Learning Search Strategies from Human Demonstrations

Guillaume de Chambrier

May 25, 2016

Abstract

This thesis lies within the domain of Programming by Demonstrations (PbD), State Space Estimation (SSE) and Reinforcement Learning (RL). A significant proportion of scientific work in PbD has focused on learning how to imitate human behaviour. Tasks such as “pick and place”, hitting motions, and bipedal locomotion have been encoded through either symbolic, statistical or dynamical system representations. In contrast there has been less focus on transferring higher cognitive behaviour, such as problem solving skills and search strategies, from humans to robots.

The aim of this thesis is to model human search strategies in spatial navigation tasks in which a high amount of uncertainty is present. We consider for instance tasks such as localising an object in a dark room or connecting a plug to a power socket in the dark, or any such situation with total suppression of visual information, and transfer this reasoning mechanism to a robot apprentice.

A difficulty in learning human reasoning mechanisms in the search scenarios we consider, is that the humans’ sensations (haptic and tactile) and beliefs are unobservable and they vary within and across subjects. We infer the human’s sensations from either the relationship between the kinematic body of the human and the known geometric description of the environment or a tool equipped with a sensor (force-torque sensor). The sensations are transformed to a binary feature vector which encodes whether contacts have been made with the environment’s features such as surfaces, edges and corners.

We model the human’s beliefs by a probability density function which we update through recursive Bayesian state space estimation using motion estimates, acquired through a tracking system (the human subjects wore markers), and the sensation estimates were obtained as described above. We make the assumption that the probability density function, representing the human’s belief, is updated by a Bayesian recursion and that this process is similar to the way in which humans integrate information.

To model the reasoning processes of human subjects performing the search tasks we learn a generative joint distribution over beliefs and actions (end-effector velocities) which were recorded during the executions of the task. The high dimensionality of the belief and its varying complexity during the searches required that we compress the belief to its most likely state and entropy.

We evaluate this methodology of learning search strategies in a task consisting of finding an object on a table. We demonstrate that multiple search strategies are encoded in the joint belief-action distribution and we compare this approach with greedy myopic and coastal navigation search algorithms. The results show that the human learned search model is the fastest of all methods.

We consider in a second setting a task in which human subjects have to demonstrate how to search for and connect a plug to a power socket to a robot apprentice deprived of visual information. We take the same approach but incorporate the learning of the policy into a reinforcement learning framework and demonstrate that by defining a simple cost function the quality of the final learned policy can be significantly improved without the need of performing exploratory rollouts which are costly and typically necessary in RL.

Both the above search tasks can be considered as active localisation in the sense there is uncertainty only in the position of the human or robot in the world. We finally consider a search setting in which both the position of the robot and aspects of the environment are uncertain. Given the unstructured nature of the belief, a histogram parametrisation of the joint distribution over the robot’s position and environmental features is necessary. However, this method quickly becomes infeasible as the computational cost is exponential in terms of the parametrisation. We demonstrate that by parametrising only the marginals and by memorising the parameters of the measurement likelihood functions we can obtain the exact same solution at a cost which is linear as opposed to exponential in space and time complexity.