4th Asia Pacific Conference of the Prognostics and Health Management,
Tokyo, Japan, September 11 – 14, 2023 P01-04

Application of Model-based Deep Reinforcement Learning Framework to Thermal Power Plant Operation Considering Performance Change

Yutaka Watanabe[1,2] and Takehisa Yairi[2]

[1]*Central Research Institute of Electric Power Industry, Yokosuka, Kanagawa, 240-0196, Japan*
yutaka@criepi.denken.or.jp

[2]*The University of Tokyo, Meguro-ku, Tokyo, 153-8904, Japan*
yutaka-w@g.ecc.u-tokyo.ac.jp, yairi@g.ecc.u-tokyo.ac.jp

Abstract

In recent years, there have been increasing expectations for the development of advanced plant operational support systems that can automate complex tasks and autonomously optimize operational procedures in thermal power plants. The performance of the equipment changes during operation and maintenance; hence, it is necessary to adjust the operating process to satisfy the operational constraints. In this study, we investigated a framework based on model-based deep reinforcement learning for acquiring control methods that are robust to changes in equipment performance using a digital twin model. A case study of the operational planning of a thermal power plant was presented and it was demonstrated that a stable control system can be constructed even when plant characteristics are changing.

1. Introduction

With the introduction of variable renewable energy into the power grid, thermal power plants will have more opportunities to start/stop and operate with varying power outputs. The internal state of a power plant changes even if the power output is the same owing to the external environment and degradation over time. Therefore, a control system adjusted based on the assumption of the plant state at a certain point in time, such as during design or after periodic repairs, may cause inefficient operation or unexpected problems owing to unexpected operating conditions resulting from changes in plant characteristics over time. Unplanned operation and shutdown due to readjustment of control parameters lead to increased costs; hence, a mechanism to adjust operation and control parameters autonomously according to the current status is necessary.

In recent years, studies have been conducted on optimizing operational methods using deep reinforcement learning (DRL), which is expected to apply to complex tasks that cannot be performed using conventional controllers. In particular, there is great interest in the possibility of achieving autonomous operations. In power plants, autonomous, efficient, and economical operations with no operator intervention are expected to lead to labor savings and advanced operation and maintenance.

Load control systems for thermal power plants are basically designed to automatically determine the amount of operation based on output command values and are complex systems that combine PID controllers, function generators, and other systems. In addition, thermal power plants are characterized by the fact that load bands, fuel properties, and changes in equipment performance over time can significantly change control characteristics, and each plant is currently undergoing a trial-and-error tuning process based on long-standing empirical rules, so there are high expectations for control systems that are capable of autonomous adaptation.

Several research of DRL has been applied to the construction of operational and control systems for power plants (Chen, et. al, 2018; Adam, et. al., 2021), and the results suggest that applying DRL to optimize driving operations is effective. However, few studies on learning methods have considered changes in the performance characteristics of environmental models. Using an environmental model that can represent changes in plant characteristics, it may be possible to obtain an operational control system that is robust against changes in equipment performance during plant operation. For example, in the research field on chemical plants, Kubozawa et al. (2022) proposed a reinforcement learning framework based on a task partitioning structure that takes into account changes in system parameters of environmental models as a method to bridge the gap between reality and simulation and showed that a robust operation control system can be obtained to reduce errors between actual plant operation and simulation in a suddenly changing environment.

In this study, we investigated the effect of the different schemes of considering equipment performance parameters on control performance for the task of acquiring a control strategy during load increase operations, as a functional framework for designing an operation and control system with such autonomy using model-based reinforcement learning.

2. Method

We considered three DRL schemes. Figure 1 illustrates the base scheme. Here, we assume a normal reinforcement learning framework in which the plant characteristics do not change. Figure 2 shows the proposed scheme 1. Here, the equipment performance parameter information is used only in the environment, and the equipment performance parameters are randomly varied during learning, but not used for agent information during the learning and testing. Figure 3 illustrates the proposed scheme 2. In this scheme, it is assumed that the equipment performance parameters of the actual plant can be obtained during testing, and that privileged information on the equipment performance parameters can be used during the learning and testing of the environment and behavioral measures. The method of acquiring equipment performance values during testing is beyond the scope of this report, but it can be achieved by using data assimilation methods that utilize dynamic physical models or machine learning models that output equipment performance values from measurement data.

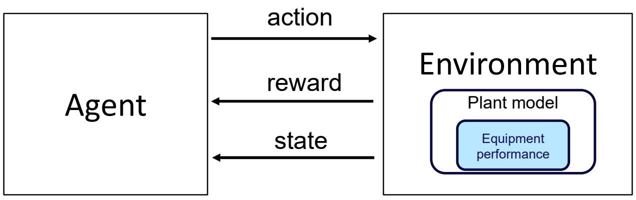


Figure 1. Base scheme

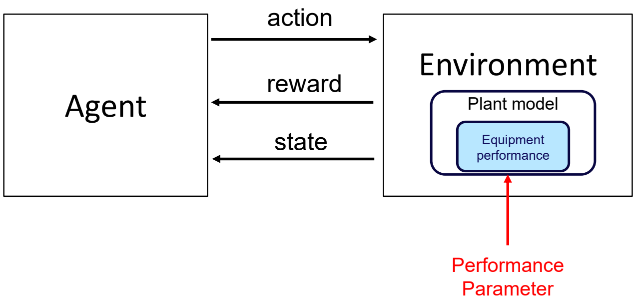


Figure 2. Proposed scheme 1

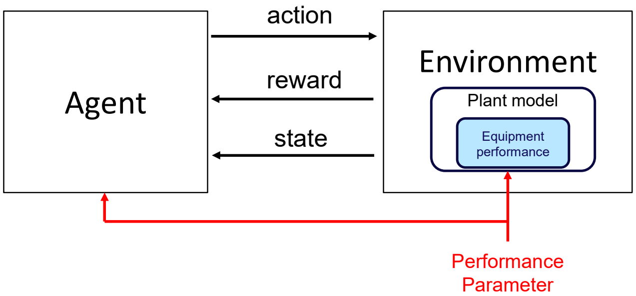


Figure 3. Proposed scheme 2

3. Case Study

The proposed method was applied to the problem of obtaining an increased power output operation for an oxygen–hydrogen combustion turbine system (Watanabe et al., 2022). We examined the challenge of finding the optimal output command profile to properly change the load when increasing the generation output from 20% to 100% while satisfying the upper-temperature limit constraint at the high temperature turbine (HTT) outlet during the load change operation.

3.1. Problem Setting

Environment: The target plant model of the system was established using a Modelica-based tool developed by the Central Research Institute of the Electric Power Industry (Takahashi et al., 2016). The plant model was constructed based on 1D-based physical modeling as shown in Figure 4, and the simulation can be performed considering parameters related to the performance of plant components. The input condition is the power output command, and the fuel flow rate was calculated from the difference between the power output command value and the calculated power output.

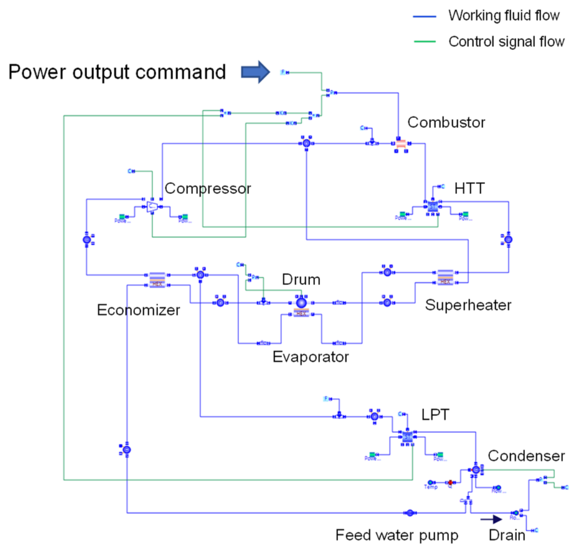


Figure 4. Dynamic model of the target system

Agent: A Deep Q Network was employed as the policy-learning algorithm. The action–value function was approximated using a deep neural network. Network inputs are states and actions, and the outputs are actions.

Action: The output change rate (in 1% increments in the range -1–10%/min for the next time interval) was set as the action for each fixed interval time (1 minute in this case) as shown in Fig. 5.

State: Three state variables were set: power output, HTT outlet temperature, and CP inlet temperature.

Performance parameter: Two equipment performance parameters were varied in this case: steam compressor and HTT. The change in equipment performance parameters is expressed by the following equation:

$$p = p_0 \cdot \alpha \quad (1)$$

where p, p_0, and α denote the performance parameter of the model, normal performance parameter of the model, and degree of degradation, which is 1.0 under normal conditions, respectively. During training, α was always fixed at 1.0 for the base scheme; for the proposed schemes 1 and 2, the value was set by randomly varying α between 0.98 and 1.0 for each episode.

Reward function: The reward function was set as follows, considering the HTT outlet temperature (upper limit: +20 °C at rated power output) as an operational constraint on the state variables:

$$R = 1 - R_1 - R_2 \quad (2)$$

For the R_1 term, a high reward was obtained by reaching the rated output as quickly as possible, and for the R_2 term, a penalty was imposed if the HTT outlet temperature exceeded the constrained temperature.

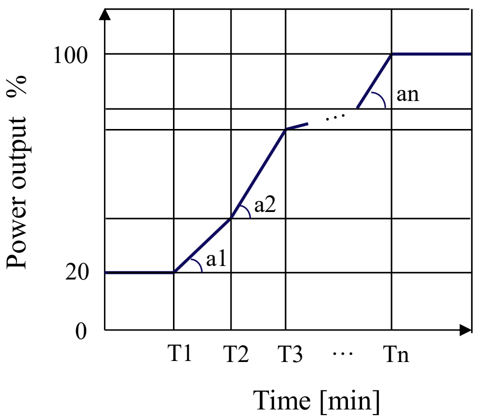
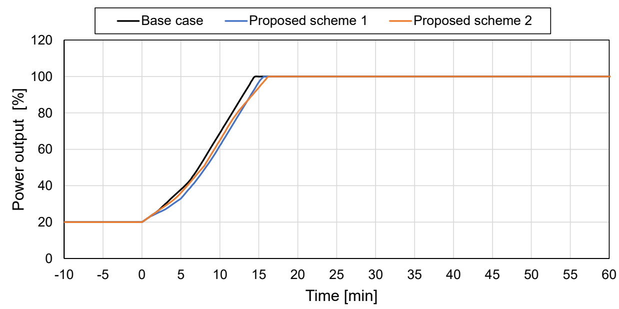


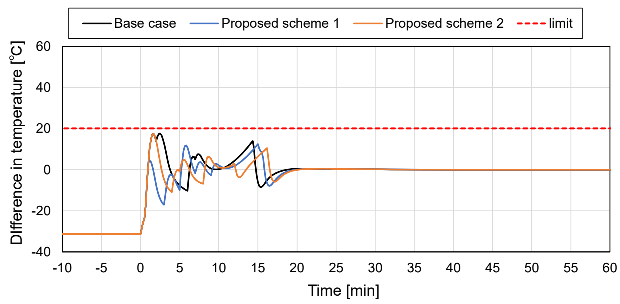
Figure 5. Operational parameter

3.2. Experimental results

As an example of the test results for the learned controller, Figure 6 shows the results for a plant model with α_{CP} = 1.0 and α_{HTT} = 1.0 for the degradation degree α of the equipment performance parameters and Figure 7 shows the results for a plant model with α_{CP} = 0.98 and α_{HTT} = 0.98 for the degradation degree α of the equipment performance parameters. As shown in Figure 6, all the methods were able to increase the load without exceeding the temperature limit. In addition, the base scheme produced the highest output change rate. However, in Fig. 7, it is observed that the operation with the base scheme and proposed scheme 1 exhibits unstable behavior around the rated output above the temperature limit, whereas proposed scheme 2 obtains stable load changes. The results suggest that by applying a reinforcement learning method that considers plant characteristics, as in the proposed scheme 2, we may be able to acquire control laws that are robust to changes in plant performance.

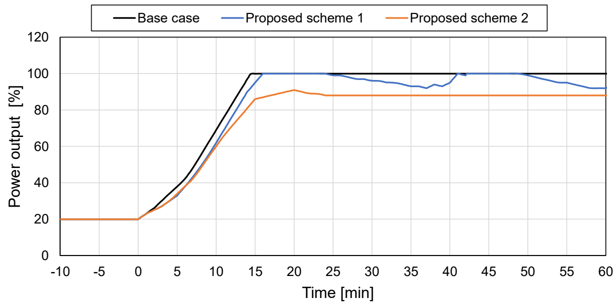


(a) Power output

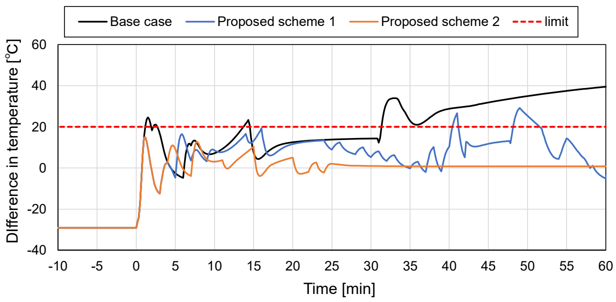


(b) HTT outlet temperature

Figure 6. Simulation result of no change in plant performance ($\alpha_{CP} = 1.0, \alpha_{HTT} = 1.0$)



(a) Power output



(b) HTT outlet temperature

Figure 7. Simulation result of changing in plant performance

$(\alpha_{CP} = 0.98, \alpha_{HTT} = 0.98)$

4. Conclusion

We proposed a reinforcement learning framework that considers the changes in plant performance. It was found that by including information on plant characteristics in the action strategy during learning and testing, as in the proposed scheme 2, a stable control system can be constructed even when plant characteristics are changing. This suggests the importance of estimating and utilizing parameters related to plant characteristics in addition to actual measurable operating data. When applying the proposed method to actual plants, combining a plant model that can accurately reproduce actual plant conditions and technology to accurately estimate the equipment performance parameters is important. In future work, we intend to study the impact of the agent's learning algorithm and structure on performance and the applicability of the proposed method to various cases that can be used in actual plants.

References

Adam, D., Oh, D., Kim, D., Lee, C., and Oh, M., (2021). *Deep reinforcement learning optimization framework for a power generation plant considering performance and environmental issues*, Journal of Cleaner Production, 291(1).

Cheng, Y., Huang, Y., Pang, B., and Zhang W., (2018). *ThermalNet: A deep reinforcement learning-based combustion optimization system for coal-fired boiler*, Engineering Application of Artificial Intelligence, 74.

Kubosawa, S., Onishi, T., Tsuruoka, Y., (2022). *Sim-to-real transfer in reinforcement learning-based, non-steady-state control for chemical plants*, SICE Journal of Control, Measurement, and System Integration, Vol.15, No1, pp.10-23. doi: 10.1080/18824889.2022.2029033

Takahashi, T., Nakamoto, M., and Watanabe, Y., (2016). *Construction of dynamic analysis tool for thermal power systems*, CRIEPI report, M15005.

Watanabe, Y., Takahashi, T., and Suzuki, K., (2022). *Dynamic Simulation of an oxygen-hydrogen combustion turbine system using Modelica*, Proceedings of Asian Modelica Conference 2022, Tokyo, Japan. doi: https://doi.org/10.3384/ecp19315

Yutaka Watanabe received M.E degreed from the University of Tokyo in Japan in 2009. He joined Central Research Institute of Electric Power Industry (CRIEPI) in 2009. He is mainly engaged in system performance analysis and dynamic simulation research on energy plant systems, such as existing thermal power plants, new thermal power systems and heat pump systems.

Takehisa Yairi is currently a professor at the Research Center for Advanced Science and Technology (RCAST) of the University of Tokyo. He received his B.Eng., M.Sc., and Ph.D. degrees in aerospace engineering from the University of Tokyo, Japan in 1994, 1996, and 1999 respectively. His research interests include anomaly detection, health monitoring, fault diagnosis, learning dynamical systems, nonlinear dimensionality reduction, as well as applications of machine learning and probabilistic inference to aerospace systems.