A Deep Neural Network Coordination Model for Electric Heating and Cooling Loads Based on IoT Data

Hongyang Jin, Yun Teng, Member, IEEE, Tieyan Zhang, Member, IEEE, Zedi Wang, and Zhe Chen, Fellow, IEEE

Abstract—As the ubiquitous electric power internet of things (UEPIoT) evolves and IoT data increases, traditional scheduling modes for load dispatch centers have yielded a variety of challenges such as calculation of real-time optimization, extraction of time-varying characteristics and formulation of coordinated scheduling strategy for capacity optimization of electric heating and cooling loads. In this paper, a deep neural network coordination model for electric heating and cooling loads based on the situation awareness (SA) of thermostatically controlled loads (TCLs) is proposed. First, a sliding window is used to adaptively preprocess the IoT node data with uncertainty. According to personal thermal comfort (PTC) and peak shaving contribution (PSC), a dynamic model for loads is proposed; meanwhile, personalized behavior and consumer psychology are integrated into a flexible regulation model of TCLs. Then, a deep Q-network (DQN)-based approach, using the thermal comfort and electricity cost as the comprehensive reward function, is proposed to solve the sequential decision problem. Finally, the simulation model is designed to support the validity of the deep neural network coordination model for electric heating and cooling loads, by using UEPIoT intelligent dispatching system data. The case study demonstrates that the proposed method can efficiently manage coordination with large-scale electric heating and cooling loads.

Index Terms—Deep neural network, electric heating and cooling load, IoT data, reinforcement learning.

I. INTRODUCTION

UBIQUITOUS electric power internet of things (UEPIoT) have been primarily implemented to develop the intelligent, synergistic and ecological construction of power systems [1], [2]. Soon enough, energy management platforms, which integrate accurate real-time situation awareness (SA) and efficient advanced application scenario services, will be greatly enhanced with an increase of IoT data [3]–[5]. Motivated by the UEPIoT, flexible demand side resources (DSRs) with large scale IoT intelligent terminals have been regarded as a sufficient means for peak shaving and renewable energy consumption. Among them, electric heating and cooling load accounts for a large proportion of demand side resources. In fact, the air conditioning load accounts for 40% of the residential energy consumption and this ratio will increase in extreme weather [6]. The time-varying characteristics and thermal inertia of electric heating and cooling loads have become the basis of the dynamic response of DSRs.

At present, the research of flexible regulation for electric heating and cooling loads based on IoT technology has been carried out, such as SA and control strategy of loads, demand side cyber-physical systems (CPS) and automatic demand response (DR) [7]–[9]. The heterogeneous recognition model was used to construct the polymerized electric heating and cooling load model under different state information by adjusting the set-point temperature in [10]. An uncertainty quantification framework was introduced in [11], where TCLs are controlled without worrying about the details of the internal mechanism. A hierarchical DR architecture (HDRA) to control different potential DR resources was proposed in [12], and the performance of power systems and communications was demonstrated in a simulation. In order to generate optimal demand response policies for HAVC systems, the learning-based home energy management system was proposed in [13], which was continuously developed in [14]. The above article primarily focuses on the state estimation and coordinated control of the load using the measurement data collected by the supervised control and data acquisition (SCADA) and the synchronous phasor measurement unit (PMU), which is essentially the measurement of the electrical quantity. However, the real-time control of the DSRs is not only related to the characteristics of the energy consumption, but also affected by the multi-source information in the natural environment, such as indoor temperature. Therefore, the awareness, understanding and prediction of massive IoT data based on multi-source data fusion are crucial to solving the problem of coordinated control of electric heating and cooling loads.

Artificial intelligence has been applied in the field of industrial process control and decision-making. Deep reinforcement learning (DRL) consisting of deep neural network and reinforcement learning is an important branch of artificial intelligence [15], [16]. To solve the intelligent demand side resource control problem with sequential decision characteristics, effective methods, especially deep Q-network (DQN) with independent optimization, are proposed in [17], [18]. In [19], a data-driven learning model for demand response

Manuscript received July 30, 2019; revised October 10, 2019; accepted December 17, 2019. Date of publication March 30, 2020; date of current version December 31, 2019. This project was supported by National Key Research and Development Plan (2017YFB0902100) and Key Project of Liaoning Natural Science Foundation under Grant (20170520292).

H. Y. Jin, Y. Teng (corresponding author, e-mail: tengyun@sut.edu.cn), T. Y. Zhang, and Z. D. Wang are with Shenyang University of Technology, Shenyang 110870, China.

Z. Chen is with the Aalborg University, Aalborg 9220, Denmark.

DOI: 10.17775/CSEEJPES.2019.01700
was contemplated in smart communities that reduced cost to the consumers and coordinated the pricing. A novel reinforcement learning coordination method of flexible resources was proposed in [20]. The results demonstrated that comfort constraints of buildings can be met by a distributed intelligent method. In [21], a demand response algorithm was proposed, which used the coordination method of reinforcement learning and deep neural networks. It was shown that the profitability of the service provider and customers was promoted by the artificial intelligence method.

The deep neural network coordination model for electric heating and cooling loads with adaptive, self-correcting and self-learning abilities is studied in this paper. The coordination model is coupled by the Q-network and deep neural network. Normally, large-scale loads with IoT data are operated by a DQN. The identifying contributions of this paper are as follows:

1) The individual thermal behavior of TCL is provided by personal thermal comfort. The dynamic set-point temperature model of electric heating and cooling loads with the performance of peak shaving and valley filling, which can consider the personal thermal comfort, is proposed using IoT data.

2) The peak shaving contribution is studied for keeping the peak shaving capacity of TCLs in the peak shaving intervals. According to the peak shaving contribution of the electric heating and cooling loads, the customers are given a certain economic compensation to ensure the expectation of optimal operation of the grid.

3) The flexible regulation method of electric heating and cooling loads based on a deep Q-network is proposed to improve the accuracy of the model and the efficiency of the algorithm.

The remainder of this paper is organized as follows. Section II illustrates the system model as well as the coordination framework. Section III proposes the DQN-based approach. Section IV verifies the performance of the proposed method with an intelligent dispatching system. Section V presents the concluding remarks.

II. SYSTEM MODEL AND PROBLEM FORMULATION

The system model of a deep neural network coordination model for electric heating and cooling loads based on IoT data is composed of a controllable load (i.e. TCLs), wind power, conventional load, fossil fuel power plant and load dispatching center (LDC) with SA and a coordinated control function, as shown in Fig. 1. The current environmental states are received through interaction with the environment. And the scheduling mode of the TCLs is determined in the load dispatching center. The online coordination control capability of the load dispatching center is a prerequisite for maximizing the dispatching potential of the intelligent load. The performance of the load dispatching center should include information filtering, a cognitive model and intelligent regulation. Therefore, the load dispatching center should incorporate the closed loop general architecture of DQN. More specifically, the information of load characteristics and real-time status of the grid are collected and preprocessed by intelligent sensors in the IoT node, such as temperature sensors and intelligent electric meters. Subsequently, the information is transmitted to the load dispatching center. The real-time optimal control commands, which maximizes personal thermal comfort and peak shaving contribution, are issued so that potential peak shaving reserve is provided.

A. IoT Data Collection and Processing

The uncertainty IoT node data may lead to system instability. The performances of the coordinated control method of electric heating and cooling loads are often adversely affected.
by data with high anomaly and low reliability. It is a strong motivation to propose a method for IoT data collection and processing. A method of sliding windows (SW) was proposed to eliminate anomaly data of TCLs in this paper. The SW is a data string window that slides on the data stream, and generally extends in the time dimension [22]. The length of the data string window remains the same, as the window slides.

Suppose the IoT node $i$ has data set $B_{i,j}^k = \{b_j, j = 0, \ldots, k\}$ at time 0 to $k$, and the sliding window $S + 1$ traverses the data set in turn. From the perspective of data statistics, the $N$-th TCL data subset is steady state when the relative standard deviation $\delta$ of the $S + 1$-th data in the window satisfies the precision of $\delta < \varepsilon$, as in (1).

$$\delta = \frac{1}{\bar{x}} \sqrt{\sum_{s=1}^{S+1} (x_s - \bar{x})^2} < \varepsilon$$

where $\delta$ is the relative standard deviation of the dimensionless quantity, $\bar{x}$ is the average value of the $(S + 1)$-th data in the sliding window and $\varepsilon$ is the accuracy requirement of the relative standard deviation.

B. Personal Thermal Comfort

Meeting the individual behavior of TCLs using IoT data is an essential factor for accurate identification of load operation characteristics. Regarding the quantitative model of individual behavior, personal thermal comfort is proposed for flexible regulation of TCLs.

The dynamic characteristic of TCLs changes with the ambient temperature and operational state. The common first-order equivalent models of TCLs are founded in (2)-(3) [23].

$$\frac{dT_{in}(k)}{dt} = \frac{1}{RC}(T_{out}(k) - T_{in}(k) - m(k)RP_{TCL}\eta)$$

$$m(k) = \begin{cases} 
0, & \text{if } T_{in}(k) \leq T_{set} - \frac{\delta}{2} \\
1, & \text{if } T_{in}(k) \geq T_{set} + \frac{\delta}{2} \\
ad(k - \varepsilon), & \text{else}
\end{cases}$$

where $T_{in}(k)$ is the indoor temperature at the time $k$. $R$ is the equivalent thermal resistance. $C$ is the equivalent thermal capacity. $P_{TCL}$ is the power of TCLs in the heating/cooling mode. $\eta$ is the energy efficiency ratio of TCLs. $m(k)$ is the switch state 0/1 of TCLs at the time $k$. In the constant set-point temperature mode, the indoor temperature is maintained within the range of $T_{set} \pm \delta/2$ by switching the variable $m(k)$ between on/off.

The thermal comfort can be calculated for (4) as follows:

$$c(k) = \exp\{-[(E(k) - 20)^2/2.5213]\}$$

where $v(k)$ is the air velocity at the time $k$, and $H(k)$ is the relative humidity at the time $k$. In order to simplify the calculation, this paper assumes that the indoor air velocity and relative humidity are the same as that outside. And the TCL operational state is primarily affected by the temperature change.

For flexible regulation of TCLs, it is essential to maintain the personal thermal comfort [25]. Therefore, the personal thermal comfort denotes the user’s somatosensory comfort, which is obtained by flexibly setting the set-point temperature $T_{set}$ of the TCL based on the real-time personalized thermal comfort. The basic idea is to find an optimal set-point temperature of the TCL by the dynamic adjustment process of the regulation variable $\Delta T_{change}(k)$ and the demand variable $\lambda(k)$, as shown in Fig. 2.

Fig. 2. The dynamic adjustment process of optimal set-point temperature.

The optimal set-point temperature of the personal thermal comfort is determined as follows:

$$T_{set}^{PTC}(k) = (T_{set}(k) + \Delta T_{change}(k))\lambda(k).$$

As far as the set-point temperature of the personalized thermal comfort is different, the thermal comfort will also have time-varying characteristics. Therefore, the critic function of the personal thermal comfort can be represented in (7).

$$c^{PTC}(k) = \lambda(k)\exp\{-[(E(k) - T_{PTC}(k))^2/2.5213]\} + (1 - \lambda(k))$$

where $T_{PTC}(k)$ is the effective temperature with the optimal individual thermal comfort based on the thermal comfort prediction model, in which $T_{PTC}(k) = E\{T_{PTC}(k)\}$. The parameters of the demand variable $\lambda(k) = 0$ or $\lambda(k) = 1$, respectively represents no requirement or requirement for thermal comfort.

In the dynamic adjustment process of the optimal set-point temperature of the TCL, the situation awareness for TCLs is finished in the load dispatching center. And the parameters of personal thermal comfort were updated, which can be used as the basis for the peak shaving and valley filling.

C. Peak Shaving Contribution

The peak shaving contribution of the unit refers to the ability of the peaking unit to follow the net load in power systems with high proportional renewable energy (RE). Through the auxiliary service market (ASM), the peaking unit is given a certain economic compensation by the utility grid [26]. The demand side load replaces the peaking unit to reduce the peak-to-valley differences, which can effectively cut the energy consumption. The utility grid should refer to the principle
of the auxiliary service market, and give a certain economic compensation for TCLs by the peak shaving contribution. Therefore, the peak shaving contribution of TCLs is that the peak shaving capacity provided by TCLs in the peak shaving intervals.

In this paper, the peak shaving intervals are divided by the power balance function of the system. There are large-scale uncertain wind power and load in the power system. The output status of the fossil fuel power plant is designed as follows:

\[ \mu(k) = \frac{\sum P_{\text{NG}}(k) - \sum P_{\text{Load}}(k)}{\sum P_{\text{NG}}} \sim N \left( \frac{\varphi_2 P_{\text{Load}}(k) - \varphi_1 P_{\text{WP}}(k)}{\left( \sum P_{\text{NG}} \right)^2} \right) \]  

where \( \mu(k) \) is the output status of the fossil fuel power plant at time \( k \). \( \sum P_{\text{NG}}(k) \) is the sum of the fossil fuel power plant power at the time \( k \), which is determined by day-ahead scheduling. \( \sum P_{\text{NG}} \) is the sum of fossil fuel power plant capacity in the system. \( \varphi_1 \) and \( \varphi_2 \) are the parameters to be identified respectively.

In this paper, the Markov chain Monte Carlo (MCMC) algorithm is used to establish the net load probability model considering the uncertainty of wind power and loads. The uncertainty of wind power and loads is reflected in the errors of wind power forecasting and load forecasting, respectively. The actual value of wind power and loads based on a normal distribution were calculated as follows [27]:

\[ P_{\text{WP}}(k) \sim N \left( \varphi_{\text{pre}} P_{\text{WP}}(k), \varphi_1^{\text{pre}} P_{\text{WP}}(k) \right) \]  

\[ P_{\text{Load}}(k) \sim N \left( \varphi_{\text{pre}} P_{\text{Load}}(k), \varphi_2^{\text{pre}} P_{\text{Load}}(k) \right) \]  

where \( P_{\text{WP}}(k) \) is the actual value of wind power at time \( k \). \( P_{\text{WP}}(k) \) is the forecast value of wind power at time \( k \). \( P_{\text{Load}}(k) \) is the actual value of load at time \( k \). \( P_{\text{Load}}(k) \) is the forecast value of the load.

If the TCL is not used as the resource of peak shaving, its values can be obtained according to the conventional method such as the equipment model. So the values of the loads obey the normal distribution.

The difference of peak shaving interval reflects the difference of the peak shaving contribution in each period. Therefore, the peak shaving intervals are divided into five categories according to the values of the output status of the fossil fuel power plant: emergency state 1 (U1), emergency state 2 (U2), critical state 1 (M1), critical state 2 (M2) and normal state (R), which are shown in Table I.

<table>
<thead>
<tr>
<th>Peak shaving interval</th>
<th>( \mu(k) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>U1</td>
<td>[0.9, 1.0]</td>
</tr>
<tr>
<td>M1</td>
<td>[0.8, 0.9]</td>
</tr>
<tr>
<td>R</td>
<td>[0.6, 0.8]</td>
</tr>
<tr>
<td>M2</td>
<td>[0.4, 0.6]</td>
</tr>
<tr>
<td>U2</td>
<td>[0.0, 0.4]</td>
</tr>
</tbody>
</table>

The probability distribution \( f(\mu_k) \) of the output state of the fossil fuel power plant can be accurately estimated, which requires the posterior probability on a large sample by the law of large numbers. However, the posterior distribution in the probability space is highly complex, which increases the difficulty of parameter estimation. Therefore, the MCMC random sampling method is adapted to solve the above problems.

The probability distribution \( f(\mu_k) \) of the output status of the fossil fuel power plant assumes that the random variable sequence \([\mu_0, \mu_1, \mu_2, \mu_3, \ldots] \) has the state transition matrix \( P(\mu_{k+1}|\mu_k) \) at each time \( k \geq 0 \). The distribution of random variables converges to a stationary probability distribution \( \pi(\mu_k) \). Since the current output state of the fossil fuel power plant is related to the output state of the fossil fuel power plant at the previous time, the Markov chain can be established based on this relationship. After \( m \)-th iteration of the state transition, the sampling point \([\mu_k, k=m+1, \ldots, n] \) is approximately independently sampled in \( \pi(\mu_k) \). The posterior distribution is obtained from the sample set \([\mu_{m+1}, \mu_{m+2}, \ldots, \mu_n] \). The mean value is used as the maximum posterior estimation result.

\[ \hat{f} = \frac{1}{n-m} \sum_{k=m+1}^{n} f(\mu_k). \]  

Thus the optimal estimation of the relevant parameters of the output status of the fossil fuel power plant is obtained, and the distribution function of the peak shaving intervals is calculated.

The operational state of the TCL is changed in the different peak shaving intervals, and the peak shaving contribution of the TCL is also varied. The different peak shaving contribution trend is illustrated in Fig. 3, in which the economic compensation is varied with the peak shaving contribution. The economic compensation is formulated as follows:

\[ e(k) = \beta_\Omega \exp[-\mu(k)] - 0.45 \]  

where \( e(k) > 0 \) denotes the excitation income. \( e(k) < 0 \) denotes the penalty cost. \( \mu(k) \) is the output status of the fossil fuel power plant. \( \beta_\Omega \) is the coefficient of peak shaving contribution, which is calculated by (13).

\[ \beta_\Omega = \begin{cases} 
1, & \mu_k \in R \\
1.5, & \mu_k \in M1 \text{ or } M2 \\
2, & \mu_k \in U1 \text{ or } U2 
\end{cases} \]  

\[ e(k) = \beta_\Omega \exp[-\mu(k)] - 0.45 \]  

\[ \beta_\Omega = \begin{cases} 
1, & \mu_k \in R \\
1.5, & \mu_k \in M1 \text{ or } M2 \\
2, & \mu_k \in U1 \text{ or } U2 
\end{cases} \]  

\[ e(k) = \beta_\Omega \exp[-\mu(k)] - 0.45 \]  

D. Problem Formulation

The optimize model of flexible regulation of TCLs (FRTCLs) is that the capacity of TCLs is varied with the maximal
personal thermal comfort and the peak shaving contribution in the next peak shaving period. Thus, FRTCLs is expressed as follows:

\[
\max \sum_{i=1}^{N} \left[ e_{PTC}^{i}(k) + P_{TCL} m_{i}(k) \cdot (e(k) - \xi(k)) \Delta k \right]
\]  

(14)

subject to: (2), (3), (4), (5), (8), (9), (10), (11).

\[
\Delta T_{\text{change},i} \leq \Delta T_{\text{change},i}(k) \leq \Delta T_{\text{change},i}^{\max}
\]  

(15)

\[
P_{\text{Gn}} + P_{\text{WPn}} - P_{\text{Ln}} - \sum_{i} V_{i} (G_{in} \cos \theta_{in} + B_{in} \sin \theta_{in}) = 0
\]  

(16)

\[
P_{\text{Gn}}^{\min} \leq P_{\text{Gn}} \leq P_{\text{Gn}}^{\max}
\]  

(17)

\[
P_{\text{WPn}}^{\min} \leq P_{\text{WPn}} \leq P_{\text{WPn}}^{\max}
\]  

(18)

\[
P_{\text{Ln}}^{\min} \leq P_{\text{Ln}} \leq P_{\text{Ln}}^{\max}
\]  

(19)

\[
P_{\text{Pn}}^{\min} \leq P_{\text{Pn}} \leq P_{\text{Pn}}^{\max}
\]  

(20)

where \( e_{PTC}^{i}(k) \) is the critic function of the personal thermal comfort of \( i \)-th TCL at the time \( k \). \( m_{i}(k) \) is the on/off status of the \( i \)-th TCL at time \( k \). \( \xi(k) \) is the time of use (TOU) at time \( k \). \( N \) is the total number of TCLs.

In (14), the stable energy supply of heating/cooling energy for the customers and the reduction of peak-to-valley ratios are confirmed respectively. Constraint (15) is the limit of the set-point temperatures of the TCLs. Constraint (16) is the nodal power balance equations. Constraint (17) is the power restriction of the fossil fuel power plant. Constraint (18) is the power restriction of wind power. Constraint (19) is the nodes voltage constraints. Constraint (20) is the limits of the line power flows.

III. EXPLOITING DQN TO SOLVE PROBLEM

In order to solve the problem of flexible regulation of TCLs, a DQN-based approach is proposed. DQN is regarded as an advanced deep reinforcement learning (DRL) based on the markov decision process, in which the model is comprised of states \( s_{k} \), actions \( a_{k} \), transition probability \( p_{k} \) and rewards \( r_{k} \). In this paper, a DQN-based approach is interpreted by the states, actions and rewards of TCLs, as follows:

1) The states of TCLs

A set of states of TCLs, which were obtained from (21), consists of the indoor temperature \( T_{m}(k) \), the state \( m(k) \) and the output status of the fossil fuel power plant \( \mu(k) \).

\[
s_{k} = \{ T_{m,1}(k), \ldots , T_{m,N}(k), m_{1}(k), \ldots , m_{N}(k), \mu(k) \}.
\]  

(21)

2) The actions of TCLs

A set of actions of TCLs, which were obtained from (22), consists of the regulation variable \( \Delta T_{\text{change}}(k) \) and the demand variable \( \lambda(k) \).

\[
a_{k} = \{ \Delta T_{\text{change},1}(k), \ldots , \Delta T_{\text{change},N}(k), \lambda_{1}(k), \ldots , \lambda_{N}(k) \}.
\]  

(22)

3) The rewards of TCLs

A reward function is calculated from (14) as follows:

\[
r_{k}(s_{k}, a_{k}) = \sum_{i=1}^{N} \left[ e_{PTC}^{i}(k) + P_{TCL} m_{i}(k) \cdot (e(k) - \xi(k)) \Delta k \right]
\]  

(23)

To access the optimal strategy, Q-learning on environment interactions is used. The goal is calculated in order to choose a policy \( \pi^{*} \).

\[
V^{*} = \max_{\pi} V_{\pi} = \max_{\pi} E_{\pi} \{ r_{\pi}(s_{k} = s) \}.
\]  

(24)

The reward function is corrected by a long-term expected reward \([28]\). \( Q^{\pi}(s, a) \) satisfies the Bellman equation based on the state-action long-term expected reward. \( Q^{*}(s, a) \) is achieved by environment interactions, in which the optimal control strategy \( \pi^{*}(s) \) is calculated as follows:

\[
\pi^{*}(s) = \arg \max_{a} Q^{*}(s, a)
\]  

(25)

\[
Q^{*}(s, a) = E_{\pi}[r(s, a, s') + \gamma \max_{a' \in A} Q^{*}(s', a')]
\]  

(26)

where \( s' \) is the transition state. \( \gamma \) is the discount factor, indicating the importance of long-term rewards compared to current rewards.

Q-values is solved by iteration as follows:

\[
Q'(s, a) = Q(s, a) + \alpha (r(s, a, s') + \gamma \max_{a' \in A} Q(s', a') - Q(s, a))
\]  

(27)

where \( \alpha \) is the learning rate.

The Q-value function cannot be obtained in tabular, which is caused by extremely high-dimensional FRTCLs problem. So the approximator \( \hat{Q}(s, a; \omega) \) of the optimal function is designed by a deep neural network. And the loss function is represented as follows:

\[
L_{\ell}(\omega) = E_{s,a}[r(s, a, s') + \gamma \max_{a' \in A} Q(s', a'; \omega_{-1}) - \hat{Q}(s, a; \omega)]^{2}
\]  

(28)

where \( \omega \) is the vector, which is updated by training until there is minimization of the loss function.

The gradient of the loss function is obtained as follows:

\[
g_{\ell}(\omega) = E_{s,a}[r(s, a, s') + \gamma \max_{a' \in A} Q(s', a'; \omega_{-1}) - \hat{Q}(s, a; \omega)] \cdot \nabla_{\omega} \hat{Q}(s, a; \omega).
\]  

(29)

The optimal policy is rewritten as follows:

\[
\tilde{\pi}^{*}(s) = \arg \max_{a \in A} \hat{Q}(s, a; \omega_{j}).
\]  

(30)

IV. CASE STUDIES

A. Experiment Setup

The intelligent dispatching system (IDS) described in Section II is simulated to illustrate the proposed DQN-based approach for flexible regulation of TCLs. There are a total of 10,000 TCLs in the real UEPIoT, located in Northeastern China. And the detailed parameters of TCLs are presented in Table II. The system of this case study is composed of wind
power of 20 MW, a fossil fuel power plant of 90 MW and a conventional load. It can be observed in Fig. 4. The real-time temperature, relative humidity, and wind speed of a typical day in summer are reflected in Fig. 5. In order to train the deep neural network, the continuous running state information of 100 days is collected from the intelligent dispatching system, including measurement data and control signals. The scheduling period is 1 day. The scheduling command control period is 0.5 hours, and a total of 4,800 sets of data are recorded.

\[
\begin{array}{|c|c|}
\hline
\text{Parameters} & \text{Value} \\
\hline
R & 2^\circ\text{C}/\text{kW} \\
C & 1.5 \text{ kWh}/^\circ\text{C} \\
P_{\text{TCL}} & 3 \text{ kW} \\
\eta & 3 \\
T_{\text{set}}(0) & 24^\circ\text{C} \\
\Delta T_{\text{min}}^\text{change} & -4^\circ\text{C} \\
\Delta T_{\text{max}}^\text{change} & +4^\circ\text{C} \\
\Delta k & 0.5 \\
\hline
\end{array}
\]

Table II: The Detailed Parameters of TCLs

The division of peak shaving intervals are illustrated in Fig. 6 by (8)–(11), which is corresponding to the economic compensation with TOU.

In the experiment, the input layer is \( s_k \), which is obtained from (21). There are 3 hidden layers having 30, 60, 120 neurons respectively, using the ReLU activation function. The output layer is \( a_k \), which is obtained from (22). The learning rate is 0.0001. The discount factor is 0.95. Let the agent be the \( \varepsilon \)-greedy strategy, and \( \varepsilon \) is 0.005. Training episodes are 100 with a mini-batch size of 32 and replay memory size of 200 in each episode. The structure of the intelligent dispatching system is illustrated in Fig. 1. The simulations are carried out on a HP Z8 G4 Workstation with Intel(R) Xeon(R) Gold 6148 CPU, 2.40 GHz and 128 GB RAM memory. Simulations are run in MATLAB R2019a.

B. Performance Evaluation

Three scenarios are conducted to compare the optimization effects of different types of rewards using a DQN-based approach, as can be seen in Table III.

The curves of the cumulative reward and average cumulative reward in Fig. 7 illustrate that the performance of the three scenarios throughout the training process. The three highest average cumulative rewards are 0.98, 0.66 and 1.67, respectively. It is obvious that the optimum average cumulative reward can be obtained during the training process in each scenario. Training episodes and computation time of the highest average cumulative reward are 78 episodes with 384 seconds, 43 episodes with 215 seconds and 56 episodes with 280 seconds, respectively.

The overall optimization results of the three scenarios in
one scheduling period are shown in Fig. 8, including the load power of TCLs and the total load power of the intelligent dispatching system.

In scenario 1, the TCLs are operated in the optimal personal thermal comfort. It is obvious that there are two peaks of power at noon and evening, which increases the peak-to-valley differences of the total load power in the intelligent dispatching system, and seriously affects the safe operation of the system.

In scenario 2, the TCLs are operated in the optimal peak shaving contribution. Most customers choose to increase the electricity consumption during the valley period for reducing the electric cost. Although the peak-to-valley differences of the system are lower than that of scenario 1, the total load power still has a peak value at 7:00 to 8:00 and 20:00 to 23:00 because of the guidance of the peak shaving contribution. This is the peak shift, not the peak shaving and valley filling, which is not conducive to the adjustment of the peak-to-valley difference.

Correspondingly, scenario 3 is the operational mode that can provide the optimal personal thermal comfort and peak shaving contribution. In this way, the peak-to-valley differences are
50.8% and 7.7% which are lower than that of scenario 1 and scenario 2, respectively.

The statistical distribution of personal thermal comfort and the electric cost of TCLs are illustrated in Fig. 9 and Fig. 10, respectively. Particularly, the positive value represents the excitation income, and the negative value represents the electric cost. In a scheduling period, the utility grid company gives the customers a certain excitation income and penalty cost according to the particular peak shaving intervals in (12). Therefore, the electric costs of TCLs are a positive value from 0:00 to 7:00 and a negative value from 8:00 to 23:00. In scenario 1, comparing the other two scenarios results, it can be seen that the electric cost is the largest at 296.2623 thousand yuan and the average personal thermal comfort is 0.98. In scenario 2, the excitation income is the largest at 31.8119 thousand yuan compared with the other two scenarios, and the average personal thermal comfort is 0.56. In scenario 3, the electric cost is 51.5997 thousand yuan and the average personal thermal comfort is 0.96. During this process, the best comprehensive effect is achieved, including the personal thermal comfort, peak shaving contribution and the effect of peak-to-valley differences. Overall, scenario 3 is more suitable for TCLs operations in peak shaving and valley filling.

Table IV shows the comparison of the three results from the DQN-based approach, the chaotic particle swarm optimization (CPSO) and the differential evolution (DE). The grid net income refers to the electricity sales cost minus the electricity purchase cost in the grid. It can be seen from the table that the comprehensive reward obtained by the DQN-based approach is higher than the other two methods at a scheduling period, with 900 and 800 respectively, in which the search ability has been enhanced. At the same time, there were also the same results in the grid net income of the three methods, indicating that the electric expenses of the DQN-based approach are less than the other two methods. The flexible operation of the power systems is improved by the method proposed in this paper with a favorable interaction. In addition, the control performance of TCLs is also improved by the DQN-based approach, especially in the reduction of the peak-to-valley differences in the grid.

### Table IV

<table>
<thead>
<tr>
<th>Descriptions</th>
<th>DQN-based approach</th>
<th>CPSO</th>
<th>DE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTC</td>
<td>9500</td>
<td>8600</td>
<td>8700</td>
</tr>
<tr>
<td>Electric cost (thousand yuan)</td>
<td>51.5997</td>
<td>85.6218</td>
<td>78.6479</td>
</tr>
<tr>
<td>Grid net income (thousand yuan)</td>
<td>159.6548</td>
<td>155.7922</td>
<td>162.2348</td>
</tr>
</tbody>
</table>

## V. Conclusion

In this paper, the deep neural network coordination model for electric heating and cooling load based on IoT data is proposed for flexible regulation of thermostatically controlled loads. The overall rewards of personal thermal comfort and peak shaving contribution can be maximized using a deep Q-network-based method. The highest average cumulative reward of the deep Q-network-based method is achieved at 1.67, which reveals the effectiveness of the comprehensive reward. The electric heating and cooling load coordination method is verified by simulation examples. The results show that the demand side resources can be optimized by the proposed model based on IoT data. The engineering requirements of peak shaving by demand response can be insured. At the same time, the performance of the search ability for the flexible regulation method based on deep reinforcement learning is enhanced, which supports the real-time demand side management method.

### References


