Paper summary: exploration by random network distillation

Marko Guberina 19970821-1256

March 16, 2022

1 Idea in few sentances

2 Explanation of the central concept

3 Methodology

4 Initial rambly notes

4.1 Abstract

Introduces exploration bonus for DRL which is easy to implement and adds minimal computational overhead to computation. The bonus is the error of a neural network which predicts features of observations given by a fixed randomly initialized neural network. The bonus is called random network distillation (RND). Also introduces method which combine intrinsic and extrinsic rewards. Achieve best results up until then on Montezuma's revenge.

4.2 Introduction

It's difficult to engineer dense rewards and sparse rewards are difficult to learn. To deal with that you need an to explicitely explore.

Vectorized environments (running the policy in parallel in the same environment) is great to solve challenging RL tasks. However exploration methods based on (pseudo)counts don't scale well to a large number of parallel environments. The method introduced in the paper only requires a single forward pass through a neural net per 1 experience batch.

Key observation: Neural networks tend to have significantly lower prediction errors on examples similar to those on which they have been trained.

Idea based on the observation Use prediction errors of networks trained on the agent's past experience to quantify the novelty of new experience.

Just using prediction loss when learning forward dynamics makes your agent stare at stochastic elements of the environment. Amending this by quantifying only the relative improvement of the prediction, rather the its absolute error, is hard to implement efficiently.

This method predicts the output of a fixed randomly initialized (deterministic) neural network on the current observation.

The method augments PPO so that it uses two value heads for the two reward streams.

4.3 Method