A Deep Reinforcement Learning based Mechanism for Cell Outage Compensation in 5G UDN

Xiao Yang, Peng Yu, Lei Feng, Fanqin Zhou, Wenjing Li, and Xuesong Qiu
State Key Laboratory of Networking and Switching Technology
Beijing University of Posts and Telecommunications
Beijing 100876, China

***Abstract*—Ultra Dense Networks (UDN) have become one of the key technologies for 5G wireless communications, which can meet the requirements of high-traffic, high-density wireless terminal access. Compared with LTE, there are a large number of heterogeneous cells in UDN. If a failure occurs and can't be alleviated its effect in time, it will lead to a significant drop in network performance. Therefore, the cell outage compensation (COC) problem in the UDN is very important. Although deep reinforcement learning (DRL) has been applied to many scenarios related to the self-organizing network (SON), there are fewer applications for cell outage compensation. In this paper, aiming at the cell outage scenario in the UDN with the goal of maximizing the sum of the throughput of all users while meeting service quality demands of each mobile user, and present a framework based on DRL to solve it. Specifically, we first allocate compensation users to adjacent BSs by using the K-means clustering algorithm, then apply a deep neural network (DNN) to approximate action-value function. The simulation results show that the algorithm converges quickly and tends to be stable, and reach 99.53% of the maximum throughput. It verifies the efficiency of the DRL-based framework and its effectiveness in meeting the requirement of user rates and handling cell interrupt compensation.**

Keywords—Deep Reinforcement Learning, Cell Outage Compensation, Ultra Dense Networks

I. Introduction

With the continuous development of wireless communication technology, as one of the key technologies for the next generation (5G) wireless communications, Ultra Dense Networks (UDN) can effectively increase system capacity and network coverage by increasing the deployment density of low-power base stations (BS) indoors and hotspots. Besides, it can reduce the delay and meet the requirements of high-traffic, high-density wireless terminal access. In UDN, the macro and small BSs constitute multiple layers, the macro BSs are responsible for the basic coverage, and the low-power small BSs are densely deployed in the indoor and hotspot areas under the coverage of the macro BS. Compared with the mature 4G network, there are a large number of heterogeneous cells in UDN. Besides, its structure is more complex, and it has a large number of deployed nodes. Therefore, its management is very complicated, which means it faces many challenges [1]. The method to maximize the overall performance of the network is to incorporate intelligence and self-adaptation into a self-organizing network (SON). The main goal of SON is to improve service quality and reduce the costs associated with network operations by reducing human involvement while enhancing network performance. Its functions include self-configuration (plug and play network elements), self-optimization (automatically optimize network elements and parameters), and self-healing (automatically detect and mitigate failures)[2]. Wireless networks are prone to errors and failures, and the most critical domain for fault management is the Radio Access Network (RAN). Each base station (BS) is responsible for serving an area with little or none redundancy. If a node is unable to perform its duties, it will result in performance degradation for a period of time, during which users can not receive proper service. This results in serious revenue loss for the operator [3]. Therefore, the self-healing function in 5G is very important [4]. It includes two stages: cell outage detection (COD) and cell outage compensation (COC). COD detects and classifies failures while minimizing detection time. COC executes actions to mitigate or at least alleviate the effect of the failure [2]. In this article, we focus on COC in UDN assuming COD is completed.

Reinforcement learning (RL) has been applied to many works related to self-organizing networks (SON), such as coverage and capacity optimization (CCO), interference management, load balancing, and hand-over management [5]. According to the reward obtained by interacting with the environment, the RL agent can generate optimal control actions. Instead of simply optimizing the current reward in a greedy manner, the RL agent takes a long-term goal into account, which is the cumulative reward. The emerging Deep Reinforcement Learning (DRL) can be considered as an enhanced version of traditional RL, which provides a better solution to handle complicated problems [6]. In traditional Q-learning (QL), the cumulative reward Q is stored in the Q table, and its rows represent states and columns represent actions. If the state and action spaces are very large, the Q table is too large to be applied to reality. Deep Q learning (DQN) improved QL in three aspects:

1) Approximating value function Q instead of Q table by using the deep Convolutional Neural Network (CNN).

2) Training the learning phase of RL by using the experience replay, which breaks the correlation between data and makes the training of neural network converge quickly and stable.

3) Updating the network parameters with the gradient descent method, and set the target network separately to deal

978-3-903176-15-7 ©2019 IFIP

with the TD deviation in the time difference algorithm, which avoids training instability.

The current research on DRL has made significant achievements in many fields, including application to SON. However, there are fewer works on COC.

In this paper, we present a DRL-based framework for cell outage compensation in UDN, which maximizes the sum of the throughput of all users while ensuring that the demand of each mobile user is satisfied [7]. Our main contribution is presenting a DRL-based framework for COC in UDN for the first time. With the aim of maximizing the sum of the throughput of all users while meeting service quality demands of each mobile user, we define the state space, action space and reward function for the DRL agent, and apply a DNN to approximate the action-value function for action decisions, which directly extracts information from the original state and avoids using handmade features.

The rest of the paper is organized as follows: We summarize the related work in Section II. In Section III, we build the system model and model the target and constraints to solve the COC problem in UDN. In Section IV, we propose a DRL-based COC framework. The simulation result in Section V shows the effectiveness of this proposed algorithm. Finally, we summarize this paper in Section VI.

II. Related Work

There are many works for COC problem in the 4G scenario. In paper [5], the authors propose a generalized Q-learning framework for cognitive cellular networks functions and apply this framework to two functions on mobility robustness optimization and mobility load balancing. However, the application scenario of this algorithm is simple, and the state and action spaces are limited. The authors of [8] propose a distributed COC management mechanism based on reference signal power adjustment. But, this algorithm considers just one parameter which is RS power, and the application scenario is LTE which is not applicable to complex scenarios. In paper [9], a coalition game based resource allocation algorithm is proposed. Small cells play coalition games to form a set of coalitions, and each coalition of small cells serves a user cooperatively with optimized power allocation. This algorithm improves cell capacity and user fairness but is applied to the simple small cell network, not suitable for complex scenarios. In paper [3], the authors present an automatic and self-organized RL-based approach for COC and select Actor-Critic to adjust automatically downlink transmission power levels and antenna tilt value in order to fill the coverage and capacity gap. This method covers more users. However, the object to be adjusted by this method is just the base station, and it does not consider each user in detail so that the services provided to compensation users are limited. Besides, the application scenario of this method is LTE, which is not applicable to complex 5G scenarios. The above works are applied to 4G scenarios and not applicable to complex scenarios. In paper [2], the authors propose a new COC method, which adds a new self-healing radio to each base station in the cloud-radio access network (C-RAN). These SHRs use the available resources of adjacent base stations only in case of fronthaul/backhaul failure of any base station in the network and do not change the antenna tilt or power of adjacent base stations. This algorithm works for complex scenes, but the fronthauling rate is lower. In paper [10], the authors proposed an enhanced immune-genetic algorithm to minimize public safety network (PSN) system outage probability, which compensates the C-RAN-based PSN using both cooperative transmission and power adjustment. This algorithm works for complex scenes, but only meets user demands without considering maximizing the sum of the throughput of all users. Besides, it is highly sensitive to initial parameters.

DRL developed by DeepMind has caused widespread concern. It is mentioned in [11] that a regression-based NN can be used to predict the path loss of a radio link to optimize the BSs' transmission power. The authors of [12] present two methods for downlink spatial filtering based on K-means clustering algorithm. The first method groups users in clusters by using the K-means algorithm then computes beam widths according to the power level of edge users. The second method also uses K-means clustering and compares the best BSs available for each user. In paper [13], the authors present the DOHM algorithm to explore Hidden Markov Model to automatically capture current states of the BSs and probabilistically estimate a cell outage. This algorithm is the first to incorporate multiple BS states for achieving an efficient COD solution, while at the same time, detecting the other sleeping cell states. The authors of [14] present an SH framework for next-generation networks using dimensionality reduction, which takes advantage of both feature selection and feature extraction techniques. This algorithm can relieve a network expert of analyzing and selecting the most relevant performance indicators in SH tasks and to reduce the storage needs of databases while optimizing the performance of these tasks. In paper [6], the authors present a novel DRL-based framework for power-efficient resource allocation in cloud RANs with the objective of minimizing power consumption and meeting demands of wireless users. It formulates the resource allocation problem as a convex optimization problem. In paper [15], the authors present a dynamic sleeping control algorithm called DeepNap, which uses a Deep Q-network (DQN) to learn effective sleeping policies from high-dimensional raw observations, and considers action experience replay and adaptive reward scaling to deal with non-stationary traffic. DRL has been applied to many solutions, but there are fewer works on COC.

In this paper, we apply DRL to COC of UDN for the first time. We consider the user connection relationship and BS transmission power allocation to maximize the sum of the throughput of all users while satisfying the demand of each mobile user.

III. System Model

As shown in Fig. 1, there is an area in a UDN, which consists of a set of N BSs and a set of M users.

The states of each BS i is modeled by the binary variable ξ_i, where ξ_i=0 represents that BS i is interrupted and ξ_i=1 represents BS i is active. The binary variable x_{ij} represents

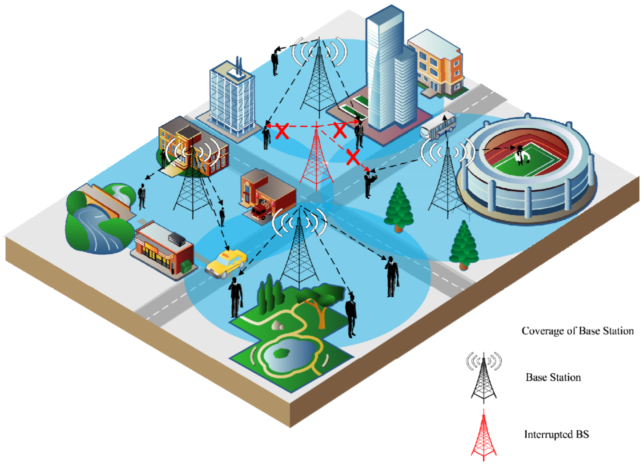

Fig. 1. BS deployment in UDN.

the connection relationship between BS i and user j, where x_{ij}=1 represents that user j is served by BS j. The throughput of user j is

$$\rho_j = \sum_{i \in B} x_{ij} \log_2(1+\gamma_{ij}) \tag{1}$$

where γ_{ij} is the signal-to-interference-plus-noise ratio (SINR) experienced by user j served by BS i. The transmit power of each BS i is

$$P_i = \sum_{j \in U} \pi_{ij} x_{ij} \tag{2}$$

where π_{ij} is the allocated transmission power from BS i to user j.

Therefore, the SINR of user j is

$$\gamma_{ij} = \frac{\pi_{ij}\sigma_{ij}x_{ij}}{\sum_{k \in B} P_k \sigma_{kj} \xi_k (1-x_{kj}) + N_0} \tag{3}$$

where σ_{ij} is the channel loss between BS i and user j, N_0 is the additive white Gaussian noise spectral density.

With the objective of maximizing the sum of the throughput of all users while ensuring that the demand of each mobile user is satisfied, the optimization problem is formulated as follows:

$$\max_{\pi,x,\xi} \quad \sum_{j \in U} \rho_j \tag{4}$$

$$\text{subject to} \quad \sum_{i \in B}\sum_{j \in U} x_{ij} = M \tag{5}$$

$$\sum_{i \in B} x_{ij} = 1 \tag{6}$$

$$\pi_{ij}\sigma_{ij} \geq P_j^{MIN} \tag{7}$$

$$\xi_i \geq x_{ij} \tag{8}$$

$$\sum_{j \in U} \pi_{ij} x_{ij} \leq P_i^{MAX} \tag{9}$$

$$\gamma_{ij} \geq \omega \tag{10}$$

In this model, the decision variables are the user connection relationship x and BS transmission power allocation π. Formula (1) considers the BSs' states ξ, so that the model is suitable for single cell outage scenarios and also for multiple cells outage scenarios. Constraint (5) ensures that every user is served. Constraint (6) ensures that each user must be served by just one BS. Constraint (7) ensures that transmission powers allocated to users meet the demand of each mobile user, where P_j^{MIN} is the minimum received power. Constraint (8) ensures that every user is served by an active BS. Constraint (9) is for the BS transmission power limitation, which is that the total allocated power is no more than maximum transmit power P_i^{MAX}. Constraint (10) ensures that SINR meets requirements.

In paper [10], the transmit power appreciation G is continuous with countless results. Therefore, it is an NP-Hard problem to solve this problem with setting RRU transmit power appreciation as the decision variables. In this paper, the allocation of transmit power is continuous with countless results. For complex scenarios, calculating all possible power allocation schemes is an NP-Hard problem. Therefore, for complex networks, it is impossible to calculate the exact maximum value. So we focus on how to quickly approach the optimal solution under the condition of satisfying the constraints. In contrast to the Q-learning, DRL trains faster and suits for large action space. In contrast to heuristics, DRL can get the global optimal solution and avoid falling into a local optimum. Therefore, we choose DRL to solve this optimization problem.

IV. DRL-BASED COC FRAMEWORK

In this section, we present the DRL-based cell outage compensation framework, which maximizes the sum of the throughput of all users while ensuring that the demand of each mobile user is satisfied. In order to reduce the action space size in the framework, we first group compensation users in clusters by using K-means clustering algorithm and allocate users to adjacent BSs. After that, we use DQN to allocate transmission powers to compensation users in order to meet the demand of each mobile user while maximizing the sum of the throughput of all users. The process of this algorithm is described in Fig. 2.

K-means clustering algorithm is an unsupervised machine learning method. The separation of the samples is performed without considering any prototypes, but rather exploiting the similarities between the data. First, initialize the cluster centroids by considering K adjacent base stations. Then, classify the compensation users according to Euclidean distances between compensation users and the cluster centroids. After that, recompute the cluster centroids, and iterate the above procedure until the cluster centroids do not change significantly. According to the result, compensation users in the same cluster are allocated to the nearest adjacent base station, and the connection relationship between BSs and users is determined.

DRL consists of an offline construction network phase and an online depth Q-learning phase. The offline phase adopts a CNN to obtain the correlation between the state-action pair (s, a) and the value function $Q(s, a)$, and the value function is the

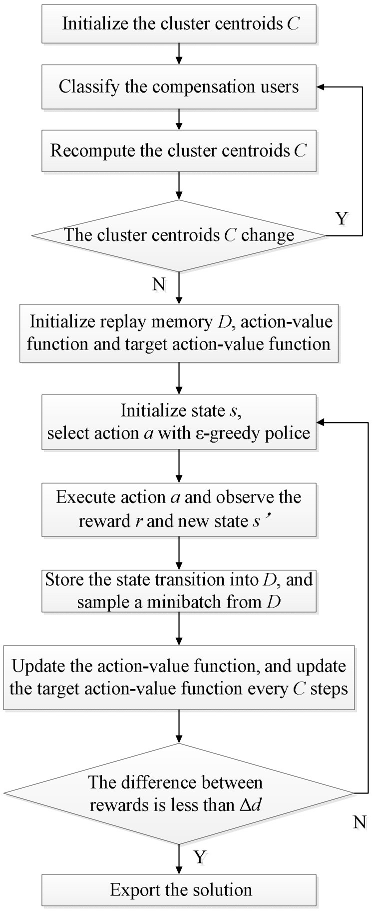

Fig. 2. The process of DRL-based COC framework.

cumulative discount reward when the action a is executed at state s. The offline construction phase needs to accumulate enough samples of value estimates and corresponding (s, a) and use experience memory to smooth the training phase. During the online learning process, in each epoch, the DRL agent uses the CNN to obtain the estimated Q value, then selects action a with ε-greedy police, which means we select an action randomly with probability ε and selects the action with the largest estimated Q value with probability 1-ε. In the interaction with the environment, the immediate reward r and the next state s' are observed, then store state transition (s, a, r, s') in the memory, and update network parameters by sampling from the experience memory. Since the rewards are different depending on the different selection of actions, the network parameters tend to be optimal.

In each decision epoch t, the network reports the current states of all BSs and user connection relationship to the DRL agent. According to the result of the last action, the DRL agent calculates the reward. Then DRL agent will make an action decision a_t based on the state s_t, and allocate transmission power π_{ij} from every BS i to each compensation user j.

We define the state space, action space and reward function of the DRL-based framework as follows:

- *State Space*: (1) Binary variable ξ_i for each BS state; (2) The connection relationship x_{ij} between BS i and user j; (3) Allocated transmission power π_{ij} from BS i to compensation user j.
- *Action Space*: Changes of transmission powers allocated to compensation users. The granularity is 0.01W, the actions of compensation users include +0.01W, 0 and -0.01W.
- *Reward*: The reward needs to represent the objective of the framework, which is simultaneously maximizing the sum of the throughput of all users and satisfying the demand of each mobile user. We define the immediate reward as

$$r_t = \begin{cases} \sum_{j \in U} \rho_j, & \text{satisfying constraints} \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

The proposed framework makes the action space relatively small and the computational complexity relatively low. It is formally presented as Algorithm 1.

Algorithm 1 The DRL-based Framework

Initialize the cluster centroids C;
Repeat:
  Classify the compensation users according to distances;
  Recompute the cluster centroids C with positions of compensation users;
Until finished();
Initialize replay memory D to capacity N;
Initialize action-value function Q with random weights θ;
Initialize target action-value function $\hat{Q}$ with weights $\theta^-=\theta$;
Repeat:
  Initialize sequence s_t and preprocessed sequence $\phi_1 = \phi(s_1)$;
  With probability ε select a random action a_t; otherwise $a_t = argmax_a Q(s_t, a_t; \theta)$, where $Q(\cdot)$ is estimated by the CNN;
  Execute action a_t and observe the reward r_t and the new state s_{t+1};
  Store the state transition (s_t, a_t, r_t, s_{t+1}) into D;
  Randomly sample a minibatch of state transitions (s_t, a_t, r_t, s_{t+1}) from D;
  Target $y_t = r_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta^-)$;
  Perform a gradient descent step $\Delta\theta = \alpha[y_t - Q(s_t, a_t; \theta)]\nabla Q(s_t, a_t; \theta)$;
  Update network parameters $\theta=\theta+\Delta\theta$;
  Every c steps reset $\hat{Q} = Q$;
Until the difference between rewards is less than Δd.

V. Simulation Results

The designed COC simulation scenario is shown as Fig. 3. The area of UDN is 1000m×1000m, which includes 11 BSs and 80 users. In Fig. 3, the 11 base stations are numbered separately, and the dashed circles indicate the coverage of BSs, which has a radius of 250 m. The users are randomly deployed on the network. The solid line connecting the user to the base station represents the connection relationship between the mobile user and the base station. At some point, BS 3 is interrupted, then the service of the user connected to the interrupted base station is interrupted. The compensation user is represented by a cross.

Here, we only consider users within the coverage of the interrupted BS 3 and its adjacent base stations. As shown in Fig. 4, there are 8 compensation users served by the interrupted base station. The compensation users are classified according to the positions of adjacent base stations and allocated to five adjacent base stations using the K-means clustering algorithm. Fig. 4 shows the connection relationship between BSs and users.

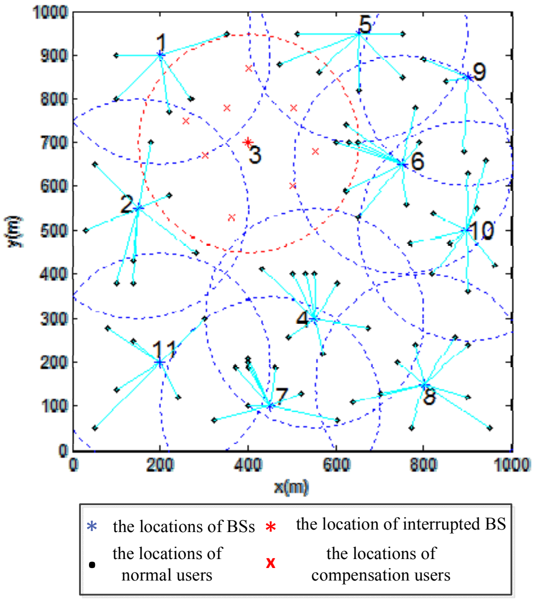

Fig. 3. COC scenario.

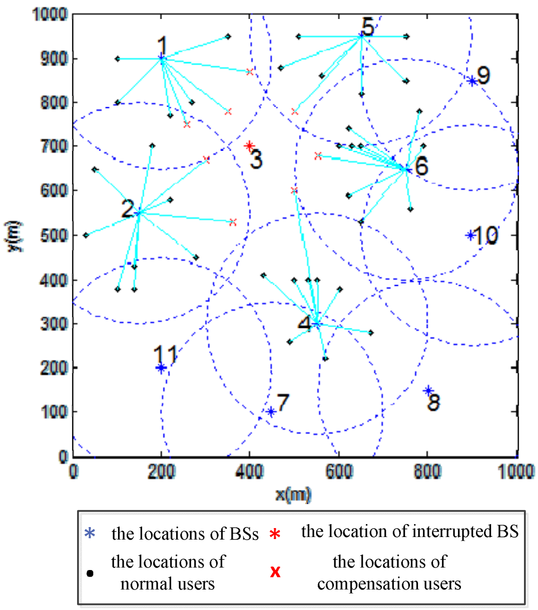

Fig. 4. The reallocated connection relationship between BSs and users.

TABLE I. SIMULATION PARAMETERS

Parameter	Value
Deployment	11 BSs
Max BS Power	5W
Path loss	15.3 + 37.6lg(d[m])
m and M	The index and the total number of users
Noise PSD	-174dBm/Hz
UE sensitivity	-90dBm
Min SINR	-6dB
Memory capacity	500
Min sampling batch	32
Reward discount	0.9
The difference between rewards	0.01

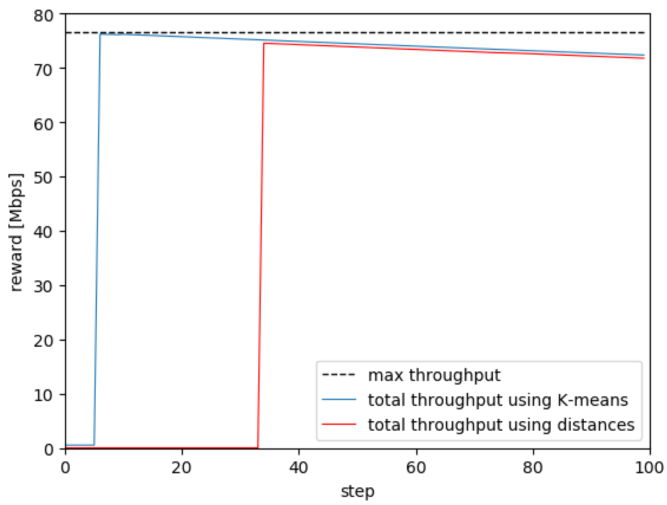

Fig. 5. Immediate reward for every step.

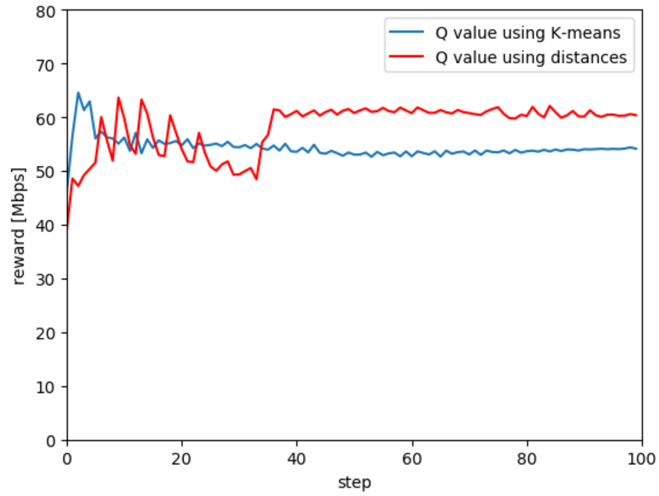

Fig. 6. Cumulative reward for every step.

It is assumed that the maximum transmission power of the base station is 5W, and the transmission power allocated to the normal user is 0.4W. While ensuring the normal users' demand, the transmission power is allocated to compensation users to maximize the sum of the throughput of all users. The simulation settings are shown in Table I.

We assume that the initial allocated transmission power of compensation users is 0, the allocated transmission power change is set to 0.01W. In this paper, we compare the results of different methods of assigning compensation users to adjacent base stations. One is to use the K-means clustering algorithm, the other is to use the way that compensation users are allocated to the nearest BSs. Fig. 5 shows the reward for each step, in which the dashed line indicates the maximum sum of throughput, the blue line indicates the sum of the actual throughput of all users using the K-means clustering algorithm and the red line indicates the sum of the actual throughput of all users using the way that compensation users are served by the nearest BSs. Fig. 6 shows the corresponding cumulative reward. As can be seen from Fig. 6, the cumulative reward tends to be stable quickly, and the convergence speed using the K-means clustering algorithm is faster than that using the nearest distance. In addition, because the Q value after stabilization using the nearest distance is bigger, we learn that the way using the K-means clustering algorithm trains faster. We can learn from Fig. 5 that the reward converges quickly and tends to be stable, 100% compensation users are compensated, and the sum of the throughput of all users using the K-means clustering algorithm is bigger than that using the nearest distance. Finally, the sum of the throughput of all users using our proposed DRL-based COC algorithm is 75.71Mbps, which reaches 99.53% of the maximum throughput 76.06Mbps.

VI. DRL-BASED COC FRAMEWORK

In this paper, the proposed DRL-based framework with the objective of maximizing the sum of the throughput of all users while meeting the demand of each mobile user solves the COC problem in UDN well. A well-trained network can quickly resolve COC problems, and its result is closed to the optimal value. Besides, this framework is also suitable for handling multiple cell outages. In this paper, the simulation scenario is relatively simple. In future work, more complex scenarios such as heterogeneous networks and networks with small cells will be considered. And we will compare our proposed DRL-based COC algorithm with others and consider channel allocation in the future. Besides, more parameters like the antenna tilt are considered to solve the COC problem. And we are looking forward to finding a better problem model with a quicker convergence rate and higher accuracy algorithm.

REFERENCES

[1] Q. Li, Y. Xu, R. Q. Hu, and Y. Qian, “Optimal Fractional Frequency Reuse and Power control in the Heterogeneous Wireless Networks”, IEEE Transactions on Wireless Communications, Vol.12, No.6, pp.2658- 2668, June 2013.

[2] M. Selim, A. E. Kamal, K. Elsayed, H. M. Abdel-Atty and M. Alnuem, "Fronthaul cell outage compensation for 5G networks," IEEE Communications Magazine, vol. 54, no. 8, pp. 169-175, August 2016.

[3] J. Moysen and L. Giupponi, "A Reinforcement Learning Based Solution for Self-Healing in LTE Networks," in 2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall), Vancouver, BC, 2014, pp. 1-6.

[4] A. Asghar, H. Farooq and A. Imran, "Self-Healing in Emerging Cellular Networks: Review, Challenges, and Research Directions," IEEE Communications Surveys & Tutorials, vol. 20, no. 3, pp. 1682-1709, thirdquarter 2018.

[5] S. S. Mwanje, L. C. Schmelz and A. Mitschele-Thiel, "Cognitive Cellular Networks: A Q-Learning Framework for Self-Organizing Networks," IEEE Transactions on Network and Service Management, vol. 13, no. 1, pp. 85-98, March 2016.

[6] Z. Xu, Y. Wang, J. Tang, J. Wang and M. C. Gursoy, "A deep reinforcement learning based framework for power-efficient resource allocation in cloud RANs," in 2017 IEEE International Conference on Communications (ICC), Paris, 2017, pp. 1-6.

[7] B. Rong, Y. Qian and H. h. Chen, "Adaptive power allocation and call admission control in multiservice WiMAX access networks [Radio Resource Management and Protocol Engineering for IEEE 802.16]," in IEEE Wireless Communications, vol. 14, no. 1, pp. 14-19, Feb. 2007.

[8] W. Li, P. Yu, M. Yin and L. Meng, "A distributed cell outage compensation mechanism based on RS power adjustment in LTE networks," China Communications, vol. 11, no. 13, pp. 40-47, Supplement 2014.

[9] S. Fan and H. Tian, "Cooperative Resource Allocation for Self-Healing in Small Cell Networks," IEEE Communications Letters, vol. 19, no. 7, pp. 1221-1224, July 2015.

[10] M. Yin, W. Li, L. Feng, P. Yu and X. Qiu, "Multi-Cell Cooperative Outage Compensation in Cloud-RANs Based 5G Public Safety Network," IEEE Access, vol. 5, pp. 17309-17321, 2017.

[11] P. V. Klaine, M. A. Imran, O. Onireti and R. D. Souza, "A Survey of Machine Learning Techniques Applied to Self-Organizing Cellular Networks," IEEE Communications Surveys & Tutorials, vol. 19, no. 4, pp. 2392-2431, Fourthquarter 2017.

[12] P. Savazzi and L. Favalli, "Dynamic Cell Sectorization Using Clustering Algorithms," in 2007 IEEE 65th Vehicular Technology Conference - VTC2007-Spring, Dublin, 2007, pp. 604-608.

[13] M. Alias, N. Saxena and A. Roy, "Efficient Cell Outage Detection in 5G HetNets Using Hidden Markov Model," IEEE Communications Letters, vol. 20, no. 3, pp. 562-565, March 2016.

[14] D. Palacios, S. Fortes, I. de-la-Bandera and R. Barco, "Self-Healing Framework for Next-Generation Networks through Dimensionality Reduction," IEEE Communications Magazine, vol. 56, no. 7, pp. 170-176, JULY 2018.

[15] J. Liu, B. Krishnamachari, S. Zhou and Z. Niu, "DeepNap: Data-Driven Base Station Sleeping Operations through Deep Reinforcement Learning," IEEE Internet of Things Journal, 2018(online).