CSE 250A Quiz 8

Thursday November 29, 2012

Instructions. You should do this quiz in partnership with exactly one other student. Write both your names at the top of this page. Discuss the answer to the question with each other, and then write your joint answer below the question. It is ok if you overhear other students' discussions, because you still need to decide if they are right or wrong. You have seven minutes.

Consider a Markov decision process (S, A, P, R) where S and A are discrete sets, P is $p(s'|s, a)$, and R is $p(r|s, a, s')$. By analogy with the directed graph for a hidden Markov model viewed as a Bayesian network, draw the directed graph corresponding to this Markov decision process, viewed as a Bayesian network. Provide labels for the nodes and edges. Use the notation "..." (ellipsis) as needed.

Answer. For each time $t \geq 1$, there are three nodes, namely S_t, A_t, and R_t. There are edges from each S_t and A_t to S_{t+1}. The label on these two edges is $p(s'|s, a)$, which is a CPT. There are also edges from each S_t, A_t, and S_{t+1} to R_t. The label on these three edges is $p(r|s, a, s')$, which is also a CPT. The sources of the directed graph are the initial node S_1 and every action node A_t.

Additional comments. The set S is a set of alternative state values, not a set of state variables. The nodes of the graph include state variables S_1, S_2, and so on. Similarly, A is the set of alternative actions, not a set of action nodes.

If the reward at time t depends on the state at time $t + 1$, as stated in the question, then there is an edge from S_{t+1} to R_t. In many MDPs, the reward at time t depends only on the state at time t and the action at time t, as a distribution $p(r|s, a)$. In this case, we only need edges $S_t \to R_t$ and $A_t \to R_t$.

When the MDP is viewed as a Bayesian network, each A_t is a node with no parents, because its value is chosen exogenously, by the agent. The MDP is a model of the environment only, not of the agent. If the agent uses a policy $\pi : S \to A$, and one wants to model the system that combines the agent and the environment, then one can say that there is an edge from each S_t node to the corresponding A_t node.

It does not matter whether the initial state is named S_0 or S_1. For the Bayesian network to be fully specified, the distribution of the initial state must be known. This distribution is usually missing from MDP descriptions.