Exploration for RL

Inductive biases in exploration strategies

Alexander Telfar

June 30th, 2019

What is RL?

Reinforcement learning is a collection of solutions to problems that;

- allow, evaluations, no feedback,
- have delayed credit assignment. (usually, but not necessarily)

Example: Bandits

The two armed bandit is the simplest problem in RL.

1. [10, -100, 0, 0, 0]
2. [2, 0]

Which arm should I pick next?

Why do exploration strategies matter?

Why not just do random search?

insert pic

- Too much exploration and you will take many sub optimal actions, despite knowing better.
- Too little exploration and you will take ‘optimal’ actions, at least you think they are optimal. . .

An example: Minecraft!

Crafting is super imporant. But has a combinatorial nature. We bring many priors to help us. We know that;

- iron is useful for making tools.
- coal and a furnace is probably needed to make iron.
- ?

What is an inductive bias?

Underconstrained problems.

Why might this matter in exploration?

Example: Matrix factorisation

Lowest rank solution

What do we require from an exploration strategy?

- Non-zero probability of reaching all state, and trying all actions in each state.
- Converges to a uniform distribution over states. (?)
- ?

Nice to have

- Scales sub-linearly with states
- ?

What are some existing exploration strategies?

- Counts / densities
- Intrinsic motivation (Surprise and Novelty)
- Optimism in the face of uncertainty
- Max entropy

Note. They all require some form of memory. Exploration without memory is just random search...

Counts / densities

In the simplest setting, we can just count how many times we have been in a state. We can use this to pick states that have have low visitation counts.

$$P(s = s_t) = \frac{\sum_{s=s_t} 1}{\sum_{s \in S} 1}$$

$$a_t = \operatorname*{argmin}_a P(s = \tau(s_t, a))$$

Intrisnic motivation

'Surprise'

$$r_t = \| \ s_{t+1} - f_{dec}(f_{enc}(s_t, a_t)) \ \|_2^2$$

'Novelty'

$$r_t = \min_{x \in M} D(s_t, x)$$

Inductive biases in exploration strategies

So my questions are;

- do some of these exploration strategies prefer to explore certain states first?
- what inductive bases do we want in exploration strageties?
- ?

Thank you!

And questions?