

Decision Making in Intelligent Systems

Deeltoets, Bachelor AI, Universiteit van Amsterdam

25 maart 2008, 9-12am

Voor alle vragen geldt: antwoord in je eigen woorden, dus kopieer/vertaal niet rechtstreeks uit de literatuur, en leg je tussenstappen en redenen voor antwoorden uit.

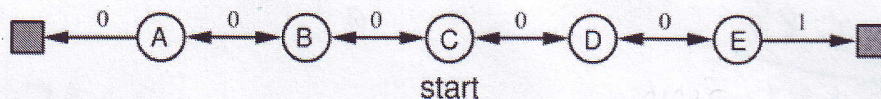
Vraag 1

1. Beschrijf het verschil tussen rewards en returns.
2. Wat is de relatie tussen aan de ene kant value functions en aan de andere kant rewards en returns?
3. Stel de discount parameter $\gamma = 0$. Wat is dan het verschil tussen rewards en returns? Beschrijf een voorbeeld van een reinforcement learning taak waarin je zo'n discount parameter $\gamma = 0$ zou gebruiken.

Vraag 2

1. Beschrijf wat bootstrapping is, alsmede de voordelen van bootstrapping versus het niet gebruiken van bootstrapping in reinforcement learning.
2. Geef één mogelijke value function update formule die gebruik maakt van bootstrapping, en één mogelijke value function update formule die geen gebruik maakt van bootstrapping.
3. Als je te maken hebt met een reinforcement learning probleem waarbij het MDP model niet gegeven is, het probleem leidt tot hele lange episodes, maar waarbij na bijna elke actie en state overgang een sterk verschillende immediate reward gegeven wordt, zou je dan een bootstrapping algoritme adviseren of niet? Leg uit.

Vraag 3



Beschouw het kleine episodische MDP van deze figuur. De agent begint altijd in state C, en de episode eindigt wanneer één van de terminal states, de grijze vierkanten, bereikt wordt. Alle immediate rewards zijn 0 behalve als de agent van state E naar de rechter terminal state gaat; daar is de immediate reward 1. Zoals gebruikelijk hebben de terminal states zelf geen value. Ga uit van een vaste policy π die, gecombineerd met de state transition function, met gelijke kans de agent elke tijdstap doet verspringen van een state naar één van de twee aangrenzende states: een random walk. Er is een discount parameter γ . Gegeven is de value van state C onder policy π : $V^\pi(C) = 0.2580$, en de value van state E onder policy π : $V^\pi(E) = 0.6925$. Wat is de waarde van γ ?