DEEP Q NETWORK IN DYNAMIC SPECTRUM ACCESS：A BRIEF SURVEY

Jing XIA[1], Zheng DOU[2] and Lin QI[3]

[1]College of Information and Communication Eng., Harbin Engineering University
(145 Nantong St., Harbin, Heilongjiang, China)
E-mail: xiajing@hrbeu.edu.cn
[2] Professor, College of Information and Communication Eng., Harbin Engineering University
(145 Nantong St., Harbin, Heilongjiang, China)
E-mail: douzheng@hrbeu.edu.cn
[3]Professor, College of Information and Communication Eng., Harbin Engineering University
(145 Nantong St., Harbin, Heilongjiang, China)
E-mail: qilin@hrbeu.edu.cn

Spectrum is a limited and non-renewable natural resource. The increasing demand for wireless communication, along with spectrum scarcity, has triggered the development of efficient dynamic spectrum access (DSA) schemes for emerging wireless network technologies. Deep Q network (DQN), as a branch of machine learning, has shown good performance in solving dynamic spectrum access problems in recent years. This paper presents the review of literature on applications of deep Q network in dynamic spectrum access. In this survey, we first reviewed the development of DQN algorithm addressing emerging issues in dynamic spectrum access. Furthermore, we categorized applications of optimized DQN algorithms in dynamic spectrum access according to environment models, algorithm models and learning models. Finally, we highlighted future prospects of applying DQN.

Key Words : *dynamic spectrum access, reinforcement learning, deep Q learning*

1. INTRODUCTION

The wireless communication business has grown exponentially due to the rapid development of wireless communication, which needs to occupy a wider spectrum of resources. However, the current allocation of spectrum resources is statically authorized. Although the static spectrum allocation method can effectively avoid the conflict and interference between different wireless services, it cannot give full play to the distribution characteristics of radio signals in the time domain, frequency domain and airspace[1-2)].

Dynamic spectrum access (DSA) is the key to realize efficient utilization of spectrum resources in cognitive radio[3)]. (In DSA, the term spectrum refers to the bandwidth of frequency points in mobile communication, which is divided in the form of channels). It allows secondary users (SUs) to sense the spectrum holes in the current frequency band without affecting the service quality of the normal communication of the primary user (PU), and to change their own access parameters under certain conditions to gain access to the spectrum, so as to improve the spectrum utilization[4-5)].

The application of deep Q network (DQN) in dynamic spectrum access can enable secondary users to adapt and learn different channel environments. Also, through modifying their own access parameters under different conditions, secondary users could achieve the optimal spectrum access effect, which is consistent with the original intention of dynamic spectrum access[6)]. In this paper, we present the survey with the development of DQN algorithm and the literature review on the applications of DQN to address issues in dynamic spectrum access.

2. DQN ALGORITHM IN DYNAMIC SPECTRUM ACCESS

(1) Deep Q network algorithm

DQN algorithm belongs to the category of deep reinforcement learning (DRL). Proposed by Google DeepMind, DRL is a creative combination of deep

learning method with strong perception and reinforcement learning method with excellent decision-making ability[7]. In 2013, DeepMind presented DQN algorithm arousing wide concern[8]. DQN algorithm is developed from Q learning algorithm, and their overall algorithm framework is similar. Q learning can deal with decision problems well[9]. An important recursive relation in reinforcement learning is Behrman equation:

$$Q^{\pi}(s_t,a_t)=E\left[r+\gamma E\left[Q^{\pi}(s_{t+1},a_{t+1})\right]\right] \quad (1)$$

Where γ is the discount factor, which represents the Q value obtained by taking action a in state s in the case of applying strategy π. From strategy π we can see, the expectation will take the randomness of future actions and the return state from the environment into account. According to equation (1), the next state is only related to the current state, which is the Markov property. Most reinforcement learning methods are based on this formula. The update rules of Q learning algorithm are as follows:

$$Q(s,a)=Q(s,a)+\alpha\left[r+\max_{a(t+1)}Q(s,a)-Q(s,a)\right] \quad (2)$$

However, more time is needed to query the Q table to obtain the optimal solution in large action and state spaces. One possible approach is function approximation. DQN uses neural networks to approximate the Q value. The weights and offsets are expressed in terms of θ. The loss function is as follows:

$$L(s,a|\theta_i)\approx\left(r+\gamma\max_{a'}Q(s',a|\theta_i)-Q(s,a|\theta_i)\right)^2 \quad (3)$$

Gradient update formula:

$$\theta_{i+1}=\theta_i+\alpha\nabla_{\theta}L(\theta_i) \quad (4)$$

The training process of DQN is shown in the figure below:

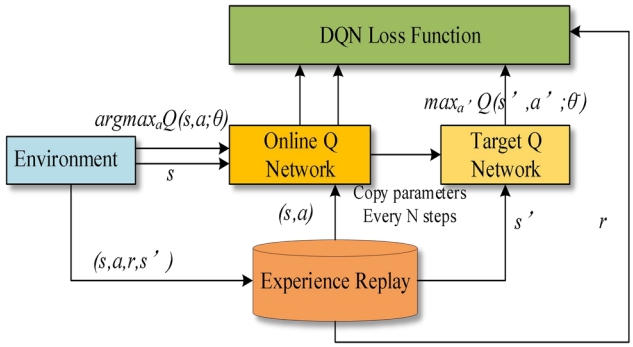

Fig.1 The training process of DQN.

Compared with the traditional Q learning algorithm, DQN has the following advantages. First, every data sample is likely to be extracted from the experience replay of DQN when the weight parameter is updated, which improves the utilization rate of data. Second, every time a sample is randomly selected, which makes the learning process smoother and avoids parameter oscillation and divergence. Third, DQN independently sets the target network and copies the parameters of the online Q network to the target Q network every once in a while, reducing the correlation between the current Q value and the target Q value.

(2) Development of DQN in dynamic spectrum access

In this section, we reviewed the history and development of DQN algorithm in dynamic spectrum access. The reviewed approaches are summarized along with the references in Table 1.

DQN algorithm belongs to reinforcement learning. In this aspect, the theoretical basis is the Markov decision process (MDP). Most reinforcement learning methods are modeling based on MDP. Through modeling the problem of dynamically choosing proper channels to access for secondary users (SUs), the optimal solution would be found.

Table 1 A summary of development for DQN in DSA.

Algorithms	Ref./ Time	Research methods	Model
MDP/ POMDP	10) 2008	A summary of development for DQN in DSA.	Gilbert-Elliot channel model
	11) 2013	Proposed and evaluated different observation policies for the spectrum framework	Centralized model
Q Learning	12) 2011	A bidding algorithm for SUs to learn from their competitors and place better bids	Auction model
	13) 2014	Adjusted reward functions for improvement of channel throughput and average system channel capacity	Multi-channel model
	14) 2015	Distributive Q learning	Cognitive cellular systems
	15) 2017	A co-operative Q-learning based spectrum sensing technique for the secondary users of an ad hoc network	Add-hoc nodes net
Deep Q Learning	16) 2012	Combined Q-learning and back propagation	CWLAN-FDA
	17) 2017	Cognitive channel selection method used deep Q learning	Distributed model
	18) 2018	Fully autonomous distributive underlay DSA algorithm with the neural network	Interference limitation model

The establishment of MDP model needs to know the prior knowledge of the channels, which is usually changing in fact and difficult to obtain. In 2008, a multi-channel access strategy based on MDP was proposed[10]. By adopting the short-sighted perception strategy, the channel selection was simplified to a cyclic process, which avoided the need of channel transfer probability. In 2013, based on the establishment of the partially observable Markov decision process (POMDP), the different

observation strategies were evaluated, and the proposed strategy that could achieve similar performance under the condition of less understanding of external signaling interference was proved[11].

An important element of MDP is the transition probability of the next action, but in reality this is difficult to obtain. So more studies turn to Q learning algorithm, which is a model-free algorithm without the needs of the transition probability. In 2011, a bidding algorithm based on Q learning was proposed[12]. The proposed method enables secondary users to learn from competitors and bid for frequency bands automatically. In 2014, the authors improved the reward function in Q learning for the single-user access strategy in multiple channels[13]. In 2015, the distributed Q learning algorithm was developed for the cognitive cellular system to improve the access density of cognitive users[14]. Although Q learning algorithm can be combined with other models (such as auction model) or improved algorithm performance by parameters (learning rate, conversion factor, reward function), it cannot handle large state and action spaces. Therefore, DQN algorithm is applied to dynamic spectrum access.

Before the DQN algorithm was proposed in 2013, an algorithm similar to DQN was successfully applied in DSA[16]. The algorithm replaces the Q table with a multilayer forward-propagation neural network by combining Q-learning and back propagation, which reduces the external signal interference and improves the network performance in cognitive wireless local area network with fibre-connected distributed antennas (CWLAN-FDA). In 2017, combining with DQN algorithm, a cognitive channel selection method for user service quality was presented. The access efficiency of the proposed algorithm was proved under the model of a single channel simplified network[17].

Summary: DQN, developed from Q learning, uses MDP theory to model the dynamic spectrum access problem. However, DQN algorithm still has some problems such as overestimation and low utilization of excellent samples. In addition to the improvement of algorithm models, existing research has also optimized the environment models and learning models. In the next section, we will summarize the optimization of these models.

3. OPTIMIZED DQN MODELS IN DYNAMIC SPECTRUM ACCESS

In recent years, there are many researches based on DQN in dynamic spectrum access. These investigations optimized the system performance of cognitive radio networks. They can be divided into environment models, algorithm models and learning models.

(1) Environment models

The dynamic spectrum access models are classified into the following categories according to different ways in Table 4 (among which there is no conflict between any two classifications).

- **Cooperative or non-cooperative** (according to whether there is information interaction between cognitive users).
- **Overlay, underlay or inter-weave** (according to how secondary users share spectrum with primary users).
- **Centralized or distributed** (according to whether the access parameters of secondary users are determined by the central controller).

At present, the most common system model of DSA algorithm is non-cooperative distributed model, because it does not need a central controller, which saves expensive overhead cost, and there is no need for information interaction between users [22]. This can reduce the complexity of the algorithm for scenes with complex models and a large number of users. In 2019, Naparstek Oshri et al. presented a centralized learning distributed access model, which effectively improved the system throughput rate and the network utilization efficiency could reach 80%[19]. Here, we present a brief description according to the system scenarios in this paper, because this paper gives a comprehensive feasible analysis of the dynamic spectrum access algorithm based on DQN. In each time slot, cognitive users use DQN network to map the current state to the spectrum access action in order to maximize the objective function, with a system model of centralized training and distributed access.

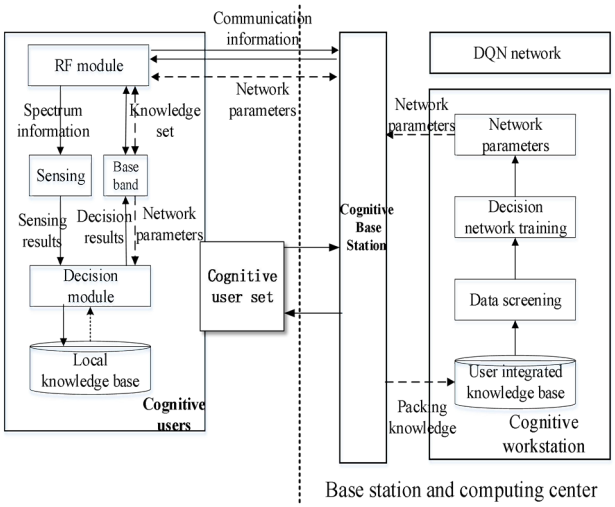

Fig.2 Cognitive network work scenarios.

Table 2 A summary of environment and algorithm DQN models in dynamic spectrum access.

Models		References	Features
Environment models	Distributive model	18-22),26)-29)	No central controller is required; SUs decide the access parameters by themselves
	Centralized model	23),25),28)	The access parameters of the SUs are determined by the central controller
	Cooperative model	28)	Users can coordinate information online
	Non-cooperative model	19),21),22),26),27),29)	Save the cost of information exchange process
	Overlay model	19),21),22)	SUs with the goal of minimizing disruption to PUs
	Underlay model	25),27)	To improve the efficiency of spectrum utilization
	Inter-weave model	23),29)	Combine overlay with underlay ones
Algorithm models	Double DQN	22),24),32)	Deal with the problem of overestimation
	Dueling DQN	22),27),29),30),32)	Improve the learning effect of SUs
	PER-DQN	24-25)	Take advantage of good samples
	DQN with dynamic learning rate	31)	To improve the convergence speed and reduce the curve jitter amplitude
	DQN with RNN	21-22),25-30),32)	Traditional DQN uses CNN. The combination of recurrent neural network can improve the performance of prediction for network.

As is shown in Figure 2, the cognitive users select frequency points independently and communicate with the cognitive base station. The cognitive workstation collects the access knowledge in the user's access process, trains the decision module through this knowledge, and transmits the trained decision parameters to the cognitive users through the cognitive base station and updates the decision module.

(2) Algorithm models

Algorithm models are classified according to the optimization of the DQN algorithm models, because there are advanced DQN models applied into dynamic spectrum access continuously.

Table 3 The performance of advanced DQN algorithm models.

Advanced DQN models	Key features	Application
Double DQN	Use 2 Q-value functions to simultaneously select and evaluate action values.	Applicable to almost MDPs
Dueling DQN	Use 2 networks to estimate the action and state value functions	For MDPs with large action and state spaces
DQN with the prioritized experience replay	Prioritize experiences in replay memory	For MDPs with prioritized experiences
Averaged-DQN	The Q estimates are averaged to reduce the error of the approximate variance	For steady training process of DQN
Distributed DQN	Distributive perspective	To observe the risks behind the actions

The performance of advanced DQN models is listed in Table 3. The first three advanced models are often used in dynamic spectrum access algorithms.

- **Double DQN** can deal with the problem of the overestimated Q value well. In DSA algorithms, although overestimation can sometimes have positive effects, secondary users cannot be effectively trained if the overestimation is uneven and is not focused on the desired state.
- **Dueling DQN** divided the Q value output by the neural network into two parts: the value of the state and the advantage value of each action. Therefore, the action value of the same reward generated by the secondary user during the access channel will be converted to the value of the calculated state. The learning effect of secondary users can be greatly improved.
- **DQN with the prioritized experience replay (PER-DQN)** can improve the role of important samples by priority sampling in the training process. In DSA algorithms, PER-DQN achieve the goal of accelerating the convergence speed and reducing the training time.

The proposed DQSA algorithm applied for general large and complex settings and does not require online coordination or message exchanges between users[19]. Double DQN algorithm was used to select the overestimated Q values which degrades performance. To solve the problems of the low frequency utilization rate, the insufficient important experience utilization rate and the slow convergence speed, a DSA algorithm based on dueling DQN with prioritized experience replay was deployed [24]. In order to break the sample correlation and make full use of the important experience samples, a priority-based sampling method was adopted when

the SUs are sampling the experience database, which improved the throughput of the cognitive system.

The above optimized DQN models have shown good results in dynamic spectrum access, but in recent years, the method combining DQN and recurrent neural network (RNN) has also been widely used. In DQN algorithm, the neural network means convolutional neural network (CNN). However, a special type of RNN, called the reservoir computing (RC) is proposed[21]. It is utilized to realize DQN by taking advantage of the underlying temporal correlation of the DSA network. Results suggest that the RC-based spectrum access strategy can help the SU to significantly reduce the chances of collision with PUs and other SUs. The author proposed a spectrum resource allocation algorithm for long short term memory (LSTM) with DQN[25]. The LSTM layer is used to for SUs to adaptively adjust the transmit power while successfully accessing the channel and the utilization of spectrum resources. Because LSTM is a kind of RNN, which can handle time sequence problems well. The combination of LSTM and DQN makes SUs have the ability of predicting channel states (Table 4).

(3) Learning models

Learning models refer to the model in which agents learn methods from the external environment in DQN algorithm. Different learning models of agents have different effects on improving the system performance of cognitive radio network. Therefore, representative learning models are selected for review in this section. From Table 4 we observe that the problems are mostly modeled as an MDP or POMDP. Channel state information (CSI) is often used as states of the algorithm. The difference is that reward functions instruct agents to learn from the environment.

Table 4 A summary of DQN learning models.

Ref.	Model	Learning Algorithms	Agent	States	Actions	Rewards	Networks
19)	Game	DDQN and Dueling DQN	SUs	Previous access actions and observations	Spectrum access actions	Data rate	Cognitive radio network (CRN)
20)	Game	DQN with LSTM	Cognitive base station	Traffic history of cognitive base station	Spectrum access probability	Throughput	LTE (long term evolution) network
27)	MDP	DQN with echo state networks	DSA users	Channel state of the link and feedback from PUs	Spectrum access and power control	Data rate enhancement and PU protections	Distributed DSA networks
29)	POMDP	DQN with LSTM	SU	Channel occupancy of PUs and the acknowledgments (ACKs) of the SUs.	the access channel and access mode	Reward +1 or -1	CRN
30)	POMDP	Deep Recurrent Q-Networks	Secondary nodes	Sequential observations of channel states	Transmission in a particular band or waiting	r_{t+1} based on successful or unsuccessful packets	CRN
33)	MDP	DQN with LSTM	Base station	Channel access history and CSI	Sensor selection for channel access	Total rate and prediction error	Internet of things

4. FUTURE PROSPECTS

In the previous sections, we presented an overview and a brief description of the proposed DQN algorithms in dynamic spectrum access. Nevertheless, several open issues remain.

(1) In practical applications, different wireless channel environments need to be considered. Most DQN-based dynamic spectrum access algorithms use the Rayleigh fading channel model, with little consideration of other models (such as the Nakagamim channel model). In this regard, it is necessary to test different wireless channel models, because CR entity communication affects the characteristics of wireless channels in fact.

(2) The dynamic spectrum access process needs to consume a large amount of perception energy, which is a huge cost for the system. It can be considered to combine spectrum prediction before spectrum perception, but there is still a lack of system-level overall scheme in this regard.

(3) Most DQN models are based on single-agent algorithm. In fact, in the process of dynamic spectrum access, secondary users play the role of cooperation, communication and competition, which is consistent with the idea of multi-agent algorithm. Collective behavior and distributed control systems are important examples of multiple autonomous agents in a dynamic environment. Therefore, the problem of multi-agent system will gradually become a hot research direction.

REFERENCES

1) Zhu J R, Zhang Z W, Wu Z G, et al. Evaluation of spectrum resource utilization in Ningbo Lingang area [J].China Radio, 2015(1): 24-25.
2) Guo Xinyuan. Current Situation and Characteristic Analysis of Radio Spectrum Resource Management in Changji Prefecture. China Radio, 2018, No.273(5): 38-40+49.
3) Kaur R, Buttar A S, Anand J. Spectrum sharing schemes in cognitive radio network: A survey[C]//2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA). IEEE, 2018: 1279-1284.
4) Xing Y, Chandramouli R, Mangold S, et al. Dynamic spectrum access in open spectrum wireless networks[J]. IEEE Journal on Selected Areas in Communications, 2006, 24(3):626-637.
5) Haykin, S. Cognitive radio: brain-empowered wireless communications[J]. IEEE Journal on Selected Areas in Communications, 2005, 23(2).
6) Luong N C, Hoang D T, Gong S, et al. Applications of deep reinforcement learning in communications and networking: A survey[J]. IEEE Communications Surveys & Tutorials, 2019, 21(4): 3133-3174.
7) Arulkumaran K, Deisenroth M P, Brundage M, et al. A brief survey of deep reinforcement learning[J]. arXiv preprint arXiv:1708.05866, 2017.
8) V. Mnih et al., “Playing Atari with deep reinforcement learning,” CoRR, vol. abs/1312.5602, pp. 1–9, Dec. 2013.
9) X. Gao, Z. Dou and L. Qi, "A New Distributed Dynamic Spectrum Access Model Based on DQN," 2020 15th IEEE International Conference on Signal Processing (ICSP), Beijing, China, 2020, pp. 351-355.
10) Q. Zhao et al., “On myopic sensing for multi-channel opportunistic access: Structure, optimality, and performance,” IEEE Trans. Wireless Communication., vol. 7, no. 12, pp. 5431–5440, Dec. 2008.
11) A. Raschellà, J. Pérez-Romero, O. Sallent and A. Umbert, "On the impact of the observation strategy in a POMDP-based framework for spectrum selection," 2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), 2013, pp. 2512-2516.
12) Chen Z, Qiu R C. Q-learning based bidding algorithm for spectrum auction in cognitive radio[C]//2011 Proceedings of IEEE Southeastcon. IEEE, 2011: 409-412.
13) Zhao B, Li O, Luan H. Application of Q learning algorithm in Chance Spectrum Access Channel Selection [J]Signal Processing,2014,30(03):298-305.
14) Morozs N, Grace D, Clarke T. Distributed Q-Learning Based Dynamic Spectrum Access in High Capacity Density Cognitive Cellular Systems Using Secondary LTE Spectrum Sharing[C]. International Symposium on Wireless Personal Multimedia Communications, 2015.
15) Das A, Ghosh S C, Das N, et al. Q-learning based co-operative spectrum mobility in cognitive radio networks[C]//2017 IEEE 42nd Conference on Local Computer Networks (LCN). IEEE, 2017: 502-505.
16) Yi L I, Hong J I. Q-learning for dynamic channel assignment in cognitive wireless local area network with fibre-connected distributed antennas[J]. The Journal of China Universities of Posts and Telecommunications, 2012, 19(4): 80-85.
17) Uyanik G S, Oktug S. Cognitive channel selection and scheduling for multi-channel dynamic spectrum access networks considering QoS levels[J]. Ad Hoc Networks, 2017, 62: 22-34.
18) Mohammadi F S, Kwasinski A. Neural Network Cognitive Engine for Autonomous and Distributed Underlay Dynamic Spectrum Access[J], 2018.
19) Naparstek O, Cohen K. Deep Multi-User Reinforcement Learning for Distributed Dynamic Spectrum Access[J], 2019.
20) U. Challita, L. Dong, and W. Saad, “Proactive resource management in LTE-U systems: A deep learning perspective,” arXiv preprint arXiv:1702.07031, pp. 1–29, Dec. 2017.
21) H. Chang, H. Song, Y. Yi, J. Zhang, H. He and L. Liu, "Distributive Dynamic Spectrum Access Through Deep Reinforcement Learning: A Reservoir Computing-Based Approach," in IEEE Internet of Things Journal, vol. 6, no. 2, pp. 1938-1948, April 2019.
22) Naparstek O, Cohen K. Deep Multi-User Reinforcement Learning for Dynamic Spectrum Access in Multichannel Wireless Networks[J], 2017.
23) J. Zhao, W. Shao, F. Li and Q. Zhou, "A Spectrum Handoff Method Based on Reinforcement and Transfer Learning," 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications, Dalian, China, 2020, pp. 572-575.
24) Pan Xiaona. A Spectrum Access Algorithm Based on Preference Experience Playback Deep Q-Learning [J]. Telecommunication Technology,20,60(05):489-495.
25) Ye Zifeng. Research on Dynamic Spectrum Allocation Method Based on Deep Reinforcement Learning [D].Guangdong University of Technology,2019.
26) Jinming Xu, Zheng Dou, and Lin Qi. Multi-User Dynamic Spectrum Access Based on Reinforcement Learning[C]. Eleventh International Conference on Graphics and Image Processing. Vol. 11373: SPIE, 2020.
27) H. Song, L. Liu, J. Ashdown and Y. Yi, "A Deep Reinforcement Learning Framework for Spectrum Management in Dynamic Spectrum Access," in IEEE Internet of Things Journal.
28) U. Kaytaz, S. Ucar, B. Akgun and S. Coleri, "Distributed Deep Reinforcement Learning with Wideband Sensing for Dynamic Spectrum Access," 2020 IEEE Wireless Communications and Networking Conference (WCNC), Seoul, Korea (South), 2020, pp. 1-6.
29) N. Yang, H. Zhang and R. Berry, "Partially Observable Multi-Agent Deep Reinforcement Learning for Cognitive Resource Management," GLOBECOM 2020 - 2020 IEEE Global Communications Conference, Taipei, Taiwan, 2020, pp. 1-6.
30) Y. Xu, J. Yu and R. M. Buehrer, "Dealing with Partial Observations in Dynamic Spectrum Access: Deep Recurrent Q-Networks," MILCOM 2018 - 2018 IEEE Military Communications Conference (MILCOM), 2018, pp. 865-870.
31) XU Jinming. Research on Non-cooperative Dynamic Spectrum Access Algorithm Based on Reinforcement Learning [D]. Harbin Engineering University,2020.
32) N. Zhao, D. Niyato, Y. Pei, M. Wu, and Y. Jiang, “Deep reinforcement learning for user association and resource allocation in heterogeneous networks,” in Proc. IEEE GLOBECOM, Abu Dhabi, UAE, Dec. 2018, pp. 1–6.
33) M. Chu, H. Li, X. Liao, and S. Cui, “Reinforcement learning based multi-access control and battery prediction with energy harvesting in IoT systems,” IEEE Internet Things J., vol. 6, no. 2, Apr. 2019.