Master's Thesis Engineering Technology

2020-2021

Application of Reinforcement Learning for Continuous Stirred Tank Reactor (CSTR) temperature control

Hendrikx Jelco

Master of Chemical Engineering Technology

1 INTRODUCTION

A new state-of-the-art algorithm, **Soft Actor-Critic** (SAC), proved to outperform other algorithms. It is hypothesized that SAC also outperforms the current used algorithms in chemical process control.

Reinforcement Learning structure [1]:

1. **Agent** takes an **action**
2. Action changes **environment**
3. Agent receives **state** and **reward**
4. Updating agent's **policy**

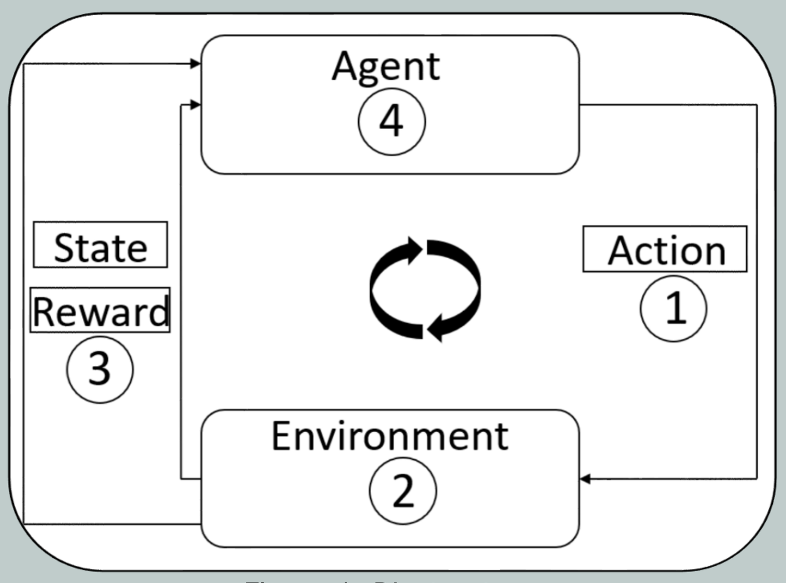

Figure 1: RL structure

The environment - CSTR with **Van de Vusse reaction** [2]:

a ➔ b ➔ c

a ➔ d

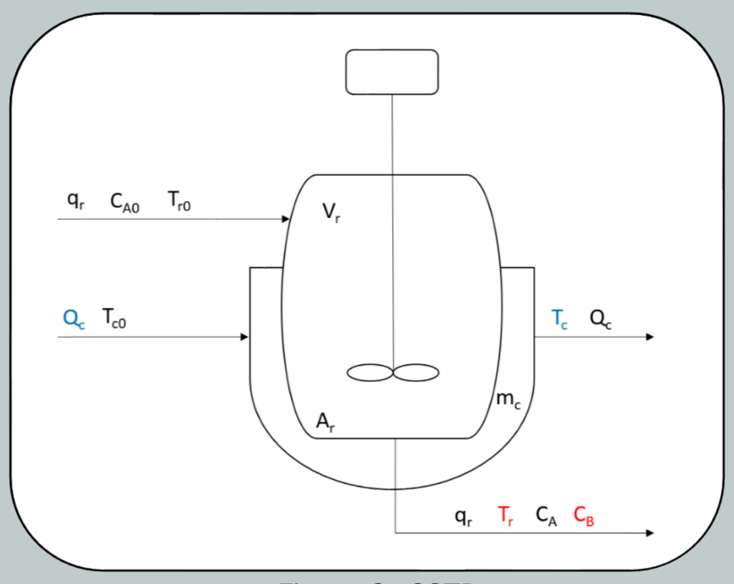

Figure 2: CSTR

2 METHODS

Soft Actor-Critic is used for **temperature** and **concentration** control of the reactor by using the **heat removal (Q_c) or cooling jacket temperature (Tc)** value as **action**

The advantages of Soft Actor-Critic [3]:

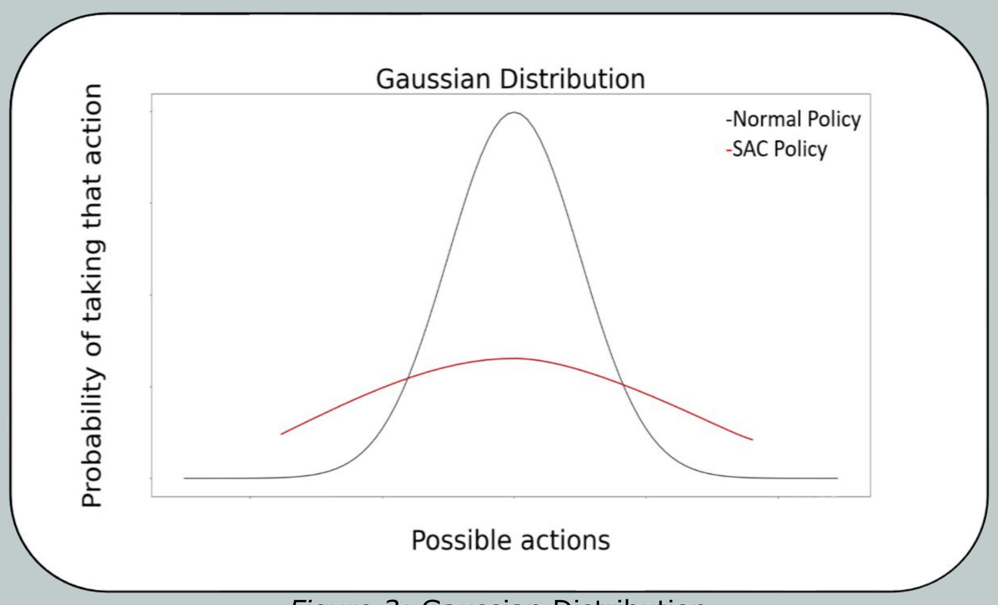

Figure 3: Gaussian Distribution

Maximum Entropy Framework (red line):

Policy = Gaussian Distribution
Broad range of actions in state
➔ more action exploration
➔ better adapted policy

Two Critic networks:
Neural Networks = collection of connected nodes
Neural Network [4] ➔ approximate Q-function
➔ Q-value
➔ action was good or bad
➔ update policy (Actor)

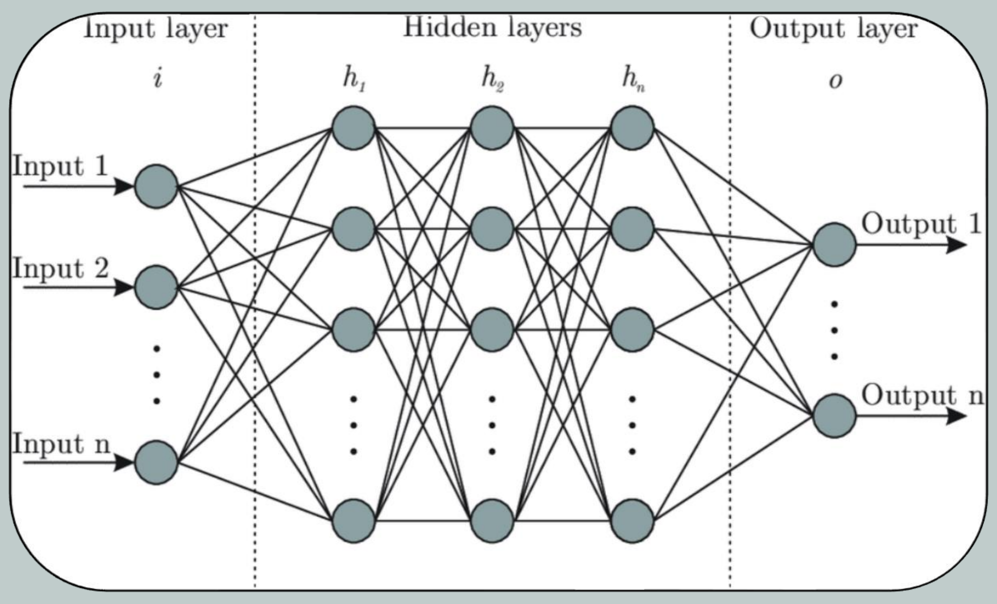

Figure 4: Neural Network

Off-policy model-free:
No pre-defined model of environment needed
Q-value estimation is based on next state and action instead of next state and current action

3 TROUBLESHOOTING

Reward function: specific for each custom environment

Amount of training steps used by SAC for learning process control

4 RESULTS

Trained results
(concentration goal= 1.10 ± 0.05 kmol/m³)

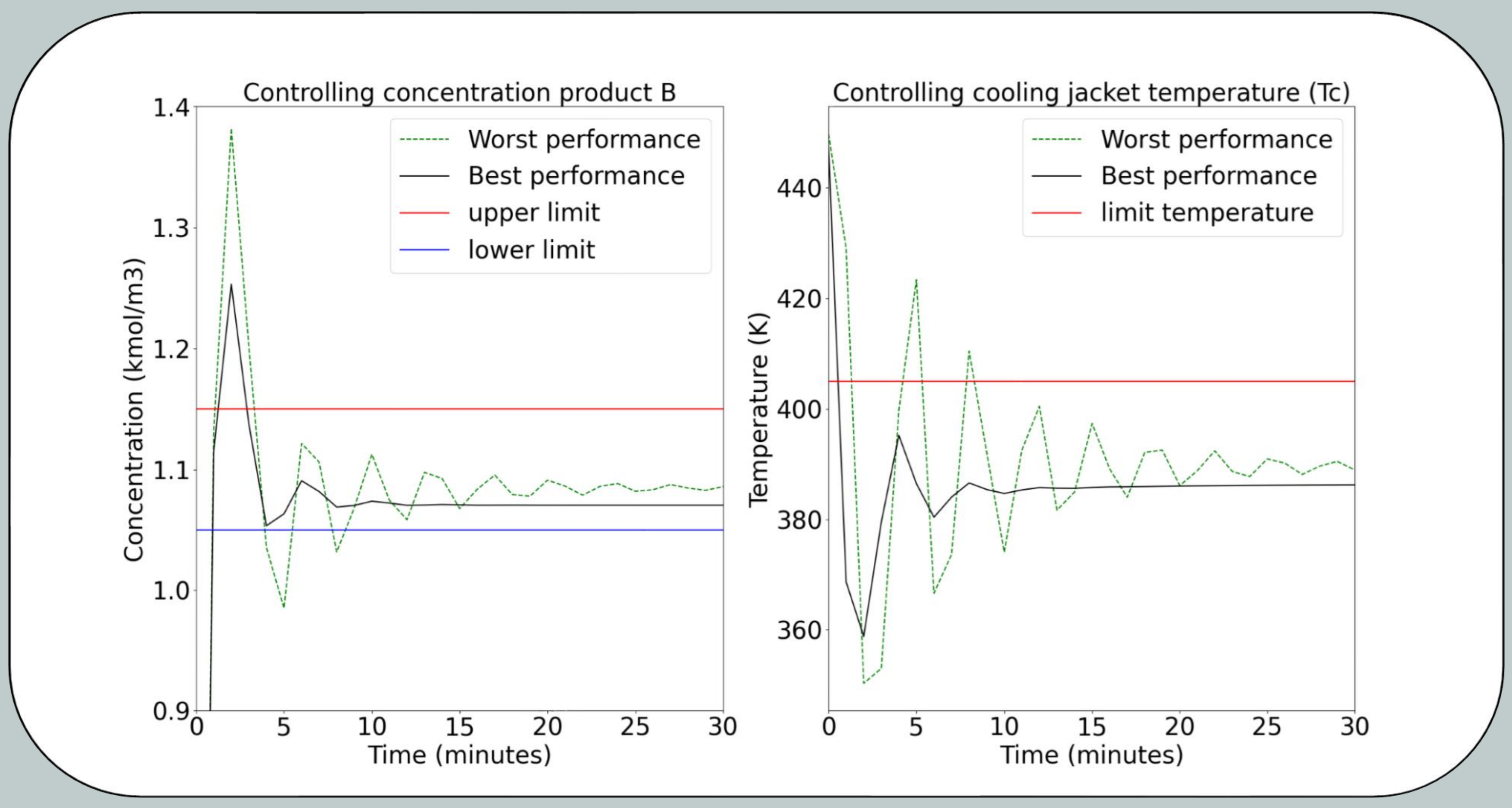

Figure 5: Trained results

Performance of SAC

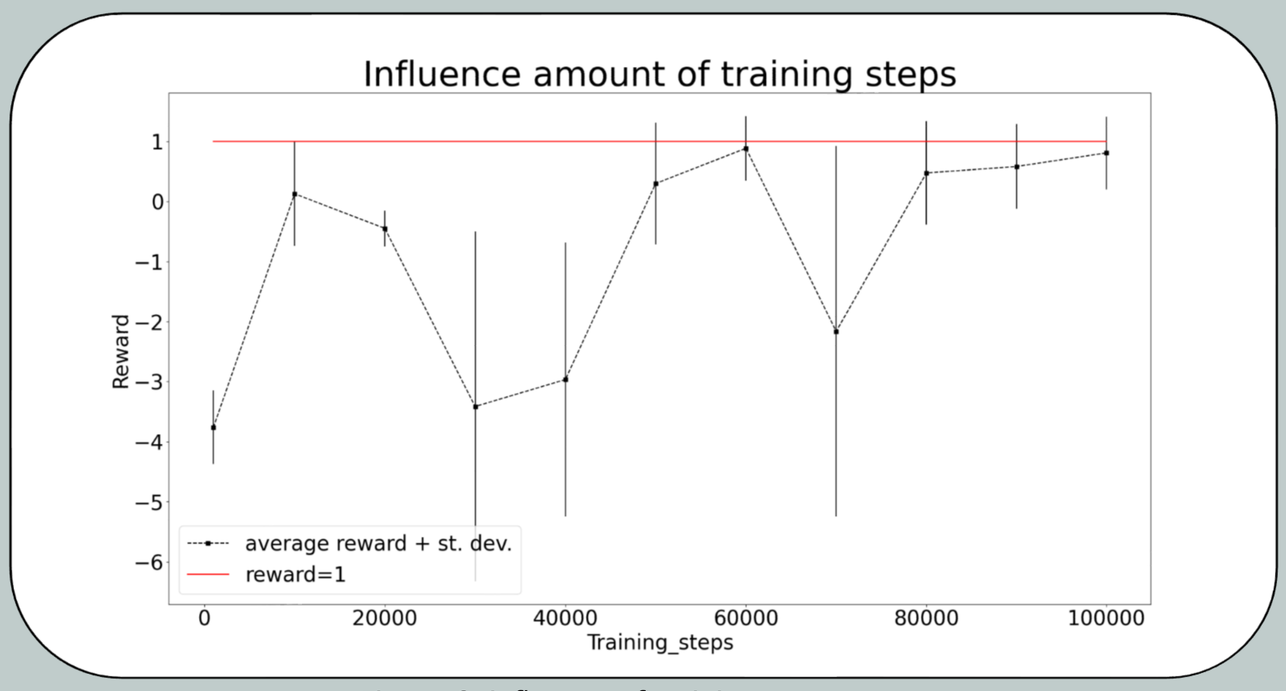

Figure 6: influence of training steps

5 CONCLUSION & FUTURE WORK

SAC can control the temperature at 375±0.05 K or concentration of product b at 1.10±0.05 kmol/m³

Best performance:
100 000 training steps, 0.2 minutes time step.

Future work:
➔ Making reward function time dependable
➔ Extensive analysis of used training steps
➔ possible better performance
➔ Use efficiency or energy consumption instead of temperature and concentration

Supervisors Prof. dr. ir. Leblebici Mumin Enis (KU Leuven)
ir. Wu Min (KU Leuven)

[1] M. Sewak, *Deep Reinforcement Learning*. Singapore: Springer Nature Singapore Pte Ltd., 2019.
[2] J. Vojtesek, P. Dostal, and V. Bobal, *Control of nonlinear system - Adaptive and predictive control*, vol. 7, no. PART 1. IFAC, 2009.
[3] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," *35th Int. Conf. Mach. Learn. ICML 2018*, vol. 5, pp. 2976–2989, 2018.
[4] F. Bre, J. M. Gimenez, and V. D. Fachinotti, "Prediction of wind pressure coefficients on building surfaces using artificial neural networks," *Energy Build.*, vol. 158, no. November, pp. 1429–1441, 2018, doi: 10.1016/j.enbuild.2017.11.045.

Center for Industrial Process Technology

UHASSELT

KU LEUVEN