Markov Decision Process

Depu Meng

Oct. 2018

1 Basic Concepts & Examples

Markov Decision Process (MDP) is a Markov Process that decisions are involved. Generally, an MDP can described by a quintuple:

- A Markov Process or an Extended Markov Process to describe the process.
- A state space.
- An action space.
- A state transition function.
- A performance function.

MDP can be divided into Continuous-Time MDP and Discrete-Time MDP according to the time factor of the Markov Process; MDP can also be divided into MDP and Semi-MDP and Partially-Observable MDP.

1.1 Policy and policy space

In this section, we will take policy and policy space in DTMDP as an example.

A DTMDP quintuple can be denoted as $\{X, \Phi, A, P_{ij}(a), f(i,a)\}$, $X = \{X_n; n \geq 0\}$ is a Discrete-Time Markov Process, $\Phi = \{i\}$ and $A = \{a\}$ are state space and action space of this process respectively. For $P_{ij}(a)$, apparently we have $P_{ij}(a) \geq 0$ and $\sum_{j\in\Phi} P_{ij}(a) = 1$. A DTMDP sample orbit can be described as $\{i_0, a_0, i_1, a_1, ...\}$. Denote $h_n = \{i_0, a_0, ..., i_{n-1}, a_{n-1}, i_n\}$ as the history before time n.

A general policy is defined as

$$v = (v_0(a|h_0), v_1(a|h_1), ...) \tag{1}$$

Infact, a general policy is a series of action defined on decision time, which is also a stochastic policy if not specified. The set that contains all policies like (1) is a policy space, denoted as Π.

For a policy, if for each $v_n(a|h_n)$, we select action a w.p.1, then we call it a determined policy, all determined polices is denoted as Π^d.

For a policy, if each $v_n(a|h_n)$ only related to initial state i_0 and state of time n i_n, i.e., for any n, we have $v_n(a|h_n) = v_n(a|i_0, i_n)$, then we call it a semi-Markov policy, denoted as Π_{sm}.