

Örjan Ekeberg	Brain Modeling and Machine Learning	Örjan Ekeberg	Brain Modeling and Machine Learning
Known Environment Unknown Environment	Bellman's Equation Solving Techniques	Known Environment Unknown Environment	Bellman's Equation Solving Techniques

1 Known Environment

- Bellman's Equation
- Solving Techniques

2 Unknown Environment

• Monte-Carlo Method

Model of the Environment

• Where does an action take us?

$$\delta(s,a) \mapsto s'$$

• How much reward do we receive?

 $r(s, a) \mapsto \Re$

The values of different states are interrelated *Bellman's Equation*:

$$\mathcal{V}^{\pi}(s) = r(s, a) + \gamma \cdot \mathcal{V}^{\pi}(\delta(s, a)) \qquad ext{where } a = \pi(s)$$

Known Environment Solving Techniques

Can we solve Bellman's equation?

$$V^{\pi}(s) = r(s, a) + \gamma \cdot V^{\pi}(\delta(s, a))$$
 where $a = \pi(s)$

• Direct solution (linear equation system)

Örjan Ekeberg

Unknown Environment

• Iteratively (value iteration)

Known Environment • Bellman's Equation • Solving Techniques

2 Unknown Environment Monte-Carlo Method

$$V_{k+1}^{\pi}(s) \leftarrow r(s,\pi(s)) + \gamma \cdot V_k^{\pi}(\delta(s,\pi(s)))$$

Brain Modeling and Machine Learning

How do we find an optimal policy π^* ? Easy if the optimal value function V^* was known:

$$\pi^{\star}(s) = \arg\max_{a} \left(r(s, a) + \gamma \cdot V^{\star}(\delta(s, a)) \right)$$

Optimal version of Bellman's equation

$$V^{\star}(s) = \max_{a} \left(r(s, a) + \gamma \cdot V^{\star}(\delta(s, a)) \right)$$

Hard to solve

Policy iteration:

Interleaved calculation of policy and values

Örjan Ekeberg Brain Modeling and Machine Learning

Monte-Carlo Method Unknown Environment





Normal scenario: Unknown environment r(s, a) and $\delta(s, a)$ are not known

 V^{π} must be estimated from experience

Monte-Carlo Method

- Start at a random *s*
- Follow π , store the rewards and s_t
- When the goal is reached, update V^π(s)-estimation for all visited states with the future reward we actually received

Very slow convergence

Örjan Ekeberg Brain Modeling and Machine Learning