[Edited.V.2]Future_ICT_2021_paper_51-p423

by Turnitin Indonesia

Proposal of Hybrid Controller Based on Reinforcement Learning for Temperature System

Mariana Syamsudin*
107221004@gm.asia.edu.tw

Abstract

The ability of a modern controller in addressing fast convergence of existing adaptive PID controllers has a restriction. This case applies to several actuators, including the thermostat system. In this project, Reinforcement Learning (RL) is applied to the challenge in the automatic tuning of a proportional-integral-derivative controller. Two means of testing procedures will be carried out in this project to achieve the objectives at investigating the unprecedented *performance of the hybrid controller*. Firstly, the researchers combine the asynchronous learning structure of the Asynchronous-Advantage-Actor-Critic (A3-C) with the incremental PID system controller. Secondly, the researchers also unite a PID system with a deep-deterministic-policy-gradient (DDPG). Both actor network structure and critic network structure are back-propagation neural networks with three-layer structure. An extensive review of the scientific literature pertaining to the hybrid controller is additionally provided.

1 Introduction

Proportional-Integral-Derivative-Controllers (PIDCs) are used in process control by the vast majority of operating controllers. The evolutionary of the adaptive PID controller has been developed and utilized to solve a wide range of control engineering problems, principally applied to various actuators, especially in temperature monitoring systems for food, medicine, plant cultivation and production processes. Constant temperature control of manufactured products is essential to maintain product quality. Moreover, it also serves as the means to prevent damage in the production process caused by instrument overheating failure and malfunction.

Nevertheless, traditional controllers are very dependent on variables. Controllers are frequently under-tuned and fail to achieve appropriate performances, especially in circumstances involving application in practice, given the environment's strong uncertainty and non-linearity. To comply with the prerequisite for self-tuning-PID variables, various controllers have been developed to address the drawbacks of the method. PID-controller advancement can be broadly categorized into three areas from the start into the latest, i.e. the fuzzy-PID controller, neural-network-PID controller and reinforcement-learning-PID controller (Sun et al., 2019).

In the recent past, Fuzzy control has evolved into a viable alternative to traditional control techniques (Shi et al., 2020). In order to cope with complex tasks and integrate the benefits of traditional controllers with human operator knowledge, the fuzzy PID controller suggested to update the parameter values by inquiring the fuzzy matrices table. The drawback of this approach is the fact it necessitates a lot more previous experience. Furthermore, this method requires a huge number of parameters to be tuned. Conventional control devices, on the other hand, contain only three parameters for tuning which can be adjusted through experimentation or by employing tuning principles found within control literature such as Ziegler Nichols techniques (Boubertakh et al., 2010).

In addition, an auto-tune-PID like controller based on Neural-Networks (NN) is proposed for an underwater vehicle. PID can be automatically estimated by NN to achieve stability of the online controller. The system managed to reach the smaller one position tracking error (Hernández-Alvarado et al., 2016). Back Propagation is an alternative NN approach that may be employed for learning and storing a large number of mapping connections of the input-output system. Furthermore, there is no necessity to reveal the formula in mathematics that defines these mapping connections in advance (Zhu et al., 2018).

Reinforcement learning (RL) algorithm has obtained current popularity and extensively implemented in the control engineering community to generate innovative control strategies. There are three out of four main types of RL methods that are frequently referred to as model-free except Model Based. They are Value-Based, Policy-Gradient, and Actor-Critic with their own distinctive advantage (Shin et al., 2019). For example, Value-Based sampling is more efficient and consistent, which is figure out by Q-Learning algorithm and all of its improvements, such as Deep-Q-Networks, Dueling-Double-Q-Networks. A substantial number of researchers have applied the RL algorithm, one of them (Hindersah & Rijanto, 2013) implemented the Q-learning algorithm to generate a control signal to control a self-balanced robot. In research (Younesi & Shayeghi, 2019), Q-learning algorithm was adopted to generate additional force for correcting the output of a pre-tuned PID controller on the fixed weight of the PID controller. Due to its additional advantages to handling control systems, numerous studies applied the Q-learning algorithm as a tuning method to find a proper set of parameters for multiple PID controllers (Shi et al., 2018).

Another approach is policy-gradient which has a faster and better convergence for continuous and stochastic environments. Moreover, after merging those two algorithms, the Actor-Critic algorithm was born. It aims to diminish all their weaknesses and keep the superiority of all features from both value-based and policy-gradient.

Many researchers have implemented reinforcement learning to optimize temperature control or heating control in numerous areas. However, the primary findings of this study are a complete assessment of the existing literature on the use of the Actor-Critic method to autonomously tune a proportional-integral-derivative controller and an outline of relevant areas for further research.

The following is an outline of this paper. Section 2 will illustrate the fundamental theory of PID and RL algorithm. A general overview of thermostat hybrid controllers will be discussed in Section 3. Last but not least, an outline of potential areas of future research and conclusion will be provided in Section 4.

2 Algorithms

2.1 Basic Structure of PID

Conventional PID controller is comprised of three parts: -proportional, -integral, and -derivative part as illustrated in Fig. 1 below:

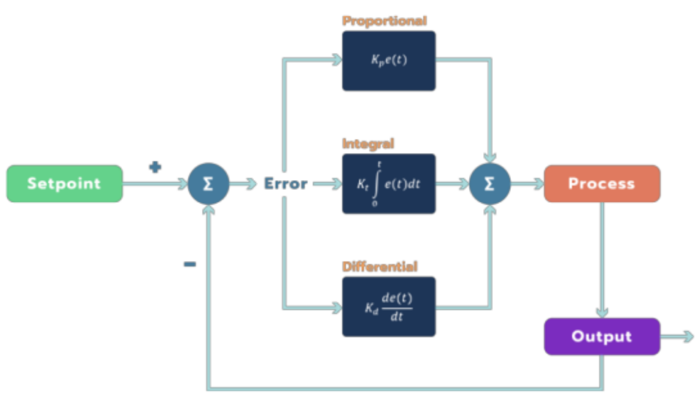

Figure 1: Structure of PID controller

The PID control system in Figure 1 is a linear controller which is based on the closed-loop control system. The PID controller process is presented in the following Equation (1). The proportional, integral, and differential parts of Equation (1) are constituted by a linear combination of the control amount used to track the target control system.

$$u(t) = K_p \left[e(t) + \frac{1}{T_I} \int_0^t e(t)dt + T_D \frac{d}{dt} e(t) \right] \tag{1}$$

Here, K_p is a proportional coefficient, T_I is an integral coefficient, and T_D is a differential coefficient.

2.2 Reinforcement Learning

Reinforcement-Learning (RL) is a subset of machine learning approaches which enables an agent to learn by trial and error utilizing feedback gathered from its behaviours and encounters in a collaborative setting. RL is more like a loop. Its objective is to achieve some goals or planning for the future by the concept of time. Figure 2 illustrates the interaction of the agent and its environment.

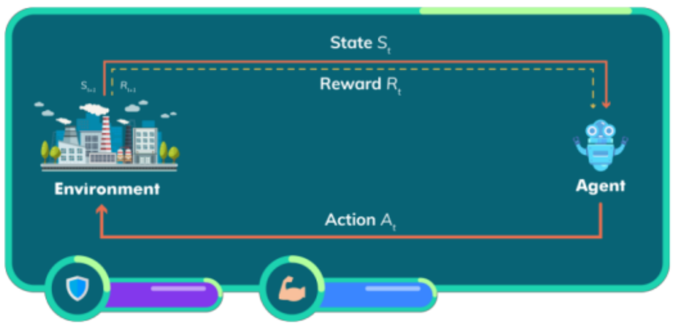

Figure 2: Reinforcement Learning in general terms

The elements of RL sequences are described as follow :

- Environment: Physical word
- State: Ongoing situation observed in the environment
- Reward: Feedback received at each step
- Policy: Method to determine agent's action to perform.
- Value: Future reward received by agent.

At certain time intervals, an Agent communicates with the surroundings. The Agent commonly has a policy (π) that determines their choice of action. After the action is successfully executed, the environment performs a one-step transition, granting the next state, S_{t+1}, along with feedback in the form of a reward, R_{t+1}. To study policies and improve them, the Agent applies the transitional state form of knowledge (S_t, A_t, S_{t+1}, R_{t+1}).

The numerical representation of a policy π in a particular state s is computed utilizing the formula (2) as follows:

$$v^{\pi}(s) = E[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \,|\, s, \pi] \tag{2}$$

Whereas E denotes the estimated future gain and γ signifies the reduction factor.

To determine the value of the present state, the agent has to examine the anticipated future rewards that the policy will provide in its current state. Meanwhile, γ represents the amount of weight given by the agent for potential rewards.

The Bellman formula determines the value function of states s for the optimal policy π* as follows:

$$v^{\pi*}(s) = maxE[r_{t+1} + \gamma V^*(s_{t+1}) | s, *] \tag{3}$$

2.3 Actor-Critic Approaches

Actor-Critic is one of the Reinforcement Learning algorithms which is formed by two agents, namely The Actor and The Critic. The former deals with selections made according to environmental observations and present policies. Meanwhile, the latter observes environment state and rewards derived from the surroundings in response to decisions made by the Actor. Furthermore, the Critic will also provide feedback to the Actor to determine the next steps or to make decisions.

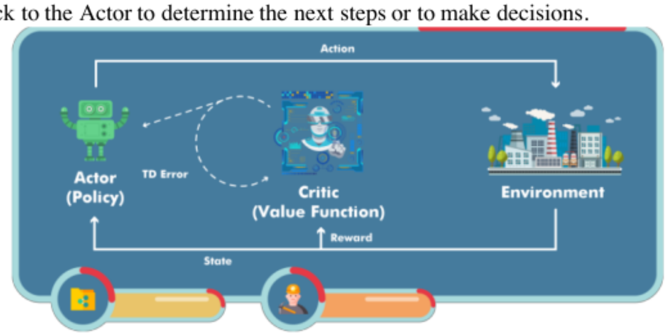

Figure 3: Actor critic environment interaction

Figure 3 shows how the Actor (policy) gets a state from the surroundings and decides which action to take. Simultaneously, the Critic (the value function) gets the previous interaction's state and reward. The Critic updates itself as the actor by using the TD error determined from the aforementioned data.

3 Method

The following part discusses the research approach that will be applied. The architecture of the hybrid temperature system based on the R-L algorithm and the PID controller is discussed in Part 1. The second part describes the learning method that employs the DDPG and A3C algorithms.

3.1 Design of Hybrid Controller

The primary purpose of the project is to develop a multi-purpose-intelligent-micro-power-control switch to attain temperature stability and power consumption monitoring, as shown in Figure 4. The smart switch platform based on NodeMCU and PID controller is consolidated with Pulse Width Modulation (PWM) (Chen et al., 2019).

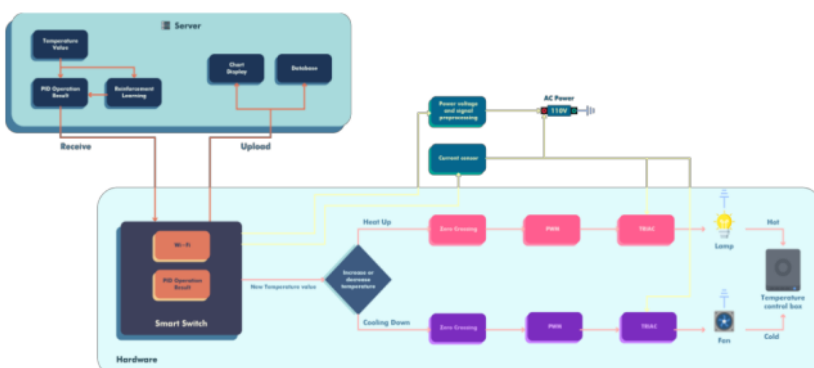

Figure 4: Temperature system architecture diagram

NodeMCU component board is used as an important bridge for data transmission. Whilst mobile device or remote server is used to input the set temperature, afterward the server receives the actual sensing temperature measured by the sensor.

In implementing the temperature control, the controller calculates the appropriate power percentage according to the RL algorithm which is then used as the basis for tuning the PID controller and drives the TRIAC solid-state electronic relay through PWM to achieve the control of heating and cooling components in order to achieve temperature control performance. The sensing value of the temperature sensor can be transmitted back.

The controller is used as the calculation basis for power adjustment. It is then transmitted to the server so that the user can see the temperature at any time and to facilitate the monitoring of the temperature.

3.2 Learning Process

Reinforcement-learning is a challenging machine-learning topic in which unexpected shifts in hyperparameters could contribute to modifications in the performance of models. Therefore, some rules must be considered, such as a speciality and which conditions necessitate a particular approach.

DDPG and A3C are an off-policy algorithm. DDPG utilises the Q-function to learn the policy and uses off-policy data and the Bellman equation to learn the Q-function. Unfortunately, DDPG is only applicable in systems with continuous activity spaces. While using A3C, actor-critic trains and updates asynchronously. Moreover, because numerous instances are running concurrently, it delivers quicker multi-processing and a higher convergence rate.

By the consideration above, two testing procedures will be done in this work to reach the aimed at investigating for a varied *performance of the hybrid controller*. The first is to combine the asynchronous learning structure of Asynchronous advantage actor-critic (A3C) with the incremental PID controller. The latter is to unite the PID system with a deep deterministic policy gradient (DDPG). Both actor network structure and critic network structure use back-propagation neural networks with a three layer structure.

In order to guarantee the results validity of the design, it is essential to do stages in the design of RLPID controller shown in Figure 5.

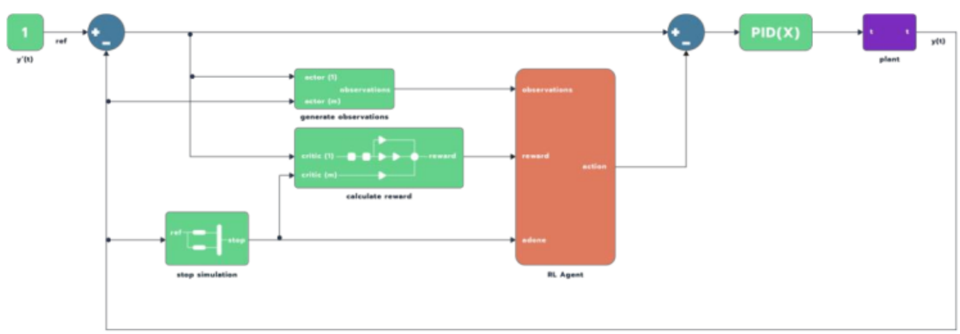

Figure 5: Hybrid control diagram based on Reinforcement Learning

4 Future Research

This prediction is obtained by a data mining process using the Scopus repository. The keywords used are Reinforcement Learning and DDPG or A3C or PID in the searching method. The data that has been analysed is part of the title or abstract, then the VOS viewer will cut the words in the title/abstract to construct and visualize co-occurrence networks of the critical relationship between the word pieces or terms.

Network visualization represents the words of deep-deterministic-policy-gradient or DDPG, deep-reinforcement-learning, reinforcement-learning and controller. These are the most dominant words, which mean those terms often appear in various research publications.

At the same time, network visualization of PID indicates the PID controller confirms the strength relationship in the research topic of reinforcement learning and robots. On the other hand, it shows a weak relationship with deep reinforcement learning, DDPG and A3C algorithm.

This data reveals that research in related fields is still thoroughly limited. It may be regarded as indisputable that the most decisive trend prediction related to the application of RL in the control field is the movement towards the deep reinforcement learning (DRL) algorithm, especially A3C. Nevertheless, there are some challenges of applying RL or DRL to control systems. One of many limitations is the correlation of the training data should be observed since the agent is profoundly dependent on its actions (Arulkumaran et al., 2017). Once the asynchronous-multi-thread training approach minimizes the association of the initial training data, it provides a more reliable and flexible controller, and vice versa. (Sun et al., 2019).

References

Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine*, *34*(6), 26–38. https://doi.org/10.1109/MSP.2017.2743240

Boubertakh, H., Tadjine, M., Glorennec, P., & Labiod, S. (2010). Tuning fuzzy PD and PI controllers using reinforcement learning. *ISA Transactions*, *49*(4), 543–551. https://doi.org/10.1016/j.isatra.2010.05.005

Chen, Y. C., Syamsudin, M., & Xu, W. (2019). An Internet of Things Thermostat Sensor Developed with an Arduino Device Using a Recursively Digital Optimization Algorithm. *Journal of Information Hiding and Multimedia Signal Processing*, *10*(3), 434–446.

Hernández-Alvarado, R., García-Valdovinos, L. G., Salgado-Jiménez, T., Gómez-Espinosa, A., & Fonseca-Navarro, F. (2016). Neural network-based self-tuning PID control for underwater vehicles. *Sensors (Switzerland)*, *16*(9), 1–18. https://doi.org/10.3390/s16091429

Hindersah, H., & Rijanto, E. (2013). *Application of R Reinforcement Learning on Self- Tuning PID C Controller for Soccer Robot ulti-Agent System*.

Shi, Q., Lam, H. K., Xuan, C., & Chen, M. (2020). Adaptive neuro-fuzzy PID controller based on twin delayed deep deterministic policy gradient algorithm. *Neurocomputing*, *402*, 183–194. https://doi.org/10.1016/j.neucom.2020.03.063

Shi, Q., Lam, H., Xiao, B., & Tsai, S. (2018). *Adaptive PID controller based on Q -learning algorithm*. *3*, 235–244. https://doi.org/10.1049/trit.2018.1007

Shin, J., Badgwell, T. A., Liu, K. H., & Lee, J. H. (2019). Reinforcement Learning – Overview of recent progress and implications for process control. *Computers and Chemical Engineering*, *127*, 282–294. https://doi.org/10.1016/j.compchemeng.2019.05.029

Sun, Q., Du, C., Duan, Y., Ren, H., & Li, H. (2019). Design and application of adaptive PID controller based on asynchronous advantage actor–critic learning method. *Wireless Networks*, *0123456789*. https://doi.org/10.1007/s11276-019-02225-x

Younesi, A., & Shayeghi, H. C. A. (2019). *Q-Learning Based Supervisory PID Controller for Damping Frequency Oscillations in a Hybrid Mini / Micro-Grid*. *15*(1), 126–141.

Zhu, J., Song, Y., Jiang, D., & Song, H. (2018). A new deep-q-learning-based transmission scheduling mechanism for the cognitive internet of things. *IEEE Internet of Things Journal*, *5*(4), 2375–2385. https://doi.org/10.1109/JIOT.2017.2759728

[Edited.V.2]Future_ICT_2021_paper_51-p423

ORIGINALITY REPORT

15%	10%	12%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Qian Shi, Hak-Keung Lam, Chengbin Xuan, Ming Chen. "Adaptive neuro-fuzzy PID controller based on twin delayed deep deterministic policy gradient algorithm", Neurocomputing, 2020 Publication	2%
2	d-nb.info Internet Source	2%
3	hal.archives-ouvertes.fr Internet Source	1%
4	www.nxtbook.com Internet Source	1%
5	Vahid Tavakol Aghaei, Ahmet Onat, Ibrahim Eksin, Mujde Guzelkaya. "Fuzzy PID controller design using Q-learning algorithm with a manipulated reward function", 2015 European Control Conference (ECC), 2015 Publication	1%
6	www.frontiersin.org Internet Source	1%

7	"Proceedings of 2021 Chinese Intelligent Systems Conference", Springer Science and Business Media LLC, 2022 Publication	1 %
8	Qifeng Sun, Chengze Du, Youxiang Duan, Hui Ren, Hongqiang Li. "Design and application of adaptive PID controller based on asynchronous advantage actor–critic learning method", Wireless Networks, 2019 Publication	1 %
9	"The 2021 International Conference on Smart Technologies and Systems for Internet of Things", Springer Science and Business Media LLC, 2023 Publication	1 %
10	Submitted to University of Glasgow Student Paper	1 %
11	medium.com Internet Source	1 %
12	www.science.gov Internet Source	1 %
13	Yeong-Chin Chen, Lon-Chen Hung, Mariana Syamsudin. "Fuzzy Q-learning Control for Temperature Systems", 2021 IEEE/ACIS 22nd International Conference on Software Engineering, Artificial Intelligence, Networking	1 %

and Parallel/Distributed Computing (SNPD), 2021
Publication

14	spiral.imperial.ac.uk Internet Source	1 %
15	iris.ucl.ac.uk Internet Source	<1 %
16	Submitted to Liverpool John Moores University Student Paper	<1 %
17	Ye Yang, Chen Chen, Jiangang Lu. "An Enhanced Full-Form Model-Free Adaptive Controller for SISO Discrete-Time Nonlinear Systems", Entropy, 2022 Publication	<1 %
18	kclpure.kcl.ac.uk Internet Source	<1 %
19	www.researchgate.net Internet Source	<1 %
20	Su Whan Sung, Jietae Lee, In‐Beum Lee. "Process Identification and PID Control", Wiley, 2009 Publication	<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography On

[Edited.V.2]Future_ICT_2021_paper_51-p423

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6