Article

The Framework of 6G Self-Evolving Networks and the Decision-Making Scheme for Massive IoT

Bei Liu [1] **, Jie Luo** [2] **and Xin Su** [3,*]

1 School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; liubei@tsinghua.edu.cn
2 China Academy of Telecommunications Technology, Beijing 100094, China; nsluojie1016@163.com
3 Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
* Correspondence: suxin@tsinghua.edu.cn

Abstract: The increasingly huge amount of device connections will transform the Internet of Things (IoT) into the massive IoT. The use cases of massive IoT consist of the smart city, digital agriculture, smart traffic, etc., in which the service requirements are different and even constantly changing. To fulfill the different requirements, the networks must be able to automatically adjust the network configuration, architectures, resource allocations, and other network parameters according to the different scenarios to match the different service requirements in massive IoT, which are beyond the abilities of the fifth generation (5G) networks. Moreover, the sixth generation (6G) networks are expected to have endogenous intelligence, which can well support the massive IoT application scenarios. In this paper, we first propose the framework of the 6G self-evolving networks, in which the autonomous decision-making is one of the vital parts. Then, we introduce the autonomous decision-making methods and analyze the characteristics of the different methods and mechanisms for 6G networks. To prove the effectiveness of the proposed framework, we consider one of the typical scenarios of massive IoT and propose an artificial intelligence (AI)-based distributed decision-making algorithm to solve the problem of the offloading policy and the network resource allocation. Simulation results show that the proposed decision-making algorithm with the self-evolving networks can improve the quality of experience (QoE) compared with the lower training.

Keywords: 6G self-evolving networks; decision-making; artificial intelligence; massive IoT

1. Introduction

The amount of connecting devices will reach 50 billion by 2030 [1], and all the devices will generate huge data traffic, which will transform the IoT into the massive IoT. The use cases of massive IoT consist of the smart city, digital agriculture, smart traffic, etc. To fulfill the constantly changing requirements of the services in different massive IoT applications, the networks are expected to be deeply integrated with AI to autonomously adjust the network configuration, architecture, resources, and other parameters according to the different scenarios, which are beyond the abilities of 5G networks. Moreover, 6G networks are expected to have endogenous intelligence, which can well support the massive IoT application scenarios. The autonomous decision-making schemes in 6G networks are expected to support the constantly changing service requirements and meet the requirements of the low latency in massive IoT.

Actually, the 6G system is expected to utilize the full spectrum, face all scenarios, and support all applications [2]. Utilizing the full spectrum means that all of the frequency bands including sub-6G, millimeter wave, terahertz, and visible light communication will be utilized in the 6G system; facing all scenarios means the 6G network will offer coverage in the air, sky, ground, and sea; and supporting full applications means the 6G network will support all the new emerging technologies including AI, machine learning (ML), big data technology, etc. In addition, 6G will offer ultra-high data rates, ultra-high

user experience rates, ultra-low latency, and so on. Compared to eight representative key performance indicators (KPIs) of 5G defined by the International Telecommunication Union (ITU) [3], 6G is expected to satisfy higher requirements. According to [4], the peak data rate of 6G is expected to reach 1 Tbps. The user experience data rate is expected to reach 10–100 Gbps. The energy efficiency (EE) and spectrum efficiency (SE) are expected to respectively reach 10–1000 times and 3–5 times compared to 5G. Furthermore, 6G is expected to provide 0.1 ms over-the-air latency, enable high mobility up to 500 km/h, and support a connection density of up to 10^7 devices/km^2. In addition to the above traditional KPIs, some new KPIs will appear in the 6G era, such as computing power for compute first networking (CFN), security level for deterministic networks [5], and intelligence level for intelligent networking, which are to be defined in follow-up research. For clarity, we give the quantitative comparison of the KPIs between 5G and 6G in Table 1.

Table 1. Quantitative comparison of the KPIs between 5G and 6G.

KPIs	5G	6G
Peak Rate	20 Gbps	≥1 Tbps
User-Experience Rate	0.1 Gbps	≥1 Gbps
Peak Spectrum Efficiency	30 bit/s/Hz	≥60 bit/s/Hz
User-Experience Spectrum Efficiency	0.3 bit/s/Hz	≥3 bit/s/Hz
Connectivity Density	10^6 devices/km^2	≥10^6 devices/km^2
Air-Interface Latency	1 ms	≤0.1 ms
Mobility	500 km/h	≥1000 km/h
Energy Efficiency	1×	10–1000×
Computing Power	-	to be defined
Reliability Level	-	to be defined
Intelligence Level	-	to be defined

In some of the massive IoT applications such as super-smart vehicles and extended reality (XR), the service requirements are constantly changing, and it is of crucial importance to achieve low latency. Traditional methods of manually configuring the network will cause the large delay and extremely high complexity. On the other hand, driven by the rapid development of mobile Internet, big data, super-computing, sensor networks, brain science, and other fields as well as the strong demands of economic and social development, AI technology is also accelerating its development [6]. Moreover, a consensus has been reached by the industry and academia that 6G networks are expected to have endogenous intelligence. In short, the 6G network with endogenous intelligence will be suited for massive IoT.

In fact, network intelligence is not a new concept. In the 4G era, a self-organizing network (SON) was proposed to realize the partial network intelligence [7]. By means of self-configuration, self-healing, and self-optimizing, SON reduced the operator resource expenditure and operating expenditure to a certain extent. Additionally, research on network intelligence has been gradually developed in the 5G era because of the higher and higher service requirements and the increasingly complex networks. The integration of AI and mobile communication wireless technology is one of the main research directions in the field in the 5G system. Various standardization organizations play extremely important roles in the promotion of 5G network intelligence. 3GPP SA2 established the research project "Study of enablers for Network Automation for 5G (eNA)" and enhanced the network data analytic functions (NWDAF) of the 5G core network in May 2017 [8]. In February 2017, the European Telecommunications Standards Institute (ETSI) also formally approved the establishment of a new industry standard group, "Experiential Networked Intelligence (ENI)" [9]. ITU-T established the "Machine Learning for Future Networks including 5G (FG-ML5G)" working group during the SG13 meeting held in November 2017.

Moreover, there are many studies about the integration of AI and 5G in academia, which focus on the intelligent resource allocation [10–16], intelligent network slicing [17–22], intelligent network operation and management [23–28], etc. From the above survey, we

can find that the integration of AI and 5G just focuses on applying AI techniques in the 5G system to enhance some network functions and only achieve partial network intelligence to a certain extent, since the wireless network environment is constantly changing, and the deployed AI components need to be continuously trained and updated, which will hinder the intelligentization of the network. It is expected that AI will be embedded in future networks.

In a word, the 6G system will undergo unprecedented innovations compared to the 5G system. AI abilities are expected to be endogenous in the 6G system. Table 2 gives the differences of AI abilities in 5G and 6G. The 6G network with the endogenous intelligence can realize autonomous sensing, autonomous decision-making, and autonomous control, which can better satisfy the service requirements of the massive IoT scenarios.

Table 2. The differences of AI abilities in 5G and 6G.

	5G	6G
Architecture	Service-based architecture	AI-embedded architecture
Network functions	Software-defined network (SDN) to realize the high efficiency and flexibility	AI-embedded to realize the totally network intelligence
AI abilities	Utilizing AI to enhance some network functions	Realize autonomous sensing, autonomous decision-making, and autonomous control

In terms of intelligent networking for 6G, much research has also been conducted. For example, Huawei proposed the intent-driven network, which was expected to adjust the network configuration based on the prediction of the user intent [29]. S. Wang et al. [30] proposed a distributed and autonomous network architecture for 6G, which was enabled by pervasive distributed intelligence. M. Peng et al. [31] proposed an extreme-intelligent and extreme-concise system architecture of radio access networks to fulfill the requirements of ultra-high data rates and ultra-low latency in 6G networks. T. Zhang et al. [32] further introduced the concept of 6G endogenous intelligence networks which was driven by artificial intelligence and discussed the characteristics and the key technologies of the endogenous intelligence networks. H. Yang et al. [33] presented an AI-enabled intelligent architecture for 6G, which consisted of the following four layers: sensing layer, data mining and analytics layer, control layer, and application layer. Furthermore, the proposed architecture can realize smart resource management, automatic network adjustment and intelligent service provisioning with a high level of intelligence. Considering the typical scenarios in 6G, N. Kato et al. [34] proposed the deep-learning-based path selection to optimize the performance of Space–Air–Ground integrated networks. Y. Xiao et al. proposed the self-learning AI architecture for 6G edge intelligence [35]. Analyzing the above works, we found that the existing research on 6G intelligent networking is not completely free from manual intervention, or considering litter on the self-update of the AI models and the self-evolution of the networks.

However, in some massive IoT scenarios, the number of the active devices and the service requirements are constantly changing, which requires the network to be scalable and evolvable. Thus, we propose the framework of self-evolving networks, which is the closed-loop framework with the abilities of autonomous sensing, autonomous decision-making, and autonomous control, and the AI models can realize the autonomous updates to support the evolution of the networks. Table 3 gives the differences among our work and the other existing works. It can be seen that our proposed framework is more suitable for the massive IoT scenarios.

Table 3. The differences of AI among the existing intelligent networks and our proposed 6G self-evolving networks.

Literature	Architecture	Characteristic
[29]	Intent-driven network	The network is driven by the user intent; in other works, the human involvement is still required.
[30]	Distributed autonomous network	The authors proposed the design principles of 6G networks, which consists of highly distributed, flat, and fully autonomous characteristics, but they did not expound how to realize the native AI.
[31]	Intelligent-concise radio access networks	The authors proposed to integrate AI with 4C (communications, computing, caching, and control).
[32]	Endogenous intelligence network	The authors proposed to induce AI to the all layers of the network.
[33]	AI-enabled intelligent architecture	The authors did not concentrate on the evolution of the AI abilities.
[34]	Space–Air–Ground integrated networks	The authors just concentrated on the path selection of Space–Air–Ground integrated networks.
[35]	Self-learning AI architecture	The authors mainly focused on 6G edge intelligence.
Our work	Self-evolving network	We propose the closed-loop framework, which has the abilities of autonomous sensing, autonomous decision-making, and autonomous control and can realize the self-evolution without the human involvement.

There are many researches on autonomous decision-making schemes. For example, reference [36] discussed the cell association problem in ultra-dense networks and proposed the Q-learning and deep Q networks (DQN)-based intelligent fast cell association algorithms. As for the task offloading and resource allocation problem, Reference [37] proposed a reinforcement learning (RL)-based joint task offloading and migration schemes to raise the total revenue of mobile users. Reference [38] proposed the DQN and deep deterministic policy gradient (DDPG)-based task offloading and resource allocation to reduce the sum cost of tasks. Reference [39] proposed an multi-agent deep reinforcement learning (MADRL)-based joint bit rate selection and radio resource allocation scheme in fog-computing based radio access networks. Reference [40] proposed the DDPG based computing offloading algorithm to reduce the total system delay cost. Additionally, many researchers focus on the autonomous decision-making of the vehicle platoon networks, and reference [41] proposed the DRL-based joint resource management algorithm to improve the communication and computing efficiency in maritime network. Reference [42] proposed the reinforcement learning-based task offloading scheme to optimize the price decision and computing resource allocation in the vehicle platoon networks. Reference [43,44] proposed DDPG algorithm for decision-making in vehicular networks. There are few works focusing on the supporting network and the distributed decision-making schemes for massive IoT scenarios. In this paper, we creatively propose the framework of 6G self-evolving networks which have the abilities of autonomous sensing, autonomous decision-making, and autonomous control and can realize the self-evolution without the human involvement. Then, we propose the distributed dueling double DQN (D3QN)-based decision-making algorithm to obtain the optimal task offloading and resource allocation policy. Compared with Q-learning, DQN and double DQN (DDQN) algorithms, and other centralized algorithms, our method can avoid the overestimation of the Q value, and the distributed decision-making mechanism is more scalable.

In the rest of the paper, we first propose the framework of 6G self-evolving networks and introduce the common decision-making mechanism in Section 2. Then, we propose the distributed decision-making scheme used in massive IoT, including the system model, the algorithm, and the simulation results in Section 3. Section 4 shows the use cases of the

proposed framework of 6G self-evolving networks. Finally, the conclusions and the future direction are given in Section 5.

2. Basic Framework and Preliminaries for Decision-Making

2.1. Edge-Computing-Based Framework

In order to fulfill the different requirements of the constantly emerging new services in massive IoT, 6G networks are expected to be deeply integrated with AI and to be able to autonomously adjust the network configuration, architecture, and other parameters to achieve the best match between the networks and the services. In this part, we propose the edge-computing-based framework 6G self-evolving network for massive IoT in Figure 1. As shown in the figure, the cloud server is deployed to manage the edge devices, and the mobile edge computing (MEC) servers are deployed to realize the data collection, computing, and communication, which can greatly reduce the response latency of the massive IoT devices. Benefiting from the development of sensors, the future IoT devices will evolve to the agents with the abilities of sensing and computation. Moreover, the AI components can be deployed in MEC servers, cloud servers, and the control center to realize autonomous decision-making and autonomous configuration. The self-evolving network includes four stages, which are autonomous sensing, autonomous decision-making, autonomous configuration, and evaluation.
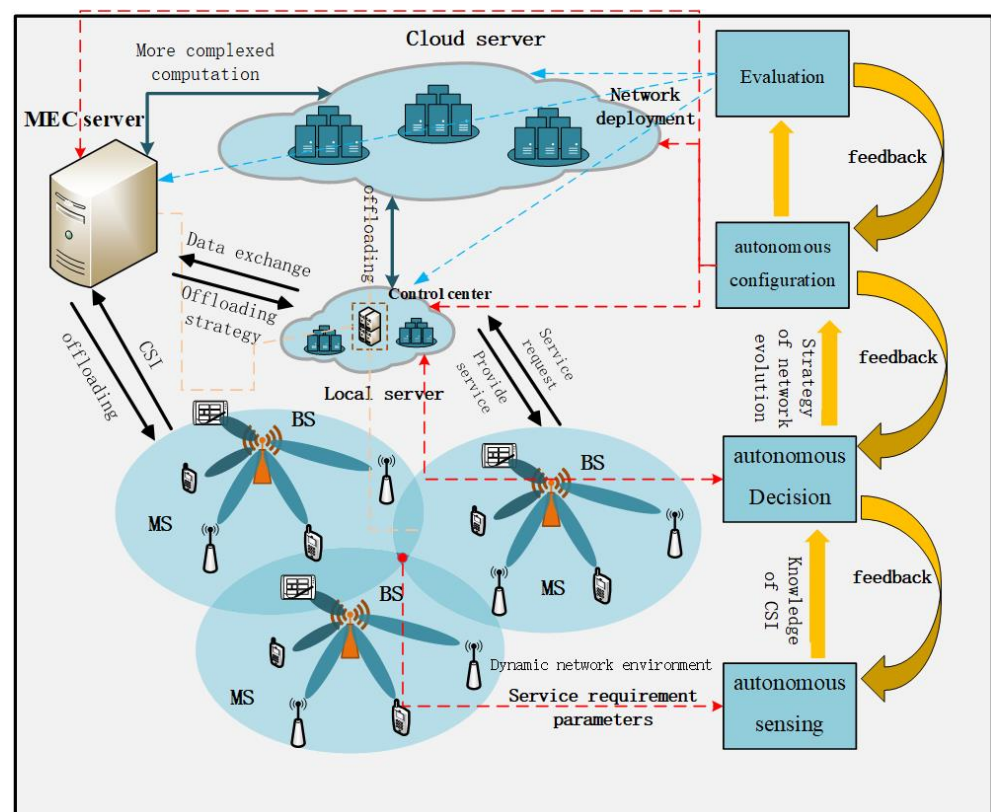

Figure 1. The framework of the 6G self-evolving network.

2.1.1. Autonomous Sensing

In the stage of autonomous sensing, benefiting from the agents, the network can sense the parameters in terms of network environment, the service, etc. The dynamic sensing model based on AI techniques (such as DRL) is designed to achieve the user-centric dynamic sense of the network data and the service requirements. Compared to the traditional sensing schemes, which collect the large data of the environment, network, user, and service at a fixed time and frequency, the AI-based sensing can adjust the sensing sets of the network parameters according to the feedback of the autonomous decision-making

stage and the evaluation stage for different scenarios so as to avoid sensing unnecessary network parameters and further avoid the waste of resources and improve the level and efficiency of sensing intelligence.

2.1.2. Autonomous Decision-Making

In the stage of autonomous decision-making, considering the scalability and flexibility of massive IoT scenarios, some distributed learning models (such as multi-agent reinforcement learning (MARL)) should be designed to realize the dynamic decision-making. First, we should judge whether the network can fulfill the service requirements, decide the direction of the network evolution, and quantify the difference between the target network and the current network. On this basis, the policy set is acquired based on the real-time sensing and analysis of the different network environments and service requirements, and networks can intelligently choose the best policy based on the prediction of the future service requirements. The output of the decision-making is used to perform the network configuration and is also feedback to the network sensing.

2.1.3. Autonomous Configuration

In the stage of autonomous configuration, based on the output of the stage of autonomous decision-making, the network can autonomously configure the network architecture, parameters, resources, etc., to ensure and optimize the user experience.

2.1.4. Evaluation

In the stage of evaluation, the status of network operation and the QoS of the users are evaluated to update the network environment and the ML models, as shown by the blue dotted line in Figure 1. The networks will autonomously evolved with the constantly updating of the ML models updating.

From the above framework, it can be seen that the decision-making scheme is one of the important parts of achieving the self-evolution of the network, and we will focus on the decision-making scheme in this paper.

2.2. Preliminaries for Decision-Making

In this section, we briefly introduce the decision-making methods. The decision-making methods can be roughly divided into two categories. One is traditionally based on numerical optimization, such as the fuzzy decision-making method, the game-theory-based decision-making method, etc. The other is based on AI, which can be divided into the following three categories according to the different training and decision-making mechanisms: centralized training and centralized decision-making, centralized training and distributed decision-making, and distributed training and distributed decision-making.

2.2.1. Centralized Training and Centralized Decision-Making

In the centralized training and centralized decision-making, the central node is set to manage all the nodes. The central node analyzes the whole data and jointly optimizes the policies of all the nodes and then sends the policies to the managed nodes. Therefore, on one hand, the central node needs to collect the information of the other nodes, which may cause the great overhead and processing delay. On the other hand, with the increase in the managed nodes, the computational overhead will become very large and unbearable. In addition, when there are new managed nodes added, the ML model should be updated, which reveals that the centralized training and centralized decision-making scheme is non-scalable.

2.2.2. Centralized Training and Distributed Decision-Making

As for the centralized training and distributed decision-making, the central node also needs to collect the information of the other nodes and complete the training of the AI models; then, it sends the trained AI models to the other nodes. The other nodes can

thus achieve the real-time decision-making with the trained AI models. Compared to the centralized training and centralized decision-making, the centralized training and distributed decision-making can reduce the processing delay once the ML models are trained, which is more suitable for the delay-sensitive massive IoT scenarios.

2.2.3. Distributed Training and Distributed Decision-Making

With regard to the distributed training and distributed decision-making, the central nodes are not needed anymore, and the distributed nodes independently implement the decision-making according to the local information. Therefore, the distributed nodes only need to exchange a small amount of information with other nodes, or even no information exchanging, which greatly reduced the processing delay. On the other hand, the distributed training and distributed decision-making scheme is scalable when there are new nodes added.

For clarity, we succinctly summarize the pros and cons of the centralized and distributed decision-making mechanisms in Table 4.

Table 4. The differences of the centralized and distributed decision-making.

Decision-Making Mechanism	Scalability	Training Overhead	Processing Delay	Output Policy
Centralized training and centralized decision-making	No	High	High	Optimal
Centralized training and distributed decision-making	Yes	High	Low	Suboptimal
Distributed training and distributed decision-making	Yes	Low	Low	Suboptimal

3. The Distributed Task Offloading Scheme for Massive IoT

In some of the massive IoT scenarios, since the number of the devices is huge and there are constantly new terminal advices added, the centralized training and centralized decision-making scheme will cause large overhead. We will concentrate on the distributed decision-making schemes in this section as well as on research in the specific cloud-edge-device scenarios.

3.1. System Model

Considering the cloud-edge-device network framework in Figure 1, we pay attention to the task offloading and the resource allocation problem. We consider the network where one cloud server and N edge servers, denoted as $\mathcal{N} = \{1, 2, \cdots, N\}$, are deployed. K users are active and generate J service requests in the network, which can be respectively denoted as $\mathcal{K} = \{1, 2, \cdots, K\}$ and $\mathcal{J} = \{1, 2, \cdots, J\}$. Assume that one user generates only one service at a certain moment, and let $(D_j(t), F_j(t), T_j(t))$ denote the attributes of the service j at time t, where $D_j(t)$ is the data size, $F_j(t)$ is the size of computing task, and $T_j(t)$ indicates the maximum tolerable delay of the service j.

3.1.1. Transmission Delay

Define $b_{j,m}(t) \in \{0, 1\}, m = \{0, 1, 2, \cdots, N\}$ to indicate the connection relationship between the services and the servers, where $b_{j,m}(t) = 1, m = 0$ indicates that the service j is offloaded to the cloud server at time t, and $b_{j,m}(t) = 1, 1 \leq m \leq N)$ indicates that the service j is offloaded to the mth edge server at time t.

When the service is offloaded to the edge servers, that is $b_{j,m}(t) = 1, 1 \leq m \leq N)$, the channel gain is denoted as $h_{j,m}(t)$, and the transmitting power is $p_{j,m}(t)$. Assuming that

the transmitting channels are orthogonal and the noise obeys the Gaussian distribution with $\mathcal{N}(0, \sigma^2)$, we can obtain the transmitting rate which can be denoted as follows:

$$r_{j,m}(t) = W \log_2 \left(1 + \frac{b_{j,m}(t) p_{j,m}(t) h_{j,m}(t)}{\sigma^2}\right) \tag{1}$$

where W is the allocated bandwidth, which is assumed as the same. Therefore, the transmitting time can be denoted as follows:

$$TR_{j,m}(t) = \frac{D_j(t)}{r_{j,m}(t)} \tag{2}$$

When the service is offloaded to the cloud server, that is $b_{j,m}(t) = 1, m = 0$, we assume that the service is offloaded to the cloud server through the edge server, which can be denoted as

$$TR_{j,m}(t) = \min\left\{\frac{D_j(t)}{r_{j,m'(1 \leq m' \leq N)}(t)}\right\} + Tr \tag{3}$$

where Tr is the needed time transmitting from the edge servers to the cloud servers.

3.1.2. Computation Delay

Assume that the $f_{j,m}(t)$ is the allocated computing resource of the j_{th} service. The computation delay can be denoted as

$$TC_{j,m}(t) = \frac{F_j(t)}{f_{j,m}(t)} \tag{4}$$

Since the edge servers have the limited computing resource, we can obtain the following constraint:

$$\sum_{j=1}^{J} f_{j,m}(t) \leq C_m \tag{5}$$

where $C_m (m = 0, 1, 2 \cdots, N)$ denotes the computation resource of the cloud server and the edge severs.

3.1.3. QoE Model

From [45], the QoE for the user generating the jth service can be modeled as

$$QoE_j(t) = \frac{1}{tM} \sum_{m=0}^{M-1} \sum_{\iota=0}^{t} \left[\alpha r_{j,m}(\iota) - \beta \frac{TC_{j,m}(\iota)}{TR_{j,m}(\iota)}\right] \tag{6}$$

where α and β are the non-negative weights.

3.1.4. Problem Formulation

In the massive IoT scenarios, users make their own decisions according the local information without considering the policies of the other users. Thus, for the jth user, we

aim to maximize the QoE by choosing the different offloading schemes and adjusting the resource allocation, and the problem is formulated as follows:

$$
\max_{b_{j,m}, f_{j,m}} \sum_{j=1}^{J} QoE_j(t)
$$

$$
s.t. C_1 : \sum_{m=0}^{M-1} b_{j,m} \leq 1 \tag{7}
$$

$$
C_2 : \sum_{j=1}^{J} f_{j,m}(t) \leq C_m
$$

$$
C_3 : TC_{j,m}(t) + TR_{j,m}(t) \leq T_j
$$

where C_1 indicates that the service j can be offloaded to only one server, C_2 indicates the constraint of the computing server of the servers, and C_3 indicates that the total delay should not exceed the maximum tolerable delay of the service j.

3.2. The Distributed DQN-Based Algorithm

The above problem is non-convex since $b_{j,m}$ are binary variables. We then transform the problem into the Markov decision process (MDP), which is denoted by $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R})$.

3.2.1. State

The state set $\mathcal{S}$ describes the computation load and the remaining resource of the servers. At time t, the computation load of the server m can be denoted as $L_m(t) = \sum_{j=1}^{j} b_{j,m}(t)$, the remaining resource is $F_m(t) = C_m - \sum_{j=1}^{j} f_{j,m}(t)$, and the state at time t can be denoted as

$$
s(t) = \{L_0(t), L_1(t), \cdots, L_N(t), F_0(t), F_1(t), \cdots, F_N(t)\} \tag{8}
$$

3.2.2. Action

According to the current policy, the agent chooses an action from the action set. The action of the user generating the jth service, denoted as $a_j(t) \in \mathcal{A}_j$, can be described as

$$
a_j(t) = \{b_{j,0}(t), b_{j,1}(t), \cdots, b_{j,N}(t), f_{j,0}(t), f_{j,1}(t), \cdots, f_{j,N}(t)\} \tag{9}
$$

3.2.3. State Transition Probability

The state is transformed to the next state by taking the action, and we can use the state transition probability to describe the process, which can be denoted as follows:

$$
P_{ss'}(\mathbf{a}) = P[\mathcal{S}_{t+1} = s | \mathcal{S}_t = s', \mathcal{A}_t = \mathbf{a}] \tag{10}
$$

where $\mathbf{a} = \{a_1, a_2, \cdots, a_J\}$ denotes the action set of all the users generating service.

3.2.4. Reward

The reward of the user generating the jth service can be described as QoE utility minus the action-selection cost, denoted as $\phi_j(t)$, which can be denoted as

$$
R_j(t) = QoE_j(t) - \phi_j(t) \tag{11}
$$

and the long term reward can be formulated as follows:

$$
\mathcal{R}_j(s, \pi_j, \pi_{-j}) = \sum_{t=0}^{T-1} \gamma^t R_i(s(t), \pi_j(t), \pi_{-j}(t) | s(0)) \tag{12}
$$

where $\pi_j(t)$ is the policy of the user generating the jth service at time t, and $\pi_{-j}(t)$ denotes the policies of the other users. $\gamma \in [0,1]$ is the discount rate to determine the weight of the future reward.

Notice that the above problem is the game process, and there exists a Nash equilibrium (NE) policy a_j^* for the user j, that is

$$\mathcal{R}_j(s, a_j^*, a_{-j}^*) \geq \mathcal{R}_j(s, a_k, a_{-k}), (k \neq j) \tag{13}$$

We first use the DQN to get the NE policy, and the target value can be denoted as

$$y_j^{DQN} = R_j + \gamma \hat{Q}_j(s', a; \theta^-) \tag{14}$$

In DQN, the neural network (NN) is used to approximate the Q-value function, and the loss function can be denoted as

$$L(\theta) = E[y_j^{DQN} - Q(s, a; \theta)]^2 \tag{15}$$

Then, to solve the problem of overestimation in typical DQN, we used DDQN [46] to solve the game problem. The target can be denoted as

$$y_j^{DDQN} = R_j + \gamma \hat{Q}_j(s', \arg\max_{a_j \in \mathcal{A}_j} Q_i(s', a; \theta); \theta^-) \tag{16}$$

Finally, we further explore the utilization of D3QN [47] to solve the game problem. Different from DQN and DDQN, the output of D3QN network consists of theses two parts: state value $V(s)$ and advantage value $A(s, a)$, and the target Q value can be denoted as

$$Q(s, a) = A(s, a) + V(s) \tag{17}$$

where $V(s) = E[R_j + \gamma V(s_{t+1})]$, $A(s, a) = A(s, a) - \frac{1}{|A|} \sum A(s, a')$, and $\frac{1}{|A|}$ denotes the advantages of the entire action space. The training stage is completed in the centralized node and is described in Algorithm 1.

**Algorithm 1** D3QN-based task offloading algorithm.

Require:
 Input the training steps T, and the episode number T'
 Initialize the experience replay memory $\mathcal{D}$
 Initialize the main network $Q(s, a; \theta)$ and the target network $\hat{Q}(s, a, \theta^-)$, and set $\theta^- \leftarrow \theta$
 Initialize the state s_0
1: **while** *episode* $\leq T'$ **do**
2: Observe the current state s_t
3: **while** *step* $t \leq T$ **do**
4:   Select the action **a**. based on the ϵ-greedy policy
5:   Obtain the current immediate reward $\mathbf{R}(t)$
6:   Renew the state $s \rightarrow s'$
7:   Store the transition tuple $(s, \mathbf{a}, \mathbf{R}(t), s')$ in $\mathcal{D}$
8:   Sample a mini-batch from $\mathcal{D}$ randomly
9:   Calculate the target Q value based on Equation (17)
10:   Calculate the loss based on Equation (15) and update θ by performing gradient descent to minimize the loss function
11:   Every T_0 steps, update the target network $\hat{Q}$
12: **end while**
13: **end while**
 Output the trained DQN and the optimal action **a**.


3.3. Simulation Results

This part shows the simulation results. We consider a square area of length 500 m. The ϵ-greedy policy is used when choosing the action, and we set $\epsilon \in [0, 0.9]$. The learning rate is set as the experience value 0.01, and the discount rate is set as $\gamma = 0.99$. The non-negative weights α and β in Equation (6) are set as the experience values which are respectively 0.5 and 0.0005. The action-selection cost ϕ_j in Equation (11) is set as 0.001. The size of the experience replay memory is 500.

Figure 2 gives the comparison of convergence performance among different algorithms. We can see that D3QN converges faster than other algorithms, which shows that D3QN has the lowest training complexity, and D3QN is more suitable for massive IoT scenarios.
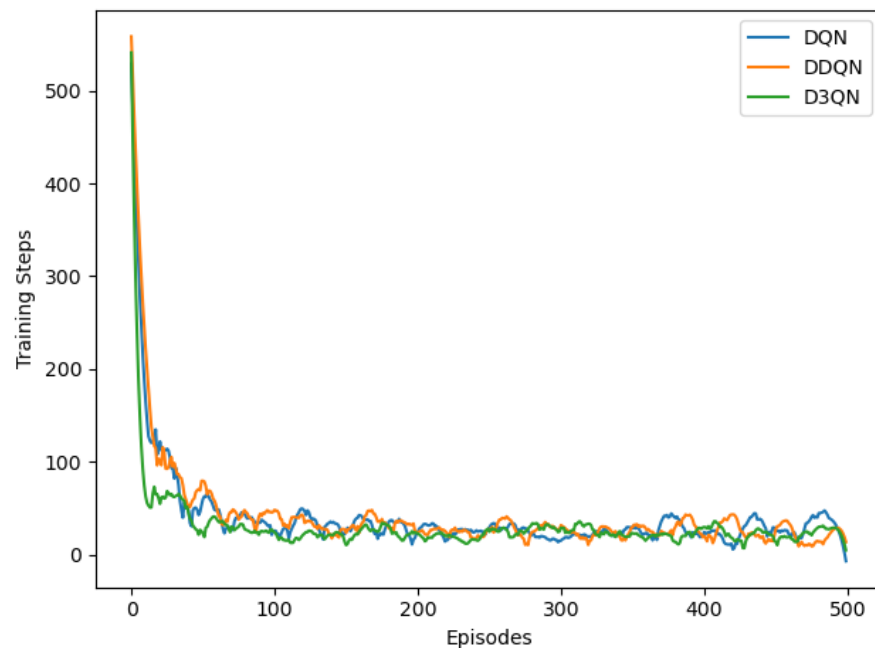
Figure 2. The comparison of convergence performance among different algorithms.

Figure 3 gives the QoE utility of the different algorithms. The horizontal axis represents the number of users who are active and have the service requests at a certain moment. On the other hand, Q-learning is the the most basic reinforcement learning algorithm, and our D3QN-based offloading and resource allocation scheme is evolved from the DQN algorithm, so we simply use the Q-learning algorithm and the DQN algorithm as the comparison methods, the same as reference [36]. From Figure 3, we can see that our proposed D3QN-based scheme can achieve the better performance in terms of the QoE utility, especially when the number of users are relatively large. The main reasons for this result are analyzed as follows: on one hand, the value of the resource allocation is continuous, and Q-learning needs do some discretization processing when meeting non-discrete action, which may cause some bias in decision-making; on the other hand, because of the overestimation of the Q valuein DQN, the output policy of DQN may not be the optimal policy.
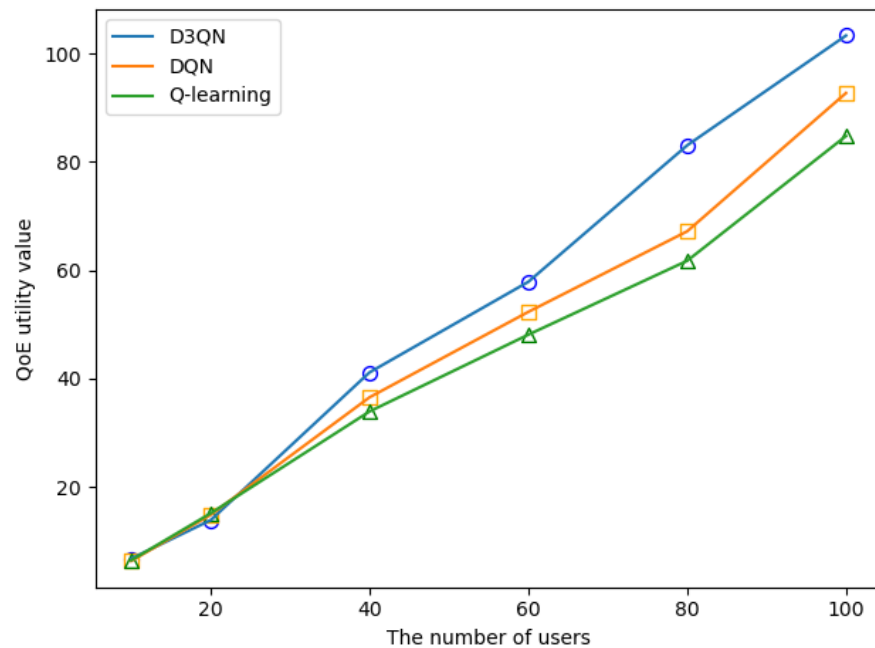
Figure 3. The comparison of QoE performance among different algorithms.

From Figures 2 and 3, we can conclude that our proposed algorithm on the basis of 6G self-evolving networks can achieve better performance. Because of the limited computing power of our personal computer, the number of the users is set as much smaller than the the actual massive IoT scenarios, but we can obtain the same conclusion as above because the performance of the total QoE utility is not affected by the number of users. Moreover, in the actual execution of the proposed decision-making scheme, the centralized training is completed in the control center, and the terminal devices can realize the real-time decision-making once the training is complete.

4. Use Cases

The proposed edge-based self-evolution framework and the autonomous decision-making algorithm can be well applied in many massive IoT scenarios. In this section, we will illustrate the application in the super-smart vehicle as an example.

Super-smart vehicle is one of the typical application scenarios of massive IoT in the 6G era, which is the upgrade of the current autonomous vehicles. Compared to the self-driving cars in the 5G era, super-smart vehicles will be more intelligent, which is reflected in the following aspects: (1) more diverse means of transportation will be used; (2) it is expected to realize point-to-point smart travel. On the other hand, the devices are always moving and the network environment is constantly changing, so it is necessary to embed the AI functions in 6G networks to realize the whole network intelligence.

Since the devices in super-smart vehicles have the features of high mobility and flexibility and the requirements for super-smart vehicles are constantly changing, it is necessary for the network to offer a flexible framework to support super-smart vehicles. By deploying AI components on the terminal devices and in the network, the super-smart vehicle systems have the abilities to autonomously sense the network environment (the states of cars, roads, people, etc.); to collect and process the traffic information, vehicle information and environmental information; and to analyze and predict traffic conditions, so that they can autonomously and quickly make decisions and control the transportation, which are fully supported by the proposed 6G self-evolving network. According to [48], the current network intelligence level is at L2–L3, and in future work, we will continue to discuss the application of the proposed 6G self-evolving network in super-smart vehicles

and research the key technologies to realize the network intelligence level at L3–L4, and we expect to give the corresponding simulation results to prove it.

5. Conclusions and Future Work

In order to realize the future intelligent wireless network to well satisfy the service requirements of the massive IoT application scenarios, it is necessary to embed the essential capabilities of AI into the wireless system. In this paper, we first propose the edge-computing-based self-evolution framework for 6G massive IoT, which is expected to realize autonomous sensing, autonomous decision-making, and autonomous configuration. Autonomous decision-making is also one of the crucial parts in the proposed 6G self-evolution framework, so we introduce the decision-making mechanism and analyze the pros and cons of the centralized decision-making schemes and the distributed decision-making schemes. Then, we consider the task offloading problem and propose the distributed D3QN-based algorithm. The simulation results show that the proposed D3QN-based scheme can converge faster than the DQN- and DDQN-based algorithm and can acquire the better performance in terms of QoE of users compared with Q-learning and DQN algorithms.

Finally, there is still much research to be done in the future. For example, this paper concentrated on centralized training and distributed decision-making. As the network and devices for massive IoT are becoming increasingly dense, the complexity of centralized training will get higher and higher. Therefore, the distributed training and distributed decision-making schemes are being urgently researched, which is also part of our future work.

Author Contributions: The work presented in this paper corresponds to a collaborative development by all authors. X.S. defined the research line, B.L. developed the proposed algorithm and wrote the paper, and J.L. used the software to simulate the algorithm and analyze the simulation results. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Piran, M.J.; Suh, D.Y. Learning-driven wireless communications, towards 6g. In Proceedings of the 2019 International Conference on Computing, Electronics Communications Engineering (iCCECE), London, UK, 22–23 August 2019; pp. 219–224.
2. Sun, Y.; Liu, J.; Wang, J.; Cao, Y.; Kato, N. When Machine Learning Meets Privacy in 6G: A Survey. *IEEE Commun. Surv. Tutor.* **2020**, *22*, 2694–2724. [CrossRef]
3. ITU-R. *IMT Vision, Framework and Overall Odjectives of the Future Development of IMT for 2020 and Beyond*; M 2083-0; ITU-R: Geneva, Switzerland, 2015.
4. Liu, G.; Huang, Y.; Li, N.; Dong, J.; Jin, J.; Wang, Q.; Li, N. Vision, requirements and network architecture of 6G mobile network beyond 2030. *China Commun.* **2020**, *17*, 92–104. [CrossRef]
5. Tang, X.; Li, F.; Zhang, Z.; Ma, J.; Liu, Q. Requirements, Architectures and Technology Trends of 6G Network. *Mob. Commun.* **2021**, *45*, 37–44.
6. China Unicom. *China Unicom Network Artificial Intelligence White Paper*; China Unicom: Hong Kong, China, 2019.
7. Gacanin, H.; Wagner, M. Artificial Intelligence Paradigm for Customer Experience Management in Next-Generation Networks: Challenges and Perspectives. *IEEE Netw.* **2019**, *33*, 188–194. [CrossRef]
8. Yao, H.; Jiang, C.; Qian, Y. *Developing Networks Using Artificial Intelligence*; Springer: Cham, Switzerland, 2019.
9. *ENI: Experiential Networked Intelligence*; ETSI GS ENI 006-2018; ETSI: Zapopan, Mexico, 2019.
10. Zhang, C.; Dong, M.; Ota, K. Fine-Grained Management in 5G: DQL Based Intelligent Resource Allocation for Network Function Virtualization in C-RAN. *IEEE Trans. Cogn. Commun. Netw.* **2020**, *6*, 428–435. [CrossRef]

11. Zhou, Y.; Fadlullah, Z.M.; Mao, B.; Kato, N. A deep-learning-based radio resource assignment technique for 5G ultra dense networks. *IEEE Netw.* **2018**, *32*, 28–34. [CrossRef]

12. Hoang, T.D.; Le, L.B.; Le-Ngoc, T. Radio resource management for optimizing energy efficiency of D2D communications in cellular networks. In Proceedings of the 2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Hong Kong, China, 30 August–2 September 2015; pp. 1190–1194.

13. Wang, J.; Huang, Y.; Jin, S.; Schober, R.; You, X.; Zhao, C. Resource management for device-to-device communication: A physical layer security perspective. *IEEE J. Sel. Areas Commun.* **2018**, *36*, 946–960. [CrossRef]

14. Liu, X.; Ansari, N. Resource allocation in uav-assisted M2M communications for disaster rescue. *IEEE Wirel. Commun. Lett.* **2019**, *8*, 580–583. [CrossRef]

15. Zhang, R.; Qi, C.; Li, Y.; Ruan, Y.; Wang, C.; Zhang, H. Towards energy-efficient underlaid device-to-device communications: A joint resource management approach. *IEEE Access* **2019**, *7*, 31385–31396. [CrossRef]

16. Bikov, E.; Botvich, D. Smart concurrent learning scheme for 5G network: QoS-aware radio resource allocation. In Proceedings of the 2017 IVth International Conference on Engineering and Telecommunication (EnT), Moscow, Russia, 29–30 November 2017; pp. 99–103.

17. Yan, M.; Feng, G.; Zhou, J.; Sun, Y.; Liang, Y.-C. Intelligent Resource Scheduling for 5G Radio Access Network Slicing. *IEEE Trans. Veh. Technol.* **2019**, *68*, 7691–7703. [CrossRef]

18. Jiang, W.; Anton, S.D.; Schotten, H.D. Intelligence Slicing: A Unified Framework to Integrate Artificial Intelligence into 5G Networks. In Proceedings of the 2019 12th IFIP Wireless and Mobile Networking Conference (WMNC), Paris, France, 11–13 September 2019; pp. 227–232.

19. Bega, D.; Gramaglia, M.; Garcia-Saavedra, A.; Fiore, M.; Banchs, A.; Costa-Perez, X. Network Slicing Meets Artificial Intelligence: An AI-Based Framework for Slice Management. *IEEE Commun. Mag.* **2020**, *58*, 32–38. [CrossRef]

20. Chergui, H.; Verikoukis, C. Big Data for 5G Intelligent Network Slicing Management. *IEEE Netw.* **2020**, *34*, 56–61. [CrossRef]

21. Chekired, D.A.; Togou, M.A.; Khoukhi, L.; Ksentini, A. 5G-Slicing-Enabled Scalable SDN Core Network: Toward an Ultra-Low Latency of Autonomous Driving Service. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 1769–1782. [CrossRef]

22. Aklamanu, F.; Randriamasy, S.; Renault, E. Demo: Intent-Based 5G IoT Application Network Slice Deployment. In Proceedings of the 2019 10th International Conference on Networks of the Future (NoF), Rome, Italy, 1–3 October 2019; pp. 141–143.

23. Jiang, Z.; Fu, S.; Zhou, S.; Niu, Z.; Zhang, S.; Xu, S. AI-Assisted Low Information Latency Wireless Networking. *IEEE Wirel. Commun.* **2020**, *27*, 108–115. [CrossRef]

24. Han, Y.; Li, J.; Hoang, D.; Yoo, J.; Hong, J.W. An intent-based network virtualization platform for SDN. In Proceedings of the 2016 12th International Conference on Network and Service Management (CNSM), Montreal, QC, Canada, 31 October–4 November 2016; pp. 353–358.

25. Comer, D.; Rastegatnia, A. OSDF: An Intent-based Software Defined Network Programming Framework. In Proceedings of the 2018 IEEE 43rd Conference on Local Computer Networks (LCN), Chicago, IL, USA, 1–4 October 2018; pp. 527–535.

26. Rafiq, A.; Afaq, M.; Song, W.C. Intent-based networking with proactive load distribution in data center using ibn manager and smart path manager. *J. Ambient Intell. Humaniz. Comput.* **2020**, *11*, 4855–4872. [CrossRef]

27. Jiang, W.; Strufe, M.; Schotten, H.D. Intelligent network management for 5G systems: The SELFNET approach. In Proceedings of the 2017 European Conference on Networks and Communications (EuCNC), Oulu, Finland, 12–15 June 2017; pp. 1–5.

28. Imran, A.; Zoha, A.; Abu-Dayya, A. Challenges in 5G: How to empower SON with big data for enabling 5G. *IEEE Netw.* **2014**, *28*, 27–33. [CrossRef]

29. Huawei. IDN Maximize Your Business Value. Available online: https://developer.huawei.com/ict/en/site-idn (accessed on 9 October 2018).

30. Wang, S.; Sun, T.; Yang, H.; Duan, X.; Lu, L. 6G Network: Towards a Distributed and Autonomous System. In Proceedings of the 2020 2nd 6G Wireless Summit (6G SUMMIT), Levi, Finland, 17–20 March 2020; pp. 1–5.

31. Peng, M.; Sun, Y.; Wang, W. Intelligent-Concise radio Access Networks in 6G: Architecture, Techniques and Insight. *J. Beijing Univ. Posts Telecommun.* **2020**, *43*, 1.

32. Zhang, T.; Ren, Y.; Yan, S.; Peng, M. Artificial intelligence driven 6G networks: Endogenous intelligence. *Telecommun. Sci.* **2020**, *36*, 14–22.

33. Yang, H.; Alphones, A.; Xiong, Z.; Niyato, D.; Zhao, J.; Wu, K. Artificial-Intelligence-Enabled Intelligent 6G Networks. *IEEE Netw.* **2020**, *34*, 272–280. [CrossRef]

34. Kato, N.; Fadlullah, Z.M.; Tang, F.; Mao, B.; Tani, S.; Okamura, A.; Liu, J. Optimizing Space-Air-Ground Integrated Networks by Artificial Intelligence. *IEEE Wirel. Commun.* **2019**, *26*, 140–147. [CrossRef]

35. Xiao, Y.; Shi, G.; Li, Y.; Saad, W.; Poor, H.V. Toward Self-Learning Edge Intelligence in 6G. *IEEE Commun. Mag.* **2020**, *58*, 34–40. [CrossRef]

36. Pan, J.; Wang, L.; Lin, H.; Zha, Z.; Kai, C. Intelligent fast cell association scheme based on deep Q-learning in ultra-dense cellular networks. *China Commun.* **2021**, *18*, 259–270. [CrossRef]

37. Wang, D.; Tian, X.; Cui, H.; Liu, Z. Reinforcement learning-based joint task offloading and migration schemes optimization in mobility-aware MEC network. *China Commun.* **2020**, *17*, 31–44. [CrossRef]

38. Liang, Y.; He, Y.; Zhong, X. Decentralized Computation Offloading and Resource Allocation in MEC by Deep Reinforcement Learning. In Proceedings of the 2020 IEEE/CIC International Conference on Communications in China (ICCC), Xiamen, China, 28–30 July 2020; pp. 244–249.

39. Chen, J.; Wei, Z.; Li, S.; Cao, B. Artificial Intelligence Aided Joint Bit Rate Selection and Radio Resource Allocation for Adaptive Video Streaming over F-RANs. *IEEE Wirel. Commun.* **2020**, *27*, 36–43. [CrossRef]

40. Chen, X.; Ge, H.; Liu, L.; Li, S.; Han, J.; Gong, H. Computing Offloading Decision Based on DDPG Algorithm in Mobile Edge Computing. In Proceedings of the 2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), Chengdu, China, 24–26 April 2021.

41. Xu, F.; Yang, F.; Zhao, C.; Wu, S. Deep reinforcement learning based joint edge resource management in maritime network. *China Commun.* **2020**, *17*, 211–222. [CrossRef]

42. Ma, X.; Zhao, J.; Li, Q.; Gong, Y. Reinforcement Learning Based Task Offloading and Take-Back in Vehicle Platoon Networks. In Proceedings of the 2019 IEEE International Conference on Communications Workshops (ICC Workshops), Shanghai, China, 22–24 May 2019; pp. 1–6.

43. Peng, H.; Shen, X.S. DDPG-based Resource Management for MEC/UAV-Assisted Vehicular Networks. In Proceedings of the 2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall), Victoria, BC, Canada, 4–7 October 2020.

44. Xu, Y.; Yang, C.; Hua, M.; Zhou, W. Deep Deterministic Policy Gradient (DDPG)-Based Resource Allocation Scheme for NOMA Vehicular Communications. *IEEE Access* **2020**, *8*, 18797–18807. [CrossRef]

45. Tao, X.; Jiang, C.; Liu, J.; Xiao, A.; Qian, Y.; Lu, J. QoE driven resource allocation in next generation wireless networks. *IEEE Wirel. Commun.* **2019**, *26*, 78–85. [CrossRef]

46. Hasselt, H.V.; Guez, A.; Silver, D. Deep Reinforcement Learning with Double Q-learning. In Proceedings of the Thirtieth Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016.

47. Wang, Z.; Guez, A.; Silver, D. Dueling Network Architectures for Deep Reinforcement Learning. In Proceedings of the 33rd International Conference on Machine Learning (ICML 2016), New York, NY, USA, 19–24 June 2016.

48. Artificial Intelligent Industry Alliance. *Artificial Intelligence Applications in the Telecommunications Industry*; White Paper; Artificial Intelligent Industry Alliance: Beijing, China, 2021.