Development of Anomaly Recovery Algorithm Based on Concept of Robust AI

Hee-Jae Lee[1], Daeil Lee[2], Jonghyun Kim[1]*
[1]*Chosun University, 309 Pilmun-daero, Dong-gu, Gwangju, 61452, Republic of Korea*
[2]*Korea Atomic Energy Research Institute, 111, Daedeok-daero 989 Beon-gil, Yuseong-gu, Daejeon, 34057, Republic of Korea*
[*]*Corresponding author: jonghyun.kim@chosun.ac.kr**

1. Introduction

When undesired transient or unexpected situation happens, nuclear power plant (NPP) operators should take appropriate action to maintain the integrity of reactor core and containment building. However, transient situation may increase operators' workload due to many parameters to be monitored, and multiple alarms. This can lead to human errors under the abnormal condition, which can further make the situation worse [1-4].

To reduce operator's workload and errors, current NPPs commonly implement automated controllers such as proportional-integral-derivative (PID), programmable logic (PI) and field-programmable gate arrays [5-9]. On the other hand, as artificial intelligence (AI) has been emerged, deep reinforcement learning (DRL) has attracted interest for controller design.

DRL-based approaches for controller have several strengths. First, DRL agents can learn from experience, whereas traditional controller relies on a model of the system to make decisions [13]. Secondly, DRL does not require manual tuning of parameters and human intervention [14]. Thirdly, DRL agents can explore different action, which can lead to better long-term performance [15]. On the other hand, traditional controllers typically rely on predefined rules such as if-then logic.

With this regard, there have been previous attempts to apply the DRL algorithm for NPP operations [16-19]. The previous attempts have trained their DRL algorithm under simulator environments so that intelligent agents interact with an environment. However, since there is a gap between a simulator and actual plant, it is difficult to guarantee whether the trained DRL model will work properly in the real NPP.

Addressing this challenge, the authors have suggested the concept of Robust AI that can adapt to a new environment where the AI model has not encountered [20, 21]. The Robust AI utilizes meta-data that can describe upper-level data on the parameter trend. By using trend image as an input, the AI network recognizes patterns on NPP states. The feasibility of this concept has been demonstrated when the working environment becomes changed from the training environment [20, 21].

This study extends the concept of Robust AI into designing a DRL-based control algorithm in abnormal conditions. The DRL algorithm uses trend image as an input and extracts meta-data on abnormal situations. The aim of this algorithm is to automatically take mitigation action under the abnormal condition, specifically, in the chemical control volume system (CVCS). To do this, the algorithm employs soft-actor-critic (SAC) agent method that can find the policy to explore more widely while giving up on clearly unpromising avenues.

2. Concept of Robust AI

The concept of a robust AI starts with the meta-data that describes other data at the upper-level. The basic idea of the Robust AI behind is that although the value of data and scale are different, the trend of values is similar between the different environments. Fig.1 shows how the Pressurizer pressure changes in two different simulators after the heat exchanger pipe break event. Exact values at the data level are different, but the trends are very similar between the different simulators. The pressure increases after the initiation of event, and then starts to decrease after the manipulation, i.e., open pressurizer spray. Considering that meta-data is data that can describe other data, the Robust AI uses parameter trends as an input format [20, 21].

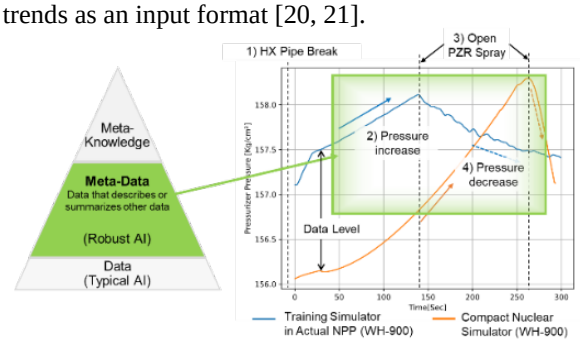

Fig. 1. Concept of robust AI.

The Robust AI uses trend image that can imply increase and decrease of plant parameters. Fig. 2 (left) shows an example of the graph for the PZR level variation in 120 seconds. The graph is then segmented into four sections, based on the state-state value as illustrated in Fig. 2 (right).

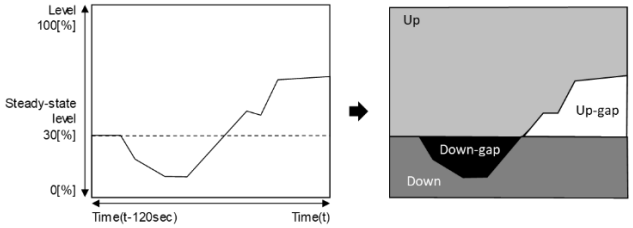

Fig. 2. Conversion process of a graph to trend image.

This study applies the concept of the Robust AI in developing DRL-based anomaly recovery algorithm in the abnormal situation. First, the algorithm will extract meta-data from trend images through the feature extractor consisting of convolutional layers. The extracted meta-data will provide information on symptoms of plant parameters. Secondly, DRL agents will be trained to take actions to recover and mitigate NPP status by controlling components identified by a work domain analysis.

3. Work Domain Analysis

To develop an anomaly recovery algorithm, this study targets the abnormal operation in CVCS. For a better understanding on key aspects of CVCS systems, a work domain analysis was performed using abstraction decomposition space (ADS).

ADS is a framework that is systematically used to analyze complex systems in the Cognitive Systems Engineering [22]. The result of the ADS analysis on the CVCS system is illustrated in Fig. 3. Through this analysis, a functional purpose of CVCS system is defined as maintaining the inventory of reactor coolant system. To meet the purpose, the physical functions available to perform recovery action were defined and then further the lowest level of abstraction such as valves, pumps, indicator, flow path, etc., is identified.

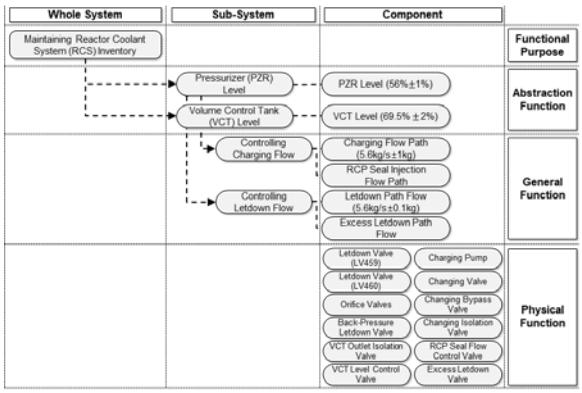

Fig. 3. ADS analysis on CVCS systems.

4. Development of Anomaly Recovery Algorithm for Abnormal Operation

This study developed an autonomous agent for the abnormal operation of the CVCS. This agent employs Soft Actor-Critic (SAC), one of DRL methods, as illustrated in Fig. 4. The SAC agent has the ability to learn effectively optimal policies in continuous action spaces. SAC agent involves Q-network and policy network. The SAC agent designed in this algorithm will take trend images on the derived physical parameters such as PZR level, VCT level, charging flow, RCP seal injection flow, letdown flow and excess letdown flow. Concurrently, the values of these six parameters will be also input to Q-network and policy network.

The Q-network in the SAC agent will first extract the meta-data from trend images in the format of vectorized matrix. Based on the extracted matrix, the Q-network then estimate the expected sum of rewards that the agent can receive starting from a given state and following a given policy. The Q-network is trained by updating the Q-value for a given state-action pair based on the immediate reward and the Q-value of the next state. The Q-network then provides the critic estimate of the value of state-action pairs to the policy network.

The policy network estimates a mapping from states to actions that the agent should take to maximize the expected sum of rewards. The policy network takes the state from trend images and outputs a probability distribution over actions. The policy is sampled from the probability distribution and the agent takes the action corresponding to the sampled policy.

While training, the agent alternates between updating the Q-network and the policy network. The Q-network is updated to minimize the mean squared error between the estimated Q-value and the target Q-value. The policy network is updated to maximize the expected reward.

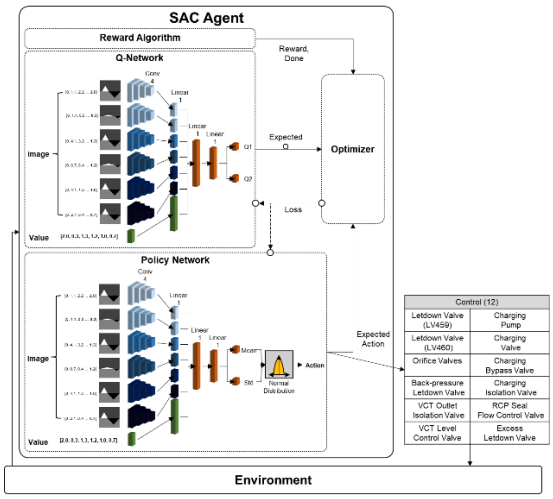

Fig. 4. ADS analysis on CVCS systems.

In DRLs, the reward is an essential element as it provides the reinforcement signal that guides the learning process. The agent in DRL takes actions and receives feedback in the form of a reward or penalty. The goal of the agent is to learn a policy that can maximize the accumulative reward. Therefore, the reward algorithm must be carefully designed to give the agent guidelines that lead to the desired outcomes.

Based on the ADS analysis (see Section 3), the reward algorithm was designed to achieve the goals of each level: 1) maintaining RCS inventory at the functional purpose, 2) satisfying success criteria of PZR level, and VCT level at the abstraction function, and 3) supplying charging line and letdown line at the general function. Table II explains how the reward is calculated for each level. Each reward for each goal has a range between 0 and 1.

Table I. Reward algorithm for each goal of abstraction level.

Abstraction Level	Goal		Reward (R) Calculation
Functional Purpose	RCS Inventory		$R(0\sim1)=\begin{cases}1, & if\ D=0\\ 1-D, & if\ D\neq 0\end{cases}$, where $D=\left\lvert\frac{\lvert PZRL_n-VCTL_n\rvert}{2}\right\rvert-\left\lvert\frac{\lvert PZRL_c-VCTL_c\rvert}{2}\right\rvert$
Abstraction Function	PZR Level (PZRL)		$R(0\sim1)=\begin{cases}1, & if\ P=0\\ 1-P, & if\ P\neq 0\end{cases}$, where $P=\lvert PZRL_n-PZRL_c\rvert$
	VCT Level (VCTL)		$R(0\sim1)=\begin{cases}1, & if\ V=0\\ 1-V, & if\ V\neq 0\end{cases}$, where $V=\lvert VCTL_n-VCTL_c\rvert$
General Function	Charging Line	Charging Flow (CGF)	$R\ (0\sim0.5)=\begin{cases}0.5, & if\ 5.6\frac{kg}{s}\leq CGF\leq 5.7\frac{kg}{s}\\ 0, & if\ CGF>5.7\frac{kg}{s}\ or\ CGF<5.6\frac{kg}{s}\end{cases}$
		RCP Seal Flow (RCPF)	$R\ (0\sim0.5)=\begin{cases}0.5, & if\ RCPF>0\\ 0, & if\ RCPF=0\end{cases}$
	Letdown Line	Letdown Flow (LDF)	$R\ (0\sim0.5)=\begin{cases}0.5, & if\ 5.6\frac{kg}{s}\leq LDF\leq 5.7\frac{kg}{s}\\ 0, & if\ LDF>5.7\frac{kg}{s}\ or\ LDF<5.6\frac{kg}{s}\end{cases}$
		Excess Letdown Path (EXLD)	$R\ (0\sim0.5)=\begin{cases}0.5, & if\ EXLD>0\\ 0, & if\ EXLD=0\end{cases}$

5. Training and Experiment

For a real-time testbed to train and validate the proposed anomaly recovery algorithm, the Compact Nuclear Simulator (CNS) was used. The CNS was developed by the Korea Atomic Energy Research Institute (KAERI), referring to a Westinghouse 900 MWe, three-loop PWR. Fig. 5. shows the display of CVCS system in CNS simulator.

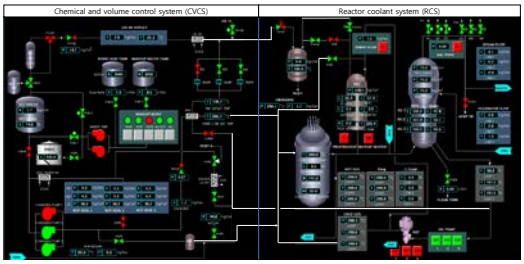

Fig. 5. Interface of CVCS system in CNS.

Training of the SAC agent was performed for more than 1,200 episodes. The obtained reward as the episode is illustrated in Fig. 6. In one episode, the theoretical maximum reward is 1,000. The feasible reward that the agent can practically achieve was observed as 900.

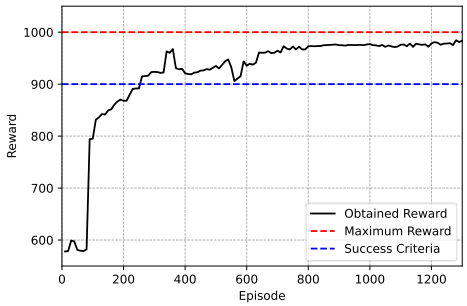

Fig. 6. Obtained reward by the SAC agent per episode.

After training, a test was conducted to confirm if the proposed algorithm can automatically take appropriate mitigation actions in case of the leakage in CVCS. Fig. 7 (upper) presents the control action suggested by the agent at the level of physical function. The performance followed by the control action is illustrated in Fig. 7 (below). When the suggested control is not considered, the PZR level and VCT level cannot meet the proper boundary as appeared on the dotted line. On the other hand, the PZR level and VCT level can be maintained with the action by the agent as seen in solid lines. It demonstrates that the proposed algorithm can effectively manage the PZR level and VCT level.

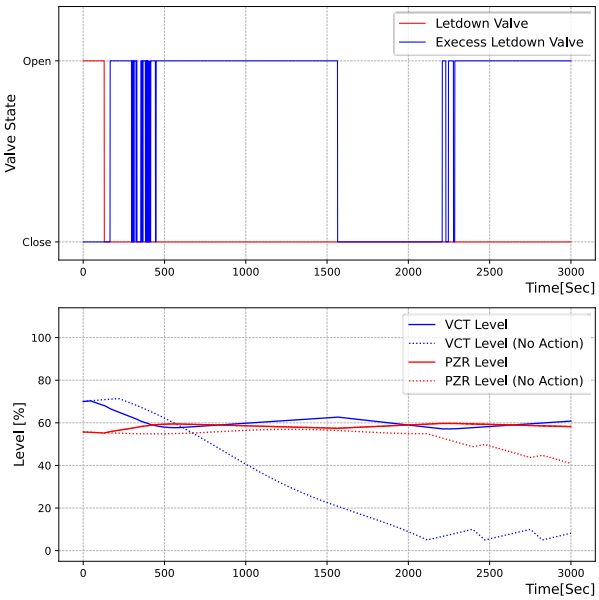

Fig. 7. The control action by the agent at the level of physical function (upper) and the followed performance (below).

5. Conclusion

This study applied the concept of the Robust AI to design anomaly recovery algorithm for the abnormal operation in the CVCS. To do this, the algorithm employs a SAC agent which is an effective approach for controlling key components in continuous action space.

The performance of the proposed algorithm was tested when leakage in CVCS is occurred. The algorithm trained in the CNS environment could make appropriate recovery action using meta-data in the CNS environment. To confirm how well this algorithm can generalize and adapt to a new environment, the proposed algorithm still needs to be tested in a new environment. As a future step, this study plans to investigate different environments that the proposed algorithm can interact with in real-time.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2022-00144042).

REFERENCES

[1] Y. Hirotsu, et al., Multivariate analysis of human error incidents occurring at nuclear power plants: several occurrence patterns of observed human errors, Cognition, Technology &, Vol. 3, pp.82-91 2001.
[2] D.N. Hogg, et al., Development of a situation awareness measure to evaluate advanced alarm systems in nuclear power plant control rooms, Ergonomics, Vol. 38, pp.2394-2413, 1995.
[3] M.-H. Hsieh, et al., A decision support system for identifying abnormal operating procedures in a nuclear power plant, Nuclear engineering and design, Vol. 249, pp.41-418 2012.
[4] J. Noyes and M. Bransby, People in control human factors in control room design, pp 344, 2001.
[5] R.T. Wood, et al., Autonomous control capabilities for space reactor power systems, Proceedings of American Institute of Physics conference, 2004.
[6] M. Zarei, et al., Robust PID control of power in lead cooled fast reactors: A direct synthesis framework, Annals of Nuclear Energy, Vol. 102, pp. 200-209, 2017.
[7] S. Khatua and V. Mukherjee, Application of PLC based smart microgrid controller for sequential load restoration during station blackout of nuclear power plants, Annals of Nuclear Energy, Vol. 151, pp. 107899, 2021.
[8] J. Yoo, et al., PLC-Based safety critical software development for nuclear power plants. in Computer Safety, Proceedings of 23rd International Conference SAFECOMP , 2004.
[9] J. She and J. Jiang, On the speed of response of an FPGA-based shutdown system in CANDU nuclear power plants, Nuclear Engineering and Design, Vol. 241, pp. 2280-2287, 2011.
[10] K. Arulkumaran, et al., Deep reinforcement learning, IEEE Signal Processing Magazine, Vol. 34, pp 26-38, 2017.
[11] K. Arulkumaran, et al., A brief survey of deep reinforcement learning, IEEE Signal Processing Magazine, Vol. 34, pp. 26-38, 2017.
[12] Q. Huang, et al., Adaptive power system emergency control using deep reinforcement learning, IEEE Transactions on Smart Grid, 11, pp. 1171-1182, 2019.
[13] A. T. Azr, et al., Drone deep reinforcement learning: A review, Electronics, Vol. 10, pp. 999, 2021.
[14] D. Lee, et al., Comparison of deep reinforcement learning and PID controllers for automatic cold shutdown operation, Energies Vol. 15, pp. 2834, 2022.
[15] Z. Dong, et al., Multilayer perception-based reinforcement learning supervisory control of energy systems with application to a nuclear steam supply system, Applied Energy, Vol. 259, pp. 114193, 2020.
[17] D. Lee, et al., Algorithm for autonomous power-increase operation using deep reinforcement learning and a rule-based system, IEEE, Vol. 8, pp. 196727-196746, 2020.
[18] D. Lee and J. Kim, Autonomous emergency operation of nuclear power plant using deep reinforcement learning. in advances in artificial intelligence, Software and Systems Engineering, Proceedings of the AHPE 2021, USA, 2021.
[19] J. Park, et al., Providing support to operators for monitoring safety functions using reinforcement learning, Progress in Nuclear Energy, Vol. 118, pp. 103123, 2020.
[20] H.-J. Lee, et al., Anomaly diagnosis of nuclear power plants using robust AI, Proceedings of American Nuclear Society, Pheonix, AZ & USA, 2022.
[21] D. Lee, et al., Concept of robust AI with meta-learning for accident diagnosis, Proceedings of Korean Nuclear Society, Jeju & Republic of Korea, 2022.
[22] N. Naikar, Cognitive work analysis: an influential legacy extending beyond human factors and engineering, Applied Ergonomics, Vol. 59, 2016.