# Reinforcement Learning Notes

## Partially Observable Markov Decision Processes

### The Finite State Partially Observable Markov Decision Processes

#### Notation

Sets

– the set of the finite values of the non-observable process

– the set of the finite “message” space

– the set of all actions at decision maker disposal

Let be a random variable defined on a sample space where ; assume takes on values in the finite set . The stochastic process called the *core process*, is assumed to be a finite state Markov chain with stationary *transition probability matrix* . The core process is completely described by and the initial distribution over denoted by , where . The core process is not directly observable; that is, the realization of is not determinable with certainty at time .

Associated with is a random variable which takes on values in a finite “message” space . By observing at time , information regarding the true value of is obtained. The probabilistic relationship between and is known to the decision maker. Suppose that if , an observation will have message with probability i.e.,

for (1)

Define the *information matrix* . The stochastic process is called the *observation process*.

A decision structure is defined which incorporates the core and observation processes. Assume that the decision maker can control both the observation and core processes by choosing actions. Assume that the decision maker can control both the observation and core processes by choosing actions. Let be a finite set denoting all of the actions available to the decision maker. Let denote the “law of motion” of the core process when action is chosen. That is, if is the current state and action is chosen, the core process moves to a new state with probability . Similarly, let denote the relationship between the observation and core processes when is chosen.

Let and denote the value of observed and the action taken at time , respectively. The data available for decision making at time is denoted by .

Define and let

is called the *information vector*. Using Bayes’ formula, the transformation of the information vector from time to is specified as:

where

# Bibliography

Astrom, K. J. (1965). Optimal Control of Markov Processes with Incomplete State Information. *Journal of Mathematical Analysis and Applications*(10), 174-205.

Braziunas, D. (2003). *POMDP Solution Methods.* University of Toronto, Department of Computer Science.

Davis, M., & Varaiya, P. P. (1971). *Dynamic Programming Conditions for Partially Observable Stochastic Systems.* Memorandum, Electronics Research Laboratory, University of California Berkeley, College of Engineering.

Monahan, G. E. (1982, January). A Survey of Partially Observable Markov Decision Processes: Theory, Models and Algorithms. *Management Science, 28*(1), 1-16.

Sawaki, K., & Ichikawa, A. (1978, March). Optimal Control for Partially Observable Markov Decision Processes Over an Infinite Horizon. *Journal of the Operations Research Society of Japan, 21*(1), 15.

Sawaragi, Y., & Yoshikawa, T. (1970). Discrete-time Markovian Decision Processes With Incomplete State Observation. *The Annals of Mathematical Statistics, 41*(1), 78-86.

Smallwood, R. D., & Sondik, E. D. (1978, Septeber-October). The Optimal Control of Partially Observable Markov Processes over a Finite Horizon. *Operations Research, 21*(5), 1071-1088.

Sondik, E. D. (1978, March-April). The Optimal Control of Partially Observable Markov Processes over the Infinite Horizon: Discounted Costs. *Operations Research, 26*(2), 282-303.

Wakuta, K. (1982, December). Semi-Markov Decision Processes with Incomplete State Observation - Discounted Cost Criterion. *Journal of the Operations Research Society of Japan, 25*(4), 351-361.

Yu, H. (2006). *Approximate Solution Methods for Partially Observable Markov and Semi-Markov Decision Processes.* PhD Thesis, Massachusetts Institute Of Technology, Department of Electrical Engineering and Computer Science, Cambridge, MA.