Final Report

Daniel Yao

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit.

1. Introduction

Def. A finite Markov Decision Process (MDP) is a five-tuple (S, A, P, R, γ) where[1][2]

1. S is the finite state space,
2. $A(s)$ is the finite action space for state $s \in S$,
3. $P : S \times A \times S$ is the transition probability function,
4. $R : S \times A \times S$ is the reward function, and
5. $\gamma \in [0, 1]$ is the discount factor.

$P(s' \mid s, a)$ is the probability that the next state is $s' \in S$ given that the current state is $s \in S$ and the action taken is $a \in A(s)$. $R(s', a, s)$ is the reward received when the current state is $s' \in S$, the action taken was $a \in A(s)$, and the previous state was $s \in S$.

Def. A policy π is a function $\pi : A \times S \to [0, 1]$ where $\pi(a \mid s)$ is the probability that an agent in state $s \in S$ takes action $a \in A(s)$. This is a probability distribution, so

$$\sum_{a \in A(s)} \pi(a \mid s) = 1$$

for all $s \in S$.

Def. The discounted return G_t at time t is the sum of all future rewards, discounted by the factor γ. That is,

$$G_t = \sum_{k=1}^{\infty} \gamma^k R_{t+k}$$

where R_{t+k} is the reward received at time $t + k$.

Def. The state-value function $V_\pi(s)$ is the expected return when starting in state s and following policy π:

2. Markov Decision Process

Lorem ipsum dolor sit amet, consectetur adipiscing elit.

3. Reinforcement Learning

4. Simulation Study

Lorem ipsum dolor sit amet, consectetur adipiscing elit.

5. Discussion

Lorem ipsum dolor sit amet, consectetur adipiscing elit.

6. Conclusion

Lorem ipsum dolor sit amet, consectetur adipiscing elit.

References

[1] P. Brothers, Risk: The Classic World Domination Game (1993). URL `https://www.hasbro.com/common/instruct/risk.pdf`

[2] M. L. Puterman, Markov decision processes: discrete stochastic dynamic programming, John Wiley & Sons, 2014.