Chapter 1

Introduction

1.1 Motivation

Taking long term decisions or spontaneous reactive actions when presented with incomplete information or partial knowledge is paramount to the survival of any biological or synthetic entity. Reasoning given a state of uncertainty is a continuously occurring event throughout our livelihood. When considering long term decisions an abundance of examples come to mind. For instance, in economic investments uncertainty is to the best of efforts quantified and minimised in order to avoid unwarranted risks. Reactive actions are just as common; when looking for the snooze button of an alarm clock, early in the morning, our hand seems to autonomously search the surrounding space picking up sensory cues gradually acquiring information guiding us towards the button. All the above types of decision require the integration of evidence and an ability to predict the outcomes of the taken decisions in order to insure a favourable end state. Abilities close to these have met with mixed levels of success in Artificial Intelligence (AI) & robotics. There is a been noticeable success in artificial agents beating humans at board games (backgammon, chess and go) but having a robot successfully climb a staircase, open a door or pick up a glass are still ongoing open problems.

It is not yet fully understood how decisions are taken, yet alone under uncertainty. The difficulty is that two processes responsible for the synthesis of our actions and decisions, our beliefs and desires, are not directly or easily measurable. There is growing interest in Neuroscience to understand the mechanisms underlying perception and decision making under uncertainty Preuschoff et al. (2013); there is not yet a consensus on the biological mechanisms involved in decision making and efforts are ongoing[1] to construct plausible models of our decision processes. At a behavioural level, early efforts to model human decision making were made in mathematics & economics (Bernoulli (1954),Von Neumann and Morgenstern (1990)), in which gambles and investments were chiefly considered. There has been considerable effort in many fields (neuroscience, cognitive science, physiology, economics, etc..) to understand how decisions and actions are taken, starting with the role of our neurons to high level decisions

[1] the human brain project: https://www.humanbrainproject.eu/

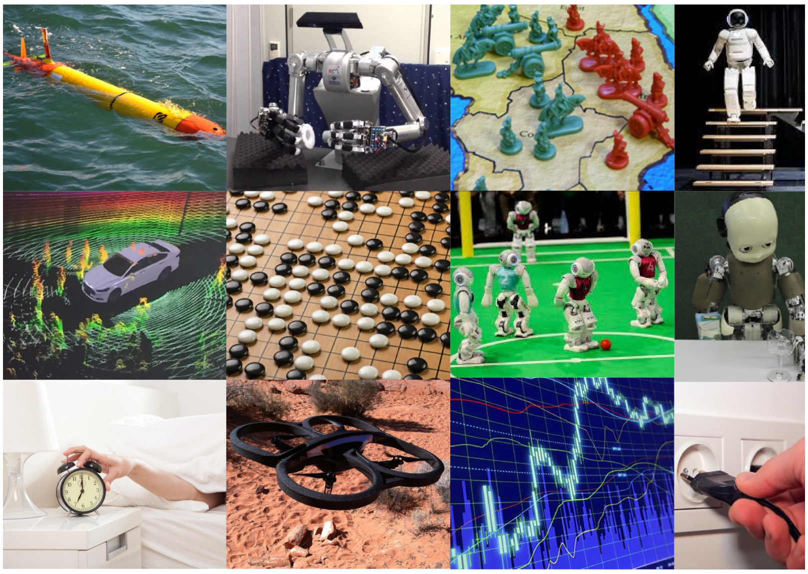

Figure 1.1: Examples of the decision making under uncertainty in both robotics and everyday life situations. Images taken from the public domain.

like gambling, orientation and navigation problems to reflexes.

Artificial intelligence & robotics considered early on uncertainty in decision making, where the predominant domain of application was spatial navigation, Cassandra et al. (1996). The problem has always been treated in two parts: the construction and representation of a world model (the map) and a planner which can reason with respect to this model in order to accomplish an objective. The world construction problem attracted a large amount of interest and has resulted in many successfu~~lly~~ applications in a wide spectrum of robotic domains (AUV, UAV, etc..). The integration of planning with mapping in a single framework is still difficult to achieve and is based on either representing the decision problem as a Partially Observable Markov Decision Process (POMDP) which is notoriously difficult to solve for large scale problems, or through search heuristics. The mapping problem can generally be solved when assuming the uncertainty is Gaussian and thus quantifiable by a few parameters.

In summary there are still open problems in decision making when considering partial observability. The mapping problem has been studied and solved within a certain set of constraining assumptions. For the mapping problem we develop a Bayesian filter which is non-parametric and has no explicit representation of a joint distribution.

Currently, both humans and animals are far better at navigation than robots, especially when uncertainty is present, Stankiewicz et al. (2006). When addressing the decision making, we leverage human foresight and reasoning in a Learning from Demonstration (LfD) framework (Billard et al. (2008)), which is used to transfer skills from an expert teacher (usually a human) to a robot. Examples include the transfer of kinematic task constraints, stiffness and impedance constraints and motion primitives, to name only a few.

In this thesis we address both problems under extreme levels of uncertainty.

1.2 Contribution

In this thesis we bring to light two main ideas. The first is the transfer of human behaviour to robots in tasks where a lot of uncertainty in present, making them difficult to solve using traditional techniques. The second is a non-parametric Bayesian state space filter which is efficient under sparse sensory information and high levels of uncertainty.

Throughout the work in this thesis we consider case studies in which vision is not available, leaving tactile and haptic information. This choice was made to induce a high level of uncertainty making it easier to study its effect on the decision making process. As a consequence the tasks we consider are by nature, haptic and tactile searches. The following three sections detail the contribution of this thesis to research decision making under sever uncertainty constraints.

1.2.1 Learning to reason with uncertainty as humans

A Markov Decision Process (MDP) allows the formulation of a decision problem in terms of states, actions, a discount factor and a cost function. Given this formulation and a suitable optimisation method (dynamic programming, temporal difference, etc..) a set of optimal decision rules are returned, known as a policy. The benefit of this approach is that the policy is non-myopic and sequences of complicated actions can be synthesised to achieve a goal which an opportunistic policy would fail to achieve. A Partially Observable Markov Decision Process (POMDP) is a generalisation of an MDP to a hidden state space and only observations are available relating to the state space. Finding an exact optimal solution to a POMDP problem is notoriously difficult due to the computational complexities involved. Sample based approaches to solve a POMDP rely heavily on a good trade-off between exploration and exploitation actions. Good explorative actions increase the chance of discovering a set of optimal decisions/actions.

In this thesis we propose a Learning from Demonstration approach to solving POMDP problems in haptic and tactile search tasks. Our hypothesis is that if we know the mental state of the human expert in terms of his believed location and observe his actions we can learn a statistical policy which mimics his behaviour. Since the human's beliefs are not directly observable we infer them by assuming that the way we integrate evidence is similar to a Bayesian filter. There is evidence both in cognitive and neuroscience that this is the case (Bake et al. (2011)). From observing the expert human performing a task we learn a cognitive model of the human's decision process by learning a generative joint distribution over his beliefs and actions. The generative distribution is then used as a control policy. By this approach we are able to have a policy which can handle uncertainty similarly to humans.

1.2.2 Non-parametric Bayesian state space filter

Simultaneous Localisation and Mapping (SLAM) is concerned with the development of filters to accurately and efficiently infer the state parameters of an agent (position, orientation) and aspects of its environment, commonly referred to as the map. It is necessary for the agent to achieve situatedness which is a precondition to planning and reasoning. The predominant assumption in most applications of SLAM algorithms is that uncertainty is related to the noise in the sensor measurements. In our haptic search tasks there is no visual information and a very large amount of uncertainty. Most of the sensory feedback is negative information, a term used to denote the non event of a sensory response. In the absence of recurrent sightings or direct measurements of objects there are no correlations from the measurement errors which can be exploited.

In this thesis we propose a new SLAM filter, which we name Measurement Likelihood Memory Filter (MLMF), in which no assumptions are made with respect to the shape of the uncertainty (it can be Gaussian, multi-modal, uniform, etc..) and motion noise. We adopt a histogram parametrisation (this is considered non-parametric because a change in a parameter has a local effect). The conceptual difference between the MLMF and standard SLAM filters, such as the Extended Kalman Filter (EKF), is that we avoid representing the joint distribution since it would entail a shattering space and time complexity. This is achieved by keeping track of the history of measurement likelihood functions. We demonstrate that our approach gives the same filtered marginals as a histogram filter. In such a way we achieve a Bayes filter which has both linear space and time complexity. This filter is well suited to tasks where the landmarks are not directly observable.

1.2.3 Reinforcement learning in belief space

We propose a Reinforcement Learning framework for the task of searching and connecting a power plug to a socket, with only haptic. We previously addressed this setup by learning a generative model of the beliefs and actions with data provided by human demonstrations following the LfD approach. However, it is usually the requirement that the teacher is an expert, with few notable exceptions (Rai et al. (2013)). Since we were solely learning a statistical controller, both good and bad demonstrations will be mixed in together. By introducing a cost function representing the task we can explicitly have a quality metric of the provided demonstrations. In this way we can optimise the parameters of our generative model to maximise the cost function. In this LfD Reinforcement Learning setup with a very simple cost function we can have a significant improvement of our a policy.

1.3 Thesis outline

The thesis is structured according~~ly~~ to the three main contributions outlined in the previous section, and all will have their individual chapter. We outline below the structure of the thesis.

Chapter 2 - Background

In this chapter we introduce and mathematically formalise the sequential decision making problem under uncertainty and we provide a detailed literature review of the related work in this domain. We provide a brief introduction to *Decision Theory* before focusing on the work in AI & robotics relevant to POMDPs whilst highlighting their relevance and contribution to our work.

Chapter 3 - Learning to reason with uncertainty as humans

In this chapter we present an approach for transferring human skills in a blind haptic search task to a robot. The belief of the human is represented by a particle filter and all subsequent beliefs are inferred from the human's motions acquired via a motion tracking system. A generative model of the joint belief and actions distribution is learned and used to reproduce the behaviour on a WAM and KUKA robot in two search tasks. Experimental evaluations showed the approach to be superior to greedy opportunistic policies and traditional path planning algorithms. The major parts of this chapter have been presented de Chambrier and Billard (2014). We also provide a review of work related to humans taking decisions under uncertainty in spatial navigation and haptic tasks with an emphasis on works which consider diminished or no visual information.

Chapter 4 - Non-parametric Bayesian state space filter

In this chapter we present an approach to perform a state space estimation of a map and agent given that there is no direct observation between the landmarks and the agent. We demonstrate that by not explicitly parametrizing the full joint distribution of the landmarks and agent but instead keeping track of the applied measurement functions we can fully reconstruct the optimal Bayesian state estimation. The advantage of our approach is that the space complexity is linear as oppose to exponential. We validate our approach in 2D search navigation tasks. This work is currently under review. We also give an overview of the literature of SLAM and emphasis the position of our filter within it.

Chapter 5 - Reinforcement learning in belief space

In this chapter we present an approach similar to the one presented in Chapter 3, "Learning to reason with uncertainty as humans", with the difference that we explicitly encode the task through the introduction of a binary objective function and we consider a peg-in-hole task under high levels of uncertainty. The task requires both high and low levels or precision to be able to accomplish it, which makes it particularly interesting. We learn a value function approximation of the belief space through locally weighted regression and approximate dynamical programming. By combining a LfD approach in this Actor-critic Reinforcement Learning framework, we demonstrate an improvement upon a purely statistical controller with nearly no additional cost. We additionally provide a review of RL methods in the context of POMDPs.

Chapter 6 - Conclusion

We conclude by providing a holistic summary of our work and achievements. We draw attention to the current open problems and directions for future work in field of uncertainty and reasoning in Artificial intelligence and robotics.