Robotics and Autonomous Systems 83 (2016) 15–31

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

Planning for robotic exploration based on forward simulation

Mikko Lauri *, Risto Ritala

Department of Automation Science and Engineering, Tampere University of Technology, P.O. Box 692, 33101 Tampere, Finland

HIGHLIGHTS

- Exploration formulated and solved as partially observable Markov decision process.
- New sampling-based approximation for mutual information in mobile robotics.
- Efficient algorithm for drawing samples for forward-simulation based planning.
- Experimental validation in simulated and real-world exploration domains.
- Software available at https://goo.gl/ENGkIf.

ARTICLE INFO

Article history:
Received 29 October 2015
Accepted 14 June 2016
Available online 23 June 2016

Keywords:
Partially observable Markov decision process
Active sensing
Robotic exploration
Mutual information
Sensor management

ABSTRACT

We address the problem of controlling a mobile robot to explore a partially known environment. The robot's objective is the maximization of the amount of information collected about the environment. We formulate the problem as a partially observable Markov decision process (POMDP) with an information-theoretic objective function, and solve it applying forward simulation algorithms with an open-loop approximation. We present a new sample-based approximation for mutual information useful in mobile robotics. The approximation can be seamlessly integrated with forward simulation planning algorithms. We investigate the usefulness of POMDP based planning for exploration, and to alleviate some of its weaknesses propose a combination with frontier based exploration. Experimental results in simulated and real environments show that, depending on the environment, applying POMDP based planning for exploration can improve performance over frontier exploration.

1. Introduction

Autonomous robotic agents performing tasks such as monitoring, surveillance or exploration must be able to plan their future information-gathering actions. Real-world environments are typically partially observable and stochastic, and planning in them requires reasoning over uncertain outcomes in the presence of sensor noise. The true state of the system is hidden, and knowledge about the state is represented by a belief state, a probability density function (pdf) over the true state. The utility of actions is measured by an appropriate reward function, and the agent's objective is to maximize the sum of expected rewards over a specified horizon of time. Such planning problems are instances of partially observable Markov decision processes [1], or POMDPs.

The solution of a POMDP is a control policy, i.e. a mapping from belief states to actions. To find policies for information-gathering and exploration tasks, several authors have proposed applying quantities such as entropy or mutual information as reward functions of the POMDP [2–6]. Although finding optimal policies for POMDPs is computationally hard (PSPACE-complete; [7]), they remain an attractive modelling choice due to the ability to simultaneously handle uncertainties in the robot's action and sensing outcomes.

In this article, we address the problem of finding control policies for robotic exploration problems formulated as POMDPs. We apply forward simulation algorithms for finding a solution to an open loop approximation of a POMDP. This approach allows general belief-dependent reward functions and with suitable choice of algorithm can avoid discretization of the continuous planning space. We derive a sampling-based approximation for mutual information that can be applied in conjunction with forward simulation based planning, and describe a method for efficiently drawing the required samples. We provide an empirical evaluation of our proposed approach in simulated and real-world exploration experiments.

This article is organized as follows. Section 2 provides a survey of related work and discusses the relation of our contribution to

* Corresponding author.
E-mail addresses: mikko.lauri@tut.fi (M. Lauri), risto.ritala@tut.fi (R. Ritala).

http://dx.doi.org/10.1016/j.robot.2016.06.008
0921-8890/© 2016 Elsevier B.V. All rights reserved.

the state-of-the-art. In Section 3, we formulate exploration as a POMDP, and discuss possible solution methods. Section 4 reviews two forward simulation based methods for non-myopic planning in POMDPs. Section 5 introduces the sample-based approximation of mutual information suitable for robotic exploration problems, and describes a method for efficiently drawing the required samples. Section 6 describes the results of simulation experiments, and Section 7 presents the software architecture for implementation of our approach and reports the results of real-world exploration trials. Finally, Section 8 provides concluding remarks.

2. Related work

Mobile robots typically collect information on both their internal state and the state of the environment they are interacting with. In simultaneous localization and mapping (SLAM) (see Durrant-Whyte and Bailey [8] for a review) the robot must jointly estimate both its pose (internal state) and the map of the environment based on its actions and observations. Robots' information-gathering actions consist of actions that affect their pose, and hence the area covered by their sensors, and actions explicitly selecting between sensors or their operating modes. In the active SLAM problem [9], the robot's actions are selected to obtain best estimates on the pose and the map. Thus, the active SLAM problem is an exploration or sensor selection problem. The goal is to maximize information on both the robot pose and the map.

Techniques to control mobile robot exploration may be categorized e.g. by whether they apply heuristic rules or formal decision theory for selection of exploration targets. Applying heuristic rules for guiding exploration spans frontier-based approaches [10], or some next-best-view approaches such as [11]. Juliá et al. [12] classify exploration methods according to the levels of multi-robot coordination and integration with SLAM algorithms. They conclude that SLAM-integrated exploration performs best with regard to quality of map information. Their results also agree with Amigoni [13], Amigoni and Caglioti [14], who found decision theoretic criteria combining both utility of exploration and its cost (e.g. time or distance) to be preferable if both extent of the explored area and exploration time were optimized. In the following, we review in more detail some single-robot exploration techniques that employ decision theoretic criteria to guide the exploration process.

Information on the location of the robot and environment features, or landmarks, may be modelled by a multivariate Gaussian distribution. The SLAM problem can then be solved for example by applying the Extended Kalman Filter. Exploration with such feature-based maps was studied by Sim and Roy [15] who describe an A-optimal exploration method, i.e. they minimize the trace of the state covariance matrix. They discretize the location of the robot to a grid and plan an informative trajectory in open loop as a sequence of discrete positions via a breadth-first search. A similar objective function was used by [16], adopting a model predictive control (MPC) approach for optimization over multiple time steps. Discretization of the action space was also applied by [17], who applied reinforcement learning to learn parameterized robot trajectories for exploration. A somewhat complementary approach was adopted in [18], where a set of candidate exploration targets were evaluated based on an utility function designed to balance exploration of unknown areas and seeing known landmarks to help maintain good localization information. However, an explicit information-theoretic quantification of the information gain was avoided.

Martinez-Cantin et al. [19] relaxed assumptions made in earlier work, such as discretization of the action space and the myopic optimization horizon. They applied a Monte Carlo search algorithm in policy space with a Gaussian process approximation of the objective function. The policies to evaluate were selected by minimizing the average mean square error of the state estimate consisting of robot and landmark locations.

A belief space planning approach investigated by [20,21] addressed many of the limitations of earlier studies. A planner architecture consisting of an estimation layer and a decision layer combined with a model predictive control strategy for non-myopic planning was applied, assuming a Gaussian belief over robot and landmark poses. Discretization was avoided by applying a gradient descent method for computing optimal actions. Possible future measurements were treated as random variables, relaxing the assumption of maximum likelihood measurements. Exploration was considered in the sense that the objective function included an A-optimality criterion for state covariance.

Another body of work in exploration employs metric map representations such as occupancy grids [22]. Bourgault et al. [23] combined occupancy grid mapping with feature-based SLAM and used mutual information as the reward function. A discretized action space was applied with myopic optimization.

Rao–Blackwellized particle filtering (RBPF) is often applied in state-of-the-art SLAM filters for occupancy grid maps [24]. Each particle represents a map and a robot trajectory hypothesis. Stachniss et al. [2] studied myopic exploration in RBPF SLAM by discretizing the action space to a set of possible waypoints and then evaluating the approximate expected information gain when travelling to the waypoints by sampling. This work was later expanded upon [25,9] by considering alternative measures of uncertainty of the SLAM solution. These approaches consider both exploration of new areas and maintaining the consistency of the particle filter approximation.

Contribution. We present a new approximation for mutual information that is useful in mobile robotics exploration problems. The approximation can be easily integrated with forward simulation planning methods, and avoids computing full SLAM filter updates during the planning phase. In contrast to e.g. [19–21], we do not assume a Gaussian belief state. We propose and empirically evaluate in simulated and real-world domains a exploration method combining strengths of decision-theoretic POMDP based exploration and classical frontier based exploration. In all cases, we concentrate on non-myopic planning instead of the greedy one-step maximization of utility.

3. Exploration as a POMDP

Consider a robot exploring a partially observable environment. Let $s \in \mathcal{S}$ denote the hidden state of the system, comprising the state of the robot and the state of the environment. At each decision epoch in the set $\mathcal{T} = \{0, 1, \ldots, H-1\}, H \in \mathbb{N} \cup \{\infty\}$, where H is the horizon of the problem, the robot selects a control action $u \in \mathcal{U}$. Consequently, the state at the next decision epoch is determined by a transition according to a Markovian state transition model $\mathbb{T}(s', s, u)$, giving the conditional probability of transitioning to $s' \in \mathcal{S}$ from $s \in \mathcal{S}$ when action $u \in \mathcal{U}$ is executed. After the state transition, the robot obtains information regarding the state in the form of an observation. The observation is modelled by a probabilistic model $\mathbb{O}(z', s', u)$ giving the conditional probability of perceiving observation $z' \in \mathcal{Z}$ in state $s' \in \mathcal{S}$ after action $u \in \mathcal{U}$ was executed. As the robot's knowledge of the true system state is incomplete, it is represented by a belief state $b \in \mathcal{B}$, a probability density function (pdf) over the state space $\mathcal{S}$. The set $\mathcal{B}$, containing all pdfs over $\mathcal{S}$, is called the belief space.

As the robot executes control actions and perceives observations, its belief state is tracked by Bayesian filtering. Given a belief state $b \in \mathcal{B}$ and an action $u \in \mathcal{U}$, the predicted belief state b^u is computed by

$$b^u(s') = \int_{s \in \mathcal{S}} \mathbb{T}(s', s, u) b(s) \mathrm{d}s. \quad (1)$$