Goal Recognition as Reinforcement Learning

Leonardo Amado[1], Reuth Mirsky, [2,3], Felipe Meneguzzi [4,1]

[1] Pontifícia Universidade Católica do Rio Grande do Sul, Brazil
[2] Bar Ilan University, Israel
[3] The University of Texas at Austin, USA
[4] University of Aberdeen, Scotland

Abstract

Goal recognition approaches often rely on manual specifications of the environmental dynamics an agent needs to overcome to achieve its goals. These specifications suffer from two key issues: they require careful design by a domain expert, and they often need costly real-time computations to compare observations with valid plans for each goal hypotheses. We overcome these limitations by combining model-free reinforcement learning and goal recognition. The resulting framework consists of two main stages: offline learning of policies or utility functions for each potential goal, and online inference. We provide a first instance of this framework using tabular Q-learning for the learning stage, as well as three mechanisms for the inference stage. The resulting instantiation achieves state-of-the-art performance against goal recognizers on standard evaluation domains and superior performance in noisy environments. A full paper describing this work has been published at AAAI 2022.

Introduction

Goal recognition (GR) is a key task in artificial intelligence, where a *recognizer* infers the goal of an *actor* based on a sequence of observations. Consider a service robot that wishes to assist a person in the kitchen by fetching appropriate utensils without interrupting the task execution or demanding attention for specifying instructions (Kautz and Allen 1986; Granada et al. 2020; Bishop et al. 2020). Most GR approaches rely on an arduous process to inform the recognizer about the feasibility and likelihood of the different actions that the actor can execute. This process might include crafting elaborate domain theories, multiple planner executions in real-time, intricate domain optimizations, or a combination of these tasks. Several limitations of this process are:

Cost of Domain Description: Crafted domain theories require deliberate design and accurate specification of domain dynamics, which is usually a process done manually by an expert. In highly complex environments, manual elicitation of such a model might even be impossible.

Noise Susceptibility: As specifying accurate domain dynamics is costly, many specifications are incomplete and cannot inform the recognizer about unlikely observations or partial observation sequences. This property makes many goal recognizers susceptible to noise.

Online Costs: Some recognizers require costly online computations, such as multiple planner executions. These computations can hinder the recognizer's real-time inference ability, especially when observations are processed incrementally and the goal of the actor needs to be re-evaluated many times throughout the plan execution.

We develop a framework to address these limitations by replacing manually crafted representations and online executions with model-free Reinforcement Learning (RL) techniques. This framework performs efficient and noise-resistant GR without the need to craft a domain model and without any planner or parser executions during recognition.

Problem and Approach Overview

We begin by defining a GR problem in a way that is consistent with existing literature (Meneguzzi and Pereira 2021; Mirsky, Keren, and Geib 2021). Given a domain theory $\mathbb{T}$, a set of possible goals $\mathcal{G}$, and a sequence of observations $\boldsymbol{O}$, a goal recognition problem consists of a goal $g \in \mathcal{G}$ that **explains** $\boldsymbol{O}$. This work proposes multiple semantics for **explains**, and we start by defining an RL-based domain theory.

A **policy** $\pi(a \mid s)$ for an MDP is a function that defines the probability of the agent taking action $a \in \mathcal{A}$ in state $s \in \mathcal{S}$. Some RL algorithms, such as Q-learning, compute the policy of an agent using a **Q-function** $Q(s, a)$, which is an estimation of the expected return starting from s after taking action a. In our new framework, a domain theory $\mathbb{T}$ consists of the state and action spaces and the transition function p of an MDP, but the reward is replaced with a set of policies or Q-functions. Unlike planning-based GR where the domain theory is decoupled from the problem instance (the set of possible goals $\mathcal{G}$), here $\mathbb{T}$ depends on the set of goals. We define two types of domain theories:

Definition 1 (Utility-based Domain Theory) *A utility-based domain theory $\mathbb{T}_Q(\mathcal{G})$ is a tuple $(\mathcal{S}, \mathcal{A}, p, \mathcal{Q})$ such that $\mathcal{Q}$ is a set of Q-functions $\{Q_g\}_{g \in \mathcal{G}}$.*

Definition 2 (Policy-based Domain Theory) *A policy-based domain theory $\mathbb{T}_\pi(\mathcal{G})$ is a tuple $(\mathcal{S}, \mathcal{A}, p, \Pi)$ such that Π is a set of policies $\{\pi_g\}_{g \in \mathcal{G}}$.*

Both domain theories consist of a *set* of MDPs with the same transitions, but have different reward functions for different goals, as implicitly dictated from the Q-functions or

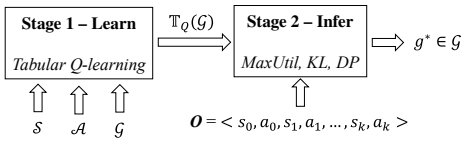

Figure 1: The Goal Recognition as RL framework.

the policies. Our aim is to learn either a good policy or a utility function that represents the expected behavior of actors under each of these MDPs. We use this formulation to provide a new definition for a goal recognition problem in which we replace the abstract notion of $\mathbb{T}$ and combine the goal set $\mathcal{G}$ into these domain theories.

Definition 3 (Goal Recognition Problem) *Given domain theory $\mathbb{T}_Q(\mathcal{G})$ or $\mathbb{T}_\pi(\mathcal{G})$ and a sequence of observations $\boldsymbol{O}$, output a goal $g \in \mathcal{G}$ that **explains** $\boldsymbol{O}$.*

Using this new problem definition, we develop our framework to solve these goal recognition problems, discuss how to learn $\mathbb{T}_Q(\mathcal{G})$ or $\mathbb{T}_\pi(\mathcal{G})$, and how to decide which goal g best **explains** observations $\boldsymbol{O}$.

Our new framework consists of two main stages: (1) learning a set of Q-functions; and (2) inferring the goal of an actor given a sequence of observations. Figure 1 illustrates this process. First, the initial inputs are state and action spaces, $\mathcal{S}$ $\mathcal{A}$, p, and a set of goals $\mathcal{G}$.

The inferred goal g^* is the one that minimizes the measured distance between its respective Q-function and the observations, as defined in Equation 1.

$$g^* = \arg\min_{g \in \mathcal{G}} \textsc{Distance}(Q_g, \boldsymbol{O}) \quad (1)$$

Amado, Mirsky, and Meneguzzi (2022) detail the specific distance measures in the Goal Recognition as Q-Learning (GRAQL) framework, but, in summary, it uses:

- MaxUtil: which consists of accumulating the utilities over the states/actions in the observations for each of the Q-value functions Q_g for $g \in \mathcal{G}$
- KL-divergence: which consists of summing the divergence between the probability distributions encoded in softmax policies π_g generated for each goal hypothesis $g \in \mathcal{G}$ and a pseudo-policy encoded for the observations; and
- Divergence Point: which consists of computing the point in which the softmax policies above diverge from the same pseudo-policy.

Figure 2 summarizes experimentation of our approaches against the approach from Ramírez and Geffner (2010) under full observability. The experimental domains use PDDLGym (Silver and Chitnis 2020) as the evaluation environment. PDDLGym is a python framework that automatically constructs OpenAI Gym environments from PDDL domains and problems. Thus, for each PDDL domain used by state-of-the-art GR algorithms, we generate the parallel representation in Gym for GRAQL. We use three domains from the PDDLGym library for their similarity with commonly used GR evaluation domains: Blocks, Hanoi, and SkGrid

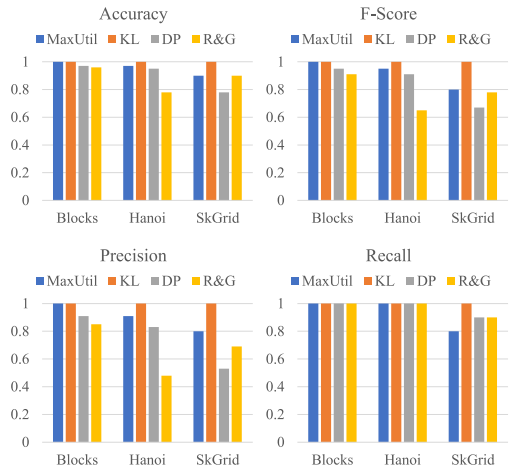

Figure 2: Comparison of R&G, MaxUtil, KL, DP by their accuracy, precision, recall, and F-score for full observability.

(The latter highly resembles common GR navigation domains such as those used by Masters and Sardina (2019)). These results show that GRAQL is able to achieve comparable results to the state-of-the-art with fully observable trajectories. Further experiments can be found in Amado, Mirsky, and Meneguzzi (2022), showing that GRAQL achieves superior performance in noisy environments.

Discussion and Conclusion

Our framework uses learned Q-values or policies, implicitly representing the agent's perceived reward under observation in lieu of explicit goals from traditional GR. This approach allows us to solve GR problems by minimizing the distance between an observation sequence and Q-values representing goal hypotheses or policies extracted from them. Our distance measures are competitive with the reference approach from the literature (Ramírez and Geffner 2009) in all experimental environments, and some distance measures outperform the reference approach in most domains, especially when the observation sequence is noisy or partial. This work paves the way for a new class of GR approaches based on model-free reinforcement learning. Future work will focus on new, more robust distance measures and mechanisms to handle noise explicitly, as well as experimenting with models learned using function approximation (e.g., neural networks).

References

Amado, L. R.; Mirsky, R.; and Meneguzzi, F. 2022. Goal Recognition as Reinforcement Learning. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI)*. AAAI Press.

Bishop, J.; Burgess, J.; Ramos, C.; Driggs, J. B.; Williams, T.; Tossell, C. C.; Phillips, E.; Shaw, T. H.; and de Visser, E. J. 2020. CHAOPT: a testbed for evaluating human-autonomy team collaboration using the video game over-

cooked! 2. In *2020 Systems and Information Engineering Design Symposium (SIEDS)*, 1–6. IEEE.

Granada, R.; Monteiro, J.; Gavenski, N.; and Meneguzzi, F. 2020. Object-Based Goal Recognition Using Real-World Data. In *Mexican International Conference on Artificial Intelligence*, 325–337. Springer.

Kautz, H. A.; and Allen, J. F. 1986. Generalized plan recognition. In *Conference on Artificial Intelligence (AAAI)*, volume 86, 5.

Masters, P.; and Sardina, S. 2019. Cost-based goal recognition in navigational domains. *Journal of Artificial Intelligence Research*, 64: 197–242.

Meneguzzi, F.; and Pereira, R. F. 2021. A Survey on Goal Recognition as Planning. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 4524–4532.

Mirsky, R.; Keren, S.; and Geib, C. 2021. Introduction to Symbolic Plan and Goal Recognition. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 16(1): 1–190.

Ramírez, M.; and Geffner, H. 2009. Plan recognition as planning. In *Twenty-First International Joint Conference on Artificial Intelligence (IJCAI)*.

Ramírez, M.; and Geffner, H. 2010. Probabilistic plan recognition using off-the-shelf classical planners. In *AAAI Conference on Artificial Intelligence*.

Silver, T.; and Chitnis, R. 2020. PDDLGym: Gym Environments from PDDL Problems. *CoRR*, abs/2002.06432.