Ad Hoc Networks 102 (2020) 102069

Contents lists available at ScienceDirect

Ad Hoc Networks

journal homepage: www.elsevier.com/locate/adhoc

A deep reinforcement learning for user association and power control in heterogeneous networks

Hui Ding[a], Feng Zhao[b,*], Jie Tian[c], Dongyang Li[a], Haixia Zhang[a,d]

[a] Shandong Key Laboratory of Wireless Communication Technologies, Shandong University, Qingdao, Shandong, 266237, China
[b] Yulin Normal University, Yulin 537000, China
[c] School of Information Science and Engineering, Shandong Normal University, Jinan 250014, P.R. China
[d] School of Control Science and Engineering, Shandong University, Jinan, Shandong, 250061, China

ARTICLE INFO

Article history:
Received 21 November 2019
Revised 16 December 2019
Accepted 20 December 2019
Available online 4 February 2020

Keywords:
Heterogeneous networks
User association
Power control
Reinforcement learning
Deep Q-learning network

ABSTRACT

Heterogeneous network (HetNet) is a promising solution to satisfy the unprecedented demand for higher data rate in the next generation mobile networks. Different from the traditional single-layer cellular networks, how to provide the best service to the user equipments (UEs) under the limited resource is an urgent problem to solve. In order to efficiently address the above challenge and strive towards high network energy efficiency, the joint optimization problem of user association and power control in orthogonal frequency division multiple access (OFDMA) based uplink HetNets is studied. Considering the non-convex and non-linear characteristics of the problem, a multi-agent deep Q-learning Network (DQN) method is studied to solve the problem. Different from the traditional methods, such as game theory, fractional programming and convex optimization, which need more and accurate network information in practice, the multi-agent DQN method requires less communication information of the environment. Moreover, for the communication environment dynamics, the maximum long-term overall network utility with a new reward function while ensuring the UE's quality of service (QoS) requirements is achieved by using the multi-agent DQN method. Then, according to the application scenario, the action space, state space and reward function of the multi-agent DQN based framework are redefined and formulated. Simulation results demonstrate that the multi-agent DQN method has the best performance on convergence and energy efficiency compared with the traditional reinforcement learning (Q-learning).

1. Introduction

In order to meet the explosive increase of mobile data traffic demands, heterogeneous networks (HetNets) have been proposed as an efficient solution due to the characteristics of the dense deployment and heterogeneity [1]. Compared with traditional homogeneous networks, HetNets consist of different types of base stations (BSs) named as micro BS, pico BS and femto BS, etc. These BSs are characterized by their transmit power, BS density and data rate [2–4]. As the number of mobile devices increases, the interference among user equipments (UEs) will be more severe under the spectrum sharing strategy for the uplink HetNets. Thus, orthogonal frequency division multiple access (OFDMA) based HetNets have been considered in major wireless communication standards [5–7]. Since the macro BS and small BS have different coverage, transmit power and processing capability, when the conventional maximum received signal strength based user association scheme is applied to the HetNets, it will result in inefficient small BS deployment because most UEs are associated with the macro BS and very few UEs are attracted by small BS. In addition, with the increase of UEs, the uplink interference is another bottleneck in HetNets [8]. A proper setting of transmit power by a power control strategy can decrease the interference among the UEs who select the same subchannel in the OFDMA based HetNets, which strongly influences the quality of service (QoS) of UEs. Thus, to further improve the system performance and user experience, the joint optimization problem of user association and power control is of great importance in the HetNets.

There are some works that studied the user association and power control problems in [9–14]. Considering the interplay within user association and power control, some literatures investigate joint optimization of user association and power control in the HetNets, such as [15–18]. The author in [15] investigated the uplink energy-efficient of the communication between the primary users and the secondary users through user association and power control and proposed an iterative algorithm to solve this

* Corresponding author.
E-mail address: zhaofeng@guet.edu.cn (F. Zhao).

https://doi.org/10.1016/j.adhoc.2019.102069
1570-8705/

problem by using convex optimization etc. Under the noncooperative game theory, a universal joint BS association and power control algorithm for HetNets was proposed by considering the system throughput in [16]. A joint user association and power control strategy for balancing the network loads by maximizing the weighted sum of long-term rate was designed in [17]. In [18], a heuristic algorithm was proposed to deal with the delay-aware uplink user association problem in conjunction with power control in HetNets. In addition, considering the non-convex and nonlinear characteristics about the joint user association and power control problem, it is difficult to obtain a global optimal solution. In order to solve the problem, some methods have recently been developed, such as the convex optimization strategy [9,10,12,13,15], game-theoretic method [11,14,16], fractional programming approach [17] and heuristic algorithm [18].

However, in order to obtain the solution of the problem by the above methods, a more and accurate network information are required, which may not be effectively and practically under the change of the communication environment in practice. For the time-varying dynamic environment, how to solve this problem more effectively and intelligently is still a challenge for the HetNets. Thus, the emerging artificial intelligence method turns into an efficient tool for the problem [19]. By constantly interacting with the environment, reinforcement learning [20,21] can solve the long-term decision and game-theoretic problems through the online learning. Among the reinforcement learning algorithms, the Q-learning algorithm is widely used because it does not need to know the state transition probability [22]. By using less prior knowledge of the environment, the Q-learning method can obtain the optimal policy to solve the intelligent decision problems. In [23], by using the Q-learning method, the author studied a joint channel allocation and power control problem for device-to-device (D2D) transmission underlaying a conventional single-cell cellular network. Then a Q-learning based method for autonomous channel and power level selection by D2D users in a multi-cell network was studied in [24]. For load balancing in the vehicular networks with heterogeneous BSs, a distributed user association algorithm based online Q-learning was studied in [25]. In [26], a Q-learning based power control scheme for energy-efficient optimization in femtocell networks was studied. The problem of joint caching and resource allocation was investigated for a network of cache-enabled unmanned aerial vehicles (UAVs) that serve wireless ground users over the LTE licensed and unlicensed bands [27].

However, the space of state and action considered in [23–27] is relatively small. For the joint user association and power control problem in the HetNets, since the space of state and action is relatively large, it is difficult to get a better performance by Q-learning method. In order to deal with the large space and make up for the deficiency of Q-learning method, a deep reinforcement learning [28] is proposed as a method to handle the large-scale problem. Based on the deep reinforcement learning approach, through the combination of Q-learning and deep neural network (DNN), the deep Q-network (DQN) [29] can effectively improve the network learning performance. In other word, the agent can learn optimal strategy from high dimensional state and action space by using DQN method. Recently, the DQN method has been studied in some works to solve the intelligent resource management and decision problem. In order to minimize the interference to vehicle-to-infrastructure (V2I) communications, a DQN based framework was proposed to optimize the joint sub-band and power level problem in [30,31]. Then, for the mobile edge computing (MEC) system, the author formulated the sum cost of delay and energy consumption for all UEs as the optimization objective and they jointly optimize the offloading decision and computational resource allocation by the DQN method [32]. The author tackled the joint caching, computing, and radio resources allocation problem in the fog-enabled internet of things (IoT), in order to minimize the service latency under the DQN method [33]. By considering the long-term system power consumption under the dynamics of edge cache states, a DQN-based joint mode selection and resource management approach was studied in [34]. However, a few recent literatures study the DQN based method to solve the joint optimization problem in HetNets, such as [35,36]. In [35], the deep reinforcement learning for user association and channel allocation in HetNets was studied, where the author considered the difference between the UE's rate and the BS's transmit power as a reward. The author in [36] studied the control of user association and power allocation to maximize UEs' sum-rate under the constraints of UE's QoS by using the DQN scheme, where a convolutional neural network (CNN) was applied. However, the above studies focus on joint user association and channel allocation (or power allocation) in HetNets without considering the analysis of energy efficiency. Considering the continuous emergence of various new business and application scenarios [37,38], the energy consumption of UE is also rising together with the growing of intensive mobile data computing and applications. Since the current battery technology cannot satisfy the energy consumption of mobile UEs, optimizing the energy efficiency of UEs becomes even more important in the HetNets.

Based on the above analysis, as deep reinforcement learning shows great potential in handling large systems, in this paper, a multi-agent deep reinforcement learning for joint user association and power control is studied. The main contributions of this paper are summarized as follows.

1) In this paper, in order to maximize the energy efficiency of all UEs, we first jointly optimize the user association and power control in OFDMA based uplink HetNets by using the multi-agent DQN method.

2) Since the problem is a mixed-integer non-linear fractional programing (MINLFP) problem, it is difficult to obtain the optimal solution by the traditional methods, and a multi-agent DQN algorithm which requires less transmission overhead information is studied. Based on the contradiction between energy consumption and battery capacity of UE, the UE's energy efficiency is redefined as the reward function in this paper. For the decentralized reinforcement learning framework, the agents are capable of intelligently making their adaptive decisions to maximize their energy efficiency under the constraints of maximum transmit power and UEs' QoS requirements without coordinating with other agents.

3) The performance of multi-agent DQN from the perspectives of the convergence, optimality and stability are analyzed. Simulation results show that the multi-agent DQN based framework achieves better convergence and energy efficiency of all UEs compared to other four methods. From the results, the multi-agent DQN algorithm shows great potential in handling large systems.

The rest of this paper is organized as follows. Section II describes the system model. Section III presents the problem formulation and the multi-agent DQN based framework. Simulation and performance analysis are included in section IV. Finally, conclusions are provided in Section V.

2. System model and problem formulation

2.1. System model

In this work, an OFDMA based two-tier HetNet is considered as shown in Fig. 1. In this scenario, a macro BS is modeled by $m = 0$ and within the coverage area of the macro BS, a set of small BSs is deployed. Without loss of generality, the set of all BSs is denoted as $\mathcal{M} = \{0, 1, 2, \ldots, M\}$. The learning process is done by the cloud server which is connected to the macro or small BSs through the

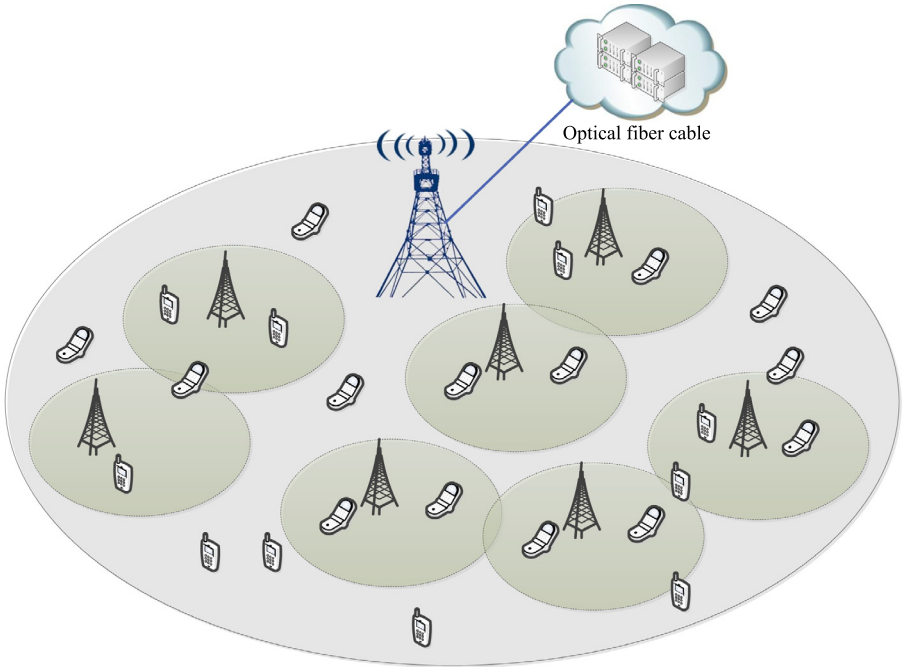

Fig. 1. A typical HetNets.

optical fiber cables. The UEs are randomly distributed in the network, and the set of the UEs is $\mathcal{U} = \{1, 2, \ldots, U\}$, where U is the total number of UEs. The OFDMA based HetNet system has N subchannels denoted by $\mathcal{N} = \{1, 2, \ldots, N\}, N < U$ and the UEs operate on the N subchannels. Since the number of UEs is greater than the number of subchannels, the N orthogonal subchannels are first assigned to the UEs, then the remaining UEs randomly access the subchannels and each UE can only access one subchannel. Assume that all BSs and UEs are equipped with one antenna. The channel gain is dominantly affected by Rayleigh fading $g_{u,m}$, log-normal shadowing (LS in dB) and path loss. The path loss for macro BS and small BS is modeled as PL_1 and PL_2 respectively. So, the channel gain $h^n_{u(m),m}$ of the uth user on the nth subchannel with the mth BS can be expressed as

$$h^n_{u(m),m} = 10^{-(\mathrm{PL}_{1/2}+\mathrm{LS})/10} g_{u,m}. \tag{1}$$

In order to describe the relationship between the UE and the BS, a set of integer binary variables $a_{u,m}, u = 1, 2, \ldots, U, m = 1, 2, \ldots, M$ are introduced. The integer variable $a_{u,m}$ denotes whether the link between BS and UE is active or not, the details as follows

$$a_{u,m} = \begin{cases} 1, & \text{if the } u\text{th UE is associated with the } m\text{th BS,} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Moreover, the power consumption of UEs is composed of two parts, static power consumption and dynamic power consumption. The static power is the power consumed when running the circuit components, such as converters, mixers, filters and so on, and the dynamic power consumption is deemed as the transmit power consumption. Since the transmit power in a digital handset can only be updated at discrete levels, the transmit power can be easily and efficiently set with a finite number of values by the practical application. Assume that $P^n_{u,m} = \{\triangle p, 2\triangle p, 3\triangle p, \ldots, K\triangle p\} = \{k\triangle p\}^K_{k=1}$ is the transmit power of the uth UE connected to the mth BS on the nth subchannel, where $K(1 \le K < \infty)$ denotes the number of power levels, and $\triangle p = P^{MAX}/K$ is the power step, with P^{MAX} denoting the maximum transmit power of UE. By assuming that the static power consumption of UE is P_{CU}, the total power consumption P^{sum}_u of the uth UE can be expressed as

$$P^{\mathrm{sum}}_u = P_{CU} + P^n_{u,m}. \tag{3}$$

Then, the signal to interference plus noise ratio (SINR) of the uth UE connected to the mth BS on the nth subchannel can be expressed as

$$\gamma_{u,m} = \frac{p^n_{u,m} h^n_{u(m),m}}{I^n_{u,m} + \sigma^2}, \tag{4}$$

where σ^2 denotes the noise power and $I^n_{u,m} = \sum_{m' \in \mathcal{M}} \sum_{i \in \mathcal{U} \setminus u} a_{i,m'} p_{i,m'} h^n_{i(m'),m}$ represents the interference of the uth UE connected to the mth BS on the nth subchannel, and the interference mainly comes from the other users who select the same nth subchannel.

According to the Shannon capacity formula, the date rate of the uth UE can be expressed as

$$R_u = \sum_{m=0}^{M} a_{u,m} \log_2(1 + \gamma_{u,m}). \tag{5}$$

2.2. Problem formulation

Based on the described system model, the optimization problem of maximizing the energy efficiency of all UEs in HetNets is formulated in this part by jointly considering user association and power control. The energy efficiency (bits/Joule) of all the UEs is defined as the sum of the energy efficiency of each UE. The individual energy efficiency of uth user selecting the mth BS on nth subchannnel is defined as the ratio of its achievable throughput to the uth user's total power consumption

$$\eta_u = \frac{R_u}{P^{\mathrm{sum}}_u}. \tag{6}$$

Thus, the sum-energy efficiency maximization problem can be formulated as follows

$$\begin{aligned}
&\mathcal{P}1: \max_{\mathbf{A},\mathbf{p}} \sum_{u=1}^{U} \eta_u \\
&\quad \text{s. t. } C1: 0 \le P_{u,m}^{n} \le P_{u}^{\mathrm{MAX}}, \\
&\qquad m = 0, 1, 2, \ldots, M, u = 1, 2, \ldots, U, \\
&\quad C2: \sum_{m} a_{u,m}\gamma_{u,m} \ge \gamma_{th}, \\
&\qquad m = 0, 1, 2, \ldots, M, u = 1, 2, \ldots, U, \\
&\quad C3: a_{u,m} \in \{0, 1\}, \\
&\qquad m = 0, 1, 2, \ldots, M, u = 1, 2, \ldots, U, \\
&\quad C4: \sum_{m} a_{u,m} = 1, \\
&\qquad m = 0, 1, 2, \ldots, M, u = 1, 2, \ldots, U.
\end{aligned} \tag{7}$$

In the above formulated problem $\mathcal{P}1$, $\mathbf{A}$ denotes the user association matrix and $\mathbf{p}$ is the vector of all UE's transmit power. The constraint $C1$ means that the transmit power of each UE cannot exceed the given maximum transmit power. $C2$ ensures that the QoS requirement for each UE can be satisfied, i.e. the predefined minimum SINR γ_{th} is guaranteed. $C3 - C4$ help to make sure that each UE can associate with only one BS. Through solving the MINLFP problem $\mathcal{P}1$, we can find an optimal control strategy about UE association with BSs and transmission power, i.e., $\mathbf{A}$ and $\mathbf{p}$.

3. Multi-agent DQN for joint user association and power control

From $\mathcal{P}1$, it can be seen that the user association and power control mechanisms are mutually involved with each other, and the problem is a mixed-integer and non-convex problem. To efficiently solve the problem, a multi-agent DQN method based on reinforcement learning is studied. The main parts of the reinforcement learning based Markov decision process are shown with a new proposed reward function before presenting the multi-agent DQN approach.

3.1. The reinforcement learning approach

For the scenario considered in this paper, similar to the existing works [39,40], the joint optimization problem $\mathcal{P}1$ is converted to a Markov decision process $(\mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{P}_{ss'})$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ defines the action space of UEs, $\mathcal{R}$ is the reward function, $\mathcal{P}_{ss'}$ is the state transition probability from state s to state s'. Then, the system state space, action space and reward function to construct the reinforcement learning process are described as follows.

State space: In the formulated problem $\mathcal{P}1$, the UE as the agent needs to select the BSs for communication and determine the transmit power, then the system state space can be defined as $s_{state} = \{s_1, s_2, \ldots s_j, \ldots, s_{(M\times K)^U}\}$, where $s_j = j$ denotes the state that the situation of all UEs association with BS and power control. Notice that the number of possible states can be very large with the increase of index U. Considering the problem in another respect, since each UE can select one subchannel and the users who select the same subchannel will have an impact on each other due to the BS selection and power control. Therefore, from the perspective of subchannel, the state space in each subchannel can be defined as $s_n = \{s_1, s_2, \ldots s_j, \ldots, s_{(M\times K)^{\mathrm{N}_n}}\}$, where N_n is the number of UEs in the nth subchannel service.

Action space: In the formulated problem $\mathcal{P}1$, the UE association with the BS and transmit power should be controlled, which leads to a composite action. The one action of the UEs in the nth subchannel at time instant t can be defined as $a_j(t) = \{\{a_{1m}(t), p_{1m}^{n}(t)\}_1, \{a_{um}(t), p_{um}^{n}(t)\}_2, \ldots, \{a_{Um}(t), p_{Um}^{n}(t)\}_{\mathrm{N}_n}\}$. Then, for the all actions in the nth subchannel, the action space at time instant t can be defined as $a_n(t) = \{a_1(t), a_2(t), \ldots, a_j(t), \ldots, a_{(M\times K)^{\mathrm{N}_n}}(t)\}$. The all actions for the N subchannels can be defined as

$$a(t) = \{a_1(t), a_2(t), \ldots, a_n(t), \ldots, a_N(t)\}. \tag{8}$$

Reward function: The learning process is driven by the reward function in the reinforcement learning framework. Under the OFDMA based HetNets, the sum-energy efficiency of all UEs is defined as the system reward function, which can be expressed as

$$r(t) = \sum_{n=1}^{N} r_n(t) = \sum_{u=1}^{U} \eta_u(t), \tag{9}$$

where $r_n(t)$ (i.e., $r_n(t) = \sum_{i=1}^{\mathrm{N}_n} \eta_i(t)$) is the reward function of the nth subchannel, which learns the optimal policy by maximizing its reward with the interactions of the environment.

Through the above analysis, the problem $\mathcal{P}1$ can be transformed into problem $\mathcal{P}2$ as follows

$$\begin{aligned}
&\mathcal{P}2: \max_{\mathbf{A}_n,\mathbf{p}_n} \quad r_n, \\
&\text{s. t.} \qquad\quad C1 - C4,
\end{aligned} \tag{10}$$

where $\mathbf{A}_n$ and $\mathbf{p}_n$ denote the user association matrix and transmit power vector on the nth subchannel, respectively.

For the reinforcement learning in each subchannel, at the instant time t, the agent knows its current state s_n^t and obtains the strategy π_n according to the s_n^t by using the learning process. The agent will decide to make an action $a_n(t)$ according to the policy π_n, i.e., $a_n(t) = \pi_n(s_n^t)$. Then the UEs in the subchannel will send signals to the BSs and get the reward $r_n(t) = r_n(t|s = s_n^t, a = a_n(t))$. As a result, the state transits to a new state $s_n^{t'}$ through the action $a_n(t)$ and continues the above operations until it reaches the maximum epoch. In the reinforcement learning, the returned reward $R_n(\tau)$ is the accumulated and discounted reward that is given by

$$R_n(\tau) = \sum_{\tau=t}^{T} \gamma^{\tau-t} r_n(\tau), \tag{11}$$

where $\gamma \in [0, 1]$ is the discount factor and T is the maximum epoch. When $\gamma = 0$, the current reward is considered and for the case of $\gamma = 1$, $R_n(\tau)$ equals to the sum of the rewards.

The objective of the agent is to maximize the expected accumulated reward under the UE's QoS satisfaction constraints, i.e., $\max E[R_n(\tau)|s_n^\tau]$. If the reward value is the largest, the optimal policy π_n^* is obtained, i.e., the highest energy efficiency of all UEs is achieved with the constraints, which is equal to the problem $\mathcal{P}2$.

In order to solve the maximization problem, a value function $V_n^{\pi_n}(s_n^\tau)$ which is the accumulated reward for the policy π_n is defined [41], i.e., $V_n^{\pi_n}(s_n^\tau) = E[R_n(\tau)|s_n^\tau]$. By considering the Markov property, $V_n^{\pi_n}(s_n^\tau)$ can be rewritten as

$$\mathrm{V}_n^{\pi_n}(s_n^\tau) = \mathrm{r}(s_n^\tau, \pi_n) + \gamma \Sigma_{s_n^{\tau'}} \mathrm{P}_{s_n^\tau, s_n^{\tau'}}(\pi_n) \mathrm{V}_n^{\pi_n}(s_n^{\tau'}), \tag{12}$$

where $\mathrm{P}_{s_n^\tau, s_n^{\tau'}}$ is the state transition probability from state s_n^τ to state $s_n^{\tau'}$.

Then, a state-action value function (Q-value function) which characterize the expected reward for choosing action $a_n(\tau)$ at system state s_n^τ by following the strategy π_n is defined as

$$Q_{\pi_n}(s_n^\tau, a_n(\tau)) = E[R_n(\tau)|s_n^\tau, a_n(\tau)]. \tag{13}$$

Based on the Bellman's equation [41], the optimal Q-value function can be expressed as

$$Q_{\pi_n^*}(s_n^\tau, a_n(\tau)) = r_n(s_n^\tau, a_n(\tau)) + \gamma \Sigma_{s_n^{\tau'}} \mathrm{P}_{s_n^\tau, s_n^{\tau'}}(a_n(\tau)) V_n^{\pi_n^*}(s_n^{\tau'}). \tag{14}$$

According the Bellman optimality equation [42], for the $V_n^{\pi_n^*}(s_n^{\tau'})$ in (14), which can be obtained as follows

$$V_n^{\pi_n^*}(s_n^{\tau'}) = \max_{a_n'} Q_{\pi_n^*}(s_n^{\tau'}, a_n'(\tau)). \quad (15)$$

Then, combining (14) and (15), the formula (14) can be expressed as

$$Q_{\pi_n^*}(s_n^{\tau}, a_n(\tau)) = r_n(s_n^{\tau}, a_n(\tau)) + \gamma \Sigma_{s_n^{\tau'}} P_{s_n^{\tau}, s_n^{\tau'}}(a_n(\tau)) \max_{a_n'} Q_{\pi_n^*}(s_n^{\tau'}, a_n'(\tau)). \quad (16)$$

For the formula (16), $P_{s_n^{\tau}, s_n^{\tau'}}(a_n(\tau))$ is very difficult to obtain. Thus, by using the Q-learning algorithm from [41], the update of Q-value function can be expressed as

$$Q_{\pi_n}(s_n^{\tau}, a_n(\tau)) = (1-\alpha) Q_{\pi_n}(s_n^{\tau}, a_n(\tau)) + \alpha [r_n(s_n^{\tau}, a_n(\tau)) + \gamma \max_{a_n'} Q_{\pi_n}(s_n^{\tau'}, a_n'(\tau'))], \quad (17)$$

where α is the learning rate.

However, the Q-learning method will select the action which has the best value and it will overestimate the selected action. In addition, the Q-learning method uses the sampling method to select the state, which will overestimate the sampled state, and the gap between the sampled state and the unsampled state will be larger. For the HetNets, it is worth mentioning that it is challenging to obtain an optimal solution with Q-learning method due to the large state and action space. That means some states are not sampled for the large-scale system state space. Thus, in order to deal with the large-scale system state space, the deep reinforcement learning is investigated to solve the problem, which will be described in detail in the next subsection.

3.2. Multi-agent DQN framework

A multi-agent DQN method is studied in this subsection. Different from the Q-learning method, a DNN is used to estimate the values of the Q-value function $Q_{\pi_n}(s_n^{\tau}, a_n(\tau))$. Thus, the Q-function at time τ approximates $Q_{\pi_n}(s_n^{\tau}, a_n(\tau)|\theta)$, i.e., $Q_{\pi_n}(s_n^{\tau}, a_n(\tau)) \approx Q_{\pi_n}(s_n^{\tau}, a_n(\tau)|\theta)$, where θ is the weight parameter of the behavior network. Then the optimal policy is given by

$$\pi_n^* = \max_{a_n'} Q_{\pi_n^*}(s_n^{\tau}, a_n'(\tau)|\theta). \quad (18)$$

Due to differences between the data samples, it is hard to get a smooth learning model. Thus, a target network with the weight parameter of θ^- is considered. For the multi-agent DQN method, there are two networks, i.e., behavior network and target network. By using the target network, the learning model for calculating the target value y_i will be constant with the weight parameter of θ^- for a certain time, which can alleviate the volatility of the learning model. In addition, a estimate value can be obtained by the behavior network and for the behavior network, which is also called online network in some literatures, and the detailed reasons will be described as follows.

In the learning process, after a certain number of iterations C, the weight parameter θ of behavior network will be synchronized to the target network, i.e. $\theta \to \theta^-$, then, the next stage of learning will get started. For the behavior network, the agent will use the ϵ-greedy policy to choose action $a_n(\tau)$ and the parameter θ is updated for each iteration by using the minimum loss function as follows

$$L(\theta) = \sum [(y_j - Q_{\pi_n}(s_n, a_n|\theta))^2], \quad (19)$$

where

$$y_j = \begin{cases} r_n(j), \text{if } \mathrm{SINR}_{s_n^{j'}} < \gamma, \\ r_n(j) + \gamma \max_{a_n'(j')} Q_{\pi_n}(s_n^{j'}, a_n'(j') \mid \theta^-), \text{otherwise.} \end{cases} \quad (20)$$

In particular, due to the correlation between the data samples, it will lead to learning instability, thus the experience replay technique is applied in the deep Q-learning. The experience replay contains two parts: stored data and sampled data. The experience data is stored into the memory D by the order of iteration. During learning, the agent will choose action $a_n(\tau)$, get a reward $r_n(\tau)$ and turn to next state $s_n^{\tau'}$. Then the vector $(s_n^{\tau}, a_n(\tau), r_n(\tau), s_n^{\tau'})$ will be stored into the experience memory. If the memory D is already full, the new experience data will cover the data which was generated by the earliest iteration. For the data sampling, the agent selects at random a mini-batches of experiences from the replay memory D to update the parameter θ. If the sampled data is the latest stored in the memory D, which is similar to online learning, that is why some literatures refer to the target network as online network. The process of the DQN strategy is shown in Fig. 2. By the multi-agent DQN strategy, the best strategy π_n^* can be derived.

The detailed process of multi-agent DQN algorithm for joint user association and power control is presented in Algorithm 1 . At the beginning, each agent initializes the memory D and the weight parameters of θ and θ^- for behavior network and target network respectively. Then, the agent initializes the starting state s_n^{τ} and the ε-greed policy is used to select an action $a_n(\tau)$. Next, the agent will send the information about user association and transmit power to the environment, if the constraints satisfy, the reward $r_n(\tau)$ and the next state $s_n^{\tau'}$ can be obtained. Otherwise, the agent will not replay anything. The experience information $(s_n^{\tau}, a_n(\tau), r_n(\tau), s_n^{\tau'})$ will be stored in the memory D. By using the sample random mini-batch for the memory D, the weight parameter of the behavior network is updated. When training a certain number of iterations C, the parameter of the behavior network will be synchronized to the target network. The next stage of learning will begin.

Algorithm 1: Multi-agent DQN algorithm for joint user association and power control.

Input: learning rate α, the parameters of action-value function θ and θ^-.
Initialize replay memory D.
Initialize behavior action-value function Q_{π_n} with weights θ and target action-value function $Q_{\pi_n'}$ with weights θ^-, $\theta^- = \theta$.
repeat
  Initialize the starting state s_n^{τ}.
  repeat
    Generate a random number $x \in (0, 1)$.
    if $x < \varepsilon$ **then**
      Select action randomly;
    else
      Select the action $a_n(\tau)$ characterized by the maximum Q-value, i.e., $a_n(\tau) = \arg\max_{\mathbf{a}_n} Q_{\pi_n}(s_n^{\tau}, \mathbf{a}_n \mid \theta)$;
    end if
    Take the action $a_n(\tau)$ and observe the next state $s_n^{\tau'}$.
    Observe reward $r_n(\tau) = \sum_{i=1}^{N_n} \eta_i(\tau)$.
    Store experience $(s_n^{\tau}, a_n(\tau), r_n(\tau), s_n^{\tau'})$ in D.
    Sample random minibatch of experience $(s_n^j, a_n(j), r_n(j), s_n^{j'})$ from D.
    if $\mathrm{SINR}_{s_n^{j'}} < \gamma$ **then**
      $y_j = r_n(j)$;
    else
      $y_j = r_n(j) + \alpha \max_{a_n'(j')} Q_{\pi_n}(s_n^{j'}, a_n'(j') \mid \theta^-)$;
    end if
    Perform a gradient descent step on $(y_j - Q_{\pi_n}(s_n^j, a_n(j); \theta))^2$ with respect to the network parameters θ.
    Every C steps reset $\theta^- \leftarrow \theta$.
  until terminal.
until terminal.

4. Simulation results and analysis

In this section, the multi-agent DQN algorithm is simulated in a two-tier HetNet where has one macro BS and some micro BSs.

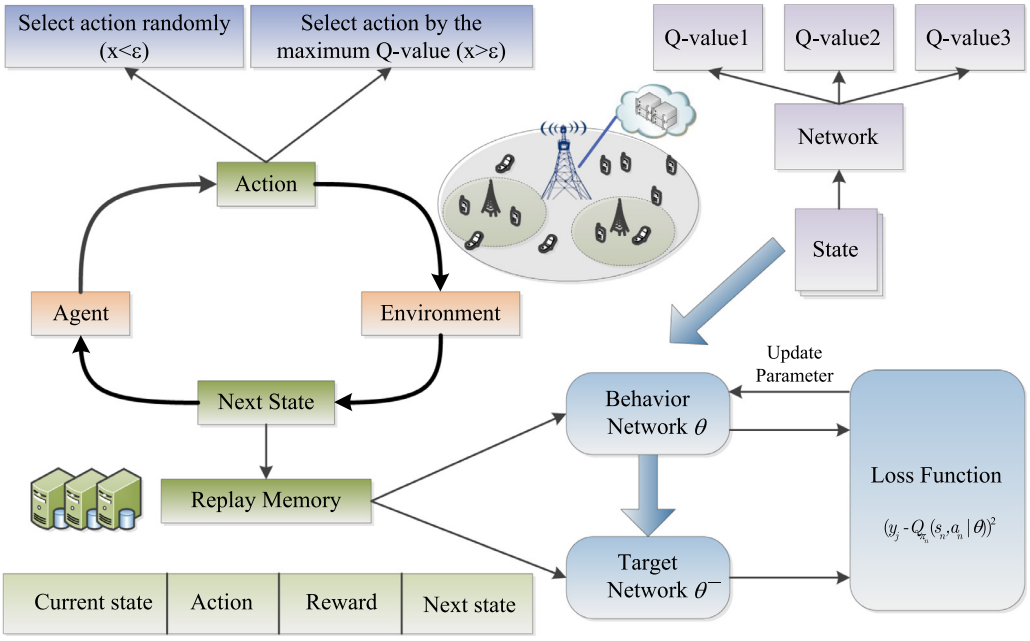

Fig. 2. The strategy of DQN.

Table 1
Parameters of DQN.

Parameters	Values
Episodes	200
Observe steps	500
The size of minimum batch	50
Minimum ε	0.0001
Learning rate α	0.01
Discount rate γ	0.8
Replay memory D	5000
Iterations C	500

There are 25 UEs who are randomly dispersed over the macro BS's coverage area with setting as a 200 m × 200 m square. Besides, the micro BSs are also randomly distributed in the considered area. The maximum transmit power of UEs is 23 dBm and the total number of subchannels is 15. The pass loss of macro BS and micro BS are $PL_1 = 34 + 40\log_{10}(d)$, $PL_2 = 37 + 30\log_{10}(d)$ respectively, where d is the distance from a BS to a UE in meters. The log-normal shadowing is 8 dB. The noise power is set as $\sigma^2 = -174$ dBm.

In order to estimate the Q-function, the DNN is adopted in the model containing two-hidden layer of fully-connected neural network with 64 and 64 neurons, and an output layer (ActionNum neurons), where the number of neurons in output layer is determined by action number in each subchannel. The detailed parameters of DQN are listed in Table 1.

Firstly, the performance of the DNN with different learning parameters (such as learning rate and the number of neuron) is studied. The training efficiency with different learning rate is studied and the results are included in Fig. 3. It can be seen that with the number of episodes increases, the energy efficiency of all UEs is gradual convergence. Moreover, with the change of learning rate α, the performance of energy efficiency of all UEs is the best for $\alpha = 0.01$ than that of $\alpha = 0.1$, $\alpha = 0.001$ and $\alpha = 0.0001$. By comparing the two cases of $\alpha = 0.01$ and $\alpha = 0.1$, it can be seen that when the learning rate α is relatively large, it is not easy to reach the optimal value. When the learning rate is relatively small, it may result in local optimum instead of global optimum. Thus, considering practical real-time execution of the algorithm, the learning rate α is set to 0.01.

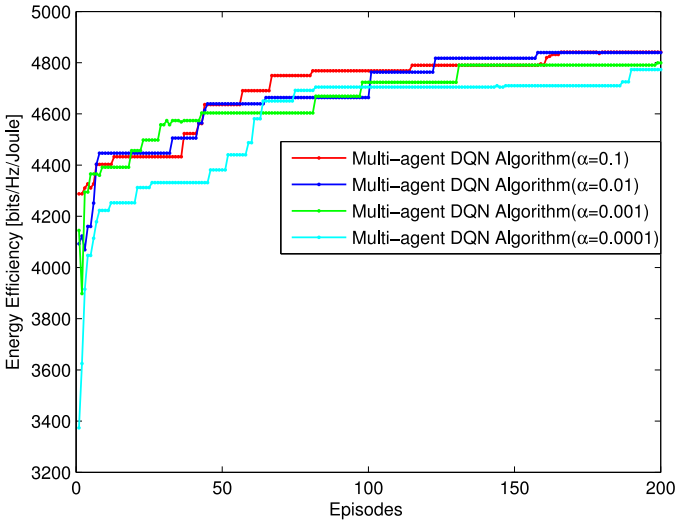

Fig. 3. The energy efficiency versus the different learning rate.

In Fig. 4, the performance of different numbers of neurons in DNN structure is studied. Fig. 4 shows that as the number of neurons increases, the energy efficiency of all UEs decreases. Due to the sparsity of the data samples, when the number of neurons becomes too large, the optimization problem may result in overfitting and more training time. It can be seen that when the neurons equals 64 and 256 of the first layer, the convergence of the two curves is almost the same, while the other cases have the less performance on convergence. Therefore, the neurons of the two-hidden layer are 64 and 64, respectively.

The convergence performance of the multi-agent DQN algorithm is investigated and compared with a classic Q-learning framework used in [25] under the scenario that the SINR requirement is $\gamma = -10$ dB. As shown in Fig. 5, it can be see that the system energy efficiency of Q-learning is lower than the system energy efficiency achieved with the multi-agent DQN method. As the

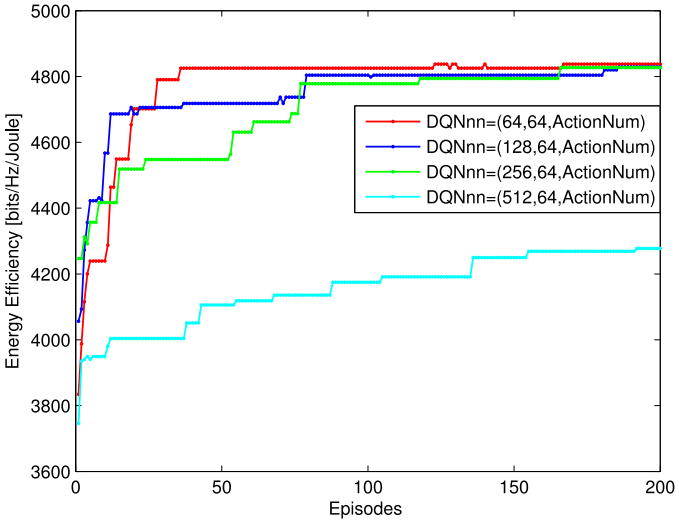

Fig. 4. The energy efficiency versus the different numbers of neurons in DNN structure.

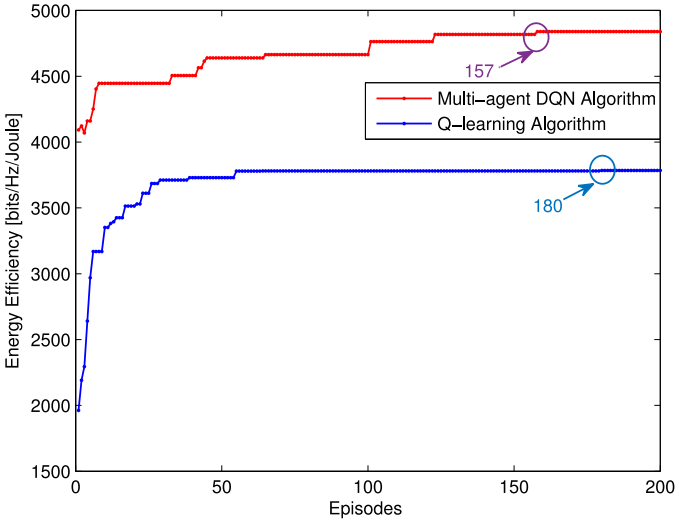

Fig. 5. The proof of convergence.

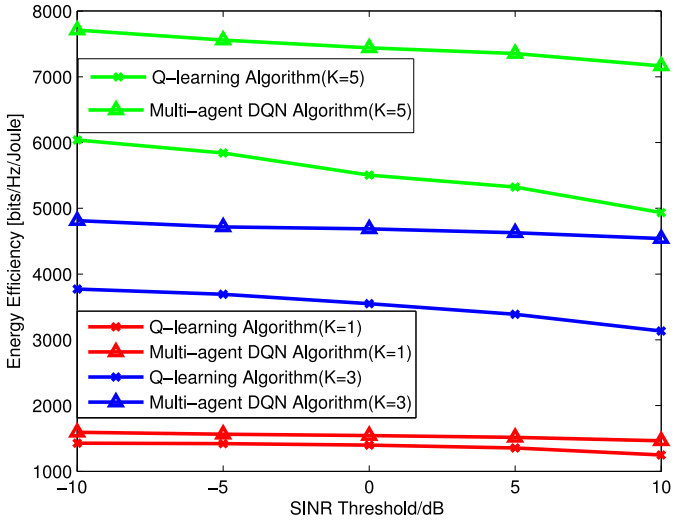

Fig. 6. The energy efficiency versus the threshold of SINR with different K.

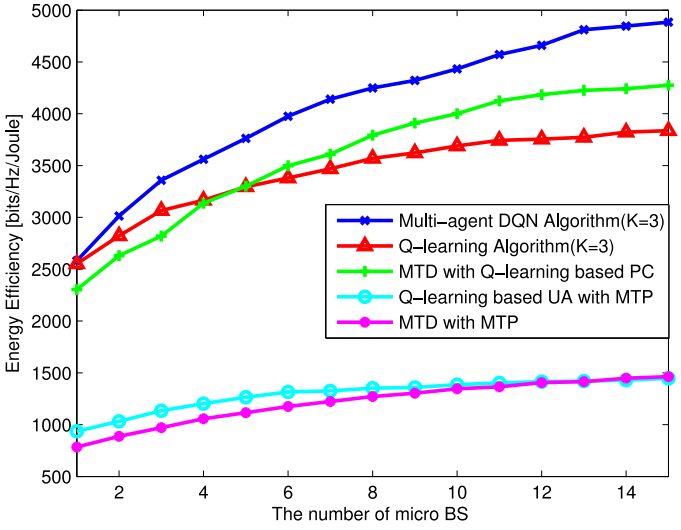

Fig. 7. The energy efficiency versus the number of micro BS.

episode increases, the system energy efficiency increases and tends to convergence for the two schemes. Moreover, the multi-agent DQN algorithm outperforms the Q-learning method on the learning speed. For the Q-learning method, there is a little improvement in the system energy efficiency when episode is approximately equal to 180, while the system energy efficiency tends to be stable when episode is approximately equal to 157 for the multi-agent DQN algorithm. For the multi-agent DQN algorithm, it is unstable at the beginning but the instability is reducing as episodes go on, and then gradually rises, that is because the agent chooses actions randomly and stores the information into replay memory. Through several iterations, the multi-agent DQN algorithm starts to learn from the experience.

Next, the energy efficiency of all UEs is simulated when employing the Q-learning algorithm and multi-agent DQN algorithm under different SINR thresholds. The results are included in Fig. 6. It can be seen that as SINR threshold of UEs increases, the energy efficiency of all the UEs decreases. This is reasonable, since more power is consumed to achieve a high SINR, which thereby will decrease the energy efficiency of all UEs. In addition, as the number of power levels increases, the energy efficiency of all UEs increases. The reason is that as the number of power levels increases, the agent can select a more suitable transmission power under the fixed user association, thereby improving energy efficiency. For the case of $K = 1$, the transmit power of UE is equal to the maximum transmit power, i.e., $P^{MAX} = 23$ dB. From the Fig. 6, when the transmit power of UE is maximum, the energy efficiency is the worst.

Furthermore, the energy efficiency of all UEs for different number of micro BS is investigated. The simulations are done by choosing the SINR requirement $\gamma = -10$ dB and $K = 3$. The results are shown in Fig. 7. To evaluate the performance of the multi-agent DQN algorithm, in addition to the Q-learning algorithm, three other schemes, i.e., MTD with Q-learning based PC, Q-learning based UA with MTP, MTD with MTP are included as reference. For MTD with Q-learning based PC scheme, a user chooses the minimum transmission distance user association scheme and adopts the Q-learning based power control algorithm. For Q-learning based UA with MTP scheme, users adopt the Q-learning based user association scheme and transmit using their maximum transmit power. Finally, for the MTD with MTP scheme, users choose the minimum transmission distance user association scheme and transmit using their maximum transmit power. From the simulation results, it can be seen that as the number of micro BSs increases, the energy efficiency of all UEs increases at first and then gradually increases. For the Q-learning and MTD with Q-learning

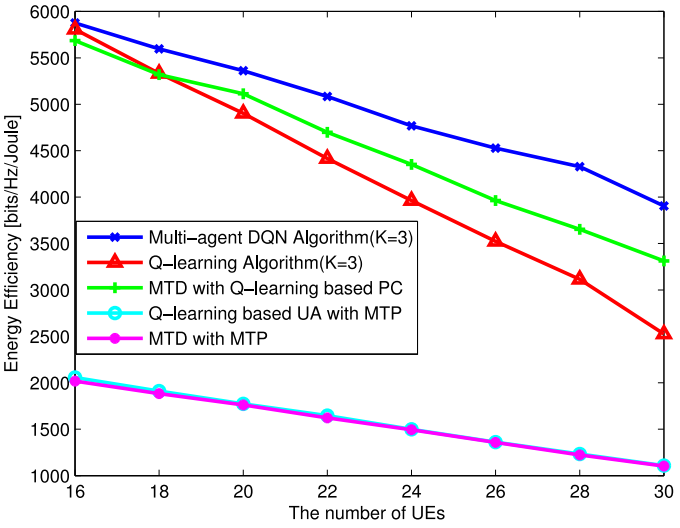

Fig. 8. The energy efficiency versus the number of UEs.

based PC schemes, when the number of micor BS is small, the performance of the Q-learning algorithm is better. As the number of micro BS increases, the performance of the Q-learning algorithm gradually deteriorates. The reason is that the space of state and action becomes larger, and some states are overestimated and unsampled. Therefore, in the OFDMA based HetNets, the number of micro BS should be carefully designed.

Fig. 8 shows the energy efficiency of all UEs under different number of UEs. Simulation results show that the multi-agent DQN algorithm obtains the best performance in terms of the energy efficiency of all UEs through comparison with the other four schemes. This is because that the multi-agent DQN algorithm optimizes not only the user association but also the transmit power compared with the MTD with Q-learning based PC, Q-learning based UA with MTP and MTD with MTP schemes. By using the DNN, the multi-agent DQN algorithm can overcome the shortcomings of Q-learning algorithm, thus the multi-agent DQN algorithm has a better performance than the Q-learning algorithm as in Fig. 8. For the Q-learning and MTD with Q-learning based PC schemes, as the number of UEs increases, the change trend of the two methods is similar to the Fig. 7. This is because that the space of state and action will grow as the number of UEs increases. In addition, as the number of UEs increases gradually, the energy efficiency performance of all UEs for the five simulated schemes decreases. This is because high user's number will incur severe interference.

5. Conclusion

In this paper, the joint optimization problem of user association and power control has been studied in the OFDMA based HetNets. The above resource management problem has been formulated as a maximum long-term uplink energy efficiency of all UEs under the constraints of maximum transmit power and UE's QoS requirements. The multi-agent DQN approach has been utilized to solve the above MINLFP problem. Different from traditional solution methods, a small communication information is needed by the multi-agent DQN algorithm. The convergence of the multi-agent DQN algorithm is analyzed and it has been shown that the multi-agent DQN algorithm has a better performance than the classical Q-learning algorithm on the convergence speed. In addition, the simulation results are done and show that the multi-agent DQN algorithm has a better performance on the energy efficiency than other four schemes, which can improve the energy efficiency of all UEs effectively.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported in part by the National Science Fund of China for Excellent Young Scholars under Grant 61622111, the National Natural Science Foundation of China (No. 61860206005, 61671278, 61871466 and 61801278) and in part by the Guangxi Natural Science Foundation Innovation ResearchTeam Project under Grant 2016GXNSFGA380002.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.adhoc.2019.102069.

References

[1] S. Cai, Y. Che, L. Duan, J. Wang, S. Zhou, R. Zhang, Green 5g heterogeneous networks through dynamic small-cell operation, IEEE J. Sel. Areas Commun. 34 (5) (2016) 1103–1115, doi:10.1109/JSAC.2016.2520217.
[2] W. Guo, S. Wang, X. Chu, J. Zhang, J. Chen, H. Song, Automated small-cell deployment for heterogeneous cellular networks, IEEE Commun. Mag. 51 (5) (2013) 46–53, doi:10.1109/MCOM.2013.6515046.
[3] J. Chen, Y. Deng, J. Jia, M. Dohler, A. Nallanathan, Cross-layer qoe optimization for d2d communication in cr-enabled heterogeneous cellular networks, IEEE Trans. Cognit. Commun. Netw. 4 (4) (2018) 719–734, doi:10.1109/TCCN.2018.2868371.
[4] C. Liu, L. Xiao, Interference precancellation for resource management in heterogeneous cellular networks, IEEE Trans. Cognit. Commun. Netw. 5 (1) (2019) 138–152, doi:10.1109/TCCN.2018.2878199.
[5] H. Yin, S. Alamouti, Ofdma: a broadband wireless access technology, in: 2006 IEEE Sarnoff Symposium, 2006, pp. 1–4, doi:10.1109/SARNOF.2006.4534773.
[6] S. Lohani, E. Hossain, V.K. Bhargava, Joint resource allocation and dynamic activation of energy harvesting small cells in ofdma hetnets, IEEE Trans. Wireless Commun. 17 (3) (2018) 1768–1783, doi:10.1109/TWC.2017.2785301.
[7] S. Rezvani, N. Mokari, M.R. Javan, E. Jorswieck, Fairness and transmission-aware caching and delivery policies in ofdma-based hetnets, IEEE Transactions on Mobile Computing (2019), doi:10.1109/TMC.2019.2892978. 1–1
[8] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, D. Malladi, A survey on 3gpp heterogeneous networks, IEEE Wireless Commun. 18 (3) (2011) 10–21, doi:10.1109/MWC.2011.5876496.
[9] X. Ge, X. Li, H. Jin, J. Cheng, V.C.M. Leung, Joint user association and user scheduling for load balancing in heterogeneous networks, IEEE Trans. Wireless Commun. 17 (5) (2018) 3211–3225, doi:10.1109/TWC.2018.2808488.
[10] Y.L. Lee, T.C. Chuah, A.A. El-Saleh, J. Loo, User association for backhaul load balancing with quality of service provisioning for heterogeneous networks, IEEE Commun. Lett. 22 (11) (2018) 2338–2341, doi:10.1109/LCOMM.2018.2867181.
[11] Y. Xu, S. Mao, User association in massive mimo hetnets, IEEE Syst. J. 11 (1) (2017) 7–19, doi:10.1109/JSYST.2015.2475702.
[12] T.M. Ho, N.H. Tran, C.T. Do, S.M.A. Kazmi, E. Huh, C.S. Hong, Power control for interference management and qos guarantee in heterogeneous networks, IEEE Commun. Lett. 19 (8) (2015) 1402–1405, doi:10.1109/LCOMM.2015.2444844.
[13] B. Xu, Y. Chen, J. Requena-Carrion, Energy-aware power control in energy-cooperation enabled hetnets with hybrid energy supplies, in: 2016 IEEE Global Communications Conference (GLOBECOM), 2016, pp. 1–6, doi:10.1109/GLOCOM.2016.7841804.
[14] J. Zheng, Y. Wu, N. Zhang, H. Zhou, Y. Cai, X. Shen, Optimal power control in ultra-dense small cell networks: a game-theoretic approach, IEEE Trans. Wireless Commun. 16 (7) (2017) 4139–4150, doi:10.1109/TWC.2016.2646346.
[15] M. Wang, H. Gao, T. Lv, Energy-efficient user association and power control in the heterogeneous network, IEEE Access 5 (2017) 5059–5068, doi:10.1109/ACCESS.2017.2690305.
[16] V.N. Ha, L.B. Le, Distributed base station association and power control for heterogeneous cellular networks, IEEE Trans. Veh. Technol. 63 (1) (2014) 282–296, doi:10.1109/TVT.2013.2273503.
[17] T. Zhou, Z. Liu, J. Zhao, C. Li, L. Yang, Joint user association and power control for load balancing in downlink heterogeneous cellular networks, IEEE Trans. Veh. Technol. 67 (3) (2018) 2582–2593, doi:10.1109/TVT.2017.2768574.
[18] Z. Chen, L. Qiu, Y. Jin, X. Liang, Delay-aware uplink user association and power control in heterogeneous cellular networks, IEEE Wireless Commun. Lett. 4 (6) (2015) 661–664, doi:10.1109/LWC.2015.2480073.
[19] R. Li, Z. Zhao, X. Zhou, G. Ding, Y. Chen, Z. Wang, H. Zhang, Intelligent 5g: when cellular networks meet artificial intelligence, IEEE Wireless Commun. 24 (5) (2017) 175–183, doi:10.1109/MWC.2017.1600304WC.

[20] L.P. Kaelbling, M. L.Littman, A. W.Moore, Reinforcement learning: a survey, J. Artif. Intell. Res. (May 1996) vol. 4, pp. 237–285.
[21] R.S. Sutton, A.G. Barto, Introduction to Reinforcement Learning, Cambridge, MA, USA: MIT Press, 1988.
[22] R.S. Sutton, A.G. Barto, R.J. Williams, Reinforcement learning is direct adaptive optimal control, IEEE Control Syst. Mag. 12 (2) (1992) 19–22, doi:10.1109/37.126844.
[23] S. Maghsudi, S. Staczak, Joint channel allocation and power control for underlay d2d transmission, in: 2015 IEEE International Conference on Communications (ICC), 2015, pp. 2091–2096, doi:10.1109/ICC.2015.7248634.
[24] A. Asheralieva, Y. Miyanaga, An autonomous learning-based algorithm for joint channel and power level selection by d2d pairs in heterogeneous cellular networks, IEEE Trans. Commun. 64 (9) (2016) 3996–4012, doi:10.1109/TCOMM.2016.2593468.
[25] Z. Li, C. Wang, C. Jiang, User association for load balancing in vehicular networks: an online reinforcement learning approach, IEEE Trans. Intell. Transp. Syst. 18 (8) (2017) 2217–2228, doi:10.1109/TITS.2017.2709462.
[26] Z. Gao, B. Wen, L. Huang, C. Chen, Z. Su, Q-learning-based power control for lte enterprise femtocell networks, IEEE Syst. J. 11 (4) (2017) 2699–2707, doi:10.1109/JSYST.2016.2535461.
[27] M. Chen, W. Saad, C. Yin, Liquid state machine learning for resource and cache management in lte-u unmanned aerial vehicle (uav) networks, IEEE Trans. Wireless Commun. 18 (3) (2019) 1504–1517, doi:10.1109/TWC.2019.2891629.
[28] K. Arulkumaran, M.P. Deisenroth, M. Brundage, A.A. Bharath, Deep reinforcement learning: a brief survey, IEEE Signal Process Mag 34 (6) (2017) 26–38, doi:10.1109/MSP.2017.2743240.
[29] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, M. Riedmiller., Playing atari with deep reinforce ment learning, arXiv preprint arXiv:1312.5602 (2013).
[30] H. Ye, G.Y. Li, Deep reinforcement learning for resource allocation in v2v communications, in: 2018 IEEE International Conference on Communications (ICC), 2018, pp. 1–6, doi:10.1109/ICC.2018.8422586.
[31] H. Ye, G.Y. Li, B.F. Juang, Deep reinforcement learning based resource allocation for v2v communications, IEEE Trans. Veh. Technol. 68 (4) (2019) 3163–3173, doi:10.1109/TVT.2019.2897134.
[32] J. Li, H. Gao, T. Lv, Y. Lu, Deep reinforcement learning based computation offloading and resource allocation for mec, in: 2018 IEEE Wireless Communications and Networking Conference (WCNC), 2018, pp. 1–6, doi:10.1109/WCNC.2018.8377343.
[33] Y. Wei, F.R. Yu, M. Song, Z. Han, Joint optimization of caching, computing, and radio resources for fog-enabled iot using natural actor critic deep reinforcement learning, IEEE Internet Things J. 6 (2) (2019) 2061–2073, doi:10.1109/JIOT.2018.2878435.
[34] Y. Sun, M. Peng, S. Mao, Deep reinforcement learning-based mode selection and resource management for green fog radio access networks, IEEE Internet Things J. 6 (2) (2019) 1960–1971, doi:10.1109/JIOT.2018.2871020.
[35] N. Zhao, Y. Liang, D. Niyato, Y. Pei, Y. Jiang, Deep reinforcement learning for user association and resource allocation in heterogeneous networks, in: 2018 IEEE Global Communications Conference (GLOBECOM), 2018, pp. 1–6, doi:10.1109/GLOCOM.2018.8647611.
[36] C. Luo, J. Ji, Q. Wang, L. Yu, P. Li, Online power control for 5g wireless communications: a deep q-network approach, in: 2018 IEEE International Conference on Communications (ICC), 2018, pp. 1–6, doi:10.1109/ICC.2018.8422442.
[37] B. Shebaro, O. Oluwatimi, E. Bertino, Context-based access control systems for mobile devices, IEEE Trans. Dependable Secure Comput. 12 (2) (2015) 150–163, doi:10.1109/TDSC.2014.2320731.
[38] J. Tian, H. Zhang, D. Wu, D. Yuan, Qos-constrained medium access probability optimization in wireless interference-limited networks, IEEE Trans. Commun. 66 (3) (2018) 1064–1077, doi:10.1109/TCOMM.2017.2775239.
[39] M. Li, X. Zhao, H. Liang, F. Hu, Deep reinforcement learning optimal transmission policy for communication systems with energy harvesting and adaptive mqam, IEEE Trans. Veh. Technol. 68 (6) (2019) 5782–5793, doi:10.1109/TVT.2019.2911544.
[40] Y. Su, X. Lu, Y. Zhao, L. Huang, X. Du, Cooperative communications with relay selection based on deep reinforcement learning in wireless sensor networks, IEEE Sensors Journal (2019), doi:10.1109/JSEN.2019.2925719. 1–1
[41] C.J.C.H. Watkins, P. Dayan, Q-learning, Mach. Learn. (1992).
[42] S.J. Bradtke, M.O. Duff, Reinforcement learning methods for continuous-time markov decision problems, in Advances in Neural Information Processing Systems 7 (1995) 393-400.

Hui dingis currently pursuing the Ph.D. degree with the School of Information Science and Engineering, Shandong University, China. His research interests include heterogeneous network in 5G, resource allocation, user association and power control, non-convex optimization and reinforcement learning.

Feng Zhaoreceived his Ph.D. degree in communication and information system from Shandong University, China in 2007. He received his B.S. degree from Guilin University of Electronic Technology, China in 1997. From April 2008 to May 2011, he worked as postdoc in Beijing University of Posts and Telecommunications (part time). He was a visiting scholar at the University of Texas at Arlington from February to August 2013. He is currently a Professor with the Guangxi Colleges and Universities Key Laboratory of Complex System Optimization and Big Data Processing, Yulin Normal University, Yulin, China. His current research interests include cognitive radio networks, MIMO wireless communications, cooperative communications, and smart antenna techniques. His research has been supported by the National Science Foundation of China. Dr. Zhao has published more than 60 papers in journals and international conferences. He was awarded the second prize of Shandong province science and technology progress twice, in 2007,2012 and 2017, respectively.

Jie tian(S'12-M' 16) received the B.E. and M.E. degrees from Shandong Normal University, China, in 2008 and 2011, respectively, and the Ph.D. degree in communication and information systems from the School of Information Science and Engineering, Shandong University, China, in 2016. She is currently an Associate Professor with the School of Information Science and Engineering, Shandong Normal University, Jinan, China. She is a member of IEEE, IEEE Communications Society, and ACM. Her research interests include cross-layer design of wireless communication networks, intelligent radio resource management in heterogeneous networks, and signal processing for communications.

Dongyang Liis currently pursuing Ph.D. degree with the Department of Information and Communications Technology, the School of Information Science and Engineering, Shandong University. His research interests include wireless big data, wireless edge caching and deep learning.

Haixia Zhang(M'08-SM'11) received the B.E. degree from the Department of Communication and Information Engineering, Guilin University of Electronic Technology, China, in 2001, and received the M.Eng. and Ph.D. degrees in communication and information systems from the School of Information Science and Engineering, Shandong University, China, in 2004 and 2008. From 2006 to 2008, she was with the Institute for Circuit and Signal Processing, Munich University of Technology as an academic assistant. From 2016 to 2017, she worked as a visiting professor at University of Florida, USA. Currently, she works as full professor at Shandong University. She has been actively participating in many academic events, serving as TPC members, session chairs, and giving invited talks for conferences, and serving as reviewers for numerous journals. She is the associate editor for the International Journal of Communication Systems and IEEE Wireless Communication Letters. Her current research interests include cognitive radio systems, cooperative (relay) communications, resource management, space time process techniques, mobile edge computing and smart communication technologies.