Leveraging Evolutionary Surrogate-Assisted Prescription in Multi-Objective Chlorination Control Systems

Rivaaj Monsia , **Risto Miikkulainen** , **Olivier Francon** and **Daniel Young**
Project Resilience
rivaaj@utexas.edu, risto@cs.utexas.edu, olivier.francon@cognizant.com, danyoung@utexas.edu

1 Introduction

This short, written report introduces the idea of Evolutionary Surrogate-Assisted Prescription (ESP) and presents preliminary results on its potential use in training real-world agents as a part of the 1st AI for Drinking Water Chlorination Challenge at IJCAI-2025. This work was done by a team from Project Resilience, an organization interested in bridging AI to real-world problems.

1.1 Background

The need for reinforcement learning algorithms has exploded as ideas for new agents and agentic systems with real-world applications have been realized. In addition, new perspectives on how to ascertain ideal agents has also expanded– namely through an evolutionary lens. Leveraging evolutionary principles in commonly multi-objective, noisy fitness landscapes has shown to be advantageous in producing diverse, creative solutions compared to standard RL algorithms. Thus, for the competition at hand, we focus on utilizing the ESP framework [Francon *et al.*, 2020] to successively train surrogates to model the reward landscape of chlorination control systems, evolve NEAT (Neuroevolution of Augmenting Topologies) [Stanley and Miikkulainen, 2002] networks via the surrogate, and collect new data by evaluating the best evolved agents on the RL environment.

1.2 Motivation

This project was undertaken by a group from Project Resilience, an organization which was initiated under the Global Initiative on AI and Data Commons (United Nations ITU) to build and promote a public utility to enable researchers, decision-makers, and domain experts to help contribute to AI initiatives focused on real-world problems. We are especially interested in evolutionary AI and have adapted much previous work, a lot undertaken by members of the team, to build up an evolution-focused RL approach for the competition.

2 Methods

In this section, we briefly present the foundation of the main idea utilized by our approach, ESP, and our specific implementation details including evolutionary configurations, the architecture of agents, etc.

2.1 Evolutionary Surrogate-Assisted Prescription

The ESP framework is comprised of two models, a prescriptor (P_s) which prescribes actions (A) from some observations (O) (1) and a predictor (P_d) which models a reward landscape and predicts a reward (R) from an observation-action (O, A) pair (2):

$$P_s(O) = A \tag{1}$$

$$P_d(C, A) = R \tag{2}$$

In turn, instead of utilizing an explicit reward function, which may be costly in terms of efficiency, a surrogate/predictor may be utilized, speeding up learning. In addition, Francon *et. al.* showed that ESP converges faster than standard RL methods (PPO, DQN) with a much lower variance and cost/regret.

ESP takes form as a cycle until agents/prescriptors reach convergence:

1. Collect initial data from a variety of sources/models.
2. Train a surrogate on said data via gradient descent.
3. Evolve prescriptors/agents, using the surrogate instead of the explicit reward function.
4. Apply the best prescriptor(s) on the environment and collect new data
5. Repeat until convergence of prescriptors.

Although this is a general framework to ESP, it may be, and likely requires to be, expounded upon.

2.2 Implementation Details

To collect initial data, we evolved NEAT networks and collected (C, A, R) triplets that may be used to train the initial surrogate. In addition, we utilized NSGA-II, a popular multi-objective evolutionary optimization algorithm based on the multi-objective nature of the chlorination systems for the reproduction scheme and fitness evaluation of the NEAT genomes.

The surrogate model was an LSTM-based architecture. 2 separate LSTM's encoded the actions and observations whose outputs were then joined and passed to fully connected layers (with dropout = 0.2, activation function = ReLU). The final output did not have an activation function. The LSTM was provided a sequence of data from 10 consecutive timesteps.

At each iteration, the LSTM was trained for 50 epochs (with early stopping) via the Adam optimization algorithm.

The general configuration for the evolution of the NEAT networks is present in the `neat-nsga2-config.init` file in the submission. In summary, we evolved a population of 100 genomes with topology mutation rates which were not too extreme. In addition, initial parameters are sampled from a normal distribution with ($\mu = 0, \sigma = 1$). To maintain efficiency, 5 scenarios were sampled from the set of scenarios provided to evaluate the best 5 prescriptors on

The final, timestep reward function was simply a composite of only two of the five metrics signified by the competition, chlorine violation bounds and cost of control. The following composite reward was used to train the surrogate and indirectly evolve the prescriptors:

$$
\begin{aligned}
r(C, A) = \frac{-10}{N} \sum_{i=1}^{N} [\max(c_i - 0.4,\ 0) + 5 \cdot \max(0.2 - c_i,\ 0)] \\
+ \frac{5}{N} \sum_{i=1}^{N} 1_{\{c_i \in [0.2,\ 0.4]\}} \\
- 0.1 \cdot \sum_{j=1}^{M} a_j \\
= \text{penalty}_{\text{Cl}} + \text{bonus}_{\text{Cl, in-range}} + \text{cost}_{\text{Cl}},
\end{aligned}
$$

where c_i are chlorine concentrations at each node i and a_j is the chlorine injection action made by the agent at injection node j. We found that, at least initially, lower bound chlorine violations dominated the chlorine penalty, so we weighed it more. To ensure that the agent was able to inject chlorine within a significant range, we also underemphasize the cost control objective.

3 Results

As was mentioned before, we initially only focused on two objectives to optimize, chlorine bound violations and cost of control. As a result, we only focus on limited results for said objectives. Analysis of the other 3 objectives showed at most little improvement throughout ESP-based evolution. ESP itself was ran for 4 iterations.

Based on the final evaluation script and corresponding function, the final solutions had cost control values ranging from $[864, 3709]$. On the other hand, chlorine violation bounds were relatively similar in terms of range $[0.173, 0.175]$, suggesting little variance within populations and/or little, if any learning.

3.1 Pareto Front Analysis

To analyze the multi-objective nature of the task, we generate a Pareto front of the 60 best solutions from the terminal generation and iteration of the ESP cycle (iteration 4). These results are shown in [1]. Note that these values are normalized to emphasize the tradeoffs between each objective rather than absolute performance on the objectives. We can see an obvious linear front, indicating proportional trade-offs between both objectives. This is to be expected as more chlorine must be injected to keep chlorine concentration within bounds, especially in the early stages of learning where, in our case, lower bound violations dominated the violation penalty.

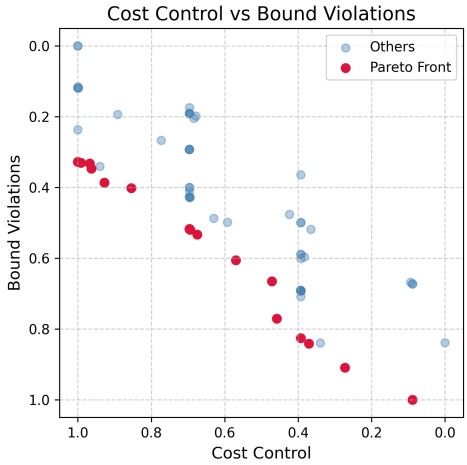

Figure 1: A 2D pareto front of the 60 best prescriptors generated from the ESP method after iteration 4. The values of each objective are normalized against the set of genomes. The pareto front is calculated and highlighted in the red.

4 Limitations and Future Work

Of course, a major limitation of the current work is the inclusion of just 2 of the 5 objectives into our evolution scheme. To demonstrate holistic performance of the evolved agent, we intend to continue shaping the reward function to directly and indirectly optimize for each objective. In addition, utilizing other trained solutions to generate initial data for ESP will likely be necessary to widen the initial reward landscape and force learning. Lastly, implementation of curricular learning, by periodically changing the reward function by way of term coefficients and injection of data from good solutions would likely help the agents learn the system sequentially instead of being overwhelmed by the complex composite fitness. Ultimately, we would like to reach a point where the objectives can be weighed equally such that tradeoffs are relatively similar between each pair of objectives.

We plan on continuing this work as a proof of concept for ESP in real-world, complex agentic systems.

5 Participation Details

A .zip file of the relevant source code was uploaded through the link provided. We will not be able to attend the IJCAI conference in-person.

Ethical Statement

There are no ethical issues.

Acknowledgments

The configuration files for NEAT evolution as well as a general framework for surrogate architecture was created by

Cognizant AI as part of the COVID-19 XPRIZE Pandemic Response Challenge. The code used for NSGA-II was written by Hugo Aboud.

References

[Francon *et al.*, 2020] Olivier Francon, Santiago Gonzalez, Babak Hodjat, Elliot Meyerson, Risto Miikkulainen, Xin Qiu, and Hormoz Shahrzad. Effective reinforcement learning through evolutionary surrogate-assisted prescription. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*, pages 814–822, Cancún, Mexico, June 2020. American Association for Artificial Intelligence.

[Stanley and Miikkulainen, 2002] Kenneth O. Stanley and Risto Miikkulainen. Evolving neural networks through augmenting topologies. *Evolutionary Computation*, 10(2):99–127, 2002.