policy (reinforcement learning)

Authors

Alexander Jung*, Department of Computer Science, Aalto University, Espoo, Finland
*Corresponding author: alex.jung@aalto.fi

Abstract

A policy in reinforcement learning (RL) is a function that specifies which action a_t to take in a Markov decision process (MDP) when the current state is s_t. A policy is often stochastic: it defines a conditional probability distribution $\mathbb{P}^{(a|s)}$ over the actions given the state, and an action is drawn from it. A policy is the hypothesis of RL: it uses features derived from the state to predict the best next action, and it is learned to maximize the expected cumulative reward. The value function of a policy is the expected cumulative reward obtained by following it from a given state.

Definition

A policy is a function that specifies how the next action a_t in a Markov decision process (MDP) is chosen when the current state is s_t. Typically, a policy is stochastic, meaning that it defines a conditional probability distribution $\mathbb{P}^{(a|s)}$ over the actions for a given current state. A policy can also be viewed as a hypothesis that uses features derived from the current state to predict the best next action (Sutton and Barto, 2018).

Cross References

action, MDP, state

References

1. Sutton and Barto (2018). Reinforcement Learning: An Introduction. 2nd ed., MIT Press, Cambridge, MA, USA.