Proposal of Hybrid Controller Based on Reinforcement Learning for Temperature System

Mariana Syamsudin*
Department of Computer Science and Information Engineering
Asia University, Taichung City, Taiwan
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 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 (A3C) with the incremental PID 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 used back-propagation neural networks with three-layer structure. A comprehensive review of the literature relating to the hybrid controller is also provided.

1 Introduction

Most of the operating controllers assuredly apply Proportional Integral Derivative Controllers (PIDCs) in process control. The evolutionary of the adaptive PID controller has been developed and utilized to solve a wide range of control engineering problems, principally applies 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, conventional controllers suffer from high dependency on parameters. The controllers usually are not optimally tuned and unable to reach satisfactory performances, especially in the situations associated with practical applications considering the high non-linearity and uncertainty of the environment. In order to satisfy the requirement of self-tuning PID parameters, various controllers have been developed to address the drawbacks of the method. Advancement of PID controller can be roughly classified into three categories 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 become an amiable alternative for conventional control algorithms(Shi et al., 2020). The fuzzy PID controller proposed to adjusts the parameters by querying the fuzzy matrix table to deal with complex processes and combine the advantages of classical controllers and human operator experience. The limitation of this method is that it needs much more

H. C. Chen et al. (Eds.): FICT 2021
https://easychair.org/cfp/FutureICT2021

prior knowledge. Moreover, this method has a large number of parameters that are required to be optimized. In contrast, classical controllers have only three tuning parameters that can be tuned by trial and error, or by using tuning rules available in control literature such as Ziegler Nichols methods (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 achieves stability of online controller. The system managed to reach the smaller one position tracking error (Hernández-Alvarado et al., 2016). Another NN algorithm is Back Propagation that can be used to learn and store plenty of mapping relations of the input-output model. Furthermore, there is no necessity to disclose in advance the mathematical equation that describes these mapping relations (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 ModelBased. They are Value-Based, Policy-Gradient, and Actor-Critic with their own distinctive advantage (Shin et al., 2019). For example, Value-Based is more sample efficient and steady, which is figure out by Q-Learning algorithm and all its enhancements like Deep Q-Networks, Double Dueling QNetworks. A substantial of researcher has 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 have a faster and better convergence for continuous and stochastic environments. Moreover, after merging those two algorithmic, 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 main contributions of this research are to give a comprehensive review of the literature relating to the application of Actor-Critic algorithm to tune a

proportional – integral - derivative controller automatically and to provide outline potential areas for future 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 controller 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 consists of three components, namely proportional, integral and derivative part as illustrated in Fig. 1 below:

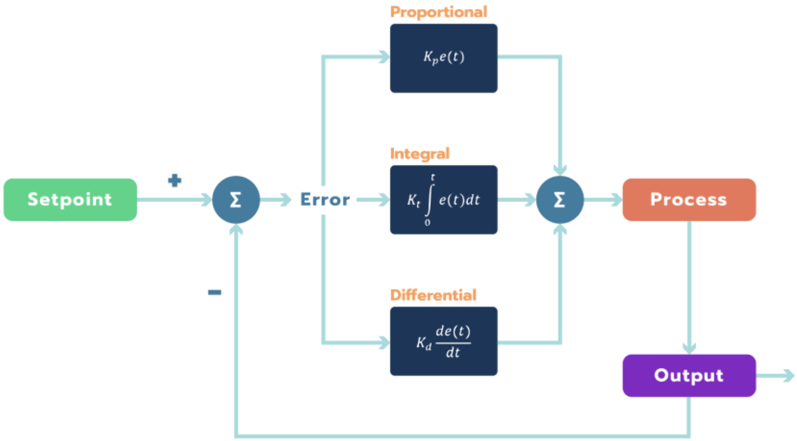

Figure 1: Structure of PID controller

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.

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

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 subgroup of a machine learning technique that enables an agent to trial and error using feedback from its actions and experiences in an interactive environment. 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.

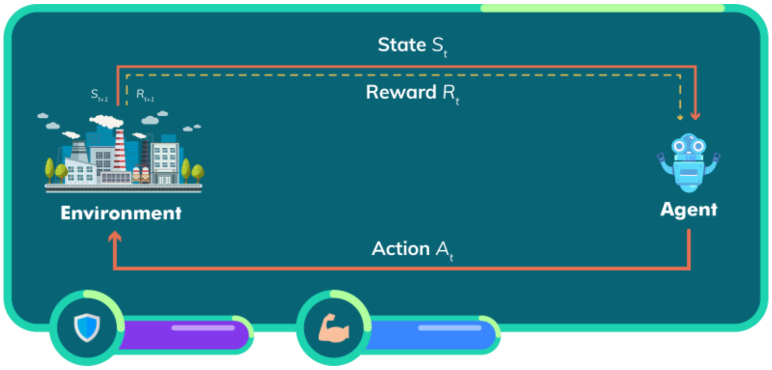

Figure 2: Reinforcement Learning in general terms

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.

The Agent interacts with the environment at different time steps. 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 value of a policy π in a given state s is calculated using the value function in (2) as follows:

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

Where E denotes the expected future return and γ signifies the discount factor.

In order to determine the value of the current state, the agent must first consider the expected future benefits that the policy will undertake at the current state. Meanwhile, γ represents the amount of weight given by the agent for future rewards.

The value function of state s for the optimum policy π^* is stated in (3) and calculated using the Bellman equation as follows:

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

2.3 Actor-Critic Methods

Actor-Critic is one of the Reinforcement Learning algorithms that consists of two agents, namely The Actor and The Critic. The former addresses decisions based on observations of the environment and current policies. Meanwhile, the latter observes environment state and rewards obtained from the environment based on the decisions made by the Actor. Furthermore, the Critic will also provide feedback to the Actor to determine the next steps or to make decisions.

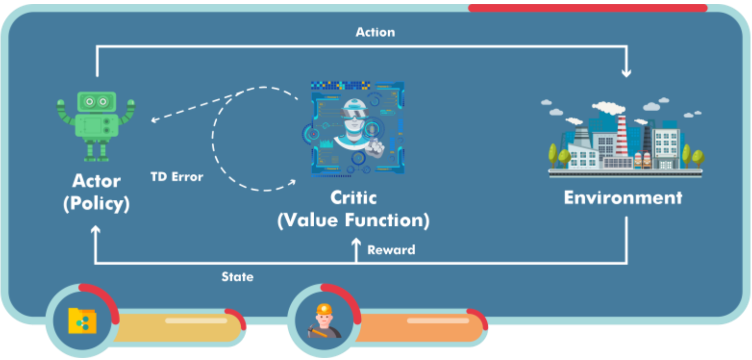

Figure 3: Actor critic environment interaction

Figure 3 illustrates that the Actor (policy) receives a state from the environment and chooses an action to perform. At the same time, the Critic (value function) receives the state and reward resulting from the previous interaction. The Critic employs the TD error calculated from this information to update itself and the actor.

3 Methodology

This section describes the details of the research methodology that will be applied. The architecture of the hybrid temperature system based on the PID controller and the RL algorithm is introduced in Section 1. On the other hand, the learning process using DDPG and A3C algorithm is explained in Section 2.

3.1 Design of Hybrid Controller

The primary purpose of the project is to develop a multi-purpose intelligent micro-power control switch to achieve constant temperature control 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).

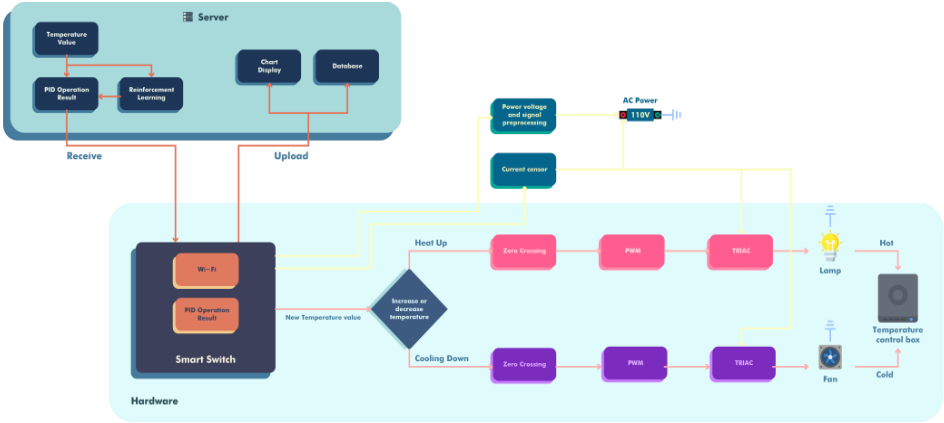

Figure 4: Temperature system architecture diagram

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 tricky machine-learning domain, whereas sudden changes in hyperparameters can lead changes in the models' performance. Therefore, some rules must be considered, such as a speciality and which situation requires which technique.

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 can only be used for environments with continuous action spaces. Whilst, in the use of A3C, actor-critic train and update in an asynchronous manner. Moreover, it provides faster multi-processing and comes up with a faster convergence rate due to multiple instances running concurrently.

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 of actornetwork structure and critic network structure is used back-propagation neural networks with a threelayer structure.

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

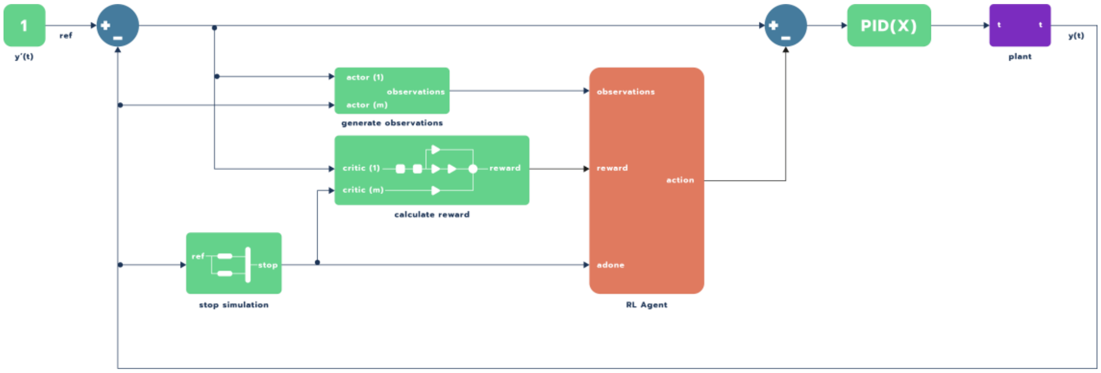

Figure 5: Hybrid control diagram based on Reinforcement Learning

4 Future Research

This prediction is obtained by 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 analyzed is part of the title or abstract, then 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.

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

In the same time, network visualization of PID indicate the PID controller confirms strength relationship in the research topic of reinforcement learning and robot. On the other hand, it shows a weakness 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 system. 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 method of asynchronous multi-thread training reduces the correlation of the training data, it proffers the more stable and adaptable controller and also 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