Reinforcement Learning Combined with Digital Twin Model for Chemical Process Control

Somayeh Mirzaei, Jia-Lin Kang\*, Zi Hang Yang

Department of Chemical and Materials Engineering, National Yunlin University of Science and Technology, Yunlin 64002, Taiwan ROC

jlkang@yuntech.edu.tw

Abstract

This study proposes reinforcement learning (RL) combined with a digital twin model to implement an environment for RL training and the stability of process control. The simulated process is established based on a sequence-to-sequence rolling model and an optimization algorithm to build a Model Predictive Control (MPC). The control stability of RL is analyzed using Selective Hydrogenation Unit (SHU) for the separation of C4 and C5 components in the column. Monte-Carlo deep deterministic policy gradient (MC-DDPG) is proposed as RL model and five types of reward functions are designed to reduce the energy consumption of process control. The value of integral absolute error (IAE) for C5 components are reduced by 100% using RL compared to MPC. For C4 components the IAE value using RL decreased by 67% and 80%, respectively indicating that the control effect of RL is better than that of MPC.

**Keywords**: Reinforcement learning; Digital twin model; StS Rolling prediction; Energy control

# 1. Introduction

Reinforcement learning (RL) is a subclass of machine learning which an agent learns by interacting with an unknown environment. The agent obtains feedback in terms of a reward from the environment, and it applies this feedback to train itself and collect experience and knowledge about the environment (Naeem, Rizvi et al. 2020, Panzer and Bender 2022). RL needs to interact with the real process for training, but due to the process safety it is not allowed to directly apply the actual process for interactive training. Kang et al. (Kang, Mirzaei et al. 2022) proposed a DDPG model with two-stage training to control the performance of boiler level control. The result showed that compared with DDPG model that directly contracts with online training, the two-stage training DDPG can effectively reduce the number of training. They also proved that control ability of three stages DDPG is better than that of 3E control.

Reinforcement learning can handle continuous or discrete control of single-input single-output systems and multiple-input multiple-output systems, and can perform well in the case of noise in the process control. However, there is not open literature to show that the simulation results are applied to the actual process for the reinforcement learning process control. Therefore, this study proposes RL combined with a digital twin model to implement an environment for RL training and the stability of process control. This study refers to Yoo et al. (Yoo, Kim et al. 2021) using the Monte-Carlo algorithm to replace the TD-error used in DDPG to update the model, this model is called MC-DDPG. Monte-Carlo deep deterministic policy gradient (MC-DDPG) is introduced as RL model and five types of reward functions are designed to reduce the energy consumption of process control.

# 2. Methodology

In this study, the real and virtual processes are implemented using Aspen Dynamics simulation and digital twin model, respectively. First, a virtual process is built by digital twin model. Then, the reinforcement learning is combined with virtual process for interactive training. The model is validated using virtual process and then the actual process can be controlled. This process is divided into training and testing parts. A sequence-to-sequence rolling model is used to establish a virtual process. The historical data is input to the encoder. Decoder is comprised the current and future operating and disturbance values. When the reinforcement learning training is completed, the virtual process is tested, analyzed the control results, and then the model is applied to the real process for control.

The architecture and training process of MC-DDPG are shown in Figure 1. The Monte Carlo algorithm is a periodic update, not a single-step update, and the actual reward can be used directly for model training without additional bootstrapping to obtain the real reward. Hence, MC-DDPG needs a set of Actor-Critic network, and Actor and Critic are each an independent ANN network.

A diagram of a process

Description automatically generated

**Figure 1.** MC-DDPG model used in this study

As shown in Figure 2, the historical data and the operation given by the optimization tool are fed into StS rolling model to predict the result. The model performs online error correction by calculating the error between the actual process data and the predicted model. Hence, the control system has the capability of self-adaptation.

A diagram of a machine

Description automatically generated

**Figure 2.** MPC used in this study

Reward functions and integral absolute error (IAE) are used as the basis to evaluate the model control. 5 types of reward functions (Eq. 3-7) are applied to calculate the energy consumption of the process.

Based on Eq. 2, C5loss and C5loss, sp indicate the upper product concentration and concentration limitation of output product in column, respectively. C4loss and C4loss, sp represent concentration of bottom product and the concentration limitation, respectively. In addition, this study tested each model for 10 rounds, each round simulated 300 time steps, and each time step was simulated for 10 minutes, so each round simulated a total of 3000 minutes.

|  |
| --- |
| (1) |
| (2) |
| (3) |
| (4) |
| (5) |
| (6) |
| (7) |

# 3. Case study

*3.1 Process description*

The industrial Selective Hydrogenation Unit (SHU) as a simulated plant was simulated using Aspen Dynamics® to illustrate the effectiveness of the proposed model. The flowsheet is shown in Figure 3. In this study, the quality control of main products including C5 and C4 at the top and bottom of column are investigated. As shown in figure, quality variables of C5 and C4 products are defined as qv1 and qv2, respectively.

A diagram of a complex diagram

Description automatically generated

**Figure 3.** Schematic illustration of SHU process.

# 4. Result and discussion

*4.1 comparison of MPC and RL*

Figure 4 indicates control performance of model predictive control and reinforcement learning models in the real process. The result compares the IAE analysis based on C5 and C4 products. The value of IAE for C5 product reduces by 100% using reinforcement learning compared to MPC. The IAE value for C4 product using RL decreases by 67% and 80% respectively, indicating that the control effect of reinforcement learning is better than that of MPC.

|  |  |
| --- | --- |
| A graph with a red square  Description automatically generated | A graph of a number of colored squares  Description automatically generated with medium confidence |
| **Figure 4.** IAE comparison for RL and MPC models  Figure 5 demonstrates IAE value of real case for C5 and C4 products using MPC and 5 types of reward functions considered in this study. It is found that the IAE value for C5 and C4 products using type 3 of reward function is higher than that of MPC models. Therefore, the energy consumption is not considered for type 3.   |  |  | | --- | --- | | A graph with different colored numbers  Description automatically generated with medium confidence | A graph of different colored bars  Description automatically generated |   **Figure 5.** IAE comparison using MPC and different reward functions  Table 1 shows energy consumption reduced using MPC and RL for 4 types of reward functions. It is obvious that type 5 of reward function has the best performance. | |

**Table1.** Energy consumption reduced by RL compared to MPC

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Type-1(%) | Type-2(%) | Type-4(%) | Type-5(%) |
| QV | 0.2641 | 1.0734 | 0.2058 | -0.3945 |
| QV&SV | 0.4422 | 0.4577 | -0.5109 | -0.4426 |

*4.2 Effect of noise on RL*

Figure 6 indicates performance of RL using 3% noise for actual process. IAE value for C5 and C4 products using ±3% noise is slightly higher than that of without noise. These results prove that reinforcement learning can stably control the process when ±3% noise is added to the process. However, the control ability of the model decreases because the information input to the model increases.

|  |  |
| --- | --- |
| A graph with a bar and a line  Description automatically generated with medium confidence | A graph with blue and orange bars  Description automatically generated |

**Figure 6.** Effect of noise on stability of RL

5. Conclusions

RL needs to interact with the real process for training, but due to the process safety it is not allowed to directly apply the actual process for interactive training. Therefore, this study proposes Monte-Carlo deep deterministic policy gradient (MC-DDPG) as RL model which is combined with a digital twin model to implement an environment for RL training and the stability of process control. Model Predictive Control (MPC) as virtual process is established based on a sequence-to-sequence rolling model and an optimization algorithm to train the reinforcement learning. StS rolling model is used due to the capability of long-term predictions. MC-DDPG model could perform better than MPC and type 5 of RL was the most stable to reduce the amount of energy consumption. Furthermore, 3% noise could not disturb the RL model stability.

References

Kang, J.-L., S. Mirzaei and J.-A. Zhou (2022). "Robust control and training risk reduction for boiler level control using two-stage training deep deterministic policy gradient." Journal of the Taiwan Institute of Chemical Engineers **130**: 103956.

Naeem, M., S. T. H. Rizvi and A. Coronato (2020). "A gentle introduction to reinforcement learning and its application in different fields." IEEE Access **8**: 209320-209344.

Panzer, M. and B. Bender (2022). "Deep reinforcement learning in production systems: a systematic literature review." International Journal of Production Research **60**(13): 4316-4341.

Yoo, H., B. Kim, J. W. Kim and J. H. Lee (2021). "Reinforcement learning based optimal control of batch processes using Monte-Carlo deep deterministic policy gradient with phase segmentation." Computers & Chemical Engineering **144**: 107133.