687 2017-09-05

Note Title 9/5/2017

Based on Sutton & Barto 1998. Area of ML

Inspired by behaviorist psych: how an agent can learn from (a sequence of) interactions with an environment.

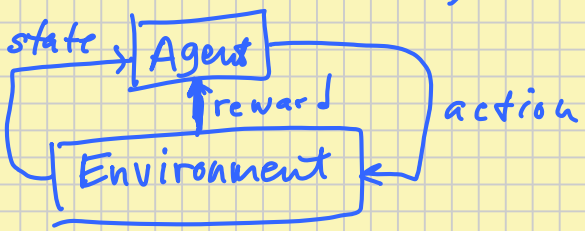
ML → Supervised, Unsupervised, Reinforcement

Key: Learn from evaluative feed back.

Not told what to do.

Operant conditioning ≈ Reinforcement learning

Agent-Environment Diagram



Q: Seems discrete - what about continuous time?

Relu to neuroscience: Where does reward come from?

Computational Neuroscience
Control Theory
etc.

↳ How do animals learn? One agent vs set of agents

vs. How can we make an agent that learns?

Can we evaluate effectiveness of different learning methods?

Ex of an RL Q: Model based vs Model free approaches

Donald Michie's Menace → Machine Educable Noughts & Crosses [Tic-Tac-Toe]

- sequence of matchboxes + colored beads

Rational Pirates Puzzle
Can develop solution by induction.

Example - Gridworld

Agent in a grid square with a) obstacles b) unpleasant places

Actions:
up, down, left, right
Succeeds w/ pr = 0.8
w/ pr .05, add 90° to direction
pr .05, subtract 90°
pr .10, does not move
Also, does not move at wall/obstacle

Model w/ a Markov Decision Process (MDP):
Math. Spec. of env. + what we want agent to learn ⇒ Problem Description

$t \in \{0, 1, \ldots\}$ time step
S_t : state of env. at time t Capital letters: random vars
A_t : action taken by agent at time t
R_t : reward given to agent at time t

MDP $M = (\mathcal{S}, \mathcal{A}, P, R, d_0, \gamma)$ $\mathcal{S}$ = set of all poss. states
$\mathcal{A}$ = set of poss. actions