Reward Preferences in Reinforcement Learning

Daniel J. Lizotte (Statistics Department – Institute for Social Research – University of Michigan, USA)

April 11, 2011

For chronic diseases like schizophrenia, the focus of treatment is management rather than cure. Disease management is a sequential decision making process, and therefore learning the best immediate action requires a non-myopic view of each action's effects. Reinforcement learning and dynamic programming in general are ideally suited to learning treatment policies in this setting.

The actual usefulness of available treatments varies depending on the decision maker's objective or "reward" of interest. For example, certain schizophrenia drugs are effective at reducing symptoms but cause significant weight gain. Others provide less symptom relief but do not cause weight gain. The choice of one treatment over another thus depends on the preferences of the decision maker. Any preference-based combination of the conflicting rewards is a potential reward function one might want a learned treatment policy to optimize.

I will introduce an algorithm for fitted Q-iteration with multiple reward functions and continuous state variables whose output gives decision makers insight into how the optimal action depends on decision maker preferences. We call this process "Inverse Preference Elicitation." The new algorithm can handle arbitrary linear regression designs for the Q-function, and is efficient in the sense that it requires computation time linear in the complexity of the stored Q-function. I will also discuss how the output of the algorithm can be used to inform clinical decision making in a way that takes into account the preferences of the decision maker.

Bio:

Dan Lizotte holds a PhD in Computing Science from the University of Alberta, where he studied sequential decision making problems in the areas of classification and global optimization. He is currently a postdoctoral research fellow at the University of Michigan Department of Statistics, where he is working on how best to use medical data to learn effective policies for treating patients.