{\Sigma p_{ii} G_{ii}}{\Sigma_{i \neq j} P_{ij} G_{ij}}$$

The IR for each state is calculated as follows:

$$IR_i = \frac{SINR_i}{SNR_i}$$

Define the Optimal policy as below:

$$V^{\pi}(S_n) = E[\Sigma_{i=1}^{\infty} r^{i-1} R(S_i, \pi(S_i)) | S_0 = S]$$

Calculate the optimized policy using the following equation:

$$V^*(S) = V^{\pi^*}(S) = \max[R(S,a) + \gamma \Sigma_{n_{i \in S}} P_{S,S'}(a), V^*(S')]$$

Calculate the optimum Q-Value as below:

$$Q^*(S,a) = \sup_{\pi} Q^{\pi}(S,a)$$

if *($|Q| \leq |Q_{TH}|$) in a Hash H* **then**

Allocate the channel to Secondary users

else

Go to step 1

return *Channel Assignment in CRNs*

SIMULATION AND RESULT ANALYSIS

The proposed channel allocation (DRLCA) has been simulated and analysed by the Python programming tools. In our simulation, *N*-primary users and *M*-secondary users are considered. Each primary user has its own channel to transmit the messages or data. Therefore, we have a total number of *N* channels. The *SINR* and *SNR* of each channel is calculated/distributed in a random fashion. The interference ratio for each state is calculated via the *SINR* and *SNR* of that particular state.

$$Interference\ Ratio(IR) = \frac{SINR}{SNR}$$

The free channels have been determined with the help of the indicator function/hidden Markov method. Suppose we have N_m-free channels that can be utilized by the secondary users. We have adopted a deep reinforcement learning scheme as already depicted in our algorithm.

IR value	**Random**
Q value	**Random**
Optimized Q value	**Generated for each and every state**
V value	**Generated through the training (after the training of Q table)**

Table 1. **Simulation parameters**

The proposed DRLCA compares with the JPCRL algorithm as in [1]. Both algorithms have been simulated and analysed by the Python programming tools. Further, it is found that the proposed DRLCA outperforms the existing JPCRL algorithm in terms of channel utilization. The channel utilization has been increased when the channels are allocated according to the DRLCA.

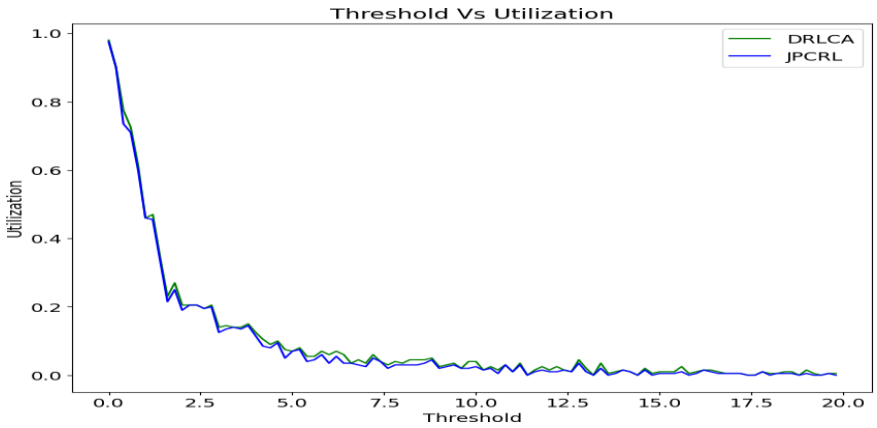

Fig-4. Threshold vs Utilization

In our simulation, we have increased the value of the threshold and observed the behavior of channel utilization. It is found that when the threshold value is low, the channel utilization is high and as the value of the threshold increased up to 20, it is observed that the utilization is going to be constant after that particular value. From figure-4, it is observed that these DRLCA schemes perform better in comparison of JPCRL scheme.

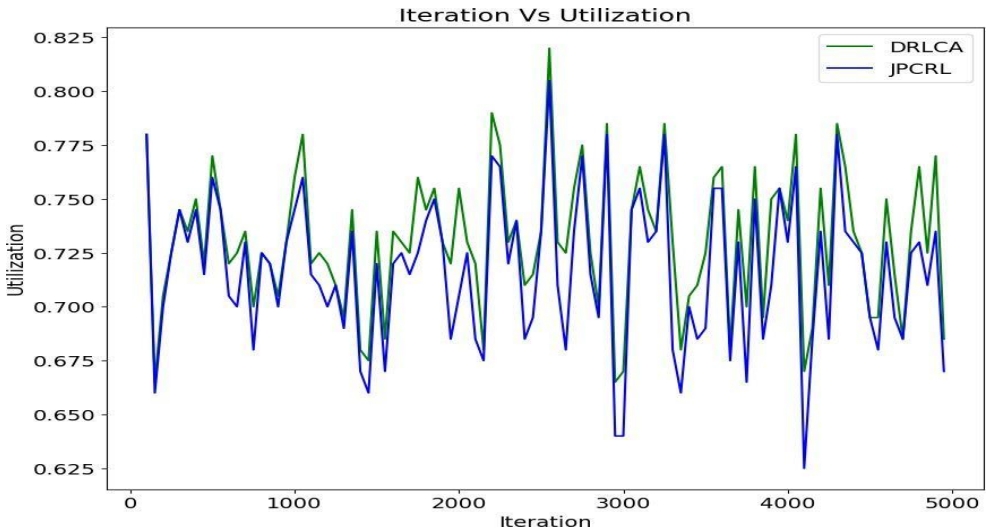

Fig-5. No of iterations vs Utilisation

The figure-5 depicts the behaviour of utilisation versus the number of iterations. It is very clear from the figure that the utilization initially around 0.6 as the iteration increased, the utilization is going in a constant zig-zag fashion. The minimum and maximum utilisation are found to be around 65% and 81% respectively by the proposed DRLCA scheme. In the other scenario, the JPCRL algorithm, the minimum and maximum utilization is 60% and 76% respectively. From the analysis, it is concluded that the DRLCA outperformed the JPCRL. There is a direct 5% improvement if the channels are allocated by the DRLCA scheme.

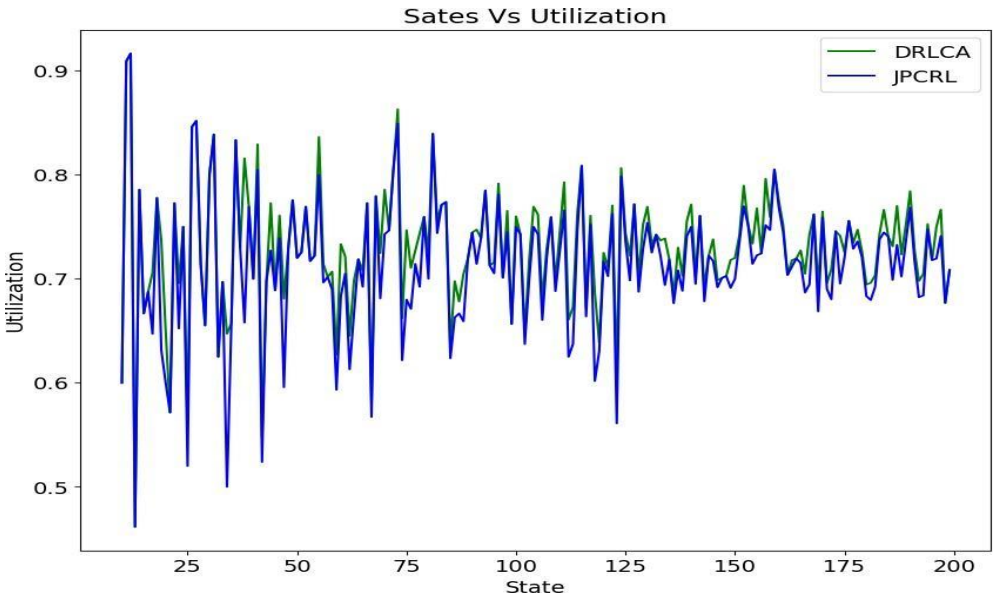

Fig. 6. States vs Iterations

The figure-6 depicts the behaviour of channel utilization and with the number of states. it is found that the mean utilization of the system is approximately 75% if the channels are allocated to the unlicensed users according to the DRLCA, the mean utilization by the JPCRL scheme is found to be 71% as the number of states varies. From the above analysis, it is noticed that the proposed DRLCA scheme performs better than JPCRL scheme.

From the above analysis of figure 4,5,6, it is found that the utilization of the system increased significantly (More than 5%) when the channels were allocated according to DRLCA as compared to DPCRL. As a result, our proposed algorithm provides better utilization and less power consumption.

CONCLUSION

A deep reinforcement learning-based channel allocation scheme (DRLCA) among unlicensed users has been presented in this paper. To develop DRLCA technique, an Interference ratio has been considered to calculate the reward, and based on the optimized reward the Q-value function has also been defined and used to allocate the channels to the secondary users as depicted in algorithm-1. It is analysed that the proposed DRLCA provides better spectrum utilization as compared to the existing JPCRL. The direct 5% improvement is observed if the channels are allocated as per DRLCA. The work has the potential to enhance the deep learning-based algorithm by considering various parameters like power, fairness Interference, etc.

REFERENCES

1. G. Zhao, Y. Li, C. Xu, Z. Han, Y. Xing, and S. Yu, “Joint power control and channel allocation for interference mitigation based on reinforcement learning,” *IEEE Access*, vol. 7, pp. 177254–177265, 2019.

2. P. Kolodzy and I. Avoidance, “Spectrum policy task force,” *Federal Commun. Comm., Washington, DC, Rep. ET Docket*, 2002.

3. L. Lu, X. Zhou, U. Onunkwo, and G. Y. Li, “Ten years of research in spectrum sensing and sharing in cognitive radio.,” *EURASIP J. Wireless Comm. and Networking*, 2012.

4. P. Yang, L. Kong, and G. Chen, "Spectrum sharing for 5g/6g urllc: Research frontiers and standards," *IEEE communications standards magazine*, vol. 5, no. 2, pp. 120–125, 2021.

5. J. Mitola, "Cognitive radio for flexible mobile multimedia communications," in *Mobile Multimedia Communications, 1999.(MoMuC'99) 1999 IEEE International Workshop on*, pp. 3–10, IEEE, 1999.

6. J. Mitola, *Cognitive Radio—An Integrated Agent Architecture for Software Defined Radio*. PhD thesis, Royal Institute of Technology (KTH), 2000.

7. I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey," *Computer Networks*, vol. 50, no. 13, pp. 2127–2159, 2006.

8. X. Li, T. Huang, H. Xiao, and F. Xiao, "Research on cognitive radio spectrum allocation strategy based on machine learning algorithm and data science technology," 2023.

9. S. Haykin, "Cognitive radio: brain-empowered wireless communications," *Selected Areas in Communications, IEEE Journal on*, vol. 23, no. 2, pp. 201–220, 2005.

10. T. Yang, H. Tang, C. Bai, J. Liu, J. Hao, Z. Meng, P. Liu, and Z. Wang, "Exploration in deep reinforcement learning: a comprehensive survey," *arXiv preprint arXiv:2109.06668*, 2021.

11. M. S. Frikha, S. M. Gammar, A. Lahmadi, and L. Andrey, "Reinforcement and deep reinforcement learning for wireless internet of things: A survey," *Computer Communications*, vol. 178, pp. 98–113, 2021.

12. T. Ohtsuki, "Machine learning in 6g wireless communications," *IEICE Transactions on Communications*, vol. 106, no. 2, pp. 75–83, 2023.

13. M. Gheisari, F. Ebrahimzadeh, M. Rahimi, M. Moazzamigodarzi, Y. Liu, P. K. Dutta Pramanik, M. A. Heravi, A. Mehbodniya, M. Ghaderzadeh, M. R. Feylizadeh, *et al.*, "Deep learning: Applications, architectures, models, tools, and frameworks: A comprehensive survey," *CAAI Transactions on Intelligence Technology*, 2023.

14. A. Bhattacharyya, S. M. Nambiar, R. Ojha, A. Gyaneshwar, U. Chadha, and K. Srinivasan, "Machine learning and deep learning powered satellite communications: Enabling technologies, applications, open challenges, and future research directions," *International Journal of Satellite Communications and Networking*, 2023.

15. O. Prakash, P. Pattanayak, A. Rai, and K. Cengiz, "Machine learning and deep reinforcement learning in wireless networks and communication applications," in *Paradigms of Smart and Intelligent Communication, 5G and Beyond*, pp. 83–102, Springer, 2023.

16. M. M. Farouk, W. L. Pang, G. C. Chung, and M. Roslee, "Critical review on machine learning in 5g mobile networks," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 8s, pp. 362–367, 2023.

17. A. M. Al-Ansi, A. Al-Ansi, *et al.*, "An overview of artificial intelligence (ai) in 6g: Types, advantages, challenges and recent applications," *Buletin Ilmiah Sarjana Teknik Elektro*, vol. 5, no. 1, pp. 67–75, 2023.

18. M. Waqas, S. Tu, Z. Halim, S. U. Rehman, G. Abbas, and Z. H. Abbas, "The role of artificial intelligence and machine learning in wireless networks security: Principle, practice and challenges," *Artificial Intelligence Review*, vol. 55, no. 7, pp. 5215–5261, 2022.

19. V. Banerjee and B. Kakde, "Optimization of resource allocation in cognitive radio network using machine learning algorithm," in *Intelligent Sustainable Systems: Selected Papers of WorldS4 2022, Volume 2*, pp. 173–184, Springer, 2023.

20. T. Chen, *Machine learning enhanced resource allocation in wireless networks*. PhD thesis, Loughborough University, 2023.

21. H. H. S. Lopes, F. G. C. Rocha, and F. H. T. Vieira, “Deep reinforcement learning based resource allocation approach for wireless networks considering network slicing paradigm,” *Journal of Communication and Information Systems*, vol. 38, no. 1, pp. 21–33, 2023.

22. G. Schwalbe and B. Finzel, “A comprehensive taxonomy for explainable artificial intelligence: a systematic survey of surveys on methods and concepts,” *Data Mining and Knowledge Discovery*, pp. 1–59, 2023.

23. P. M. Pawar and A. Leshem, “Distributed deep reinforcement learning for collaborative spectrum sharing,” *arXiv preprint arXiv:2104.02059*, 2021.

24. U. Student, “Machine learning based spectrum detection in cognitive radios,”

25. V. Saravanan, P. Sreelatha, N. R. Atyam, M. Madiajagan, D. Saravanan, H. P. Sultana, *et al.*, “Design of deep learning model for radio resource allocation in 5g for massive iot device,” *Sustainable Energy Technologies and Assessments*, vol. 56, p. 103054, 2023.

26. H. Mao, M. Alizadeh, I. Menache, and S. Kandula, “Resource management with deep reinforcement learning,” in *Proceedings of the 15th ACM workshop on hot topics in networks*, pp. 50–56, 2016.

27. J. C. Clement, K. Sriharipriya, P. Prakasam, *et al.*, “Throughput enhancement in a cognitive radio network using a reinforcement learning method,” *Multimedia Tools and Applications*, pp. 1–23, 2023.

28. M. Shin, T. Mahboob, D. M. Mughal, and M. Y. Chung, “Deep reinforcement learning–based multi–channel spectrum sharing technology for next generation multi–operator cellular networks,” *Wireless Networks*, vol. 29, no. 2, pp. 809–820, 2023.

29. W. Lee and J.-B. Seo, “Deep learning-aided channel allocation scheme for wlan,” *IEEE Wireless Communications Letters*, 2023.

30. O. T. H. Alzubaidi, M. N. Hindia, K. Dimyati, K. A. Noordin, A. N. A. Wahab, F. Qamar, and R. Hassan, “Interference challenges and management in b5g network design: A comprehensive review,” *Electronics*, vol. 11, no. 18, p. 2842, 2022.

31. T. Oyedare, V. K. Shah, D. J. Jakubisin, and J. H. Reed, “Interference suppression using deep learning: Current approaches and open challenges,” *IEEE Access*, 2022.

32. A. Doshi, S. Yerramalli, L. Ferrari, T. Yoo, and J. G. Andrews, “A deep reinforcement learning framework for contention-based spectrum sharing,” *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 8, pp. 2526–2540, 2021.

33. M. Bkassiny, Y. Li, and S. K. Jayaweera, “A survey on machine-learning techniques in cognitive radios,” *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 1136–1159, 2012.

34. Y. Wang, Z. Ye, P. Wan, and J. Zhao, “A survey of dynamic spectrum allocation based on reinforcement learning algorithms in cognitive radio networks,” *Artificial Intelligence Review*, vol. 51, no. 3, pp. 493–506, 2019.

35. P. Zhu, J. Li, D. Wang, and X. You, “Machine-learning-based opportunistic spectrum access in cognitive radio networks,” *IEEE Wireless Communications*, vol. 27, no. 1, pp. 38–44, 2020.

36. N. Hosey, S. Bergin, I. Macaluso, and D. O’Donoghue, “Q-learning for cognitive radios,” in *Proceedings of the China-Ireland Information and Communications Technology Conference (CIICT 2009). ISBN 9780901519672*, National University of Ireland Maynooth, 2009.

37. Z. Yin, Y. Wang, and C. Wu, “Reinforcement learning spectrum management paradigm in cognitive radio using novel state and action sets,” *Procedia Computer Science*, vol. 129, pp. 433–437, 2018.

38. E. C. Santos, “A simple reinforcement learning mechanism for resource allocation in lte-a networks with markov decision process and q-learning,” *arXiv preprint arXiv:1709.09312*, 2017.

39. Y. Yao and Z. Feng, “Centralized channel and power allocation for cognitive radio networks: A q-learning solution,” in *2010 Future Network & Mobile Summit*, pp. 1–8, IEEE, 2010.

40. M. Li, Y. Xu, and J. Hu, “A q-learning based sensing task selection scheme for cognitive radio networks,” in *2009 International Conference on Wireless Communications & Signal Processing*, pp. 1–5, IEEE, 2009.

41. F. Li, K.-Y. Lam, Z. Sheng, X. Zhang, K. Zhao, and L. Wang, “Q-learning-based dynamic spectrum access in cognitive industrial internet of things,” *Mobile Networks and Applications*, vol. 23, no. 6, pp. 1636–1644, 2018.

42. K.-L. A. Yau, G.-S. Poh, S. F. Chien, and H. A. Al-Rawi, “Application of reinforcement learning in cognitive radio networks: models and algorithms,” *The Scientific World Journal*, vol. 2014, 2014.

43. R. Politanskyi and M. Klymash, “Application of artificial intelligence in cognitive radio for planning distribution of frequency channels,” in *2019 3rd International Conference on Advanced Information and Communications Technologies (AICT)*, pp. 390–394, IEEE, 2019.

44. B. S. Awoyemi, B. T. Maharaj, and A. S. Alfa, “Solving resource allocation problems in cognitive radio networks: a survey,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, p. 176, 2016.

45. M. N. Pavan, S. Kumar, and G. Nayak, “Interference aware resource allocation (iara) in cognitive radio networks,” in *2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS)*, pp. 202–206, IEEE, 2018.\

46. A. Gattami, Q. Bai, and V. Aggarwal, “Reinforcement learning for constrained markov decision processes,” in *International Conference on Artificial Intelligence and Statistics*, pp. 2656–2664, PMLR, 2021.

47. Q. Hu and W. Yue, *Markov decision processes with their applications*, vol. 14. Springer Science & Business Media, 2007.

48. M. Van Otterlo and M. Wiering, “Reinforcement learning and markov decision processes,” in *Reinforcement learning: State-of-the-art*, pp. 3–42, Springer, 2012.

49. T. P. Le, N. A. Vien, and T. Chung, “A deep hierarchical reinforcement learning algorithm in partially observable markov decision processes,” *Ieee Access*, vol. 6, pp. 49089–49102, 2018.

50. E. Mizutani and S. Dreyfus, “A tutorial on the art of dynamic programming for some issues concerning bellman’s principle of optimality,”