Reinforcement Learning Agents

Frozen lake game

Alexandre Marques up202106956
Beatriz Sá up202105831
João Gonçalves up202107438

About gym environment

The board is represented by a 4x4 map, where each cell is represented by:

- S: Initial state where the agent starts.
- F: Frozen state that the agent can step on safely.
- H: Hole in the ice, if the agent falls here, he loses the game.
- G: Objective state, the place where the agent needs to reach to win the game.

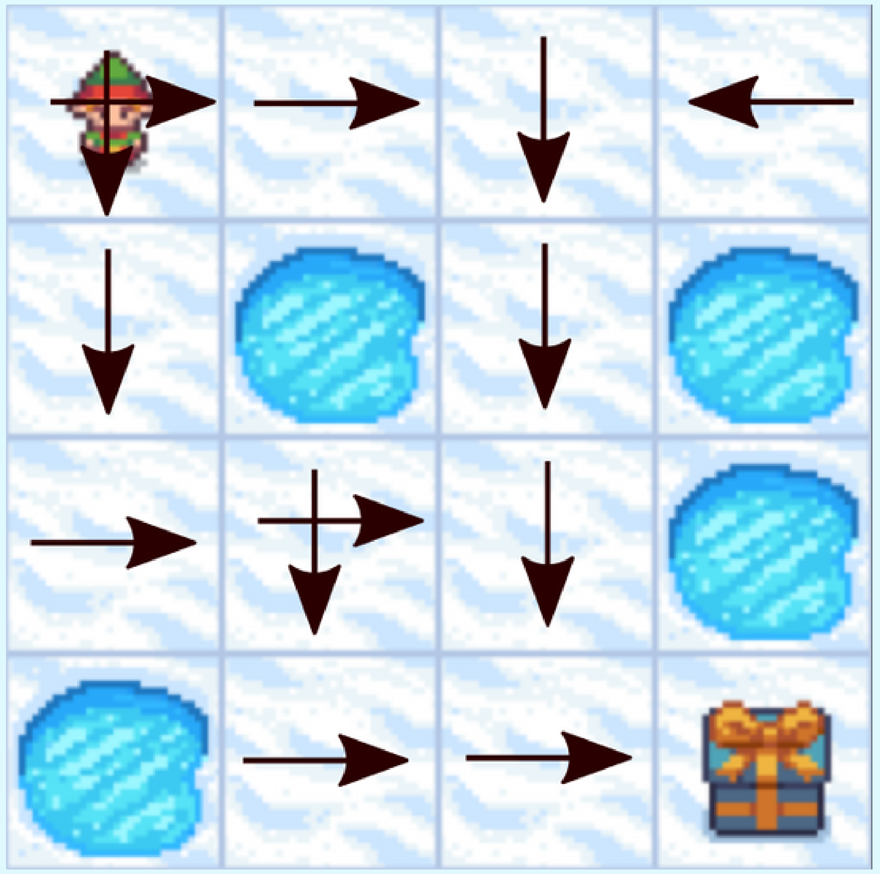
The agent may notice:

- Your current position on the board.
- The rewards or penalties associated with each action taken (moving left, right, up or down).
- Information about the environment around you, such as the presence of a hole (H), the objective (G), or a safe space to move around (F).

About gym environment

The action the player will take is indicated by a value between 0 and 3:

- 0: Move left
- 1: Move down
- 2: Move right
- 3: Move up



Throughout the game the player receives the following rewards:

- Reach goal: +1
- Reach hole: 0
- Reach frozen: 0

Changes in the gym environment

Changing the reward system

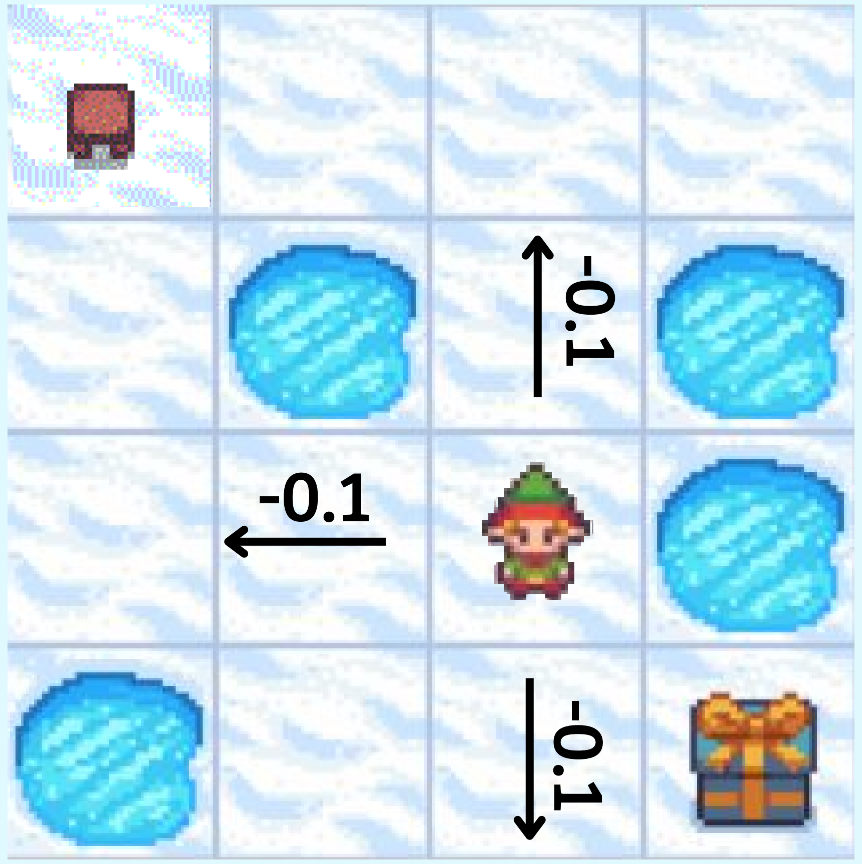
The rewards system is sparse, meaning that rewards are only attributed when achieving the goal

Initially we changed it to:

- Reach goal: +100
- Reach hole: -100
- Reach frozen:-1

After some research, we changed it to:

- Reach goal: +1.6
- Reach hole: -1 0
- Take a step : -0.1



Changes introduced in the agent's actions

Change in floor slippage

By deactivating the frozen cells, the randomness factor will not be considered in this problem, and we can compare the two scenarios to see how this impacts the agent.

RL algorithms

- A2C
- PPO
- DQN
- TRPO
- QR-DQN
- Markable PPO
- ARS

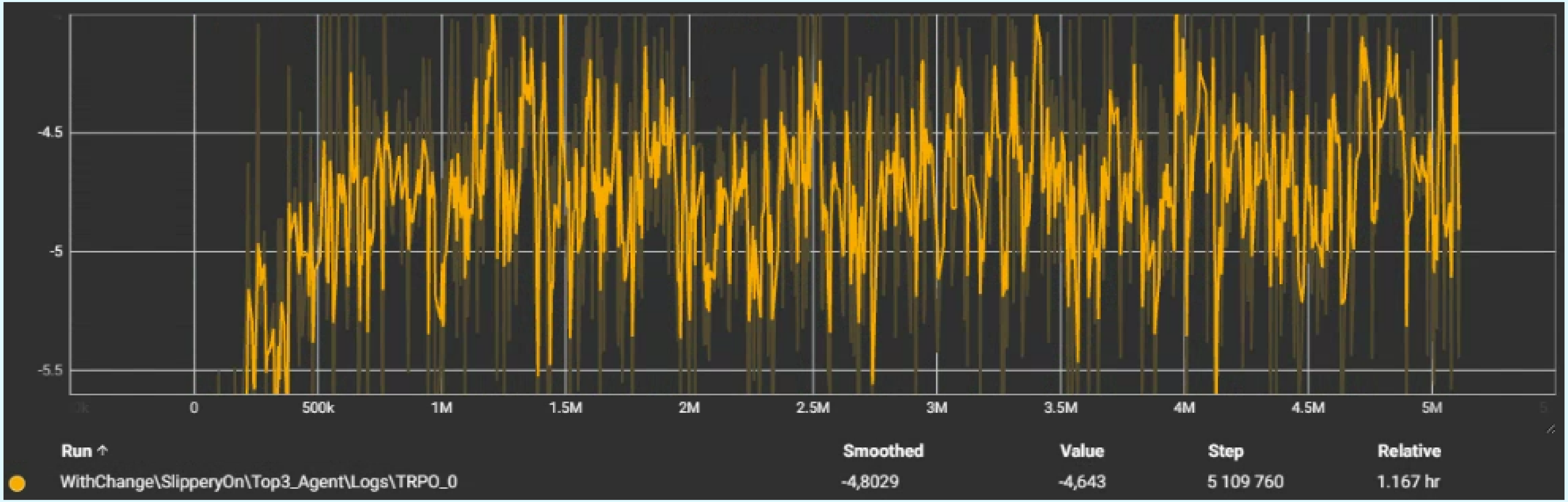
Best 3

- PPO
- TRPO
- A2C

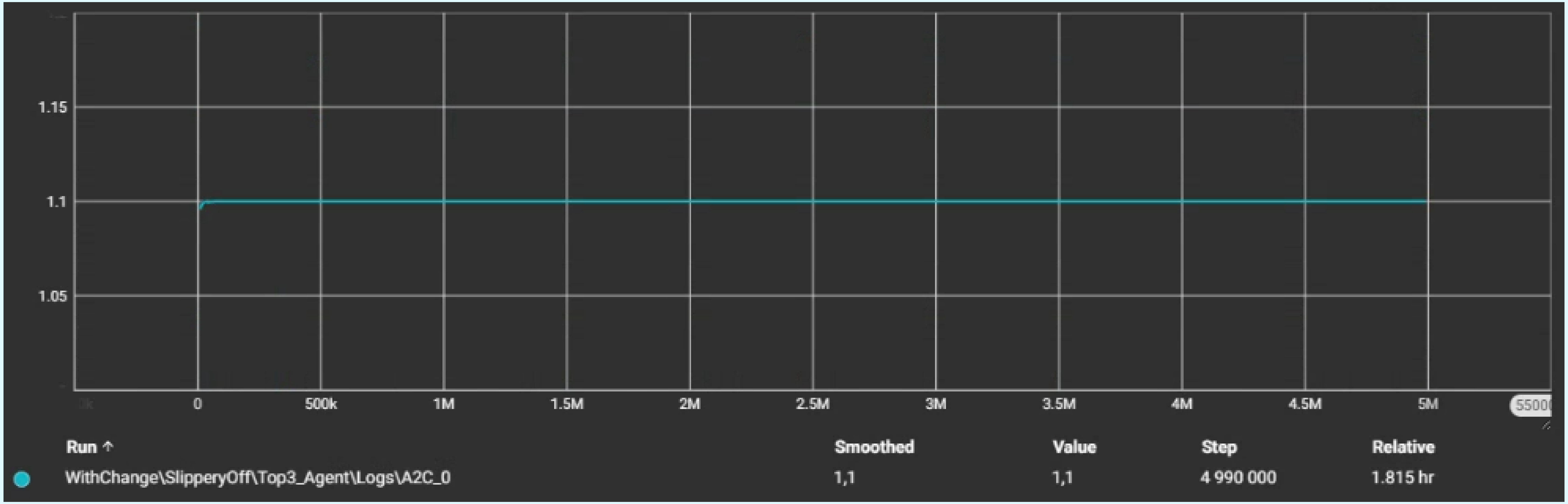
Best with slippery

Best without slippery

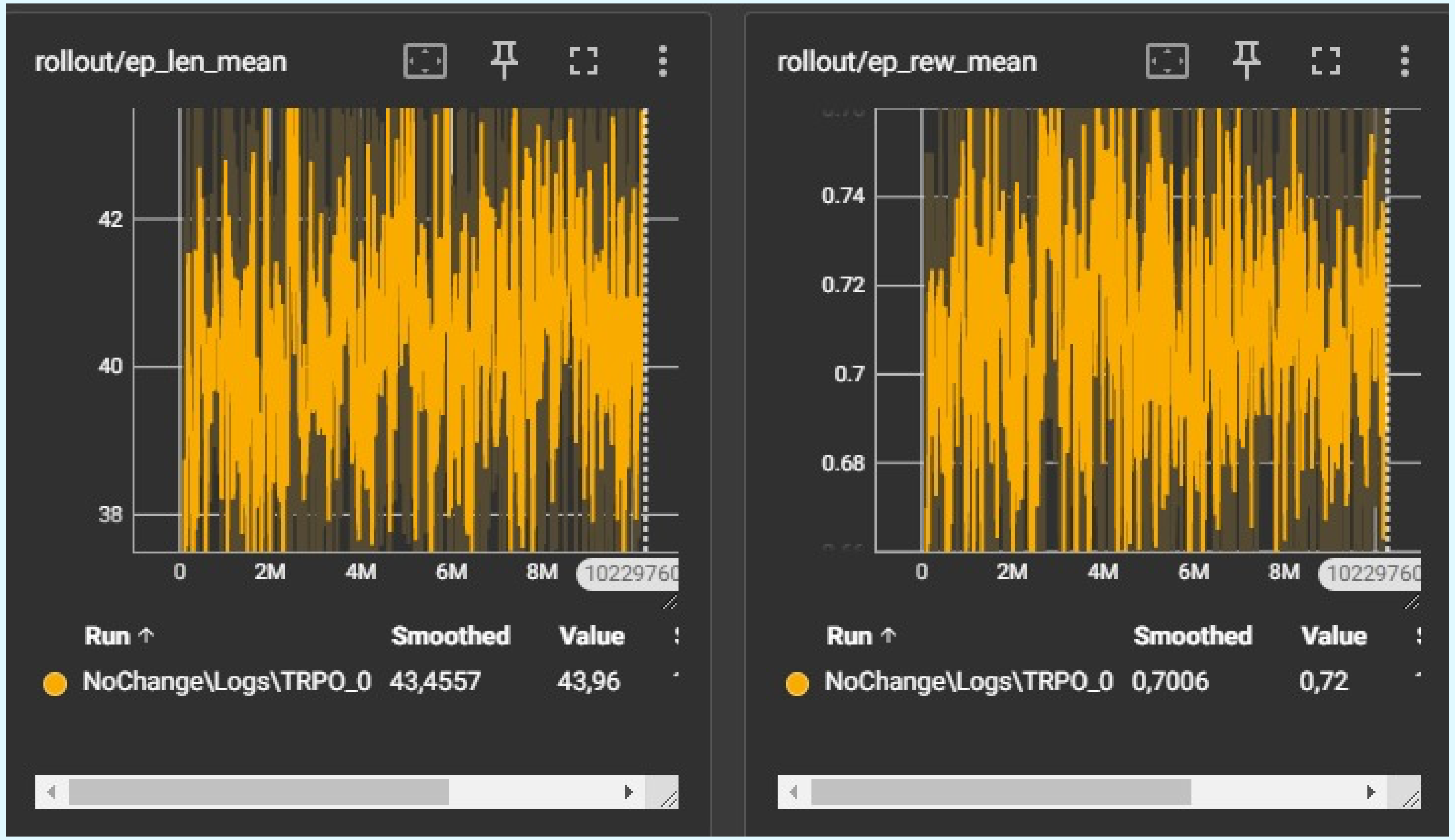
Results with slippage

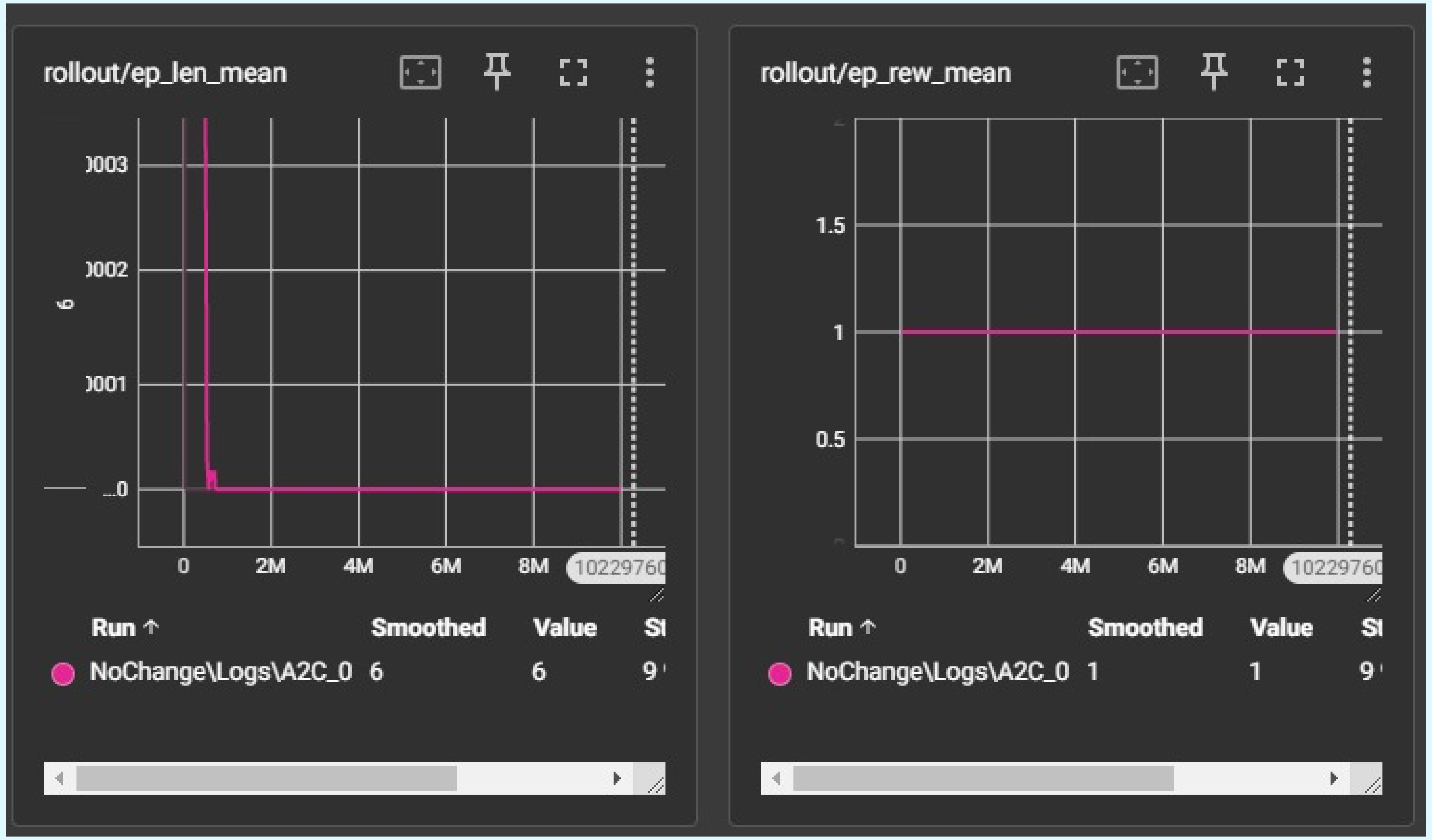


Results without slippage



Comparison of results with the default game





Conclusion

With this work, we obtained better results compared to the initial game. By removing the slipping from the floor, the agent obtained effective results and in a few episodes managed to learn the path very quickly since it always has the same initial and objective state. With the slipping of the floor, he obtained slightly worse results due to the fact that his actions are not always based on his decisions. Nevertheless the agent still performs better with our changes by taking fewer steps to get to the solution.