Utility Decomposition for Planning under Uncertainty for Autonomous Driving

Doctoral Consortium

Maxime Bouton
Stanford University
Stanford, California
boutonm@stanford.edu

ABSTRACT

The objective of this research is to provide scalable decision making algorithms for autonomously navigating urban environments. The vehicle must plan in a stochastic environment with many entities to avoid, rapid changes in driver behavior, and partial observability. Partially observable Markov decision processes (POMDP) offer a theoretically grounded framework to model such problems. We aim at developing a scalable POMDP formulation that takes into account dynamic occlusions, interaction between entities, and can generalize to a variety of different scenarios. This work demonstrates utility fusion and deep reinforcement learning methods to efficiently find optimal policies to navigate occluded urban environments.

KEYWORDS

POMDP; Reinforcement Learning; Autonomous Driving; Decision Making

ACM Reference Format:
Maxime Bouton. 2018. Utility Decomposition for Planning under Uncertainty for Autonomous Driving. In *Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), Stockholm, Sweden, July 10–15, 2018,* IFAAMAS, 2 pages.

1 INTRODUCTION

Autonomous vehicles are expected to drive safely and reliably in crowded urban environments. Decision making for autonomous vehicles is challenging because they must anticipate rapid changes in the intentions of human drivers and pedestrians. In addition, they only have partial observations of the environment due to sensor noise and occlusion. Providing appropriate behavior requires reasoning about the potential locations of pedestrians and other vehicles along with their motion over time. Hand-engineering strategies to navigate such environments would require anticipating the space of possible situations and finding a suitable behavior for each, which places a large burden on the designer and is unlikely to scale to complicated situations. Instead, our POMDP planning and utility decomposition method automatically finds suitable behavior by optimizing decision policies.

Previous work has explored modeling autonomous urban navigation scenarios as POMDPs, which is a mathematical framework for modeling dynamic, uncertain scenarios with imperfect state measurements. Most POMDP approaches in the literature provide tailored solutions considering only a small number of agents, focusing either on intention prediction or occlusions but rarely both, and still face tractability challenges [2, 5]. Some of them rely on discretization, which imposes a rigid and suboptimal representation. Recent advances in POMDP planning techniques have demonstrated their practical use on a robotic platform navigating in a crowd [1]. Although these results are promising, it is still unclear how the approach would perform in multi-modal scenarios with sensor occlusions. In addition, large problems require simplified models that fail to exploit interactions between traffic participants.

We propose a generic representation of driving scenarios as POMDPs, considering sensor occlusions and interaction between entities. The limitation of this framework is the computational tractability, which we are addressing through utility fusion techniques [9]. We demonstrate how to combine state-of-the art deep Reinforcement Learning (RL) methods in a POMDP formulation. Future research directions involve generalizing the approach to a variety of scenarios and safety verification of the resulting policies.

2 PROPOSED APPROACH

Autonomous driving can be described as a sequential decision making problem where the autonomous vehicle receives a partial observation of the environment and must take an action to maximize a high level objective. We first explain how to model urban navigation scenarios as POMDPs and then demonstrate a decomposition method to efficiently solve large decision making problems.

2.1 Modeling autonomous driving scenarios

In a POMDP, the environment is fully described by a state and evolves through time following an underlying probabilistic model. The agent only has partial observability of this state. At every time step, the autonomous vehicle receives a noisy observation containing partial information about the state and must decide which action to take. Since the state is not fully observable, the agent must estimate it using the history of its past observations and actions. It aggregates the resulting information in a belief state, which is a distribution over the possible states of the environment. Solving a POMDP involves finding a mapping between the belief state and the action to execute that maximizes the expected accumulated discounted reward.

To model such scenarios, we suggest an entity-based representation of the state. An entity can be a pedestrian, another vehicle, or a stationary obstacle such as a parked truck. Each entity is represented by its physical state (its position, heading, velocity, and

Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), M. Dastani, G. Sukthankar, E. André, S. Koenig (eds.), July 10–15, 2018, Stockholm, Sweden.

geometry). Including stationary obstacles in the state representation allows the formulation to generalize to a variety of occluded environments. The road topology is not explicitly present in the state space but is assumed to be known by the planner. The autonomous vehicle receives a noisy measurement of the physical state of all the visible entities. However, some road-users might be occluded by fixed obstacles or other moving entities. We have demonstrated that belief state planners are well suited to reason about the location of undetected road-users [4].

2.2 Utility Decomposition

To scale the POMDP approach to complex environments involving a large number of entities, we propose decomposing the problem into several tractable subproblems. Utility decomposition involves approximating the global utility function as a combination of the optimal utility function of each subproblem [6, 9]. A full scale navigation scenario involving cars, pedestrian and stationary obstacles can be decomposed in subgroups of three interacting entities, one of them being the ego vehicle, to account both for interaction between traffic participants and dynamic occlusions. Since the subproblems are lower dimensional, they can be solved using offline POMDP planners. For each of them, the solution method gives an approximately optimal belief-action utility function. The utility function of the full scale problem is then approximated by combining the solutions of the individual subproblems. This operation significantly reduces the computational effort while enabling the planner to consider a very large number of road-users. The resulting policy outperforms rule-based methods, but the utility decomposition sacrifices optimality.

To address this suboptimality, we proposed a technique inspired by multi-fidelity optimization to derive a corrective term using a neural representation [3]. This technique relies on the deep Q-learning algorithm to learn the correction term gearing the solution from utility decomposition towards optimality. Initializing the policy using the solution from the decomposition method significantly reduces the exploration needed in deep Q-learning. The correction term is sparse and easy to learn compared to learning the full-scale utility function. We have shown empirically, through simulation of an occluded crosswalk scenario, that learning the correction term improves the quality of the policy in terms of the number of collisions and the efficiency of the trajectory.

The decomposition method enables scaling POMDP techniques to very large problems currently intractable by state-of-the-art offline POMDP planners [8].

3 FUTURE DIRECTIONS

Thus far, This work has applied POMDP techniques and utility decomposition to planning for autonomous driving in spite of sensor occlusions. Additionally, we proposed an improvement on existing approaches through learning an additive correction using a neural representation. These works showed promising results in POMDP methods for automotive applications.

The next steps of this research include extending the experiments to different scenarios involving dynamic occlusions and interaction between traffic participants. Considering subproblems of three entities allows us to model interactions between two traffic participants as well as the ego vehicle. We first consider simple heuristics to model this interaction and analyze the benefits of this modeling on the resulting policy. An extension would be to model the other agent using game theoretical approaches such as level-k models. Enabling interaction-awareness in the planning model is likely to improve the quality of the policy.

Another open problem is the generalization to different scenarios. Given the recent advances in deep RL algorithms, we decided to investigate the integration of neural representations in the POMDP planning procedure. Our first consideration is to use recurrent neural networks to carry out the state estimation part (belief update) of the planning. Using a neural representation as the state estimator is likely to provide better generalization performance than conventional Bayesian filters. Since obstacles are part of the state description, the agent can keep a belief over possible obstacle configurations. The resulting policy should be robust to any obstacle shape.

Finally, another challenge to address is the lack of performance and safety guarantees. Our current approach uses the reward function of the POMDP to balance conflicting objectives such as safety and efficiency by varying the cost associated with collisions [2]. Metrics, such as time to reach the goal and collision rate, are then used to choose a suitable operating point. Tools from formal methods can be used to verify properties of a system with high confidence and to constrain the search in RL or POMDP algorithms [7].

REFERENCES

[1] Haoyu Bai, Shaojun Cai, Nan Ye, David Hsu, and Wee Sun Lee. 2015. Intention-aware online POMDP planning for autonomous driving in a crowd. *IEEE International Conference on Robotics and Automation (ICRA)*.

[2] Maxime Bouton, Akansel Cosgun, and Mykel J. Kochenderfer. 2017. Belief state planning for autonomously navigating urban intersections. In *IEEE Intelligent Vehicles Symposium (IV)*.

[3] Maxime Bouton, Kyle Julian, Alireza Nakhaei, Kikuo Fujimura, and Mykel J. Kochenderfer. 2018. Utility Decomposition with Deep Corrections for Scalable Planning under Uncertainty. *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

[4] Maxime Bouton, Alireza Nakhaei, Kikuo Fujimura, and Mykel J. Kochenderfer. 2018. Scalable Decision Making with Sensor Occlusions for Autonomous Driving. *IEEE International Conference on Robotics and Automation (ICRA)*.

[5] Sebastian Brechtel, Tobias Gindele, and Rüdiger Dillmann. 2014. Probabilistic decision-making under uncertainty for autonomous driving using continuous POMDPs. In *IEEE International Conference on Intelligent Transportation Systems (ITSC)*.

[6] James P Chryssanthacopoulos and Mykel J Kochenderfer. 2012. Decomposition methods for optimized collision avoidance with multiple threats. *AIAA Journal of Guidance, Control, and Dynamics*.

[7] Javier García and Fernando Fernández. 2015. A comprehensive survey on safe reinforcement learning. *Journal of Machine Learning*.

[8] Hanna Kurniawati, David Hsu, and Wee Sun Lee. 2008. SARSOP: Efficient Point-Based POMDP Planning by Approximating Optimally Reachable Belief Spaces. *Robotics: Science and Systems*.

[9] Stuart J. Russell and Andrew Zimdars. 2003. Q-Decomposition for Reinforcement Learning Agents. In *International Conference on Machine Learning (ICML)*.