

Curriculum Reinforcement Learning^a

Marius Lindauer



Winter Term 2021

^aBased on a blog by Lilian Weng

From Easy to Hard

- ▶ We humans also learn step by step
 - ▶ E.g., in math, we learn first basic arithmetics before we later learn complex derivatives and integrals
 - ▶ E.g., in this lecture, we first talked about simple planning on MDPs, before we talked about complex meta-RL ideas
- ▶ **Idea:** break down complex concepts into simpler concepts s.t. we can start from the easy ones and build up the complex one
- ▶ **Challenge:** How can we design a curriculum starting from simple to hard tasks?
 - ▶ A poorly designed curriculum might even harm learning.
- ▶ **Challenge:** How do we avoid catastrophic forgetting by training on another instance?

Task-Specific Curriculum Learning [Bengio et al. 2009]

1. Cleaner Examples may yield better generalization faster.
 2. Introducing gradually more difficult examples speeds up online training.
- ▶ Results by [Zaremba and Sutskever. 2014] indicated that one should mix in easy tasks to not forget how to solve these.

How to Quantify Complexity/Difficulty of an Env?

- ▶ Human specification

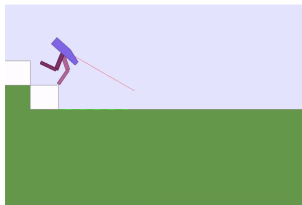
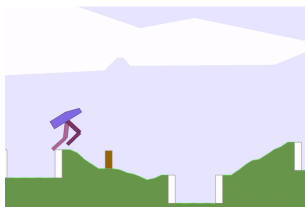
- ▶ parameterized generator for environments that allow to control the complexity by hand
- ▶ or we can measure meta-features of the different envs, e.g.
 - ▶ Size of maze
 - ▶ distance between start state and end state
 - ▶ Fraction of floor space to walls
 - ▶ ...
- ▶ Note: We cannot quantify the complexity of the optimal policy for a given MDP, because we have to measure these before actually solving the MDP

Automatic Curriculum Generation

- ▶ In contrast to a hand-designed curriculum before training starts, the curriculum is generated on the fly for any instances in the domain
- ▶ Criteria for the ordering have to be provided or computed during runtime
- ▶ While criteria need to be computed during runtime, they do not need to be hand-designed

Example Criterion 1: Manual Difficulty Ratings

1. Idea: define a function to rate env difficulty
2. Example: POET [Wang et al. 2019]¹



¹Image source: Uber AI

Example Criterion 2: External Model

1. Minimal loss wrt another policy being pretrained on other tasks as criterion
2. Sort the instances s.t. they go from easy to hard wrt this pretrained agent [Weinshall et al. 2018]