Infomax Control for Social Robots

Ian R. Fasel
Department of Computer Science
University of Arizona
ianfasel@cs.arizona.edu

Paul Ruvolo[1,2], **Tingfan Wu**[1,2], **Javier Movellan**[2,3]
[1]Department of Computer Science, [2]CalIT2
[3]Institute for Neural Computation
University of California, San Diego
{ruvolo, ting, movellan}@mplab.ucsd.edu

Robotics problems involve updating and acting on the basis of beliefs about the world state. In social robots, it is important to maintain beliefs not just about objects in the room and the locations of the walls, but also beliefs about peoples' locations, their head directions, who they are interacting with, and their emotional states. Each of these aspects of the social world may change rapidly and with complex dynamics. Social robots are also often designed with human-like affordances, including oriented sensors such as cameras in the eyes and microphones in the ears. This means that the robot must have a policy for sequentially orienting these sensors by changing its head direction over time. The problem is made more difficult by the fact that unlike special-purpose robots designed for a single task, a social robot may need to execute several different tasks depending upon the actions of the people it encounters. For instance a museum docent robot needs to be ready to answer questions or direct guests towards any number of exhibits when someone approaches it, and the rest of the time should look for new people who look confused or may be interested in a tour. Some possible strategies in this case might be to track any face that comes into view until it leaves some specified region, look for "interesting" events, fix its gaze at the doorway, or shift its gaze randomly.

When the next task is not easy to predict beforehand, a reasonable strategy for a robot may be to act in such a way that the uncertainty about the social world state is always kept as small as possible, so that it is, in some sense, "ready for anything". In the current work, we formalize this idea with an Information Maximization (InfoMax) POMDP, in which the reward function at each time step is the negative entropy of the probability distribution over the state of the social world (belief state). Because negentropy is an intrinsic reward derived from the belief state at every time-step, it is a good automatic "shaping" reward that does not depend on a specific human-provided reward or goal. This makes InfoMax an ideal "default" strategy in the absence of another external task demand.

We explored this idea by implementing an InfoMax control policy on a robotic head during the "Neuromorphic Engineering" summer workshop in Telluride, Colorado. The head, which is designed to look like Albert Einstein (complete with facial expressions) has directional cameras in its eyes which are fed to computer vision algorithms that can locate human faces, detect their head pose, and recognize their facial expressions. The robot was also equipped with a microphone array which could localize sources of sound, and a wide-field silicon retina which can rapidly detect a variety of subtle motion cues. The robot was placed in a large room in which people could enter and leave randomly, and who sometimes would approach the robot and try to interact with it. There was no high-level "goal" for the robot – that is, Einstein didn't have a job. Rather, his intrinsic desire was to always maintain as much certainty about the state of the social environment as possible.

We experimented with a variety of methods for combining the sensors to form probabilistic beliefs about the locations and facial expressions of people, ultimately settling on a simple recurrent network-based belief model. An InfoMax control policy was then learned using a natural actor-critic reinforcement learning (RL) algorithm in a simulated environment, using a convolutional radial-basis-function network as the controller to capture symmetry and shift-invariances. We found that depending on the reliability of the sensors and the dynamics governing the people in the social world, a number of different behaviors emerged automatically from the learning algorithm, typically within a few hundred timesteps. These included tracking individual faces, occasionally switching gaze between people, and turning to look towards sudden outbursts of sound. While these are all sensible ideas a human might have hand-coded, the RL algorithm learned them automatically, and also learned how to sequence them optimally. Interestingly, a key feature of the learned controllers were center-surround-like weight vectors that operated on probabilistic belief maps.