Energy Consumption Control Method for Oven based on DDPG

Xudong He [1, a], Wenhua Zhang [1, b], Qian Zhang [2, c], Tengbiao Fang [1, d]

[1] Manufacturing Department (PF), SAIC VW Automobile Co., Ltd Shanghai, China

[2] Security Department (HS), SAIC VW Automobile Co., Ltd, Shanghai, China

[a] hexudong@csvw.com, [b] zhangwenhua2@csvw.com, [c] zhangqian6@csvw.com,

[d] fangtengbiao@csvw.com

Abstract. Oven equipment wastes a lot of energy to ensure the process and production requirements during car manufacturing. To solve this problem, this paper proposes an energy consumption control method for oven based on DDPG. This paper collects equipment data in real time by IoT technology; To solve the problems of abnormal data and inconsistent dimensions, data cleaning is applied; Then, the key features of energy consumption are extracted by integrated feature selection and dimension reduction algorithm to improve the upper limit of the model; After that, this paper creates an energy consumption reward function to evaluate energy consumption effect of parameter setting; Finally, the accuracy of the model can be improved by setting appropriate state space and action space of the model, building the network structure of the model and adjusting the super parameters; The energy consumption of oven can be controlled by this model and experiments show that the algorithm can reduce about 18% energy consumption of oven with process requirements, which the energy saving effect is good.

Keywords: Energy Control; Feature Selection; Feature Dimension Reduction; DDPG.

1. Introduction

The 4 processes of car manufacturing include stamping, welding, painting and assembly. In terms of energy consumption, paint shop is the major energy consumption of the whole car plant. In order to realize energy conservation and emission reduction of car manufacturing, it is particularly critical to reduce the energy consumption of the paint shop. Oven equipment is the key energy consumption equipment of the paint shop. Reducing the energy consumption of oven equipment is an important point for energy saving of paint shop.

The oven equipment is used to dry car body, which can ensure that car body meets baking process requirements through four energy kinds including natural gas, electric energy, chilled water and compressed air. The oven equipment has two characteristics. 1. The parameters of oven are set according to the maximum load requirements. There is a large allowance under the condition of non-maximum load, which result in the high temperature of the discharged gas; 2. The oven equipment takes a long time to start and shut down. The equipment is in production state during a short time of shift changing and the energy consumption is completely discharged directly to the outside.

Oven equipment wastes a lot of energy consumption to ensure the process and output requirements. Through the reasonable control of equipment parameters, the problem of massive energy consumption waste can be solved. At present, the control algorithm of oven equipment is mainly PID control, which is easy to result oscillation and has poor control effect. In addition, the energy kind of the oven device is diverse and there is a coupling relationship between natural gas, electric energy, chilled water and compressed air. It is difficult to realize accurate energy saving and ensure the process and output requirements by one parameter control.

Reinforcement learning (RL) [1] maximizes the reward signal through the interaction between agent and environment, which can realize the global optimal control for complex system. At present, there is no research on the energy-saving control method based on reinforcement learning for oven equipment. However, many scholars have studied energy-saving control through reinforcement learning in other fields. Yuan et al. [2] combined RL and rule control to optimize the air conditioning system. Compared with PID and rule-based strategy, RL performs better in energy consumption and comfort of air conditioning system, and the total energy consumption is reduced by 7.7% and 4.7% respectively; Chen et al. [3] took the office building as the research object and used the Q-learning algorithm to input the start and stop control strategy of HAVC system. Compared with the rule control strategy, the energy consumption of the two simulation experiments was reduced by 13% and 23% respectively; Dalamagkidis et al. [4] used free model reinforcement learning method to control ventilation fan speed and window opening degree. The results show that this strategy has better performance than switch controller and fuzzy PD controller.

From the above research results, reinforcement learning can effectively improve the energy-saving effect. However, there is still a problem in the application of the above algorithm for oven equipment. The state space and action space of oven equipment are multidimensional and continuous. However, at present, most relevant studies only deal with the problem of limited parameter space and control only for a single discrete variable, which limits their applicability to complex systems. DDPG algorithm can adapt to multi-dimensional continuous action space system and can solve the problem [5]

Finally, the paper takes the oven equipment in the paint shop of Anting No. 2 car plant of SAIC Volkswagen Co., Ltd. as the research object. This paper puts forward an energy consumption control method for oven based on DDPG to reduce the overall energy consumption of the system when ensure the body process and output requirements. Figure 1 is the basic flow of the algorithm.

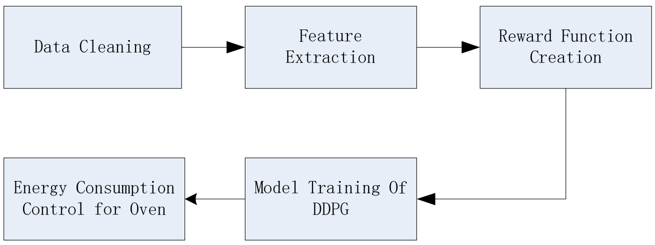

Fig 1. The basic flow of energy consumption control method for oven based on DDPG

2. Data Cleaning

Data cleaning is the basis of feature extraction, which the quality of data cleaning will affect the effect of feature extraction. Therefore, it is necessary to do data cleaning for oven equipment to extract the data features better for oven equipment.

The oven equipment data has two characteristics. 1. A large number of abnormal data may be caused under the conditions of equipment on-off, power-off, sensor calibration and so on; 2. The dimensions of oven equipment's process, production, conveyor and energy consumption data are inconsistent. Based on these two characteristics of oven equipment, data cleaning includes abnormal data processing and feature scaling.

The process of abnormal data processing includes abnormal data location, abnormal cause matching and the right medicine. Abnormal data is located by drawing visual graphics such as histogram and scatter diagram; It is necessary to match the abnormal causes because that abnormal data may be caused by equipment startup, shutdown, power failure, sensor calibration and so on. After matching, the measurement can be taken to the data cleaning. The value of the previous time can be replaced for the abnormal data of on-off machine; The abnormal data of equipment power failure can be processed by linear interpolation or data deletion; The abnormal data of sensor calibration can be processed by linear interpolation.

Feature scaling is to solve the influence of different dimensions and improve the convergence speed. The main methods of feature scaling include linear function normalization, mean normalization, standardization, unit length scaling. This paper selects standardization for data processing.

After data cleaning, the data has 36 dimensions including 4 energy consumption values and 32 features affecting energy consumption.

3. Feature Selection and Dimension Reduction Algorithm based on Integration

The cleaned data has two characteristics. 1. Data dimension is high and some dimensions have no effect on energy consumption; 2. Some dimensions overlap partially, which is easy to lead to model overfitting. Combined with the characteristics of oven equipment, this paper proposes a feature selection and dimension reduction algorithm based on integration. The algorithm flow is shown in Figure 2. The algorithm includes integrated feature selection and feature dimensionality reduction algorithm based on comparison.

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections A-D below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

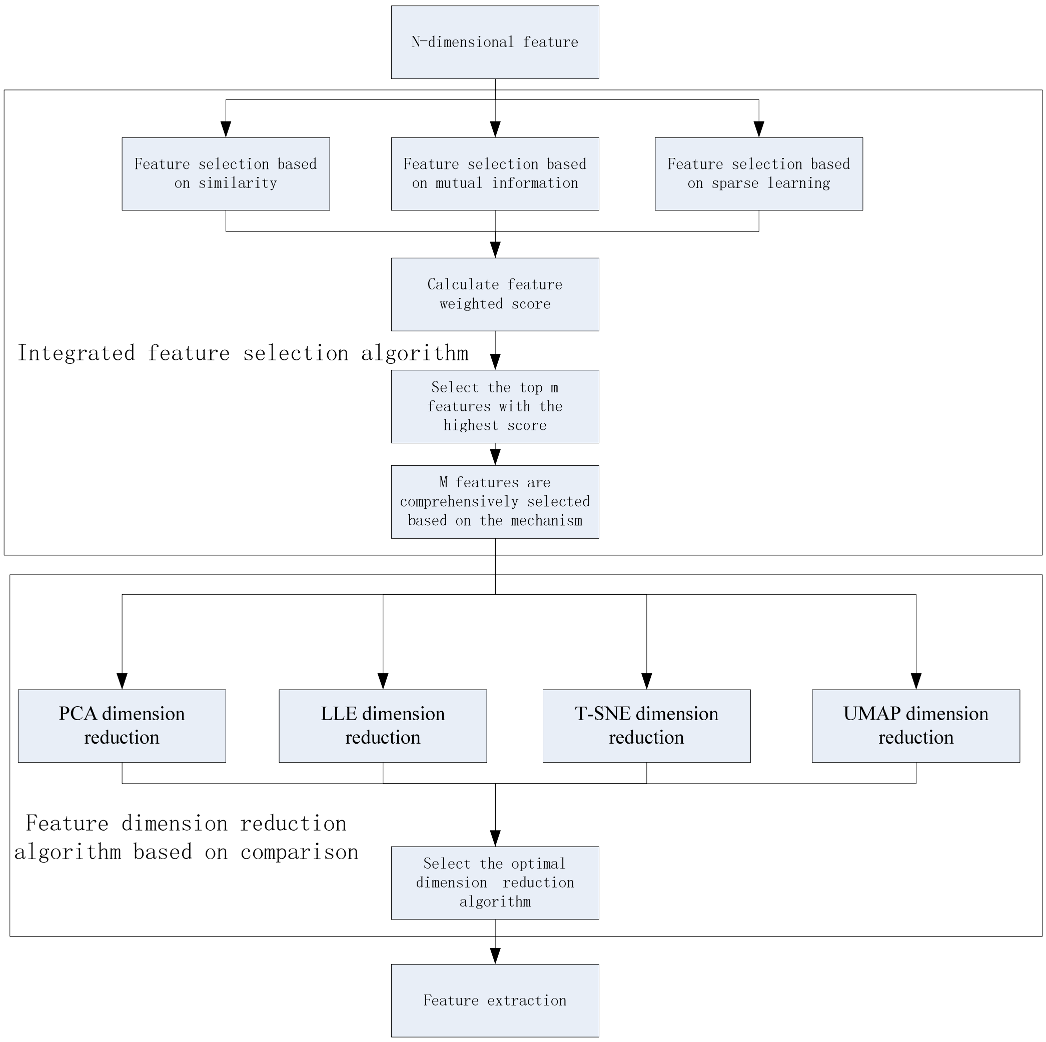

Fig 2. The flow of feature selection and dimensionality reduction algorithm based on integration

3.1 Integrated Feature Selection Algorithm

The integrated feature selection algorithm is used to eliminate completely useless features. The specific implementation steps are as follows:

Assume the sample set after data cleaning, $X = \{x_1, x_2, \dots, x_n\}$

1. The $\text{score}_i(f_k)$ are obtained respectively for feature K by feature selection based on similarity, mutual information and sparse learning

2. The weighted sum of the scores for each feature in the three methods is shown in formula (1)

$$\text{score}(f_k) = \sum_{j=1}^{3} score_j(f_k) \tag{1}$$

3. Select the top m features with the highest score;

4. According to the equipment mechanism, comprehensively select the first m key features that affect energy consumption.

Through this algorithm, this paper selects 20 important energy consumption features from 32 energy dimensions.

3.2 Feature Dimension Reduction Algorithm based on Comparison

Feature dimension reduction algorithm based on comparison can combine key features, reduce spatial dimension and reduce the role of model overfitting. The specific implementation steps are as follows:

1. PCA, LLE, T-SNE and UMAP dimension reduction are carried out respectively for the first m key features obtained by the integrated feature selection algorithm;

2. The effect of each algorithm is evaluated through the model evaluation index, and the optimal dimension reduction algorithm is selected as the final feature extraction result

The system evaluates each algorithm through three model evaluation indexes including MAE, RMSE and MVE. Finally, the results of the feature dimension reduction algorithm based on comparison are shown in Figure 3. In the figure, ①, ②, ③ and ④ respectively represent PCA, LLE, T-SNE and UMAP dimension reduction. It can be seen that UMAP dimension reduction has the best effect

The paper reduces the dimension from 20 important features to 13 key features through UMAP dimension reduction.

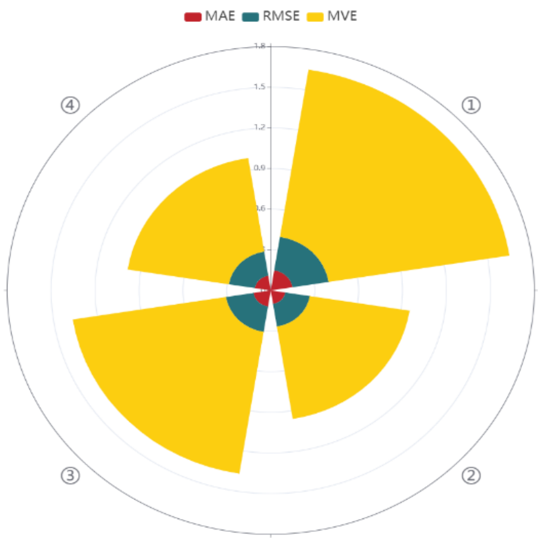

Fig 3. Feature dimension reduction algorithms based on comparison

3.3 Reward Function Creation

The reward function is very important for reinforcement learning. The model is difficult to converge if the designed reward function can't distinguish good or bad actions; On the contrary, a good reward function helps to strengthen the convergence of the learning model and output high-quality actions.

The goal of this paper is to save energy for oven equipment. The energy consumption of oven equipment is equal to the process energy consumption plus the emission energy consumption, as shown in formula (2). Process energy consumption must be met, so our goal can be changed to minimize emission energy consumption. The emission energy consumption includes air volume and temperature, and the trade-off between the two needs to be considered in the design.

$$\mathrm{Energy_{all}} = \mathrm{Energy_{process}} + \mathrm{Energy_{discharge}} \tag{2}$$

According to the above objectives, the definition of reward function includes the following:

1. The reward value is positive when the temperature and frequency are lower than their average values, otherwise it is negative;

2. The reward value $\mathrm{Reward_1}$ is defined by comparing the discharge temperature with the average value, as shown in formula (3);

$$Reward_1 = \begin{cases} \dfrac{1}{\frac{\sigma(temperature)}{mean(temperature)-y_{model}}} \\ -\dfrac{\lambda}{\frac{\sigma(temperature)}{y_{model}-mean(temperature)}} \end{cases}$$
$$, if\ T_{model} < mean(temperature)$$
$$, if\ T_{model} > mean(temperature) \tag{3}$$

3. The reward value $\mathrm{Reward_2}$ is defined by comparing the frequency and average value of each heating box and exhaust fan, as shown in formula (4);

$$Reward_2 = \begin{cases} \dfrac{1}{\frac{\sigma(frequency)}{mean(frequency)-y_{model}}} \\ -\dfrac{\lambda}{\frac{\sigma(frequency)}{y_{model}-mean(frequency)}} \end{cases}$$
$$, if\ F_{model} < mean(frequency)$$
$$, if\ F_{model} > mean(frequency) \tag{4}$$

4. Pull-in penalty item. The maximum temperature and frequency with no intervention were set as thresholds. At the same time, compare the temperature and frequency with the threshold value. If the control exceeds the thresholds, punish it punishment item $\mathrm{Reward_3}$ as reward, as shown in formula

$$Reward_3 = \begin{cases} -100,\ if\ T_{model} \geq max_{temperature} \\ \quad 0,\ if\ T_{model} < max_{temperature} \end{cases}$$
$$or\ F_{model} \geq max_{frequency}$$
$$or\ \ F_{model} < max_{frequency} \tag{5}$$

Sum the reward values to obtain the final reward value Reward, as shown in formula (6).

$$\mathrm{Reward} = \mathrm{Reward_1} + \mathrm{Reward_2} + \mathrm{Reward_3} \tag{6}$$

The reward value can better evaluate the energy consumption for oven equipment.

4. Model Training based on DDPG Algorithm

DDPG defines the problem as a standard reinforcement learning problem, which model environment E as a Markov decision process [6] with state space S, action space A, initial state $\mathrm{p(s_1)}$, transition probability $\mathrm{p(s_{t+1}|s_t, a_t)}$, and reward function $\mathrm{r_t = r(s_t, a_t)}$. The agent selects different actions $\mathrm{a_t \in R^N}$ at the time t interactting with the environment. After each interaction, the environment releases a new state s_{t+1}, the behavior of the agent can be defined π by the random policy, which refers to the action mapping according to the environmental state $\pi: s_t -> a_t$ taken by the agent.

4.1 Integrated Feature Selection Algorithm

The paper needs to define the state space (environment) to model the problem as a situation that reinforcement learning can deal with. 12 parameters to be optimized are obtained by the feature selection and dimension reduction algorithm based on integration. The state space needs enough environmental variables to output accurate parameter values by neural network. There are 574 environmental variables in the state space in this paper, which are the first 2 hours plus the current time data (one point in 3 minutes, 41 points contain the current time data). The data includes 12 parameters to be optimized, 1 parameter of body data and 1 parameter of energy emission data.

4.2 Action Space Definition

DDPG action space is composed of 12 adjustable key features in this paper for model optimization. It can be divided into two agents according to the number of car bodies in view of the different process parameter requirements between the oven equipment state with car and no car, which the agents define different action spaces. The action space of the two agents is shown in Table 1.

Table 1. Action space of agent

data name	production agent	standby agent
incinerator furnace temperature	[690,710]	[500,710]
frequency of incinerator exhaust fan	[40,46]	[35,46]
frequency of circulating fan of heating preservation box zone 1	[42,48]	[37,48]
temperature of heating preservation box zone 1	[180,220]	[160,220]
frequency of circulating fan of heating box zone 4	[47,50]	[41,50]
temperature of heating box zone 4	[190,220]	[160,220]
frequency of circulating fan of heating box zone 3	[47,50]	[41,50]
temperature of heating box zone 3	[185,220]	[160,220]
frequency of circulating fan of heating box zone 2	[43,49]	[37,49]
temperature of heating box zone 2	[120,150]	[100,150]
frequency of circulating fan of heating box zone 1	[43,47]	[37,47]
temperature of heating box zone 1	[85,115]	[85,115]

4.3 Model Network Structure

The DDPG model includes actors and critic [7]. Actor is composed of three hidden layers and one full connection layer. The network structure is $574X32X64X32X12$, of which 574 is the input dimension that is the number of environment variables in the state space and 12 is the output dimension that is the number of action variables in the action space; The excitation function of forward propagation is a rectified linear unit (relu). Critic is also composed of three hidden layers and one full connection layer. The network structure is $586X32X64X32X12$, of which 586 is the input dimension that is the sum of the number of environment variables in the state space and the number of action variables in the action space; The excitation function of forward propagation is also relu. The structure of the target network of actor and critic is the same as their respective network structures.

4.4 Model Super Parameter

To improve the accuracy of the model, the paper need to continuously optimize the super parameter of the model. After optimization, the values of model super parameters are shown in Table 2.

Table 2. Super parameter value

super parameter	value	remarks
optimizer	Adam	update network structure
ϒ	0.99	calculate the reward value of the target reviewer network
learning rate	0.01	learning step
τ	0.05	update target network structure
batch	128	
training cycle	100	
length of trace	100	number of environment iterations
buffer capacity	300	the sample size that the buffer can hold

4.5 Model Training

After completing the above steps, this paper starts DDPG model training and participates in the optimization of oven equipment parameters in real time.

5. Experimental Result

Figure 4 shows the score (total reward value) of energy consumption control algorithm for oven equipment based on DDPG during 1000 training times. It can be seen from the figure that the reward value fluctuates each time during the training process. There are two main reasons for this phenomenon. 1. The initial environment of each training is different. 2. The algorithm adds random noise to each strategy exploration. However, from the change trend of the overall reward value, the total reward value shows a steady upward trend during the training process and the reward value reaches the saturation value after about the 118th training that the total reward value is close to 3000, which indicates that the agent has completed the training.

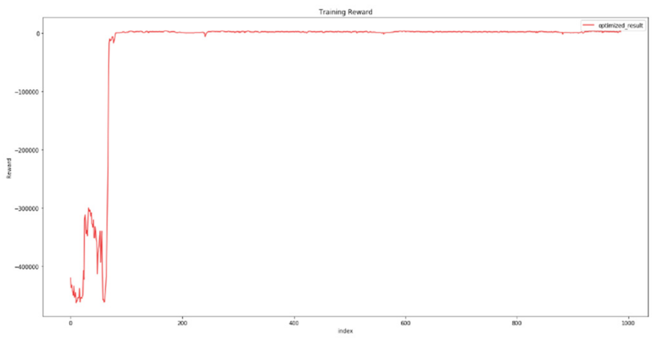

Fig 4. Reward value tracking

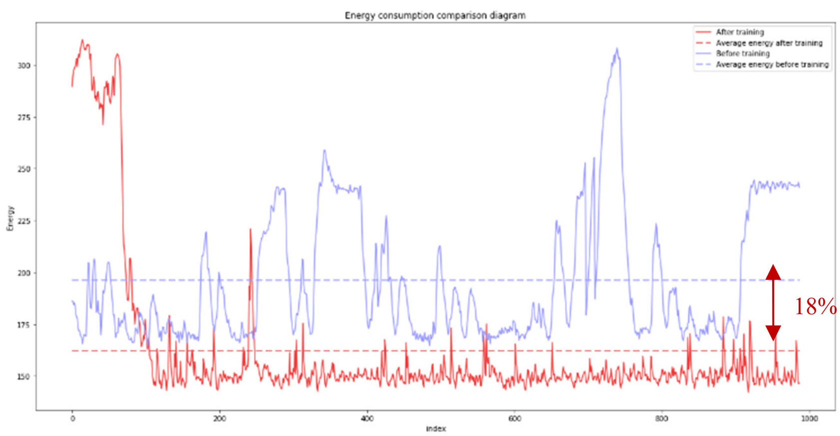

Fig 5. Comparison of energy consumption before and after control

The change of energy consumption during the training process is shown in Figure 5. It can be seen that the energy consumption fluctuated greatly, the agent was constantly exploring and seeking greater rewards before the 118th training. After that, the energy consumption tended to stabilize at the value of about 160. As can be seen from the figure, the parameters have been continuously explored to obtain the lowest energy consumption by the training process. Finally, after being controlled by DDPG algorithm, the energy consumption can be saved by about 18%.

6. Summary

This paper realizes an energy consumption control method for oven based on DDPG and experiments verify its effectiveness and practicability. This paper proposes a feature selection and dimension reduction algorithm based on integration to handle the data with characteristics of oven equipment's high dimension and easily lead to overfitting, which can extract the key features for oven equipment and improve the upper limit of the model. At the same time, this paper can realize the real-time control of the key parameters for oven equipment by creating the reward function and building the DDPG model, which can save energy by about 18%. However, the real-time control of the oven equipment by this algorithm will cause the fluctuation of key parameters and increase the risk of equipment failure.

References

[1] R.S. Suttion, and A.G. Barto, " Reinforcement Learning," A Bradford Book, vol. 15, no. 7, pp. 665-685, 1998.

[2] X. Yuan, Y. Yang, and J. Yang, " Study on the application of reinforcement learning in the operation optimization of HVAC system," Building Simulation, pp. 1-13, 2020.

[3] L.Y. Hu, J.P. Chen, and Q. M. Fu, "Building energy efficiency-oriented reinforcement learning adaptive control method," Computer Engineering and Applications, vol. 53, no. 21, pp. 239-246, 2017.

[4] K. Dalamagkidis, D. Kolokotsa, and K. Kalaitzakis, " Reinforcement learning for energy conservation and comfort in buildings," Building and Environment, vol. 42, no. 7, pp. 2686-2698, 2007.

[5] H.Y. Jiao, H.D. Feng, and D. Wei, "Energy saving control of subway station air conditioning system based on Reinforcement Learning," Control and Decision.

[6] F. Zhang, L.Y. Liu, and X.X. Guo, "Multistage group decision making model based on improved Q-learning algorithm," Control and Decision, vol. 34, no. 9, pp. 1917-1922, 2019.

[7] L. Chen, C. Liang, and J. Y. Zhang, " A Multi-agent reinforcement learning algorithm based on improved ddpg under actor critical framework," Control and Decision.