Meta Reinforcement Learning

Definition of Meta-RL[1]

Marius Lindauer

tnt

Leibniz Universität Hannover

[1] Based on a blog by Lilian Weng

RL Problem Setting

- Definition of MDP (S, A, P, R, γ)
  - S is a (finite) set of Markov states $s \in S$
  - A is a (finite) set of actions $a \in A$
  - P is dynamics/transition model for each action, that specifies $P(s_{t+1} = s' \mid s_t = s, a_t = a)$
  - R is a reward function $R(s_t = s, a_t = a) = \mathbb{E}[r_r \mid s_t = s, a_t = a]$
    - ⋆ Sometimes R is also defined based on (s) or on (s, a, s')
  - Discount factor $\gamma \in [0, 1]$

- Task: Compute the optimal policy

$$\pi^*(s) \in \arg\max_{\pi} V^{\pi}(s)$$

Meta-RL Problem Setting

- No single formal problem setting
- Usually based on a set of different MDPs $\mathcal{M}$
- Commonalities in $\mathcal{M}$ vary
- Often: shared state & action space
- Task: Compute optimal policy over all of $\mathcal{M}$

Example Setting 1: Two Walking agents

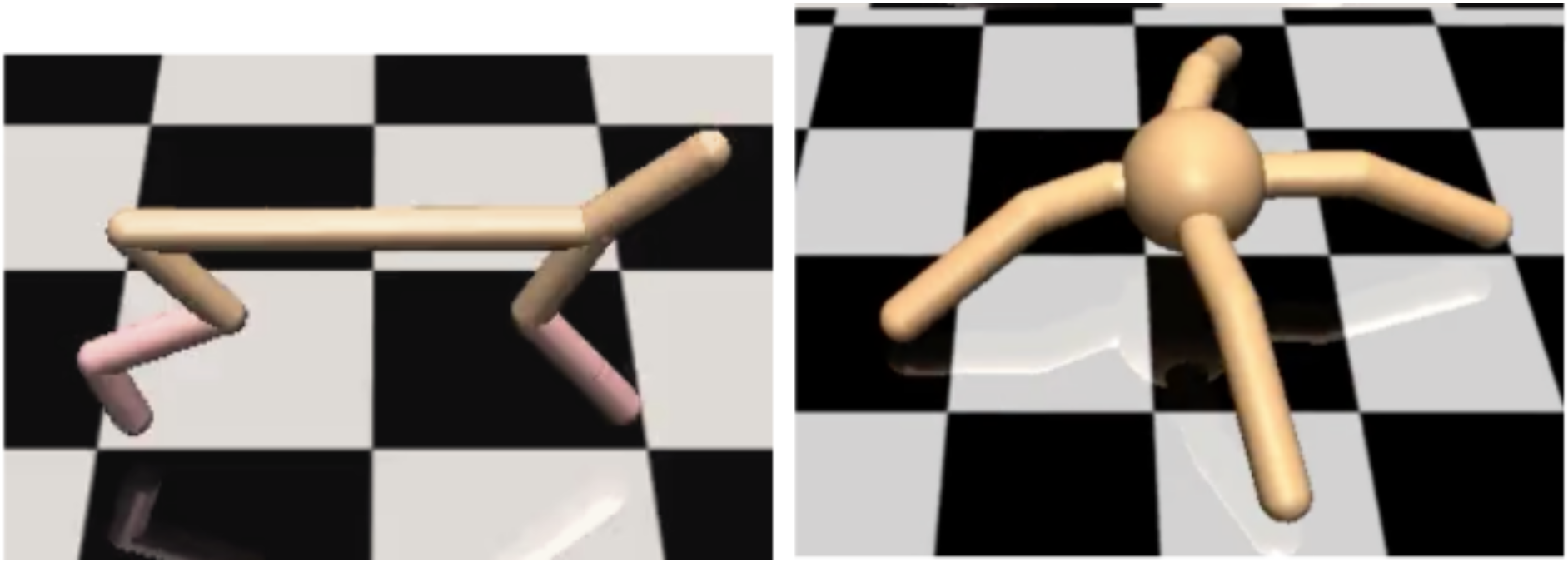

Figure: Two robots with different behavior.[2]

[2]Image Source: MAML

Example Setting 2: 100 Mazes

Figure: 100 Mazes, each with a different goal to reach.[3]

[3] Image Source: VDS

Example Setting 3: 200 Mazes

Figure: Mazes of two layouts and 100 goals each.[4]

[4] Image Source: VDS

Key Approaches in Meta-RL

- Meta-Learning Hyperparameters
  - Hyperparameter values for good performance
  - Alternative optimization methods
- Meta-Learning the Training Dynamics
  - Credit Assignment for state-action pairs
  - Problem-specific Exploration Strategies
- Task Generation
  - Task diversification for generalization
  - Curriculum Learning