Reward shaping

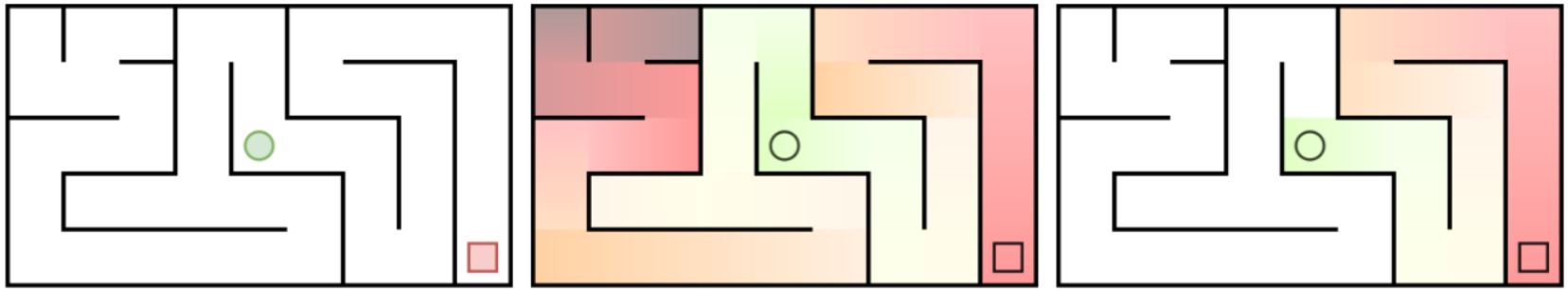
Reward shaping

Ostbayerische Technische Hochschule Amberg-Weiden

Reward shaping

- covers all methods that modify the reward given to the agent in order to optimize (speed up) training
- incorporates prior knowledge about the task into the reward function

Example: Different levels of reward shaping depending on the distance to the goal



https://arxiv.org/pdf/2210.09579.pdf

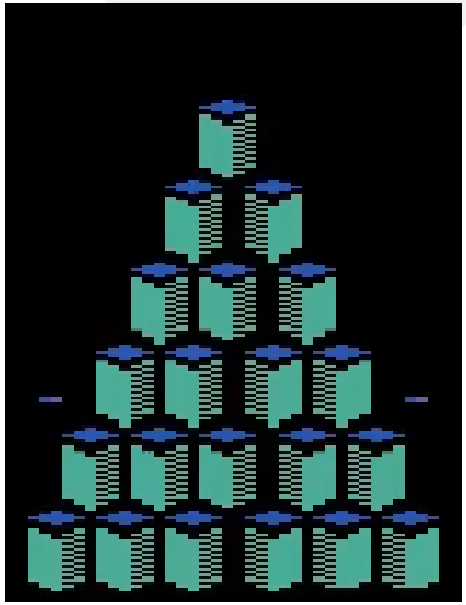
Reward shaping

Problem with RL in general

- RL is extremely good in finding bugs/loopholes in your program, this effect may get aggravated through reward shaping

Example 1: Q-Bert

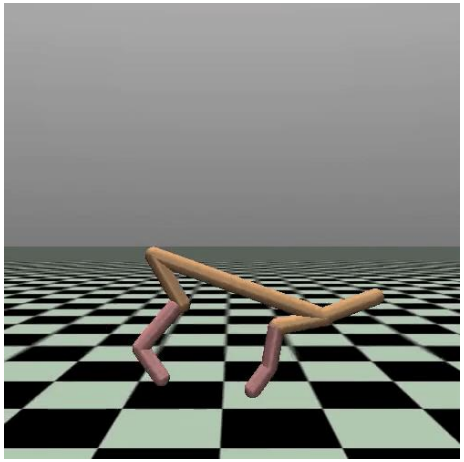
- Old Amiga game, RL agent found a (previously unknown) bug during training

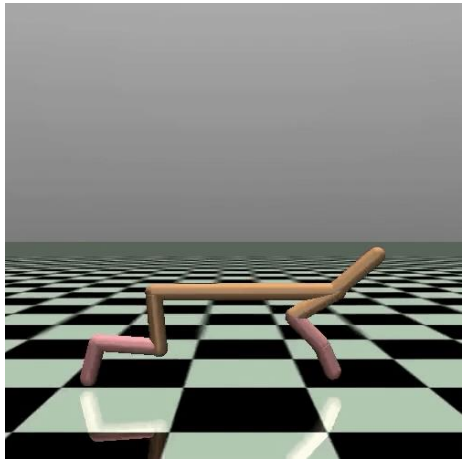


https://www.youtube.com/shorts/uVHAAIA3F_U

Example 2: Half Cheetah

- Through optimized parameters and a more positive reward if the agent moves faster to the roll, it learned that it is faster to roll instead of running





https://bair.berkeley.edu/blog/2021/04/19/mbrl/

Reward shaping

Problem with reward shaping: Cobra effect

- Modifying a (initially simple) reward function in order to speed up training may lead to undesired consequences

Example: CoastRunners

- Real goal: Finish the race as quickly as possible
- Reward assigned to agent: Collect as many targets along the way

https://www.youtube.com/watch?v=tlOIHko8ySg&t=5s

Reward shaping

Ostbayerische Technische Hochschule Amberg-Weiden

Types of rewards

immediate rewards get reward immediately after taking an action	vs.	**delayed rewards** get reward some time after taking an action
dense rewards get reward every timestep	vs.	**sparse rewards** do not get reward every timestep
positive rewards rewards $R_t \geq 0$	vs.	**negative rewards** rewards $R_t \leq 0$
individual rewards focus on one task aspect	vs.	**additive rewards** focus on many task aspects

Reward shaping

Ostbayerische Technische Hochschule Amberg-Weiden

Immediate vs. delayed rewards

Immediate rewards

- Get a reward (associated with that action) immediately after taking the action
- Examples
  - Autonomous driving: Distance to target
  - Learning to walk: Distance traveled

Delayed rewards

- Get a reward (associated with that action) some time after taking the action
- Examples
  - Games (get a reward of e.g. +1/-1 when winning/losing the game)
  - Stock market (buy stocks now, see how it performs over the years)

→ Delayed rewards are way more difficult for RL than immediate rewards, training can be sped up if suitable immediate rewards are added to delayed rewards.

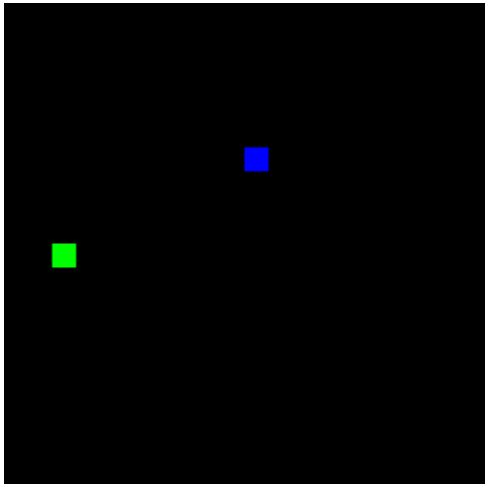
Reward shaping

Ostbayerische Technische Hochschule
Amberg-Weiden

Dense vs. sparse rewards

Dense rewards

- Get a reward at every timestep
- Examples
  - Robot maze
  - Autonomous driving: Distance to target update every second



https://medium.com/ml-everything/reinforcement-learning-with-sparse-rewards-8f15b71d18bf

Sparse rewards

- Don't get a reward at every timestep (e.g. every N steps)
- Examples
  - Snake
  - Soccer: Goals scored

→ Sparse rewards are usually more difficult than dense rewards but by far not as difficult as massive delayed rewards

Reward shaping

Ostbayerische Technische Hochschule Amberg-Weiden

Positive vs. negative rewards

Positive rewards

- Get a positive reward when taking a good action or reaching the goal state
- Examples
  - Learning to walk: Distance traveled

Negative rewards

- Get a negative reward for every action that does not reach the goal state
- Examples
  - Robot maze

→ Negative rewards encourage the agent to reach a goal state as quickly as possible, positive rewards may stimulate the robot to accumulate reward rather than reaching the goal state (useful for e.g. non-episodic tasks)

Reward shaping

Ostbayerische Technische Hochschule Amberg-Weiden

Individual vs. additive rewards

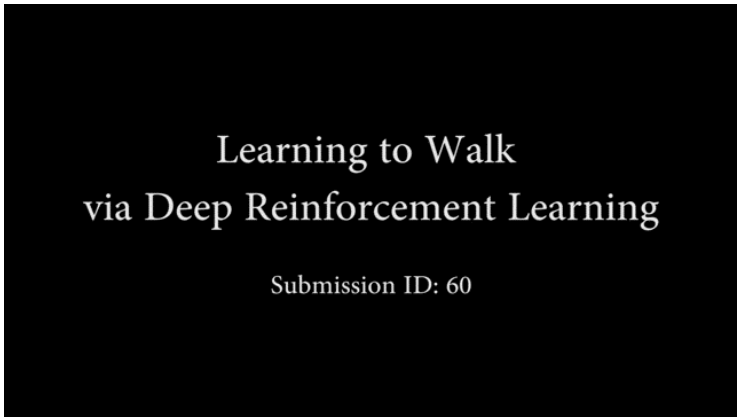
Individual rewards

- Rewards focusing only on the main aspect of the task
- Examples
  - Learning to walk: Distance travelled

Additive rewards

- Rewards focusing on multiple task aspects (with $\alpha, \beta, \ldots$ as weights): $R = \alpha R_a + \beta R_b + \cdots$
- Examples
  - Learning to walk: Distance travelled and torso height

→ Additive rewards can be a good addition to individual rewards if learning progress is slow



https://www.youtube.com/watch?v=n2gE7n11h1Y

Reward shaping

Ostbayerische Technische Hochschule
Amberg-Weiden

Task: How would you shape an reward for

- running (with a simulated human)
- driving a car
- playing chess

Reward shaping

Example: Robot maze