Re-exploration of pretrained artificial intelligence model for a nuclear power plant autonomous operation

Jae Min Kim, Hong Jun Yang and Seung Jun Lee*
Ulsan National Institute of Science and Technology, 50, UNIST-gil, Ulsan, 44911
** Corresponding author: sjlee420@unist.ac.kr*

1. Introduction

An autonomous system is a technology that is in the spotlight in industrial fields because of an advantage of excluding human error. Nuclear industry has also implemented automated systems that support nuclear power plants (NPPs) operation in full power mode. However, tasks of adjusting key parameters to suit the situation as the plant temperature and pressure gradually decrease or increase are performed by human operators. Thus, it is difficult to stabilize an NPP only with automatic systems during startup or shutdown operation.

In the previous work, the framework to develop autonomous system for startup and shutdown operation was suggested [1]. As an extension of the previous study, actor-critic algorithm was implemented to stabilize a pressurizer (PZR) pressure, one of tasks during startup operation. Training process consists of two stages: first, training is performed from the beginning, then the model is saved, and second, the saved model is started again as a starting point. Through this process, a PZR pressure control model was successfully obtained for a system operating block.

2. Nuclear power plant environment

Reinforcement learning (RL) is a field of machine learning, in which agents learn better behavior by receiving feedback on their behavior from an environment [2].

Markovian decision process (MDP) is a mathematical model for solving RL problems. In MDP problem, an agent learns optimal policy to achieve a goal through interaction with an environment.

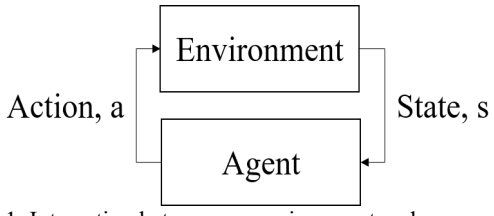

Fig. 1. Interaction between an environment and an agent.

It is natural that giving as much information as possible is advantageous for training an optimal policy. However, how much information an agent receives from the environment and how many actions an agent can choose has a great influence on the training development. This is because the computational process increases exponentially as the number of input nodes increases due to the nature of a neural network structure. In addition, even a small number of variables having continuous values create exponentially many combinations of states. For this reason, the efficiency of RL varies depending on which variables are input as a state.

In case of NPP operation, plant parameters are mainly composed of physical properties such as temperature, flow rate, and pressure. These parameters do not provide information on the relationship between each variable. In other words, an agent must empirically learn physical phenomena such as pressure and volume change due to temperature change. Purpose of startup operation is to gradually increase temperature and pressure of reactor coolant system (RCS) for preparing nuclear reaction. Therefore, RL models required for startup and shutdown operations should always include RCS pressure and temperature. Trend value is added to an input state to enable an agent to detect laten changes in the environment for its action early.

3. Methodology

Actor-critic algorithm was implemented to build an autonomous operation module for controlling PZR pressure. This algorithm consists of an actor and a critic to develop policy and value of actions [3]. An action is selected according to a policy by an actor and a critic evaluate the action when the policy updating. As the number of input states increases, there are more paths that can be taken within an episode. Therefore, it is important to induce an agent to experience as much as possible.

At each step, actions are selected probabilistically based on the policy function, which is called Boltzmann approach. In addition, noise that decays as the episode progresses is added to induce random selection in the beginning and action selection that follows the policy function in the later episode. In this way, an agent is encouraged to explore as many paths as possible.

As the number of episodes increases, the path that the agent experiences within one episode tends to become fixed. To compensate for this, the following training process is taken. Training process consists of two stages to provide the agent with more paths. First, the model that is obtained after an agent performs 500 episodes is stored. Then, the noise of the stored model is initialized so that the policy established performs exploration again.

4. Experiment setting

A test was conducted using compact nuclear simulator (CNS) for the application. The CNS models three loops Westinghouse PWR, 993MWe, developed by Korea Atomic Energy Research Institute (KAERI) [4].

Input states consists of RCS temperature, PZR pressure, status of two control valves, and trend value. Trend value is obtained change of PZR pressure from the previous value to the current one.

Action sets are consisted of two control valves: FV122, charging flow control valve, and HV142, letdown flow control valve. Each valve is controlled with a signal activating a toggle, not directly changing the parameters. For example, if FV122 should open, the agent selects a command to adjust the valve by a certain amount.

There are a total 9 actions with change of FV122 and HV142. FV122 and HV142 are control valves with a change of 1.5% once. In addition, there is commands for 'no change' and combined control of FV122 and HV142 simultaneously.

As shown in equation (1), reward is defined as a function to induce successful training rather than discrete values like zero and one:

$$R = \begin{cases} 0.5 * (BC_u - BC_d) - (abs(P_t - P_c)) \\ (-1) * (step_{max} - step_c) \ (if\ done) \end{cases} \quad (1)$$

During training, the closer PZR pressure is to the target, the greater the reward. However, if the training is terminated early, the penalty is as great as the remaining steps.

5. Result and Discussion

Figure 2 shows records of PZR pressure through 1911 to 1920 episodes of training. The agent achieved optimal policy to maintain PZR pressure in a target area around 25 kg/cm^2 to maximize the rewards. Figure 3 shows the change of PZR pressure and valves in the last episode.

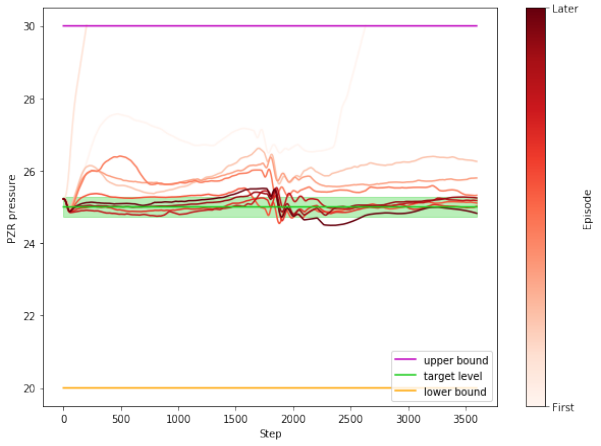

Fig. 2. The change of PZR pressure for last 10 episodes of training.

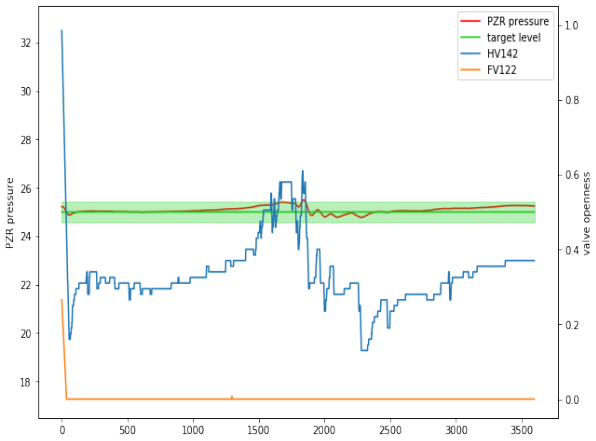

Fig. 3. The change of PZR pressure and valves according to the timestep in the last episode.

Figure 4 shows records of FV122 and HV142 valves of the last 10 episodes. The operation method of closing FV122 as much as possible and adjusting the pressure using HV142 is a result similar to the operation method performed by an NPP operator in general.

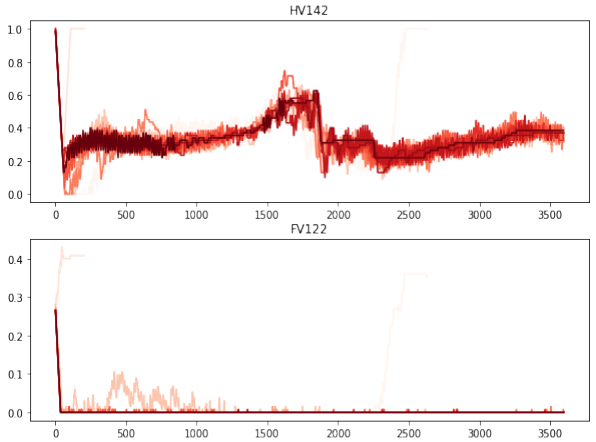

Fig. 4. The change of FV122 and HV142 for the last 10 episodes.

The achieved model will be applied to the system operation module. Future work aims to implement a system that autonomously operates a part of the startup operation by integrating the trained models.

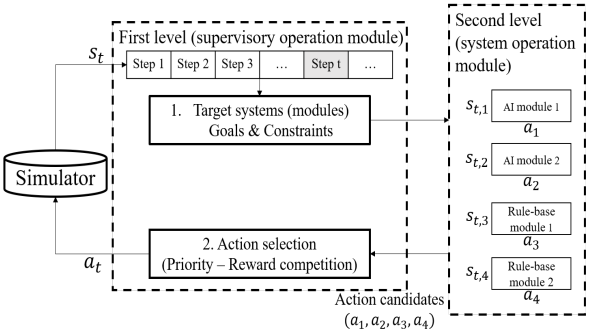

Fig. 4. Framework of autonomous operation for an NPP.

6. Conclusion

This paper deals with a study to induce an agent to quickly obtain an optimal operation policy for controlling PZR pressure as part of a study for the development of an NPP autonomous operation model. The trend value was added in consideration of the characteristics of an NPP environment. In addition, the model training was divided into the following two stages. After 500 trainings the stored model was initialized as a starting point, and then training was performed until an optimal policy was obtained.

Through this study, it was possible to contribute to the research on development of an NPP autonomous operation model by improving the training method considering an NPP environment. In the future work, it is necessary to develop a model that integrates the trained models.

7. Acknowledgement

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (No. 20171510102040).

REFERENCES

[1] J. M. Kim and S. J. Lee, Framework of Two-level Operation Module for Autonomous System of Nuclear Power Plants during Startup and Shutdown Operation, Transactions of the Korean Nuclear Society Autumn Meeting, 2019.
[2] R.S. Sutton, and A.G. Barto, Reinforcement Learning: an introduction, 2nd, The MIT Press, 2018.
[3] A. G. Barto, R. S. Sutton and C. W. Anderson, Neuronlike adaptive elements that can solve difficult learning control problems, in IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-13, no. 5, pp. 834-846, 1983.
[4] KAERI, Advanced Compact Nuclear Simulator Textbook, Nuclear Training Center in Korea Atomic Energy Research Institute, 1990.