Name: James Onyejizu

Project Name: Off-Policy Evaluation for Safe Offline Reinforcement Learning

Pillar: Artificial Intelligence

Contact Details:

James.Onyejizu@ibm.com
James.Kel71@gmail.com

James Onyejizu

Social Media:

James Onyejizu

@kel074

techh_gk

Off-Policy Evaluation for Safe Offline Reinforcement Learning

Manager: Dharmashanker Subramanian

Mentors: Long Vu, Lan Hoang

Why Safe RL?

- Respect safety constraints for improved/realistic system performance(Thermal control in buildings).

Project Overview

- Off-policy evaluation leverages offline log data in decision making.

Why OPE:

- high stakes and expensive settings.

Approach:

- Fitted Q-Evaluation(FQE): uses iterative regression in decision making.

Our Contribution(Safe FQE):

Why Safe FQE? Current FQE takes a single metric – In real world situations, we need to consider multiple scenarios + safety-critical measures.

Formulation(Energy Efficiency Use Case)

- Energy efficient RL environment development.
- Safety function to indicate environment safety
- Train RL agents(with safety columns)
- Train FQE with benchmark agents and dataset

Reward/Safety Function

$r_t = -W * \lambda_E * P - (1 - W) * \lambda_T * (\max(T - T_{low}, 0) + \max(T_{up} - T, 0))$ (1)

$R = \sum_{t=0}^{\infty} \gamma^t r_t$ (2)

where $\Upsilon, W \in [0, 1] =$ discount factor, weight

Subject to ci ∈ C,

$c_1 = \{T_z \leq T_{ex}\}$ (3) $c_2 = \{P < P_{ex}\}$ (4)

Current Results

Fig 1-2: Training using a DRL agent(PPO)

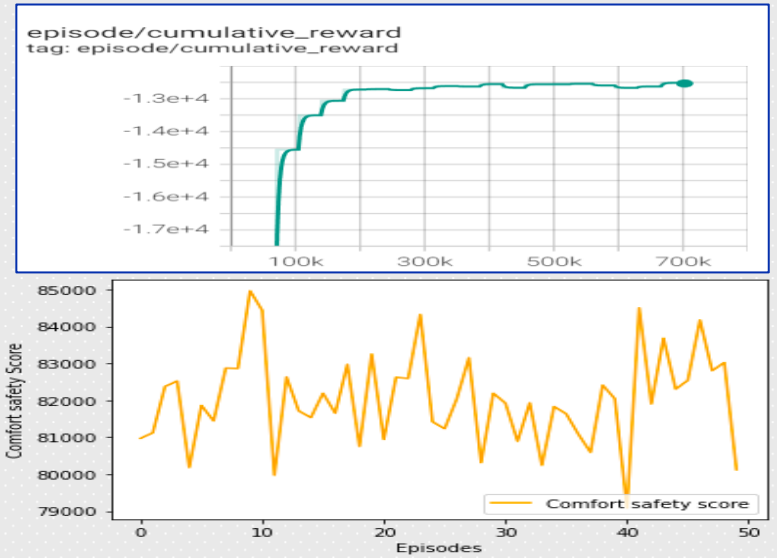
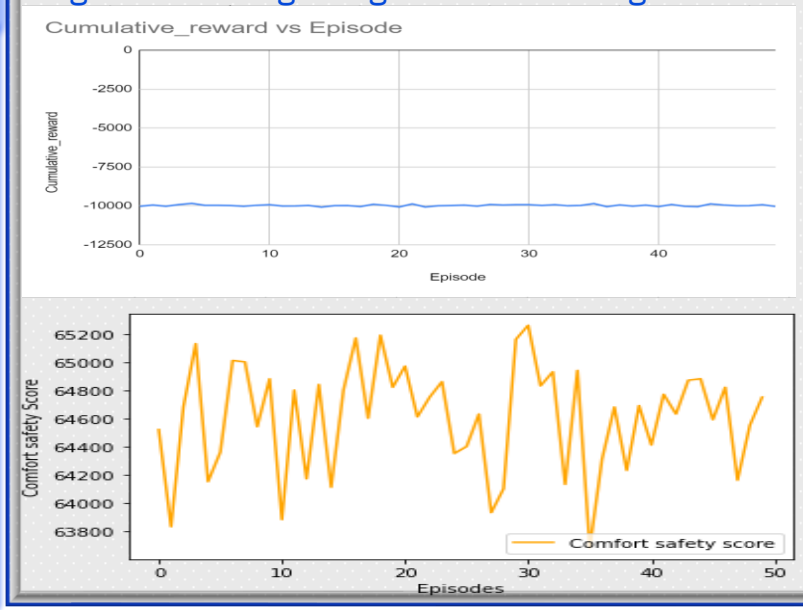


Fig 3-4: Training using a Rule-based agent



Analysis

- The cumulative reward for the PPO agent was comparable to the rule based agent, but received a lower safety score.

Next step

- Train FQE with benchmark agents and generated dataset.

Project's Current State

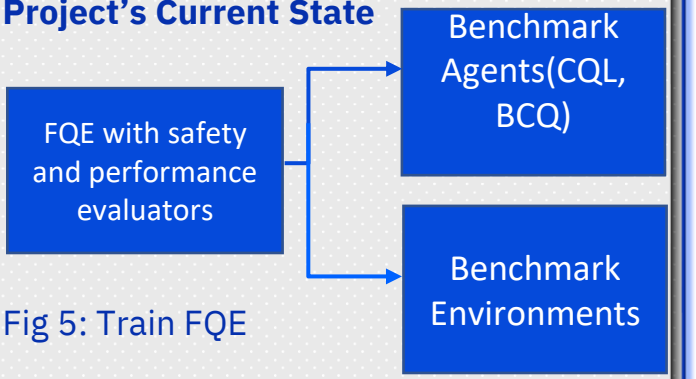


Fig 5: Train FQE

Future Work

How to combine the objectives? (weighted + safety threshold?)

Can FQE be viewed in a distributional way i.e. potential different dataset for FQE fit + test and observations to be evaluated

References

https://realworldsdm.github.io/paper/34.pdf

https://arxiv.org/pdf/2002.03478.pdf -> adding expert knowledge

Name: James Onyejizu

Email: James.Onyejizu@ibm.com