WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 3, Number 1, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v3n1.01
Journal homepage: https://wepub.org/index.php/IJCSIT/index

A Review of Personalized Federated Reinforcement Learning

Gaofeng Chen, Qingtao Wu

Henan University of Science and Technology, Luoyang 471000, China

ABSTRACT

Reinforcement learning and federated learning both provide strong theoretical support for the study of artificial intelligence. In recent years, an emerging federated reinforcement learning paradigm has been proposed and widely studied and applied. However, in the federated reinforcement learning architecture, the environment, data type and device performance of different agents may be different, which is called heterogeneity. The existence of heterogeneity factors may lead to slow convergence speed of the algorithm, poor generalization, and suboptimal quality of the trained model. Therefore, how to solve the negative impact of the heterogeneity problem on model training has become a hot content of research, and the most important method is to train personalized models for clients. This paper introduces the theory of federated reinforcement learning, as well as methods to cope with heterogeneity in federated reinforcement learning, and provides an overview of the applications of federated reinforcement learning. We conclude the paper with a summary and future perspectives.

KEYWORDS

Federated reinforcement learning; Heterogeneity factors; Personalized model training

1. INTRODUCTION

With the development of society, the application of artificial intelligence has become an essential part of life, greatly promoting the development of industry, agriculture, education and other industries, such as smart home [1], driverless cars [2], intelligent robots and so on [3]. And behind these smart devices, it is inseparable from the support of efficient Machine Learning (ML) algorithms. The goal of ML algorithms is to solve an optimal function model that can approximate these data by analyzing the existing data, through which an accurate prediction of the future posture can be achieved [4]. Therefore, designing an efficient and stable ML algorithm is an important goal in the field of ML research. ML is classified into supervised learning, unsupervised learning, deep learning, and reinforcement learning [5]. Early ML is usually a centralized framework, where a large amount of private data from dispersed users needs to be sent to a central service node in order to train accurate algorithmic models. However, this is prone to the problems of data leakage and increased communication overhead, and with the strengthening of people's awareness of privacy protection and the introduction of some national legal protection measures for related Internet data, it becomes more difficult to obtain local user data for training algorithmic models.

To solve the above dilemma, Google's research team proposed the algorithmic architecture of Federated Learning (FL) in 2016 [6, 7]. FL is actually a special kind of distributed ML framework [8]. The basic idea is that in each round of training, a central server is used to aggregate model parameter information from multiple local clients, and the aggregated global model parameters are fed back to the local clients for them to conduct the next round of training based on the aggregated global model, and this process is iterated for many times until a globally optimal model is obtained. Since the local client exchanges with the central server not local data but model parameters obtained

from local training, FL has the advantages of facilitating model training, protecting data privacy and security, and saving network resources [9]. FL has been continuously optimized and improved by subsequent studies, and has also been proven to be significantly effective and superior in a large number of practical applications.

Reinforcement Learning (RL) is one of the fields of ML, in which the agent can learn the optimal policy through the interaction with the environment to maximize the long-term reward or achieve the predetermined optimal goal [10]. Classical RL algorithms include Q-Learning, Sarsa, Policy Gradient, etc. [11]. In order to obtain better model training effects and protect user data privacy, some scholars have proposed to combine RL and FL to form a new Federated Reinforcement Learning (FRL) framework [12, 13]. However, in practice, due to reasons such as differences in the environment or equipment, resulting in most of the data belonging to Not Independent and Identically Distributed (Non-IID) among them [14]. These heterogeneity may cause some new problems in FRL aggregation. For example, multiple agents get the global optimal model after FRL training, but the effect after applying it to the local area is worse than that of the model trained by agents only using local data, thus slowing down the speed of model training of agents [15]. This is contrary to the original purpose of designing FRL to speed up model training, so it is urgent to consider solving the heterogeneity problem in FRL. This paper first introduces the theoretical framework and classification of FRL, then generalizes the approach to correspondence heterogeneity, and summarizes its applications and future developments.

2. THEORETICAL BASIS OF FRL

2.1. Introduction to FRL

FRL is a new paradigm in the field of ML. By applying the federated learning framework to RL, multiple reinforcement learning clients can collaborate to train to obtain the global optimal policy, which can not only speed up the policy model training speed and improve the sample efficiency, but also protect the client data security and privacy, and reduce the communication pressure of client interaction [16]. The traditional centralized ML training model needs to collect data from all clients and store them uniformly in the central server device, and the model training is carried out by the central server. However, when centralized ML collects user data, a large amount of data needs to be transmitted over the network, which may lead to data privacy leakage and occupy a large communication bandwidth problem. To solve this problem, the federated learning paradigm was proposed in 2017 and has been widely used in related optimization problems, and the framework is shown in Figure 1.

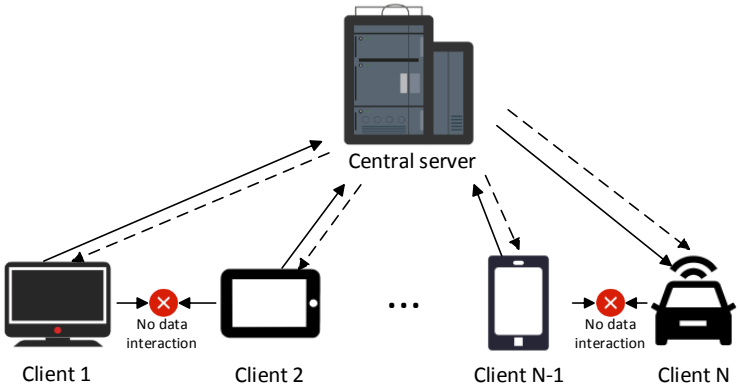

Figure 1. Federated architecture diagram.

In RL, training a model on a single client often faces the challenges of low sample efficiency and slow policy model training. One way to address these challenges is to have clients communicate with each other so that they can learn from each other's experiences. However, due to the increased

awareness of privacy precautions and the constraints of relevant policies and regulations, it is obviously not feasible to collect client data directly. To this end, the FRL paradigm is formed by introducing the federated architecture, so that the agent client trains the policy model locally, and then uploading the trained policy parameters to the central service node. The central service node aggregates the model parameters of all clients to form global model parameters, and sends them to each client. The client continues to train locally based on the global model parameters. This process is iterative, and clients learn from each other in this way to speed up model training.

2.2. FRL Optimization Goals

Traditional RL is usually modeled as a Markov decision process, and FRL is no exception. Information such as state, action and reward in the environment of the agent can also be modeled as a Markov decision process quintuple [17], which is expressed as: $\{S, A, R, P, \gamma\}$, where S denotes the state space set of the environment, A denotes the set of action spaces of an intelligent, R is the reward function of taking action $a, a \in A$ in state $s, s \in S$, P is the probability that an action $a, a \in A$ is taken in state $s, s \in S$ such that the state of the environment is transferred to s'. $\gamma \in (0,1)$ is a constant discount factor. Because RL is a process of constant exploration, the cumulative reward from the current time k to time K can be expressed as:

$$R_k = \gamma^0 r_{k+1} + \gamma^1 r_{k+2} + \gamma^2 r_{k+3} + \cdots + \gamma^{K-1} r_K = \sum_{k=1}^{K} \gamma^{k-1} r_k \tag{1}$$

where the discount factor γ is used to determine the importance of the reward relative to the current reward at different moments. The policy π represents the probability of selecting a particular action given a certain state, guiding the entire exploration process in RL. The cumulative reward of state transition under policy π, that is, the state value function, can be expressed as follows:

$$V^{\pi}(s) = \mathbb{E}[\sum_{k=1}^{K} \gamma^{k-1} r_k | s_k = s] \tag{2}$$

Then the cumulative reward from taking action a in state s according to policy π is called the state-action value function, denoted as:

$$Q^{\pi}(s, a) = \mathbb{E}[R_k | s_k = s_0, a_k = a_0] = \mathbb{E}[\sum_{k=1}^{K} \gamma^{k-1} r_k | s_k = s_0, a_k = a_0] \tag{3}$$

where s_0 and a_0 denote the initial state and action, respectively.

In FRL, because multiple clients are required to participate in training together to obtain a global optimal policy model, the optimization objective of FRL can be expressed by the following formula:

$$\max_{\pi} R(\pi) = \frac{1}{N} \mathbb{E}[\sum_{i=0}^{N} V_i^{\pi}(s_i)] \text{ or } \max_{\pi} R(\pi) = \frac{1}{N} \mathbb{E}[\sum_{i=0}^{N} Q_i^{\pi}(s_i, a_i)] \tag{4}$$

where R represents the global policy reward, N represents the number of agents participating in federated training.

2.3. The Architecture and Properties of FRL

FRL can be classified into Horizontal Federated Reinforcement Learning (HFRL) and Vertical Federated Reinforcement Learning (VFRL) based on the differences in organizational structures [8]. In HFRL, all agents participating in federated training run in parallel, each agent interacts with the environment independently, and trains the policy model locally according to the environment state and feedback information. Each agent encrypts the trained model parameters and uploading them to

the central server, and the central server decrypts and consensus the model parameters of all agents after getting them. After that, the server sends the consensus parameters encrypted to all agents, and the agents update the consensus model parameters locally after receiving them. The basic architecture of HFRL is shown in Figure 2, and the framework can effectively protect user data privacy. At the same time, the agents can draw on the knowledge learned by other agents in different environments through the central server, so as to enrich the agents' own experience and improve the sample efficiency, and the process is conducive to improving the model learning rate.

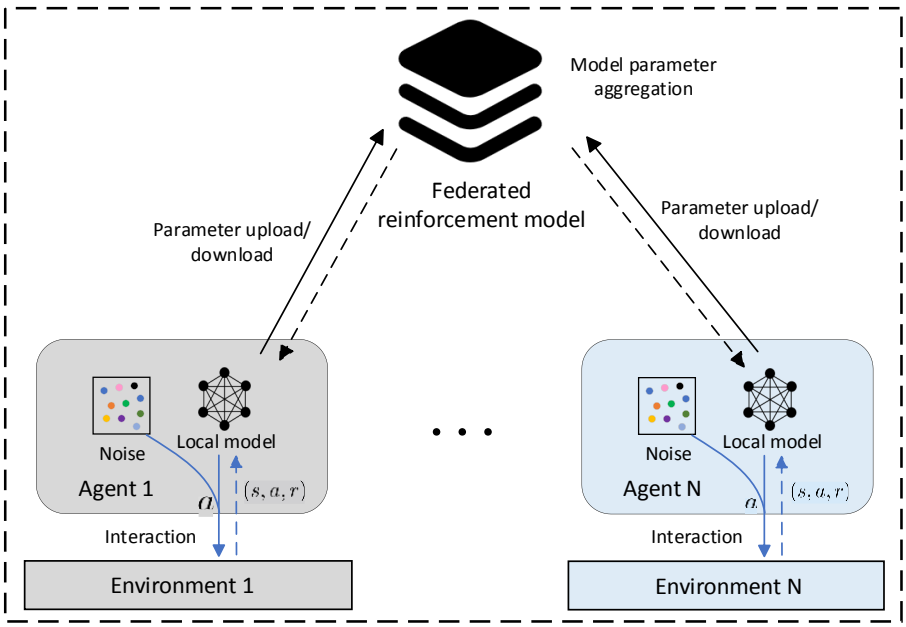

Figure 2. Horizontal federated reinforcement learning architecture graph.

Different from HFRL, VFRL eliminates the central server dedicated to model consensus, but selects an agent from multiple agents participating in the training to maintain an additional Q neural network, which is used to consensus the model parameters of other agents. The agent is called Q network agent, and the others are called cooperative agents. VFRL aims to train a more effective RL agent, and its basic architecture is shown in Figure 3. In this architecture, all agents are located in the same environment, each agent performs different observation interaction on the same environment, and maintains action policy within itself. Agents can make decisions regardless of the state of the environment, but they all need to observe and record the state, reward, and other information of the environment. According to the recorded information, the network parameters are trained in the local neural network, and then the training results are encrypted and sent to the Q network agent. After decrypting the collected training results, the Q network agent trained the Q network and returned the obtained parameter information to the cooperative agents, and each agent continued to update the local network. This method can not only protect data privacy, but also train a more efficient reinforcement learning agent.

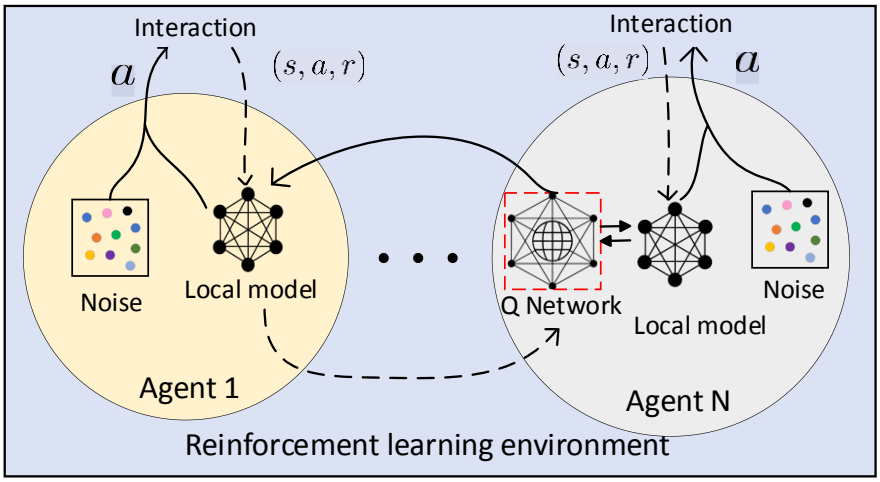

Figure 3. Horizontal federated reinforcement learning architecture graph.

Although FRL has different compositional architectures, it has some common characteristic properties, the advantages of which are that they all have user data protection, distributed computing capabilities, generalizability of their trained models and reduced client communication pressure. However, FRL also faces some challenges, such as increased communication pressure on the central server node, which is due to the need for the central server node to interact with multiple clients. Moreover, the importance of the center node makes it vulnerable to illegal attacks. In addition, there are factors such as heterogeneous data, environment and computational resources in practice, which may have an impact on the performance of FRL model training. Among them, heterogeneity is one of the main challenges facing FRL. It can be seen that solving the above challenges is important for the development of FRL.

3. METHODS AND RELATED RESEARCH TO COPE WITH HETEROGENEITY IN FRL

3.1. Solution Oriented

The personalized method can effectively deal with the heterogeneity problem of FRL. Its main goal is to train multiple different algorithm models according to the heterogeneous reasons such as the environment of local agents while considering the global optimal model of FRL. These models are optimal for specific agents and perform better than ordinary federated global aggregation models. In order for personalized FRL to work in practical applications, the following three issues need to be paid attention to simultaneously [18]. Firstly, how to develop advanced personalized models so that clients can benefit from federated model consensus; The second is how to build an accurate global model so that customers with limited privacy data can benefit from personalization; The third is how to achieve fast model convergence in a small number of training rounds. Therefore, designing personalized algorithms usually needs to be oriented to these three problems.

3.2. Related Personalized Methods Research

As FRL is an emerging field, there are relatively few studies to deal with the heterogeneity problem in FRL. Among them, Nadiger et al. [19] proposed a method that uses FRL to accelerate the training of an agent's personalized model of its environment. For the first time, the FL framework was combined with the Deep Q-Network (DQN) method in RL, and the grouping policy, learning policy and joint policy were proposed. The method of classification aggregation according to data type was used to speed up the model aggregation and improve the accuracy of the model. Wu et al. [20] proposed a Personalized Federated learning framework (DPFed) considering the data differences

between participating clients and the problems of uneven model accuracy and slow convergence speed of existing methods under non-IID data. It used Deep Reinforcement Learning (DRL) method in the central server to identify the relationship between different clients. It makes the clients with similar relationship perform interactive collaborative learning. This method trains a personalized model for each client to speed up the convergence of the model, and reduces the variance between different client models by regularizing the DRL reward function to further improve the quality of the model. The personalized method is shown in Figure 4. Jin et al. [12] studied the problem of multi-agent cooperation in federated reinforcement learning, and this work mainly emphasized the constraint problem of environmental heterogeneity. Firstly, two FRL algorithms based on value function and policy function are proposed, and it is proved that the existence of environmental heterogeneity will lead to the convergence of the two algorithm models to suboptimal solutions. The training of the personalized model is then achieved by embedding the environment information into an additional network vector layer.

4. FRL APPLICATION AREAS

4.1. Edge Computing

With the development and popularization of smart devices, the demand for various edge devices has been further enhanced, such as base stations and roadside units, etc. The application of edge computing equipment can integrate, classify, preprocess and calculate a large amount of data directly at the data acquisition end or the edge part. However, data security and privacy protection are still one of the main challenges faced by edge computing [21], and FRL provides a solution for it. In recent years, FRL applied to edge computing has received attention and research. In order to solve the problems of mobile edge computing, caching and communication, Wang et al. [22] proposed a potential FRL framework, which has lower overhead and higher performance, and makes the mobile communication system have cognitive and environmental adaptability. Zhang et al. [23] optimized the joint solution of cooperative edge caching and proposed a cooperative edge caching method based on dueling deep Q-network to improve caching performance and the curse of dimensionality. Zhu et al. [24] proposed a resource allocation scheme for edge hosts, called Concurrent Federated Reinforcement Learning (CFRL), which not only has the privacy protection and complex problem solving ability of FRL, but also adds concurrency in the form of joint decision making. Wang et al. [25] effectively fused model-based RL and ensemble knowledge distillation into FL to create an ensemble of dynamics models for the client, and then train the policy by using only the ensemble model without interacting with the environment. In order to solve the problem of allocating resources to maintain the efficiency and timeliness of data and tasks, Wang et al. [26] proposed a new framework for UAV-assisted MEC system based on federated multi-agent RL.

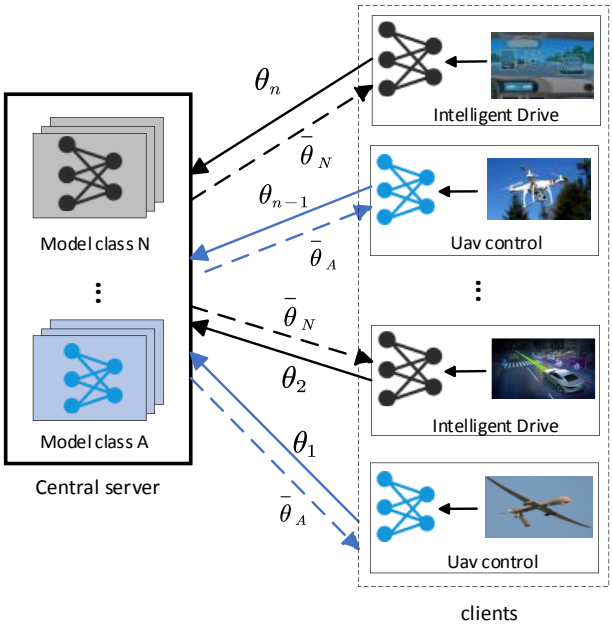

Figure 4. The central server classification aggregation trains the personalized model.

4.2. Control Optimization

RL can solve sequential decision-making problems, so it has significant advantages in complex scenarios such as robot control. However, the optimal training of a single agent is usually limited by factors such as data, exploration and sampling, which will lead to problems such as low efficiency and poor quality of model training. Therefore, the introduction of FRL will be an effective solution. Liu et al. [3] proposed a cooperative learning architecture called LFRL for navigation in cloud robotic systems. Enabling robots to transfer experience to each other so that they can leverage prior knowledge and quickly adapt to changing environments. Liang et al. [27] proposed a Federated Transfer Reinforcement learning (FTRL) framework for knowledge extraction, so that all vehicles can learn from each other while ensuring data privacy. Seongin et al. [28] proposed a new federated learning DDPG algorithm (FLDDPG) based on FL's deep reinforcement learning training strategy, which was applied to the swarm robot system to solve the problem of unstable or limited communication between robots and servers. Experimental results show that the FLDDPG method has higher robustness and generalization ability. Tai et al. [29] used Proximal Policy Optimization (PPO) to achieve the optimal task scheduling policy based on the Deep Reinforcement Learning (DRL) method. Then, a federated learning based algorithm is proposed to improve the performance of PPO agent.

4.3. Communications Network

With the popularization of 5G communication and the emergence of various communication technologies, specific network communication systems can be designed for different scenarios, which leads to the heterogeneity of network systems. Therefore, the management of network system has been widely studied. In order to prevent the traditional management methods from causing data security and inefficiency problems, the introduction of FRL method can be an effective coping method. Among them, Huang et al. [30] studied the Network Function Virtualization (NFV) part of communication network services, and proposed a Scalable Service Function Chain Orchestration (SSCO) scheme by using FRL in NFV supporting networks. They train global models through federated frameworks as well as time-varying local model exploration for scalable SFC orchestration. Open Radio Access Network (O-RAN) has been widely used in 5G networks. Cao et al. [31] proposed a joint DRL-based scheme to train the user access control model in O-RAN in order to solve the

problem of load balancing and handover control. The independent agents make decisions with the help of the global model server, and the server aggregates the DQN parameters of the selected agents to update the global DQN parameters. Krouka et al. [32] considered that in a distributed RL environment, updates of multiple agents may cause interference when communicating with limited bandwidth. This paper proposes A distributed reinforcement learning algorithm based on alternating direction multipliers method (ADMM) and simulated transmission "in-air aggregation" (A-RLADMM). The algorithm enables the agent to transmit each element of its updated model over the same channel using analog communication. Gupta et al [33] to enable resource constrained in-vehicle edge nodes to learn their communication parameters from a central parameter server. The use of FRL accelerates the problem of long training time for convergence of models caused by non-independently and Non-IID data samples in Cellular Vehicle-to-Everything (C-V2X).

5. SUMMARY AND PROSPECT

This paper focuses on an overview and analysis of personalized FRL. First, the developmental context and the challenges of heterogeneity faced by federated reinforcement learning are analyzed. Then, the basic theory and optimization goal of federated reinforcement learning are introduced, and its different classifications are expounded. Subsequently, an overview of the current personalized approaches to the problem of environmental heterogeneity in FRL is given. Finally, the application of FRL in different fields is discussed, which further illustrates the importance of FRL in related fields.

Although FRL has significant advantages for model training and privacy protection, it can be seen that there are few relevant researches on the heterogeneity problem faced by FRL, and further exploration is needed. It is one of the research directions to solve the problem of model training in the continuous high-dimensional state-action space while paying attention to the influence of heterogeneity. Additionally, in policy-based federated reinforcement learning, the challenge of not available of first-order gradients information is noteworthy and merits further research attention.

ACKNOWLEDGEMENTS

This work was supported in part by the Key Technologies R & D Program of Henan Province under Grant No. 222102210049 and 242102211024, in part by the Longmen Laboratory Frontier Exploration Project of Henan Province under Grant No. MQYTSKT035.

REFERENCES

[1] Liu X, Fu X, Du X, et al. Machine learning based non-intrusive digital forensic service for smart homes [J]. IEEE Transactions on Network and Service Management, 2023, 20(2): 945-960.

[2] Kiran B R, Sobh I, Talpaert V, et al. Deep reinforcement learning for autonomous driving: A survey [J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(6): 4909-4926.

[3] Liu B, Wang L, Liu M. Lifelong federated reinforcement learning: A learning architecture for navigation in cloud robotic systems [J]. IEEE Robotics and Automation Letters, 2019, 4(4): 4555-4562.

[4] Zhou Z H, Yu Y, Qian C. Evolutionary learning: Advances in theories and algorithms [M]. Singapore: Springer Singapore, 2019.

[5] Mjolsness E, DeCoste D. Machine learning for science: State of the art and future prospects [J]. Science, 2001, 293(5537): 2051-2055.

[6] McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data [C]. Proceedings of the 20th International Conference on Artificial In-telligence and Statistics, 2017: 1273-1282.

[7] Konečný J, McMahan H B, Ramage D, et al. Federated optimization: Distributed machine learning for on-device intelligence [J]. arXiv preprint arXiv:1610.02527, 2016.

[8] Yang Q, Liu Y, Chen T, et al. Federated machine learning: Concept and applications [J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2019, 10(2): 1-19.

[9] Konečný J, McMahan H B, Yu F X, et al. Federated learning: Strategies for improving com-munication efficiency [J]. arXiv preprint arXiv:1610.05492, 2016.

[10] Sutton R S, Barto A G. Reinforcement learning: An introduction [M]. MIT press, 2018.

[11] Den Hengst F, Grua E M, el Hassouni A, et al. Reinforcement learning for personalization: A systematic literature review [J]. Data Science, 2020, 3(2): 107-147.

[12] Jin H, Peng Y, Yang W, et al. Federated reinforcement learning with environment heterogenei-ty [C]. International Conference on Artificial Intelligence and Statistics, 2022: 18-37.

[13] Lin Q, Ling Q. Byzantine-robust federated deep deterministic policy gradient [C]. IEEE Inter-national Conference on Acoustics, Speech and Signal Processing, 2022: 4013-4017.

[14] Wu Q, He K, Chen X. Personalized federated learning for intelligent IoT applications: A cloud-edge based framework [J]. IEEE Open Journal of the Computer Society, 2020, 1: 35-44.

[15] Kulkarni V, Kulkarni M, Pant A. Survey of personalization techniques for federated learn-ing [C]. IEEE Fourth World Conference on Smart Trends in Systems, Security and Sustainability, 2020: 794-797.

[16] Sheng X, Gao Z, Cui X, et al. Federated reinforcement learning technology and application in edge intelligence scene [C]. International Conference on Emerging Internetworking, Data & Web Technologies. Cham: Springer International Publishing, 2023: 284-291.

[17] Lauer M, Riedmiller M A. An algorithm for distributed reinforcement learning in cooperative multi-agent systems [C]. Proceedings of the 17th International Conference on Machine Learning. 2000: 535-542.

[18] Jiang Y, Konečný J, Rush K, et al. Improving federated learning personalization via model agnostic meta learning [J]. arXiv preprint arXiv:1909.12488, 2019.

[19] Nadiger C, Kumar A, Abdelhak S. Federated reinforcement learning for fast personalization [C]. IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering . 2019: 123-127.

[20] Wu J, Liu X, Liu J, et al. DPFed: Toward fair personalized federated learning with fast convergence [C]. Proceedings of the 18th IEEE International Conference on Mobility, Sensing and Networking, 2022: 510-517.

[21] Lim W Y B, Luong N C, Hoang D T, et al. Federated learning in mobile edge networks: A comprehensive survey [J]. IEEE Communications Surveys & Tutorials, 2020, 22(3): 2031-2063.

[22] Wang X, Han Y, Wang C, et al. In-edge AI: Intelligentizing mobile edge computing, caching and communication by federated learning [J]. IEEE Network, 2019, 33(5): 156-165.

[23] Zhang M, Jiang Y, Zheng F C, et al. Cooperative edge caching via federated deep reinforcement learning in fog-rans [C]. IEEE International Conference on Communications Workshops. 2021: 1-6.

[24] Tianqing Z, Zhou W, Ye D, et al. Resource allocation in IoT edge computing via concurrent federated reinforcement learning [J]. IEEE Internet of Things Journal, 2021, 9(2): 1414-1426.

[25] Wang J, Hu J, Mills J, et al. Federated ensemble model-based reinforcement learning in edge computing [J]. IEEE Transactions on Parallel and Distributed Systems, 2023.

[26] Wang C, Yao T, Fan T, et al. Modeling on resource allocation for age-sensitive mobile edge computing using federated multi-agent reinforcement learning [J]. IEEE Internet of Things Journal, 2023.

[27] Liang X, Liu Y, Chen T, et al. Federated transfer reinforcement learning for autonomous driving [M]. Federated and Transfer Learning. Cham: Springer International Publishing, 2022: 357-371.

[28] Na S, Rouček T, Ulrich J, et al. Federated reinforcement learning for collective navigation of robotic swarms [J]. IEEE Transactions on cognitive and developmental systems, 2023.

[29] Ho T M, Nguyen K K, Cheriet M. Federated deep reinforcement learning for task scheduling in heterogeneous autonomous robotic system [J]. IEEE Transactions on Automation Science and Engineering, 2022.

[30] Huang H, Zeng C, Zhao Y, et al. Scalable orchestration of service function chains in NFV-enabled networks: A federated reinforcement learning approach [J]. IEEE Journal on Selected Areas in Communications, 2021, 39(8): 2558-2571.

[31] Cao Y, Lien S Y, Liang Y C, et al. Federated deep reinforcement learning for user access control in open radio access networks [C]. IEEE International Conference on Communications. 2021: 1-6.

[32] Krouka M, Elgabli A, Issaid C B, et al. Communication-efficient and federated multi-agent reinforcement learning [J]. IEEE Transactions on Cognitive Communications and Networking, 2021, 8(1): 311-320.

[33] Gupta A, Fernando X. Co-operative edge intelligence for C-V2X communication using federated reinforcement learning [C]. IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications. 2023: 1-6.