Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 6013816, 9 pages
https://doi.org/10.1155/2022/6013816

*Research Article*

# Location Mechanism and Perception Scheme of Remote Transmission Equipment Layout in Smart Water

**Shikai Xing and Xiang Xie**

*School of Economics and Management, Beijing Jiaotong University, Beijing, China*

Correspondence should be addressed to Xiang Xie; xxie@bjtu.edu.cn

Received 7 April 2022; Revised 22 April 2022; Accepted 2 June 2022; Published 26 August 2022

Academic Editor: Aboul Ella Hassanien

Copyright © 2022 Shikai Xing and Xiang Xie. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Whether the traditional water equipment is equipped with remote transmission unit depends on whether the power consumption of the installed equipment can meet the expected life cycle. In the paper, by using intelligent strategies to save the power consumption of sending data, more sites can be selected to install remote transmitting units. In the process of transmitting data packet sequence, the remote transmission device needs to perceive the environment, interact with the environment and make decisions, and adjust the strategy according to the effect of the decision actions. Therefore, in this paper, the transmission process is modeled as Markov sequence decision model. And the real time signal interference noise ratio of the channel and the transmission delay of the data packet are defined as the state space. The decision action space consists of immediate transmission and delayed one, and the minimum total power consumption is taken as the objective function. The model is solved by Proximal Policy Optimization (PPO) algorithm, and the optimal decision sequence of site selection threshold is obtained.

## 1. Introduction

Different from solving multi-device game problems through reinforcement learning, solving single-device decision-making problems through reinforcement learning can be modeled as a Markov decision process [1]. When the state transition probability of a single device is known, the iterative equation of dynamic programming can be used to solve its decision-making problem. Due to the wireless environment of the installation point of water equipment, its state transition probability is often unknown, and reinforcement learning algorithm can explore the environment through trial and error, and obtain a good strategy to maximize the overall expected return. In this paper, a deep reinforcement learning algorithm will be used to solve the Markov sequence decision problem of a single intelligent device.

## 2. Literature Review

In recent years, wireless communication technology is developing rapidly, especially the Internet of Things communication technology represented by NB-IoT (Narrowband Internet of Things) [2], which has the characteristics of deep coverage, low power consumption, and low remote maintenance cost, which is very suitable for the water industry, the deployment location, data transmission, battery power supply, and other characteristics of the remote transmission equipment. With the full coverage of the NB-IoT network, a large number of IoT devices have been installed or replaced in the urban water supply and drainage network and copy collection system. The number of remote transmission equipment such as water meters and manhole cover monitors installed with NB-IoT modules in urban pipeline networks and copy collection systems is also rapidly increasing.

These remote transmission devices have intelligent features such as on-demand data reporting, intelligent sleep and access to the platform, and online detection with the platform. These intelligent devices are mainly composed of three modules, namely, data acquisition and calculation module, compression and storage module, and data transmission module. These modules are located in the application layer, storage layer, and communication layer of the smart device, respectively.

According to the agreement between the device and the platform [3], the device collects data at a fixed frequency every day, and the amount of data accumulated in each time period is basically fixed. The sensors of the device collect data according to certain rules and frequencies. The basic data is the measurement data of the user's water flow. In addition, there may be additional data such as water pressure and water temperature. The collected data is stored in the remote transmission device. Metering and sensing accumulated data to the platform in the form of incremental or absolute amount through the communication unit on the equipment at a predetermined time every day. The data accumulated during upload will be split into multiple packets and sent.

After the data is sent on the same day, in order to save power consumption and prolong the working life, the remote transmission device will power off itself to sleep or enter the PSM (power save mode) state after sending data, and wake up on the next day or when data is sent, the unsent data will continue to be sent the next day.

Most of the equipment of intelligent water affairs is installed in the weak coverage area of wireless channels such as corridors and wells. On the coverage benchmark, there will also be various frequency shifts, random disturbances generated by interference, and superposition of noises, such as white noise, occlusion, and co-frequency interference. The result of superposition makes the channel quality change randomly.

Therefore, the channel quality of wireless communication is jointly determined by signal coverage and noise interference. Generally, two quantities are used to evaluate the channel quality: RSRP (Reference Signal Receiving Power) and SINR (Signal to Interference plus Noise Ratio).

## 3. Problem definition

Most of the equipment in the urban pipe network system and the copying system are deployed in underground pipelines or corridors. The remote transmission equipment is battery powered and power consumption sensitive. Due to building block reflection or stray interference, the wireless signal will be greatly attenuated, and the battery life is very dependent on the installation environment and data transmission strategy. If the installation location and transmission strategy are not good, some batteries will be consumed quickly from the start of the remote transmission equipment to the use of the battery, which cannot support the data collection and transmission business needs of a period of 6–10 years. Since the installed equipment is unattended, when the equipment needs to replace the battery, it generally needs to be manually operated on-site. When the battery is manually replaced on-site or the traditional manual reading is used, it is often difficult to enter the home and the equipment maintenance cost is relatively high.

Due to the relatively high labor cost of replacing batteries for these devices, from the perspective of investment and cost, it is expected that these devices can maintain a longer life cycle and send more data, thus resulting in the problem of installation and location selection of remote transmission devices.

The installation site selection of the remote transmission equipment depends on two aspects [4]: the wireless signal coverage quality of the installation point, and the ability of the equipment to perceive the installation environment.

Before remote transmission equipment is installed, it is necessary to measure the wireless signal coverage quality at the installation site to ensure compliance with a certain threshold. The determination of the threshold is related to the ability of the device to perceive the installation environment. That is, the stronger the device's ability to perceive the installation environment, the lower the requirements for the quality of wireless signal coverage; on the contrary, the weaker the device's ability to perceive the installation environment, the higher the requirements for the quality of wireless signal coverage. By improving the device's ability to perceive the environment, the requirements of the device on the quality of on-site signal coverage at the installation point are reduced, so that more installation points have the conditions to install remote transmission equipment.

When the channel state changes, the SNR fluctuates. It will affect the power (power consumption) consumed by the remote device when sending data. When the signal gets worse, it takes longer to send data and requires more power consumption. The relationship between power consumption and signal quality presents a nonlinear relationship, that is, when the signal quality gradually deteriorates, the power consumption gradually increases. When the signal quality falls below a certain threshold, the power consumption starts to increase rapidly.

Based on the above-mentioned nonlinear relationship between channel quality and power consumption, there is no clear judgment value, and it is necessary to determine the best action value that should be taken according to the environment and a certain strategy. That is, the remote transmission device detects the channel quality of the wireless communication through the sensor to determine whether to send the data immediately or temporarily. Under the condition of a certain battery capacity of the remote transmission equipment, after all data of the day is sent correctly, the power consumption is the smallest, so that the service life of the remote transmission equipment is the longest. Therefore, when the life cycle of the equipment is 6–10 years, it is necessary to determine the location range of whether the remote transmission unit can be installed according to the amount of data to be sent.

Therefore, the installation site selection and perception improvement of remote transmission equipment is essentially an optimization problem for smart equipment to adopt a certain decision-making mechanism in an unknown environment to minimize the power consumption of the equipment.

## 4. System Model

In this paper, the environment-aware decision-making process of data packets sent by remote equipment is modeled as MDP [5] (Markov decision process) to describe the state transition of the system. And the derivation of the process is

equivalent to the layout and location selection mechanism of intelligent water remote transmission equipment.

Remote device is divided into $n$ data packets each time, the data packet set $T_i = \{t_1, t_2, \ldots, t_n\}$, is a sample of task distribution, and each task sample corresponds to an MDP process. This MDP process is defined by a quintuple $\{S, A, P, R, \gamma\}$, where $S$ is the state space, $A$ is the action space, $P$ is the state transition matrix, and $R$ is the state transition benefit, which $\gamma$ is the forward benefit discount.

*4.1. Channel Status.* Remote transmission device sends each data packet in turn. According to the agreement, the amount of data sent is fixed. However, due to the random characteristics of wireless channels, the channel quality will vary randomly. Some packets need to be resent if poor channel quality results in higher BER or packet loss.

The signal-to-noise ratio (SINR) of the channel obeys a small-scale Rayleigh distribution, and its probability density function [6] is expressed as

$$p(snr) = \frac{1}{\overline{snr}} \exp\left(-\frac{snr}{\overline{snr}}\right). \tag{1}$$

Among them, $snr$ the threshold value set of SINR is $\overline{snr}$ expressed, and the expectation of SINR is expressed. The NB-IoT network $snr$ divides the channel quality into $N$ grades according to the value, and sets a total of $N$ -1 signal quality thresholds, $snr = \{snr_1, snr_2, \ldots, snr_{N-1}\}$. According to the $snr$ threshold, the channel quality can be divided into N states, $C = \{c_0, c_1, \ldots c_{N-1}\}$. The channel state probability is

$$p_c(c_i) = \int_{snr_i}^{snr_{i+1}} p(snr) d(snr). \tag{2}$$

Assuming that the next moment of the channel can only be transferred to the adjacent interval, the state transition probability of the channel [6] is

$$\begin{aligned} p_c(c_i, c_{i+1}) &= \frac{N(snr_{i+1}) * T_f}{p_c(c_i)}, \quad i \in \{0, 1, \ldots, N-2\}, \\ p_c(c_i, c_{i-1}) &= \frac{N(snr_i) * T_f}{p_c(c_i)}, \quad i \in \{1, 2, \ldots, N-1\}. \end{aligned} \tag{3}$$

Among them, $N(snr_i) = \sqrt{2 * \pi * snr_i / \overline{snr}}$, $*$, $f_D *$ $\exp(-snr/\overline{snr})$, $f_D$ is the maximum Doppler frequency shift.

*4.2. Device Latency Status.* Device latency status of the device is described by the total delay after the device sends the *i-th* data packet, and the total delay corresponds to the power consumption. Suppose the sending time of the first data packet is, and the sending time of $t_0$ the *i-th* data packet is $t_i$. Obviously, the total delay after the device transmits the *i-th* data packet is not only *related* to the immediate decision, but also depends on the delay accumulation of the previous $i$−1 data packets. The solution of the total transmission delay of the *i-th packet* can be transformed into a combination of immediate decision-making and solution of the total delay of the *i-1-th* packet.

$$\tau_i = \tau_{i-1} + \Delta\tau_i + T_f, \tag{4}$$

where the device latency status is $\tau = \{\tau_1, \ldots, \tau_n\}$, $\Delta\tau_i$ denote the waiting time determined by the immediate decision is related to the channel state and $\tau_{i-1}$ depends on the state of the previous $i$ -1 data packets, $T_f$ is a constant which is the transmission time of one data packet.

*4.3. System Status.* System state $S$ can be defined as the combined state of the channel state and the device delay state, i.e., $S \triangleq C \otimes \tau$. The above equation is a recursive process and cannot be solved by dynamic programming because the transition probability is unknown.

*4.4. Action.* $A$ denotes the remote transmission device taken when the new data packet is to be transmitted and it can be expressed as $A \triangleq \{0, 1\}$: When $a = 0$, it means to transmit immediately; when $a = 1$, it means to suspend transmission.

## 5. MDP Process Analysis of Remote Equipment

*5.1. System State Transition.* When the remote transmission device sends data packets in sequence in a queuing manner, according to the channel state, the remote transmission device may choose to send the data packets immediately. This strategy works in the connected state, corresponding to the instantaneous SINR, and uses the transmit power to transmit with a certain probability. And transfer to the new channel state and device delay state.

The remote device may also adopt a policy of suspending sending a packet. This strategy waits in an idle state, expecting a better channel state to transmit, and the smart device $\Delta\tau_i$ continues to transmit after an interval according to the change of the channel state. After waiting for transmission, the system state transitions to the new channel state and in-device delay state. Since each data packet is small, the time to transmit a data packet is very short compared to the waiting time, which is defined as one frame $T_f$.

$\Delta\tau_i$ set as the single-step delay time of the $i$ -th data packet, which is randomly determined $\Delta\tau_i$ by the intelligent device according to the environmental state. It is the pause waiting time. This waiting will cause the overall delay of the time for the device to complete the transmission of all data packets $\Delta\tau_i$. $\tau_i$ is the total delay of sending the first $i$ data packets. $t_i$ In order to send the time to be transmitted for the $i$ - *th* data packet, the delay before the i-1 data packet $t_{i-1} - t_0$ is to be transmitted is

$$\tau_{i-1} = \begin{cases} t_{i-1} - t_0 + T_f, & \text{sentting the } (i-1)\text{th Packets immediately,} \\ t_{i-1} - t_0 + \Delta\tau_{i-1} + T_f, & \text{suspending the } (i-1)\text{th packet,} \end{cases} \tag{5}$$

the *ith* data packet is decided, the recursive formula of the delay state of the device is

$$\tau_i = \begin{cases} \tau_{i-1} + T_f, & \text{sentting the ith Packets immediately,} \\ \tau_{i-1} + \Delta\tau_i + T_f, & \text{suspending the ith packet.} \end{cases} \tag{6}$$

Set $\tau_0 = 0$. Then, the one-step delay of the $i$ -th packet is

$$\Delta\tau_i = \left\{ \text{int}\left[ \frac{N - snr_i}{(snr_N - snr_1/N)} \right] \right\} * \delta. \tag{7}$$

*SINR* is related to the delay. The better the *SINR*, the smaller the delay. $\delta$ In order to suspend the parameter of the minimum time segment, this paper assumes that the channel quality remains unchanged in the minimum time segment to ensure that the retransmission is an incoherent channel, and $N$ is the number of states. And *int* is a rounding operation function.

*5.2. System Benefits and Costs.* After the device sends a data packet, if it does not receive the confirmation message from the platform within a certain period of time, it will resend the data packet. Due to the different acknowledgment mechanisms of different platforms, the additional power consumed by the device to resend the data packet is different. Therefore, an equivalent model is established: in different environments, the device increases the transmit power to ensure the same bit error rate during the transmission and transmission of data packets.

NB-IoT network adopts two rates to transmit data according to the *SINR*, corresponding to BPSK modulation (Binary Phase Shift Keying) and Q PSK modulation (Quadrature Phase Shift Keying). When the channel quality is relatively poor, the NB - IoT network uses the BPSK modulation method to obtain the minimum transmit power consumption when the bit error rate requirements are met under different channel states [7],

$$P_{ber}(s_i) \le 0.5 * erfc\left( \sqrt{\frac{snr_i * P_i')}{\sigma}} \right), \tag{8}$$

where $P_{ber}(s_i) = ber_i/\sigma$ is the bit error rate probability and $erfc(x) = 2/\sqrt{\pi} * \int_0^x e^{-\eta^2} d\eta$ is the Gaussian error function.

*Ber* is the allowable bit error rate of the wireless channel, $\sigma$ is interference noise power. It is the $P_i'$ equivalent transmit power that guarantees no packet loss or retransmission when the bit error rate is met. When the channel quality is relatively good, the NB-IoT network uses the QPSK modulation method to obtain the minimum transmission power consumption when the bit error rate requirements are met under different channel states, namely,

$$P_{ber}(s_i) \le 0.2 * exp\left( \frac{-1.6 * snr_i * P_i'}{3 * \sigma} \right). \tag{9}$$

When the device suspends sending the first $i$ data packet, the power consumption is $I_0 * \Delta\tau_i + P_i'/v * T_f$, where $I_0$ is the idle state working current of the device, and when the device is in the waiting state, the idle state current of the $I_0$ device within $\Delta\tau_i$ the time is. $v$ is the device voltage and $T_f$ is the length of time in the connected state to transmit a data packet.

When the device sends immediately, $\Delta\tau_i = 0$, the power consumption of a single packet is

$$W_i = \Delta\tau_i * I_0 + \frac{P_i'}{v} * T_f. \tag{10}$$

After sending each data packet of all samples, the total power consumption of one day can be expressed as

$$W_{\text{day}} = \sum_{m}^{M} \sum_{i=1}^{n} \left( \Delta\tau_i * I_0 + \frac{P_i'}{v} * T_f \right). \tag{11}$$

Among them, $M$ is the number of samples sent in one day, each sample is the data to be sent once agreed in the protocol, and $n$ is the number of data packets corresponding to one sample. The goal of the system is to find the most efficient decision to send packets for maximum benefit, i.e., minimum total power consumption.

Defined $t_i$ immediate reward function is the negative value of the power consumption of a single data packet, that is, $R_i = -W_i$, the immediate reward function of the amount of data sent per day is the negative value of the total power consumption, i.e., $R_{\text{day}} = -W_{\text{day}}$.

*5.3. Location Range of Equipment.* The expected life days that the remote device can use is $d$, and the battery capacity is set to be $W_{\text{total}}$, then $d = W_{\text{total}}/W_{\text{day}}$, the expected life years $D = d/365$.

If the remote device only has data to be sent once a day, the location decision can be calculated by the following formula

$$W_{\text{total}} = \sum_{j=0}^{d} \sum_{i=1}^{n} \left( \Delta\tau_i * I_0 + \frac{P_i'}{v} * \tau_0 \right). \tag{12}$$

Since the battery capacity of the device $W_{\text{total}}$ is a known value, when the expected number of days or years of use is given, an optimal environment-aware decision chain needs to be obtained to obtain the optimal $\Delta\tau_i$ sum $P_i'$ scheduling.

Under the condition that the battery capacity, expected service life, and data volume of the device are all determined, whether a remote transmission device can be installed at an installation point depends entirely on $\Delta\tau_i$ and $P_i'$ there is a

feasible solution. In order to delay the service life of the equipment, it is also necessary to seek the optimal strategy $A$ to obtain the optimal solution.

According to the data collection and transmission rules of the device and the platform, each time data is sent every day during actual operation, it will be divided into $n$ data packets and sent in a sequence. During the sample learning period, data can be sent continuously for $M$ times, parameter training can be performed, battery power consumption can be measured, and it can be converted into whether the battery life meets the expected number of years.

## 6. Optimal Decision-Making Based on Deep Reinforcement Learning

For the above MDP problem, this paper will use deep reinforcement learning algorithm to solve. In the random and aligned MDP process, the state $s$ and the decision action $a$ have randomness and can $\pi(a|s)$ be represented by a conditional distribution, representing $\pi$ the mapping from the state to the decision action $a$, that is, the policy. The following defines a class of neural networks to fit policies.

*6.1. Sequence Representation of Neural Networks.* The smart device divides the data sent each time into $n$ data packet sequences to send, and needs to make a sending decision on the $n$ data packets in turn to form a decision sequence. This paper uses seq2seq (sequence-to-sequence) [8] neural network to express the above process sequence, that is, input a sequence and output a sequence. The network structure is shown as Figure 1.

Seq2Seq neural network consists of an encoder and a decoder, both of which are R NN networks (Recurrent Neural Network, recurrent neural network). Assuming that the parameters of the neural network are $\theta$, then when the state $s$ is input, the conditional probability of outputting the optimal decision $a$ can be rewritten $\pi_\theta(a|s)$. According to the input $t_i$, the neural network first encodes the learning and memory through the encoder, and then outputs through the decoder according to the memory of the network $d_j$, and then passes through two different activation functions, corresponding to the output state value function $v(s)$ and decision sequence probability $\pi_\theta(a|s)$.

In order to reflect the characteristics of the data to be sent, the sequence of data packets can be converted into a sequence of embedded vectors and input to the neural network. Input $t_i$ = [data packet sequence number $i$, data packet size], which is a 2-dimensional vector. For different types of devices, the amount of data to be sent each time may be different, so the size of the data packet is different. The size of the last data packet of each device is also different from the previous $n-1$ data packets, which is the margin of the data amount divided by $n-1$.

Denote the encoder and decoder [9] as' $f_{dec}$, respectively, $f_{enc}$, then.

The output of the encoding part is

$$e_i = f_{enc}(t_i, e_{i-1}). \tag{13}$$

The output of the decoding part is

$$d_j = f_{dec}(z_j, s_{j-1}, a_{j-1}). \tag{14}$$

The input of the decoder consists of 3 parts [10], including the weighted sum of the output of the encoder $z_j$, and the decision execution result of the previous step $s_{j-1}, a_{j-1}$. $z_j$ is the context of the decoder in step $j$, which contains the attributes of the data to be sent.

Seq2seq neural network is an $n$ -dimensional vector $d$. After nonlinear activation of this vector, an $n$ -dimensional probability vector $\pi_\theta$ and a value function $v$, $(s_i)$, respectively, are obtained. The probability vector corresponds to $\pi_\theta$ the probability that the decision action takes a certain $a$, and the sum is 1. Then, the decision action of the $a_j = \text{argmax}_a(\pi_\theta)jth\ step$ can be obtained through the greedy algorithm.

*6.2. Parameter Update and Optimal Decision.* To update the neural network parameters, this paper adopts the PPO2 [11] algorithm (Proximal Policy Optimization, proximal policy optimization). The remote transmission equipment may send 1 or more samples per day, and the daily objective function can be defined as

$$J = \text{ER}_{\text{day}}. \tag{15}$$

Based on the sequence representation of neural network, the objective function $J$ for each sample $(\theta_m)$ can be rewritten as

$$J(\theta_m) = E_\theta(R_{\text{day}}(\theta_m)). \tag{16}$$

According to the law of large numbers,

$$J(\theta_m) \approx \sum_{m=1}^{M} (R(\theta_m)) = E(-W_{\text{day}}), \tag{17}$$

where $M$ is the number of samples, the data volume of each sample is divided into $n$ data packet sequences and sent, $m$ is the sample sequence number, that is, the data packet sequence. $R(\theta_m)$ is the power consumption of each sample, namely,

$$R(\theta_m) = \sum_{i=1}^{n} \min(\omega_i * A_i, \text{clip}_{1-\Delta}^{1+\Delta}(\omega_i) * A_i), \tag{18}$$

$\Delta$ is the clipping constant, and the trend function [12] is

$$A_i = \sum_{j=0}^{n-i+1} (\gamma * \omega_i)^j * (r_{i+j} + \gamma * v_\pi(s_{i+j+1}) - v_\pi(s_{i+j})). \tag{19}$$

Among them, $\gamma$ is the advantage function discount coefficient, $r$ is the immediate return, and $v_\pi$ is the state value function. $\omega$ are the adjustment coefficients for variance and bias,

$$\omega_i = \frac{\pi_{\theta_m}(a_i|s_i)}{\pi_{\theta_m^0}(a_i|s_i)}. \tag{20}$$

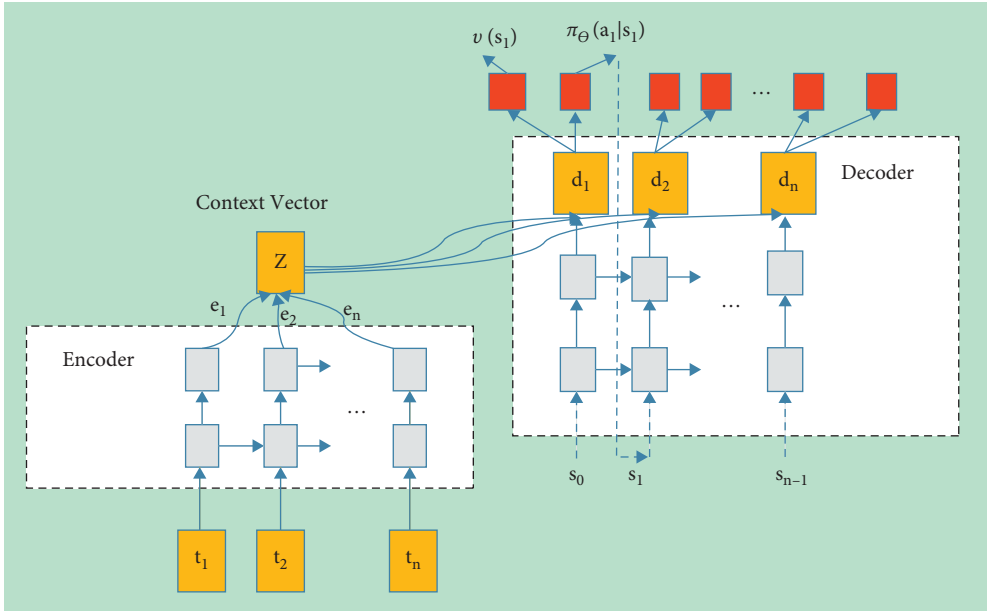


Figure 1: Architecture of the seq2seq neural network.

According to the formula (10), $t_{i+j}$ the immediate reward function that can be obtained is

$$r_{i+j} = E\left(-W_{i+j}\right). \tag{21}$$

gradient update formula of the neural network parameters is

$$\theta_i' = \theta_i + \alpha * \nabla_{\theta_i}\left(J\left(\theta_m\right)\right). \tag{22}$$

Among them, $\alpha$ is a training coefficient, which is the learning rate, that is, the gradient descent factor. When the neural network parameters converge iteratively, the output of seq2seq is the optimal decision.

## 7. Experiment and Result Analysis

The list of simulation parameters used in this paper is as follows, including model parameters and algorithm parameters. In the simulation process, a sample is divided into 10, 12, 15, and 20 data packet sequences for training. In order to reduce the channel correlation, the minimum delay time segment of data packet transmission is taken as 5 s.

The water model and neural network, as well as the parameter settings of PPO2 are shown in Table 1.

After encoding the delay state and channel state, respectively, they are input into the neural network. Figure 2 shows the change of the device delay state. When the decision is to send immediately, the delay is much smaller than the delay caused by suspending transmission. Figure 3 shows the change of the channel state. The channel state, as the environment in which the smart device is located, is relatively stable when the signal-to-noise ratio is good, and is more likely to change randomly when the signal-to-noise ratio is poor.

Each time the far-transmission device sends a data packet, it may either suspend the sending or send it immediately. The transmit power each time a data packet is sent is related to the state and environment it is in. Figure 4 shows the power when different data packets are transmitted, and the different power levels reflect the equivalent model of repeated transmission. The power consumption of suspending the transmission of data packets is related to the idle current and the waiting time. The power consumption of sending a packet immediately is related to the transmit power/device voltage, and the transmit duration. The immediate return in Figure 5 is the negative value of power consumption, and the minimum power consumption is equivalent to the maximum return.

After the water MDP model is established, it is solved by the PPO2 algorithm. For the sake of convenience, the reference objective function value is -6 in the experiment, and

Table 1: Simulation parameters.

| Parameter | Value description |
|---|---|
| Number of packets | $n = \{10, 12, 15, 20\}$ |
| Suspend the smallest time segment | $\delta = 5$ s |
| Packet transmission time | $T_f = 1$ s |
| Packet size | <200 bytes |
| SNR threshold | snr = [1.28 3.28 5.28 6.28] |
| Equipment voltage | $V = 3$ |
| Device idle current | $I_0 = 250\mu A$ |
| Number of samples | $M = 5$ |
| Expected life cycle | $d = 10 * 365$ days |
| Learning rate | $\alpha = [0.002, 0.005]$ |
| Return discount factor | $\gamma = 0.9$ _ |
| Advantage function discount factor | $\phi = 0.95$ |
| Clipping constant | $\Delta = 0.2$ |
| Number of neurons | unit $s = 2$ 56; |
| Overfitting factor | 0.5 _ |
| Hidden unit | Layers = 256 |
| Coding layer | Layer 1 = 2 |
| Decoding layer | Layer 2 = 2 |

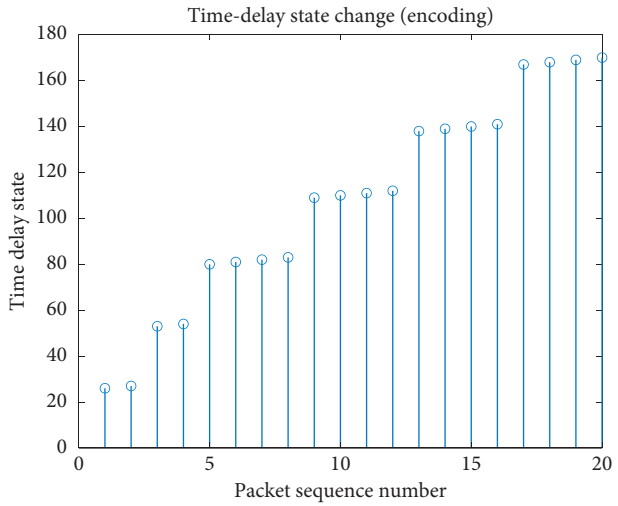


Figure 2: Delay change process of packet sequence.

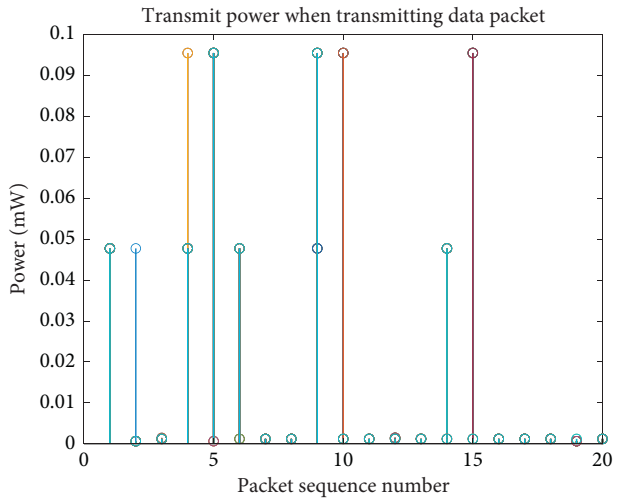


Figure 4: Transmit power of data packet sequence.

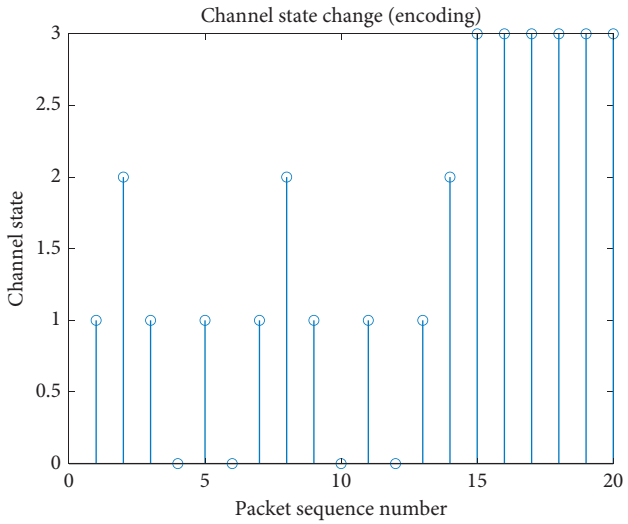


Figure 3: Channel change process of packet sequence.

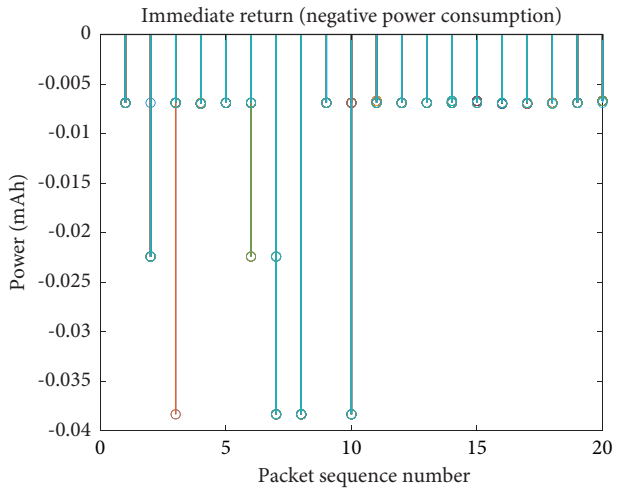


Figure 5: Immediate reward of packet sequence decision.

the difference between the training objective and the reference objective value is defined as the loss function. The parameters of the neural network converge as the number of iterations increases, and the strategy is also more optimized. When the loss function is stable, the parameters of the neural network converge to the optimal, and the corresponding strategy is the optimal expression of the neural network. Figure 6 is the parameter iteration process when the learning rate is 0.002, and Figure 7 is the parameter iteration process when the learning rate is 0.0005. When the learning rate is increased, the convergence is faster, but the oscillation is also larger.

The agreement between the water remote transmission equipment and the platform stipulates that a single data should be less than 200 bytes, so when the amount of sample data to be sent is large, it will be split into more data packets. On the other hand, when the amount of data to be transmitted is constant, the more data packets are split, that is, the smaller the data packets, the lower the cost of retransmission, which is more suitable for the requirements of the narrowband Internet of Things. At this time, the loss function is smaller and easier to converge. Figures 6 and 7 are, respectively, the convergence of neural network parameters for 10, 12, 15, and 20 data packets.

Figure 8 randomly distributes the full connection bias range of the neural network from 0 to 1. After adjusting to a normal random distribution, the convergence of the loss function will be more oscillating due to the increase in the dynamic range.

As mentioned above (Figure 5), if the perception model and algorithm in this paper are not used, the power consumption of one data packet is 0.04 mAh. If 20 data packets are sent every day and the life cycle of 10 years is calculated,

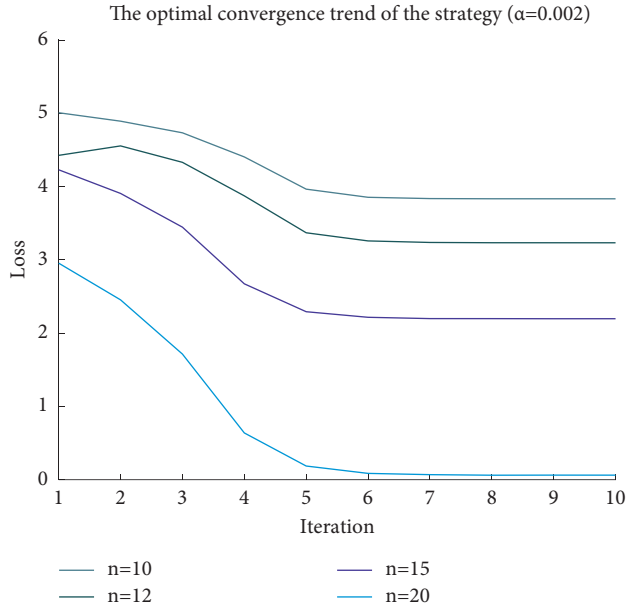


Figure 6: Convergence process of policy (learning rate 0.002).

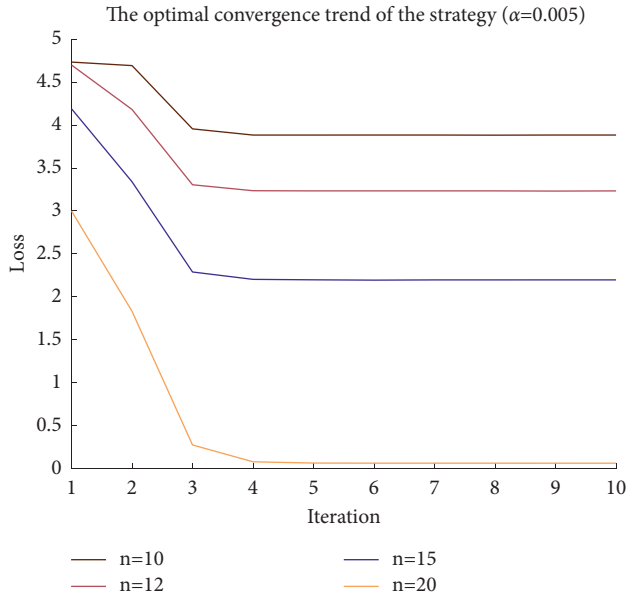


Figure 7: Convergence process of policy (learning rate 0.005).

the required battery capacity is $0.04 * 20 * 365 * 10 = 2920$ mAh. Generally, the medium battery capacity is 5000 mAh, which can support the device to send 1.71 samples per day, and each sample includes 20 data packets. After adopting the perception model and algorithm in this paper, the average power consumption of a data packet in Figure 5 is 0.018 mAh, which can support the device to send 3.7 samples per day. With the increase in the amount of data collected by equipment, the perception model and algorithm in this paper can support more installation point equipment with remote transmission units.

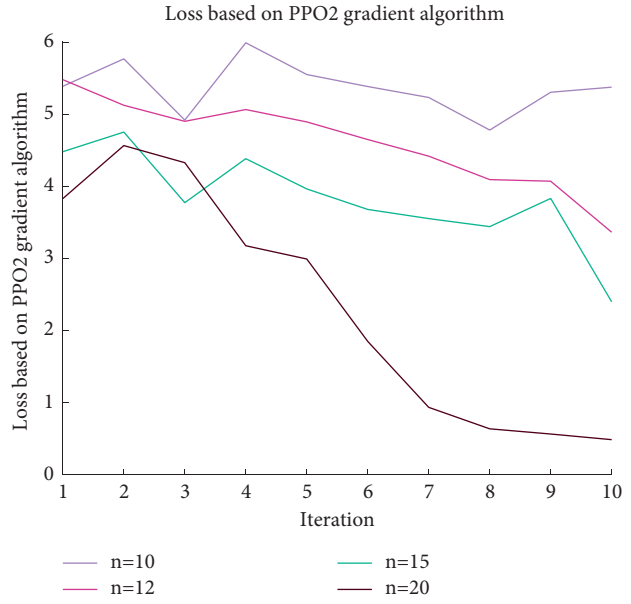


Figure 8: Convergence process with wider bias range (normal random distribution).

## 8. Conclusion

Starting from the actual problem of the layout and location mechanism of remote transmission equipment of intelligent water affairs, this paper establishes a system model based on MDP, namely, the Markov sequence decision model. The state, action, and reward function definitions of the model are given in this paper, and the simulation results of these definitions are given in the experimental part. The sequence decision is expressed based on the Seq2Seq neural network, and the PPO2 algorithm is used to solve the MDP sequence decision problem. The solution process and results are given in the experimental part. And when the model and method are applied to the actual water management problem, based on the life cycle of the remote equipment and the battery capacity value, the application scope of the equipment layout and location selection is solved.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was partially supported by National Natural Science Foundation of China (71874011).

## References

[1] X. Lin, Yu Tan, J. Zhang, C. Yang, and J. Liu, "High energy efficiency policy based on Markov decision in wireless

transmission," *Systems Engineering and Electronics*, vol. 36, no. 07, pp. 1433–1438, 2014.

[2] Bo Dai and G. Yuan, *NB-IoT Standards and Key Technologies*, People's Post and Telecommunications Publishing House, China Gongxin Publishing Group, Beijing.

[3] *TCMA SB 040-2019, Technical Guide for Layout and Location, Acceptance and Use of NB-IoT Automatic Meter Reading System*.

[4] T. IoT Industry Alliance, *5G Smart Water White*, p. 12, Paper 2020 Edition, China, 2020.

[5] X. Zhao, Q. Wang, and K. Zhao, "Research and application of reinforcement learning based on constraint Markov," *Computer Science*, vol. 4, no. 4, p. 36, 2009.

[6] Z. Jiang, T. Wang, Y. Song, and Y. Liu, "Transmission scheduling scheme based on deep Q learning in wireless network," *Journal of Communication*, vol. 39, no. 04, pp. 35–44, 2018.

[7] S. T. Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: a unified view," *IEEE Transactions on Communications*, vol. 49, no. 9, pp. 1561–1571, 2001.

[8] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," 2017, https://arxiv.org/abs/1707.06347.

[9] W. Zhan, J. Wang, Q. Zhu, H. Duan, and Y. Ye, "Computational unloading scheduling method based on deep reinforcement learning in moving edge computing," *Computer Application Research*, vol. 38, no. 01, pp. 241–245+263, 2021.

[10] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *Proc. ICLR*, pp. 1–15, 2015.

[11] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbel, "High-demensional continuous control using generalized advantage estimation," *Proc. ICLR*, pp. 1–14, 2016.

[12] J. Wang, J. Hu, G. Y. Min, A. Y. Zomaya, and N Georgalas, "Fast adaptive task offloading in edge computing based on meta reinforcement learning," *IEEE Transactions on Parallel and Distributed Systems*, vol. 32, no. 1, pp. 242–253, 1 Jan 2021.