2021 IEEE/ACIS 22nd International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD) | 978-1-6654-0403-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/SNPD51163.2021.9704994

Fuzzy Q-learning Control for Temperature Systems

Yeong-Chin Chen
Department of Computer Science and Information Engineering,
Asia University, Taiwan
e-mail: ycchenster@gmail.com

Lon-Chen Hung*
Department of Electronic Engineering,
Lunghwa University of Science and Technology, Taiwan
*e -mail: hlc02468@gm.lhu.edu.tw

Mariana Syamsudin
Department of Computer Science and Information Engineering,
Asia University, Taiwan
e-mail: 107221004@gm.asia.edu.tw

***Abstract*—In this paper, the reinforcement learning algorithm applied to temperature control of the internet of things (IoT), which aims to develop a multi-purpose intelligent micro-power control switch to achieve advanced temperature control research. This paper is based on the fuzzy Q-learning PID control algorithm based on reinforcement learning, with LinkIt Smart 7688 Duo platform. The error value between the set temperature and the actual sensed temperature is exposed to the reinforcement learning PID control operation. Specifically, a temperature sensor will provide temperature feedback to the LinkIt Smart 7688 Duo in order to achieve the stated temperature control. Finally, the suggested control approach will be compared to PID control to illustrate its efficacy and performance.**

Keywords—reinforcement learning, IoT, fuzzy q-learning, multi-agent, temperature

I. INTRODUCTION

Over recent years, due to the popularization of the Internet and mobile devices, the development of Internet of Things (IoT) in various fields has been accelerated, combining resources from different industries to extend existing technologies. Among the wide-ranging applications of the IoT, the one that is closest to life and has the most related applications is undoubtedly the smart home, which combines daily life with the concept of the IoT to achieve the goal of interconnecting people and things. Smart home is to embed home electrical equipment with control chips and sensors, thereby it has the ability of smart sensing to achieve the concept of interconnection of things.

In terms of electric power control, by adjusting the input power of the heat or heat dissipation (cooling) element, the output conditions of the element can be adjusted to change the temperature of the controlled environment to achieve the purpose of constant temperature control [1-4]. In view of the PID algorithm [5,6] control parameters (k_p , k_i , k_d) due to temperature control components, controlled temperature environment, different settings or adjustments are required for factors such as the height of the controlled temperature to increase the temperature control efficiency.

Although PID control is a commonly used control theory in general manufacturing, however, as the control application of IoT system, different environmental conditions and temperature control systems, the adjustment of its control parameters is quite time- consuming. Since the PID controller needs to adjust its three gain parameters. There are many ways to find proper control of these gain parameters. The most famous is the Zeigler-Nichols adjustment method. This benefit method is easy to implement. Conversely, an empirical technique generally results in a substantial overshoot. Intelligent PID controllers have been created using integrated intelligent technology to address this restriction. However, the present application of intelligent PID control methods in temperature control systems is restricted and only exists in off-line approaches [7-9].

These technologies are mainly offline genetic algorithms, ant colony algorithms, particle swarm optimization technology, multi-objective extreme value optimization, fuzzy and neural network algorithms, etc. These approaches compute the PID controller gain in relation to the fitness function. However, the intelligent algorithms, such as genetic algorithms, have been utilized to enhance the performance of PID by maintaining the temperature control operation close to the ideal state. The primary drawback of these approaches is that the controller can only be fine-tuned off-line and cannot be adapted to environmental changes.

To overcome these constraints, reinforcement learning algorithms are used to train effective adaptive gain adjustment techniques without the need for human intuition; in the instance of temperature, no prior knowledge of the temperature platform is required. The adaptive fine-tuning method described in this study is based on the following two concepts: 1) it should not rely on an adaptive gain adjustment strategy based on human intuition, 2) it should be able to have better temperature heating without further control of efficiency rate of change. In order to realize this fine-tuning mechanism, one of the reinforcement learning (Q-learning) is applied. It develops an efficient adaptive gain adjustment technique by interacting with a particular environment since it does not require any prior knowledge of the temperature platform.

Given a target state arbitrarily, let the control algorithm adjust the state of the system to the target state as fast as possible. The *Q*-learning is value-based algorithm in the reinforcement learning algorithm. The environment responds to this action, and get the next observation, through the continuous cycle of the above process. [10,11].

Since the *Q*-learning algorithm can only search in discrete action spaces, the fuzzy *Q*-learning is an extension method of fuzzy logic control and *Q*-learning, which can process continuous action spaces without precise mathematical models. According to the requirements of classic controller design, modeling the control platform representation through closed differentiable equations will make the system more complicated. The adaptability of *Q*-learning can be used in fuzzy inference systems, where actions and *Q* functions can be inferred from fuzzy rules.

The main motivation used the fuzzy Q-learning (FQL) PID control for temperature systems that could perform in the micro-power IoT platform. Using the fuzzy Q-learning method, which the gain of the PID controller by adjusting online. The MAS is composed of three main agents, each of which can only modify one gain, which allow the gain value to be updated online without any need for specialized expertise, through contact with the environment.

This paper uses the low-cost LinkIt Smart 7688 Duo component board as the controller to design a fuzzy Q-learning PID control method based on reinforcement learning, then use the mobile device or input the set temperature control on the remote server. The controller finds the appropriate PID parameters according to the fuzzy Q-learning multi-agent, and then calculates the appropriate power percentage.

II. METHODOLGY

This LinkIt Smart 7688 Duo component board connected to the cloud platform through a Wi-Fi device, and the temperature of the constant temperature plant is sent back to the cloud for storage in real time to achieve the purpose of monitoring the constant temperature plant, as shown in Fig. 1.

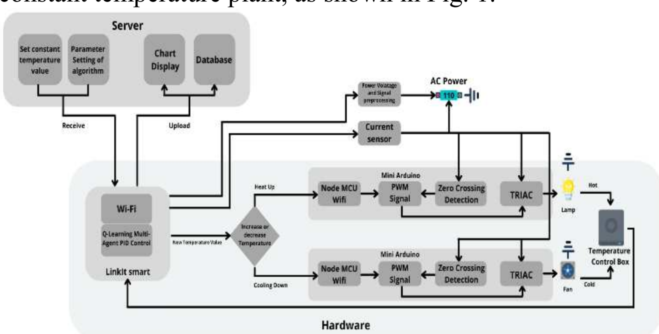

Fig. 1. The Temperature Control Process of IoT Architecture

In order to allow the heating or cooling device to quickly and stably adjust the temperature to the set desired value, this research establishes an intelligent temperature control environment. The LinkIt smart 7688 Duo can measure the current ambient temperature through a temperature sensor and upload the data to the database via Wi-Fi, users see the current temperature through a browser or mobile device and set the target temperature through API and then return it, the LinkIt smart 7688 Duo can control the output PWM wave through the algorithm to control the TRIC solid-state switch with the control ratio, so that the fan and the light bulb can raise and lower the temperature; In addition, a step-down circuit is also set up to process the power signal through step-down and bias processing, and the ADC (analog-digital converter) function captures the voltage signal, and uses the Hall current sensor module to measure the current. Next, calculate the electric power, and finally display various information on the server webpage with remote.

2.1 The Reinforcement Learning

Reinforcement learning is a type of algorithmic learning which being inspired by living creatures. The purpose of reinforcement learning is to devise a method that optimizes expectations while minimizing reward. The maximum expected discount reward is explored in the space of possible state-behavior pairs. The performance of beneficial actions when executed in a given state will be rewarded by positive enhancement signals [12,13].

2.2 The *Q*-learning

The Q-learning(reinforcement learning) with the agent calculates the Q function, which is used to evaluate the future discount rewards of the candidate actions. Given state x and action a, the function output is indicated by a letter $Q(x)$. When the Q value is executed in the state, the update of each action is as follows:

$$Q^{'}(x,a) := Q(x,a) + \eta(R(x,a,x^{'}) + \gamma_{a}^{\max} Q^{'}(x,a) - Q(x,a)) \quad (1)$$

whereas $Q^{'}(x,a)$ is the reward $R(x,a,x^{'})$ obtained from the new value of the newly updated state-action combination by having out the activity in the state x. The Q-learning model implies that the agent will proceed from state x by executing the optimum strategy, so it is in the state $\max_{a} Q^{'}(x,a)$. The learning rate η defines the new learning level information covering the old information, and the discount coefficient γ defines future rewards[14].

Furthermore, the Q-learning agent selects the action to be performed out in the state and determines the outcome state x. It gets rewards $R(x,a,x^{'})$, and assuming the optimal policy from the state x is implemented, it updates the combined value (x,a). This is how the Q-learning agent uses the exploration strategy to try to maximize feedback as a strategy. The feature of using Q-learning is a model-free method. However, it has the drawback of not being applicable to a continuous action state space. The fuzzy function is used as an approximation to address this problem.

2.3 The Fuzzy Q-learning

One of the primary benefits of fuzzy logic systems is the ability to produce better approximations in the Q function and use fuzzy Q-learning in continuous state space issues. In fuzzy Q-learning, x is the input that defines the state of the agent as a clear set, which is transformed into fuzzy values, with each fuzzy rule representing a state. Therefore, the excitation intensity of each rule represents the extent to which the agent is in a certain condition, and the result is generated by each rule through the exploration algorithm. Therefore, the rules of the fuzzy logic system have the following form:

$$\begin{aligned} &\text{If } x \text{ is } \mathbf{x}_j \text{ then } \alpha[j,1] \\ &\qquad \text{or } \alpha[j,2] \\ &\qquad \vdots \\ &\qquad \text{or } \alpha[j,k] \end{aligned}$$

The fuzzy Q-learning algorithm are described [15,16]:

1. Observe the state x
2. By using exploration algorithm, choose an action for each trigger rule.
3. Through using following formula, compute the global output $a(x)$:

$$a(x)=\sum_{j=1}^{N} w_j(x)\alpha_j \Big/ \sum_{j=1}^{N} w_j(x) \tag{2}$$

where α_j relates to the determined action of the rule.

4. As follows, determine the equivalen value $Q(x,a)$:

$$Q(x,a)=\sum_{j=1}^{N} w_j(x)q[j,j^{'}] \Big/ \sum_{j=1}^{N} w_j(x) \tag{3}$$

where $q(j,j')$ is the value that stimulates the rule and used the exploration strategy.

5. Observe the new state x' after performing action $a(x)$.
6. Determine the reward $R(x,a, x')$
7. Update the value of q as shown below:

$$\Delta q(j,j^{'})=\eta\Delta Q\{\sum_{j=1}^{N} w_j(x)q[j,j^{'}] \Big/ \sum_{j=1}^{N} w_j(x)\} \tag{4}$$

where

$$\Delta Q=R(x,a,x^{'})+\gamma(Q^{'}(x^{'},a^{*})-Q(x,a,Q(x^{'},a^{*}))$$

$$=\sum_{j=1}^{N} w_j(x)q[j,j^{'}] \Big/ \sum_{j=1}^{N} w_j(x) \tag{5}$$

and $q[j,j']$ is the selection triggering the action with the largest Q value in the rule j'.

$$q^{'}[j,j^{'}]:=q[j,j^{'}]+\eta(R(x,a,x^{'})+\gamma Q(x^{'},a^{*})-Q(x,a))[w_j(x)/\sum_{j=1}^{N}w_j(x)] \tag{6}$$

A multi-agent system (MAS) is a computing system composed of multiple agents interacting in an environment. It can be used to solve problems that are difficult to solve by separate agents and single-layer systems. According to MAS, it can be expressed as a single subject with multiple states and action vectors, and traditional Q-learning can be applied. According to this algorithm, the disadvantage is that the state action space of the agent will increase exponentially due to the increase in the number.

In order to overcome this shortcoming, various research methods have been proposed, the most famous of which are coordinated learning, distributed value function and independent learners. This will make the agent not know the Q value or strategy of other agents. Each agent stores and updates its Q table, and each Q_j is based on the global state x.

$$Q_j(x,a_j)\leftarrow Q_j(x,a_j)+\eta(R_j(x,a,x^{'})+\gamma_a^{\max}Q^{'}(x,a)-Q(x,a)] \tag{7}$$

The MAS of the temperature control system in this study is composed of three agents. As shown in Fig. 2, the action of each agent adjusts the gain of each parameter of the PID controller [16]. The MAS has two input signals, which are the temperature error. These signals determine the global state of the MAS [17].

The MAS consisted of three agents, each one for one gain of the PID controller. There were the three agents to adjust the gains (k_p, k_i, k_d), respectively. The input domain of the state variable in the range [1,-1] that the present states of the MAS are specified using fuzzy logic. The seven membership functions are NB, NM, NS, ZE, PS, PM, and PB denote Negative Big, Negative Medium, Negative Small, Zero, Positive Small, Positive Medium, and Positive Big, respectively.

The reinforcement signal of reward (R) is defined by utilizing a specific qualitative knowledge focused on a system behavior description.

$$R=\begin{cases} \dfrac{1}{1+|e'|}-\dfrac{1}{1+|e|}, & |e(t)|\le\varepsilon, |e(t)|\le|e(t-1)|\,(reward) \\ 0, & |e(t)||y(t)|>0 \\ -R(t)t_p, & otherwise\ (punishment) \end{cases} \tag{8}$$

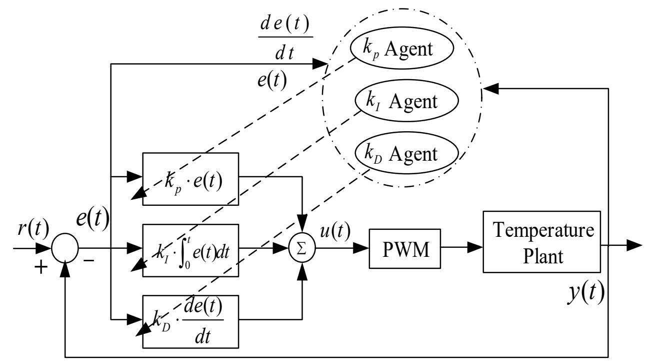

Fig. 2. The Fuzzy Q-learning PID Control for Temperature System

III. RESEARCH RESULTS

Use WAMP (Windows + Apache + MySQL + PHP) to set up a server integration package, and to write RWD web pages to provide a mobile device interface. The user interface, the main settings include constant temperature value and constant temperature schedule. The cloud software part uses PHP to compose a user interface to store data in a self-built server, as shown in Fig. 3. The heating rate determined by the FQLC is utilized to manage the heating power of the heating device through the PWM duty cycle, allowing the temperature of the constant temperature bath to achieve the specified temperature in the fastest time possible.

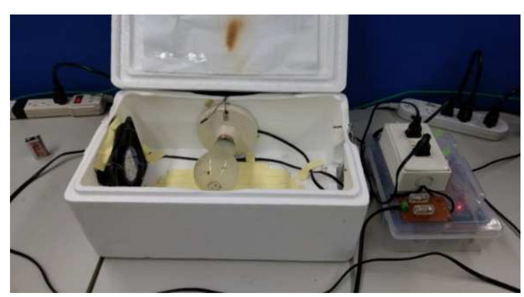

Fig. 3. The Intelligent Temperature Control Based on IoT System

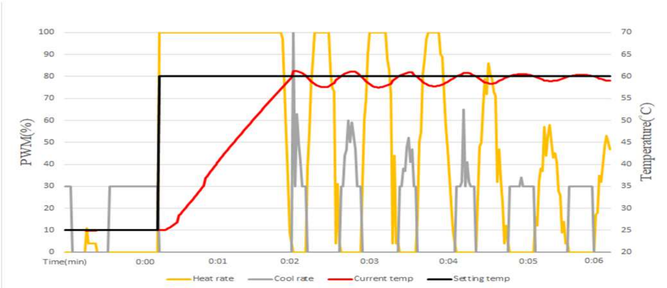

Fig. 4. The PID controller for the Temperature of 60 Degrees Response

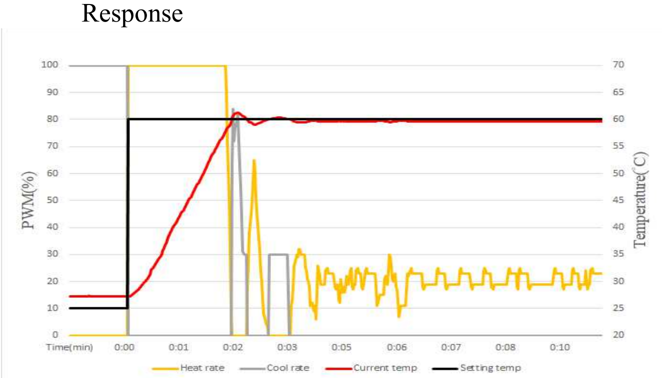

Fig. 5. The Fuzzy Q-learning PID Controller for the Temperature of 60 Degrees Response

Table 1 Comparison of Temperature Control Output Response

Spec. / Controller	Rise time(s)	Settling time(s)	Peak time(s)	Max overshoot
PID control	126	120	132	11.67%
FQL PID control	118	105	126	10.33%

As shown in Fig. 4, the PID controller employs a proportionate effect to minimize temperature deviation and accelerate the adjustment process; when the proportional impact is weak, the PID adjustment process become slow. Similarly, when the temperature control platform changes, its PID control response is relatively slow.

The proposed the fuzzy *Q*-learning PID controller to intelligently adjust the temperature in real time in Fig. 5. The results show that, compared with the PID controller, the fuzzy Q-learning PID controller provides better performance by reducing the time to reach the steady state in Table 1.

IV. CONCLUSIONS

This paper uses low-cost components, cooperates with Arduino development chip, based on fuzzy *Q*-learning PID control algorithm and adjusts PWM technology for intelligent micro power control switch device. Place the smart thermostat in multiple environments that need to be used, and use the built-in cloud server to present it in the form of web pages. A complete monitoring system has been achieved, allowing users or maintenance personnel to access the Internet. The ambient temperature can be monitored without going to the site. The cloud system will automatically store the data in the database for future data analysis and environmental research. The alarm system allows system maintenance personnel to use the system's intelligent alarm prompts to quickly find problems and efficient maintenance, so that the system can maintain high-precision control, and the constant temperature can be more stable.

Special Session 4: AI Applications for Power and Energy Engineering

REFERENCES

[1] H. Yan, Z. Chao, J. Q. Shen, L. X. Gao, L. V. Wei, "Design of water temperature monitor dystem based on MCU," Chinese Journal Full-Text Database, Communications Technology, 2007.

[2] N. Pachauri, A. Rani, V. Singh, "Bioreactor temperature control using modified fractional order IMC-PID for ethanol production chemical Engineering Research and Design, Vol. 122, pp. 97-112, 2017.

[3] M. Kumar, D. Prasad, B. S. Giri, R. S. Singh, "Temperature control of fermentation bioreactor for ethanol production Using IMC-PID Controller," Biotechnology Reports, Vol. 22, 2019.

[4] J. L. Song, W. L. Cheng, Z. M. Xu, S. Yuan, M. Hou Liu, "Study on PID temperature control performance of a novel PTC material with room temperature curie point," International Journal of Heat and Mass Transfer, Vol. 95, pp. 1038-1046, 2016.

[5] B. S. Taysom, C. D. Sorensen, John D. Hedengren, "A comparison of model predictive control and PID temperature control in friction stir welding," Journal of Manufacturing Processes, vol. 29, pp. 232-241, 2017.

[6] Visioli, Antonio, Practical PID Control, Springer Verlag, 2006.

[7] H. Liu, Y. Li, Y. Zhang, Y. Chen, J. Qian, "Intelligent tuning method of PID parameters based on iterative learning control for atomic force," Microscopy Micron, Vol. 104, pp. 26-36, 2018.

[8] S. Dequan, G. Guili, G. Zhiwei, X. Peng, "Application of expert fuzzy PID method for temperature control of heating furnace," Procedia Engineering, vol. 29, pp. 257-261, 2012.

[9] V. Tavakol, I. Eksin, "Fuzzy PID controller design using q-learning algorithm with a manipulated reward function," European Control Conference, pp. 2502-2507, 2015.

[10] I. Carlucho, M. D. Paula, S. A. Villar, G. G. Acosta, "Incremental q-learning strategy for adaptive PID control of mobile robots," Expert Systems with Applications, vol. 80, pp. 183-199, 2017.

[11] I. Carlucho, M. D. Paula, G. G. Acosta, "Double q-PID algorithm for mobile robot control," Expert Systems with Applications, vol. 137, pp. 292-307, 2019.

[12] I. Grondman, L. Busoniu, G. Lopes, R. Babuska, "A survey of actor-critic reinforcement learning: standard and natural policy gradients," IEEE Trans. Syst. Man Cybern. Part C, vol. 42, no. 6, pp. 1291-1307, 2012.

[13] I. Grondman, M. Vaandrager, L. Busoniu, R. Babuska, E. Schuitema, "Efficient model learning methods for actor critic control," IEEE Trans. Syst. Man Cybern. Part B: Cybern. vol. 42 no. 3, 591-602, 2012.

[14] Y. P. Pane, S. P. Nageshrao, J. Kober, R. Babuka, "Reinforcement learning based compensation methods for robot manipulators," Engineering Applications of Artificial Intelligence, vol. 73, no.10-12, pp. 2274-2279, 2019.

[15] P. Kofinas, A. I. Dounis, "Fuzzy q-learning agent for online tuning of PID controller for DC motor speed control," Algorithm, vol. 11, no.10, pp. 1-13, 2018.

[16] P. Kofinas, A. I. Dounis, "Online tuning of a PID controller with a fuzzy reinforcement learning MAS for flow rate control of a desalination unit," Electronics, vol. 8, no. 231, pp. 1-16.

[17] D.P. Kwok, Z.D. Deng, C.K. Li, T.P. Leung, Z.Q. Sun, J.C.K. Wong, "Fuzzy neural control of systems with unknown dynamic using q-learning strategies," IEEE International Conference on Fuzzy Systems, vol. 2, pp. 25-28, 2003.