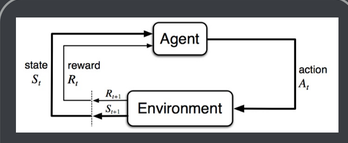
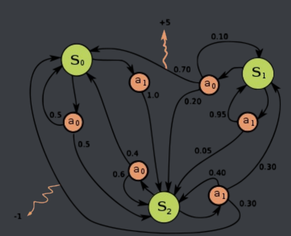
Reinforcement Learning in a nutshell

Basic RL Framework



Block Digram

Math Framework



Markov Decision Process

Details

$\{(S_t, A_t, R_{t+1}, S_{t+1})\}$

Navigating the MDP Graph produce traces of the form

$P(S_{t+1}, R_{t+1}|S_t, A_t)$

MDP Dynamic

Probabilistic Representation is more general as it includes the deterministic one as a special case where distributions are Dirac Delta

$\pi(A|S)$

Policy

Probabilistic Representation is more general as it includes the deterministic one as a special case where distributions are Dirac Delta

Assume you know the First Order MDP Graph exactly which is equivalent to knowing $P(S_{t+1}, R_{t+1}|S_t, A_t)$ then navigating it with a given policy $\pi(A|S)$ produce traces $\{(S, A, R)_t\}_{t\in[0,T]}$, which means collecting a set of Rewards $\{R_t\}_{t\in[0,T]}$.

Navigating the MDP with a Policy

Discounted Reward

So the policy π in a state s will get this collection $\{R_t\}_{t\in[0,T];\pi;s}$

Collecting Rewards

We can then define the Discounted Reward $G(\tau;\gamma,\pi,s) = \sum_{t=0}^{\infty} \gamma^t R_t$, where γ is theit discount factor for the future rewards, as the function projecting into the present the amount of Rewards the policy will be able to collect starting from that state

Projecting the Amount of Rewards the Policy is able to collect into the present

Hypothesis

> ***The Reward Hypothesis***: *That all of what we mean by goals and purposes can be well thought of as the maximization of the expected value of the cumulative sum of a received scalar signal (called reward).*

Reward Hypothesis

Questions

How efficient is to compress all of the useful info for the agent into a single scalar

Important implications

Reward Design

Reinforcement Learning in a nutshell

1. Basic RL Framework

1.1. Block Digram

2. Math Framework

2.1. Markov Decision Process

2.1.1. Details

2.1.1.1. Navigating the MDP Graph produce traces of the form

2.1.1.2. MDP Dynamic

2.1.1.2.1. Probabilistic Representation is more general as it includes the deterministic one as a special case where distributions are Dirac Delta

2.1.1.3. Policy

2.1.1.3.1. Probabilistic Representation is more general as it includes the deterministic one as a special case where distributions are Dirac Delta

2.1.1.4. Discounted Reward

2.1.1.4.1. Navigating the MDP with a Policy

2.1.1.4.2. Collecting Rewards

2.1.1.4.3. Projecting the Amount of Rewards the Policy is able to collect into the present

3. Hypothesis

3.1. Reward Hypothesis

3.1.1. Questions

3.1.1.1. How efficient is to compress all of the useful info for the agent into a single scalar

3.1.2. Important implications

3.1.2.1. Reward Design