International Journal of Research (IJR)

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Vol. 8 Issue 12
December 2021

Deep Reinforcement Learning-Based Modulation and Coding Scheme Selection in Cognitive Heterogeneous Networks

Sharab Shravani[1], N.Vinod Kumar[2], Dr.S.A Siva Kumar[3]
[1]P.G. Scholar, [2]Assistant Professor, [3]Head of the Department
[1,2,3]M.Tech.,(DECS)
[1,2,3]Dr.K V Subbareddy College Of Engineering For Women
Email: [1]sravanisharabu@gmail.com,[2]vinod.nayakallu@gmail.com

ABSTRACT

We consider a cognitive heterogeneous network (HetNet), in which multiple pairs of secondary users adopt sensing-based approaches to coexist with a pair of primary users on a certain spectrum band. Due to imperfect spectrum sensing, secondary transmitters (STs) may cause interference to the primary receiver (PR) and make it difficult for the PR to select a proper modulation and/or coding scheme (MCS). To deal with this issue, we exploit deep reinforcement learning (DRL) and propose an intelligent MCS selection algorithm for the primary transmission. To reduce the system overhead caused by the MCS switching's, we further introduce a switching cost factor in the proposed algorithm.

The simulation results show that the primary transmission rate of the proposed algorithm without the switching cost factor is 90% ~ 100% of the optimal MCS selection scheme, which assumes that the interference from the STs is perfectly known at the PR as prior information, is 30% higher than that of the upper confidence bandit (UCB) algorithm, and is 100% higher than that of the signal-to-noise ratio (SNR)-based algorithm. Mean while, the proposed algorithm with the switching cost factor can achieve a higher primary transmission rate than those of the benchmark algorithms without increasing system overheads.. Here we added an extension to current model like considering a number of frames in Energy Efficiency (EE) system where the users have different subcarrier spacing (SCS). Unlike in single numerology EE systems, mixed frames EE systems suffer from inter-numerology interference (INI). We first derive the interference pattern and find that the variance of interference energy increases due to the difference in SCS. This increase in variance negatively affects decoding performance, since the interference energy is unbalanced between subcarriers

Keywords:-Cognitive HetNet, intelligent DRL, MCS selection, spectrum sharing, switching cost.

INTRODUCTION

FUELED by the exponential growth of smart phones and tablets, recent years have witnessed an explosive increase of data traffics in wireless networks [1]–[3]. It is envisioned that wireless data traffics will continue increasing in the next few years. To accommodate these traffics, it is urgent to improve the network capacity. Two typical approaches to improve the network capacity include enhancing the wire- less link efficiency and optimizing

the network architecture. Nevertheless, the wireless link efficiency is approaching the fundamental limit with the development of the multiple-input- multiple-output and orthogonal frequency division multiplexing technologies. As such, a heterogeneous network(HetNet) is emerging as a promising network architecture to improve the network capacity [4]–[10].

Different from a conventional cellular network, the HetNet typically consists of a macro base station (BS), multiple small cell BSs, and numbers of users [6]. The macro BS is deployed to provide a wide coverage for users with low data- rate requirements and the small cell BSs are to extend the coverage of the macro BS as well as to support high data-rates for the users in a relatively small area. In the deployment of the HetNet, one major challenge is the coexistence among multiple wireless links of different users. On the one hand, if dedicated spectrum bands are assigned to different wireless links to avoid interference, large amount of spectrum resource is required to satisfy massive transmission demands in the network. On the other hand, if all the wireless links share the same spectrum band, different wireless links may cause severe interference to each other. To deal with this issue, the cognitive radio technology has been introduced to the HetNet, namely, cognitive HetNet[11]–[15].In particular, the cognitive HetNet consists of two types of users, i.e., primary users with high priorities to the spectrum bands and secondary users with low priorities to the spectrum bands.

To protect primary transmissions, the secondary transmitter(ST) usually adopts a sensing-based approach to determine whether to access a target spectrum band or not. In particular, the ST first measures the energy of the signal on the target spectrum band. If the measured energy exceeds a certain threshold, the target spectrum band is declared to be occupied by primary users and the ST keeps silent. Otherwise, the target spectrum band is idle and the ST can access it directly. However, the complicated environment in a cognitive HetNet may lead to imperfect spectrum sensing.

In fact, the imperfect spectrum sensing issue commonly exists in a cognitive network. To reduce the impact of imperfect spectrum sensing on the primary transmission performance, the authors of [16] suggested guaranteeing a high detection probability, e.g., 90%, for a relatively low strength of the received primary signal at the ST, e.g., the signal-to-noise ratio (SNR) of the received primary signal at the ST is as low as 15 dB. This method can is able to reduce effectively the impact of imperfect spectrum sensing on the primary transmission performance and thus is widely adopted in cognitive networks [17].

A. Motivations :

It is clear that the effectiveness of the method in [16] diminishes for the scenario in which the PT is transparent to the ST, i.e., the strength of the received primary signals at the ST is extremely low (much lower than 15 dB), and the channel from the ST to the PR is non-ignorable. In this scenario, the ST may cause severe interference to the PR after imperfect spectrum sensing and degrade the primary transmission performance. This scenario is particularly relevant to the uplink transmission of a cognitive HetNet, in which a user(PT) transmits uplink data to the BS (PR) on a certain spectrum band, and multiple pairs of secondary users adopt a sensing- based approach to coexist with the primary users on the same spectrum band. Since the antenna height of the user terminal is relatively low while that of the macro BS is high, the wireless links between the PT and STs may be heavily blocked

by buildings while the line of sight propagations exist between the PR and the STs. To avoid severe interference from the STs to the primary transmission, existing literature suggested each ST adopt a conservative transmit power. However, the question still lies in whether the primary transmission performance and the network capacity can be further enhanced.

We notice that the starting time of the secondary transmission is later than that of the primary transmission according to the sensing-based protocol. As such, the interference from STs is unknown at the PT at the starting time of the primary transmission, and the PT cannot adapt its transmission with the interference information. In fact, the interference from STs typically follows a certain pattern, and it is possible for the PT to learn the interference pattern by analyzing the historical interference information and in further interference in the future frames. In this paper, we adopt the deep reinforcement learning (DRL) for the PR to learn the interference pattern from STs and infer the interference in the future frames [18]. With the inferred interference, the PT can adapt its transmission to enhance the primary transmission rate as well as the network capacity. In the following, we first provide related work on the applications of both RL and DRL in wireless communications and then elaborate the contributions of the paper.

B. Related Work :

Recently, RL is widely applied in wireless communication networks, especially in decision-making scenarios [19]–[27]. Specifically, [19] proposed two RL-based user handoff algorithms in a Millimeter wave HetNet. [20] Developed an efficient RL-based radio access technology selection algorithm in a HetNet. [21] studied the energy-efficiency in a HetNet, and proposed a RL-based user scheduling and resource allocation algorithm.[22]and[23]investigated the spectrum sharing problem in cognitive radio networks and developed RL-based spectrum access algorithms for cognitive users.

[24] and [25] focused on the self-organization network and adopted RL to deal with the request coordination problem and the user scheduling problem, respectively. In addition,

[26] applied RL in the physical layer security and proposed an RL-based spoofing detection scheme. [27] formulated the wireless caching as an optimal decision-making problem and developed an RL-based caching scheme to reduce the energy consumption.

It is proved that RL works well in decision-making scenarios when the size of the state-action space in the wireless system is relatively small. However, the effectiveness of RL diminishes as the size of the state-action space becomes large. Then, DRL emerges as a good alternative to solve the decision- making problem in wireless systems with a large size of state-action space [28]–[34]. In particular, [28] developed a DRL-based user scheduling algorithm to enhance the sum-rate in a wireless caching network. [29] proposed a DRL-based channel selection algorithm to improve the transmission per- formance in a multi-channel wireless network. [30] adopted DRL to learn the jamming pattern in a dynamic and intelligent jamming environment and proposed an efficient algorithm to obtain the optimal anti-jamming strategy. [31] adopted DRL to learn the power adaption strategy of the primary user in a cognitive network, such that the secondary user is able to adaptively control its power and satisfy the required quality of services of both primary and secondary users. [32] studied the handover problem in a multi-user multi-BS wireless network and proposed a DRL-based handover algorithm to reduce the handover rate of each user under a

minimum sum- throughput constraint. In addition, [33] proposed a distributed DRL-based multiple access algorithm to improve the uplink sum-throughput in a multi-user wireless network. [34] applied DRL to the power allocation problem in an interference channel and proposed a DRL-based algorithm to enhance the sum-rate.

C. Contributions of the Project

In this paper, we consider a cognitive HetNet, in which a user (PT) transmits uplink data to the BS (PR) on a certain spectrum band and multiple STs adopt a sensing-based approach to access the same spectrum band. In particular, the PT is transparent to the STs, and the channel from each ST to the PR is non-ignorable. As a result, each ST may access the spectrum band with imperfect spectrum sensing and cause interference to the PR. Since the interference is unknown at the PT due to time causality, it is difficult for the PR to select a proper MCS to improve the primary transmission performance with conventional optimization techniques (e.g., convex optimization). It is worth noting that MCS has a significant impact on the performance (e.g., throughput) of a wireless system, and the MCS selection problems in a variety of wireless systems have been studied based on conventional optimization techniques (e.g., [35] and [36]). Thus, it is important and practical to investigate the considered MCS selection problem which cannot be efficiently addressed with conventional optimization techniques. In this paper, we adopt the DRL technique and develop a DRL-based MCS selection algorithm for the considered MCS selection problem. The effectiveness and the advantages of the proposed algorithm are demonstrated through simulation results. Note that MCS refers to modulation and coding scheme in a coded system, and is reduced to modulation scheme in an un coded system. For consistency, we use MCS to represent modulation and coding scheme in a coded system, and represent modulation scheme in an un coded system. We summarize the major contributions of the paper as follows:

We propose an intelligent DRL-based MCS selection algorithm for the PR. Specifically, we enable the DRL agent at the PR to learn the interference pattern from the STs. With the learnt interference pattern, the PR can infer the interference from the STs in the future frames and adaptively select a proper MCS to enhance the primary transmission rate.

Considering that the DRL agent may switch the MCS frequently among different MCS's to maximize the transmission rate. On the one hand, since each MCS switching requires the negotiation between the PT and the BS and system reconfiguration, frequent MCS switching may increase significantly both signalling overheads and system reconfiguration costs [40]. On the other hand, since the information exchange of each MCS switching needs both spectrum resource and energy consumption, frequent MCS switching may degrade considerably both spectra land energy efficiencies. To deal with this issue, we take the system overhead caused by MCS switchings in to consideration and introduce a switching cost factor in the proposed algorithm. By adjusting the value of the switching cost factor, we can achieve the desired balance between the primary transmission rate and system overheads.

Simulation results show that the transmission rate of proposed algorithm without the switching cost factor is 90% ~ 100% to that of the optimal MCS selection scheme, is 30% higher than that of the upper confidence bandit (UCB) algorithm, and is 100% higher than that of the signal-to-noise ratio (SNR) based algorithm. Meanwhile, the proposed algorithm with the switching cost factor can achieve a higher primary transmission rate than those of the benchmark algorithms without increasing system

overheads.

EXISTING SYSTEM :

We consider a cognitive heterogeneous network (HetNet), in which multiple pairs of secondary users adopt sensing-based approaches to coexist with a pair of primary users on a certain spectrum band. Due to imperfect spectrum sensing, secondary transmitters (STs) may cause interference to the primary receiver (PR) and make it difficult for the PR to select a proper modulation and/or coding scheme (MCS). To deal with this issue, we exploit deep reinforcement learning (DRL) and propose an intelligent MCS selection algorithm for the primary transmission. To reduce the system overhead caused by the MCS switchings, we further introduce a switching cost factor in the proposed algorithm. The simulation results show that the primary transmission rate of the proposed algorithm without the switching cost factor is 90% ~ 100% of the optimal MCS selection scheme, which assumes that the interference from the STs is perfectly known at the PR as prior information, is 30% higher than that of the upper confidence bandit (UCB) algorithm, and is 100% higher than that of the signal-to-noise ratio (SNR)-based algorithm. Meanwhile, the proposed algorithm with the switching cost factor can achieve a higher primary transmission rate than those of the benchmark algorithms without increasing system overheads. Index Terms—Cognitive HetNet, intelligent DRL, MCS selection, spectrum sharing, switching cost.

PROPOSED SYSTEM :

We consider a number of frames in Energy Efficiency (EE) system where the users have different subcarrier spacing (SCS). Unlike in single numerology EE systems, mixed frames EE systems suffer from inter-numerology interference (INI). We first derive the interference pattern and find that the variance of interference energy increases due to the difference in SCS. This increase in variance negatively affects decoding performance, since the interference energy is unbalanced between subcarriers.

SYSTEM MODEL

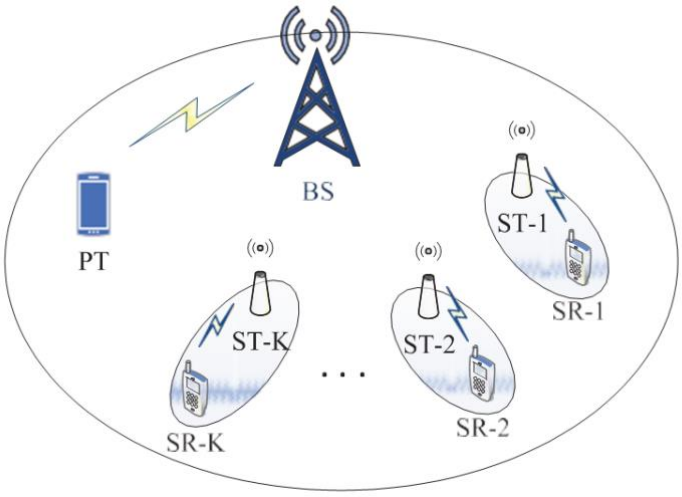

Figure 1: Considered cognitive HetNet, which consists of a PU, a macro BS, and K pairs of secondary users.

LITERATURE SURVEY

Applications of Deep Reinforcement Learning in Communications and Networking: A Survey. This paper presents a comprehensive literature review on applications of deep reinforcement learning in communications and networking. Modern networks, e.g., Internet of Things and Unmanned Aerial Vehicle (UAV) networks, become more decentralized and autonomous. In such networks, network entities need to make decisions locally to maximize the network performance under uncertainty of network environment. Reinforcement learning has been efficiently used to enable the network entities to obtain the optimal policy including, e.g., decisions or actions, given their states when the state and action spaces are small. However, in complex and large-scale networks, the state and action spaces are usually large, and the reinforcement learning may not be able to find the optimal policy in reasonable time. Therefore, deep reinforcement learning, a combination of reinforcement learning with deep learning, has been developed to overcome the shortcomings. In this survey, we first give a tutorial of deep reinforcement learning from fundamental concepts to advanced models. Then, we review deep reinforcement learning approaches proposed to address emerging issues in communications and networking. The issues include dynamic network access, data rate control, wireless caching, data offloading, network security, and connectivity preservation which are all important to next generation networks such as 5G and beyond. Furthermore, we present applications of deep reinforcement learning for traffic routing, resource sharing, and data collection. Finally, we highlight important challenges, open issues, and future research directions of applying deep reinforcement learning

Spectrum Access In Cognitive Radio Using a Two-Stage Reinforcement Learning Approach

With the advent of the 5th generation of wireless standards and an increasing demand for higher throughput, methods to improve the spectral efficiency of wireless systems have become very important. In the context of cognitive radio, a substantial increase in throughput is possible if the secondary user can make smart decisions regarding which channel to sense and when or how often to sense. Here, we propose an algorithm to not only select a channel for data transmission but also to predict how long the channel will remain unoccupied so that the time spent on channel sensing can be minimized. Our algorithm learns in two stages - a reinforcement learning approach for channel selection and a Bayesian approach to determine the optimal duration for which sensing can be skipped. Comparisons with other learning methods are provided through extensive simulations. We show that the number of sensing is minimized with negligible increase in primary interference; this implies that lesser energy is spent by the secondary user in sensing and also higher throughput is achieved by saving on sensing.

Spectrum Sharing for Internet of Things: A Survey

The Internet of Things is a promising paradigm to accommodate massive device connections in 5G and beyond. To pave the way for future IoT, the spectrum should be planed in advance. Spectrum sharing is a preferable solution for IoT due to the scarcity of available spectrum resource. In particular, mobile operators are inclined to exploit the existing standards and infrastructures of current cellular networks and deploy IoT within licensed cellular spectrum. Yet, proprietary companies prefer to deploy IoT within unlicensed spectrum to avoid any

licence fee. In this paper, we provide a survey on prevalent IoT technologies deployed within licensed cellular spectrum and unlicensed spectrum. Notably, emphasis will be on the spectrum sharing solutions including the shared spectrum, interference model, and interference management. To this end, we discuss both advantages and disadvantages of different IoT technologies. Finally, we identify challenges for future IoT and suggest potential research directions.

SIMULATION RESULTS

In this section, we provide simulation results to evaluate the performance of the proposed intelligent DRL-based MCS selection algorithm. For comparison, we consider the optimal MCS selection algorithm, which assumes that the BS knows the average SINR $\bar{\gamma}$ at the beginning of each frame and solves (7) to obtain the optimal solution. As aforementioned, it is impractical for the BS to know the average SINR $\bar{\gamma}$ at the beginning of each frame due to time causality. Thus, the performance of the optimal MCS selection algorithm is the theoretical upper bound. Meanwhile, we provide two benchmark algorithms, namely, SNR-based algorithm and upper confidence bandit (UCB) learning algorithm *[26] [35]*. In particular, SNR-based algorithm replaces the average SINR $\bar{\gamma}$ in (7) with the measured SNR $\gamma 0$ and solves (7) to obtain a solution. The selected MCS with the

UCB learning algorithm at the beginning of frame t is as follows:

$$m*=\arg\max_{m\in\{1,2,\cdots,M\}}(\mu m+\sqrt{2\ln t\Gamma m(t-1)}), \quad (22)$$

where $\Gamma m(t-1)$ is the number of times that MCSm has been selected in the previous $t-1$ frames, μm is randomly initialized and updated by

Table I: Considered MCSs and the corresponding SERs.

MCS	SER [36]
BPSK	$f1(\bar{\gamma})=Q(\sqrt{2\bar{\gamma}})$
QPSK	$f2(\bar{\gamma})=2(1-1\sqrt{4})Q(\sqrt{3\log2(4)\bar{\gamma}4-1})$
16QAM	$f3(\bar{\gamma})=2(1-1\sqrt{16})Q(\sqrt{3\log2(16)\bar{\gamma}16-1})$
64QAM	$f4(\bar{\gamma})=2(1-1\sqrt{64})Q(\sqrt{3\log2(64)\bar{\gamma}64-1})$

V-A*ASSUMPTIONS AND SETTINGS IN THE SIMULATION*

It is clear that secondary transmissions do not have any impact on the primary transmission when the primary transmission is inactive (i.e., the PU does not transmit data to the BS). When the primary transmission is active (i.e., the PU is transmitting data to the BS), the secondary transmissions may interfere with the PU transmission due to imperfect spectrum sensing. To investigate the effectiveness of the proposed algorithm in the presence of imperfect spectrum sensing, we assume that the primary transmission is always active. Then, the miss-detection/interference probability of ST-k is also the active probability of ST-k.

In the simulation, we consider an uncoded system and assume that the PU supports four MCS levels as shown in Table I, although the proposed algorithm can be easily applied to coded systems. The DQN is composed of an input layer with 4Φ+1 ports, which correspond to 4Φ+1 elements in s(t)
, two fully connected hidden layers, and an output layer with four ports, which correspond to four MCS levels in Table I. In particular, each hidden layer has 100 neurons with the Relu activation function. We apply an adaptive
ϵ-greedy algorithm, in which ϵ follows ϵ(t+1)=max{ϵmin,(1−λϵ)ϵ(t)} *[28]*. An intuitive explanation of adopting a varying ϵ is as follows. At the beginning frames of the proposed algorithm, the number of the experienced state-action pairs is small and the DRL agent needs to explore more actions to improve the long-term reward. As the number of the experienced state-action pairs increases, the DRL agent does not need to perform so many explorations. We set ϵ(0)=0.3, ϵmin=0.005, and λϵ=0.0001. Besides, the batch size of experience samples in the proposed algorithm is Z=32, and the local memory at the DRL agent is NE=500. Furthermore, we set γ=0.5
, and the RMSProp optimization algorithm with a learning rate
0.01 is used to update θ *[37]*. In addition, we set τ−τpT−τp=0.1 and T−τT−τp=0.9 in (4), L=100, and each frame contains N=1000 symbols.

V - *PERFORMANCE COMPARISON IN QUASI-STATIC AND DYNAMIC INTERFERENCE SCENARIOS*

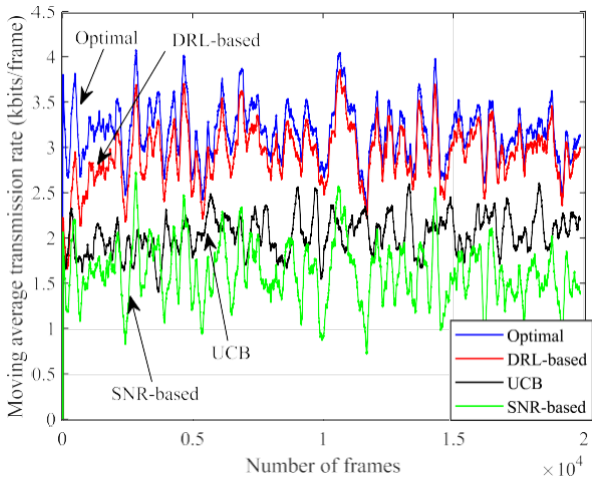

Fig. 4 .Transmission rate comparison in a quasi-static interference scenario. Each value is a moving average of the previous 200 frames and each curve is the average of20 trials.

Fig. 4 compares the transmission rates of different algorithms in a quasi-static interference scenario. We consider two pairs of secondary users, i.e., namely, (ST-1, SR-1) and (ST-2, SR-2), although the proposed algorithm can handle more than two pairs of secondary users. The wireless links from the PU to both STs are completely blocked and STs cannot detect PU's signal at all. As such, we set the miss-detection probability of each ST to be 1. Meanwhile, we assume that the correlation coefficient of the Rayleigh fading in two successive frames is 0.99. In this scenario, the interference from each ST to the BS changes slowly. In the simulation, we set that the received average SNR pp¯gpσ2 of the PU signal at the BS to 20 dB and each received average interference-to-noise ratio pk¯gkσ2 (k∈{1,2}) at

the BS to 5 dB. From the figure, the optimal transmission rate fluctuates around 3 kbits/frame, the transmission rate of the UCB learning algorithm increases from around 1.8 kbits/frame to around 2.1 kbits/frame, and the transmission rate of the SNR-based algorithm varies between 1 kbits/frame and 1.4 kbits/frame. Meanwhile, the proposed DRL-based MCS selection algorithm gradually achieves the optimal transmission rate of the optimal MCS selection algorithm, and is around 50% higher than that of the UCB learning algorithm, and is 100% higher than that of the SNR-based algorithm. This figure indicates that the proposed DRL-based algorithm is able to learn almost the perfect information of the quasi-static interference.

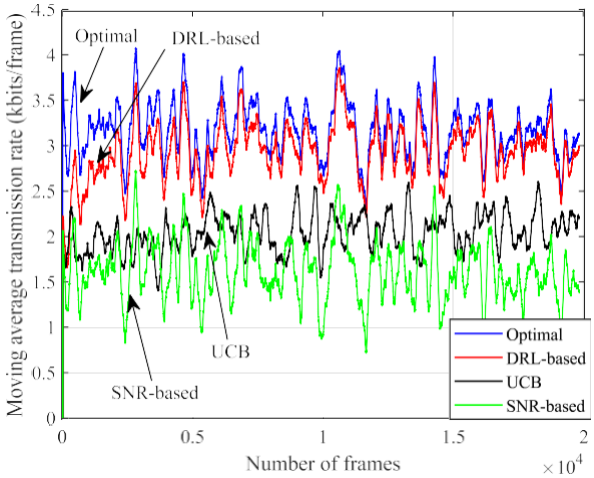

Fig.5. Transmission rate comparison in a dynamic interference scenario. Each value is a moving average of the previous 200 frames and each curve is the average of 20 trials.

Fig. 5 illustrates the transmission rates of different algorithms in a dynamic-interference scenario. We consider three secondary users, i.e., namely, (ST-1, SR-1), (ST-2, SR-2), and (ST-3, SR-3). The wireless links from the PU to ST-1 and ST-2 are completely blocked and ST-1/ST-2 cannot detect PU's signal at all, and the wireless link from the PU to ST-3 is not completely blocked but is extremely weak. As such, we set the miss-detection probabilities of the three STs to be 1, 1, and 0.5, respectively. Meanwhile, we set the correlation coefficient of the Rayleigh fading in two successive frames to be 0. In this scenario, the interference from each ST to the BS changes rapidly. In the simulation, we set that the received average SNR pp¯gpσ2 of the PU signal at the BS is 20 dB and each received average interference-to-noise ratio pk¯gkσ2 (k∈{1,2,3}) at the BS is 5 dB. From the figure, the optimal transmission rate is around 2.9 kbits/frame, the transmission rate of the UCB learning algorithm converges to around 2 kbits/frame, and the transmission rate of the SNR-based algorithm is around 1.3 kbits/frame. Meanwhile, the transmission rate of the proposed DRL-based MCS selection algorithm converges to around 2.6 kbits/frame, and is around 90% of the optimal transmission rate, and is around 30% higher than the transmission rate of the UCB learning algorithm, and is 100% higher than that of the SNR-based algorithm. This figure verifies the effectiveness of the proposed DRL-based algorithm when the interference from STs to the BS is highly dynamic.

V-CPERFORMANCE OF THE PROPOSED ALGORITHM WITH DIFFERENT Φ

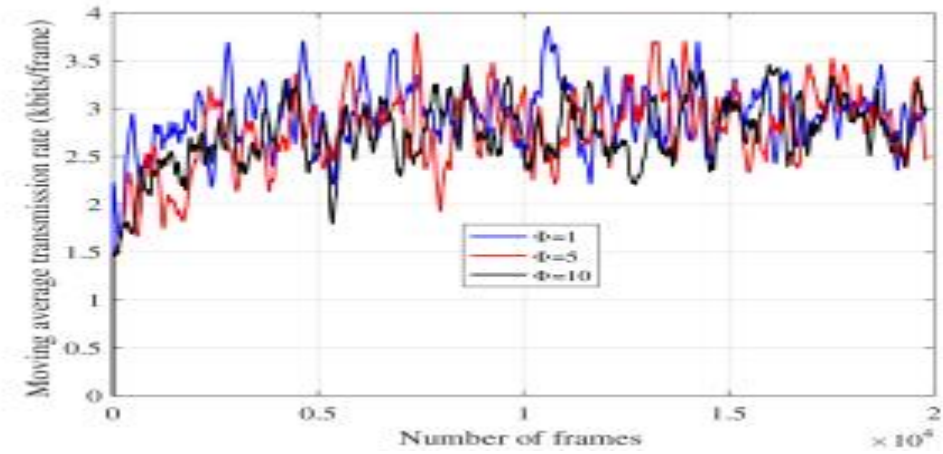

Figure 6: Transmission rate of the proposed algorithm with different Φ in a quasi-static interference scenario. Each value is a moving average of the previous 200 frames and each curve is the average of 20 trials.

Fig. 6 illustrates the transmission rate of the proposed algorithm with different Φ in a quasi-static interference scenario. In the simulation, the receive SNR of the PU signal at the BS, the number of secondary users, the miss-detection probability of each ST, and each receive interference-to-noise ratio at the BS are the same as those in Fig. 4. Besides, we set Φ=1, Φ=5, and Φ=10. From the figure, the transmission rate of the proposed algorithm remains almost the same when Φ increases from 1 to 10. The reason is as follows. The interference pattern from STs to the BS is dominated by the variation pattern of the corresponding channel gains. Since each interference channel gain changes slowly in a quasi-static interference scenario, the historical data in multiple previous frames provides almost the same interference pattern information as the historical data in the last frame for the DRL agent to infer the interference in the future. This figure indicates that it is unnecessary to put the historical data in multiple previous frames in each state when the interference from STs to the BS is quasi-static.

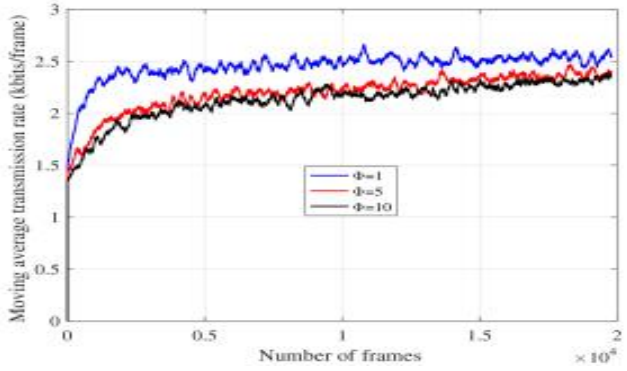

Figure 7: Transmission rate of the proposed algorithm with different Φ in a dynamic interference scenario. Each value is a moving average of the previous 200 frames and each curve is the average of 20 trials.

Fig. 7 investigates the transmission rate of the proposed algorithm with different Φ in a dynamic interference scenario. In the simulation, the receive SNR of the PU signal at the BS,

the number of secondary users, the miss-detection probability of each ST, and each receive interference-to-noise ration at the BS are the same as those in Fig. 5. Besides, we set Φ=1, Φ=5, and Φ=10. From the figure, the transmission rate of the proposed algorithm decreases as Φ increases from 1 to 10. The reason is as follows: As aforementioned, the interference pattern from STs to the BS is dominated by the variations of the corresponding channel gains. Since the channel model in the considered system is a first-order Markov process, each interference channel gain is only related to the interference channel gain in the previous frame. Note that each interference channel gain varies rapidly in a dynamic interference scenario. Then, the historical data in multiple previous frames cannot provide more interference pattern information than the historical data in the last frame for the DRL agent to infer the interference in the future, but in turn causes confusions to the DRL agent. As such, the DRL agent needs more frames to extract useful information about the interference pattern and infer the interference in the future. This figure indicates that it is harmful to put the historical data in multiple previous frames at each state when the interference from STs to the BS is highly dynamic.

V-D BALANCE BETWEEN TRANSMISSION RATE AND SYSTEM OVERHEADS

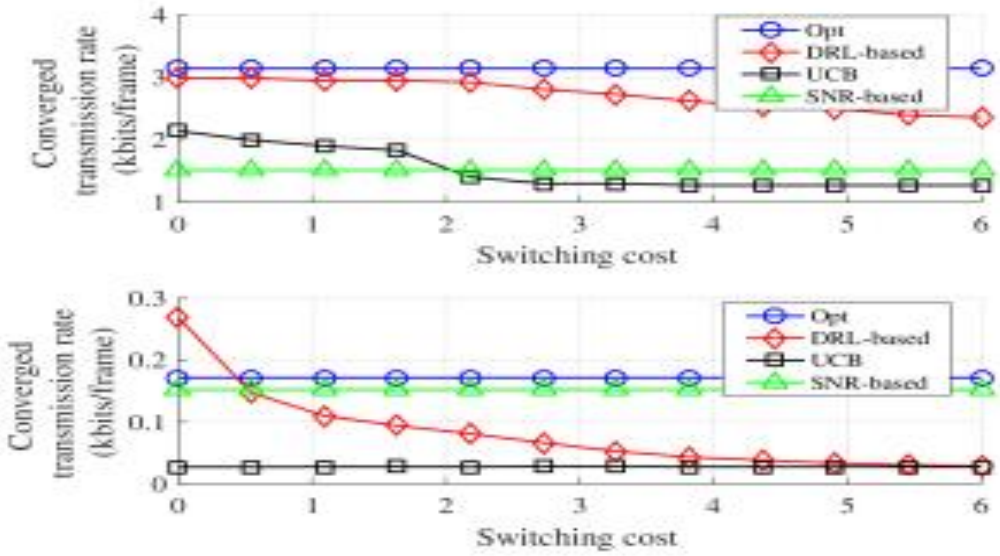

Figure 8: Converged transmission rates and switching rates of different algorithms in a quasi-static interference scenario.

Fig. 8 provides the converged transmission rates and the switching rates with different switching costs in a quasi-static interference scenario. In the simulation, the receive SNR of the PU signal at the BS, the number of secondary users, the miss-detection probability of each ST, and each receive interference-to-noise ration at the BS are the same as those in Fig. 4. For a given switching cost, each algorithm runs 20,000 frames similar to Fig. 4. The converged transmission rate is obtained by averaging the latest 5000 moving average transmission rates and the switching rate is Nswitching5000, where Nswitching is the number of switchings in the latest 5000 frames. In this figure, the converged transmission rates and the switching rates of the optimal algorithm and the SNR-based algorithm remain constant as the switching cost c grows. This is reasonable since the switching cost has no impact on both algorithms. Besides, The converged transmission rate of the UCB algorithm decreases from around 2.15 kbits/frame to around 1.4 kbits/frames as the switching cost increases from c=0 to c=6, and the corresponding switching rate remains around 0.03. The converged

transmission rate of the DRL algorithm decreases from around 3 kbits/frame to around 2.4 kbits/frames as the switching cost increases from c=0 to c=6, and the corresponding switching rate decreases from around 0.26 to around 0.03.

Fig. 8 indicates that, by adjusting the switching cost c, the DRL-based algorithm can achieve a higher converged transmission rate and a lower switching rate simultaneously than the SNR-based algorithm. For instance, when the switching cost is between 0.5 and 6, the converged transmission rate of the DRL-based algorithm is always higher than that of the SNR-based algorithm. Meanwhile the switching rate of the DRL-based algorithm is always lower than that of the SNR-based algorithm. Besides, by adjusting the switching cost c, the DRL-based algorithm can achieve a larger converged transmission rate than that of the UCB algorithm with a comparable switching rate. For instance, when the switching cost is between 4 and 6, the converged transmission rate of the DRL-based algorithm is always higher than that of the UCB algorithm, and the switching rates of both algorithms are almost identical. Therefore, when the interference from STs to the BS is quasi-static, the DRL-based algorithm can achieve a better balance between the primary transmission rate and system overheads than those of the optimal algorithm, the UCB algorithm, and the SNR-based algorithm.

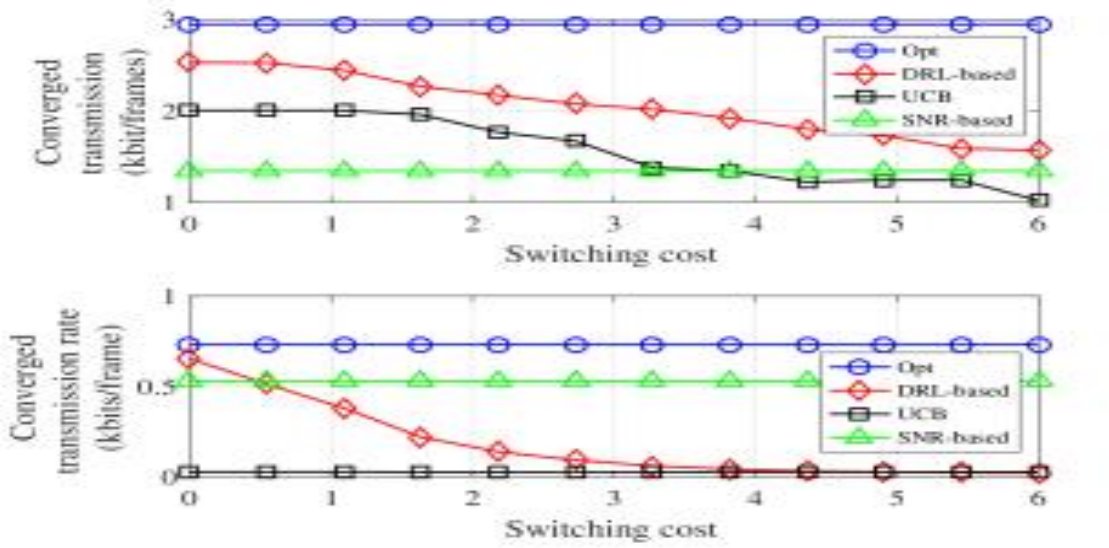

Figure 9: Converged transmission rates and switching rates of different algorithms in a dynamic interference scenario.

Fig. 9 provides the converged transmission rates and the switching rates with different switching costs in a dynamic interference scenario. In the simulation, the receive SNR of the PU signal at the BS, the number of secondary users, the miss-detection probability of each ST, and each receive interference-to-noise ratio at the BS are the same as those in Fig. 5. The converged transmission rate and switching rate are calculated with the same method as that in Fig. 8. In general, the trend of each curve in Fig. 9 is similar to that in Fig. 8. Specifically, by adjusting the switching cost c, the DRL-based algorithm can achieve a higher converged transmission rate and a lower switching rate simultaneously than those of the SNR-based algorithm. For instance, when the switching cost is between 0.5 and 6, the converged transmission rate of the DRL-based algorithm is always higher than that of the SNR-based algorithm, and the switching rate of the DRL-based algorithm is always lower than that of the

SNR-based algorithm. Besides, the converged transmission rate of the UCB algorithm ranges from 1.25 kbit/frame to 2 kbits /frame with a constant switching rate around 0.03. The performance of the UCB algorithm can be achieved by the DRL-based algorithm through adjusting the switching cost c between c=3.5 and c=6. Additionally, the DRL-based algorithm can also achieve a converged transmission rate higher than 2 kbits/frame with a switching rate higher than 0.03 when the switching cost c is between c=0 and c=3.5. In other words, when the interference from STs to the BS is highly dynamic, the DRL-based algorithm can achieve a converged transmission rate similar to the UCB algorithm for a tight switching rate constraint scenario, and achieve a higher converged transmission rate than the UCB algorithm for a loose switching rate constraint scenario. To summarize, when the interference from STs to the BS is dynamic, the DRL-based algorithm can achieve a better balance between the primary transmission rate and system overheads.

CONCLUSIONS

In this paper, we studied a cognitive HetNet and proposed an intelligent DRL-based MCS selection algorithm for the PR to learn the interference pattern from STs. With the learnt interference pattern, the DRL agent at the PR can infer the interference in the future frames and select a proper MCS to enhance the primary transmission rate. Besides, we took the system overhead caused by MCS switching into consideration and introduced a switching cost factor in the proposed algorithm to balance the primary transmission rate and system overheads. Simulation results showed that, the transmission rate of the proposed algorithm without the switching cost is 90% 100% to that of the optimal MCS selection scheme, is 30% higher than the UCB algorithm, and is 100% higher than that of the SNR-based algorithm. Meanwhile, the proposed algorithm with the switching cost factor can achieve a higher transmission rate than thõse of the benchmark algorithms without increasing system overheads.

BIBLOGRAPHY

1. Y.-P. E. Wang *et al.*, "A primer on 3GPP narrowband Internet of Things,"
2. *IEEE Communication. Mag.*, vol. 55, no. 3, pp. 117–123, Mar. 2017.
3. L. Zhang, M. Xiao, G. Wu, and S. Li, "Efficient scheduling and power allocation for d2d-assisted wireless caching networks," *IEEE J. Sel. Areas Commun.*, vol. 64, no. 6, pp. 2438–2452, Jun.2016.
4. L. Zhang, Y.-C. Liang, and M. Xiao, "Spectrum sharing for Internet of Things: A survey," *IEEE Wireless Commun.*, to be published. doi:10.1109/MWC.2018.1800259.
5. C. Yang, J. Li, M. Guizani, A. Anpalagan, and M. Elkashlan, "Advanced spectrum sharing in 5G cognitive heterogeneous networks," *IEEE Wireless Commun.*, vol. 23, no. 2, pp. 94–101, Apr.2016.
6. L. Zhang, J. Liu, M. Xiao, G. Wu, Y.-C. Liang, and S. Li, "Performance analysis and optimization in downlink NOMA systems with cooperative full-duplex relaying," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2398–2412, Oct.2017.
7. R. Q. Hu and Y. Qian, "An energy efficient and spectrum efficient wireless heterogeneous network framework for 5G systems," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 94–101, May2014.
8. Q. Zhang, H. Guo, Y.-C. Liang, and X. Yuan, "Constellation learning- based signal detection for ambient backscatter communication systems," *IEEEJ.Sel.AreasCommun.*,vol.37,no.2,pp.452–463,Feb.2019.

9. Q. Zhang, L. Zhang, Y.-C. Liang, and P. Y. Kam, "Backscatter-NOMA: A symbiotic system of cellular and Internet-of-Things networks," *IEEE Access*,vol.7,pp.20000–20013,2019.
10. G. Yang, Y.-C. Liang, R. Zhang, and Y. Pei, "Modulation in the air: Backscatter communication over ambient OFDM carrier," *IEEE Trans. Commun.*, vol. 66, no. 3, pp. 1219–1233, Mar.2018.
11. G. Yang, Q. Zhang, and Y.-C. Liang, "Cooperative ambient backscatter communications for green Internet-of-Things," *IEEE Internet Things J.*, vol. 5, no. 2, pp.1116–1130, Apr. 2018.
12. H. ElSawy, E. Hossain, and D. I. Kim, "HetNets with cognitive small cells: User offloading and distributed channel access techniques," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 28–36, Jun.2013.
13. J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3122–3134, Nov.2011.
14. W.Cheung, T. Q. S. Quek, and M. Kountouris, "Throughput opti- mization, spectrum allocation, and access control in two-tier femtocell networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 561–574, Apr.2012.
15. S. Mukherjee, "Distribution of downlink SINR in heterogeneouscellular networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 575–585, Apr.2012.
16. S.-Y. Lien, K.-C. Chen, Y.-C. Liang, and Y. Lin, "Cognitive radio resource management for future cellular networks," *IEEE Wireless Commun.*,vol.21,no.1,pp.70–79,Feb.2014.
17. Y.-C. Liang, Y. Zeng, E. C. Y. Peh, and A. T. Hoang, "Sensing- throughput tradeoff for cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1326–1337, Apr.2008.
18. L. Zhang, M. Xiao, G. Wu, M. Alam, Y.-C. Liang, and S. Li, "A survey of advanced techniques for spectrum sharing in 5G networks," *IEEE Wireless Commun.*, vol. 24, no. 5, pp. 44–51, Oct.2017.
19. N. C. Luong *et al.* (2018). "Applications of deep reinforcement learning in communications and networking: A survey." [Online]. Available: https://arxiv.org/abs/1810.07862
20. Y. Sun, G. Feng, S. Qin, Y.-C. Liang, and T.-S. P. Yum, "The SMART handoff policy for millimeter wave heterogeneous cellular networks," *IEEETrans.MobileComput.*,vol.17,no.6,pp.1456–1468,Jun.2018.
21. D. Nguyen, H. X. Nguyen, and L. B. White, "Reinforcement learning with network-assisted feedback for heterogeneous RAT selection," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6062–6076,Sep. 2017.
22. Y. Wei, F. R. Yu, M. Song, and Z. Han, "User scheduling and resource allocation in HetNets with hybrid energy supply: An actor-critic rein- forcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 17, no.1,pp.680–692,Jan.2018.
23. N.Morozs,T.Clarke,andD.Grace,"Heuristicallyacceleratedreinforce- ment learning for dynamic secondary spectrum sharing," *IEEE Access*, vol. 3, pp. 2771–2783,2015.
24. V. Raj, I. Dias, T. Tholeti, and S. Kalyani, "Spectrum access in cognitive radio using a two-stage reinforcement learning approach," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 20–34, Feb.2018.
25. O.-C. Iacoboaiea, B. Sayrac, S. B. Jemaa, and P. Bianchi, "SoNcoor- dination in heterogeneous networks: A reinforcement learning frame- work," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 5835–5847, Sep.2016.
26. M. Qiao, H. Zhao, L. Zhou, C. Zhu, and S. Huang, "Topology-transparentschedulingbasedonreinforcementlearninginself-organized wireless networks," *IEEE Access*, vol. 6, pp. 20221–20230,2018.

27. L. Xiao, Y. Li, G. Han, G. Liu, and W. Zhuang, "PHY-layer spoofing detection with reinforcement learning in wireless networks," *IEEETrans. Veh. Technol.*, vol. 65, no. 12, pp. 10037–10047, Dec.2016.
28. S. O. Somuyiwa, A. György, and D. Gündüz, "A reinforcement-learning approach to proactive caching in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 6, pp. 1331–1344, Jun.2018.
29. Y.He*etal.*,"Deep-reinforcement-learning-basedoptimizationforcache- enabled opportunistic interference alignment wireless networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 10433–10445, Nov.2017.
30. S. Wang, H. Liu, P. H. Gomes, and B. Krishnamachari, "Deep rein- forcement learning for dynamic multichannel access in wireless net- works," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 2, pp. 257–265, Jun.2018.
31. X. Liu, Y. Xu, L. Jia, Q. Wu, and A. Anpalagan, "Anti-jamming communications using spectrum waterfall: A deep reinforcement learn- ingapproach," *IEEE Commun. Lett.*, vol. 22, no. 5, pp. 998–1001, May2018.
32. X. Li, J. Fang, W. Cheng, H. Duan, Z. Chen, and H. Li, "Intelligent power control for spectrum sharing in cognitive radios: A deep rein- forcement learning approach," *IEEE Access*, vol. 6, pp. 25463–25473, Apr.2018.
33. Z.Wang,L.Li,Y.Xu,H.Tian,andS.Cui,"Handovercontrolinwireless systems via asynchronous multi-user deep reinforcementlearning," *IEEE Internet Things J.*, to bepublished.
34. Y. Yu, T. Wang, and S. C. Liew. (2018). "Deep-reinforcement learning multiple access for heterogeneous wireless networks." [Online]. Avail- able:https://arxiv.org/abs/1712.00162
35. Y. S. Nasir and D. Guo. (2019). "Multi-agent deep reinforcement learning for dynamic power allocation in wireless networks."[Online].
36. Available: https://arxiv.org/abs/1808.00490
37. Kim, B. C. Jung, H. Lee, D. K. Sung, and H. Yoon, "Optimal modulation and coding scheme selection in cellular networks with hybrid-ARQ error control," *IEEE Trans. Wireless Commun.*, vol. 7, no.12,pp.5195–5201,Dec.2008.
38. J. Meng and E. H. Yang, "Constellation and rate selection in adaptive modulation and coding based on finite blocklength analysis and its application to LTE," *IEEE Trans. Wireless Commun.*, vol. 13, no. 10, pp. 5496–5508, Oct.2014.
39. L. Zhang and Y.-C. Liang, "Average throughput analysis and optimiza- tion in cooperative IoT networks with short packet communication," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 11549–11562, Dec. 2018.
40. T. Kim, D. J. Love, and B. Clerckx, "Does frequent low resolution feedback outperform infrequent high resolution feedback for multiple antenna beamforming systems?" *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1654–1669,Apr. 2011.
41. L. Zhang, M. Xiao, G. Wu, S. Li, and Y. C. Liang, "Energy-efficient cognitive transmission with imperfect spectrum sensing," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1320–1335, May2016.
A. Farrokh, V. Krishnamurthy, and R. Schober, "Optimal adaptive modulation and coding with switching costs," *IEEE Trans. Commun.*, vol. 57, no. 3, pp.697–706, Mar. 2009.
42. R. Alnwaimi and H. Boujemaa, "Adaptive packet length and MCS usingaverageorinstantaneousSNR,"*IEEETrans.Veh.Technol.*, vol.67, no. 11, pp. 10519–10527, Nov.2018.

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Vol. 8 Issue 12
December 2021

43. V. Mnih*et al.*, "Human-level control through deep reinforcement learn- ing," *Nature*, vol. 518, pp. 529–533,2015.
44. C. Shen, C. Tekin, and M. van der Schaar, "A non-stochastic learning approach to energy efficient mobility management," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3854–3868, Dec.2016.
45. J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. New York, NY, USA: McGraw-Hill,2007.
46. T.TielemanandG.Hinton,"Lecture6.5-RMSPROP:Dividethegradient by a running average of its recent magnitude," *COURSERA, Neural Netw. Mach. Learn.*, vol. 4, no. 2, pp. 26–31,2012.