Distributional Metareasoning for Heuristic Search

Tianyi Gu
University of New Hampshire
gu@cs.unh.edu

Abstract

Heuristic search methods are widely used in many real-world autonomous systems. Yet, people always want to solve search problems that are larger than time allows. To address these challenging problems, even suboptimally, a planning agent should be smart enough to intelligently allocate its computational resources, to think carefully about where in the state space it should spend time searching. For finding optimal solutions, we must examine every node that is not provably too expensive. In contrast, to find suboptimal solutions when under time pressure, we need to be very selective about which nodes to examine. In this work, we will demonstrate that estimates of uncertainty, represented as belief distributions, can be used to drive search effectively. This type of algorithmic approach is known as metareasoning, which refers to reasoning about which reasoning to do. We will provide examples of improved algorithms for real-time search, bounded-cost search, and situated planning.

1 Introduction

Heuristic search methods are widely used in many real-world autonomous systems. However, people always want to solve search problems that are larger than time allows. To solve the challenging problems, even suboptimally, a planning agent better be smart enough to intelligently allocate its computational resources, to think carefully about where in the state space it should spend time to search. In contrast to finding optimal solutions, in which we must examine every node that is not provably too expensive, we need to be very selective about which nodes to examine when under time pressure. This type of algorithmic approach was named metareasoning, which refers to reasoning about what to reason about. An agent equipped with a metareasoning component would solve a meta-level-reasoning problem in addition to the conventional object-level-reasoning problem [Horvitz, 1990]. These two problems typically differ in their utility functions. The meta-level utility is the expected utility associated with inference-related cost (i.e, deliberation cost), while the object-level utility is the expected utility associated with the state of the world (i.e, solution cost) without regard to the cost of reasoning. The relation of different types of cost is illustrated in Figure 1. In my dissertation, I propose practical metareasoning methods that would carefully allocate effort due to time pressure and thus optimize combined utility that takes into account both reasoning cost and object-level cost.

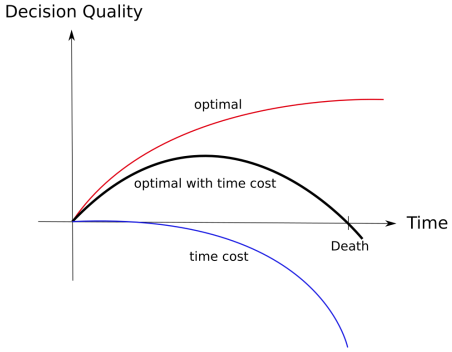

Figure 1: Illustration of optimal action selection with time cost.

Historically, the metareasoning problem has been discussed for a long time since it was proposed [Good, 1971; Russell and Wefald, 1991; Zilberstein, 2008]. However, there only have been a few practical search algorithms that actually do metareasoning. It ought to be beneficial to design such an agent that explicitly optimizes both object-level utility and meta-level utility. For an agent to plan under time pressure, obviously, it should not only optimize the solution cost but also the planning time to achieve the best performance. Therefore, I pursue practical metareasoning components that can enhance various families of traditional search algorithms.

Conventional metareasoning approaches follow decision theory that tells the agent to select the action that maximizes the expected utility. However, when modeling an uncertain value, scalar expected value often is not as powerful as a belief distribution. Recently, distributional methods have been proposed in the RL community [Bellemare *et al.*, 2017; Dabney *et al.*, 2020], and it has been shown that distributional-informed methods can often outperform scalar-expected-value-informed methods by taking advantage of reasoning on value uncertainty. Therefore, taking inspiration from this prior work from RL, my work pursues an alternative class of rational metareasoning that takes advantage of distributional methods to have a better estimate model instead of only relying on the expected utility. The distributions can be constructed through offline learning or online learning. In my dissertation, I aim to improve algorithms for the following three problem settings: (1) real-time search (2) bounded-cost search, and (3) situated planning.

2 Overview of Dissertation

Firstly, I will address real-time decision-making. In this setting, when the time bound is reached, the agent has to commit to the best action on hand even if it only has a partial solution plan. Because the time bound tightly limits the computation that an agent can perform, metareasoning could play an important role in this setting. Traditional real-time search methods were adapted directly from off-line search methods like A* and make online action selection decisions based on a lower bound rather than an expected cost, which is not appropriate as a basis for rational action selection. To do a rational real-time search, it might be worth it for the agent to gather information about the value uncertainty due to the bounded rationality and make online decisions based on the value uncertainty as well as the expected utility. The traditional real-time search approaches are lacking this kind of metareasoning component. The first part of my dissertation makes contributions to designing a rational real-time search approach [Fickert *et al.*, 2020b; Fickert *et al.*, 2020a].

Secondly, I propose algorithms for bounded-cost search settings where the agent is given a specific cost bound along with the search problem. The goal is to find a complete solution within the cost bound as quickly as possible. Bounded-cost search is also very useful since its users can have control over the solution quality. Traditional methods for bounded-cost search are focused on designing inadmissible heuristics that could guide the search toward the search nodes that have a high chance for finding a solution within the bound [Stern *et al.*, 2011]. BEES [Thayer *et al.*, 2012] explicitly tries to find a solution within the bound as quickly as possible, which is a meta-level problem. However, the performance of BEES can be very sensitive to the error of its estimate. In the second part of my dissertation, I propose a distributional method to not only explicitly optimize the time to find a solution within bound but also take advantage of knowing the uncertainty of the estimate and thus better guide the search. (This work is currently under review.)

Thirdly, I also propose a metareasoning algorithm for online planning, specifically, answering the question of when to commit an action. When the planner commits to an action, it re-roots its search at the node representing the outcome of that action. We assume that the system cannot be uncontrolled, so the planner must commit to a new action (perhaps a no-op) before the previously chosen action completes. In this setting, it can be beneficial to commit early, in order to devote more lookahead search focused below an upcoming state. In the third part of my dissertation, we propose a principled method for making this commitment decision. (This work is still in progress.)

References

[Bellemare *et al.*, 2017] Marc G Bellemare, Will Dabney, Munos, and Remi. A distributional perspective on reinforcement learning. In *the International Conference on Machine Learning*, 2017.

[Dabney *et al.*, 2020] Will Dabney, Zeb Kurth-Nelson, Naoshige Uchida, Clara Kwon Starkweather, Demis Hassabis, Rémi Munos, and Matthew Botvinick. A distributional code for value in dopamine-based reinforcement learning. *Nature*, pages 671–675, 2020.

[Fickert *et al.*, 2020a] Maximilian Fickert, Tianyi Gu, Leonhard Staut, Sai Lekyang, Wheeler Ruml, Jörg Hoffmann, and Marek Petrik. Real-time planning as data-driven decision-making. In *the ICAPS-20 Workshop on Bridging the Gap Between AI Planning and Reinforcement Learning (PRL-20)*, 2020.

[Fickert *et al.*, 2020b] Maximilian Fickert, Tianyi Gu, Leonhard Staut, Wheeler Ruml, Jörg Hoffmann, and Marek Petrik. Beliefs we can believe in: Replacing assumptions with data in real-time search. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 9827–9834, 2020.

[Good, 1971] I J Good. Twenty-seven principles of rationality. In *Godambe, V.P. & Sprott, D.A. (Eds), Foundations of Statistical Inference*, pages 108–141, 1971.

[Horvitz, 1990] Eric J Horvitz. *Rational metareasoning and compilation for optimizing decisions under bounded resources*. Knowledge Systems Laboratory, Medical Computer Science, Stanford University, 1990.

[Russell and Wefald, 1991] Stuart Jonathan Russell and Eric Wefald. *Do the right thing: studies in limited rationality*. MIT Press, 1991.

[Stern *et al.*, 2011] Roni Tzvi Stern, Rami Puzis, and Ariel Felner. Potential search: A bounded-cost search algorithm. In *Twenty-First International Conference on Automated Planning and Scheduling*, 2011.

[Thayer *et al.*, 2012] Jordan Tyler Thayer, Roni Stern, Ariel Felner, and Wheeler Ruml. Faster bounded-cost search using inadmissible estimates. In *Twenty-Second International Conference on Automated Planning and Scheduling*, 2012.

[Zilberstein, 2008] Shlomo Zilberstein. Metareasoning and bounded rationality. In *AAAI Workshop on Metareasoning: Thinking about Thinking*, Chicago, Illinois, 2008.