Reinforcement Learning for Goal Recognition

Gabriel Waengertner Henrique
Polytechnic School
PUCRS
Porto Alegre, Rio Grande do Sul, Brasil

Abstract

In this work, we aim to integrate Reinforcement Learning with Goal Recognition in planning domains, learning robust policies for inferring goals from sequences of observations. We intend to develop both Reinforcement Learning and Inverse Reinforcement Learning methods, and assess their performance using a variety of widely used domains for planning.

Introduction

Goal recognition is a relevant topic on artificial intelligence research, as it is a common task on many applications with user interaction. Intelligent agents have an implicit ability of inferring the objective of others by simply observing what they are doing, and it presents as a helpful ability in multiagent systems.

Reinforcement Learning is currently a widely used methodology to learn how to behave in environments, learning by interacting with it directly. Another approach that has seen advancements in recent years is Inverse Reinforcement Learning, a class of methods that learns the reward structure of environments, instead of learning how to act directly.

In this work, we propose an application of Reinforcement Learning in Goal Recognition. We intend to develop policies that are capable of discerning which goal an agent is intending to reach, by measuring the distance of each policy to the trajectory the agent is taking. We plan to test different Reinforcement Learning techniques, including standard Reinforcement Learning approaches, and also Inverse Reinforcement Learning to infer the underlying reward structure for each goal. The performance of these methods will be assessed in commonly used planning domains.

Reinforcement Learning

Reinforcement Learning is a framework used to learn robust policies that are capable of maximizing a reward signal. Reinforcement Learning depends upon a Markov Decision Process $\mathcal{M} = \langle S, A, P, R, \gamma \rangle$, where

- S is a finite set of states,
- A is a finite set of actions,

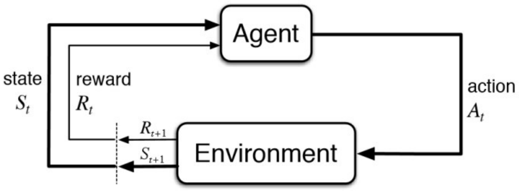

Figure 1: Reinforcement Learning iterative cycle. The agent performs an action A_t at state S_t, and the environment outputs the next state S_{t+1} and reward R_{t+1}.

- P is a transition function,
- R is a reward function,
- $\gamma \in [0, 1]$ is a discount factor.

The most common Reinforcement Learning methodology is based on direct interaction with the environment, where the agent has a simulator of the environment and is capable of performing actions directly on it. As they perform actions in a state, they transition to another one and receive a reward, indicating if their actions were relevant to the goal that R formulates. A policy π is then learned based on this interaction, after a number of timesteps or until convergence, optimizing the expected discounted return. Figure 1 illustrates this iterative cycle.

Reinforcement Learning methods can be implemented using either tabular approaches or by approximating functions. A tabular implementation consists of optimizing a table that indicates the probability of performing an action at a state, or the value (expected return) of performing it. With function approximation, a parameterized function is optimized to approximate the optimal behavior (or value). Tabular methods are expensive for large space states, while function approximators can deal well with them, but rely mainly on gradient descent, and are not guaranteed to converge.

Inverse Reinforcement Learning

A standard Reinforcement Learning problem requires the reward function R to be provided, but there are cases where R is missing, and what we have are only traces of agents interacting in the environment. The Inverse Reinforcement Learning class of methods, instead of learning a policy by interacting with the environment, tries to recover the missing reward function R by analysing past traces of another agent's policy or traces.

These methods are considered to be ill-defined problem (Ng, Harada, and Russell 1999), since there are many optimal policies that can explain a trajectory, and many reward functions that can explain an optimal policy. This was a recurring problem since the problem was first defined, but state-of-the-art methods present a way to learn rewards disentangled from the environment dynamics (Fu, Luo, and Levine 2017).

Goal Recognition

A Goal Recognition problem is a tuple $G_P = \langle \Xi, \mathcal{I}, \mathcal{G}, \mathcal{O} \rangle$, where

- Ξ is the domain definition,
- $\mathcal{I}$ is the initial state of a problem,
- $\mathcal{G}$ is the set of candidate goals, where one of them is an assumed hidden goal g,
- $\mathcal{O}$ is a sequence of observations $\langle o_1, o_2, ..., o_n \rangle$.

A solution to a Goal Recognition problem is the hidden goal $g \in \mathcal{G}$ that is most consistent with observation sequence $\mathcal{O}$.

The sequence of observations can be represented as actions or states. The observation sequences can also be full or partial, meaning that the observation sequence could be missing a percentage of observations. Also, sequences can contain noisy observations, i.e., observations that were not performed or seen.

Current approaches to Goal Recognition rely either on executing a planner for a number of times (Ramírez and Geffner 2009) or landmarks heuristics (Pereira, Oren, and Meneguzzi 2020). From what we know, there are no approaches that use any kind of learning in order to perform goal recognition.

Reinforcement Learning in Goal Recognition

Goal Recognition relates to Inverse Reinforcement Learning in the sense that the sequence of observations $\mathcal{O}$ can be used as the dataset to learn a policy for the hidden goal $g \in G$ that $\mathcal{O}$ intends to reach. In this work, we propose to exploit this relation, studying if it is capable of recognizing goals efficiently.

Although a planning problem does not have an explicit reward function, we can easily see that a reward function could be defined as 1 when reaching the goal state, and 0 otherwise. This is already implemented in the PDDLGym library (Silver and Chitnis 2020), where Reinforcement Learning-like environments are created from PDDL domain definitions. This not only enables the addition of reward functions, but also simplifies the actions an agent can perform by eliminating grounded parameters that are only present in an action for its preconditions, and could be inferred by the current state or by the action itself. The library does not present a negative reward, but could be extended to expose it when dead-ends are found, for example.

The main drawback of this approach is the necessity of training a different policy for each existing goal, and for each problem. As the state space changes when the problem definition changes, the policy also will need to be learned from scratch, thus adding an additional step for each domain. If results are robust enough, this could be a tradeoff for efficiency.

Schedule

As this project is due in 5 weeks, the current plan to develop this work is as follows:

- Week 1 - Develop tabular Reinforcement Learning methods to learn policies in small state space domains
- Week 2 - Start testing Inverse Reinforcement Learning as a solution do Goal Recognition
- Week 3 - Use functions approximators, starting with linear ones
- Week 4 - Expand to larger state space domains
- Week 5 - Compare with other works and report results on a new paper.

Conclusion

We propose to develop a Reinforcement Learning and Inverse Reinforcement Learning approach that is capable of recognizing goals. We do not expected to completely finish the Inverse Reinforcement Learning implementation, as this will be an ongoing research and there are no similar works to what we are trying to accomplish.

References

Fu, J.; Luo, K.; and Levine, S. 2017. Learning robust rewards with adversarial inverse reinforcement learning. *arXiv preprint arXiv:1710.11248*.

Ng, A. Y.; Harada, D.; and Russell, S. J. 1999. Policy invariance under reward transformations: Theory and application to reward shaping. In *Proceedings of the Sixteenth International Conference on Machine Learning*, ICML '99, 278–287. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Pereira, R.; Oren, N.; and Meneguzzi, F. 2020. Landmark-based approaches for goal recognition as planning. *Artificial Intelligence* 279.

Ramírez, M., and Geffner, H. 2009. Plan recognition as planning. In *Proceedings of the 21st International Jont Conference on Artifical Intelligence*, IJCAI'09, 1778–1783. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Silver, T., and Chitnis, R. 2020. Pddlgym: Gym environments from pddl problems. In *International Conference on Automated Planning and Scheduling (ICAPS) PRL Workshop*.