Reinforcement Learning based Adaptive Handover in Ultra-Dense Cellular Networks with Small Cells

Qianyu Liu, Chiew-Foong Kwong, Wei Sun, Lincan Li, Haoyu Zhao

University of Nottingham
UK | CHINA | MALAYSIA

University of Nottingham Ningbo China, 199 Taikang East Road, Ningbo, 315100, Zhejiang, China.

First published 2020

University of Nottingham
UK | CHINA | MALAYSIA

Reinforcement Learning based Adaptive Handover in Ultra-Dense Cellular Networks with Small Cells

Qianyu Liu[a], Chiew-Foong Kwong*[b], Wei Sun[b], Lincan Li[b], Haoyu Zhao[c]
[a]University of Nottingham Ningbo China, International Doctoral Innovation Centre, Ningbo, China;
[b]University of Nottingham Ningbo China, Department of Electrical and Electronic, Ningbo, China;
[c]University of Nottingham, Department of Electrical and Electronic, Nottingham, UK;

ABSTRACT

The dense deployment of the small base station (BS) in fifth-generation commination system can satisfy the user demand on high data rate transmission. On the other hand, such a scenario also increases the complexity of mobility management. In this paper, we developed a Q-learning framework exploiting user radio condition, that is, reference signal receiving power (RSRP), signal to inference and noise ratio (SINR) and transmission distance to learn the optimal policy for handover triggering. The objective of the proposed approach is to increase the mobility robustness of user in ultra-dense networks (UDNs) by minimizing redundant handover and handover failure ratio. Simulation results show that our proposed triggering mechanism efficiency suppresses ping-pong handover effect while maintaining handover failure at an acceptable level. Besides, the proposed triggering mechanism can trigger the handover process directly without HOM and TTT. The respond speed of triggering mechanism can thus be increased.

Keywords: handover, reinforcement learning, ultra-dense networks

1. INTRODUCTION

The ultra-dense networks (UDNs) consisting of massive small base stations (BSs) is a promising approach to cope with the demand of mobile user for higher data transmission rate and broader bandwidth. On the other hand, the dense deployment of smalls BSs could increase the complexity of cellular networks and lead serious of new challenges. Handover management is one of the challenges, which has become one of the main barriers to overall network performance. During the movement of user equipment (UE), the UE needs to perform the handover process to enable seamless data connection. According to the third-generation partnership project (3GPP), the A3 event is defined as the triggering mechanism for UE handover in fourth (4G) [1] and fifth-generation communication system (5G) [2].

The triggering decision from the A3 event only relies on a single criterion known as reference signal received power (RSRP). As shown in Fig.1, the UE compares RSRP values between its serving and neighbouring BSs. The handover is triggered if UE's RSRP from neighbouring BS is higher than servicing BS and remain a specific pre-defined condition, that is, handover margin (HOM) and time to trigger (TTT). The HOM and TTT are used in A3 event to avoid unnecessary and frequent handover that incurred by noise and interference. However, the A3 event is initially developed for handover in macro BS. Since small BS with less coverage area and stronger inter-cell interference, simply apply A3 event in small BS could result in handover frequent occur between two BSs that is known as ping-pong effect. Moreover, with the increasing deployment of small BSs, the workload for configuration and optimization of HOM and TTT for BSs is also dramatically increased[3].

To address the challenges for handover in UDNs, many works try to adjust HOM and TTT dynamically based on different approaches. The work in [4] proposed a threshold comparison based approach to auto tune HOM and TTT on the basis of user speed and reference signal receiving power (RSRP). The proposed scheme can effectively increase mobility robustness of user in UDNs by minimizing frequent handover and handover failure ratio. Meanwhile, the authors in [5] also introduced a threshold comparison based approach to optimize HOM and TTT with the objective of mobility load balancing. Simulation results in [5] show that the proposed algorithm can provide a more balanced load among networks by considering network load status and load estimation.

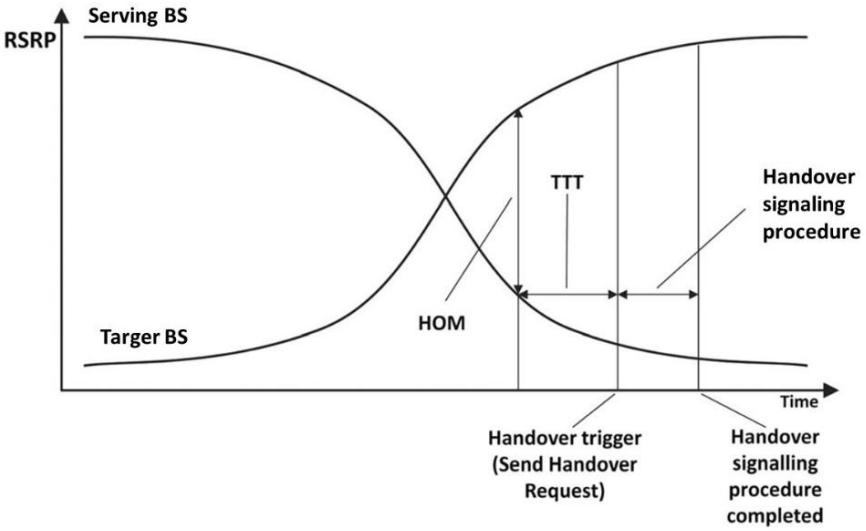

Figure 1 A3 event with HOM and TTT

On the other hand, since machine learning has robust learning and reasoning ability to enable an algorithm with the adaptive feature in many fields [6]–[9]. Some researches has also incorporated fuzzy logic or reinforcement learning to optimize handover parameter such as [10]–[15]. In [10], a fuzzy logic based adaptive handover optimization method was proposed. The user velocity, RSRP and reference signal receiving quality (RSRQ) are adopted as input for fuzzy logic to adjust HOM dynamically. The simulation results in [10] indicated that this fuzzy logic based solution could almost eliminate ping-pong handover and reduce unnecessary handover by comparing with the state of art algorithms. Some research has also incorporated reinforcement learning to optimize handover parameter such as [11]–[15]. The authors in [11], [12] developed a handover optimization algorithm based on Q-learning frameworks. The speed of UE is utilized as a state vector, and system key performance indicators (KPIs) such as throughput, latency and number of handovers are adopted to formulate reward function. Paper [13], [14] integrated both advantages of fuzzy logic and Q-learning into handover optimization. The handover ratio, call drop ration, and HOM are used as state vector in Q-learning to learn the fuzzy rules for fuzzy interference system. The fuzzy inference system is then considered handover ratio, call drop ratio and HOM as input to update HOM for each BS. The work in [15] proposed a reinforcement learning based handover policy to select optimal BS as a handover target for different UE density circumstances. The works in [11]–[15] shown that reinforcement learning could learn the characteristic from a different environment and obtain an optimal optimization policy for handover. The simulation results show that the proposed algorithm can minimize the number of handovers, ping-ping effect and call drop rate while improving system throughput.

Since the coverage of small BS is much lower than the macro BS, the residence time of users in a small cell is relatively short. The handover procedure system also needs to complete in a short moment. Under the policy of A3 event with HOM and TTT, the handover process is only executed after these two pre-defined conditions. The existence of HOM and TTT reduces the response speed of the algorithm and hence can easily lead to handover failure. In this paper, we aim to develop a handover triggering mechanism with the adaptive feature that can trigger the handover process directly without HOM and TTT. The proposed algorithm should be able to increase mobility robustness of user by minimizing the number of handover and ping-pong effect while retaining the handover failure rate at a low level. To achieve these objectives, the reinforcement learning framework is adopted to learn the optimal handover triggering policy from a different environment. The trained policy from reinforcement learning is used to select the most suitable triggering point for UE. The performance of the proposed algorithm is evaluated and compared with the current A3 event and other algorithms.

The remainder of this paper is organized as follows. Section 2 introduces the framework of Q-learning and formulation of the algorithm. Simulation environment and evaluation results are shown in Section 3. Finally, the paper is concluded in Section 4.

2. PROPOSED METHOD

In the proposed environment, there are three parameters, that is, RSRP, SINR, and transmission distance used to form the states vector for the reinforcement learning framework and determine the triggering point of handover.

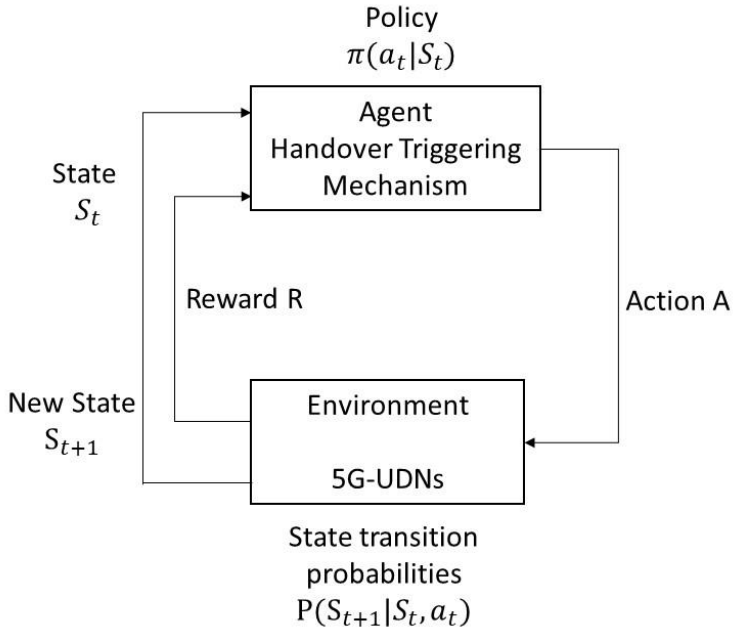

Figure 2 the framework of Q-learning based handover

One of reinforcement learning algorithm - Q-learning is adopted in this work. The core idea of Q-Learning is to acquire information from the environment and obtains feedback (reward) to improve the policy. The basic structure of Q-learning is shown at Fig.1. There are four elements in the Q-learning, including states, actions, a policy, and value functions. A policy, π, is a set of rules of performing an action for an agent in each state. The value of performing action a in state s under the policy π is denoted by $Q^{\pi}(s, a)$, which is also called action-value function or Q-value. The main objective of Q-learning is to lean optimal policy from the environment that can select the action with the highest value in each state to receive a maximum accumulated reward [16].

Due to the Q-learning only can store limited state-action pair, the state in this paper is defined by the combination of three indexed based on the level of UE's RSRP, SINR and transmission distance. The value of the input parameter will be normalized between 0 to 1 by Eqs 1 and 2.

$$Z_i = \frac{[x_i - min\{x_i\}]}{[max\{x_i\} - min\{x_i\}]} \quad (1)$$

$$Z_i = \frac{[max\{x_i\} - x_i]}{[max\{x_i\} - min\{x_i\}]} \quad (2)$$

where, x_i is the value of the parameter in the data set. Eq.1 is for benefit parameter (higher is better), and Eq.2 is for cost parameter (lower is better). The RSRP and SINR are the benefit parameter, and transmission criteria is a cost parameter.

Each parameter is divided into four levels with 0.25 as the interval and arranged in ascending order with 1-4 as the index. For example, if normalized RSRP, SINR and transmission distance equal to 0.2,0.7 and 0.3 receptivity, then the state will be represented by [1 3 2]. Moreover, the index in this state will be used to formulate reward value. For example, for the state [1 3 2], the reward value is the sum of index equal to 6. For each state, there are two actions "handover (HO)" or "maintain the current connection (NHO)" can be selected. If policy chooses action "handover" to perform at time step *t*,

the UE's connection will be transferred to a BS with highest SINR at the time step *t+1*. Otherwise, UE will maintain its connection with its current serving BS at time step *t+1*.

After define state, action and reward, the Q-value for each state-action pair is updated as,

$$Q_{t+1}(s_t,\ a_t) = Q_t(s_t,\ a_t) + \alpha \times \{r + \lambda \times \mathrm{Q}\ (s_{t+1},\ a_{t+1}) - Q_t(s_t,\ a_t)\} \tag{3}$$

where α is learning rate, λ is a discount factor, and r is the reward that received after a_t perform at s_t.

Where, the ϵ greedy policy is utilized in Q-learning to trade-off between exploration and exploitation to obtain the best strategy. In this project, the value ϵ is reduced from 1 to 0.1 from the beginning of each episode, which means that the Q-learning will pay more attention to exploration in the beginning. The learning stage of Q-learning is demonstrated in Table 1.

Table 1 Learning stage of Q-learning

Input: the data sets of RSRP, SINR and transmission distance
Output: Q table
Convert the input parameters to state
Initialize $\forall\ s \in$ S, $a \in$ A, Q (s, a) =0
A = [HO, NHO]
Initialize *number_of_points_measured m*
episodes=0
loop for *n* episodes (epochs) // Start learning
s=0
    time_in_episode=0
    loop for *m* time steps (points)
        Generate a random number δ *(0* $\leq\delta \leq$ *1) at the time step t*
        if $\delta > \epsilon$ **then** (ϵ-greedy policy)
            Select an action randomly
        else
            Select the action with the maximum q-value
        end if //choose action
        Move to new time step *t+1* and update Q-value Eq.3
        Determine the action at the next state //choose_action
        Update the Q table (Q (s, a) and q-values) //update
        time step+=1
    end
    episodes+=1
end

After the learning stage of Q-leaning, a table known as Q-table that store the Q-value for each state-action pair can be obtained. The trained Q-table is used to trigger the handover process for UE.

The measured RSRP, SINR and transmission distance will first be converted to state vector to find the corresponding state at Q-table. The UE will then select an action with the highest Q-value to perform. If action "handover" is selected, then a handover request will be sent by UE to its serving BS. The connection of UE will thus subsequently be transferred to target BS. If action "maintain the current connection" is chosen, then UE will keep a connection with its serving BS.

3. SIMULATION SETUPS AND RESULTS

3.1 Evaluation environment

A communication environment with the dense deployment of 16 small BSs is developed to evaluate the performance of the proposed handover algorithm. There are 16 small BSs are evenly placed in a square area with a side length of 1200 meters. There are 40 UEs randomly moving at a constant speed of 30 km/h. Three KPIs that is, the number of handovers, ping-pong handover ratio and handover failure ratio is adopted in this paper. The A3 event –RSRP based and fuzzy logic based handover algorithm are used as the competitive algorithms. Where, the A3 event only relies on RSRP to trigger the handover process, and fuzzy logic uses RSRP, SINR and transmission distance as inputs. The other simulation setups are shown in Table 2.

Table 2 Simulation Setup

Parameters	***Specification***
Carrier frequency (GHz)	28
Subcarrier spacing (KHz)	30
System bandwidth (MHz)	100
Physical resource block	275
Number of BSs	16
BS transmitted power (dBm)	35
Subcarriers per PRB	12
Duration of simulation	10000 s
Mobility model	Random direction
Number of UE	40
UE speed (km/h)	30
Type of noise	AWGN, Rayleigh
HOM	5dB
TTT	50ms

3.2 Simulation results

The first KPI in Fig.3 the number of handovers per UE during the entire simulation. The second KPI in Fig. 4 is the ping-pong handover ratio, which is used to quantify the occurrence of unnecessary handovers. The ping-pong handover will be detected if UE repeatedly triggers handover between two base stations within 10s. The triggering mechanism should minimize the number of unnecessary handovers while maintaining high service quality for the user. According to the Fig.1 and 2, the traditional RSRP based triggering mechanism has the highest number of handovers and ping-pong handover ratio. The A3 event –RSRP based only relies on a single metric, that is, RSRP in handover decision making. Due to the existence of noise and interference, RSRP always fluctuates, which will reduce the accuracy of handover decision and lead to many unnecessary handovers.

Compared with RSRP based approach, the fuzzy logic based handover triggering mechanism has a lower number of handovers and ping-pong handover rate. The fuzzy logic can effectivity incorporate multiple inputs to estimate a suitable triggering point. The triggering decision made by fuzzy logic is under the restriction of several fuzzy rules, and it can reduce unnecessary handovers. However, the reliability of fuzzy logic is difficult to guarantee, because its fuzzy membership function and fuzzy rules in fuzzy logic need to be designed based on the practical experience. It is hard to establish optimal membership function and rule for different application scenario.

The proposed Q-learning based handover triggering mechanism has the lowest number of handovers and ping-pong handover ratio. These results indicate that the Q-learning framework can effectivity learn the optimal handover policy based on the characteristic of the environment. The Q-learning framework could enable the proposed handover triggering mechanism with an adaptive feature to update its policy with the changes of environment. The adaptive feature allows the proposed method to select the best triggering point under noise and interference

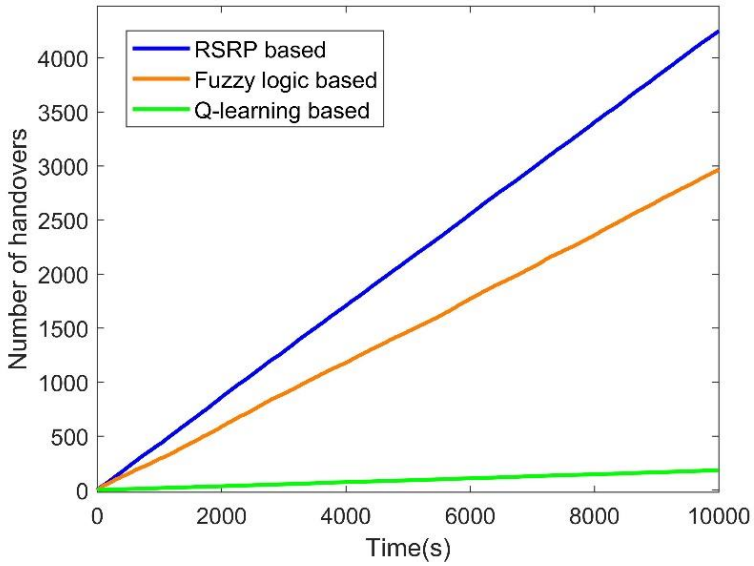

Figure 3 Number of handovers under different approaches

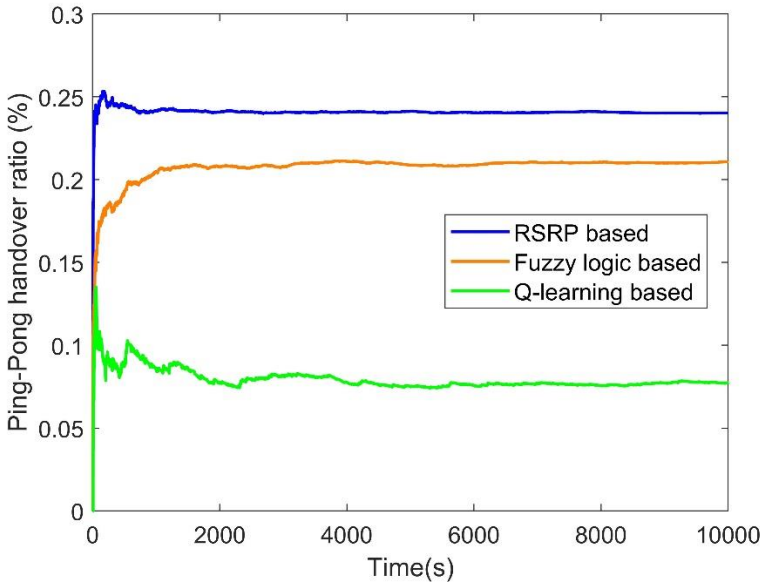

Figure 4 Ping-ping handover rate under different approaches

The last KPIs in Fig.4 is the handover failure ratio and used to test the reliability of the handover triggering algorithm. The handover failure is detected when SINR of UE is lower than a certain threshold, which occurs when the handover process is triggered too early or too late.

According to the Fig.4, the traditional RSRP based triggering mechanism has the lowest handover failure rate. The RSRP based is always select the BS with highest RSRP to connect. RSRP is the critical factor to affect SINR, and high SINR could ensure the link connection as well as the success of handover. Moreover, the HOM and TTT are set as 5dB and 50ms in this paper. Under these setups, the handover triggering condition can easily be satisfied.

The fuzzy logic based triggering mechanism has the highest handover failure rate. Under the restriction of several fuzzy rules, the handover decision is easily triggered late by fuzzy logic, resulting in a higher handover failure rate. In this condition, the fuzzy rules need to be carefully designed to balance the number of handover and handover failure rate.

The handover failure ratio of proposed Q-learning based triggering mechanism is around 1%, which is slightly lower than the RSRP- based approach. Due to Q-learning framework considers the multiple metrics as input, the weights of RSRP and SINR are weakened. Therefore, the handover failure of Q-leaning is slightly lower (0.5%) than the RSRP based approach, but it is still at a low level.

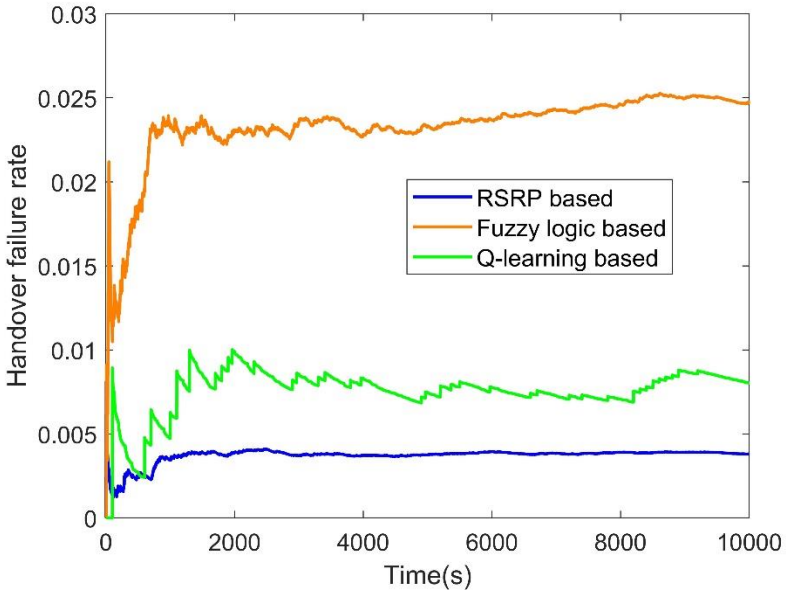

Figure 5 Handover failure rate under different approaches

4. CONCLUSION

In order to minimize the redundant handover and increase the mobility robustness of UE in UDNs, this paper adopted Q-learning framework to establish a triggering mechanism with the adaptive feature. The Q-learning framework can incorporate multiple parameters to learn the optimal handover triggering policy from the environment. Under the Q-learning framework, the proposed algorithm could adaptively update its triggering policy with the changes of environment. Simulation results show that the proposed triggering mechanism can reduce approximately 90% redundant handover caused by noise and interference. The proposed approach outperforms the other two competitive algorithms in terms of the number of handovers and ping-ping handover. In addition, the proposed method can also retain the handover failure rate at a low level.

REFERENCES

[1] the 3GPP Organizational Partners, *Evolved Universal Terrestrial Radio Access (E-UTRA),Radio Resource Control (RRC),Protocol specification,document TS 36.331*. 2018.

[2] the 3GPP Organizational Partners, *Radio Resource Control (RRC) protocol specification, document TS 38.331*. 2018.

[3] M. Kamel, S. Member, W. Hamouda, and S. Member, "Ultra-Dense Networks : A Survey," vol. 18, no. 4, pp. 2522–2545, 2019.

[4] A. Alhammadi and G. S. Member, "Auto Tuning Self-Optimization Algorithm for Mobility Management in LTE-A and 5G HetNets," *IEEE Access*, vol. 8, pp. 294–304, 2020.

[5] M. M. Hasan, S. Kwon, and J. H. Na, "Adaptive mobility load balancing algorithm for LTE small-cell networks," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 4, pp. 2205–2217, 2018.

[6] H. Lu, Y. Li, S. Mu, D. Wang, H. Kim, and S. Serikawa, "Motor anomaly detection for unmanned aerial vehicles using reinforcement learning," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2315–2322, 2018.
[7] Y. Zhang, Y. Li, R. Wang, M. S. Hossain, and H. Lu, "Multi-Aspect Aware Session-Based Recommendation for Intelligent Transportation Services," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–10, 2020.
[8] H. Lu, Y. Li, M. Chen, H. Kim, and S. Serikawa, "Brain Intelligence: Go beyond Artificial Intelligence," *Mob. Networks Appl.*, vol. 23, no. 2, pp. 368–375, 2018.
[9] H. Lu, Q. Liu, D. Tian, Y. Li, H. Kim, and S. Serikawa, "The Cognitive Internet of Vehicles for Autonomous Driving," *IEEE Netw.*, vol. 33, no. 3, pp. 65–73, 2019.
[10] K. Da Costa Silva, Z. Becvar, and C. R. L. Frances, "Adaptive Hysteresis Margin Based on Fuzzy Logic for Handover in Mobile Networks with Dense Small Cells," *IEEE Access*, vol. 6, no. c, pp. 17178–17189, 2018.
[11] A. Abdelmohsen, M. Abdelwahab, M. Adel, M. Saeed Darweesh, and H. Mostafa, "LTE handover parameters optimization using Q-learning technique," *Midwest Symp. Circuits Syst.*, vol. 2018-Augus, no. 4, pp. 194–197, 2019.
[12] T. Goyal and S. Kaushal, "Handover optimization scheme for LTE-Advance networks based on AHP-TOPSIS and Q-learning," *Comput. Commun.*, vol. 133, no. September 2018, pp. 67–76, 2019.
[13] J. Wu, J. Liu, Z. Huang, and S. Zheng, "Dynamic fuzzy Q-learning for handover parameters optimization in 5G multi-tier networks," *2015 Int. Conf. Wirel. Commun. Signal Process. WCSP 2015*, 2015.
[14] P. Munoz, R. Barco, J. M. Ruiz-Aviles, I. De La Bandera, and A. Aguilar, "Fuzzy rule-based reinforcement learning for load balancing techniques in enterprise LTE femtocells," *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 1962–1973, 2013.
[15] Y. Sun, G. Feng, S. Qin, Y. C. Liang, and T. S. P. Yum, "The SMART Handoff Policy for Millimeter Wave Heterogeneous Cellular Networks," *IEEE Trans. Mob. Comput.*, vol. 17, no. 6, pp. 1456–1468, 2018.
[16] Richard S. Sutton， Andrew G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, Massachusetts: The MIT Press, 2018.