Paper summary: Contrastive unsupervised representations for reinforcement learning (CURL)

May 20, 2022

1 Idea in few sentances

2 Explanation of the central concept

3 Methodology

4 Initial rambly notes

4.1 Abstract

CURL extracts high-level features from raw pixels by using contrastive learning and performs off-policy control on top of the extracted features. It works well.

4.2 Introduction

To make image-based RL more efficient, people tried:

1. do auxiliary tasks on observations
2. learn world models that predict the future

CuRL is in the first camp of methods and it uses contrastive learning to learn the representations.

4.3 Method

The architecture works like follows. A batch of samples is taken from the replay buffer. A stack of frames (4 for Atari, 3 for continuous control) is as single sample. Two different image transformations are done on the batch, but the same transformation is applied accross a single stack of frames. One augmented batch goes to the query encoder, while the other goes to the key encoder. Contrastive

unsupervised learning is the applied to those encoders. The batch of feature vectors from the query encoder is then used as an input to the reinforcement learning algorithm. Those gradients do go through the query encoder (altough this isn't strictly necessary, you can sacrifice performance for cross-task representations). Instance discrimination similar to SimCLR, MoCo and CPC is used and it is performed accross frame stacks. Momentum encoding similar to Moco is used and that's applied on keys. A bi-linear inner product is used to compare keys and queries. If I read this right, the key and query network are the same network.

4.4 Other stuff