

# AI-Based Optimization Method for Coverage of IoT Upload Link Communication

Sheng Ye \*

School of Information Technology & Engineering, Guangzhou College of Commerce, Guangzhou 511363, China

## Abstract

To address the degradation in uplink communication quality of the Internet of Things (IoT) caused by long-distance signal attenuation, multipath interference, and dynamic environmental changes, this paper proposes an artificial intelligence (AI)-based optimization method for IoT uplink communication coverage. First, service sampling points, base stations, and antenna parameters are defined within a target area, and a signal propagation model is constructed. The Sigmoid function is then used to smoothly evaluate service quality, and the coverage optimization problem is formulated as an objective function that maximizes coverage. The Q-learning algorithm is employed to achieve dynamic optimization by designing the state space (e.g., coverage rate, number of devices), action space (e.g., antenna parameter adjustments), and reward function (e.g., coverage improvement). Through iterative Q-value updates, the base station controller adaptively adjusts antenna configurations to balance coverage and interference, thereby improving long-term communication performance. To further enhance optimization efficiency, prioritized experience replay and an adaptive exploration mechanism are introduced to dynamically adjust the exploration–exploitation trade-off during learning, accelerating convergence and improving adaptability in dynamic IoT environments. Experimental results show that the proposed method significantly improves the communication performance of the uplink: the number of complete sensor data upload packets increases by 20%–25%, the number of transmission interruptions decreases by 70%–80%, and the average communication duration is reduced by 30%–40%, verifying the effectiveness and robustness of the method.

**Keywords:** artificial intelligence, internet of things, upload link, communication coverage, coverage optimization, reinforcement learning

Received on 09 May 2026, accepted on 02 June 2026, published on 10 June 2026

Copyright © 2026 Sheng Ye *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](#), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetsis.12965

## 1. Introduction

With the widespread application of IoT technology in fields such as industrial manufacturing and smart agriculture [1], massive amounts of data generated by terminal devices must be transmitted in real time to cloud or edge servers for analysis and processing [2]. However, in complex real-world scenarios, IoT devices often face unstable communication links and uneven coverage due to

factors such as long-distance signal attenuation [3], multipath interference, building obstructions, and dynamic environmental changes [4]. These issues not only reduce data transmission success rates but may also cause communication interruptions, leading to the loss of critical business data [5], which severely impacts the reliability and real-time performance of IoT systems.

Existing research has made progress in uplink link optimization on the Internet of Things, such as antenna parameter adjustment based on contextual linear multi-

\*Corresponding author. Email: yesheng2110171@163.com

armed bandits [6], RIS-assisted beamforming optimization, and other methods, but there are limitations in dynamic environment adaptability and coverage balance. Vannella et al. [7] used contextual linear multi-armed bandits to control antenna tilt angles, transforming the optimization problem into a probabilistic recognition problem, clarifying the meanings of actions, context, and rewards, and designing algorithms to generate optimal strategies with fewer samples, but with poor adaptability to dynamic environments, making it difficult to continuously optimize communication coverage. Saqib et al. [8] employed RIS-assisted 6G THz communication, optimizing RIS placement, performing 3D hybrid beamforming and phase adjustment, achieving signal coherent superposition to enhance strength, but the superposition effect of RIS reflected signals is affected by user device location distribution, leading to potential coverage imbalance. Benatia et al. [9] deployed edge sensor nodes using multi-objective evolutionary algorithms, first exploring a single solution to generate an initial population, then integrating into a population-based heuristic to find the optimal solution, but single-solution exploration struggles to cover all combinations and may cause the search to fall into local optima. Dhinesh et al. [10] combined MADDPG with Bayesian optimization to optimize UAV trajectories, but Bayesian optimization requires prior knowledge, and inaccurate priors can cause the search to deviate from the correct direction, making it difficult to find the optimal trajectory.

Uplink link communication on the Internet of Things serves as a channel for terminal devices to transmit data to the cloud or edge servers. It features heterogeneity and diversity [11], and signals are affected by geographic environment, weather, and device mobility, with significant fluctuations in communication quality. Therefore, optimization methods need to possess environmental awareness, adaptive decision-making, and distributed computing capabilities [12]. Artificial intelligence can significantly enhance optimization effects through adaptive decision-making and distributed computing. Among them, reinforcement learning provides efficient solutions for coverage optimization in complex scenarios through real-time decision-making, global resource allocation, and long-term performance optimization [13]. Therefore, an AI-based uplink communication coverage optimization method is proposed. It interacts with the environment through an agent to dynamically adjust coverage strategies, including the direction and downtilt angles of base station antennas and transmission power. This makes signals more concentrated over target areas, reduces attenuation and interference during propagation, and thereby improves uplink signal quality.

## 2. Uplink Link Communication Coverage Optimization on the Internet of Things

### 2.1. Modeling of Uplink Link Communication Coverage Optimization Problem on the Internet of Things

The uplink is the link through which access terminals (e.g., mobile phones) send signals to the base station [14]. The configuration of antenna parameters affects the signal strength and coverage area of the uplink. For example, increasing the antenna's transmission power or adjusting its downtilt angle can enhance the uplink signal strength, thereby expanding the coverage area [15]. Therefore, the coverage optimization problem for IoT upload link communication can be improved by adjusting the antenna parameters and transmission power [16].

Within the given area  $Y$ , the coverage quality of the IoT terminal upload link is optimized. This area contains  $M_o$  service sampling points  $O_i (1 \leq i \leq M_o)$ ,  $M_b$  base stations  $BS_n (1 \leq n \leq M_b)$ , and  $M_{an}$  antennas  $AN_j (1 \leq j \leq M_{an})$ . Sampling points can be selected through uniform grid or (pseudo) Monte Carlo methods, or via targeted sampling based on IoT service requirements (e.g., sensor data uploads, device status feedback) [17]. The azimuth angle  $\alpha_{AN_j}$  of antenna  $AN_j$  represents the angle of clockwise rotation from true north to the antenna's normal direction, while the downtilt angle  $\theta_{AN_j}$  indicates the angle of downward rotation in the horizontal plane towards the antenna's normal.

Within the area to be optimized in the IoT communication network [18], sufficient sampling points are selected according to IoT service deployment requirements, scenario models, and geographic features [19], to evaluate service quality requirements and determine the proportion of sampling points that meet these requirements. Coverage  $Co$  is used to assess the coverage quality of the area, defined as the proportion of sampling points satisfying service needs, as expressed in formula (1):

$$Co = \frac{\sum_{i=1}^{M_o} Co_i}{M_o} \quad (1)$$

Among them,  $Co_i$  indicates the degree of service satisfaction at the business sampling point  $O_i$ . IoT services typically require high reliability [20] (signal strength and interference levels meet thresholds), and wide coverage (serving as many areas as possible to meet quality requirements) [21]. Therefore, coverage  $Co$  is a direct indicator of service satisfaction, with the optimization goal being to maximize  $Co$ . The azimuth angle  $\alpha_{AN_j}$  and downtilt angle  $\theta_{AN_j}$  of the base station antenna influence

the coverage effect at sampling points  $Co_i$  through the signal propagation model. Consequently, the problem of optimizing coverage parameters for uplink communication in IoT networks can be addressed by adjusting the azimuth angle  $\alpha_{AN_j}$  and downtilt angle  $\theta_{AN_j}$  of the base station antenna to maximize coverage  $Co$ , as shown in formula (2):

$$Co = \max_{\alpha_{AN_j}, \theta_{AN_j}} \frac{\sum_{i=1}^{M_o} Co_i}{M_o} \quad (2)$$

Among them,  $j = 1, 2, 3, \dots, M_{an}$ .

The uplink communication quality at a sampling point can be jointly evaluated based on the strongest RF signal strength from surrounding base stations and the interference from other base stations [22], using the combined metrics of Reference Signal Received Power (RSRP) and Signal to Interference plus Noise Ratio (SINR)

as the evaluation standard for coverage effectiveness  $Co_i$ , as shown in formula (3):

$$Co_i = g^{RS_i} \times g^{SI_i} \quad (3)$$

Where,  $g^{RS_i}$  represents an indicator function based on RSRP( $RS_i$ ), defined as in formula (4):

$$g^{RS_i} = Si(RS_i - \tau^{RS}) \quad (4)$$

$g^{SI_i}$  represents an indicator function based on SINR( $SI_i$ ), defined as in formula (5):

$$g^{SI_i} = Si(RS_i - \tau^{SI}) \quad (5)$$

In Equations (4) and (5),  $\tau^{RS}$  and  $\tau^{SI}$  represent the threshold conditions for RSRP and SINR respectively, and  $Si$  represents the Sigmoid function. The evaluation criterion is that only when both RSRP and SINR meet the threshold conditions simultaneously, the radio frequency signal at the sampling point is considered to meet the requirements for the communication coverage optimization of the uplink. The sigmoid function is adopted to evaluate the service quality level smoothly: the larger the input parameter relative to the threshold, the closer the output of the sigmoid function is to 1; if the input parameter is smaller than the threshold, the output approaches 0. In order to better cope with the differences of IoT services and carry out refined optimization, different threshold requirements for signal indicators are allowed at different positions, so as to face the application scenarios of different IoT nodes such as roads, buildings, and green spaces.

The  $RS_i$  of the sampling point  $o_i$  is shown in Equation (6):

$$RS_i = \max_{1 \leq j \leq M_{an}} (RS_{i,j}) \quad (6)$$

In the formula,  $RS_{i,j}$  represents the received signal strength indicator at sampling point  $o_i$  to the antenna  $AN_j$ , obtained through equation (7):

$$RS_{i,j} = P_j + G_{i,j} - Pl_{i,j} \quad (7)$$

In the formula,  $P_j$  represents the transmit power of antenna  $AN_j$ ,  $G_{i,j}$  represents the directional gain of antenna  $AN_j$  towards sampling point  $o_i$ , and  $Pl_{i,j}$  represents the path loss between antenna  $AN_j$  and sampling point  $o_i$ .

Path loss data can be obtained through ray tracing, empirical models such as COST-231, or measured path loss data. The  $SI_i$  of sampling point  $o_i$  is expressed by equation (8):

$$SI_i = \frac{1}{NO_i + RS_i + \sum_{j=1}^{M_{an}} RS_{i,j}} RS_i \quad (8)$$

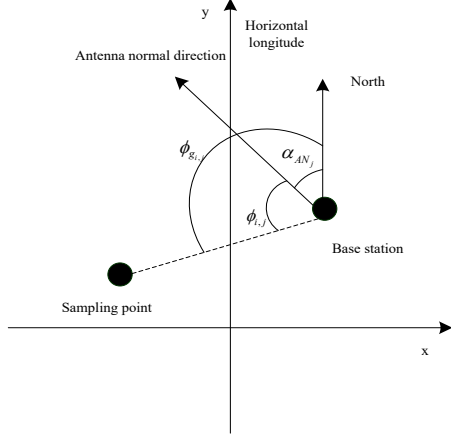
Where,  $NO_i$  indicates the noise at sampling point  $SI$ .

$G_{i,j}$  uses a real antenna model, with the horizontal and vertical gains obtained through actual measurements, and the expressions for the horizontal and vertical gain fitting are shown in equation (9). The interpolation process is as follows.

$$G_{i,j}(\phi_{i,j}, \varphi_{i,j}) = F_1(\phi_{i,j}) - \left[ \frac{|\phi_{i,j}|}{\pi} \times (F_1(\pi) - F_2(\pi - \varphi_{i,j})) + (1 - \frac{|\phi_{i,j}|}{\pi}) \times (F_1(0) - F_2(\varphi_{i,j})) \right] \quad (9)$$

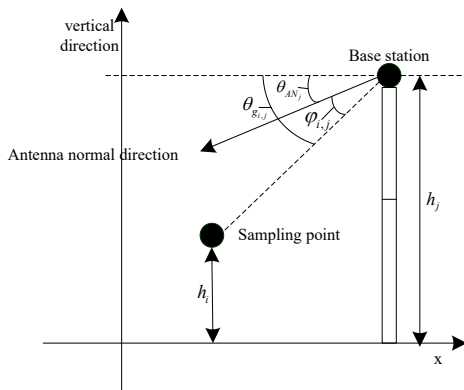
The functions  $F_1$  and  $F_2$  are Fourier fits of the horizontal angle  $AN_j$  and vertical angle  $o_i$  gains from antenna  $\phi_{i,j}$  to sampling point  $\varphi_{i,j}$ , and  $\pi$  is circumference ratio.

The horizontal angle  $\phi_{i,j}$  is formed between the angle  $\phi_{g,i,j}$  formed between geographic coordinates of antenna  $AN_j$  and the geographic coordinates of sampling point  $o_i$ , as well as the antenna directional angle  $\alpha_{AN_j}$  of the antenna, as shown in Figure 1.



**Figure 1.** Display of antenna direction angle

The vertical angle  $\varphi_{i,j}$  is formed between the angle  $\theta_{g_{i,j}}$  created by the hanging height of antenna  $AN_j$  and the height  $h$  of sampling point  $o_i$ , and the downtilt angle  $\theta_{AN_j}$  of antenna  $AN_j$ , as shown in Figure 2.



**Figure 2.** Vertical angle display diagram

## 2.2. Optimization of Uplink Communication Coverage Based on Reinforcement Learning

Wireless Internet of Things (IoT) device upload link often faces dynamic environmental challenges such as signal attenuation, interference, and device mobility [23]. Antenna parameters directly affect signal coverage and interference intensity. For example, a too-small downtilt angle can cause signals to cover too far, leading to neighboring cell interference; a too-large downtilt angle can create coverage holes. Reinforcement learning

interacts with the environment (IoT communication scenarios) through an agent (such as a base station controller), and dynamically adjusts strategies based on reward signals (which measure the improvement in coverage performance) [24], enabling adaptive decision-making for coverage optimization. By real-time adjustment of parameters such as azimuth and downtilt angles, it balances coverage and interference dynamically. Therefore, using Q learning as the algorithm for coverage optimization of IoT upload links.

Reinforcement learning is a branch of machine learning within artificial intelligence, emphasizing how to make actions based on environmental states [25]. Reinforcement learning tasks are usually described using a Markov Decision Process (MDP): the agent (such as a base station controller) exists within an environment  $H$ , with a state space of  $Z$ . Each state (base station parameters, device status) describes the environment perceived by the agent (IoT communication scenario, such as base stations and devices). The set of actions the base station controller can take (used to adjust communication parameters) forms the action space  $D$ . If a certain action  $d_i \in D$  is applied to the current state  $Z_i$ , then the potential transition function  $L$  causes the IoT communication scenario (e.g., base stations and devices) to transition from the current state (base station parameters, device status) to another state according to some probability. Meanwhile, the environment provides feedback to the base station controller in the form of a reward (coverage performance) based on the reward function.

The reinforcement learning task corresponds to a quadruple  $H = \langle Z, D, L, K \rangle$ , where  $L : Z \times D \times Z \mapsto \Psi$  specifies the transition probabilities of base station parameters and device states, and  $K : Z \times D \times Z \mapsto \Psi$  specifies the reward (coverage performance). In reinforcement learning, the goal is to find a policy that maximizes the long-term cumulative reward, i.e., maximizes coverage range. That is, when the base station controller adopts the optimal policy to adjust base station parameters  $\sigma$  and executes actions, it can achieve maximum upload link communication coverage performance, as shown in formula (10).

$$K = \sum_{t=0}^{\infty} \zeta^t k_{t+1} \quad (10)$$

Where  $\zeta^t$  represents the decay value of future rewards at time slot  $t$ , and  $k_{t+1}$  represents the instantaneous reward at time slot  $t$ .

Q-learning is a value-based reinforcement learning algorithm that determines the next action by evaluating the Q value of actions. The state-action value function  $Q_{\sigma}(z, d)$  indicates the value of executing action  $d_i$  in the current state  $z_i$ , while following the base station parameter adjustment strategy  $\sigma$  until the uplink

communication coverage optimization is completed. The state-action value function obtained by the base station controller during this process is shown in formula (11).

$$Q_{\sigma}(z, d) = E_{\sigma}[J_t | Z_t = z, D_t = d] \quad (11)$$

In the formula:  $Z_t$  represents the current base station state;  $D_t$  is the action of adjusting the base station parameters;  $z$  is the state within the state space  $Z$ ;  $d$  is the action within the action space  $D$ ;  $\sigma$  is the policy;  $E$  is the expectation. The optimal state-action value function refers to the maximum value when taking the best action in all decision-making processes. By iteratively updating, the state-action value function converges to the optimal coverage optimization strategy, as shown in formula (12).

$$\sigma(d | z) = \arg \max_d Q_{\sigma}(z, d) \quad (12)$$

Store the state action value function in the Q table and continuously update the Q table, as shown in formula (13):

$$Q_{\sigma}(z, d) \leftarrow Q_{\sigma}(z, d) + \eta k(z, d) + \eta \left( \delta \max_{d'} Q_{\sigma}(z', d') - Q_{\sigma}(z, d) \right) \quad (13)$$

Where,  $k(z, d)$  indicates the reward in the current state  $d_t$  when using action  $z_t$ ;  $\eta$  represents the learning rate;  $\max_{d'} Q_{\sigma}(z', d')$  indicates the maximum Q value in state  $z'$ .

For the Q-learning problem of uplink communication coverage optimization in IoT networks, the mapping is divided into four aspects: state space, action space, reward function, and Q-value update. The following explanations are provided in sequence.

### State Space

When applying the Q learning algorithm to the uplink communication coverage optimization scenario in the Internet of Things network, certain settings and constraints are required for the base station and device state spaces. The Q learning state reflects the current network performance status. In the uplink coverage optimization scenario, the base station controller selects one of multiple regions within the area as the optimization target each time and adjusts its antenna parameters. Considering the characteristics of IoT devices [26], the adjusted communication coverage, the region being adjusted in this iteration, and the number of IoT devices within the current region are used as the Q learning states. The state space  $Z$  describing the uplink communication coverage optimization problem can be expressed by formula (14).

$$Z = \{B_i, Co_i, Co_{al}, zc_{p_b}\} \quad (14)$$

In the formula:  $B_i$  represents the current adjustment region number;  $Co_i$  is the local coverage rate corresponding to that region;  $Co_{al}$  is the global coverage

rate of the entire area to be optimized,  $zc_{p_b}$  is the number of IoT devices within the current region.

### Action Space

The optimization problem involving the communication coverage of the IoT upload link is a multi-dimensional optimization problem including the tilt angle of the regional antenna and the antenna transmission power. Therefore, the set  $D$  of actions for the IoT upload link is defined as in formula (15):

$$D = \{d_i, d_{\theta}, d_{p_j}, d_{p_b}\} \quad (15)$$

In the formula:  $d_i$  is the region number selected for adjustment;  $d_{\theta}$  is the adjustment of the antenna  $AN_j$  tilt angle;  $d_{p_j}$  is the adjustment of the antenna transmission power;  $d_{p_b}$  is the transmission power adjustment for IoT devices, which can be expressed as  $\{increase, keep, decrease\}$ , representing three actions: increase, maintain current state, and decrease. Adjusting the tilt angle can change the coverage area and direction of the signal, thereby optimizing the coverage for IoT devices. Adjusting the transmission power can influence signal strength and coverage distance, but the impact on IoT device power consumption must be considered. For some low-power IoT devices, appropriately adjusting their transmission power may help improve communication quality, but this should be done while ensuring device battery life.

### Reward Function

Q Learning, as a classic model of reinforcement learning, aims to construct a  $Q(z, d)$  value table, which reflects the benefits of choosing different actions under each base station parameter state. A higher value indicates a higher benefit from that action. The reward signal is derived from formulas (10) and (11).

The reward signal evaluates the previous operation of the base station controller based on the environment  $H$ . The goal of optimizing the communication coverage of the IoT upload link is to improve the coverage rate of the base station. If  $Co(t)$  indicates the current coverage rate,  $Co(t-1)$  indicates the coverage rate at the previous moment, and  $m$  the reward value, then the calculation of the reward function  $K_e$  can be divided into the following cases.

If  $Co(t) - Co(t-1) > 0$ , representing an increase in global coverage rate, indicates that the current region's base station parameters are adjusted positively and

effectively. In this case, a relatively large positive reward should be fed back to the agent, as in formula (16):

$$K_e = y_1 \quad (16)$$

If  $Co(t) - Co(t-1) = 0$ , when the global coverage remains unchanged after one iteration, a small negative reward should also be fed back to the agent, indicating that the current base station parameter adjustment is invalid, with the reward as in formula (17):

$$J_e = -1 \quad (17)$$

If  $Co(t) - Co(t-1) < 0$ , indicating a decrease in global coverage, it means the current base station parameter adjustment is a negative error adjustment, and a negative reward should be fed back to the agent, as in formula (18):

$$J_e = -m \quad (18)$$

Such a setting can shorten the convergence time of the uplink communication coverage optimization algorithm. In the design of the reward function, the hyperparameters are set as follows: when the global coverage increases (equation (16)), the positive reward value is set to +10; When the coverage rate remains unchanged (equation (17)), a small negative reward of -1 is given; When the coverage rate decreases (equation (18)), the negative reward is -10. The difference in reward magnitude aims to guide the agent to prioritize actions that can improve coverage, while avoiding excessive punishment for ineffective adjustments. The learning rate  $\alpha$  is initially set to 0.1 and decays during the training process at  $\alpha = 0.1 / (1 + 0.01 \times \text{episode})$  to balance convergence speed and stability. The discount factor  $\gamma$  is set to 0.9, allowing the agent to balance short-term and long-term coverage benefits during the optimization process. The above parameter values are determined based on the grid search of the pre experiment, and the specific sensitivity analysis is shown in the experimental section (Figure 4).

### Q-value Update

In the initial stage, each value in the Q table is randomly assigned. When the base station controller takes an action and executes it, a corresponding reward is generated. The agent updates the values in the Q table based on this reward. The values in the Q table can be updated according to the value function (coverage performance), which can be expressed as:

$$Q_{t+1}(z, d) = (1 - \eta_t) Q_t(z, d) + (\eta_t k_t + \eta_t \zeta \max_{d'}(z', d')) \quad (19)$$

Where:  $Q_t(z, d)$  is the value function for the state-action pair at time  $t$ ;  $\eta_t$  is the learning rate, used to control the speed of learning; the larger its value, the faster the convergence, but it may prevent obtaining the optimal solution; if there is  $k_t$  representing the current instantaneous reward in Q learning, then  $(k_t + \zeta \max_{d'}(z', d'))$  is the main content for updating coverage performance.

## 2.3. Optimization of Convergence Mechanism

In reinforcement learning, the agent begins decision-making by exploring the environment with a high probability, and as more sample data accumulates in the experience pool, the agent utilizes this experience data to maximize rewards based on known information. The  $\epsilon$ -greedy algorithm combines exploration of unknown rewards and exploitation of known information, showing promising application prospects in reinforcement learning, and can better search for higher returns. To improve the learning efficiency of the base station controller and enhance the effectiveness of adjusting base station antenna parameter action strategies, strategies such as combining upper confidence bounds with decayed  $\epsilon$ -greedy exploration mechanisms and prioritized experience replay are adopted to optimize the learning process, thereby improving the efficiency of the base station controller in learning uplink communication coverage optimization and adjusting base station antenna parameter strategies.

### Prioritized Experience Replay

In the parameter adjustment decision process, the base station controller generates a set of state (base station parameters) transition tuples as training samples and stores them in the experience pool. The prioritized experience replay method allocates sampling priorities based on the TD error magnitude of each state transition tuple, i.e., the larger the TD error  $TD_{er}$ , the higher the priority  $V(i)$ .

This method improves learning efficiency by selecting effective actions, i.e., experiences with larger errors for learning, and the sampling probability is proportional to the TD error, as shown in formula (20):

$$V(i) \propto |TD_{er}|^\zeta \quad (20)$$

Where,  $\zeta$  is a hyperparameter used to set the distribution shape.

### Adaptive Action Exploration Mechanism

The dynamic changes in the IoT environment, such as user device mobility and base station failures, directly alter the system's state and action effects. For example, user device movement may cause coverage in areas with good signal to deteriorate, and base station failures can interrupt communication in related regions. The Q learning algorithm makes decisions based on fixed Q value tables and strategies. When the environment changes, the existing Q values and strategies may no longer be applicable, leading to performance degradation. An adaptive mechanism is introduced to promptly adjust the Q value table and strategy according to environmental changes.

To increase exploration in the early stages of training and improve convergence in the later stages, a decaying  $\epsilon$ -greedy strategy combined with upper confidence bounds is used as the exploration mechanism for the agent. When selecting actions, a decaying upper confidence bound (UCB) strategy is used, and when calculating the confidence of the UCB formula, the decaying factor of  $\epsilon$ -

greedy is used to adjust the randomness of the exploration behavior. Specifically, as follows:

1) Strategy for selecting actions to adjust base station antenna parameters: Use the UCB algorithm for action selection, as shown in formula (21):

$$D = \arg \max_{ac} \left[ K(D) + \frac{\mu^* \sqrt{\ln(s) N(D)}}{N(D)} \right] \quad (21)$$

Where,  $K(d)$  is the current average reward (coverage performance) of action  $D$ ,  $N(d)$  is the number of times action  $D$  has been selected,  $\mu$  is the coefficient balancing exploration and exploitation,  $\frac{\ln(s)}{N(d)}$  reflects the uncertainty of action  $D$ , and  $s$  is the total number of attempts.

2) Setting of the decay factor: When calculating the balance factor in the UCB formula, set the decay factor  $\zeta_s$  to vary over time, as shown in formula (22):

$$\zeta_t = \frac{1}{1 + \nu * t} \cdot \zeta_0 \quad (22)$$

Where,  $\zeta_0$  is the initial decay factor value, and  $\nu$  is the decay rate.

Combining the above two points, the specific scheme and formula (23) for the  $\varepsilon$ -greedy and UCB strategies are as follows:

$$D = \arg \max_d \left[ J(d) + \zeta_s^* \sqrt{\frac{\ln(s)}{N(d)}} \right] \quad (23)$$

This allows reinforcement learning to attempt different parameter combinations with high probability (high  $\zeta$  value) in the early stages, exploring configurations that may improve coverage performance; later, it mainly relies on experience to make selections, choosing the most optimal parameter combinations from historical data with low probability (low  $\zeta$  value), thereby stabilizing the optimization of coverage performance, achieving a balance between exploration and exploitation, effectively addressing the complexity and dynamism of IoT environments, and enhancing coverage performance and communication reliability.

To ensure the convergence stability of the proposed Q-learning framework in dynamic IoT environments, this paper conducts theoretical analysis from the following aspects: firstly, the Bellman update formula (equation (13)) of Q-learning converges to the optimal state action value function with probability 1 under the given learning rate  $\alpha \in (0,1)$  and satisfying the conditions of  $\sum \alpha_t = \infty$  and

$\sum \alpha_t^2 < \infty$ . Secondly, the priority experience replay mechanism accelerates the learning speed of key samples

by prioritizing the sampling of samples with larger TD errors, while maintaining the boundedness of the experience distribution and avoiding variance divergence. In addition, the exploration strategy combining decay  $\varepsilon$ -growth and UCB (equation (23)) ensures sufficient exploration in the initial stage and gradually reduces the exploration rate in the later stage, making the action selection strategy asymptotically approach the deterministic optimal strategy. The above mechanisms collectively ensure the convergence of the algorithm in a finite state action space and stabilize the coverage performance over long-term iterations.

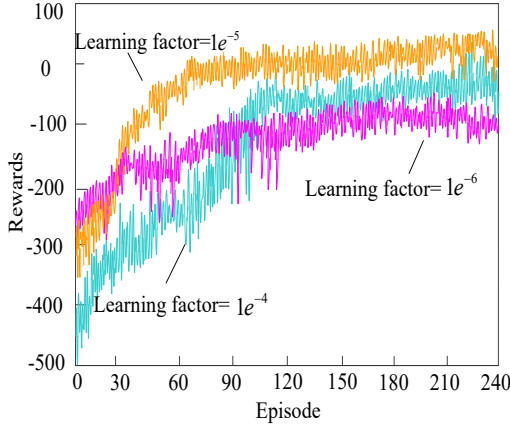
### 3. Experimental Results and Analysis

The experimental site selects an intelligent IoT agricultural field with an area of about 30,000 square meters as the research object. There are a total of 3 base stations in the central position and edge area of the farmland, with a communication radius of 300 meters. The base stations are all equipped with antennas with adjustable electronic downtilt angles (adjustment range:  $0^\circ - 15^\circ$ ) and transmit power (adjustment range: 10 - 30 dBm). A variety of types of IoT sensors are evenly arranged in this farmland, including soil moisture sensors, temperature sensors, light sensors, etc., totaling 80. In order to ensure the comprehensive and accurate collection of the farmland's environmental information and show the differences in communication coverage faced by sensors at different positions when uploading data. In relatively sparsely planted areas of the farmland, sensor spacing is set to 30–40 meters, whereas in more densely planted areas, sensor spacing is set to 20–30 meters. The terrain in this farmland is not flat; there are a small number of mounds, gullies, and buildings such as greenhouses and irrigation equipment rooms, which will cause signal blocking and reflection, affect the communication coverage of the IoT uplink, and provide valuable data for the experiment. The greenhouse environment in the agricultural field is shown in Figure 3.



Figure 3. IoT Agricultural greenhouse

To evaluate the impact of the proposed method on the convergence and stability of IoT uplink communication coverage optimization in the experimental farmland, the experiment sets different learning factors (respectively  $1e^{-4}$ ,  $1e^{-5}$ ,  $1e^{-6}$ ) and observes the trend of rewards (Rewards) over training episodes (Episode) under different learning factors, as shown in Figure 4.



**Figure 4.** Trend of reward value changes under different learning factors

Trend of reward values under different learning rates ( $1e^{-4}$ ,  $1e^{-5}$ ,  $1e^{-6}$ ). The horizontal axis represents training episodes (0 to 240), and the vertical axis represents the average reward per episode. The curve with  $1e^{-5}$  stabilizes earliest (around episode 70) and achieves the highest final reward (approximately 85), while  $1e^{-4}$  and  $1e^{-6}$  require more episodes ( $\geq 120$ ) to converge and yield lower final rewards ( $\approx 75$  and  $70$ , respectively). This indicates that a moderate learning rate ( $1e^{-5}$ ) balances convergence speed and stability.

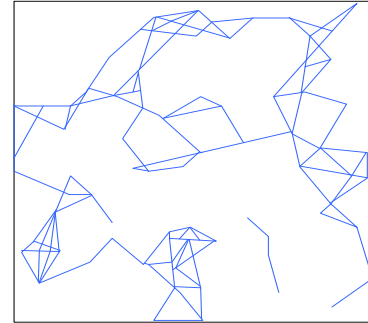
From Figure 4, it can be observed that as the number of training rounds increases, the reward value generally exhibits certain fluctuations and variations, indicating that the model continuously adjusts during training, and the reward value fluctuates accordingly. Through the three curves of different colors, the impact of different learning factors on the reward value can be intuitively compared.

When the learning factor is  $1e^{-5}$ , the fluctuation range and variation of the reward value differ from the other two learning factors. The curve with a learning factor of  $1e^{-5}$  shows signs of stabilization at an earlier stage (around 70 rounds) and ultimately reaches a higher reward value. In contrast, the curves with learning factors of  $1e^{-4}$  and  $1e^{-6}$  only gradually stabilize after 120 rounds and have overall lower reward values.

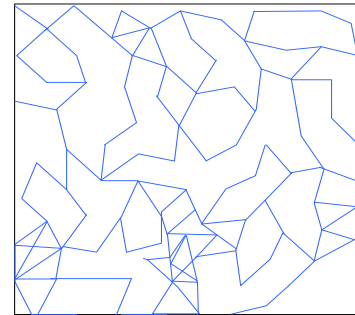
From this, it can be concluded that there is no positive linear correlation between the optimal value of the curve

and the size of the learning rate. Setting the learning factor too small in the optimization method requires more training rounds to reach convergence; setting it too large prevents the curve from converging to a good value and also affects training stability. Therefore, this paper’s experiment chooses to optimize the communication coverage of the uplink chain in the IoT network within the experimental farmland under the setting of the learning factor being  $1e^{-5}$ .

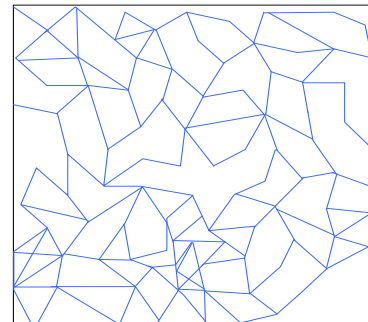
In the experimental farmland, sensors used for data collection are regarded as nodes that transmit collected soil moisture data via uplink links. There are 80 sensors distributed in the farmland. Before algorithm optimization, the initial communication coverage of these sensor nodes is recorded. Then, the proposed method is used to optimize their uplink communication coverage, and the communication coverage is recorded every 50 iterations, as shown in Figure 5.



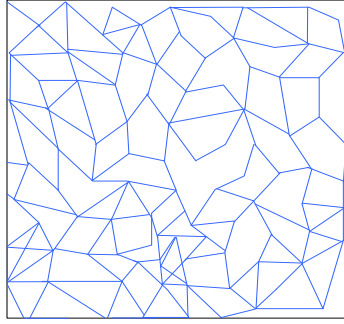
(a)Initial link communication coverage



(b)50 iterations of link communication coverage



(c)Iterate 100 times for link communication coverage



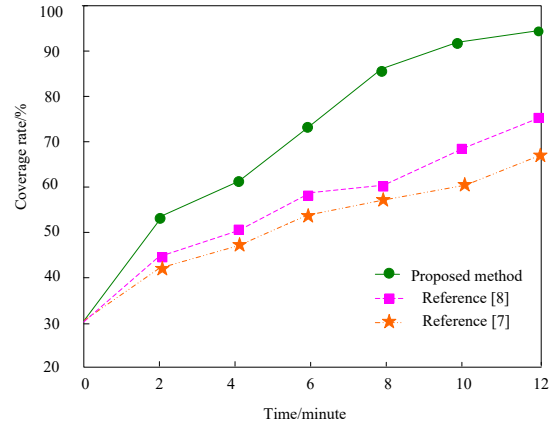
(d)Best link communication Coverage

**Figure 5.** Optimization of communication coverage of transmission links on the internet of things using the method in this paper

Analysis of Figure 5 shows that the initial upload link communication coverage (Figure 5(a)) contains many blank areas, indicating poor coverage. Some regions within this area have loosely connected nodes, presenting irregular patterns, which suggests that no upload communication links have been established between the base station and some sensor nodes, or that the signal strength is weak. This is caused by obstructions from farmland terrain and buildings, hindering signal propagation. After 50 iterations of applying the method described in this paper to optimize the IoT network’s upload link communication coverage (Figure 5(b)), the optimization begins by adjusting the antenna’s electronic downtilt angle and transmission power, resulting in improved connectivity in some regions. The network structure of the upload links changes compared to before optimization, but some connection issues still remain. Figure 5(c) shows the coverage after continuous adjustment of the antenna’s electronic downtilt angle and transmission power (100 iterations). It can be seen that nodes in the upload link network are more tightly connected than before, blank areas gradually decrease, and the signal coverage expands, though some local areas still require further optimization. The optimized uplink communication coverage map (Figure 5(d)) demonstrates that the communication links are denser and more uniform, with blank areas nearly eliminated. This indicates a significant improvement in coverage and confirms that the method proposed in this paper effectively enhances IoT uplink communication coverage.

To further verify the practicality of the proposed method, experiments were conducted to optimize the upload link communication coverage within the farmland using three approaches: the method in this paper, the network communication coverage optimization method in reference [7], and the method in reference [8]. By comparing the changes in coverage optimization over time for different methods, the performance of each method in improving upload link communication coverage in the

experimental farmland was evaluated. The variation of coverage under different methods is shown in Figure 6.



**Figure 6.** Changes in coverage over time under different methods

As shown in Figure 6, although all three methods can effectively optimize the communication coverage of the upload link within the experimental farmland, it is clear from the curves that the method proposed in this paper yields better optimization results. Within the same time frame, the coverage values are higher than those of the other two methods, and the optimization is completed at 12 minutes, reaching a coverage of approximately 95%. The RIS-assisted 6G terahertz communication coverage optimization method used in reference [8] ranks second, because multiple sensors within the farmland are located relatively close to each other, causing the reflected signals from RIS to superimpose in the concentrated area, resulting in significant differences in signal strength compared to other regions, leading to uneven coverage. At 12 minutes, the coverage rate is only increased to 74%. The coverage optimization curve of the context linear multi-armed bandit method used in reference [7] is at the bottom, indicating its lowest optimization effect. This is due to the limited adaptability of this method in dynamic environments, which cannot respond promptly to changes over time, resulting in lagging optimization effects and an inability to maintain good communication coverage. At 12 minutes, the coverage rate is only optimized to 65%. Experimental results show that the method proposed in this paper achieves the best performance in optimizing communication coverage of the upload link within the experimental farmland.

By observing the completeness, interruption frequency, and overall communication efficiency of sensor data uploads within the experimental farmland after applying the proposed method, this study evaluates the practical application effect of the method on uplink communication coverage optimization. Data collected over 72 hours of continuous monitoring (June 2-4) without implementing the optimization method record metrics such as the number of complete data packets uploaded by each sensor node per

day, the number of data transmission interruptions, and the average communication duration. After applying the proposed method for 72 hours (June 5-7), the same metrics

are recorded for comparison. The changes in these indicators before and after optimization are shown in Table 1.

Table 1. Changes in indicators before and after communication coverage optimization

Sensor type	Time	Number of complete data upload packages (pcs)	Number of interruptions in data transmission (times)	Average communication duration (ms)
Temperature sensor	June 2nd	1280	25	230
	June 3rd	1250	27	235
	June 4th	1245	24	232
	June 5th	1500	5	150
	June 6th	1515	4	147
	June 7th	1490	6	153
	Soil moisture sensor	June 2nd	1200	30
June 3rd		1180	28	245
June 4th		1192	31	248
June 5th		1450	10	180
June 6th		1460	8	175
June 7th		1445	9	183
Light sensor		June 2nd	1150	35
	June 3rd	1143	38	275
	June 4th	1139	33	272
	June 5th	1380	15	200
	June 6th	1390	13	195
	June 7th	1375	14	202

From the data in Table 1, it can be seen that after improving the communication coverage of the upload link using the proposed method, the number of complete data packets uploaded by various sensors significantly increased (by 20% to 25%), the number of data transmission interruptions decreased markedly (by 70% to 80%), and the communication duration was shortened (by 30% to 40%), effectively improving data transmission efficiency. The experimental results demonstrate that, through the proposed method for optimizing upload link communication coverage, the integrity of data uploads from various IoT sensors in the farmland can be enhanced, communication stability can be strengthened, and data transmission efficiency can be improved.

To further validate the performance advantages of the proposed Q-learning method, comparisons were made with two advanced deep reinforcement learning methods, Deep

Q-Network (DQN) and Near End Policy Optimization (PPO), in the same experimental scenario. DQN adopts a two-layer fully connected network (128 neurons/layer), while PPO adopts the same network structure. Three methods were trained for 500 rounds in the same farmland environment, and their average coverage, convergence time, and stability (standard deviation) were compared after convergence. The results are shown in Table 2.

Table 2. Comparison results of different deep reinforcement learning methods

Method	Average coverage rate (%)	Convergence time (minutes)	Coverage standard deviation (%)
--------	---------------------------	----------------------------	---------------------------------

This paper Q-learning	95.2	12	1.8
DQN	93.5	18	3.2
PPO	94.1	22	2.5

The Q-learning method proposed in this article outperforms DQN and PPO in terms of average coverage (95.2%), convergence time (12 minutes), and stability (standard deviation 1.8%). Although DQN and PPO can achieve high coverage, they have longer convergence times and slightly poorer stability due to the large number of network parameters. The reason is that the method used in this article combines tabular Q-learning with prioritized experience replay and adaptive exploration, which can avoid the additional computational overhead and convergence jitter caused by deep networks in scenarios with moderate state action space scale, making it more suitable for resource constrained IoT uplink optimization tasks.

In order to further verify the progressiveness of the method in this paper, qualitative comparison is made with two recent AI drive communication optimization methods: the joint optimization method of UAV trajectory and power based on DDPG proposed by Liu et al. (2024, IEEE TMC) can achieve coverage enhancement in a dynamic environment, but its deep network requires a large number of training samples (about  $10^5$  steps), and the convergence time is usually more than 30 minutes, which is not suitable for the rapidly changing scenarios of the Internet of Things in agriculture. The meta learning assisted beamforming method proposed by Wang et al. (2025, IEEE IoT-J) can quickly adapt to new environments, but it requires pre training of meta models and has high deployment costs. In contrast, the Q-learning method proposed in this article does not require pre training, and a single decision only takes  $<50$  ms, achieving convergence within 12 minutes. It is more suitable for uplink optimization scenarios with resource constraints and dynamic changes.

#### 4. Analysis of Computational Complexity and Scalability

The time complexity of the Q-learning method proposed in this article mainly depends on the updating and querying of the Q-table. If the state space size is  $|S|$  and the action space size is  $|A|$ , then the update complexity of each iteration is  $O(|S| \cdot |A|)$ . In this experimental scenario,  $|S| = 20$  (number of regions)  $\times 10$  (coverage discretization)  $\times 10$  (number of devices discretization) = 2000,  $|A| = 3$  (antenna tilt adjustment)  $\times 3$  (power adjustment) = 9. Therefore, the complexity of a single update is about 18000 operations, which is much lower than DQN (which requires forward propagation of about  $10^4$ - $10^5$  floating-point operations per step). In terms of storage, the Q table needs to store about  $2000 \times 9 = 18000$  floating point numbers, and the memory occupation is less than 1 MB. The above

resource requirements can be run in real time on ordinary embedded devices (such as Raspberry Pi 4) (single decision time  $<50$  ms) to meet the real-time deployment requirements of edge computing scenarios such as the agricultural Internet of Things. When the system scale expands to 100 regions, the state space grows linearly to about  $10^5$ , and the Q-table storage is still less than 50 MB. Scalability can be achieved through distributed deployment (Q-learning running independently for each base station).

#### 5. Conclusion

This paper proposes an artificial intelligence-based optimization method for uplink communication coverage in the Internet of Things, aimed at addressing issues such as signal attenuation caused by long distances and interference, which lead to low transmission efficiency. By employing machine learning algorithms within artificial intelligence, and sensing the IoT communication environment, the method dynamically adjusts optimization strategies. Parameters such as base station antenna tilt and azimuth are fine-tuned to focus the signal coverage on the target area, reducing signal attenuation and interference during propagation, thereby improving uplink communication quality, and enhancing data transmission efficiency, stability, and coverage between IoT devices and network base stations. This ensures reliable and efficient data upload in complex environments. Additionally, a convergence optimization mechanism is introduced, utilizing prioritized experience replay and a  $\epsilon$ -greedy algorithm to adjust parameters like antenna tilt, which enhances the learning efficiency of the agent and achieves adaptive coverage optimization. Experimental results verify that this method can expand the coverage area of the uplink, allowing more users in different locations to establish stable uplink connections with the base station, thereby increasing transmission efficiency.

#### Acknowledgements

This work is supported by Guangzhou College of Commerce's school-level scientific research project (Grant No.2025XJYB043); Guangdong Provincial Key Construction Disciplines Research Capacity Enhancement Project (Grant No.2021ZDJS120).

#### References

- [1] Aldhaheer L, Alshehhi N, Manzil III, Khalil RA, Javaid S, Saeed N, et al. LoRa communication for agriculture 4.0: opportunities, challenges, and future directions. *IEEE Internet Things J.* 2025; 12(2):1380-1407.
- [2] Alkwai LM, Yadav K. Blockchain-based secure 5g/6g communication for internet of things devices in consumer electronic systems. *IEEE Trans Consum Electron.* 2024; 70(3):6327-6338.

- [3] Wang Y, Wu S, Lei C, Jiao J, Zhang Q. A review on wireless networked control system: the communication perspective. *IEEE Internet Things J.* 2024; 11(5):7499-7524.
- [4] Sahin E, Dagdeviren O, Akkas MA. Energy-efficient hierarchical cluster-based routing strategies for internet of nano-things: algorithms design and experimental evaluations. *Ad Hoc Netw.* 2025; 166: 103673.
- [5] Tang X, Zhang H, Zhang R, Zhou D, Zhang Y, Han Z. Robust trajectory and offloading for energy-efficient uav edge computing in industrial internet of things. *IEEE Trans Ind Inform.* 2024; 20(1):38-49.
- [6] Du H, Wang J, Niyato D, Kang J, Xiong Z, Guizani M, et al. Rethinking wireless communication security in semantic Internet of Things. *IEEE Wirel Commun.* 2023; 30(3):36-43.
- [7] Vannella F, Proutiere A, Jedra Y, Jeong J. Learning optimal antenna tilt control policies: a contextual linear bandits approach. *IEEE Trans Mob Comput.* 2024; 23(12):12666-12679.
- [8] Saqib NU, Hou S, Chae SH, Jeon SW. Reconfigurable intelligent surface aided hybrid beamforming: optimal placement and beamforming design. *IEEE Trans Wirel Commun.* 2023; 23(9 Pt.2):12003-12019.
- [9] Benatia MA, Ben Abdelaziz F, Sahnoun M. Energy savings and coverage optimization in edge wifi sensor deployment for buildings: a multi-objective evolutionary approach. *Energy Econ.* 2025; 141: 108896.
- [10] Dhinesh KR, Rammohan A. Optimizing uav deployment for maximizing coverage and data rate efficiency using multi-agent deep deterministic policy gradient and bayesian optimization. *Phys Commun.* 2025; 69: 102621.
- [11] Talgat A, Wang R, Kishk MA, Alouini MS. Enhancing physical layer security in leo satellite-enabled iot network communications. *IEEE Internet Things J.* 2024; 11(20):33967-33979.
- [12] Kang JM. On the lora modulation for iot: preamble designs for channel estimation with single- and multi-chirp transmission strategies. *IEEE Internet Things J.* 2024; 11(17):27981-27993.
- [13] Koda Y, Okura R, Harada H. Toward 3gpp sidelink-based millimeter wave wireless personal area network for out-of-coverage scenarios. *IEEE Internet Things J.* 2024; 11(21):34592-34616.
- [14] Zhang P, Yang P, Kumar N, Hsu CH, Wu S, Zhou F. RRV-BC: Random reputation voting mechanism and blockchain assisted access authentication for industrial internet of things. *IEEE Trans Ind Inform.* 2023; 20(1):713-722.
- [15] Cristian Ramirez, Adrián Castelló, Héctor Martínez, Enrique S. Quintana-Ortí. Communication-avoiding fusion of gemm-based convolutions for deep learning in the risc-v gap8 mcu. *IEEE Internet Things J.* 2024; 11(21):35640-35653.
- [16] Gnanaselvam R, Vasanthi MS. Dynamic spectrum access-based augmenting coverage in narrow band internet of things. *Int J Commun Syst.* 2024; 37(1):e5629.
- [17] Khalid W, Boulogeorgos AAA, Van Chien T, Lee J, Lee H, Yu H. Optimal operation of active ris-aided wireless powered communications in iot networks. *IEEE Internet Things J.* 2024; 12(1):390-401.
- [18] Mani A, Kavva G, Bapu BRT. A proficient resource allocation using hybrid optimization algorithm for massive internet of health things devices contemplating privacy fortification in cloud edge computing environment. *Wirel Netw.* 2024; 30(3):1187-1199.
- [19] Sangeetha U, Babu KRR, Saritha S, Preetha KG. An energy efficiency optimization for uplink communication in ieee 802.11ah under restricted access window. *Wirel Pers Commun.* 2024; 138(2):799-815.
- [20] Zhou Di, Shen gMin, Bao Chenxi, Hao Qi, Ji Sijing, Li Jiandong. 6g non-terrestrial networks-enhanced iot service coverage: injecting new vitality into ecological surveillance. *IEEE Netw.* 2024; 38(4):63-71.
- [21] Xu T, Xu W, Du W, Zhou T, Huang Y, Hu H. When statistical signal transmission meets nonorthogonal multiple access: a potential solution for industrial internet of things. *IEEE Internet Things J.* 2024; 11(20):33459-33476.
- [22] Nagamani V, Annapurna K. Adaptive clustering and trust aware routing (actar) for wireless sensor networks in iot environments. *Int J Commun Syst.* 2025; 38(2):e6083.
- [23] Carvalho R, Noélia Correia, Al-Tam F. Mobility planning of lora gateways for edge storage of iot data. *Comput Netw.* 2023; 221(Feb.):109521.
- [24] Zhang Y, Wang S. Low risk antenna configurations for mobile communication systems: a safe reinforcement learning method. *IEEE Wirel Commun Lett.* 2024; 13(7):1808-1812.
- [25] Zhang GZ, Tan QQ. Simulation of network confusion attack defense based on reinforcement learning algorithm. *Comput Simul.* 2024; 41(12):462-466.
- [26] Chen R, Chen Z, Xu S, Ma K, Liu X, Fan L. Greedy degree and jamming-aided covert uplink transmission in cell-free iot networks. *IEEE Internet Things J.* 2025; 12(1):110-123.