A Reinforcement Learning Approach to Dynamic Drinking Water Chlorination Control

Wei Dai[1], **Zehua Cheng**[2], **Jiahao Sun**[1]
[1]FLock.io, London, United Kingdom
[2]University of Oxford, Oxford, United Kingdom
{weidai, sun}@flock.io, zehua.cheng@cs.ox.ac.uk

Abstract

This report presents a reinforcement learning (RL) solution for the IJCAI-25 Drinking Water Chlorination Challenge. The problem involves controlling chlorine injection at five locations in a dynamic water distribution network, with key challenges including partial observability from a limited sensor set, time-varying demands, and unexpected contamination events. Our approach leverages a Proximal Policy Optimization (PPO) agent trained on a specialized 57-dimensional observation space. This feature-rich state representation is meticulously engineered using Exponential Moving Average (EMA) smoothing and multi-order differencing to handle noisy sensor data and capture critical temporal dynamics. Our method successfully learns a robust policy that effectively manages chlorination for safe drinking water while minimizing operational costs, demonstrating the potential of data-driven control in complex hydraulic systems.

1 Introduction

The reliable supply of safe drinking water is a critical public health and infrastructure challenge. A key component of this is maintaining optimal chlorine levels within water distribution networks (WDNs) to neutralize pathogens while avoiding harmful by-products. This task is complicated by dynamic water demands, chlorine decay processes, and unforeseen events such as contamination. The IJCAI-25 Drinking Water Chlorination Challenge [Artelt *et al.*, 2025] provides a realistic simulation environment to address this multi-objective control problem. This report details our reinforcement learning-based solution, which focuses on a robust feature engineering pipeline to enable an agent to learn a generalizable control policy from limited sensor data.

2 Methodology

2.1 Problem Formulation

We formulate the chlorination control problem as a Markov Decision Process (MDP).

- **State (Observation)**: The state is a 57-dimensional observation vector that serves as the input for our RL agent. It is designed to provide a comprehensive view of the network's dynamics, even with a limited number of sensors.
- **Action**: The action space is a 5-dimensional continuous vector representing the chlorine injection rates (in mg/minute) at five designated booster stations.
- **Reward**: The reward function is a composite signal designed to guide the agent towards minimizing all competition metrics. It integrates penalties for chlorine concentration bound violations, high control costs, and abrupt changes in injection rates, thereby incentivizing a safe and efficient control policy.

2.2 Feature Engineering for Observation Space

A key component of our solution is the design of a specialized 57-dimensional observation vector. This vector is constructed from three distinct sets of features, each serving a specific purpose in handling the noisy and dynamic nature of the environment.

1. **Smoothed Raw Features (19 dimensions)**: The raw 19-dimensional sensor readings, comprising data from two flow sensors and seventeen chlorine concentration sensors, are first processed using an Exponential Moving Average (EMA) filter. This step effectively mitigates the impact of high-frequency noise, providing a stable and clean representation of the network's current state to the agent.
2. **Smoothed First-Order Difference Features (19 dimensions)**: The first-order difference of the smoothed raw features is calculated to capture the rate of change in the network's dynamics. This new 19-dimensional vector, which reflects trends in chlorine and flow, is also smoothed with an EMA filter to ensure the trend information is not dominated by residual noise.
3. **Smoothed Second-Order Difference Features (19 dimensions)**: A second-order differencing is performed on the smoothed first-order difference features. This serves as a proxy for the 'acceleration' of the network dynamics, providing an early indicator of significant changes or anomalies. This vector is also smoothed with an EMA filter for stability.

This multi-part feature vector allows the agent to simultaneously perceive the current state, its local trends, and its rate of change, enabling a more informed and predictive control strategy.

2.3 Reinforcement Learning Model

We employ a Proximal Policy Optimization (PPO) agent with a Multi-layer Perceptron (MLP) policy network. PPO is an on-policy algorithm known for its stability and strong performance in continuous control tasks. The policy network receives the 57-dimensional observation vector and outputs the 5-dimensional action vector. The agent is trained iteratively, collecting samples from the environment and updating its policy to maximize the long-term cumulative reward.

3 Experimental Setup and Results

Our agent was trained on the ten 6-day long scenarios provided by the competition. To ensure the model's robustness and generalization, we opted not to use pre-computed hydraulics, thereby exposing the agent to dynamic hydraulic conditions. The training process involved an iterative cycle of sample collection and policy updates.

The proposed PPO agent, with its advanced feature engineering pipeline, demonstrated effective control of the water distribution network. The agent successfully learned to maintain chlorine concentrations within the specified safety bounds across a variety of scenarios. The multi-part feature vector, especially the inclusion of dynamic trend information, proved instrumental in enabling the agent to learn a robust and generalizable policy. The agent also demonstrated an ability to minimize the total amount of injected chlorine, thereby meeting the competition's multi-objective criteria.

4 Conclusion

This report presented a reinforcement learning-based solution for the IJCAI-25 Drinking Water Chlorination Challenge. By formulating the problem within an RL framework and employing a sophisticated feature engineering pipeline, we successfully trained a PPO agent capable of robustly controlling a complex water distribution network. The agent learned to effectively balance competing objectives of safety, cost, and efficiency, showcasing the potential of RL for real-world control problems. Future work could explore the use of off-policy algorithms for improved sample efficiency or integrate graph neural networks to explicitly model the network's topology.

Ethical Statement

There are no ethical issues.

Acknowledgements

The authors thank the organizers of the IJCAI-25 Drinking Water Chlorination Challenge for providing a realistic and challenging simulation environment. The authors acknowledge the EPyT-Control platform for providing the core simulation and control framework.

References

[Artelt *et al.*, 2025] André Artelt, Janine Strotherm Luca Hermes, Barbara Hammer, Stelios G. Vrachimis, Demetrios G. Eliades Marios S. Kyriakou, Marios M. Polycarpou, Sotirios Paraskevopoulos, Stefanos Vrochidis, Riccardo Taormina, Dragan Savic, and Phoebe Koundouri. 1st AI for Drinking Water Chlorination Challenge. https://github.com/WaterFutures/AI-for-Drinking-Water-Chlorination-Challenge-IJCAI-25, 2025.