

Learning Search Strategies from Human Demonstrations

Guillaume de Chambrier

May 24, 2016

Abstract

This thesis lies within the domain of Programming by Demonstrations (PbD), State Space Estimation (SSE) and Reinforcement Learning (RL). A large body 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.

Our aim is to model how humans reason with respect to their beliefs and the role uncertainty plays during spatial navigation search tasks. We consider for instance tasks such as localising an object in a room or connection a plug to a power socket in the dark, or any such situation with total suppression of visual information, and transfer this reasoning mechanisms to a robot apprentice.

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

We model the humans 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. An assumption we make is that probability density function, representing the humans 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 process of human subjects doing the search tasks we learn a generative joint distribution over beliefs and actions (end-effector velocities) both of which were recorded during the executions of the task by the human subjects. Because of high dimensionality of the belief and its varying complexity during searches we compress the belief to its most likely state and entropy.

We evaluated this methodology of learning search strategies in a task consisting of finding an object on a table. We demonstrated that multiple search strategies were encoded in the joint belief-action distribution and compared this approach against a greedy myopic and coastal navigation search algorithms and found that the human learned search model is faster than the other two methods.

We consider in a second setting a task in which human subjects have to demonstrate to a robot apprentice how to search for and connect a plug to a power socket, whilst also 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 search tasks above can be considered as active localisation in the sense that uncertainty originates from the position of the human or robot in the world. We now consider search setting in which both the position of the robot and aspects of environment are uncertain. Given the unstructured nature of the belief a histogram parametrisation of the joint distribution over the robot and the environment is necessary. However, naively doing so becomes quickly infeasible as the computational cost is exponential in terms of the parametrisation. We demonstrate that by only parametrising the marginals and by memorising the parameters of the measurement likelihood functions we can recover the exact same solution as the naive parametrisations at a cost which is linear in space and time complexity as oppose to exponential.