Christina Schweikert
Reinforcement Learning

May 18, 2004
Dr. Parsons

Notes for Tuesday, May 18, 2004

Paper: "Cobot: A Social Reinforcement Learning Agent"
Presented by: Josh Waxman

Recently, Reinforcement learning has been applied to human-computer interactions, such as spoken dialog systems. The problems faced in this arena include: data sparsity, violations of Markov property, irreproducibility of experiments, users' lack of understanding of how Cobot works, inconsistency of reward, and choosing an appropriate state space. Cobot is an RL-based agent created for LambdaMOO environment. LambdaMOO is a virtual world, multi-user chat environment where users talk and can express actions and emotional states. The long term goal of Cobot is to develop an agent that can learn to perform useful, interesting, and entertaining actions in LambdaMOO on the basis of user feedback. Cobot was originally a Social Statistics agent that monitored how users interact, and only acted in a Reactive manner. Modifications have been made to make Cobot proactive and able to start conversations, introduce users, and word play. Decision making is modeled as a Markov Decision Process since the agent only needs to consider the current state to make a decision. An agent's action effects change in the environment and the agent receives an appropriate reward or punishment. The goal is to maximize expected rewards over time; users give Cobot rewards or punishments according to their feelings about Cobot's actions. Proactive actions of Cobot include: null action, topic starters, roll call, social commentary, and introductions. Cobot chooses an action according to a distribution based on Q values in current state and receives rewards or punishments accordingly. Cobot was trained for the individual user to make learning fast and significant, also to make sure that less interactive users are not overshadowed by more active users. Cobot has been tested with 254 human users with a total of 63123 RL actions and 3171 reward, punishment events. For most users, the policy learned was close to uniform distribution, and policies mostly depended on states. Cobot responded most to those who had most training impact, as far as their preferred reactions from Cobot, etc. Cobot is an example of what can be accomplished by applying Reinforcement learning to create an agent that can learn to interact meaningfully in a human social environment.