Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 9653334, 12 pages
https://doi.org/10.1155/2022/9653334

Review Article

Deep Q-Learning for Intelligent Band Coordination in 5G Heterogeneous Network Supporting V2X Communication

Hua-Min Chen,[1] **Shou-Feng Wang**,[2] **Peng Wang**,[3] **Shaofu Lin**,[1] **and Chao Fang**[1,4]

[1]*Faculty of Information Technology, Beijing University of Technology, Beijing, China*
[2]*Beijing Samsung Telecommunication R&D Center, Beijing, China*
[3]*Beijing Institute of Remote Sensing Equipment, Beijing, China*
[4]*Purple Mountain Laboratory: Networking, Communications and Security, Nanjing, China*

Correspondence should be addressed to Shaofu Lin; linshaofu@bjut.edu.cn

Received 12 October 2021; Accepted 8 March 2022; Published 6 April 2022

Academic Editor: Abdul Basit

A heterogeneous or hybrid 5G network is required to support connected vehicles to implement the full range of cooperative ITS (intelligent transport system) services in diverse scenarios. In order to enhance data rate or reduce latency by increasing transmission bandwidth, 5G utilizes frequency bands below and above 6 GHz. The challenge is that multiple band coordination in 5G will be essential to mobile network operators. Even worse, traditional strategies could not meet the demand. Most current 5G research is focused in 5G network optimization. However, frequency coordination in 5G, as one of the most important requirements from operators, is left untouched. In this paper, a multi-agent deep Q-learning network (DQN) is developed as coordination solution. Transfer learning is introduced in DQN to decrease the deployment complexity of the proposed solution on 5G gNB (next-generation NodeB). By deploying the proposed solution in the system level simulation, the simulation shows an average 10% throughput enhancement, an about 24% accessed user number increasing, and around 70% training time saving, compared with normal Q-learning solution, and it enables the operators to optimally utilize all the valuable frequency resources to the best commercial value.

1. Introduction

With the coming era of 5G, the topic of frequency band coordination strategy renewal becomes one of the hottest concerns in 5G instruction. According to 3GPP TS.38.104 [1], typical assigned frequency bands for existing networks are appropriate for 5G. Some operators who have multiple networks already consider shifting partial frequency resource from non-5G networks to 5G network. Aside strategy defining within already assigned frequency bands, 5G could utilize frequency bands above 6 GHz, which is also named as millimeter-wave (mm-wave or MMW) band to support extremely high throughput and low latency since much more wide bandwidth is available in MMW [2–4]. Then, how to utilize 5G frequency bands in 5G heterogeneous network is regarded as an emerging key requirement for operators in network optimization.

In particular, in the case of V2X communication [5–8], the corresponding intelligent transport system (ITS) relies on a hybrid or heterogeneous network providing full range of necessary services and has diverse requirements on the latency, throughput, and access number of vehicles. According to the ISO standard [9] and ETSI specification [10], a heterogeneous network is required. It means that higher frequency bands are for higher data rate, and lower frequency is for coverage. Moreover, the interference increases with user number and more resource is needed to guarantee the performance. According to the traffic characteristics [5, 11–13], V2X traffic is real time and radio frequency is a big challenge. Therefore, band coordination is necessary for V2X service in a heterogeneous network.

1.1. Unspoken Concern from Operators in Networking. Network coverage and capacity are the two main characters in networking. Each operator tries the best way to fulfill the requests on both coverage and capacity. However, it is mutually contradictory to realize both of them with one single network. To satisfy coverage request, the coverage of one base station needs to be as large as possible. To guarantee capacity request, that is to say, to provide as many frequency resources as possible to obtain throughput within a small area, the coverage of one cell should be small. To solve this contradiction, heterogeneous network is widely utilized among operators.

A heterogeneous network has at least two overlay networks covering the same area. To make the overlay networks work together, different frequencies should be adopted for different overlay networks. Due to the limited frequency bands assigned to mobile network operators, a typical heterogeneous network has two frequency bands at least with each for one layer, wherein the lower frequency band is usually applied for coverage guarantee layer and the higher frequency band is equipped in the layer for hotspot capacity usage.

This simple strategy is widely used among operators for 2G, 3G, and 4G [14, 15]. Some optimizations on this strategy mainly focus on the manners of base station deployment within one layer to enhance network performance. Few strategy optimizations are found for updating frequency usage. However, 5G frequency band coordination will be more complex as each layer in a 5G heterogeneous network will have more frequency bands to select. It is a pity that few researches contribute to this field.

1.2. The New Requirement from Operators on Network Optimization for 5G Heterogeneous Network. Some mobile network operators, such as LG, U+, SK, AT&T, and CMCC, already own more frequency bands in their 3G/4G networks among all the operators. Before 5G commercialization, these operators already consider shifting partial frequency resource from non-5G networks. The earlier to be ready for 5G frequency resource deployment and networking strategy, the more opportunity to gain the leading position in 5G area. Recently, raised hot topics related to frequency refarming are mainly consequences under this motivation [16, 17].

More importantly, frequency utilization solution has been one of the top-level business strategies in the operators. Besides the published papers on reusing the frequency for 5G, the topic of how to utilize multiple frequency bands in 4G/5G heterogeneous network to gain more network throughput and how to fulfill users' quality of experience (QoE) has been discussed in the past 2 to 3 years among the operator top managers. Because the frequency bands are different in the world, the frequency usage strategies vary in regions.

Take widely accepted strategy in 3G/4G for example. In America, Band 17 (uplink (UL), 704-716 MHz; downlink (DL), 734-746 MHz) is more likely used for coverage, and Band 4 (UL, 1710-1715 MHz; DL, 2110-2115 MHz) is welcomed to be employed for hotspots. Most countries in Europe prefer Band 20 (UL, 832-862 MHz; DL, 791-821 MHz) as the frequency band for coverage, and they have three bands for hotspots including Band 3 (UL, 1710-1785 MHz; DL, 1805-1880 MHz), Band 7 (UL, 2500-2570 MHz; DL, 2620-2690 MHz), and Band 38 (UL, 2570-2620 MHz; DL, 2570-2620 MHz; it is time division duplex (TDD)). Japan has different choices. Band 1 (UL, 1920-1980 MHz; DL, 2110-2170 MHz) is used for coverage while Band 41 (UL, 2496-2690 MHz; DL, 2496-2690 MHz; TDD) is regarded as the frequency band for hotspots in Japan.

However, the strategies for 5G are not generated now. One cause is that the channel condition in different MMW above 6 GHz (such as 28 GHz and 49 GHz) varies, which results in the strategies for 5G really hard to be made. The other cause is that current researches are attracted by high frequency in 5G, lacking the study on 5G heterogeneous networking optimizing since MMW bands are never used by mobile network operators before. Recently, the operators put their priority concern back to 5G networking optimization techniques.

1.3. Challenges and Complexity in 5G Multiple Bands Coordination in Heterogeneous Network. The flourishing of frequency bands of 5G, growing demand for QoE of users, and network performance enhancement result in the frequency selection strategy hard to make. According to the analysis above, the design for 5G multiple band coordination in a heterogeneous network faces many challenges:

(1) The optimal frequency band selection among multiple candidates in each layer is hard to evaluate. Traditional heterogeneous networking only considers maximum two layers of network, each with one single frequency band. But for 5G, more than two layers would be normal case. Different frequency bands have various coverage capabilities; more cells are requested for cell handover between different layers. In the meantime, networking should always keep good communication continuity. It makes networking policy very hard to choose the balance among coverage, throughput, and continuity. According to the analysis in [18, 19], the resource allocation is affected by the power allocation which differs considerably for different UEs (user equipment) as a result of significant of shadowing and pathloss. Therefore, the overall system throughput cannot be guaranteed.

(2) QoE in 5G network request needs further enhancement than other elder generation networks. Besides eMBB (Enhanced Mobile Broadband) service, high reliability and low latency services, such as uRLLC (ultrareliability and low latency) [2, 20–22], are also highly valued. Therefore, traditional heterogeneous networking should reevaluate its strategies on cosite design for different layers of network. A joint resource allocation is presented in [20] to maximize system throughput with different service types, but the cell association is restricted by assigning UEs to

specific carrier without freedom to select appropriate bands. Further, the networking of too many cosited high-frequency band cells with low-frequency band cells may not work with requirement from high reliability and low latency services. As analyzed in [23, 24], some advanced resource allocation is required with the fact that traffic status and channel state of UEs vary in different bands.

(3) The conflict requirement between operator targets and individual requests. Operators always need the whole network with high overall throughput and low energy consumption, while individual subscribers want their own throughput as high as possible and latency as low as possible. This factor should also be carefully taken into account when 5G heterogeneous networking introduces more diverse frequency bands [24].

1.4. Main Contribution. A novel solution for 5G multiple band coordination is proposed in this paper. The main contribution is as follows:

(1) A multi-agent deep Q-learning network (DQN) is designed for multiple frequency band decision-making for each user. "Experience replay" method is then deployed to accelerate learning speed. By introducing DQN, our solution could intelligently perform the optimal frequency band solution. To the best of our knowledge, the proposed solution is a pioneer in the field of utilizing DQN and transfer learning for intelligent frequency band coordination in 5G heterogeneous network, especially in V2X system

(2) QoE is used in the optimization evaluation process. Both mean opinion score (MOS) and throughput are used as key performance indicators. Signal to interference plus noise ratio (SINR) is used as a reference of received signal condition. In this way, the request to enhance subscriber experience is fulfilled

(3) "Transfer learning" in DQN is adapted to reduce the complexity of calculation for each user. Then, both the balance of user level performance and network level performance can be obtained

The rest of this paper is organized as follows. Section 2 depicts the system model of 5G heterogeneous network with multiple bands, as well as the quantity model of MOS for multiple services. The model of multi-agent DQN, Q-learning model, and the DQN with "experience replay" method and "transfer learning" are exposed in Section 3 in detail. The evaluation and result analysis are taken in Section 4, and the conclusion is drawn in Section 5.

2. System Model

The rest of this section will show the 5G wireless resource and networking and user QoE-related model. The 5G wireless resource consists of 5G frequency bands, heterogeneous networking, and the deployment scenario. These contents correspond to Section 2.1. The user of QoE-related model includes the SINR calculation and transmission rate estimation and MOS models. They are introduced in Sections 2.2 and 2.3, respectively.

2.1. 5G Heterogeneous Network. Frequency band coordination is first introduced in traditional heterogeneous network. A heterogeneous network usually consists of two layers. Each layer is composed of base stations assigned with non-overlapping frequency band (or bands). Traditionally, one layer is designed for network coverage guarantee and the other layer is to fulfill hotspot capacity requirement. The base station in each layer could either be cosite or non-cosite, as shown in Figure 1. In the figure, all vehicles can be accessed to the macro cell with a low-frequency band F1 and can be offloaded to the micro cells with higher frequency bands (F2 and F3). Further, different UEs can be allocated different resources within different frequency bands. For example, data transmission for UE1 is performed over F2 and F4, while UE2 and UE3 are served through an aggregation of F2 and F3, according to the traffic status and resource allocation of each band.

With the abovementioned features, the proposed intelligent band coordination function is performed per gNB (next-generation NodeB) in this paper. The action point of frequency band coordination function in gNB working flow is shown in Figure 2. As depicted, the proposed intelligent band coordination is triggered after channel condition estimation. In channel condition estimation, gNB select a target SINR for each user, from a discrete target SINR list. The intelligent band coordination will calculate the best frequency band for each user according to the chosen target SINR. The detailed procedure of our proposal will be introduced in detail later in Section 3. The coordination result is sent back to access control. In this way, gNB could intelligently select suitable frequency for each user.

2.2. SINR Calculation and Transmission Rate Estimation. The SINR for a user could be calculated by

$$\mathrm{SINR} = \frac{G_i \cdot P_i}{\sigma^2 + \sum_{\substack{j=1 \\ j \neq i}}^{N_{\mathrm{UE}}} G_j \cdot P_j}, \tag{1}$$

where G_i and G_j are the channel gain for user i and user j; P_i and P_j are transmission power for user i and user j; σ^2 stands for the background noise power and N_{UE} is the user number. The user could obtain the channel state information (CSI) via active learning, from which the channel gain is estimated [25].

To guarantee the user traffic quality, the SINR shall satisfy a minimum threshold. There is a discrete target SINR list for a user to select according to its channel condition. If the channel condition is good, the user could select high SINR to obtain high data rate. If the channel condition is poor, the user could select low SINR, but the chosen target SINR shall not be smaller than the minimum SINR

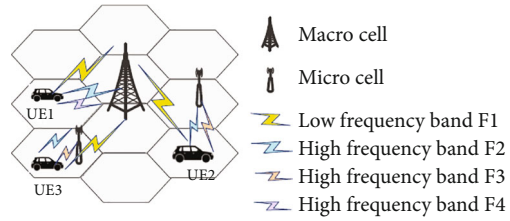

Figure 1: System model.

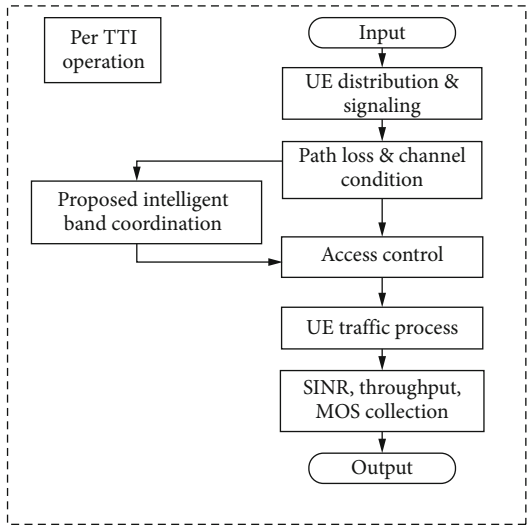

Figure 2: The action point of frequency band coordination in gNB working flow.

threshold for the user's traffic. Let β_i denote the chosen target SINR for user i, and the transmission rate of the user is [26]

$$r_i = W \log_2(1 + k \cdot \beta_i), \tag{2}$$

where $M(\beta_i) = (1 + k \cdot \beta_i)$ refers to the bit number per modulation symbol; $k = -1.5/\ln(5 \cdot \mathrm{BER})$ is a constant decided by the requested target maximum bit error rate (BER); W denotes the channel bandwidth.

2.3. Quantity Model of MOS for Multiple Services. Traditionally, MOS is only used for voice service. However, data service lacks key parameter indicator to reflect QoE. Reference [27] proposed a method to evaluate QoE for data service based on the transmission rate and packet loss ratio. The model is derived from various real data service samples, which makes the quantity model of data service MOS reliable. For one user utilizing data service, the MOS could be

$$M_D = a \log_{10}(br_i(1 - p_{\mathrm{e2e}})), \tag{3}$$

where p_{e2e} represents the end-to-end packet error rate and a and b are two constants for specific scenarios.

Video service, as one of the most promoting data services, will be widely used in V2X communication. Reference [28–32] proposed the experimental MOS model for video service as

$$M_V = c\left(1 + e^{d(k\log_{10}(r_i)+p-h)}\right)^{-1}, \tag{4}$$

where c, d, k, p, and h are constants obtained by fitting real video data set.

3. DQN for Intelligent Band Coordination

This sector presents the DQN for intelligent band coordination in 5G heterogeneous network. Specifically, this solution is considered of multiple steps, each of which optimizes the performance of the former step. The deep Q-learning problem space is illustrated in Section 3.1. We set up the DQN for each user under the concern of different QoE for different traffic, and users may request different types of traffic. Therefore, a multi-agent DQN is set up in Section 3.2, as the implementation for each user's DQN. To further extend the ability of deep learning, an "experience replay" method is adopted in multi-agent DQN. This "experience replay" is explained in detail in Section 3.3. Under the concern of a large number of users may simultaneously connect to the same gNB in 5G, the calculation resource request for multi-agent DQN (i.e., each DQN for one user) may be a burden for gNB. Thus, optimization solution named "transfer learning" method is introduced in Section 3.4. Therefore, the solution and methods through Section 3.1 to Section 3.4 form the whole solution of our proposal.

3.1. Problem Space. Reinforcement learning is regarded as one of the effective solutions to resource allocation in many application areas [31, 32]. A reinforcement learning agent has near-optimal control action via interactions with the environment and receiving immediate reward. As one of the well-known reinforcement learning methods, DQN is regarded as the reprehensive work. In coordination with the system model in Section 2, we could set up the problem space including the state space and action space as follows.

3.1.1. State Space. Since the band selection is based on given target SINR, it is clear to use the target SINR list as the state set. In the state set, one user in one state means that the user's traffic requesting for the corresponding target SINR is met. Thus, let $S = \{s_1, s_2, \cdots, s_{N_{\mathrm{UE}}}\}$ be the state space; s_i represents a given target SINR β_i. The target SINR is discrete and limited; thus, the state set S is also a limited discrete set. To be noted that the user might not find any frequency band for its target SINR, then the state for this condition is marked as Θ. Then, the state space could be rewritten as

$$S = \{\Theta, \beta_1, \beta_2, \cdots, \beta_i\}. \tag{5}$$

3.1.2. Action Space. The action is of course the choice of specific frequency band from all the possible frequency bands. Then, the action set $A = \{a_1, a_2, \cdots, a_{N_{\mathrm{F}}}\}$ is defined based on the possible frequency bands, where a_i represents a frequency band and N_{F} is the number of frequency bands. Note that the action could be same at adjacent time t and $t + 1$, if

it is no need to change the frequency band. Therefore, the action space is reformulated as

$$A = \{f_1, f_2, \cdots, f_K\}, \quad K \le N_F, \tag{6}$$

wherein f_k is a frequency band a_j.

3.1.3. Reward Space. The reward generation processing is as follows. At time t, the agent takes an action $f(t) \in A$ while it is in state $\beta(t) \in S$. During this interaction with the environment, the agent achieves an immediate reward $R(\beta, f)$ and the system transitions into a new state $\beta(t+1) \in S$. According to (3) and (4), the MOS value is a function of user's transmission rate. It is reasonable to use MOS value as a positive reward, and the reward could be formulated as

$$R_t^i(\beta_t, f_t) = \begin{cases} \varepsilon, \beta_{t+1} < \beta_t, \text{or } \beta_{t+1} = \Theta, \\ M_D^i \text{or} M_V^i, \text{others}. \end{cases} \tag{7}$$

3.2. Multi-agent DQN. The architecture of the proposed solution is shown in Figure 3. The proposed method is performed per gNB.

3.2.1. Multi-agent DQN Structure. As indicated in Figure 3, the multi-agent DQNs with transfer learning receive renewed channel condition information and made frequency band selection. The upper part of Figure 3 depicts the structure for multi-DQN agents and the connection of DQN agents to the neural network for classification. The neural network is designed for transfer learning, which will be further described in Section 3.4.

Within a DQN agent as shown in the upper part in Figure 3, the DQN agent collects channel condition information and the given target SINR as input, and outputs the chosen frequency band information. The input data is first sent to the online Q-learning unit. The online Q-learning unit learns from the input data and optimizes its Q values. The input and the Q values are restored in a replay memory, and Timer 1 is triggered when data is updated in the replay memory. The data in the replay memory will be sent to the target Q-learning unit, when Timer 1 expires. The target Q-learning unit trains its Q values based on the data in the replay memory. That is to say, the target Q-learning unit learns from the trained result of the online Q-learning unit. When the target Q-learning unit starts to work, Timer 2 is triggered. The output Q values of the target Q-learning unit are forwarded to the online Q-learning unit when Timer 2 expires. The abovementioned process in the DQN agent is regarded as the "experience replay" method, and the corresponding detailed description is in Section 3.3. As described, there are only two parts in the DQN agent. Actually, the online Q-learning and the target Q-learning share the same Q-learning structure. The only difference is that they use different data as their input. The detailed introduction of the Q-learning is as follows.

3.2.2. Q-Learning. According to the problem space in Section 3.1, each user selects a discrete target SINR in terms of channel conditions. Then, MOS and throughput could

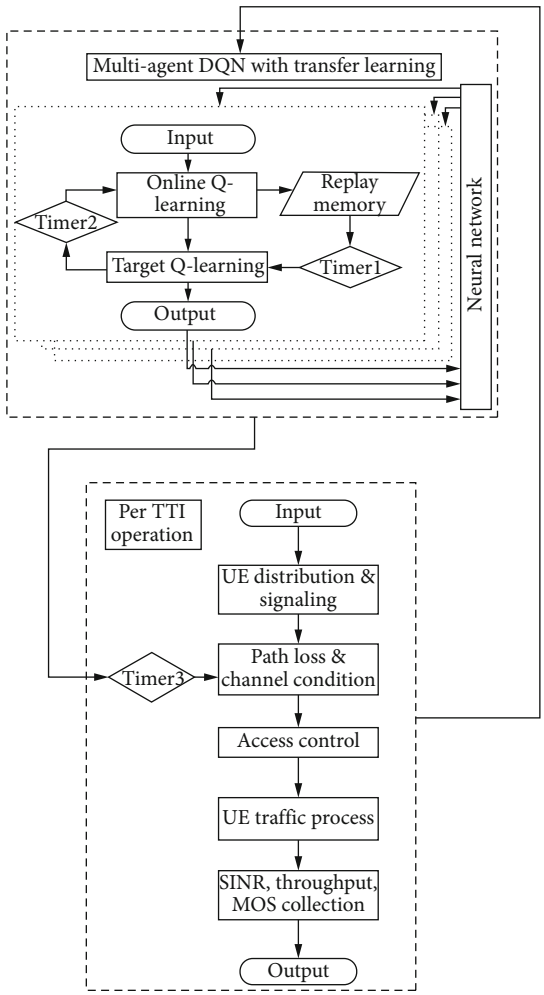

Figure 3: The architecture of proposed 5G intelligent multiple band coordination. (a) Multiagent DQNs with transfer learning. (b) gNB working flow.

be calculated based on selected SINR. The target SINRs could be satisfied, and the user could change to another frequency. Then, frequency is selected as an action, and a Q-learning module should be built. The user repeatedly makes its decision and finally obtains it optimal policy (i.e., how to select a proper frequency band) to maximize the expected sum of discounted reward

$$\begin{aligned} \left\{\left(\widehat{f}_t\right)\right\} &= \text{argmax} \frac{1}{N_{UE}} \sum_{i=1}^{N_{UE}} \left\{M_{DorV}^t \middle| f_t\right\} \\ \text{s.t.} \sum_{i=1}^{N_{UE}} \Psi_i\left(\beta_t^i \middle| f_t\right) &\le 1, i \in N_{UE} \\ r_{i,t} &= \max\left(r_{i,t}^{f_k}\right), k \in K, i \in N_{UE} \end{aligned} \tag{8}$$

where

$$\Psi_i = (1 + 1\beta_i)^{-1}. \tag{9}$$

The first boundary condition $\sum_{i=1}^{N_{\mathrm{UE}}} \Psi_i(\beta_t^i | f_t) \le 1, i \in N_{\mathrm{UE}}$ in (8) means the transmission power of user is no less than zero. The boundary condition is derived from (1) with the assumption of the target SINR β_i is satisfied. Based on this boundary condition and (9), the transmission power of the user i could be derived as

$$P_i = \frac{\varphi_i \sigma^2}{G_i\left(1 - \sum_{j=1}^{N_{\mathrm{UE}}} \varphi_j\right)}, \quad i = 1, 2, \cdots, N_{\mathrm{UE}}. \tag{10}$$

The obtained transmission power in (10) could be used in calculation of SINR by (1) for each user. The second boundary condition $r_{i,t} = \max\left(r_{i,t}^{f_k}\right), k \in K$, in (8) reflects the object of pursuing high throughput. The user will select the available frequency band which will provide the highest throughput at time t.

Because MOS is always positive, maximizing the overall MOS and solving (8) could be achieved by maximizing the individual MOS from each user. Thus, the user seeks to obtain an optimal policy through the DQN learning to maximize its own reward, which reflects its MOS and throughput. It could be seen from the objective function in (8) that different types of traffic will share the same quality metric. Both video and data traffic could use MOS as the uniform measurement scale. This property enables the seamless integration for different types of traffic. Maximizing the overall network MOS could be done by optimizing each user's action to maximize the user's cumulative future reward. The future reward is the optimal Q value function of the Q-learning module as

$$Q_i^*(\beta_i, f_i) = \max\left\{\sum \gamma^t \cdot E(R_i(\beta_i, f_i)) | \pi\right\}, \tag{11}$$

where β_i denotes discrete target SINR, f_i respects to difference frequency, π depicts the frequency selection strategy, $R_i(\beta_i, f_i)$ means the reward for user i in state β_i to perform action f_i, and γ^t stands for the coefficient at current time t.

3.3. Algorithm 1: DQN with "Experience Replay" Method. Based on Q-learning module, a multi-agent DQN is built for each user to perform frequency band coordination. This solution is marked as Algorithm 1, and the flowchart is shown in Figure 4.

As depicted in Figure 3, one DQN agent is composed of two Q-learning modules and a replay memory. Online Q-learning calculates Q values directly from instant inputs, and target Q-learning applies the results in replay memory. Timer 1, Timer 2, and Timer 3 control the time when well-learned strategies are applied in the frequency band coordination process. Equation (11) for Q-learning is value iteration algorithm that converges to the optimal active-value function in case of $\bar{t} \longrightarrow \infty$. A neural network, a fully connected feed-forward multilayer perception (MLP) network, has been proposed to efficiently approximate nonlinear action-value function.

The DQN utilizes MLP network as an action-value function approximation. "Experience replay" method is realized in DQN to improve the learning performance. At each time

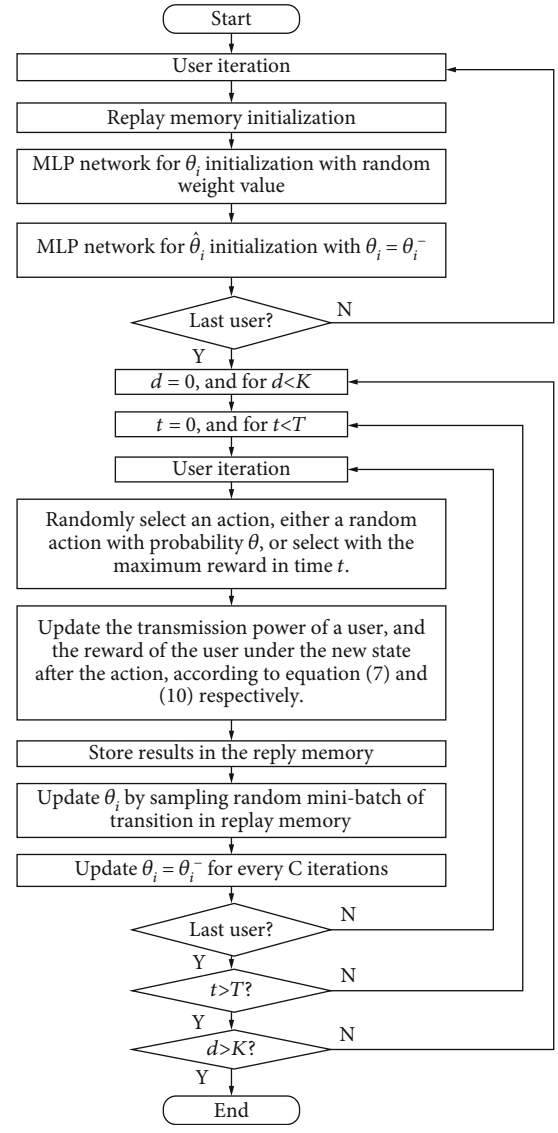

Figure 4: Flowchart of DQN with "experience replay" method (Algorithm 1).

step in experience replay, the environment is stored as in the replay memory. $e_i(t) = (\beta_i(t), f_i(t), R_i(t), s_i(t+1))$ stands for the data in the replay memory. Thus, user (agent) i will have its data $D_i(t) = \{e_i(1), e_i(2), \cdots e_i(t)\}$ as in the replay memory. Moreover, each agent utilizes two separate MLP networks. One is an action-value function approximation $Q_i(\beta_i, f_i, \theta_i)$ and the other is a target action-value approximation $\widehat{Q}_1(\beta_i, f_i, \theta_i^-)$, wherein θ_i and θ_i^- reflect the current and old parameters, respectively. For every step (Timer 2), θ_i is updated via minibatch of random samples of transitions $(\beta_i, f_i, \theta_i, \xi_i)$ from the replay memory D_i. θ_i^- of the target action-value function at every C iterations (Timer 1). θ_i is updated via a gradient descent algorithm based on the cost function of

$$L(\theta_i) = E\left\{R_i(\beta, f) + \gamma \cdot \max\left(\widehat{Q}_i(\beta_i, f_i, \theta_i^-)\right) - (Q_i(\beta_i, f_i, \theta_i))^2\right\}. \tag{12}$$

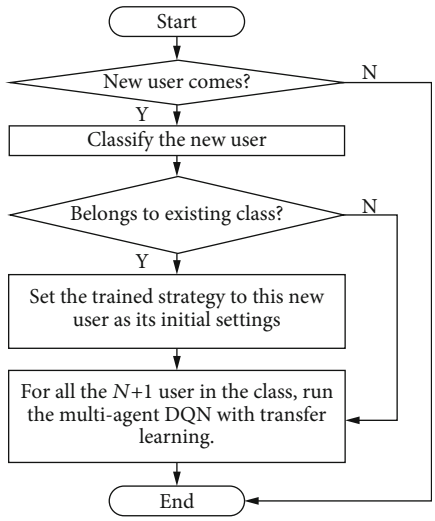

Figure 5: Flowchart of "transfer learning" method (Algorithm 2).

3.4. Algorithm 2: "Transfer Learning" Method. It will be a burden for a gNB to train every agent when the number of active users grows big, since each DQN agent is for one user. To alleviate this training computational complexity, a neural network for classification is built to collect each DQN agent's inputs, outputs, etc. This neural network is employed for the classification of users, and such whole process is remarked as Algorithm 2 as illustrated in Figure 5. Based on the above parameters, the users could be classified into different groups. Each group has a unique setting of these parameters. The learning parameter setting of users in the same group is always similar to each other. The average of each learning parameter in the same user group is used as the default learning parameter value.

As the frequency selection strategy trained by one user is done along with the environment information where this user is in, the trained strategy is adaptive for this environment (i.e., wireless conditions). Thus, if a new user comes into this environment, this new user may reuse the trained strategy as its initial settings. We do this by classifying the new user to one of the identified classes, via the neural network. The new user uses the same trained strategies in the classification as initial settings. Thus, the training complexity is degraded.

The 5G gNB will recognize the new user when the new user tries to access the network or handover from another gNB. The gNB also has the information of all users who are already connected with this gNB. Then, the gNB will estimate this new user with the likelihood with its connected users. Here, this estimation is taken by classification users into different groups. The users in the same group will share similar environment as well as similar parameters. Thus, the new user will use the settings of the group which the new user is estimated to belong to. The settings are just used as the new user's initial settings. The user could train and change the settings in the future. One thing to be noticed is that the classification of users is based on the following parameters: (1) the terminal capacity class, for example, type 0 stands for CPE (Customer Premise Equipment) and type 3 is usually regarded as widely used cell phones; (2) the traffic type, the concerned parameters include package delay and average data burst size; and (3) the layer number used in downlink and uplink transmission. More layers provide high transmission rate for one user. It is reasonable to refer to the users which have the same layer in data transmission.

3.5. Complexity Analysis of Algorithm 1 and Algorithm 2. As described in Sections 3.3 and 3.4, the complexity of the proposed algorithms is analyzed in this part. Let $|\cdot|$ be the cardinality of a set; the size of state space and action space is $|S|$ and $|A|$ according to (5) and (6), respectively. Then, the size of policy space and Q value space is $|S| \times |A|$, and the size of transition space and reward space is $|S|^2 \times |A|$. According to the model in Section 3.2, let $\mathscr{M}$ be the size of transition space, i.e., $\mathscr{M} = |S| \times |A|$, T be the attempt number in one Q-learning, $\mathbb{T} = \mathrm{O}(T, \mathscr{M})$ be time complexity, and then, the complexity for each agent for the proposed Algorithm 1 with the architecture in Figure 3 can be expressed as

$$T_{\text{peragent}} = \mathrm{O}(T_1 \times \mathscr{M}) + \mathrm{O}(T_2 \times \mathscr{M}) = O((T_1 + T_2) \times \mathscr{M}), \tag{13}$$

wherein T_1 and T_2 stand for the attempt number threshold for Timer 1 and Timer 2, respectively. Assuming the training time boundary for Timer 3 noted as T_3, the time complexity for multi-agent DQN in Figure 3 can be obtained as

$$T_{\text{ma-DQN}} = \mathrm{O}(T^* \times \mathscr{M}), \tag{14}$$

wherein $T^* = \min(T_1 + T_2, T_3)$. Please note that the obtained complexity is the time complexity of Algorithm 1.

According to the architecture depicted in Figure 3, the total execution complexity of Algorithm 1 in the gNB serving N_{UE} could be obtained by

$$T_{\text{gNB}}^{\text{Algo-1}} = \mathrm{O}(T^* \times \mathscr{M} \times N_{\text{UE}}). \tag{15}$$

Based on Algorithm 1, a transfer learning based on a simple neural network with 3 layers is introduced in Algorithm 2. Denote L_1, L_2, and L_3 as the node number of layer 1 (input layer), layer 2 (hidden layer), and layer 3 (output layer), wherein the node number for the input and the output layer is constant. Further, let C_{NN} represent the training time, and then, the time complexity of this nature network can be calculated as

$$T_{\text{NN}} = \mathrm{O}(C_{\text{NN}} \times L_2). \tag{16}$$

With this simple neural network, a new accessed user to the gNB could be classified into a same UE class and resource allocation is performed according to one of the well-trained DQN parameters. Thus, the training time at gNB side is reduced. According to the description in Algorithm 2, K_1, K_2, and K_3 stand for the number of UE type, the number of UE traffic type, and the number of transmission layer, respectively. C_{UE} is the number of UE class obtained by the neural network classification. Then, the time

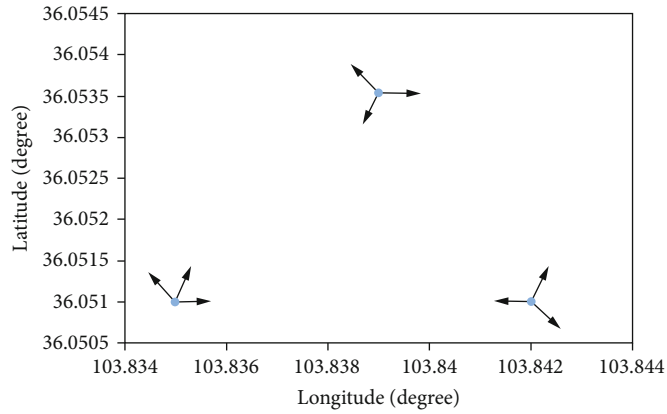

Figure 6: Cell layout of simulation scenario.

Table 1: Simulation setting.

Parameter	Value
$\{\beta_1, \beta_2, \cdots, \beta_{N_{UE}}\}$	$\{-10,-8,-6,-4,-2, 0, 2, 4, 6, 8, 10\}$ dB
$A = \{f_1, f_2, \cdots, f_K\}$	{2.6,3.5,28,38,45} GHz
P_{max}	23 dBm
α	0.01
γ	0.9
ε	0.8
L_{replay}	100
$L_{minipatch}$	10
K	500
N_{drop}	30

complexity of gNB with transfer learning could be finally obtained as

$$T_{gNB}^{Algo-2} = O(C_{NN} \times L_2) + C_{UE} \times O(T^* \times \mathscr{M}). \quad (17)$$

Based on (15) and (17), the complexity without transfer learning can be expressed as

$$\frac{T_{gNB}^{Algo-2}}{T_{gNB}^{Algo-1}} = \frac{O(T_{NN} \times L_2) + C_{UE} \times O(T^* \times \mathscr{M})}{O(T^* \times \mathscr{M} \times N_{UE})}. \quad (18)$$

Considering the fact that $T_3 \gg T_1$ and $T_3 \gg T_1$ in the real system and the classification has limited UE types as $C_{UE} \ll L_1 \times L_2 \times L_3$, then $O(T_{NN} \times L_2)]$ can be neglected and (18) can be approximated as

$$\frac{T_{gNB}^{Algo-2}}{T_{gNB}^{Algo-1}} \approx \frac{C_{UE} \times O(T^* \times \mathscr{M})}{O(T^* \times \mathscr{M} \times N_{UE})} = \frac{C_{UE}}{N_{UE}}. \quad (19)$$

It is clear that $C_{UE} \ll N_{UE}$, according to the definition of C_{UE} and N_{UE}. There, the complexity of Algorithm 2 with transfer learning is lower than that of Algorithm 1, especially under large accessed UE number.

4. Simulation Results

4.1. Scenario Selection. To better reflect the performance of the proposed algorithm in real network, we use the cell layout according to the current deployed networks. The chosen area is a central business district (CBD) area with many vehicles, as shown in Figure 6. There are totally 9 outdoor cells within the CBD area. The positions and the antenna directions of the cells remained in coordination with this real deployment. This setting will reflect the real world environment and scenario.

As depicted in the figure, there are 3 physical base station sites in the area, and the 9 outdoor cells are marked with arrows indicating the antenna directions. The displayed positions of the cells have been processed with a given shift; however, the relative distances among the cells are kept the same as the original data.

4.2. Benchmark Algorithm and Comparison. The benchmark algorithm is a standard Q-learning method. To compare the differences in considering "transfer learning," we need comparisons among three solutions. The first solution is marked as "DQN new user transfer learning" based on a combination of proposed Algorithm 1 and Algorithm 2. The gNB performs DQN with experience replay method for each user, as a multi-agent DQN solution. The parameters for each agent are trained by Algorithm 1. When a new user comes into the gNB, it is evaluated and classified into a certain group and uses Algorithm 2 to obtain the corresponding trained parameters as the new user's initial settings.

The second solution is marked as "DQN new user individual learning" based on Algorithm 1 only. The second solution utilizes Algorithm 1 as the first solution, but a new user will train its parameters with the initial value being zero. Thus, the comparison between the first and second solutions is mainly focused on the difference of transfer learning.

The third solution is regarded as "Q-learning new user individual learning," i.e., the benchmark. The third solution

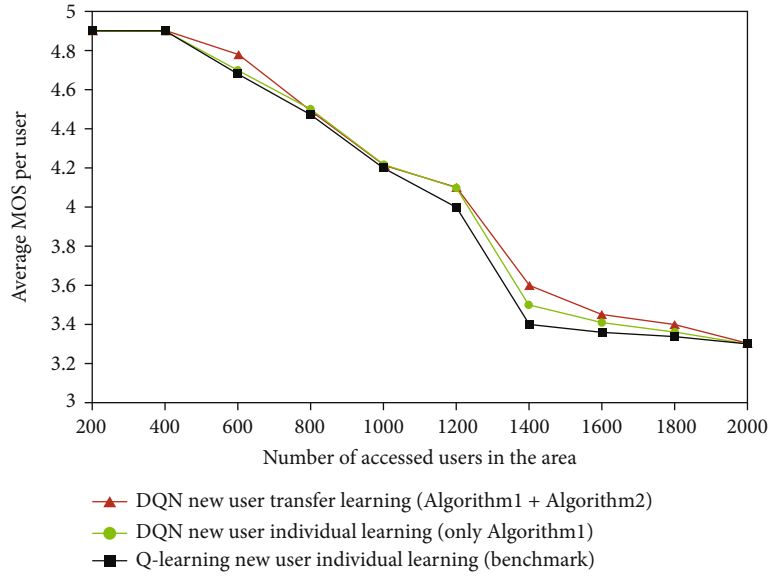

Figure 7: The average MOS along with the increase of users in the area.

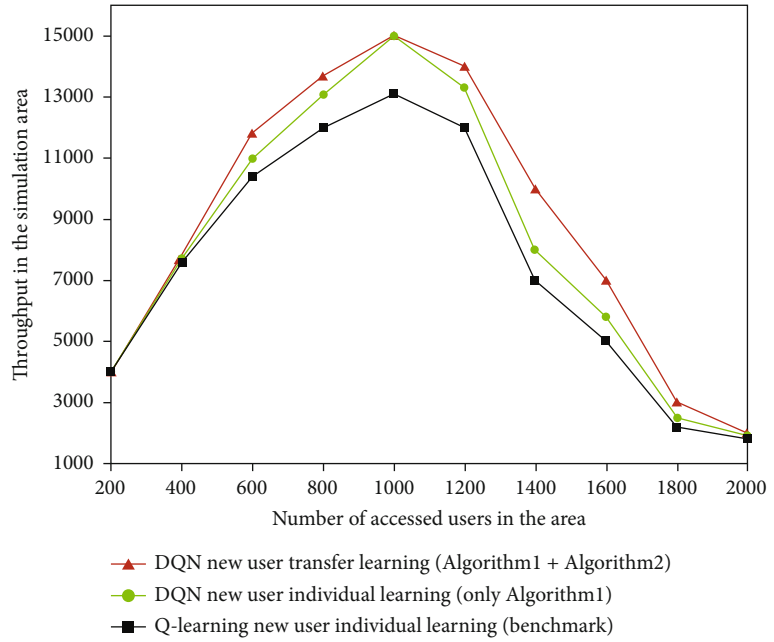

Figure 8: The change of network throughput along with the increase of user number.

uses Q-learning as the method to train the parameters, instead of Algorithm 1 used by the first solution and second solution. Thus, the comparison between the first solution and second solution is mainly for performance evaluation of Algorithm 1. The comparison between the third solution and the first solution is used for overall performance enhancement of the whole proposed solution in this paper.

The performance is evaluated based on the measuring, as a function of the accessed users with guaranteed QoE in the network, the change of average MOS of the accessed users, the change of throughput in the simulation area, the maximum number of accessed users, and the efficiency of learning algorithm.

4.3. Simulation Method and Parameter Setting. The detailed parameter settings are listed in Table 1. The performance of the presented intelligent band coordination in 5G heterogeneous network is studied via Monte Carlo simulation. The scenario is selected based on real network layout as Section 4.1 depicted. The users are randomly dropped within the simulation area and try to access to the network. The channel module refers to urban micro scenario in [33], because

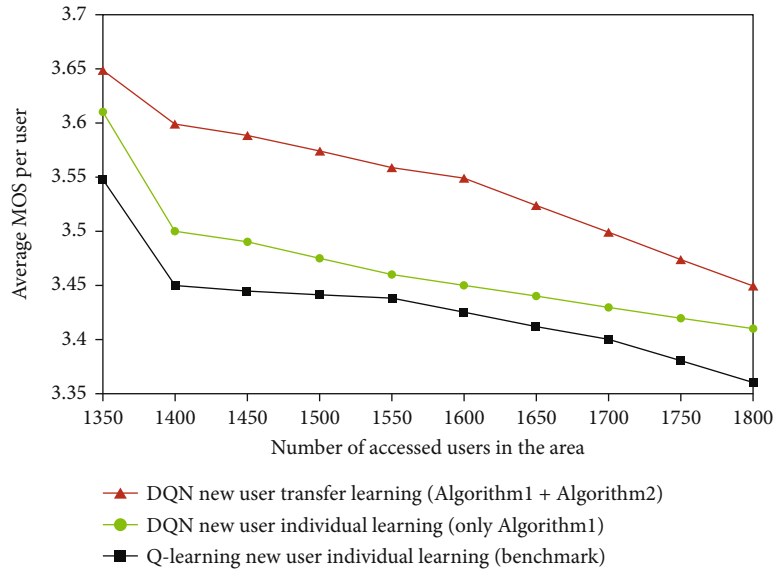

Figure 9: The maximum accessed user number for different solutions when their average MOS drop to around 3.5.

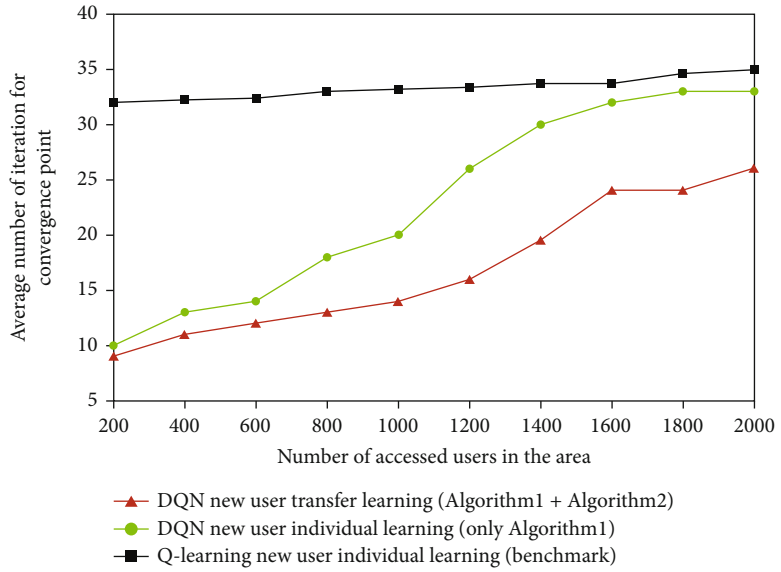

Figure 10: The learning efficiency among the three solutions for comparison.

this channel module could fit the frequency range from 0.5 GHz to 100 GHz. The ratio of data traffic to video traffic is 8 to 2. Additionally, for DQN, it has two Q-learning networks. Each Q-learning network is made of two hidden layers of 3 neurons and 2 neurons, respectively. The input layer is consisted of two nodes indicating the state and the action. The output layer contains one node.

4.4. Simulation Results and Analysis

4.4.1. Average MOS. Figure 7 illustrates the average MOS per user along with the increase of user number in the area. At the beginning, there are few users and plenty of resources; thus, the maximum MOS is easily obtained. With the increase of the users, the potential interference on each available frequency grows. Then, available SINR for each user drops beneath the chosen high target SINR, as interference grows along with increasing user number. As illustrated in Figure 7, the average MOS per user shows a sharp decrease when the number of users is greater than 1200. This is because the MOS of the user will degrade when the target SINR falls lower than the minimum requirement for the given MOS level. According to the figure, it can be found that the MOS drop rate under the first solution is smaller than the other two compared solutions when the user number is larger than 1200.

4.4.2. Throughput in the Simulation Area. The real network deployment is adopted as Section 4.1 describes. Because real network deployment is not evenly distributed, the statistics

on per cell base throughput will not fit this condition well. In addition, the environment for each outdoor cell is different, as well as the choice of frequency bands affected the capacity for different cells. Thus, we consider the whole simulation area as the evaluation scenario and statistic the overall achieved throughput in this area as "throughput in the simulation area." The statistics is drawn in Figure 8.

According to Figure 8, it shows a clear trend that the throughput increases dramatically with the number of accessed users, when user number is less than 600. The increasing trend slows down when user number grows high in the range from 600 to 1000 in the area. This is due to more users on the same frequency bring more interferences among the users on the frequency, and it leads to changing users' target SINR to be low target SINR. Thus, the total throughput in the simulation area increases, but the increasing rate drops down. As the number of users keeps growing greater than around 1200, as shown in Figures 8 and 7, the achievable SINR becomes too low to maintain a high MOS value. When the total accessed users are greater than 1200, the overall throughput starts to decrease. Because the obtained target SINR is corresponding to a certain throughput for each user, too many users result in severe interference and huge degradation of throughput.

Figure 8 shows that the throughput obtained by the first solution is always the highest. The effect of transfer learning is evident, since it saves much iteration times for new user in training and result in more opportunity to transmit with high SINR and QoE. This difference is easily proved by comparing the throughput of the first solution and the second solution in Figure 8. By comparison of the second solution with the third solution in Figure 8, the throughput is also greater when DQN is adapted. It is about 10% throughput enhancement among different user numbers when using the proposed solution compared with the benchmark.

4.4.3. Maximum Accessed Users. While the average MOS decreases with the increases of accessed user number, some accessed user will face the condition that the minimum MOS requested will not be satisfied, if one or more new users are accessed. It is also the boundary condition in (8). It means that the network has reached the maximum number of accessed users. It is a wide concern that average MOS being 3.5 corresponding to the QoE of the user being not satisfied.

Figure 9 shows the maximum accessed user numbers for different solutions when their average MOS drops to around 3.5. It can be found that the maximum accessed user number for the first, the second, and the third solutions are 1700, 1400, and 1370, respectively. That is to say, the network could obtain the most users when utilizing the proposed method. When applying transfer learning on the network that equips DQN, the capacity will increase about 20%. By utilizing both transfer learning and DQN, the capacity will increase about 24% compared to the network that equips Q-learning only.

4.4.4. Learning Efficiency. Figure 10 illustrates the learning efficiency among the three solutions for comparison. The first solution utilizes not only DQN but also transfer learning that could share the common environment reflection already trained parameters to the new coming user. This is effective in reducing the iterations requested for new user (around 25% in comparison with the second solution). The second solution also utilizes DQN in learning, but a new user has to train its own parameters from zero. According to Figure 10, the first solution always outperforms the third solution in the number of iterations requested for convergence by about 70% when the user number is smaller than 1400.

5. Conclusion

In this paper, we proposed an intelligent band coordination solution in 5G-based V2X heterogeneous network that equips DQN as its decision-making core in 5G gNB. Both real-time and regular V2X data traffic are considered in our proposal. By introducing the MOS as QoE metric for all types of traffic, the seamless integration of similar and dissimilar traffic is realized via a single common measuring scale. Furthermore, transfer learning is applied in our proposal to enhance the learning process of the DQN-based solution (i.e., Algorithm 1), reducing the complexity of deployment in 5G base station. To further alleviate the training burden, the trained models of other users by Algorithm 2 are used to enhance the stability and efficiency of the system, wherein the V2X traffic characteristics of different users within a cell are quite similar.

The provided simulation results show that our proposed solution "DQN new user transfer learning" based on the combination of Algorithm 1 and Algorithm 2 outperforms the compared solutions ("DQN new user individual learning" based on Algorithm 1 only and benchmark as "Q-learning new user individual learning") in average MOS, the throughput in the area, number of accessed users, and learning efficiency. Simulation results indicate that the proposed solution has smaller MOS drop rate with the increasing of user number, since interference can be alleviated after dynamic spectrum sharing. Further, the proposed solution decreases the iteration number by approximately 70% compared to the benchmark when the accessed number is smaller than 1000. Benefiting from less iterations and better reward searching process in the proposed solution, an average 10% throughput enhancement and an about 24% accessed user number increasing are also obtained, compared with the benchmark solution.

Data Availability

Data are available upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This paper was supported by the National Key Research and Development Program of China (2020YFF0305401).

References

[1] 3GPP, "Technical specification group radio access network: base station (BS) radio transmission and reception, 3GPP," 2018, TS 38.104 V15.2.0.

[2] ITU, *Recommendation ITU-R M.2083-0:IMT Vision ¨C Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond*, Recommendation ITU, 2015.

[3] C. Seker, M. T. Güneser, and T. Ozturk, "A review of millimeter wave communication for 5G," in *2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, Ankara, Turkey, 2018.

[4] T. Rappaport, S. Sun, R. Mayzus et al., "Millimeter wave mobile communications for 5G cellular: it will work," *IEEE Access*, vol. 1, pp. 335–349, 2013.

[5] K. Zheng, Q. Zheng, H. Yang, L. Zhao, L. Hou, and P. Chatzimisios, "Reliable and efficient autonomous driving: the need for heterogeneous vehicular networks," *IEEE Communications Magazine*, vol. 53, no. 12, pp. 72–79, 2015.

[6] M. Sepulcre, J. Gozalvez, O. Altintas, and H. Kremo, "Exploiting context information for estimating the performance of vehicular communications," in *2013 IEEE Vehicular Networking Conference*, Boston, MA, USA, 2013.

[7] J. M. Anderson, K. Nidhi, K. D. Stanley, P. Sorensen, C. Samaras, and O. A. Oluwatola, *Autonomous vehicle technology: a guide for policymakers*, Rand Corporation, 2014.

[8] J. Bierstedt, A. Gooze, C. Gray, J. Peterman, L. Raykin, and J. Walters, *Effects of Next Generation Vehicles on Travel Demand and Highway Capacity*, FP Think Working Group, 2014.

[9] ISO/TC 204, *Intelligent Transport Systems- Communications Access for Land Mobiles (CALM)-Architecutre*, vol. 2014, ISO 21217, 2014.

[10] ETSI TC ITS, "Intelligent transport systems (ITS); communications RArchitecutre," 2010, ETSI EN 302 665, v1.1.1.

[11] A. Torres, Y. Ji, C. T. Calafate, J. C. Cano, and P. Manzoni, "V2X solutions for real-time video collection," in *2014 11th annual conference on wireless on-demand network systems and services (WONS)*, Obergurgl, Austria, 2014.

[12] J. A. Sanguesa, M. Fogue, P. Garrido et al., "An infrastructureless approach to estimate vehicular density in urban environments," *Sensors*, vol. 13, no. 2, pp. 2399–2418, 2013.

[13] R. Aissaoui, H. Menouar, A. Dhraief, F. Filali, A. Belghith, and A. Abu-Dayya, "Advanced real-time traffic monitoring system based on V2X communications," in *2014 IEEE international conference on communications (ICC)*, Sydney, NSW, Australia, 2014.

[14] M. Makni, M. Baklouti, S. Niar, M. Biglari-Abhari, and M. Abid, "Heterogeneous multi-core architecture for a 4G communication in high-speed railway," in *2015 10th International Design and Test Symposium (IDT)*, Amman, Jordan, 2015.

[15] S. K. Padaganur and J. D. Mallapur, "A neural network based resource allocation scheme for 4G LTE heterogeneous network," in *2017 2nd International Conference for Convergence in Technology (I2CT)*, Mumbai, India, 2017.

[16] Y. Wang, "Research on GSM-H1 upgrade technology based on GSM refarming frequency re-cultivation," *Information and Communications*, vol. 7, pp. 1–5, 2017.

[17] Y. Wang, B. Xia, Y. Chen, and X. Zhu, "Analysis and verification of GSM band refarming to LTE FDD strategy," *Telecom Engineering Technique and Standardization*, vol. 7, pp. 23–28, 2017.

[18] X. Han, H. Chen, L. Xie, and K. Wang, "A resource allocation scheme for the heterogeneous OFDMA system with ad hoc relay," in *2011 IEEE 13th International Conference on Communication Technology*, pp. 637–641, Jinan, China, Sep. 2011.

[19] D. Fooladivanda and C. Rosenberg, "Joint resource allocation and user association for heterogeneous wireless cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 1, pp. 248–257, 2013.

[20] R. Liu, G. Yu, J. Yuan, and Y. Li, "Resource management for millimeter-wave ultra-reliable and low-latency communications," *IEEE Transactions on Communications*, vol. 69, no. 2, pp. 1094–1108, 2021.

[21] S. R. Pandey, M. Alsenwi, Y. K. Tun, and C. S. Hong, "A downlink resource scheduling strategy for URLLC traffic," in *2019 IEEE International Conference on Big Data and Smart Computing (BigComp)*, Kyoto, Japan, Feb. 2019.

[22] S. Xing, X. Xu, Y. Chen, Y. Wang, and L. Zhang, "Advanced grant-free transmission for small packets URLLC services," in *2019 IEEE International Conference on Communications Workshops (ICC Workshops)*, Shanghai, China, May 2019.

[23] J. Wang, J. Weitzen, O. Bayat, V. Sevindik, and M. Li, "Interference coordination for millimeter wave communications in 5G networks for performance optimization," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, Article ID 46, 2019.

[24] C. Yang, J. Li, M. Guizani, A. Anpalagan, and M. Elkashlan, "Advanced spectrum sharing in 5G cognitive heterogeneous networks," *IEEE Transactions on Wireless Communications*, vol. 23, no. 2, pp. 94–101, 2016.

[25] R. Zhang, "On active learning and supervised transmission of spectrum sharing based cognitive radios by exploiting hidden primary radio feedback," *IEEE Transactions on Communications*, vol. 58, no. 10, pp. 2960–2970, 2010.

[26] X. Qiu and K. Chawia, "On the performance of adaptive modulation in cellular systems," *IEEE Transactions on Communications*, vol. 47, no. 6, pp. 884–895, 1999.

[27] O. Dobrijevic, A. J. Kassler, L. Skorin-Kapov, and M. Matijasevic, "Q-point: Qoedriven path optimization model for multimedia services," in *International Conference on Wired and Wireless Internet Communications, Springer*, pp. 134–147, 2014.

[28] P. Hanhart and T. Ebrahimi, "Calculation of average coding efficiency based on subjective quality scores," *Journal of Visual Communication and Image Representation*, vol. 25, no. 3, pp. 555–564, 2014.

[29] Y. Chen, K. Wu, and Q. Zhang, "From qos to qoe: a tutorial on video quality assessment," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 1126–1165, 2015.

[30] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on image processing*, vol. 15, no. 11, pp. 3440–3452, 2006.

[31] J. Zhu, Y. Song, D. Jiang, and H. Song, "A new deep-q-learning-based transmission scheduling mechanism for the cognitive internet of things," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2375–2385, 2018.

[32] V. Mnih, K. Kavukcuoglu, D. Silver et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[33] 3GPP, "Technical specification group radio access network. Study on channel model for frequencies from 0.5 to 100 GHz, 3GPP," 2018, TS 38.901 V 15.0.0.