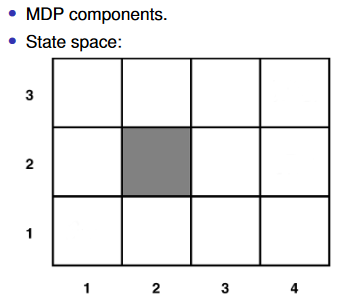
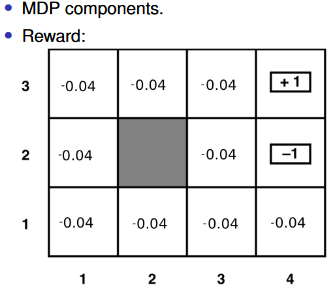
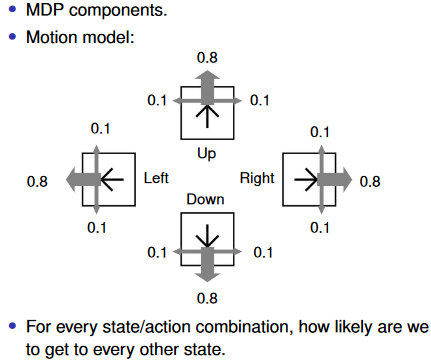
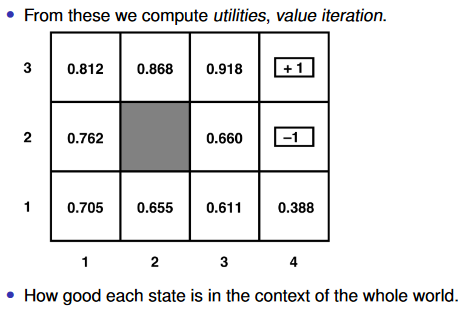
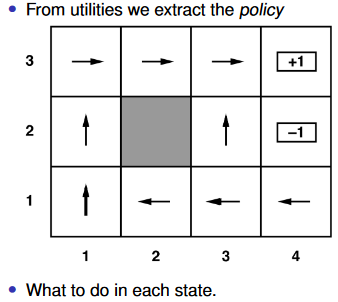
Lecture 9: Reinforcement Learning

Simons Parsons

Markov Decision Processes (MDPs) are methods that can be used to make decisions when actions are non-deterministic

* Decisions about what to do
* Sequential decision problems
* Sequences of decisions
* Use probability and expectation
* 
* 
* If you are in a particular state it has some value to you and the AI gets rewarded to remain in that state or leave that state
* The reward is an intrinsic value of each state.
* 
* 
* 
* A policy is the result of a decision marking(“carry out this action in this state”)

How do we get policies from utilities?

We pick the action with the highest expected utility. This can also have a big downside and so may not be as good as another outcome.

Policy iteration

The utilities are only helpful in getting us to the policy

We can also go direct – policy iteration

Partially observable MDPs