Applying deep learning algorithms for autonomous emergency operation in nuclear power plant

YooJoon Seoung, Junyong Bae, Seung Jun Lee

Department of Nuclear Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea, yjs0427s@unist.ac.kr

EXTENDED ABSTRACT

Nuclear power plants have increasingly incorporated autonomous systems to improve safety and efficiency.[1] However, critical operations still rely heavily on human operators, especially during emergencies. In emergency scenarios, for example, operators must promptly bring the reactor coolant system to a stable state. Given the complexity and time-sensitive nature of such situations, the likelihood of human error remains significant. One example is an aggressive cooldown process. Aggressive cooldown is a post-trip emergency operating procedure intended to rapidly reduce the temperature and pressure of the reactor coolant system (RCS) following specific accident scenarios, such as a small break loss-of-coolant accident (SBLOCA) combined with the failure of Safety Injection Systems (SIS).

The primary objective of this strategy is to lower the RCS pressure sufficiently to enable the operation of low-pressure injection systems such as the Shutdown Cooling System (SCS), while simultaneously maintaining effective core cooling to ensure that the fuel cladding temperatures remain below safety limits.

This operation is typically executed by opening the Atmospheric Dump Valves (ADV) on the secondary side to discharge steam and facilitate rapid depressurization, while supplying feedwater through the Auxiliary Feedwater (AFW) system to sustain heat transfer in the steam generators. Simultaneously, Steam Generator cooldown through ADV opening drives a gradual reduction in RCS pressure and temperature, eventually allowing automatic or manual injections from accumulators such as the Safety Injection Tanks (SIT).

During aggressive cooldown, strict attention must be paid to maintaining the cooldown rate below 55.6 °C/hr, to avoid thermal stress or damage to critical components such as the reactor pressure vessel, steam generators, and piping systems. Additionally, the coolant temperature must be lowered to at least 177 °C, which is the prerequisite for initiating SCS injection.[2] Achieving this balance between rapid cooldown and system integrity is essential to the success of the procedure. Table I shows the states of the pressure, volume, temperature of the coolant before and after aggressive cooldown operation.

TABLE I. Initial and final state of reactor coolant before and after aggressive cooldown operation

	Initial state	Final state
Pressure(P)	P_0	20kg/cm^2
Volume(V)	V_0	50%
Temperature(T)	T_0	177^{o}C

In this study, we propose a DRL-based framework to optimize emergency cooling strategies for aggressive cooldown operations. To this end, we introduce a soft actor-critic (SAC) algorithm for continuous control of components and hindsight experience replay (HER) for overcoming sparse positive feedback problems. It has already been applied in reactor power control. For instance, Bae et al. (2023) successfully automated the reactor heat-up process using a DRL-based approach that incorporates Soft Actor-Critic (SAC) and Hindsight Experience Replay (HER) by optimizing reactor coolant state control.[3]

SAC is an off-policy deep reinforcement learning algorithm that optimizes a stochastic policy by maximizing both expected return and policy entropy, promoting exploration and improving stability. It uses two Q-networks to mitigate overestimation bias and employs an entropy temperature parameter α or adaptive trade-offs between exploration and exploitation.[4] Its ability to output continuous actions and reuse past experiences makes it well-suited for controlling complex systems like nuclear power plants. HER enhances learning in sparse-reward environments by modifying past experiences with alternative goals. By replacing the original goal with an achieved state and recalculating the reward, HER transforms failed episodes into informative learning samples, significantly improving sample efficiency.[5]

The reward function was redesigned to guide the DRL agent toward achieving a safe and effective aggressive cooldown strategy by emphasizing precise control of the cooldown rate while enforcing strict operational constraints. The reward structure focuses on minimizing overshoot beyond the target cooldown rate and encourages the agent to reach the target state within a specified time limit.

The primary reward signal is derived from the deviation between the current cooldown rate and the target rate of 55.6 °C/hr.[6] If the agent maintains a cooldown rate less than or equal to the target, the reward is defined as the negative square of the error, which gently penalizes underperformance while encouraging proximity to the target. Conversely, if the cooldown rate exceeds the target, which poses a risk of structural damage to the reactor system, a significantly harsher penalty is applied using a large negative linear coefficient to discourage aggressive overcooling:

$$r_{cool} = \begin{cases} -(r_t - 55.6)^2 & if\ r_t \leq 55.6 \\ -1{,}000\ (r_t - 55.6) & if\ r_t > 55.6 \end{cases} \quad (1)$$

At the end of each episode, a terminal reward is assigned based on whether the agent successfully reduces the reactor coolant temperature to 177 °C or below within 6 hours. If the condition is met, a small positive bonus (+1.0) is given to reinforce successful operations. If the target is not achieved, a large terminal penalty (−50,000) is applied to strongly penalize failure:

$$r_{time} = \begin{cases} 1 & if\ T \leq 177^{o}\mathrm{C} \\ -50{,}000 & if\ T > 177^{o}\mathrm{C} \end{cases} \quad (2)$$

This reward formulation reinforces successful completion of the cooldown process within realistic operational constraints.

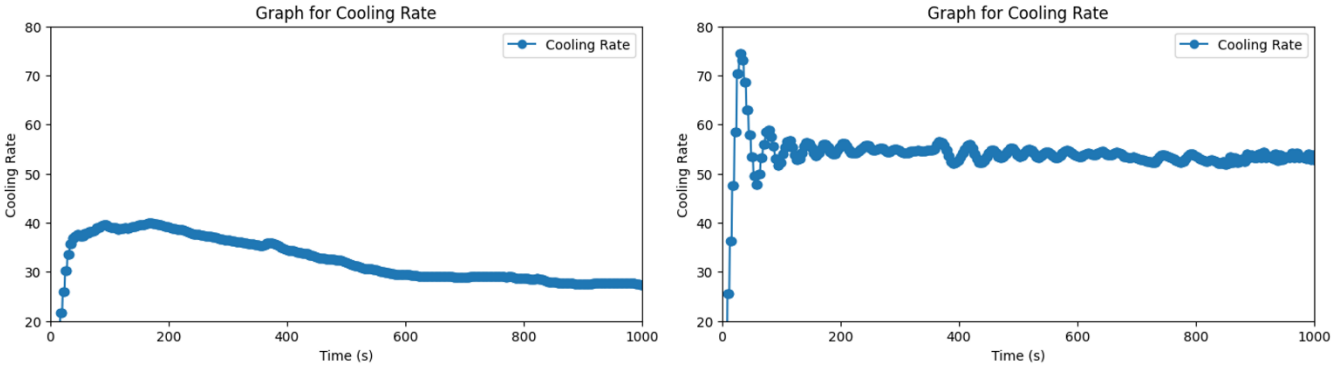

Fig. 1. Cooling rate difference over time (Left: trial 10, Right: trial 1000)

Figure 1 compares the cooling rate trajectories of the DRL agent at trial 10 (left) and trial 1000 (right). In the early stage of training (trial 10), the agent exhibits unstable control behavior, failing to maintain the target cooldown rate consistently. The cooling rate fluctuates and often falls below the desired level, indicating a lack of learned policy optimization.

By contrast, in trial 1000, the agent demonstrates significantly improved control performance, maintaining the cooling rate steadily around the target value of 55.6 °C/hr. This highlights the effectiveness of the reward function and training framework in guiding the agent to learn a stable and safe aggressive cooldown strategy.

This study aims to demonstrate that autonomous operation based on reinforcement learning with SAC and HER can be effectively applied not only under normal conditions but also in emergency scenarios within nuclear power plants. An experimental framework has been established to support model training and evaluation.

In future work, the scope of experiments will be expanded by gradually increasing both the number of control variables and target variables, allowing the agent to adapt to a wider range of operating conditions and emergency situations. The trained agent's performance will be assessed in various scenarios to evaluate its applicability in realistic reactor environments. Furthermore, the generalization capability of the model will be examined, and comparative studies with conventional control approaches will be conducted to validate the effectiveness of the proposed autonomous operation strategy.

ACKNOWLEDGMENTS

This work was supported by Korea Institute of Energy Technology Evaluation and Planning(KETEP) grant funded by the Korea government(MOTIE)(RS-2024-00403194, Next-Generation Nuclear Technology Creation IP-R&D Talent (Human Resources) Development Project)

This research was supported by the National Research Council of Science & Technology(NST) grant by the Korea government (MSIT) (No. GTL24031-400)

REFERENCES

[1] Brown, David. Artificial Intelligence for Accelerating Nuclear Applications, Science, and Technology. No. BNL-223196-2022-INRE. Brookhaven National Laboratory (BNL), Upton, NY (United States), 2022.
[2] Han, Seok Jung, Ho Gon Lim, and Joon Eon Yang. "Thermal hydraulic analysis of aggressive secondary cooldown in a small break loss of coolant accident with a total loss of high pressure safety injection." (2003).
[3] Bae, Junyong, Jae Min Kim, and Seung Jun Lee. "Deep reinforcement learning for a multi-objective operation in a nuclear power plant." *Nuclear Engineering and Technology* 55.9 (2023): 3277-3290.
[4] Haarnoja, Tuomas, et al. "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor." *International conference on machine learning*. PMLR, 2018.
[5] Andrychowicz, Marcin, et al. "Hindsight experience replay." *Advances in neural information processing systems* 30 (2017).
[6] Kim, Man Cheol, and Sang Hoon Han. "Variability of plant risk due to variable operator allowable time for aggressive cooldown initiation." *Nuclear Engineering and Technology* 51.5 (2019): 1307-1313.