

Artificial Intelligence 5M - Loch Lomond Lake

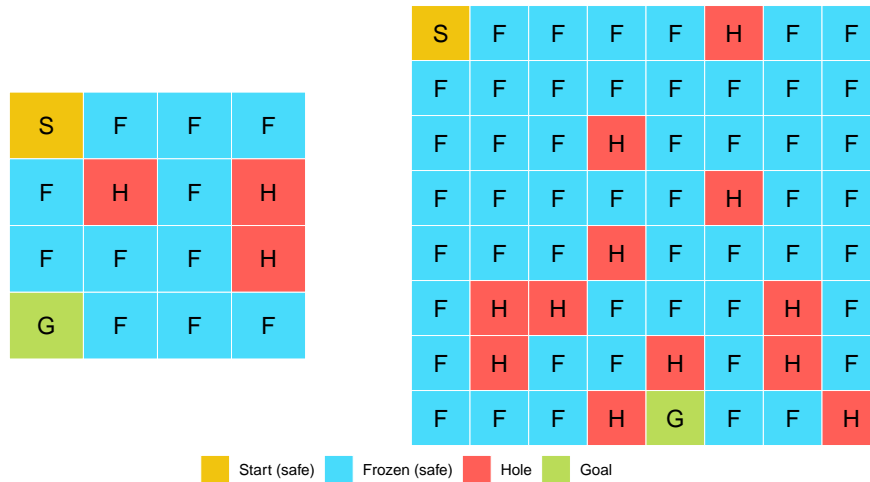
Mayra A. Valdes Ibarra - 2419105v

1 Introduction

The Loch Lomond Frozen Lake environment is a customized Open AI Gym derived from FrozenLake (https://gym.openai.com/envs/#toy_text).

The goal of this report is to design, implement and evaluate three different virtual agents which are able to navigate across the Loch Lomond Frozen Lake grid and retrieve the frisbee disc. Three different agents are analyzed: a senseless agent, a simple agent and a reinforcement agent. This report includes the analysis for the 8x8 grid, but the analysis of the different variants of the 4x4 grids are included in the Appendices.

2 Analysis



3 Methodology

4 Implementation

4.1 Senseless Agent

4.2 Simple Agent

4.3 Reinforcement Learning Agent

This agent was implemented through an active reinforcement learning algorithm. This agent, as stated before, does not have prior knowledge of the environment, but its goal is to learn it via the

sensors/actions each episode. Given our specifications, we need a model-free method for our agent, thus we decided to use Q-learning method. Using a temporal-difference approach¹, we make use of the update equation:

$$Q(s, a) = Q(s, a) + \alpha(R(s) + \gamma \cdot \max_{a'} Q(s', a') - Q(s, a))$$

Where: * $Q(s, a)$ denotes the value of doing action a in state s

$$Q(s, a) = Q(s, a) + \alpha(R(s) + 0.95 \cdot \max_{a'} Q(s', a') - Q(s, a))$$

- s is the current

Our exploration functions are as follows:

$$U(s) = \max_a Q(s, a)$$

The relation between Q-values and the utility values are as follows[]:

$$U(s) = \max_a Q(s, a)$$

Additionally, a custom mapper (`EnvMDP` within `helpers.py` file) was implemented to map the environment from `Open AI Gym` to a `Grid MDP` in order to produce. It is important to note that the results obtained from the *Policy Iteration* and *Value Iteration* algorithms are completely independent from our reinforcement learning agent, and they are only provided as a measure of comparison of final results from our active reinforcement learning agent.

4.4 Environment Modifications

In order to add additional flexibility and generic support to the agents, slight changes were made to the `uofgsocsai.py` file, the one containing the main `LochLomondEnv` environment class. The changes below were approved as long as justification was provided. The changes and justifications are as follows:

- Parameters `map_name_base`, `reward` and `path_cost` were added to the `LochLomondEnv` constructor. The default values are `8x8-base`, `1.0` and `0` respectively. The default values do not alter the functionality from the original file provided.
- Attributes `is_stochastic`, `reward_hole`, `reward` and `path_cost` were added to the class

The reason of the changes for the constructor was to add flexibility to be able to test different scenarios without the need to modify the file every time a different variant was analyzed. The attributes were added to the class in order to be able to access them via the object (e.g. `env.path_cost`) and create a Markov Decision Process out of it. Even though a Markov Decision Process was out of the scope of this project, an inhouse mapper from `Open AI Gym` environment to `Grid MDP` with the only purpose of doing a sanity check between the final U and policy from the Reinforcement Learning agent and the ones that a *Policy Iteration* and *Value Iteration* algorithm would provide.

¹Russell, Stuart J, and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Englewood Cliffs, N.J: Prentice Hall, 1995. Print.

Finally, the way to assign an environment grid was changed from `MAPS_BASE[map_name_base]` to `copy.deepcopy(MAPS_BASE)[map_name_base]`, with the only purpose of being able to instantiate the `LochLomondEnv` more than once in a single run (e.g. `python run_rl.py 1,2,3,4,5,6,7`), which runs all the variants in a single run.

There may be other better ways of accomplishing the same without code changes, but due to the current lack of experience/knowledge in Python programming, time did not permit to find better ways for it.

5 Evaluation

Every agent produces different evaluation files that will be created inside the `out` folder.

6 Conclusions

7 References

8 Appendices

8.1 Appendix A: Title Here

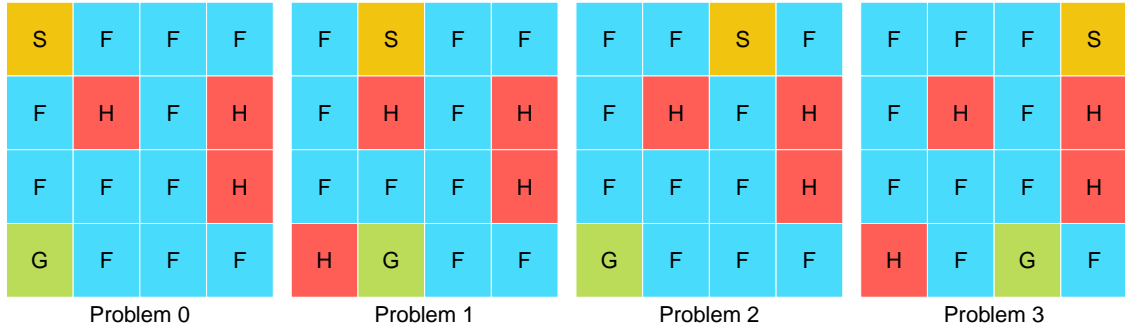


Figure 1: My caption here

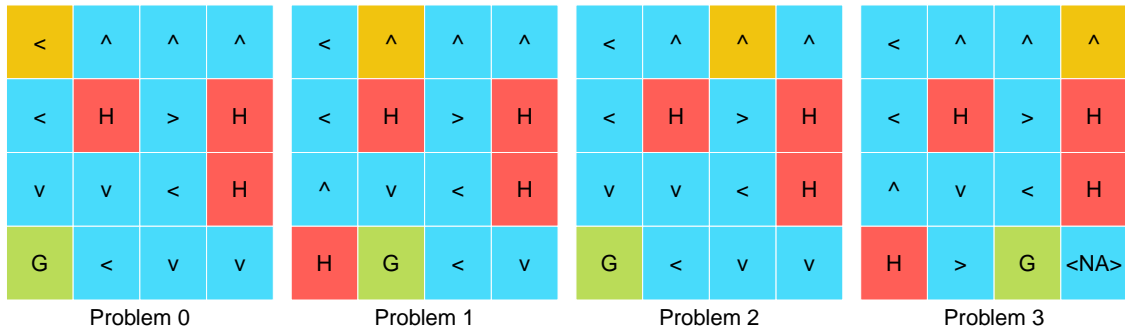


Figure 2: My caption here

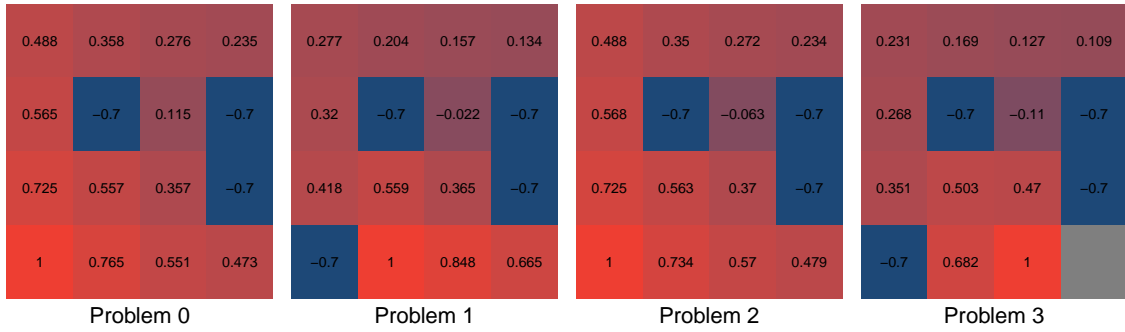


Figure 3: My caption here

8.2 Appendix B: Title Here

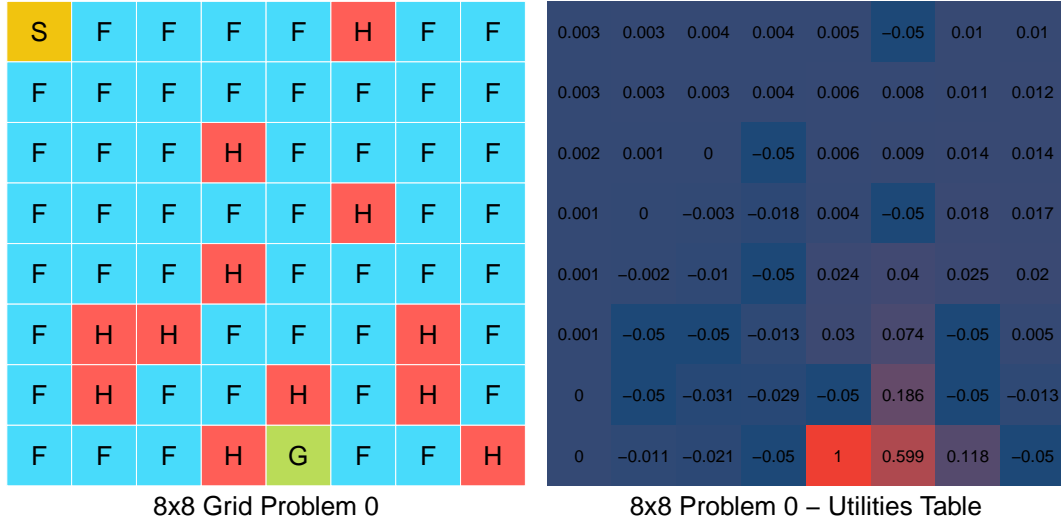


Figure 4: My caption here

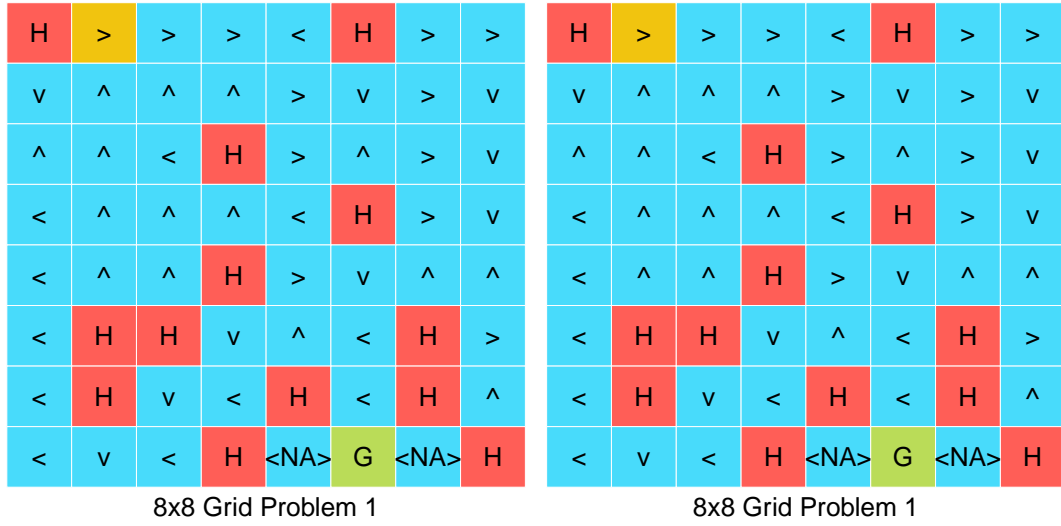


Figure 5: My caption here