

observation

O_t

r_t

reward

r_{t+1}

O_{t+1}

Agent

Environment

action
 a_t

