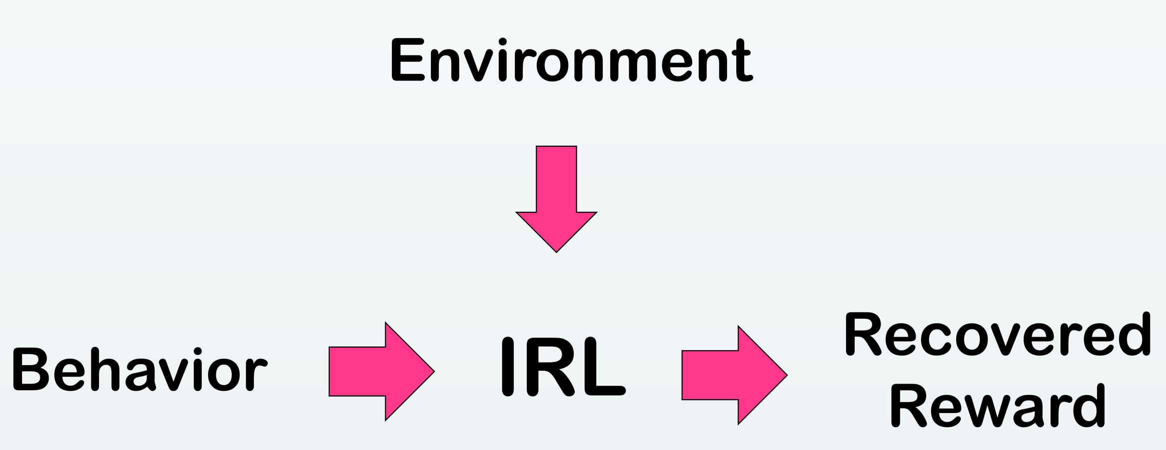
Investigating the Extent to which Inverse Reinforcement Learning can Learn Rewards from Noisy Demonstrations

Charalampos Perdikis – charalampos@perdikis.ac.cy
Responsible Professor: Dr. Luciano Cavalcante Siebert
Supervisor: Angelo Caregnato Neto
TU Delft Delft University of Technology

1. Introduction

Inverse Reinforcement Learning (IRL) infers the underlying reward function from observed behavior. Hence, IRL extracts implicit knowledge from expert demonstrations to generate a reward function.



2. Main Research Question

This research aims to **investigate the extent to which IRL can learn rewards from noisy demonstrations**.

3. Maximum Entropy IRL

Maximum Entropy Inverse Reinforcement Learning (MaxEnt IRL) [1] aims to find a reward function that not only replicates the observed behavior but also maximizes the entropy or uncertainty of the expert's actions.

This allows for a broader range of possible policies that could explain the expert's demonstrated actions.

4. Methodology

To answer the research question, we followed the steps below:

1. Use an implementation of MaxEnt IRL [2].
2. Decide and set up a Markov Decision Process (MDP) environment .
3. Create optimal and noisy expert demonstrations for input to the IRL.
4. Compare the noisy and optimal recovered rewards.

5. Experiments

For our experiments we:

- Set up a 5x5 Grid World MDP as in Figure 1.
- Constructed optimal demonstrations for the defined reward in Figure 1.
- Similarly created noisy demonstrations for three types of noise.
- Compared the optimal and recover rewards according to some metrics.

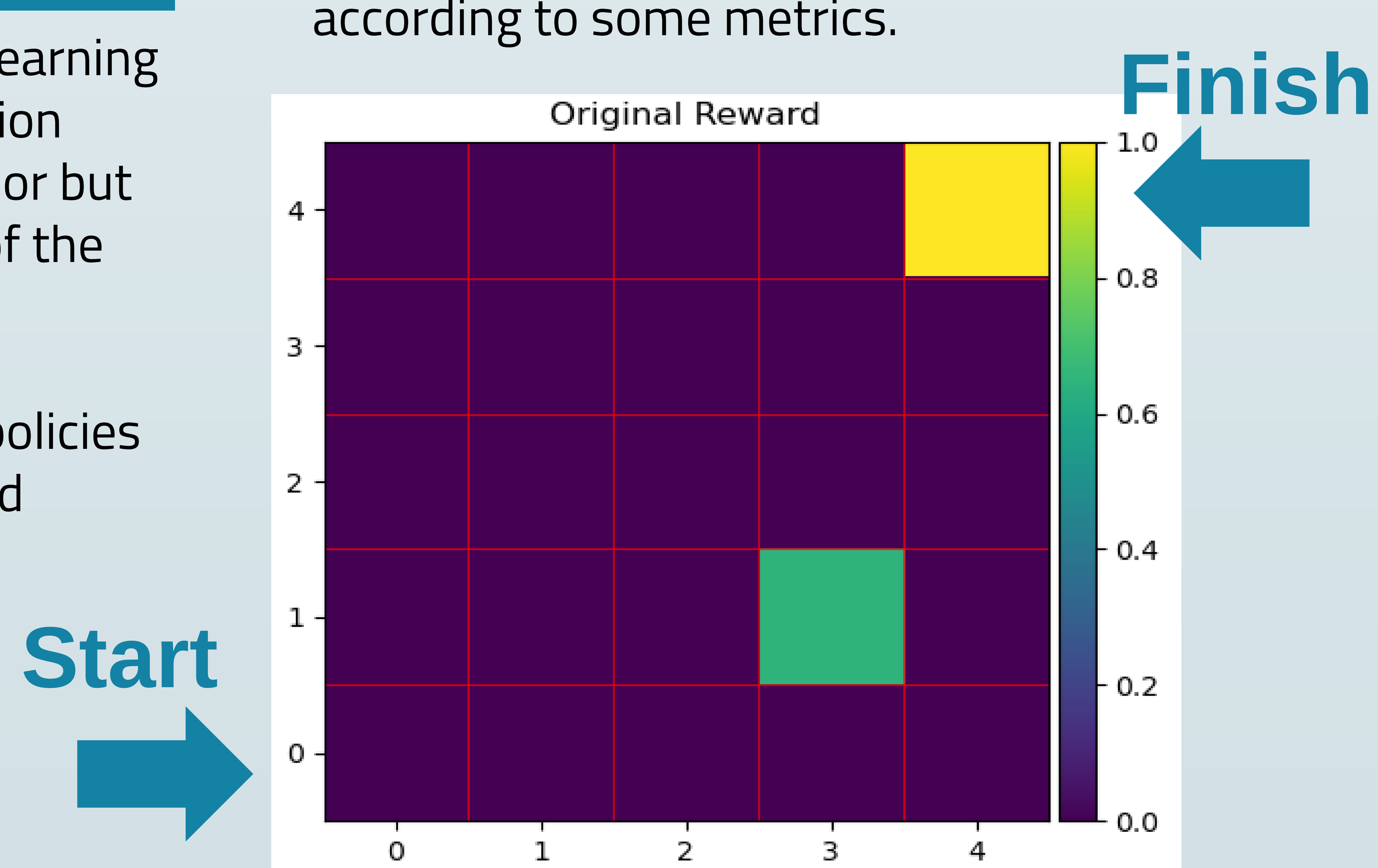


Figure 1: 5x5 Grid World with the defined reward used in generate expert trajectories.

Random Events Noise (REN)

This noise refers to unexpected and unpredictable events that could occur during the execution of actions in an environment.

Random Bias Noise (RBN)

This noise introduces random behavior observed in all demonstrations in a similar way, resulting in a form of bias.

Sparse Noise (SN)

Describes demonstrations were a proportion of them is considered optimal, while the rest have significant anomalies.

6. Results

In the tables 1-3 below, we show results using the metric Failure to Achieve Goal, which counts the number of times the recovered reward failed to produce a path that reaches the final state of the grid, from 100 iterations:

REN probability – Number of failures

0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
0	0	0	0	0	0	2	17

Table 1: Probabilities of REN in the upper row and the number of failures in the lower row.

RBN probability – Number of failures

0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40
2	0	7	13	10	25	36	47

Table 2: Probabilities of RBN in the upper row and the number of failures in the lower row.

SN influence factor – Number of failures

0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
0	0	0	0	0	0	0	0

Table 3: Influencing factor of SN in the upper row and the number of failures in the lower row.

7. Conclusion

From the above results we concluded that:

- Random Bias Noise is detrimental to MaxEnt IRL even in low probabilities.
- Random Events Noise is tolerable with some problems in high probabilities.
- MaxEnt IRL appears to be robust to our simulated Sparse Noise.

Results from other metrics we used, also suggest the above conclusions.

8. Future Work

This research can be extended by:

- Modelling more noise types.
- Mixing noises.
- Changing the IRL algorithm.
- Increasing the complexity of the grid.

References

[1] Brian Ziebart, Andrew Maas, J. Bagnell, and Anind Dey. Maximum entropy inverse reinforcement learning. pages 1433–1438, 01 2008.
[2] Maximilian Luz. Maximum entropy and maximum causal entropy inverse reinforcement learning implementation in python. https://github.com/qzed/irl-maxent, 2019. Accessed: May, 2023.