Meta Reinforcement Learning

The Big Picture

Marius Lindauer

Winter Term 2021

Can We Generalize beyond a given MDP?

- What happens if the environment changes? (non-stationary environments)
  - Can we efficiently adapt our policy to changed transitions or reward functions?
- After a human player learned how to play Super Mario Bros in the first levels, they will also play fairly well the upcoming levels.
- However, an RL agent potentially will fail.

⇝ Strong limitations regarding the applications of a trained agent

Main Ideas

1. Can we use easy environments to learn how to behave in hard environments?

Main Ideas

1. Can we use easy environments to learn how to behave in hard environments?
2. Can we train a policy that is easily adaptable to new environments?

Main Ideas

1. Can we use easy environments to learn how to behave in hard environments?
2. Can we train a policy that is easily adaptable to new environments?
3. Can we find better training dynamics across a set of environments?

Main Ideas

1. Can we use easy environments to learn how to behave in hard environments?
2. Can we train a policy that is easily adaptable to new environments?
3. Can we find better training dynamics across a set of environments?
4. Can we train a policy that generalizes to new environments without any new training?

- Assumption: We sample our environments i.i.d. from a fixed distribution
  - Similar to the assumption in supervised learning, but on a meta-level
  - Training environments to train our agent on and test environments to check how well it performs.
  - We might have control how we sample from this distribution;
    we might don't.