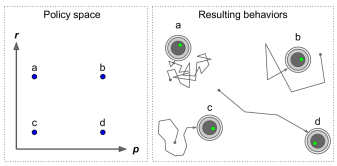
Systemy agentowe

Przykład polityki



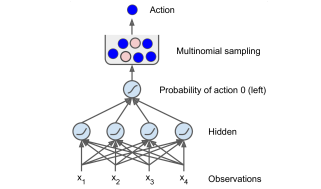
A. Géron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow* 2017

p to prawdopodobieństwo jechania prosto (czyli $1 - p$ to pr. zakrętu), a r to kąt zakrętu

2017-12-17

Systemy agentowe

Polityka za pomocą sieci neuronowej



A. Géron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow* 2017

Losowanie, żeby zapewnić eksplorację

Q-Value iteration

- $Q_k(s,a)$ wartość akcji a w stanie s w kroku k
- $T(s,a,s')$ prawdopodobieństwo przejścia $s \to s'$ przy akcji a
- $R(s,a,s')$ nagroda za przejście $s \to s'$ przy akcji a
- γ discount ration

$$Q_{k+1}(s,a) \leftarrow \sum_{s'} T(s,a,s') \left[R(s,a,s') + \gamma \max_{a'} Q_k(s',a') \right]$$

$$\pi^*(s) = \arg\max_a Q_\infty(a)$$

T i R są nieznane, a do tego liczba możliwych stanów zwykle będzie gigantyczna, więc nie da się wykonać odpowiednich obliczeń.

Q-Learning

$$Q_{k+1}(s,a) \leftarrow (1-\alpha) Q_k(s,a) + \alpha \left(r + \gamma \max_{a'} Q_k(s',a') \right)$$

α to prędkość uczenia. Wciąż niepraktyczne, bo większości par (s,a) nigdy nie odwiedzimy.