Planning Under Uncertainty and Temporally Extended Goals

Alberto Camacho *
Department of Computer Science
University of Toronto. Canada.
acamacho@cs.toronto.edu

1 Introduction

In the last decade, we have seen an exponential increase in the number of devices connected to the Internet, with a commensurate explosion in the availability of data. New applications such as those related to smart cities exemplify the need for principled techniques for automated intelligent decision making based on available data. Many decision-making problems require reasoning in large and complex state spaces, sometimes under stringent time constraints. The nature of these problems suggests that planning approaches could be used to find solutions efficiently. Automated planning is the basis for addressing a diversity of problems beyond classical planning such as automated diagnosis, controller synthesis, and story understanding. Nevertheless, many planning paradigms make assumptions that do not hold in real-world settings.

Our work focuses on exploring planning paradigms that capture properties of real-world decision-making applications. These properties include the ability to model nondeterminism in the outcome of actions, the ability to deal with complex objectives that are temporally extended (in contrast to final-state goals) some of which may be necessary and other simply desirable to optimize for. Finally, we are interested in dealing with incomplete information. Addressing this class of problems presents challenges related to problem specification, modeling, and computationally efficient techniques for generating solutions.

Illustrative Example Consider the problem of designing a tourist route to visit a set of touristic attractions in London. The tour is subject to certain constraints. For example, an individual may feel it's mandatory for the tour to include the London Eye and the Houses of Parliament, and also desirable to visit the Maritime Museum and the Greenwich Observatory or other highly-rated attractions, if these can be included. The tour must be realizable via a combination of walking and public transit. If it is raining, then walking should be minimized. These are examples of temporally extended goals. Following from our example, aspects of the dynamics of the environment are not controllable by the agent, such as traffic, punctuality of public transport, and the weather. If the stochastic model for these events is available, we can quantify the *expected quality* of the plan according to a certain metric (e.g. probability of visiting the noted touristic attractions at the end of the journey) and attempt to produce plans that maximize this objective. When the stochastic model is not available, we may want to produce plans that are robust to *any* contingency (e.g. a plan that suggests visiting a museum, at any moment, if it starts to rain).

2 Progress to the Date

In our work to date, we have advanced the state of the art in planning problems with non-deterministic actions and temporally extended goals. In this section, we introduce the FOND and probabilistic planning models, and describe the high-level contributions of our work. We refer the reader to the respective publications for further details.

A *Fully Observable Non-Deterministic* (FOND) planning problem is a tuple $\mathcal{P} = \langle S, s_I, \mathcal{A}, F, S_{\mathcal{G}} \rangle$, where S is a finite set of *states*, $s_I \in S$ is the *initial state*, $S_{\mathcal{G}} \subseteq S$ is a set of *goal states*, and $\mathcal{A}$ is a finite set of actions. For each action $a \in \mathcal{A}$, and state $s \in S$, the result of applying a in s is one of the states in the set $F(s, a) \subseteq S$. Solutions to FOND planning problems are *policies*, or mappings $\pi : S \rightarrow \mathcal{A}$ from states into actions. In concrete, *strong-cyclic* solutions are those that lead the agent to a goal state with complete guarantees [Cimatti *et al.*, 2003].

A *probabilistic* planning problem is a tuple $\mathcal{P} = \langle S, s_I, \mathcal{A}, T, S_{\mathcal{G}} \rangle$. Different than the FOND model, for each action $a \in \mathcal{A}$, and pair of states $s, s' \in S$, $T(s, a, s')$ is the *transition probability* of reaching s' when a is applied in s. Solutions to probabilistic planning problems are policies. In goal-oriented probabilistic planning models such as MaxProb, solutions are policies that lead the agent to a goal state with maximal probability.

2.1 ProbPRP

In [Camacho *et al.*, 2016a] we present ProbPRP, a probabilistic planner that finds solutions to probabilistic planning problems where the objective is to attempt to maximize the probability of reaching a goal state. We formalize this class of problems and call it HighProb.

ProbPRP has two important merits. First, it overcomes scaling difficulties that previous offline algorithms experi-

*The contributions presented in this paper are joint work with (in alphabetical order) Jorge Baier (jabaier@ing.puc.cl), Sheila McIlraith (sheila@cs.toronto.edu) (supervisor), Christian Muise (cjmuise@mit.edu), and Eleni Triantafillou (eleni@cs.toronto.edu).

enced. And second, it offers superior optimality guarantees with respect to the previous state of the art in HighProb, the online planner RFF [Teichteil-Königsbuch *et al.*, 2010]. Despite being an offline algorithm, ProbPRP outperforms RFF in general and solutions are of better quality.

ProbPRP leverages core similarities between probabilistic and FOND planning, making use of state-of-the-art FOND planning techniques from PRP [Muise *et al.*, 2012] in its underlying algorithm. The *partial state* representation obtained via plan regression facilitates states entailment during the search process, and results in considerable improvements in the algorithm convergence. Besides, the compact representation of state results in smaller policies. The *deadend detection* mechanism prunes the search space effectively by means of forbidden state-action pairs (FSAPs) generated automatically during the search process, and guarantees optimality of the algorithm when deadends are avoidable. ProbPRP extends the state-of-the-art FOND planner PRP [Muise *et al.*, 2012] with techniques that leverage probabilistic information to produce high quality solutions. Some of these enhancements to ProbPRP include the bias towards exploration of *high-likelihood plans*, and the *final FSAP-free round*. The search bias produces policies that have smaller expected plan length – orders of magnitude smaller in some instances. A final search round is performed to extend the best incumbent policy found by the algorithm, this time with the FSAP mechanism disabled. We observed the final FSAP-free round increments the probability of reaching a goal state up to 30%.

2.2 LTL FOND Translations

In [Camacho *et al.*, 2016b] we address the problem of planning with non-deterministic actions and temporally extended goals. We assume goals are specified as LTL formulas [Pnueli, 1977], and call the model LTL FOND. LTL formulae can be interpreted over finite or infinite state trajectories. Solutions to different interpretations are not always equivalent. A number of techniques exist to solve planning with LTL goals, a subset in the presence of non-deterministic actions, and with finite and infinite LTL interpretations. A common approach is to compile the problem into one with a final-state goal, and solve the resulting problem with state-of-the-art planning technology (e.g. [Baier and McIlraith, 2006; Patrizi *et al.*, 2013; Torres and Baier, 2015]). Related work attempts to maximize reward in MDPs with finite LTL goals and preferences (e.g. [Lacerda *et al.*, 2015]), and in decision processes with non-markovian rewards (e.g. [Thiébaux *et al.*, 2006]).

We present two different techniques for compiling LTL FOND into FOND, each addressing both the case of finite LTL interpretations, and the case of infinite LTL interpretations. Remarkably, we are the first to solve the full spectrum of LTL FOND planning interpreted on inifite state trajectories. Equipped with strong-cyclic planner, PRP, our system proves competitive with other state-of-the-art algorithms for LTL FOND, with the advantage of being able to solve the full spectrum of LTL FOND problems.

Our translations leverage ideas from [Baier and McIlraith, 2006; Torres and Baier, 2015; Patrizi *et al.*, 2013], and use Non-deterministic Finite Automata (NFA) and Alternating Automata (AA) representations of the LTL formula to monitor progression, and strong-cyclic planning to synthesize solutions. The size of NFA-based translations is worst-case exponential in the size of the formula, and the size of AA-based translations is worst-case polynomial. Interestingly, PRP performance was better with NFA-based translations, with smaller policies and lower run-times than with AA-based translations.

3 Discussion and Future Work

The techniques we are developing are applicable to a diversity of real-world problems from the control of collections of smart-home devices, to applications in transportation planning and industrial process planning. A natural next step is to extend our recent work to address the class of probabilistic planning problems with LTL goals which we believe can be done via our existing translations and ProbPRP. We are also interested in extending our work to capture LTL preferences and rewards. Finally, we plan to explore extensions to our models to include both propositional and real-valued variables since such hybrid models are prevalent in many of the real-world applications we've encountered.

References

[Baier and McIlraith, 2006] J. Baier and S. McIlraith. Planning with first-order temporally extended goals using heuristic search. In *AAAI*, pages 788–795, 2006.

[Camacho *et al.*, 2016a] Alberto Camacho, Christian Muise, and Sheila A. McIlraith. From fond to robust probabilistic planning: Computing compact policies that bypass avoidable deadends. In *ICAPS*, 2016.

[Camacho *et al.*, 2016b] Alberto Camacho, Eleni Triantafilou, Jorge A. Baier, Christian Muise, and Sheila A. McIlraith. LTL Synthesis for Non-Deterministic Systems on Finite and Infinite Traces. In *HSDIP*, 2016.

[Cimatti *et al.*, 2003] Alessandro Cimatti, Marco Pistore, Marco Roveri, and Paulo Traverso. Weak, strong, and strong cyclic planning via symbolic model checking. *AIJ*, 147:35–84, 2003.

[Lacerda *et al.*, 2015] Bruno Lacerda, David Parker, and Nick Hawes. Optimal Policy Generation for Partially Satisfiable Co-Safe LTL Specifications. *IJCAI*, pages 1587–1593, 2015.

[Muise *et al.*, 2012] Christian Muise, Sheila A. McIlraith, and J. Christopher Beck. Improved Non-deterministic Planning by Exploiting State Relevance. In *ICAPS*, pages 172–180, 2012.

[Patrizi *et al.*, 2013] Fabio Patrizi, Nir Lipovetzky, and Hector Geffner. Fair LTL synthesis for non-deterministic systems using strong cyclic planners. In *IJCAI*, pages 2343–2349, 2013.

[Pnueli, 1977] Amir Pnueli. The temporal logic of programs. In *FOCS*, pages 46–57, 1977.

[Teichteil-Königsbuch *et al.*, 2010] Florent Teichteil-Königsbuch, Ugur Kuter, and Guillaume Infantes. Incremental plan aggregation for generating policies in MDPs. In *AAMAS*, volume 1, pages 1231–1238, 2010.

[Thiébaux *et al.*, 2006] Sylvie Thiébaux, Charles Gretton, John K Slaney, David Price, Froduald Kabanza, et al. Decision-theoretic planning with non-markovian rewards. *JAIR*, 25:17–74, 2006.

[Torres and Baier, 2015] Jorge Torres and Jorge A. Baier. Polynomial-time reformulations of LTL temporally extended goals into final-state goals. In *IJCAI*, pages 1696–1703, 2015.