Artificial Intelligence Empowered Power Allocation for Smart Railway

Jianpeng Xu and Bo Ai

Smart railway is in the process of vigorous development based on the advancement of numerous emerging and advanced techniques, such as artificial intelligence and big data. However, a smart railway network is exceedingly time-varying and complicated, which presents great challenges to power allocation. Deep reinforcement learning is capable of effectually improving the intelligence as well as cognition of high-speed railway network, which helps to optimize the power allocation problems with both time-varying and complicated characteristics. In this article, we first provide an overview of the existing power allocation methods, including advantages, disadvantages, and complexity, as well as characteristics. Then we investigate an innovative power allocation algorithm based on multi-agent deep recurrent deterministic policy gradient (MADRDPG), which is capable of learning power decisions from past experience instead of an accurate mathematical model. Finally, numerical results indicate that the performance of the MADRDPG-based method significantly outperforms existing state-of-the-art methods in terms of spectrum efficiency and execution latency.

Smart railway is in the process of vigorous development based on the advancement of numerous emerging and advanced techniques, such as artificial intelligence and big data. However, a smart railway network is exceedingly time-varying and complicated, which presents great challenges to power allocation.

Introduction

Being a comfortable and green way of transportation, high-speed railway (HSR) is growing to enter the new era of smart railway, where more and more interconnection among infrastructures, passengers, and high-speed trains will be established. Smart railway makes railway transportation safer, greener, as well as more convenient by utilizing numerous emerging and advanced techniques, such as artificial intelligence (AI), 5G, and big data [1, 2]. Nowadays, a growing number of countries have presented and proceeded with the construction of smart railway. To accelerate construction of smart railway, it is significant to provide high data transmission rate in HSR scenarios. Nevertheless, according to the measurements of China Mobile, spectrum efficiency in realistic HSR scenarios is only 1 b/s/Hz [3], which needs to be greatly improved to satisfy the growing requirements for smart railway communication services.

Achieving high data rate over HSR networks has always been a rather challenging task. On one hand, from the perspective of the transport layer, multi-path TCP is capable of increasing the data rate in HSR environments, guaranteeing service dependability [4]. On the other hand, millimeter-wave (mmWave) communication is capable of providing multi-gigabit data rate, which is regarded as a promising technique to provide great experience for high-speed mobile users [5]. The mmWave signals experience high path loss, which results in the emergence of large-scale antenna arrays in mmWave communications. Utilizing a beamforming scheme for large-scale antenna arrays is considered as a kind of up-and-coming technique to enhance the spectrum efficiency in HSR environments, and in recent years has been extensively researched [3]. Nevertheless, few works consider the joint optimization of power allocation and hybrid beamforming in mmWave HSR systems. In [5], the authors claimed that they were the first to investigate the power allocation problem in mmWave HSR systems to improve energy efficiency based on a precise mathematical model. However, the optimization-based methods are very difficult to make work well in the realistic HSR environment with time-varying characteristics [13], where there are a few disadvantages as follows. First, they assume that the critical factors such as channel conditions can be accurately known. Nevertheless, the train travels at a high speed, which intuitively leads to highly dynamic as well as complicated patterns of HSR networks. Consequently, it is impossible to accurately acquire the channel conditions of the HSR environment in practice. Second, by leveraging most of these approaches, optimal or near optimal solutions can be obtained only for one snapshot of the HSR communication systems. They ignore the long-term impact of present decisions on power allocation. Third, they need a considerable number of iterations to obtain the optimal or near optimal solutions, which are inapplicable for making real-time power allocation decisions in time-varying HSR channels.

Recently, AI has been taken as a significant direction in future HSR networks [1–3]. Compared to optimization-based methods, AI, especially deep reinforcement learning (DRL), is capable of effectivelly improving the intelligence as well as cognition of HSR networks, which helps to address the resource allocation problems under uncertainty, such as the environment with both time-varying and complicated characteristics [2, 13]. Since the railway transportation lines are fixed, it is possible to make use of the previously acquired runtime statistics for the same train or other trains travelling on the same transportation line. DRL is beneficial for learning and establishing the characteristic profiles of the HSR channels, which then facilitates resource allocation.

Digital Object Identifier:
10.1109/MCOM.001.2000634

The authors are with Beijing Jiaotong University.

0163-6804/21/$25.00 © 2021 IEEE

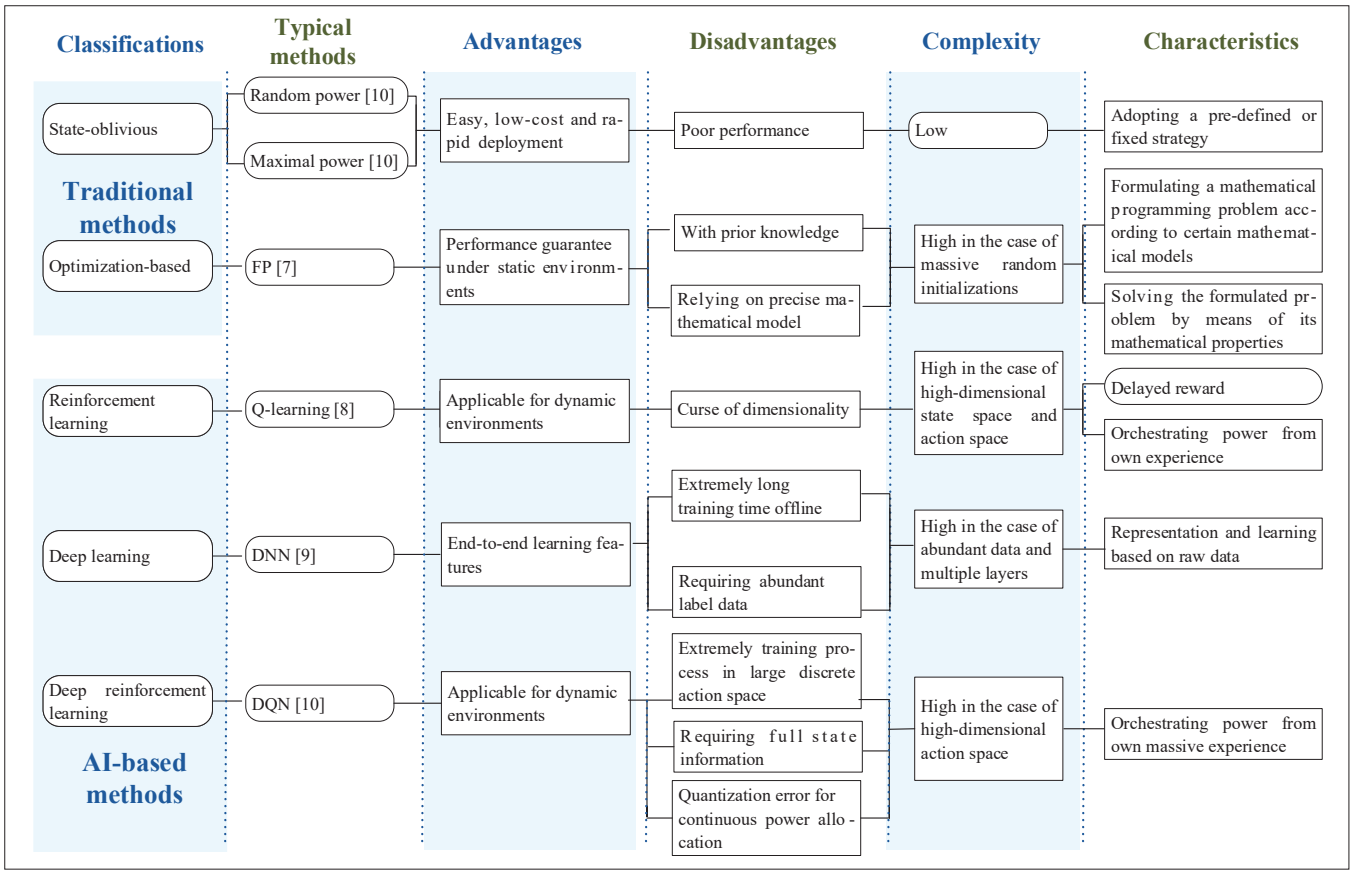

FIGURE 1. A summary of existing approaches in power allocation.

Note that although it takes a lot of time to train the AI model, the resource allocation decision can be inferred within a few computations after the AI model is trained to be stable [10]. Consequently, we endeavor to investigate the power allocation problem from a new point of view based on AI-empowered mmWave HSR network architecture. We investigate an innovative power allocation algorithm based on multi-agent deep recurrent deterministic policy gradient (MADRDPG), which makes the agent have the capacity of intelligently orchestrating power resource from the past experience to achieve maximum spectrum efficiency even under partial observations.

The remainder of this article is structured as follows. We first provide an overview of the existing power allocation methods, including advantages, disadvantages, complexity, as well as characteristics. Next, we formulate a power allocation problem in an mmWave HSR system to maximize achievable sum rate and exploit MADRDPG to develop an intelligent power allocation method. After that, numerical results demonstrate that the MADRDPG-based method outperforms other state-of-the-art methods in spectrum efficiency and execution latency. At last, the article is concluded.

State-of-the-Art Approaches for Power Allocation

In this section, we summarize the traditional power allocation methods and state-of-the-art AI-based power allocation methods, including advantages, disadvantages, complexity, and so on, which are listed in Fig. 1.

Traditional Power Allocation Methods

Traditional power allocation methods can be split into two classes: state-oblivious and optimization-based. A state-oblivious scheme generally allocates the power by adopting a pre-defined or fixed strategy, such as maximal power [10]. An optimization-based approach is usually made up of two steps. The first step is that the power allocation problem is formulated into a mathematical programming problem on the basis of certain mathematical models.

The second step is to solve the formulated problem by means of its mathematical properties, such as convex programming. Fractional programming (FP) [7] is a representative optimization-based method.

It seems to us that the two sorts of methods will not work well for highly dynamic HSR networks. Although the state-oblivious approach is simple and feasible to solve power allocation problem, it results in poor performance, such as low spectrum efficiency. This is a result of not taking into account runtime states of the mmWave HSR communication systems. In addition, under both low-speed and slowly varying scenarios, optimization-based approaches can exhibit good performance with the following two requirements. First, exact prior knowledge of user requirements and network parameters is acquired, such as exact instantaneous channel state information (CSI). Second, network environments as well as user requirements can accurately be characterized in a mathematical way. Nevertheless, it is very challenging to satisfy these two requirements in time-varying mmWave HSR networks. In addition, FP generally needs numerous iterations to converge, making real-time power allocation infeasible.

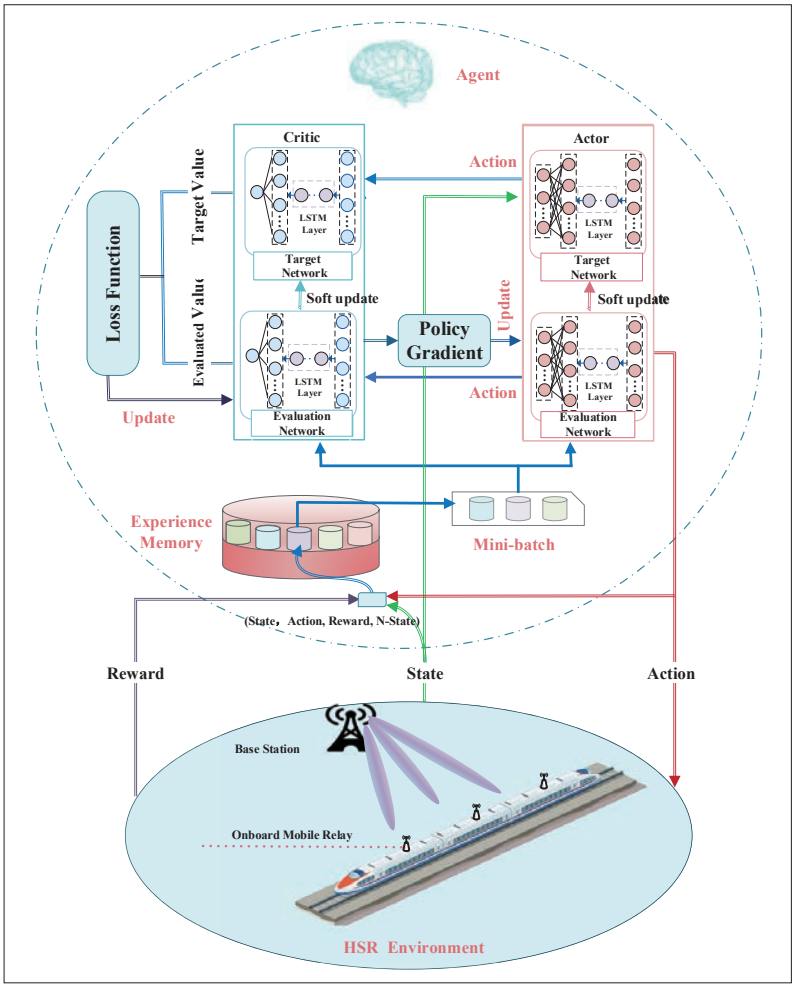

FIGURE 2. The procedure of MADRDPG to settle the power allocation problem in smart railway.

AI-Based Power Allocation Methods

AI has been proven to be an effective method to make clever decisions under uncertain circumstances [13]. Moreover, AI is also considered as the prospective technique to settle the complex power allocation problem [8–10]. There are three popular machine-learning-based power allocation methods.

Deep learning (DL): DL is a representation learning approach, in which features required for either classification or prediction can be automatically extracted from large amounts of raw data. The authors of [8] applied deep neural network (DNN), to settle the power allocation problem in device-to-device (D2D) communications. They used the transmit power level achieved by exhaustive search as label data to accomplish multiple objectives, including maximizing either the energy efficiency or spectrum efficiency, or minimizing the total transmit power. DNN is trained in an end-to-end way. Thus, it can derive the optimal strategy for power allocation from practical channel data rather than the analytical system model, achieving much better performance in the real world. Unfortunately, DL demands optimal labeled data in a realistic training model, while it is difficult to obtain perfect training data in time-varying as well as complicated wireless systems. Furthermore, it needs to spend a lot of training time due to physical constraints.

Reinforcement learning (RL): RL is a branch of machine learning whose advantage is primarily that it is capable of enabling the agent to make wise decisions with low demand for prior knowledge [13]. Unlike optimization-based methods that neglect the long-term impact of the current decision on power allocation, RL concentrates on maximizing long-term rewards, which is advantageous to the decision making problem of power allocation in dynamic as well as time-varying systems. In [9], the authors investigated the subchannel and power allocation problem in non-orthogonal multiple access systems with the help of Q-learning, which is a typical RL algorithm whose target is to maximize energy efficiency. Nevertheless, due to the mmWave HSR environments' high time variability, it will generate high-dimensional state and action space, significantly increasing computational complexity (i.e., the curse of dimensionality problem). In addition, handcrafted features have a great impact on performance of such a power allocation strategy.

DRL: DRL combines DL with RL. A typical DRL algorithm is deep Q-network (DQN). DQN, which leverages DNN to represent state-action pairs and approximates the Q-function, is able to yield better performance than Q-learning for systems with high-dimensional state and action space. The authors of [10] used multi-agent DQN-based power allocation aiming to maximize the weighted sum rate for a multi-cell system, where all base stations (BSs) and mobile users are equipped with a single antenna. The simulation results verified that applying DQN techniques in power allocation is particularly suitable for the both inaccurate and non-negligible CSI delay system model. However, since DQN can only output discrete decisions, it will lead to quantization error for power allocation. Furthermore, DQN only has good performance when observing full state information. However, due to time-varying characteristics, the agent in HSR networks only observes partial state information with high probability, which is harmful to the performance of DQN. Moreover, with the increasing number of actions, it will be extremely time-consuming for training.

In a word, these above works may not work well over realistic HSR networks since they do not consider the high-speed mobility of smart railway. In this article, we investigate an innovative power allocation method, referred to as MADRDPG for smart railway, which is shown in Fig. 2.

Power Allocation Based on MADRDPG

System Model

In this article, we consider power allocation for the mmWave multiuser multiple-input multiple-output (MU-MIMO) HSR system. According to the mmWave HSR scenario recommended by 3rd Generation Partnership Project (3GPP) [11], a potential mmWave HSR system model is shown in Fig. 2.

With the purpose of reducing the penetration losses caused by the Faraday cage characteristics of a high-speed train, the mobile relay (MR) connected to the in-cabin wireless access point is deployed on top of each carriage. The MR communicates with the BS, playing the role of forwarding data between the passengers on the train and the broadband wireless networks. MRs as well as BSs take advantage of the mmWave beamforming technique to

further improve the quality of signals. In this article, we concentrate on the power allocation problem between the MRs and the BS with the objective of maximizing system achievable sum rate.

A BS equipped with multiple antennas employs hybrid beamforming to serve an MR with analog-only beamforming. Due to power consumption as well as hardware limitations, every MR with multiple antennas communicates with the BS via only one data stream. The mmWave HSR channel in [6] is adopted. In the investigated power allocation, this work considers two constraints. First, the power allocated to each MR should not only be non-negative, but also be no more than maximum transmitting power. Second, the minimum quality of service (QoS) requirements of each MR should be met.

MADRDPG Structure

Power allocation in the mmWave HSR systems, where multiple MRs attempt to communicate with the BS, can be characterized as the multi-agent DRL issue. Each MR plays the part of an agent and interacts with the HSR environment to accumulate adequate experiences employed to guide its own power allocation strategies. Specifically, at each time slot, after observing the state that includes the features of the HSR environment, each agent takes an action based on its policy, forming joint actions, and then receives a reward. Afterward, the environment turns into the new state. We introduce the critical elements of the MADRDPG-based power allocation algorithm in great detail.

State: The state in DRL denotes a space to indicate the circumstances of the HSR environment. The elements of state are every MR agent's channel gain of current time step, beamforming design of the current time step, and achievable rate and emitting power of the last time step. Be aware that this article primarily aims to study the power allocation problem in mmWave HSR systems. How to design an efficient and reliable beamforming scheme in mmWave HSR systems based on DRL is an open issue [2, 3] and a very interesting direction worthy of further study. However, it is beyond the scope of this work.

Action: As shown in Fig. 2, each agent's action, which indicates the power allocated to each MR, is directly determined by the actor network of MADRDPG.

Reward: The reason DRL has an advantage in addressing problems that are hard to optimize lies in the flexible design of its reward function [13]. In this work, the reward function of each agent at each time step is as follows. When the MR's QoS constraints are satisfied, the reward is system achievable sum rate; otherwise, the reward is zero.

Be aware that in order to make all agents behave cooperatively, all MRs have a share in the identical reward of the system. In addition, in this article, the achievable sum rate is taken as spectrum efficiency.

In realistic HSR environments, the agent only observes partial state information instead of all the state information with high probability. In the case that the agent observes partial state information, the interaction among agents will also change, which makes the training process of multiple agents more complex. That is because each agent observes the results related to its own action, as well as the actions of other agents. To alleviate the effect of instability, which is due to observing partial state in the HSR scenario, we combine multi-agent deep deterministic policy gradient (MADDPG) [14] with RNN structure, specifically long short-term memory (LSTM) [12], which is named MADRDPG, shown in Fig. 2. With the help of LSTM structure, the agent is able to maintain an internal state and aggregate previous observation information, which is beneficial for predicting the system state. More details of LSTM can be found in [12]. Note that all parameters of MADRDPG are updated in the training stage. The specific workflow of MADRDPG is as follows.

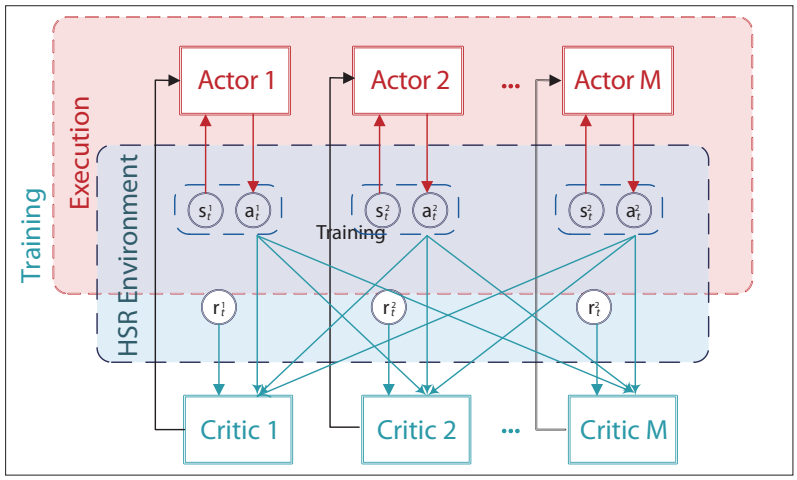

FIGURE 3. MADRDPG framework of centralized training with distributed execution.

First, the parameters of each agent's actor network and critic network are randomly initialized. The structures of each agent's target networks are the same as those of the evaluation networks. After that, the parameters of evaluation networks are slowly updated through training and are cloned to the target networks periodically. After the state is inputted to each agent, it is first processed by the LSTM layer. Then each agent's actor network derives the action. In this work, we also use experience memory to improve learning stability. After saving samples into the experience memory, we randomly sample a mini-batch to train both critic and actor networks. The parameters of the actor network are updated with the policy gradient, while the parameters of the critic network are updated according to minimizing the loss function.

Training and Execution

There are three inherent problems in the MADRDPG framework.

High complexity: Due to the fact that the joint state-action space extension strictly takes place based on the number of diverse variables coming from agents, the complexity of MADRDPG increases exponentially with the increase of agents.

Credit assignment: In the MADRDPG framework, we generally only focus on the cumulative reward, which corresponds to actions taken by all agents, while ignoring the contribution of every agent.

Nonstationary equilibrium: All agents maintain dynamic evolution over time as a result of the stress that comes from the cooperation among all agents.

To overcome these drawbacks, the MADRDPG framework is categorized into two stages, training and execution, as illustrated in Fig. 3.

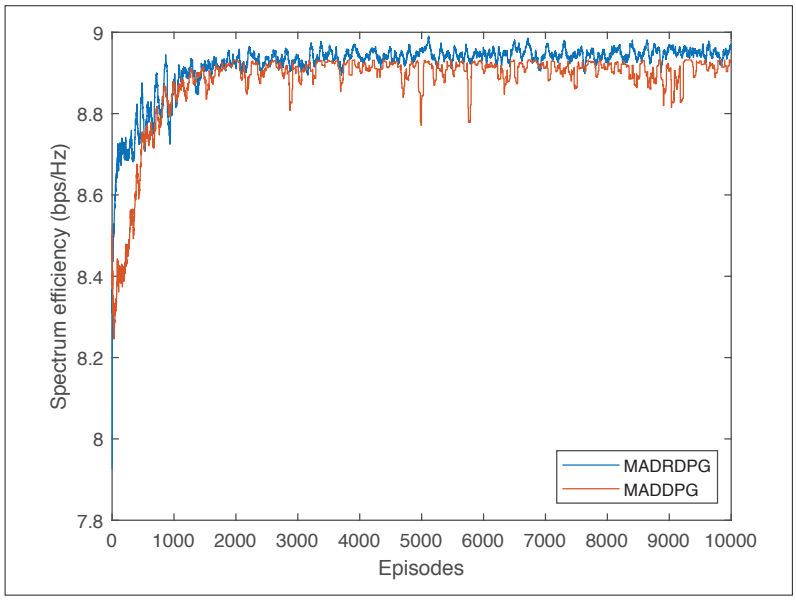

FIGURE 4. Spectrum efficiency comparison of MADRDPG and MADRDPG schemes during the training stage

Since a stable agent can not only reduce the computational complexity of MADRDPG effectively, but also strengthen the stability of other agents [15], we use a distributed strategy during the execution stage, which enables each agent to adaptively adjust its power policy effectively and efficiently based on past experience. Multiple actor networks are capable of intelligently orchestrating sound power strategies via interacting with the HSR environment. All available state information and actions are input into centralized critic networks to evaluate the generated actions during the training stage. It is clear that the MADRDPG framework is constrained by past experience. Consequently, for the sake of gaining enough experience to make timely and effective decisions under similar circumstances, it is essential to collect a sufficient number of samples for training the agents.

Be aware that the training stage, with a large amount of computation, can be executed offline in a wide range of HSR channel environments, while the low-cost execution process is performed online in the deployment. Because the transportation line of the train is constant within the rigorous scheduled timing, it is intuitive to generate the space-time rules in HSR wireless systems. Based on environment dynamics [13], unless the HSR environment properties change considerably, there is no need to retrain all agents at short intervals. That is to say, it is only necessary to retrain all agents once a week or even a month.

Numerical Results

In this section, the performance of the MADRDPG-based power allocation scheme in the mmWave HSR system is verified using TensorFlow and Python by carrying out extensive simulation experiments. In the simulations, the maximum transmission power of the MR is set to 23 dBm. The number of MRs is 4. We set the additional white Gaussian noise power as −174 dBm. The MR and BS are equipped with 16 and 64 antennas, respectively. The carrier frequency is set as 28 GHz [6].

Next, hyper-parameters employed for the framework of MADRDPG are presented. Each agent's actor and critic networks contain a two-layer LSTM and three hidden layers. The rectified linear unit (ReLU) is regarded as the activation function between hidden layers. The network parameters of both the actor and critic networks are updated by using the Adam algorithm with a learning rate of 0.0001 and 0.001, respectively. We compare the MADRDPG-based method with existing methods, including FP [7, Algorithm 3] with instantaneous CSI, DQN [10], maximal power, as well as random power.

Figure 4 compares the spectrum efficiency results of MADRDPG and MADRDPG without LSTM (namely MADDPG) during the training stage. From the figure, we can observe that the trend curve of MADRDPG is the same as that of MADDPG. However, the performance of MADRDPG is not only more stable, but also slightly better than MADDPG in the final stage of convergence, which indicates that with the help of LSTM, the agent can learn a more effective power allocation strategy.

Figure 5 compares the spectrum efficiency of different power allocation algorithms over 500 samples of the mmWave HSR channel realizations when the speed of the train is 360 km/h, where MADRDPG as well as DQN only observe partial state information with a certain probability. Note that before the testing, we have trained MADRDPG with 16,000 independent mmWave HSR wireless channel realizations, and MADRDPG's power decisions tend to be stable. It is easy to see that in the time-varying mmWave HSR system, MADRDPG outperforms other schemes in terms of spectrum efficiency. That is because MADRDPG enables each MR agent to adaptively adjust its power policy effectively and efficiently based on past experience, instead of depending on either any predefined strategy or any precise mathematical model. We also observe that the curves of the MADRDPG, DQN, and FP methods are not as stable as those of maximal power and random power, which is decided by the time-varying characteristics of the HSR channel. However, the performance of DQN is slightly worse than FP in a few mmWave HSR channel realizations, while MADRDPG is always better than FP as well as DQN. That is because DQN can only take action based on limited discrete power levels, while MADRDPG can orchestrate power resources based on continuous action space. Moreover, compared to DQN, MADRDPG can make use of previous observations with the help of LSTM, which helps MADRDPG to predict the future states.

Last but not least, the performance of both the random power and maximal power schemes is always poorer than MADRDPG, DQN, and FP, which illustrates that the joint optimization of power allocation and beamforming is beneficial to improve the achievable sum rate of an mmWave HSR system. It is worth mentioning that in the execution stage of MADRDPG and DQN, each agent can intelligently orchestrate power within 0.63 ms, while FP requires about 36 times longer execution latency, which is conducive to the realization of real-time power allocation.

Figure 6 compares the spectrum efficiency of different power allocation algorithms with varying speeds of trains — low speed (50 km/h), middle speed (160 km/h), and high speed (360 km/h) — which verifies the advantages of the MADRDPG-based algorithm. Furthermore, we find that with increasing speed, the spectrum efficiency achieved

by all power allocation algorithms significantly decreases. The reason is that the faster the speed of a train, the more serious is the inter-channel interference caused by Doppler effect to which it leads.

Conclusions

Railway is envisaged to evolve into the era of smart railway. In this article, we first provide an overview of the existing power allocation methods, including advantages, disadvantages, complexity as well as characteristics. Furthermore, we formulate a power allocation problem for smart railway to maximize spectrum efficiency, which is settled by leveraging an innovative multi-agent DRL method, referred to as MADRDPG. Numerical results demonstrate its effectiveness and advantage over other state-of-the-art methods in the mmWave HSR communication system in terms of spectrum efficiency and execution latency. In future work, we will investigate the capability of extension of the MADRDPG-based approach to mobile edge computing and caching resource allocation in HSR scenarios.

Acknowledgments

This work was supported in part by the NSFC under Grant U1834210, 61725101 and 6196113039, in part by the Royal Society Newton Advanced Fellowship under Grant NA191006, in part by the National Key Research and Development Program under Grant 2016YFE0200900, in part by the Beijing Natural Haidian Joint Fund under Grant L172020, in part by the Major Projects of Beijing Municipal Science and Technology Commission under Grant Z181100003218010, in part by the State Key Lab of Rail Traffic Control and Safety under Grant RCS2018ZZ007, RCS2020ZT010 and RCS2019ZZ007 and in part by the Fundamental Research Funds for the Central Universities 2020YJS201, and in part by the Open Research Fund from Shenzhen Research Institute of Big Data under Grant 2019ORF01006.

References

[1] B. Ai *et al.*, "5G Key Technologies for Smart Railways," *Proc. IEEE*, vol. 108, no. 6, June 2020, pp. 856–93.
[2] L. Yan *et al.*, "AI-Enabled Sub-6-GHz and mm-Wave Hybrid Communications: Considerations for Use with Future HSR Wireless Systems," *IEEE Vehic. Tech. Mag.*, vol. 15, no. 3, Sept. 2020, pp. 59–67.
[3] S. Han *et al.*, "Achieving High Spectrum Efficiency on High Speed Train for 5G New Radio and Beyond," *IEEE Wireless Commun.*, vol. 26, no. 5, Oct. 2019, pp. 62–69.
[4] J. Xu *et al.*, "When High-Speed Railway Networks Meet Multipath TCP: Supporting Dependable Communications," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, Feb. 2020, pp. 202–05.
[5] L. Wang *et al.*, "Energy-Efficient Power Control of Train-Ground mmWave Communication for High-Speed Trains," *IEEE Trans. Vehic. Tech.*, vol. 68, no. 8, Aug. 2019, pp. 7704–14.
[6] M. Cheng *et al.*, "A Fast Beam Searching Scheme in mmWave Communications for High-Speed Trains," *IEEE ICC*, Shanghai, China, 2019, pp. 1–6.
[7] K. Shen *et al.*, "Fractional Programming for Communication Systems — Part I: Power Control and Beamforming," *IEEE Trans. Signal Processing*, vol. 66, no. 10, 15 May15, 2018, pp. 2616–30.
[8] W. Lee *et al.*, "Intelligent Resource Allocation in Wireless Communications Systems," *IEEE Commun. Mag.*, vol. 58, no. 1, Jan. 2020, pp. 100–5.
[9] H. Zhang *et al.*, "Artificial Intelligence-Based Resource Allocation in Ultradense Networks: Applying Event-Triggered Q-Learning Algorithms," *IEEE Vehic. Tech. Mag.*, vol. 14, no. 4, Dec. 2019, pp. 56–63.
[10] Y. S. Nasir and D. Guo, "Multi-Agent Deep Reinforcement Learning for Dynamic Power Allocation in Wireless Networks," *IEEE JSAC*, vol. 37, no. 10, Oct. 2019, pp. 2239–50,.
[11] 3GPP TR 38.913 V0.3.0, "Technical Specification Group Radio Access Network: Study on Scenarios and Requirements for Next Generation Access Technologies (Release 14)," Jan. 2016..

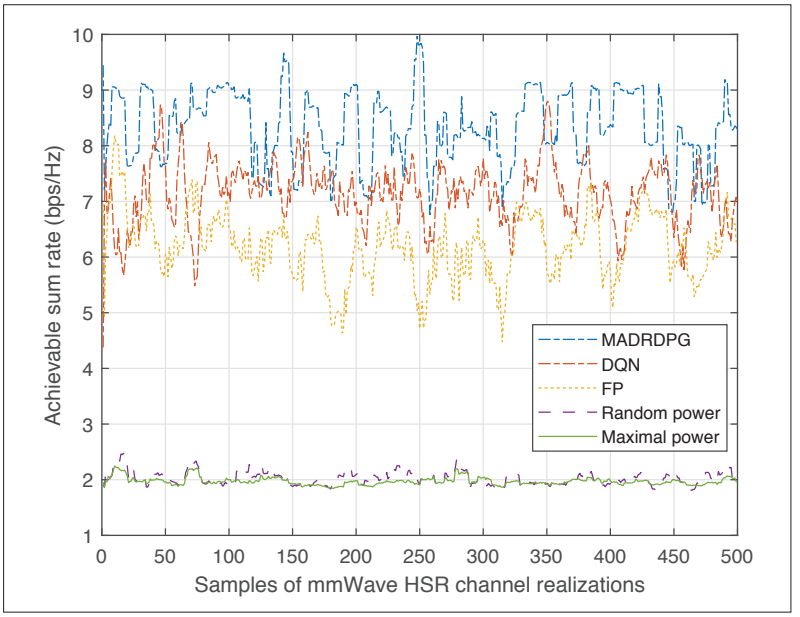

FIGURE 5. Spectrum efficiency comparison of different algorithms over HSR networks.

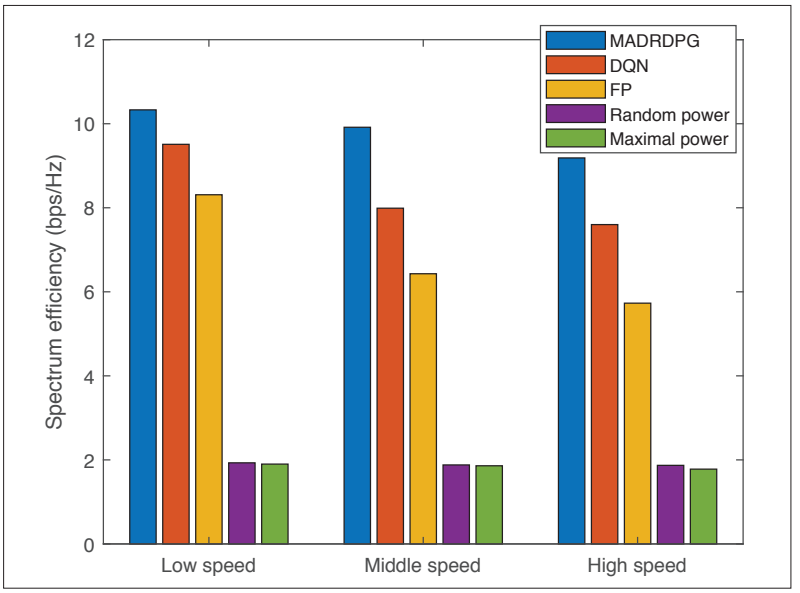

FIGURE 6. Spectrum efficiency comparison of different algorithms with varying speeds of trains.

[12] K. Greff *et al.*, "LSTM: A Search Space Odyssey," *IEEE Trans. Neural Networks and Learning Systems*, vol. 28, no. 10, Oct. 2017, pp. 2222–32.
[13] L. Liang, H. Ye, and G. Y. Li, "Spectrum Sharing in Vehicular Networks Based on Multi-Agent Reinforcement Learning," *IEEE JSAC*, vol. 37, no. 10, Oct. 2019, pp. 2282–92.
[14] R. Lowe *et al.*, "Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments," *Advances in Neural Info. Processing Systems*, 2017, pp. 6379–90.
[15 B. Deng *et al.*, "The Next Generation Heterogeneous Satellite Communication Networks: Integration of Resource Management and Deep Reinforcement Learning," *IEEE Wireless Commun.*, vol. 27, no. 2, Apr. 2020, pp. 105–11.

Biographies

Jianpeng Xu [S'20] (18111037@bjtu.edu.cn) is currently pursuing a Ph.D. degree with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, China. His research interests include mobility management, and deep reinforcement learning.

Bo Ai [SM'10] (boai@bjtu.edu.cn) is a professor with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University. He is the deputy director of the State Key Lab of Rail Traffic Control and Safety in China. His research interests include rail traffic mobile communications and channel modeling.