Learning Probabilistic Models via Bayesian Inverse Planning

Abdeslam Boularias
Department of Computer Science
Laval University, Quebec, G1K 7P4, Canada
boularias@damas.ift.ulaval.ca

Brahim Chaib-draa
Department of Computer Science
Laval University, Quebec, G1K 7P4, Canada
chaib@damas.ift.ulaval.ca

Consider two robots navigating in an environment where they should collaborate in order to perform a given task. Assume that this task can be performed only if the two robots are in the same location, so the robots should first plan to meet somewhere. We consider the case where the two robots are not able to communicate any information between them. This can be the case of robots fabricated by different companies for example. The two robots are also heterogenous, they have different capabilities and dynamics. For example, one robot is legged and can move across low obstacles, while the other is wheeled and cannot do so. Although each robot is provided with a precise model of its own dynamics, planning a meeting is not simple since each robot should also know the dynamics and the capabilities of the other in order to choose a plan accordingly. A direct solution to this problem consists in using prior probabilistic models representing what each robot believes the dynamics of the other robot is. The robots start executing plans according to these models, and gradually improve them according to the observed behaviors. However, the robots need to repeat each action in each state several times before the models converge. Moreover, the robots cannot learn anything about the state-actions that are being avoided, like moving toward obstacles.

In this work, we propose to use inverse planning, along with direct learning, in order to infer the dynamics of a robot by observing its optimal behavior. Inverse planning refers to the problem of answering the question: which model of the dynamics makes a given observed behavior optimal? If one robot observes that the other one prefers taking longer paths than moving across low obstacles, then the only possible explanation for this behavior is that the other robot simply ”cannot” move across such obstacles. This information leads to an improvement of the upcoming meeting plans.

This approach is related to inverse reinforcement learning [1] where an agent learns a reward function based on an observed optimal behavior and a known dynamics model. In the current work, the agent learns a dynamics model based on a known reward function and an observed optimal behavior.

Formally, we consider that the dynamics of each robot can be modeled as Markov Decision Process (MDP), where the transition functions for different agents are independent, and the agents get the same reward for their joint action. The problem of finding a transition function from an observed optimal policy is ill-posed, therefore, we adopt a bayesian regularization to solve this issue. We assume that the transition parameters are distributed according to a Dirichlet prior. After observing an agent executing an action in a given state and transiting to another state, the other agents calculate the maximum a posteriori (MAP) of its transition function by taking into account the evidence regarding the transition that happened, as well as the fact that the executed action was optimal.

We used a gradient ascent algorithm to calculate the MAP of the transition function, and the preliminary results on simple meeting problems show that the robots learn better models of their teammates when inverse planning is used.

Reference

[1] Ng, A., & Russell, S. (2000). Algorithms for Inverse Reinforcement Learning. *Proceedings of the Seventeenth International Conference on Machine Learning (ICML'00)* (pp. 663–670).