Robotics & Automation Engineering Journal
ISSN: 2577-2899

Juniper Publishers
key to the Researchers

Mini Review
Volume 2 Issue 1 - February 2018
DOI: 10.19080/RAEJ.2018.02.555580

Robot Autom Eng J

Autonomous Robotic Manipulation as a Partially Observable Markov Decision Process

Thibault Rouillard[1], Ian Howard[2] and Lei Cui[3]*

Department of Science and Engineering, Curtin University, Australia

Submission: November 22, 2017; **Published:** February 02, 2018

***Corresponding author:** Lei Cui, Curtin University, Kent St, Bentley WA 6102, Australia, Tel: +61892667594; Email: Lei.Cui@curtin.edu.au

Abstract

Object manipulation is the most significant and challenging way for a robot to interact with its environment. It often requires computation on different channels of sensory information and the use of sophisticated controllers that can quickly lead to intractable solutions. In recent years, increasingly uncertain work spaces have lead researchers to turn to Partially Observable Markov Processes for robotic manipulation. The frame works provide methods to find an optimal sequence of actions to follow based on state belief and a partially observed world. This paper reviews the use of POMDP in the literature and points to fundamental papers on the subject.

Keywords: POMDP; Robotic manipulation

Abbreviations: POMDP: Partially Observable Markov Decision Process; RL: Reinforcement Learning

Introduction

Robots are increasingly expected to be more than mindless tools that are programmed to do one task in one particular way. They must perform in domestic/human environments and respond intelligently to high level command such as “make me dinner” or “find my phone”. Nonetheless, the uncertainty of these environments, unreliable sensory data and the difficulty in transforming a high level task to a low level manipulation problem makes this premise a challenging one. Manipulation and task planning capabilities for robotic systems are by consequence a vibrant subject of research at the present time.

POMDP [1] is a promising framework that integrates robotic manipulation and task planning and allows robots to develop sophisticated behaviour based on the user defined goals. It is formulated as the tuple <S; A; T; R; O; Z> where S, the state space of the process, is the collection of all system states. Depending of the application, the state space may be multi dimensional containing both information on the robot (configuration, joint torque, etc...) and on the workspace acquired through sensors (Tactile, Force/Torque, Depth Camera, etc...). A is the action space, consisting of all the actions available to the robot such as input torques on a joint or the movement of a certain object (macro actions [2]). An action a ϵ A leads to a state transition which is non-deterministic and described by T. In scenarios where the transition cannot be described analytically, they can be learned through experience [3]. R, the reward function, assigns a positive or negative numerical value to actions taken in specific parts of the state space. Rewards allow the robot to converge towards the desired goal and to stay away from undesirable situations, the process known as reward shaping. 0 ∈ O is the observation space, that generally consist of information received from the sensors on the robots. The reliability of an observation o 2 O is embedded in Z, the observation function, which gives the probability of an observation being accurate. It corresponds to the partially observable aspect of the process.

Since the states of the system are only partially observable, POMDPs reason on a belief, b(s) = Pr (s\o,a) rather than the states themselves. The belief is a probability density function over the state space making it continuous. The axioms of probability dictate that $0 \le b(s) \le 1$ and that $\Sigma_{s \in S} b(s) = 1$ [1]. Upon taking an action and receiving an observation, a new belief b′ can be recovered using a Bayes filter:

$$b'(s') = \frac{o(s', a, 0) \Sigma s \in s T(s, s', a) b(s)}{\Pr(o / s, a)}$$

This part of the process is often referred to as the state estimator SE.

The second part of a partially observable Markov decision process is to find the optimal action to take based on the current belief. The Bellman equation [1] can be used to compute the value of a belief state which is based on the immediate reward received for taking an action in a certain state and the expected future reward over a defined horizon. Keeping track of the belief and searching for an optimal action through the state, observation and action space becomes

rapidly expensive. Recent algorithms such as point-based value iteration [4] and SARSOP [5] only consider parts of the belief to make the computations tractable.

Section II of this paper elaborates on a few applications of POMDPs specifically for manipulation tasks and section III concludes the paper.

Pomdp for Robotic Manipulation

Autonomous robotic manipulation can be addressed on different levels: low level controls [6]. Where the robot's configuration becomes the state space of the decision maker and the action space controls the configuration directly or, higher level task planning [7] where macro actions are used to act on a state space that is constructed from sensory information. Both can be addressed using the POMDP framework.

In [6], Vien and Toussaint use a Willow garage PR2 platform for object manipulation using tactile feedback. They split the state space into two distinctive parts, one with smooth dynamics: the joint space of the robot and one with non-smooth ones: object location and table height. On the opposite, in [7], the authors consider manipulation of objects in a cluttered environment. Only the objects location, orientation and type are part of the state space. It is the macro actions used (e.g MOVE-OBJECT-TO (i,x,y)) that allow the robot to earn rewards. The authors do not explicitly state their procedure for grasping behaviour. The most common assumption in this case is access to a grasping library that contains pre-programmed grasp configuration for known objects.

One of the challenging aspects of POMDPs is to define the state transitions in scenarios with non-smooth dynamics. In the absence of a model, the transitions can be learned though techniques such as reinforcement learning (RL). In Chebotar et al. [8] use the POMDP framework for a re grasping task using a 3-fingered robotic hand. From an initial failed grasp, they use RL to learn the mapping between the data acquired during the failed grasp and the grasp adjustment. Sung et al. [9], tackle a scenario where a robotic gripper must turn a button until it clicks. In this case, they point out that only tactile feedback is useful in completing the task. They use a deep recurrent recognition network to learn to represent the haptic feedback.

In most robotic applications, the scale of the state space and action space is too large to be acted upon directly. It must be appropriately reduced whilst maintaining the cognitive aspect of the robot. Using the POMDP framework, rewards may be assigned to actions that help to disambiguate the state space [1]. Monso et. al. [10] use a robotic manipulator for clothes separation. The state space in their paper is reduced to 2 dimensions with 4 possible states per dimension. Most of the intelligence is transferred to the actions. Finally in [11], the authors use the POMDP framework to plan grasping behaviour. They use several abstractions on the state space and action space to work around the continuity of the belief space.

Conclusion

Robots today need to be intelligent in order to cope with uncertainty when interacting with their environment. Robotic manipulation is an impactful way for robots to affect their workspaces and partially observable Markov decision processes provide a solid framework to do so. It is continuously used thorough the literature and allows the development of sophisticated behaviours for autonomous robotics manipulation.

Acknowledgement

Dr Lei Cui is the recipient of an Australian Research Council Australian Discovery Early Career Award (project number: DE170101062) funded by the Australian Government.

References

1. Kaelbling LP, Littman ML, Cassandra AR (1995) Partially observable Markov decision processes for artificial intelligence, Springer Berlin Heidelberg, Germany, p. 1-17.
2. Pajarinen J, Kyrki V (2015) Robotic manipulation of multiple objects as a pomdp. Artificial Intelligence.
3. Pape L, Oddo CM, Controzzi M, Cipriani C, Förster A, et al. (2012) Learning tactile skills through curious exploration. Front Neurorobot 6: 6.
4. Pineau J, Gordon G, Thrun S (2003) Point-based value iteration: An anytime algorithm for pomdps. IJCAI 3: 1025-1032.
5. Kurniawati H, Hsu D, Lee WS (2009) Sarsop: Efficient point-based pomdp planning by approximating optimally reachable belief spaces. Robotics: Science and systems, Zurich, Switzerland.
6. Vien NA, Toussaint M (2015) Pomdp manipulation via trajectory optimization. Intelligent Robots and Systems (IROS), International Conference on, Daejeon, South Korea.
7. Li JK, Hsu D, Lee WS (2016) Act to see and see to act: Pomdp planning for objects search in clutter. Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on, Daejeon, South Korea.
8. Chebotar Y, Hausman K, Su Z, Sukhatme GS, Schaal S (2016) Self-supervised regrasping using spatiotemporal tactile features and reinforcement learning. Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on, Daejeon, South Korea.
9. Sung J, Salisbury JK, Saxena A (2017) Learning to represent haptic feedback for partially-observable tasks. arXiv preprint arXiv:1705.06243.
10. Monsó P, Alenyà G, Torras C (2012) Pomdp approach to robotized clothes separation. Intelligent Robots and Systems (IROS), IEEE International Conference pp. 1324-1329.
11. Hsiao K, Kaelbling LP, Lozano-Perez T (2007) Grasping pomdps. Robotics and Automation, IEEE International Conference on, pp. 4685-4692.

How to cite this article: Thibault Rouillard, Ian Howard, Lei Cui. Autonomous Robotic Manipulation as a Partially Observable Markov Decision Process. Robot Autom Eng J. 2018; 2(1): 555580. DOI: 10.19080/RAEJ.2018.02.555580.

DOI: 10.19080/RAEJ.2018.02.555580

How to cite this article: Thibault Rouillard, Ian Howard, Lei Cui. Autonomous Robotic Manipulation as a Partially Observable Markov Decision Process. Robot Autom Eng J. 2018; 2(1): 555580. DOI: 10.19080/RAEJ.2018.02.555580.