



Contribution ID: 46

Type: **not specified**

## Reinforcement Learning for Batch Optimization

*Monday, 17 May 2021 15:20 (20 minutes)*

Reinforcement Learning (RL) is one of the three basic machine learning paradigms, alongside supervised and unsupervised learning. RL focuses on training an agent to learn an optimal policy, maximizing cumulative rewards from the environment of interest [1]. Recent developments in RL have achieved remarkable success in various process optimization and control tasks, where multiple applications have been reported in the literature, including parameter tuning for existent single PID control loops [2], supply chain management [3] and robotics operations [4].

The main challenge in applying RL in industrial settings concerns the training of the agent. In the training phase, the agent improves its policy based on many input-output experimentations. However, the number of experimentations that can be collected from real manufacturing processes are prohibitively high. In addition, the explorable operational spaces are often limited due to quality constraints. Therefore, the only feasible alternative is to utilize a model, either a first-principles based or a data-driven machine learning surrogate. In this work, we tested and compared three state-of-the-art RL approaches to optimize an industrial batch case study: Proximal Policy Optimization (PPO), Advantage Actor Critic (A2C), and Soft Actor Critic (SAC). These RL methods optimize the batch process by controlling the reaction temperature and raw material feed rate in order to maximize the total reward (the reward is defined as the profit margin, subject to certain process and safety constraints). Both a first-principle model and a surrogate model are used to generate the required data for the training of the agent.

The aforementioned RL methods were compared based on their convergency rates and sample efficiency, as well as their proposed optimized trajectory. These trajectories are further compared to the batch profiles currently employed in the plant. The different solutions obtained lead to a better understanding of critical batch periods, whereas the different convergency rates allow the identification of the best RL learning algorithm for this process. This information is critical for developing real-time control strategy that can lead to batches with maximum profit margin.

### References

- [1] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing atari with deep reinforcement learning. arXiv preprint arXiv:1312.5602.
- [2] Badgwell, T. A., Liu, K. H., Subrahmanya, N. A., & Kovalski, M. H. (2019). U.S. Patent Application No. 16/218,650.
- [3] Gokhale, A., Trasikar, C., Shah, A., Hegde, A., & Naik, S. R. (2021). A Reinforcement Learning Approach to Inventory Management. In *Advances in Artificial Intelligence and Data Engineering* (pp. 281-297). Springer, Singapore.
- [4] Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. arXiv preprint arXiv:1801.01290.

**Primary authors:** RENDALL, Ricardo (Dow Inc.); CASTILLO, Ivan (Dow Inc.); WANG, Zhenyu (Dow Inc.); H. CHIANG, Leo; PENG, You (Dow Inc.)

**Presenter:** RENDALL, Ricardo (Dow Inc.)

**Session Classification:** Modelling / DoE for optimization

**Track Classification:** Data Science in Process Industries