THERMAL STATION MODELLING AND OPTIMAL CONTROL BASED ON DEEP LEARNING

by

Min CAO*

Zhengzhou University of Technology, Zhengzhou, China

Original scientific paper
https://doi.org/10.2298/TSCI2104965C

To solve the mismatch between heating quantity and demand of thermal stations, an optimized control method based on depth deterministic strategy gradient was proposed in this paper. In this paper, long short-time memory deep learning algorithm is used to model the thermal power station, and then the depth deterministic strategy gradient control algorithm is used to solve the water supply flow sequence of the primary side of the thermal power station in combination with the operation mechanism of the central heating system. In this paper, a large number of historical working condition data of a thermal station are used to carry out simulation experiment, and the results show that the method is effective, which can realize the on-demand heating of the thermal station a certain extent and improve the utilization rate of heat.

Key words: *deterministic strategy, short- and long-term memory networks gradient thermal power station optimization control, deep learning algorithm*

Introduction

With the popularization of the central heating system in northern Cities of China, how to achieve the optimal control objectives of energy conservation and emission reduction under the premise of ensuring the heating quality has become a hot research topic among scholars. A great deal of research has been done on the modelling and optimal control of the central heating system by Chinese and foreign scholars. Some scholars proposed establishing a simulation model of urban secondary heating networks by combining structure mechanisms with experimental identification methods. Some scholars have established a simulation model of *power-power-power-Load-Energy storage* in the whole process of heat energy production, transmission, distribution, and consumption of heating systems. A multi-objective and multi-cycle optimization model of regional energy system design and operation strategy has been proposed. Some scholars proposed a real-time optimization model of multi-source and complementary urban heating system load scheduling based on particle swarm optimization algorithm and used PHP to get each index's feasible weight region. Because the central heating system of non-linear, multivariable, strong coupling, large hysteresis properties, such as the traditional mechanism, mechanism of combining experimental method cannot accurately thermal station model is set-up, but the neural network has strong ability of fitting, can learn characteristics of the internal relation between data, so he *uses* the length of time memory (LSTM) neural network and back propagation (BP) neural network, respectively thermal station model is set up, prepare

* Author's e-mail: caomin54213@163.com

the way for subsequent optimization control. Deep reinforcement learning is a new end-to-end algorithm that combines deep learning with the perceptual ability and reinforcement learning with decision-making ability. Among them, a deep deterministic strategy gradient (DDPG) based on actor-critic is an algorithm to solve continuous state space tasks. It *uses* experiences a playback mechanism to reduce the correlation of continuous action. At present, some scholars have applied DDPG to financial stock management. Some scholars use the DDPG algorithm to realize session scheduling [1]. Some scholars have introduced DDPG to energy management. Some scholars combine DDPG with a power grid cutting machine. It shows the advantage of the DDPG algorithm in solving problems in continuous action space. Therefore, because of the phenomenon that the heat supplied by heat stations is sometimes much more considerable than or far less than the demand, this paper proposes an optimal control strategy based on the DDPG algorithm for the primary side of heat stations and realizes the on-demand heating of heat stations by adjusting the water supply flow on the primary side.

Principle of DDPG algorithm

The actor-critic algorithm

Actor-critic (AC) is made up of the strategic network (actor) and the evaluation network (critic), which evaluates whether the actions generated by the actor-network are good or bad. The actor-network modifies the actions according to the evaluation of the critic network. Its structure is shown in fig. 1.

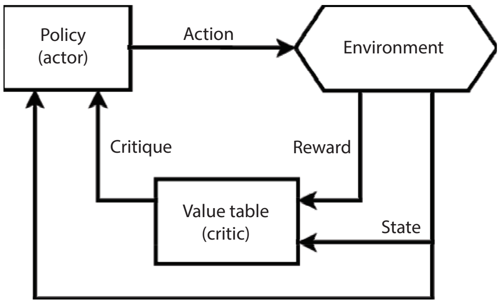

Figure 1. Structure of AC

The workflow of the AC network is the actor-network generates actions based on the current state of the environment. The environment according to the action give return. The critic network evaluates actions. The actor-network adjusts the next output action according to the evaluation of the critic network, that is, adjusts the strategy. The network revises the evaluation criteria based on return R. The cycle continues until the network converges or reaches a set threshold. The AC network *uses* time difference (TD) method to update the network step by step. The TD uses the next state's value function estimate the current value function, which is characterized by low variance, low deviation, and fast convergence. The mathematical expression:

$$V(s_t) \leftarrow V(s_t) + a[r_{t+1} + \gamma V(s_{t+1}) - V(s_t)] \tag{1}$$

where s_t, s_{t+1} are the state of the agent at time t, $t + 1$, respectively, $V(s_t)$, $V(s_{t+1})$ represent t, $t + 1$ are the value function at the time, α means update step size, r_{t+1} – the return at time $t + 1$, γ stands for discount factor, $r_{t+1} + \gamma V(s_{t+1})$ stands for TD target, which can be estimated by using boot strapping method. The $\delta_t = r_{t+1} + \gamma V(s_{t+1}) - V(s_t)$ is the TD deviation. The TD calculates the TD error based on the value function output from the actor-network, and then it *uses* the strategy gradient to update the actor-network parameters to realize the optimization of the AC network.

The DDPG algorithm

The AC involves two neural networks, and in the continuous state, there is correlation before and after each parameter update, which sometimes leads to the one-sidedness of neural network update and even the problem that it is impossible to learn things. To solve this problem, the AC algorithm is modified, and the DDPG algorithm is obtained. The DDPG network comprises the AC leading network and the AC target network and adds memory libraries. Each time the leading network interacts with the environment, it produces a set of samples and puts them into the memory bank, then taken out at random when needed, thus reducing the samples' correlation. The AC target network has the same structure as the AC target network, but with different parameters. The AC target network regularly passes parameters to the AC target network, which updates the adversary network's arguments by calculating the error values of the corresponding value functions generated by the adversary network and the target, thereby optimizing the arguments of the adversary network. In the actor-arguments leading network, the actor leading network implements parameter optimization according to the leading network's argument. The DDPG network structure successfully solves the problem that the AC network cannot learn anything in continuous action prediction, thus realizing the control of continuous action spatial sequence [2].

Operation principle of the central heating system

The central heating system is composed of a heat source, heat station, and heat user. The heating network is responsible for connecting each part, as shown in fig. 2. The primary side is between the thermal station and the heat source, and the secondary side is between the thermal station and the heat user. The heat source generates heat and enters the heat station through the primary side heating network. The heat is transferred to the secondary side heating network in the heat station through the heat exchanger, and the heat enters the heat users along with the secondary side heating network [3]. The primary side of the heat station is mainly studied here. Therefore, a simplified treatment is made for the heat source and heat users, assuming that the heat source produces enough heat, and the number of heat users is fixed.

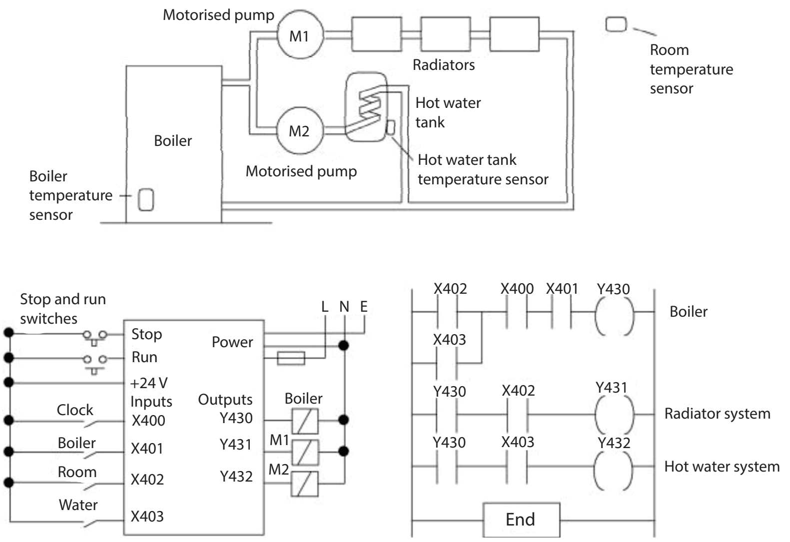

Figure 2. Structure of the central heating system

As can be seen from fig. 2, there is a regulating valve at the primary side inlet of the thermal station, which can control the water supply flow, and the water supply flow affects the heat supply quantity. Therefore, the problem of optimal control of heat supply quantity is transformed into optimizing the set value of water supply flow at the primary side of a heat station. Local regulation is adopted here for the thermal power station.

Primary side model of the heating station of central heating system

Establishing a relatively accurate thermal station model is based on optimal control of the central heating system. According to the thermal station system's working principle, the experimental system's dynamic model is determined:

$$y = f(u_1, u_2, u_3, t) \tag{2}$$

where u_1 is the opening degree of the regulating valve on the primary side of the heat power station (primary water supply flow), u_2 – the water supply temperature at the primary side of the thermal power station, u_3 – the outdoor temperature, and y – the heat supply of a heat station. The primary water supply flow, primary water supply temperature, and outdoor temperature are taken as input, and the heat supply quantity is taken as output. The thermal station is treated as a black box to learn the relationship between its input and output and establish the thermal station's heating model. The BP neural network and LSTM neural network were, respectively used to model the thermal power station. Finally, the modelling effect was compared, and an appropriate modelling method was found [4].

The LSTM neural network

The LSTM is an improved network-based on a circulating neural network, and its cell unit structure is shown in fig. 3. The LSTM neural network controls cell state by deliberately adding input gate, forgetting gate and output gate, and makes information pass selectively through *gate*, that is, adding information or removing information. The c_{t-1}'s information, h_{t-1} and x_t enter the input gate at the same time to provide updated candidate data for the cell. The forgetting gate determines the data be forgotten based on c_{t-1}, h_{t-1} and x_t. The c_t replaces the candidate data in the input gate with the forgotten data selected by the forgetting gate in a particular proportion complete the update c_t. Finally, c_t outputs the result h_t at time T. The LSTM neural network limits the cell's capacity through the gate so that necessary information can be stored in the cell's limited memory capacity.

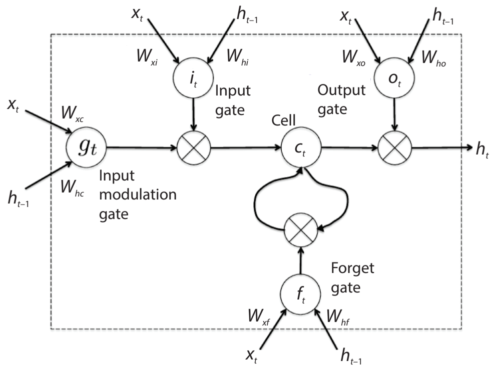

Figure 3. Schematic diagram of LSTM unit structure

Unlike the traditional neural network, LSTM neural network takes into account the influence of the data of the previous moment on the current moment's data. When processing time-series data characteristics, it can better mine the internal relationships among the data. The thermal station produces data with time-sequence characteristics. The LSTM model of the thermal station reflects the data timing characteristics of the thermal station and solves the problem of long-term data dependence [5].

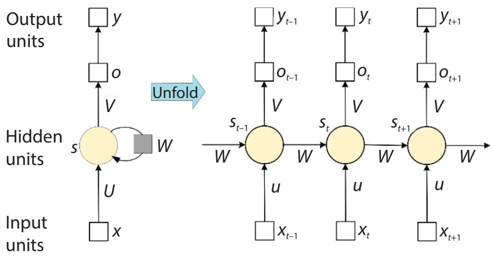

Figure 4. The LSTM-based thermal station model

Thermal station model based on LSTM

The model adopts a structure of 3-3-1, that is, three input variables, three intermediate hidden layers, and one output variable, as shown in fig. 4.

Figure 4, u_1 and u_2 are the primary water supply flow and primary side water supply temperature at T of the thermal power station, u_3 is the outdoor temperature at time T, y – the heating quantity at time $T + 1$ of the heat station. After many experiments, we found that when the number of hidden layers is 3, and the number of nodes is 20, the output result is better. The setting of other parameters is shown in tab. 1.

Table 1. Parameters of the thermal station model based on LSTM

Parameter names	Parameter value
Number of input variables	3
Number of output variables	1
Size of input data for each group	24
Time step	1
Hidden layer node	20
Number of hidden layers	3
Vector	0. 01
Number of iterations	1 000

The BP based thermal station model

The model adopts a 3-1-1 structure: 3 input variables, 1 hidden layer, and 1 output variable. Its structure is shown in fig. 5. Relevant parameters of the model were obtained through a large number of experimental analyses, as shown in tab. 2 [6].

Table 2. The BP based thermal station model parameters

Parameter names	Parameter value
Input variables per variable	3
Output variables per variable	1
Hidden layer nodes	7
Vector	0. 1
Number of iterations	1000
Training deviation	0.001

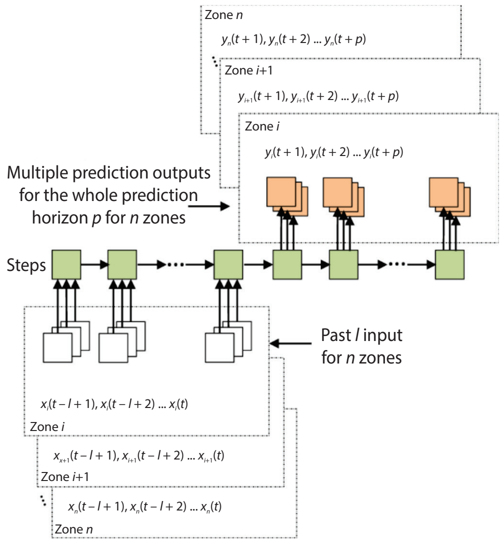

Figure 5. The BP based thermal station model

Analysis of modelling results

The data adopted the historical working condition data of a heat power station. The paper selected the data from December 1 to 30th, 2018, for 30 days in a row, and sampled the data at an interval of 1 hour, 24 times a day, as shown in tab. 3. The outdoor temperature was obtained from China Weather Net. We processed the collected data and screened the data that differed significantly from other data for elimination [7]. The values of three of the input variables are not of the same order of magnitude. To avoid the prediction result being dominated by some eigenvalues with too large dimensions, the input data is standardized.

Table 3. Training data

Time	Primary side water supply flow [TH^{-1}]	Primary side water supply temperature [°C]	Outdoor temperature [°C]	Heating load [GJ]
2018/12/1 0:00	74.33	82.4	–1	11.74
2018/12/1 1:00	76.29	82.5	–2	11.96
2018/12/30 22:00	79.8	82.29	–12	12.18
2018/12/30 23: 00	80.3	82.29	–13	12.41

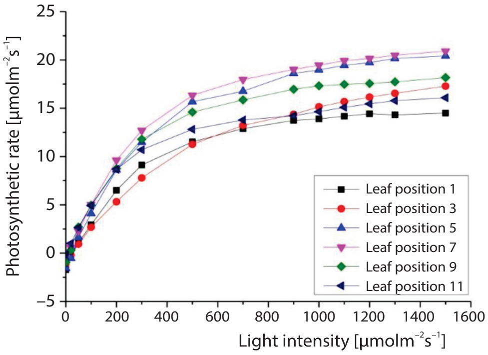

Figure 6. The BP based thermal station model results

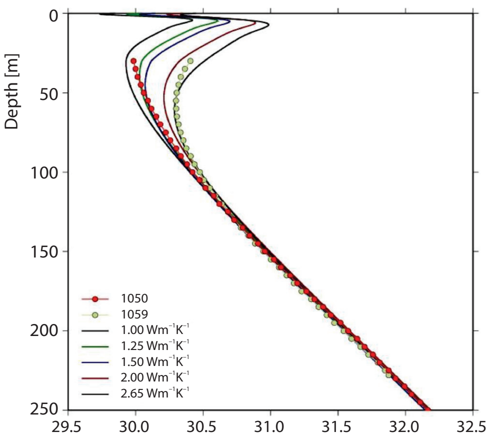

Figure 7. The LSTM thermal station model results

A total of 744 sets of data were used for the modelling experiment of the model structure. We selected 24 sets of data on December 31, 2018, as the test set and visualized the data's test effect. The BP based thermal station model results are shown in fig. 6. The model's predicted value is roughly consistent with the trend of the real value change, and the maximum relative error is 3%.

Modelling results based on LSTM thermal station are shown in fig. 7. The predicted value trend is closer to the real value, and the maximum relative error is 2.1%. We can see that the thermal station model based on LSTM is more accurate than that based on BP, which meets the heating quantity control precision required by the heating company. Therefore, the LSTM algorithm is used to model the thermal station [8].

The DBP based primary side optimization control of thermal power station

Data selection

In the DDP based primary side optimization control for thermal power stations, data from January 1 to February 27, 2019, were selected as training data, and the sampling interval was 1 hour, as shown in tab. 4.

Table 4. Training data table

Time	Primary side water supply flow [TH^{-1}]	Primary side water supply temperature [°C]	Outdoor temperature [°C]	Heating load [GJ]
2019/1/1 0:00	82.29	79.7	–13	12.54
2019/1/1 1:00	82.29	81.0	–13	13.02
2019/2/27 22:00	64.01	80.3	–6	10.18
2019/02/27 23: 00	63.36	80.7	–6	10.45

Among them, the outdoor temperature adopts the data of China Weather Network. The historical working condition data of a thermal power station is used to supply the primary side's water supply temperature. We take the KTH + 1dst side water supply temperature, the $k + 1$dst outdoor temperature, and the KTH + 1dst side heating quantity generated by the actor main network as the heat station model's input. In the central heating system, the difference between the two days before and after the primary side water supply temperature is relatively small. Here the $k + 1$ day primary side water supply temperature is replaced by the k day primary side water supply temperature [9].

Short-term thermal load prediction of heat stations

The short term heat load prediction of a heat station is set to make the heat user get suitable heating needs. Based on the 24 hours outdoor temperature sequence of the same day, the 24 hours outdoor temperature sequence of the previous day, and the heat supply quantity sequence of the previous day as inputs, a BP heat load prediction model was established to predict the change of heat load in the future 24 hours range, and the results of this model were taken as a part of the performance index function of the primary side optimization control of the heat station.

Design of performance index functions

The performance index function is set to determine the heating quantity obtained by the DDPG control algorithm:

$$U_i = \left| Q_1 - Q_p \right| \tag{3}$$

where U_i is the absolute value of the heating quantity error of the heat station, Q_p – the heating quantity value output by DDPG control algorithm, and Q_1 – the short-term thermal load forecast output value of the heat station. The DDPG control algorithm outputs r according to U_i. As U_i gets smaller, r gets bigger. The larger r indicates that the water supply flow sequence at the primary side of the heating station output by the actor leading network in the DDPG control algorithm is more reasonable. According to the thermal power station's fundamental operating principle, it also needs to meet the constraints:

$$\begin{gathered} T_{1g} < T_{\text{heat source}} \\ T_{1g\min} < T_{1g} < T_{1g\max} \\ Q_p < Q_{\text{heat source}} \end{gathered} \tag{4}$$

where T_{1g} is the water supply temperature on the primary side and meets the upper and lower limits, $T_{\text{heat source}}$ – the outlet temperature of the heat source, and Q_p – the amount of heat added to a heat station.

The overall design of the DDPG control algorithm

We combine the DDPG control algorithm with the thermal station's primary side model to form a coherent system. The system learns the continuous action space to realize the optimized water supply flow sequence of the primary side of the thermal station by optimizing the primary side's control. The block diagram of the optimal control of the thermal station's primary side based on the DDPG algorithm is shown in fig. 8.

Actor main and target networks are composed of two full connection layers, one using the ReLU activation function, the other using Sigmoid activation function, according to the wa-

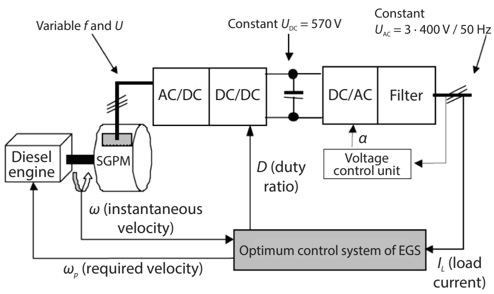

Figure 8. Block diagram of primary side optimization control of thermal power station based on DDPG algorithm

ter supply flow limit scope worth to the primary network water supply flow. The main and target networks use a full connection layer to evaluate the primary water supply flow and the primary heat supply, using ReLU activation functions. The learning rate of the actor host and target network is set as 0.00001, the learning rate of the critic and target network is set as 0.00001, the updating rate of four network parameters is 0.01, and the memory storage capacity is 3000.

The flow of the DDPG control algorithm is as follows.

- (1) Randomly initialize actor and critic main network parameters and assign them to the corresponding parameters of actor and critic target network, respectively, and initialize state S_t simultaneously.
- (2) Actor generates a_t based on s_t input into the thermal station model and the outdoor temperature and primary water supply flow at the corresponding time, and output s_{t+1}, which is compared with the predicted thermal load (performance index function) to meet the performance index function generate a return r_t.
- (3) Store the sample data (s_t, a_t, r_t, s_{t+1}) at time T in the memory bank, then assign the value of s_{t+1} to s_t continue the execution.
- (4) Generate another set of (s_t, a_t, r_t, s_{t+1}) and put it in the memory bank until the number of data reaches the capacity set by the memory bank.
- (5) When the memory database is full, randomly sample the conversion data (s_i, a_t, r_i, s_{i+1}) and train it as a unit input data set of actor, target network. Actor tirget network outputs a_{i+1} according to S_{i+1}, and the target network outputs future discount return γq_{i+1} and value function $r_i + \gamma q_{i+1}$ according to s_i+1 and N.
- (6) Calculate the loss function value $L = E[r_i + \gamma q_i + q_{i+1} - q_i)^2$ of a_{i+1} converted data by using the value function q_{i+1} output from the leading critic network and the value function $r_i + \gamma q_{i+1}$ output at the corresponding moment of the target network, then use Adams method to optimize the leading critic network. The main critic network *uses* the strategic gradient to optimize the actor leading network. The actor leading network assigns corresponding parameters to the actor target network, respectively.
- (7) Loop (4)~(6) until the time step is completed.

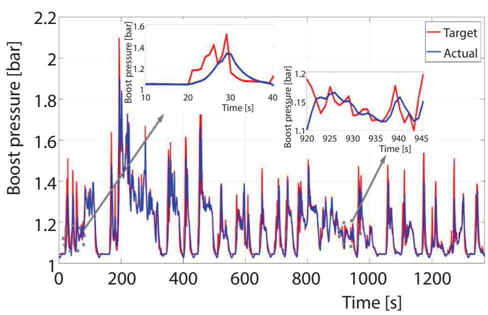

Figure 9. Output results of thermal power station based on DDPG algorithm

Simulation verification

Software environment: the computer operating system for Windows 10, operating environment for Tensorflow1.0, the programming language for Python3.5. We selected 24 sets of data on February 28, 2019, as the test data to verify the DDPG based primary side optimization control model's generalization ability for thermal power stations. The output results of the thermal power station based on the DDPG algorithm are shown in fig. 9.

In fig. 9, the blue line represents the predicted value of the thermal station's short-term heat load, and the maximum error between the predicted value and the actual heat load of the thermal station is guaranteed to be within 5%. Therefore, the predicted value of the thermal station's short-term heat load is approximately the actual heat load for reference. The green line represents the heating quantity obtained by the DDPG control algorithm. It can be seen that the heating quantity controlled by the DDPG algorithm is basically consistent with the predicted value of the short-term heat load of the heat station.

Conclusion

Aiming at the problem of heat supply and demand do not match, through a large number of actual data and combined with a variety of factors affecting thermal station is established a lateral heat DDPG control model of optimization, the model of the output thermal station heating quantity compared with prediction heat load, according to meet the degree of the sexual index function adjust a net flow of water supply, thus achieve thermal heat to each according to his need. The simulation results show that the control scheme can obtain the optimized flow sequence of primary water supply and realize the heating station's goal on demand.

References

[1] Schwalbe, K., Hoffmann, K. H., Optimal Control of an Endoreversible Solar Power Plant, *Journal of Non-Equilibrium Thermodynamics*, *43* (2018), 3, pp. 255-271
[2] Delfino, F., *et al*., An Energy Management Platform for the Optimal Control of Active and Reactive Powers in Sustainable Microgrids, *IEEE Transactions on Industry Applications*, *55* (2019), 6, pp. 7146-7156
[3] Ouammi, A., Achour, *et al*., Supervisory Model Predictive Control for Optimal Energy Management of Networked Smart Greenhouses Integrated Microgrid, *IEEE Transactions on Automation Science and Engineering*, *17* (2019), 1, pp. 117-128
[4] Wu, S., Study and Evaluation of Clustering Algorithm for Solubility and Thermodynamic Data of Glycerol Derivatives, *Thermal Science*, *23* (2019), 5, pp. 2867-2875
[5] Yang, Y., *et al*., Optimal Control Methods for Photovoltaic Enabled Grid with Abnormal Power Loads , *IEEE Systems Journal*, *15* (2019), 1, pp. 852-861
[6] Ali, U., *et al*., Motion-Communication Co-Optimization with Cooperative Load Transfer in Mobile Robotics: An Optimal Control Perspective, *IEEE Transactions on Control of Network Systems*, *6* (2018), 2 pp. 621-632
[7] Yin, L., *et al*., Real-Time Thermal Management of Open-Cathode PEMFC System Based on Maximum Efficiency Control Strategy, *Asian Journal of Control*, *21* (2019), 4, pp. 1796-1810
[8] Fontes, F. A., *et al*., Optimal Control of Thermostatic Loads for Planning Aggregate Consumption: Characterization of Solution and Explicit Strategies, I*EEE Control Systems Letters*, *3* (2019), 4, pp. 877-882
[9] Wu, S., Construction of Visual 3-D Fabric Reinforced Composite Thermal Perfomance Prediction System, *Thermal Science*, *23* (2019), 5, pp. 2857-2865

Paper submitted: November 16, 2020
Paper revised: December 28, 2020
Paper accepted: January 16, 2021