Paper summary: Data-efficient reinforcement learning with self-predictive representations

June 1, 2022

1 Idea in few sentances

2 Explanation of the central concept

3 Methodology

4 Initial rambly notes

4.1 Abstract

Learning from a small number of examples is a problem for RL. Proposal: leverage self-supervised objectives based on visual input and sequential interaction with the environment. The resulting method, Self-Predictive Representations (SPR), trains an agent to predict its own latent space representations multiple steps into the future. Target representations for future states are computed with an encoder which is an exponential moving average of the agent's parameters, and the predictions are made using a learned transition model. This alone outperforms other methods, but is made even better by adding data augmentation to the future prediction loss. The benchmark is 100K Atari (400K frames).

4.2 Introduction

Dyamics model does not rely on reconstructing states. The algorithm starts with Data-efficient rainbow with the SPR loss and is evaluated with and without data augmentation. Notably, SPR outperforms human expert on 7/26 games using the same amount of experience.

4.3 Method

4.3.1 Self-Predictive representations

Let $(\boldsymbol{s}_{t:t+K}, \boldsymbol{a}_{t:t+K})$ be a sequence of $K+1$ previously experienced states and actions sampled from a replay buffer, where K is the maximum number of steps

into the future which we want to predict.

Online network An *online encoder* f_o transforms observed states $\boldsymbol{s}_t$ into representatins $\boldsymbol{z}_t \triangleq f_o(\boldsymbol{s}_t)$. These representations are used in an objective that encourages them to be *predictive* of future observations up to some fixed temporal offset K, given a sequence of K actions to perform.

Rather than predicting representations produce by the online encoder, it JUST FUCKING BOOTSTRAPS EVERYTHING?!?!?!? This is some crazy stuff man. The bootstrapping is explained in: bootstrap you own latent: a new approach to self-supervised learning. Will have to look into this more

4.4 Other stuff