Continuous Control of Complex Chemical Reaction Network with Reinforcement Learning

Khalid Alhazmi[1*], S. Mani Sarathy[1]
[1]*King Abdullah University of Science and Technology (KAUST)*
*Contact: khalid.alhazmi@kaust.edu.sa, phone +966-540 000 8254

Abstract— **The goal of process control is to maintain a process at the desired operating conditions. Disturbances, measurement uncertainties, and high-order dynamics in complex and highly integrated chemical processes pose a challenging control problem. Even though advanced process controllers, such as Model Predictive Control (MPC), have been successfully implemented to solve hard control problems, they are difficult to develop, rely on a process model, and require high performance computers and continuous maintenance. Reinforcement learning presents an appealing option for such complex systems, but little work has been done to apply reinforcement learning in chemical reactions with practical significance, to discuss the structure of the RL agent, and to evaluate the performance against benchmark measures. This work (1) applies a state-of-the-art reinforcement learning algorithm (DDPG) to a network of reactions with challenging dynamics and practical significance. (2) Disturbances and measurement uncertainties have been simulated. In addition, (3) we defined an observation space that is based on the working concept of a PID controller, optimized the reward function to achieve the desired controller performance, and evaluated the performance of the RL controller in terms of setpoint tracking, disturbance rejection, and robustness to parameter uncertainties.**

I. Introduction

The goal of process control is to maintain a process at the desired operating conditions to ensure safety and profitability. Disturbances, measurement uncertainties, and high-order dynamics in complex and highly integrated chemical processes pose a challenging control problem. The major issue with PID controllers, which are the workhorse controller in chemical plants, is that they are reactive (action is taken after error is detected) due to the lack of knowledge about the process. Hence, the process industry has invented Model Predictive Control (MPC).

Model Predictive Controllers utilize a dynamic model of the process in order to anticipate future system behaviour and act accordingly. MPC's find an optimal sequence of actions by solving a constraint optimization problem (COP), which requires an objective function, at each time step. This highlights two challenges with MPC: one is the high computational cost that results from solving the optimization problem at each time step, and the other is that the performance of MPC relies on the accuracy of the dynamic model. Hence, alternative solutions to MPC are needed.

A promising alternative to MPC is model-free Reinforcement Learning (RL). The goal of RL is to find the sequence of actions that will generate the optimal outcome as defined by a reward function. Hoskins and Himmelblau were the first to propose the application of reinforcement learning to chemical engineering problems [1]. They used an artificial neural network (ANN) as a function approximator and reinforcement learning to adjust the weights of the ANN. The algorithm they employed is an actor/critic algorithm applied to a Continuous Stirred Tank Reactor (CSTR) with a very simple reaction: A→B. Even though the system is nonlinear, the dynamics are still much simpler that what is found in practice. In addition, much progress has been made since 1992 in developing more efficient reinforcement learning algorithms.

Other studies that implemented reinforcement learning on a CSTR have been published, but only applied to simple A →B reactions. A highly cited article that was published in late 2019 in the flagship chemical engineering journal, Computers & Chemical Engineering [2] confirms that little progress has been made since Hoskins and Himmelblau proposed using RL as process controllers. When discussing an example of RL in continuous control, the authors simulated a bipedal robot from the OpenAI Gym.

Much work must be done in applying reinforcement learning in chemical engineering problems. In addition to demonstrating the performance of RL algorithms in problems with complex dynamics, there is a need to study the structure of the observation space and reward engineering for problems relevant to chemical engineers. Moreover, the performance of RL-based controllers has to be evaluated according to benchmark performance measures.

II. Methods

A. Algorithms, Observations, Reward

The reinforcement learning algorithm that is used in this work is called Deep Deterministic Policy Gradient (DDPG) [3]. The algorithm aims to find an optimal policy that maximizes the long-term reward. The reasons for selecting this algorithm include its speed and its continuous observation and action space. Since DDPG is online and model-free, a policy can be learned without requiring a static dataset or a process model.

Table I shows the variables that are included in the observation space of the RL agent and the purpose of including each variable. The observation space that includes the differential error is compared with the that that does not in the Results section.

TABLE I
OBSERVATION SPACE

	Quantity	Purpose
e	Proportional error	Accounts for present error value
$\int e\,dt$	Integral error	Accounts for cumulative error
$\frac{d}{dt}e$	Differential error	Anticipates future error
T_r	Reactor temperature	To perceive the current state

The reward function is chosen as follows:

$$r_t = 10(|e| < 0.05) - 1(|e_t| \geq |e_{t-1}|)$$

Where e is the error between the process variable and the setpoint. Ten points are rewarded if the error is less than 5%, and one point is subtracted if the error increases. This reward function has the advantage of being simple, so that it's applicable to different environments.

B. *Reactor and Reactions*

The environment to which reinforcement learning is applied to in this work is a complex reaction network in Continuous Stirred Tank Reactor (CSTR). Table II shows the mathematical model of the reactor, where *c*:concentration, T: temperature, V: volume, ρ: density, q: volumetric flow rate, U: heat transfer coefficient, Qc: cooling rate, A: surface area, k: rate coefficient, E: activation energy, h: enthalpy, k0: pre−exponential coefficient.

This system was chosen because it is 4th order nonlinear dynamical system, has unstable zero dynamics, cannot be controlled by linear controllers, resembles real world problems, and is commonly used as a benchmark in the process control community. The system is described in detail in [5].

$$\underset{Cyclopentadiene}{C_5H_6\ (A)} \xrightarrow{k_1} \underset{Cyclopentanol}{C_5H_7OH\ (B)} \xrightarrow{k_2} \underset{Cyclopentanediol}{C_5H_8(OH)_2\ (C)}$$

$$\underset{Cyclopentadiene}{2\ C_5H_6\ (A)} \xrightarrow{k_3} \underset{Dicyclopentadiene}{C_{10}H_{12}(D)}$$

TABLE II
REACTOR MODEL DESCRIPTION

	Equation
Species balance	$\frac{dC_A}{dt} = \frac{q_r}{V_r}(c_{A0} - c_A) - k_1 c_A - k_3 c_A^2$
	$\frac{dc_B}{dt} = -\frac{q_r}{V_r} c_B + k_1 c_A - k_2 c_B$
Energy balance	$\frac{dT_r}{dt} = \frac{q_r}{V_r}(T_{r0} - T_r) - \frac{\Delta H_r}{\rho_r c_{pr}} + \frac{A_r U}{V_r \rho_r c_{pr}}(T_c - T_r)$
	$\frac{dT_c}{dt} = \frac{1}{m_c c_{pc}}(Q_c + AU(T_r - T_c))$
Rate coefficients	$k_j = k_{0j} \cdot \exp\left(-\frac{E_j}{RT_r}\right),\ for\ j = 1,\ 2,\ 3\text{To}$
Reaction enthalpy	$\Delta H_r = h_1 \cdot k_1 \cdot c_A + h_2 \cdot k_2 \cdot c_B + h_3 \cdot k_3 \cdot c_A^2$

III. RESULTS AND DISCUSSION

A. *Training*

Fig. 1 compares the training performance of the DDPG algorithm when the observation space includes the differential error (PID) vs. when it does not (PI). When the differential error is included, the training performance is significantly improved as illustrated by the higher reward achieved in a smaller number of steps. The reason for this improvement is due to the information we are inserting by including the differential error, which tells the agent about the rate of change of error. This improvement illustrates that knowledge of control theory can help in optimizing the structure of the RL agent for control applications.

Fig. 1 Training of reinforcement learning agent with DDPG.

B. *Setpoint Tracking*

Fig. 2 shows the performance of the RL agent with respect to setpoint tracking. The figure on the top right shows the closed-loop response of the CSTR to step changes of the setpoint (Tr), while the bottom figure shows the controller's action to adopt to the new setpoint. The error between the setpoint and the process variable is less than 5%, which demonstrates the controller's ability to track the setpoint with relatively fast response time and no overshoot.

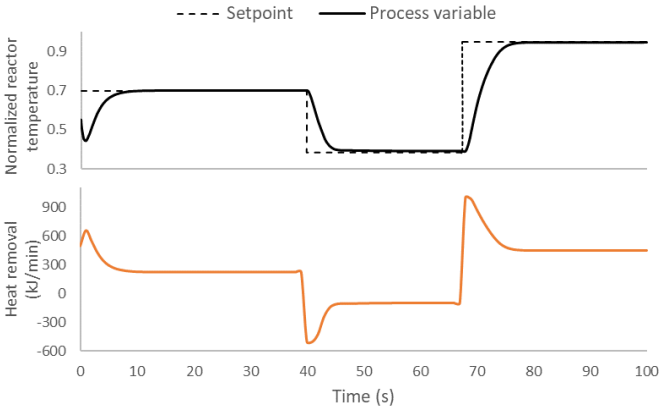

Fig. 2 Performance of RL agent with respect to setpoint tracking. Top: controlled variable. Bottom: controller's action

C. *Disturbance Rejection*

To represent measurement uncertainty, white noise is introduced to the temperature sensor. In addition, to evaluate the controller's disturbance rejection, a sudden drop in the feed concentration is simulated at time 50sec, as shown in Fig 3. The top figure shows a drop in the temperature, while the bottom figure shows the controller's action. The

controller was able to recover very fast to the disturbance, despite the presence of noise.

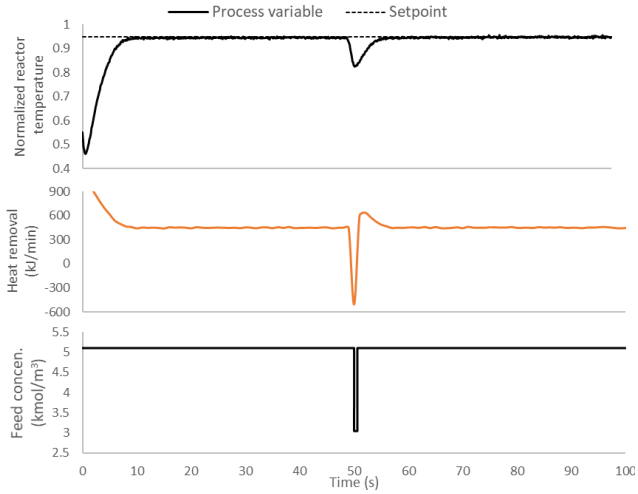

Fig. 3 Performance of RL agent with respect to disturbance rejection. Top: controlled variable (Tr). Middle: controller's action. Bottom: drop in feed concentration.

D. Robustness to Parameter Uncertainties

To study the robustness of the controller, the activation energies were increased by 5% and the overall heat transfer coefficient (U) was decreased by 38%. The change in U resembles real life situations when the heat exchange efficiency decreases as a result of fouling. As Fig. 4 reveals, the controller demonstrates robustness toward plant mismatch as evidenced by its ability to maintain the setpoint.

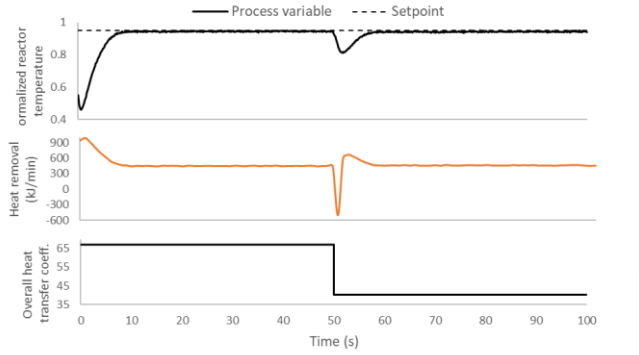

Fig. 4 Performance of RL agent with respect to robustness. Top: controlled variable. Middle: controller's action. Bottom: change in model parameter.

IV. Conclusions

This work showed that the choice of the observation space and the design of the reward function strongly influences the training efficiency and the controller's performance. The RL-based controller showed excellent setpoint tracking, disturbance rejection, and robustness despite being trained without noise or disturbance. This performance demonstrates that RL controllers present a promising process control option in cases where the control law is difficult to derive. The choice of observation space and reward function in this work is not specific to the reaction network presented, which allows its implementation to other systems (chemical or otherwise). Future work will include a case for multiple-input and multiple-output (MIMO) control. In addition, the deployment of the controller to a lab-scale reactor will be studied. Finally, the performance of the RL controller will be compared in detail with nonlinear model predictive controllers according to the same performance measures discussed in this paper.

References

[1] Hoskins, J. C., & Himmelblau, D. M. (1992). Process control via artificial neural networks and reinforcement learning. Computers & chemical engineering, 16(4), 241-251.

[2] Shin, J., Badgwell, T. A., Liu, K. H., & Lee, J. H. (2019). Reinforcement Learning–Overview of recent progress and implications for process control. Computers & Chemical Engineering, 127, 282-294.

[3] Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., ... & Wierstra, D. (2015).

[4] Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971.

[5] Chen, H., Kremling, A., & Allgöwer, F. (1995, September). Nonlinear predictive control of a benchmark CSTR. In Proceedings of 3rd European control conference (pp. 3247-3252).