Resolving Multiple-Dynamic Model Uncertainty in Hypothesis-Driven Belief-MDPs

Extended Abstract

Ofer Dagan
Aerospace Engineering Sciences,
University of Colorado, Boulder
Boulder, United States
ofer.dagan@colorado.edu

Tyler Becker
Aerospace Engineering Sciences,
University of Colorado, Boulder
Boulder, United States
tyler.becker-1@colorado.edu

Zachary N. Sunberg
Aerospace Engineering Sciences,
University of Colorado, Boulder
Boulder, United States
zachary.sunberg@colorado.edu

ABSTRACT

This work explores the set of planning problems we define as hypothesis-driven planning. In these problems, a human with potentially no access to the system's underlying planning model (e.g., the reward or transition functions) is interested in answering a set of questions (hypotheses) that are not reflected in the underlying planning problem. To reason about the possible hypotheses and autonomously decide which one is most likely, we develop a new multiple-dynamic hypothesis belief Markov decision process (MDH-BMDP). This enables adding multiple hypotheses for existing planning problems and reasoning over arbitrary belief shapes using existing sparse tree search solvers. Lastly, we suggest a reward function that supports balancing the objectives of determining the correct hypothesis in time and performing well in the underlying problem; and test the framework's applicability and performance of the new reward function in simulations.

KEYWORDS

Hypothesis; Multiple-dynamic model; POMDP; Belief-MDP

ACM Reference Format:
Ofer Dagan, Tyler Becker, and Zachary N. Sunberg. 2025. Resolving Multiple-Dynamic Model Uncertainty in Hypothesis-Driven Belief-MDPs: Extended Abstract. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 3 pages.

1 INTRODUCTION

Consider a human operator that detects an unexpected behavior while monitoring a cyber-physical system, e.g., a drone performing some inspection task while suddenly suffering from a possible rotor failure [10], or a space-object tracking system for space domain awareness (SDA) that observes some anomaly in the orbit of an object [7]. The operator, with potentially no access to the planning algorithm or model of the system, is interested in answering a set of questions (hypotheses) that are not reflected in the underlying problem definition (Figure 1). For example, is a space object currently making a surprise maneuver, or is it malfunctioning?

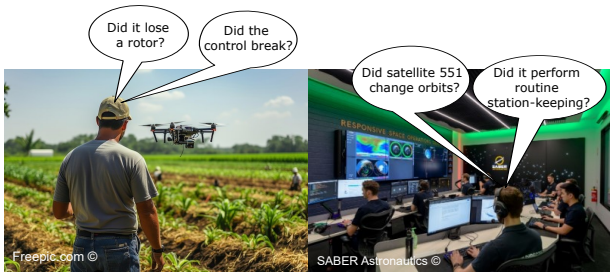

Figure 1: Examples of possible questions/hypotheses when monitoring cyber-physical systems

In some cases, taking information-gathering actions such as additional measurements or control inputs given to the system can help resolve uncertainty and determine the most accurate hypothesis. To resolve the uncertainty over the correct hypothesis, the original planning problem is augmented to form a new planning problem, with objectives that often compete with the objectives of the base problem. We define the task of optimizing actions that can help resolve uncertainty and determine the most probable hypothesis as *hypothesis-driven planning*.

Unfortunately, this problem suffers from the 'curse of history', similar to a partially observable Markov decision process (POMDP). To plan in continuous domains, an agent must reason over countlessly many possible action-observation histories, each resulting in a different belief over the unknown state. The problem is exacerbated in the hypothesis-driven context since each action-observation pair spawns a different belief for each hypothesis. This research explores the hypothesis-driven planning problem to find a formulation that enables reasoning over multiple hypotheses while allowing tractable solutions using existing sparse tree search algorithms. We focus on hypotheses that are spawned due to different dynamic models in an underlying POMDP and seek a reward function that balances the goals of determining the (most likely) correct hypothesis and performing well in the underlying POMDP.

2 HYPOTHESIS-DRIVEN PLANNING

Many planning problems can be framed as sequential decision-making problems and modeled as partially observable Markov decision processes (POMDPs). A POMDP is defined by the tuple $(\mathcal{S}, \mathcal{A}, T, O, Z, \mathcal{R}, \gamma)$, where $\mathcal{S}$, $\mathcal{A}$ are the sets of all possible states and actions, respectively. $T(s, a, s') = p(s'|s, a)$ is a stochastic state transition model, which defines the probability of transitioning to state $s' \in \mathcal{S}$ from state $s \in \mathcal{S}$ after taking action $a \in \mathcal{A}$, O is

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

the set of all possible observations, and $Z(s', a, o) = p(o|s', a)$ is the stochastic observation function. Finally, the reward function $\mathcal{R}(s, a)$ determines the immediate reward the agent receives when taking action a at state s, and $\gamma \in [0, 1)$ is a discount factor.

As the true state s of a POMDP is unknown, an agent maintains a *belief* over states $s \in \mathcal{S}$, which summarizes the history $h_t = (b_0, a_0, o_1, a_1, ..., o_t)$ of all actions taken and observations received up to and including time step t and starting from initial belief b_0. During planning, the belief is updated using Bayes' rule

$$b'_t(s') = p(s_t = s'|h_t) \propto Z(s', a_t, o_t) \cdot \int_{s \in \mathcal{S}} b_{t-1}(s) \cdot T(s, a_t, s')ds.$$

This results in a different belief $b'_t(s') = p(s'|a_t, o_t)$ for each possible action–observation pair, which makes the POMDP problem intractable in continuous domains.

For multi-hypothesis problems, with uncertainty about the model itself, each action–observation pair generates multiple beliefs that must be reasoned over. These might stem from uncertainty in (i) which object generated a measurement signal, also known as the data association problem [2], [5], or (ii) which transition model generated the next state. The latter has been considered under three different formulations: planning with hybrid dynamics [3], [6], Robust POMDPs [11], [12], and Bayes-adaptive POMDPs [13], [9]. The common denominator for the multi-hypothesis planning literature is the implicit assumption that the planning and inference objectives coincide. The task of planning for the base POMDP thus aligns with the task of reasoning over multiple possible models. For the set of hypothesis-driven problems considered in this work, this means that the task of determining which hypothesis is most probable does not affect the planning optimization process. Nevertheless, determining the correct hypothesis can be crucial for explaining a surprising behavior or understanding certain outcomes, thus requires a new formulation and treatment of the problem.

3 PROBLEM STATEMENT AND APPROACH

Consider a 'base' partially observable Markovian planning problem $\mathcal{P}$, that can be framed as a POMDP. Denote the system's state space as $\mathcal{S}_x$ and the corresponding belief space as $\mathcal{B}_x$. Assume that there is uncertainty regarding the transition model that derives the base state space, that is, there are $n_{\mathcal{H}}$ distinct possible transition models T_i (with no transitions between them), each corresponding to hypothesis $\mathcal{H}_i$. We define the set of questions of whether model i is correct as $\mathcal{H}$. Then the belief over all possible hypotheses states $\mathcal{S}_{\mathcal{H}}$ is represented by $\mathcal{B}_{\mathcal{H}}$. The hypothesis-driven POMDP problem $\bar{\mathcal{P}}$ considered in this work searches for an optimal policy π^* that maximizes the expected cumulative reward over two potentially competing requirements – deciding which hypothesis $\mathcal{H}_i$, with transition model T_i, is (most likely) correct, and still performing well in the original underlying problem $\mathcal{P}$.

To solve the hypothesis-driven POMDP problem $\bar{\mathcal{P}}$ we choose a belief-MDP formulation. Any POMDP is a belief-MDP, with a set of belief states $b \in \mathcal{B}$, a belief transition model $\mathcal{T}(b, a, b')$ [8] and a belief-dependent reward function [1], $\rho(b, a) = \int_{s \in \mathcal{S}} b(s)\mathcal{R}(s, a)ds + \varrho(b, a)$. Where the first term is the expected reward when the state s is distributed according to b, and $\varrho(b, a)$ is a belief-dependent reward such as Shannon's entropy or Kullback-Leibler divergence. A belief-dependent reward is advantageous for hypothesis-driven planning, as it naturally enables reasoning about uncertainty over possible hypotheses.

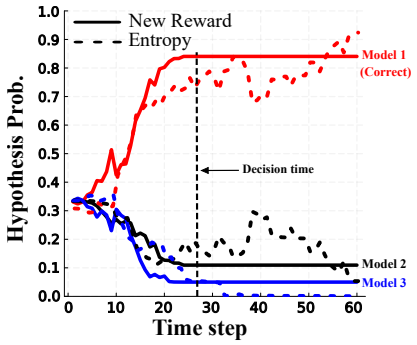

Figure 2: Simulation results showing the effect of different reward functions

Formally, we define the new *multiple-dynamics hypothesis belief MDP* (MDH-BMDP) with the tuple, $(\bar{\mathcal{B}} = \mathcal{B}_x \times \mathcal{B}_{\mathcal{H}}, \mathcal{A}, \bar{\mathcal{T}}, \bar{\rho}, \gamma)$, where $\bar{\mathcal{B}}$ is the augmented belief state, which includes the base belief state $\mathcal{B}_x$ and the hypothesis state $\mathcal{B}_{\mathcal{H}}$. $\mathcal{A}$ and γ are the same as in the base problem $\mathcal{P}$, $\bar{\mathcal{T}}$ is the set of transition models including all belief transition models $\mathcal{T}_i$ for $i = 1 : n_{\mathcal{H}}$ for all dynamics hypotheses. Lastly $\bar{\rho}$ is the new reward function, which is defined as a convex combination of the underlying problem reward, ρ_x and the hypothesis-based reward $\rho_{\mathcal{H}}$,

$$\bar{\rho} = (1 - w) \cdot \rho_x(b_x, a) + w \cdot \varrho_{\mathcal{H}}(b_{\mathcal{H}}, a),$$

where $w \in [0, 1]$ is a weight parameter to prioritize the hypothesis task versus following the base problem behavior.

4 CONTRIBUTION SUMMARY

This work defines the *hypothesis-driven* POMDP problem as the set of multi-hypothesis POMDP problems where explicitly determining the most accurate hypothesis is one of the optimization objectives. These problems naturally arise in human-robot collaboration, when the human detects a surprising behavior or unexpected outcome and wants to explore it. In that case, we want to reason over different possible models, that might justify the observed behavior and provide an alternative explanation. We formulate the problem as belief-MDP, dubbed MDH-BMDP, and solve it using existing sparse tree search algorithms.

To motivate actions that can help resolve uncertainty and determine the most probable hypothesis while still performing well in the underlying POMDP problem, we explore new reward functions, explicitly rewarding in-time decisions. Simulation results demonstrate the advantage of the new reward functions over entropy-based reward, balancing between timely hypothesis decisions and the underlying problem objectives (Figure 2). More details on this work, such as the effect of the reward weight w on the quality of the solution, and the connection between the base reward and the hypothesis belief reward can be found in the arXiv version [4].

ACKNOWLEDGMENTS

This work is supported by the Air Force Office of Scientific Research (AFOSR), Grant number FA9550-23-1-0726.

REFERENCES

[1] Mauricio Araya, Olivier Buffet, Vincent Thomas, and Françcois Charpillet. 2010. A POMDP Extension with Belief-dependent Rewards. In *Advances in Neural Information Processing Systems*, Vol. 23. Curran Associates, Inc. https://papers.nips.cc/paper_files/paper/2010/hash/68053af2923e00204c3ca7c6a3150cf7-Abstract.html

[2] Moran Barenboim, Moshe Shienman, and Vadim Indelman. 2023. Monte Carlo Planning in Hybrid Belief POMDPs. *IEEE Robotics and Automation Letters* 8, 8 (Aug. 2023), 4410–4417. https://doi.org/10.1109/LRA.2023.3282773 Conference Name: IEEE Robotics and Automation Letters.

[3] Emma Brunskill, Leslie Kaelbling, Tomas Lozano-Perez, and Nicholas Roy. 2008. Continuous-State POMDPs with Hybrid Dynamics. In *ISAIM*.

[4] Ofer Dagan, Tyler Becker, and Zachary N. Sunberg. 2024. Resolving Multiple-Dynamic Model Uncertainty in Hypothesis-Driven Belief-MDPs. https://doi.org/10.48550/arXiv.2411.14404

[5] Ming Hsiao, Joshua G. Mangelson, Sudharshan Suresh, Christian Debrunner, and Michael Kaess. 2020. ARAS: Ambiguity-aware Robust Active SLAM based on Multi-hypothesis State and Map Estimations. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Las Vegas, NV, USA, 5037–5044. https://doi.org/10.1109/IROS45743.2020.9341384

[6] Ajinkya Jain and Scott Niekum. 2017. Belief Space Planning under Approximate Hybrid Dynamics. In *Robotics: Science and Systems (RSS) Workshop on POMDPs in Robotics*.

[7] A. D. Jaunzemis, M. J. Holzinger, M. W. Chan, and P. P. Shenoy. 2019. Evidence gathering for hypothesis resolution using judicial evidential reasoning. *Information Fusion* 49 (Sept. 2019), 26–45. https://doi.org/10.1016/j.inffus.2018.09.010

[8] Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra. 1998. Planning and acting in partially observable stochastic domains. *Artificial Intelligence* 101, 1 (May 1998), 99–134. https://doi.org/10.1016/S0004-3702(98)00023-X

[9] Sammie Katt, Frans A. Oliehoek, and Christopher Amato. 2017. Learning in POMDPs with Monte Carlo Tree Search. In *Proceedings of the 34th International Conference on Machine Learning*. PMLR, 1819–1827. https://proceedings.mlr.press/v70/katt17a.html ISSN: 2640-3498.

[10] Zakariya Laouar, Qi Heng Ho, Rayan Mazouz, Tyler Becker, and Zachary N. Sunberg. 2024. Feasibility-Guided Safety-Aware Model Predictive Control for Jump Markov Linear Systems. http://arxiv.org/abs/2310.14116 arXiv:2310.14116 [cs, eess].

[11] Arnab Nilim and Laurent El Ghaoui. 2005. Robust Control of Markov Decision Processes with Uncertain Transition Matrices. *Operations Research* 53, 5 (Oct. 2005), 780–798. https://doi.org/10.1287/opre.1050.0216

[12] Takayuki Osogami. 2015. Robust partially observable Markov decision process. In *Proceedings of the 32nd International Conference on Machine Learning*. PMLR, 106–115. https://proceedings.mlr.press/v37/osogami15.html ISSN: 1938-7228.

[13] Stephane Ross, Brahim Chaib-draa, and Joelle Pineau. 2007. Bayes-Adaptive POMDPs. In *Advances in Neural Information Processing Systems*, Vol. 20. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2007/hash/3b3dbaf68507998acd6a5a5254ab2d76-Abstract.html