Reinforce vs DQN: Does REINFORCE converge faster in Atari games?

Niklas Baack
https://github.com/nikl234

Oscar Tylkowsky
https://github.com/ostylk

Thibaut Klenke
https://github.com/thibautklenke

1 Motivation

As we learned that policy-gradient methods have better convergence garantees, we want to create a study over mulitple games in which we compare REINFORCE(policy-gradient methods) with DQN and see if the convergence garantees hold true.

2 Related Topics

What does this relate to from the lecture or the literature? Mention 2-3 concepts or cite 2-3 papers that are relevant to what you want to do.

3 Idea

Briefly describe your idea. We do not necessarily need details at this stage, but a rough notion of what you want to accomplish and why.

Algorithm 1 A great RL algorithm.

Require: environment e, algorithm A **return** policy π
  while TRUE **do**
    Train A on e
  end while

$$\pi \in \Pi, \pi : \mathcal{S} \mapsto \mathcal{A} \tag{1}$$

4 Experiments

We want to know roughly what you are planning to show in terms of experiments.

Environments & Metrics What environments will you consider and what are you planning to measure?

We plan to take between 10 and 50 environments from the atari games library. We want to measure over many seeds how long it will take to converge. Convergence is measured by the percentual change of the loss over mulitiple episodes.

Experimental Scope How many experiments are you running? Include seeds, hyperparameter optimization, different environments, ablations, etc. here.

Algorithms=2 Environments=ca 30 Seeds=ca 50 NN_Sizes=ca 3

We plan to run at least: 2x30x50x3=900 runs

Estimated Computational Load How long will your experiments take and how will you run them? These do not need to be exact numbers, but we want to see that you know roughly how much time, memory and other resources you will need.

900 runs x 10min = 15 hours As we should be able to parallelize the task it should be as fast as 15h/8(cores) $\approx$ 90min

As the atary games have a low memory footprint, memory should not be a problem.

We dont plan to make the NN really large, therefore RAM or VRAM shouldnt be a problem either.

This is a confirmatory research project.

We believe that REINFORCE will converge significant faster than dqn(a=0.05).

5 Timeline

Please tell us how long you think it will take to accomplish all parts of your project. This includes:

- Research
- Implementation
- Experiments
- Analysis
- Reporting