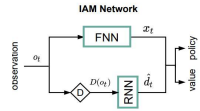
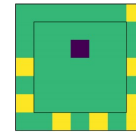
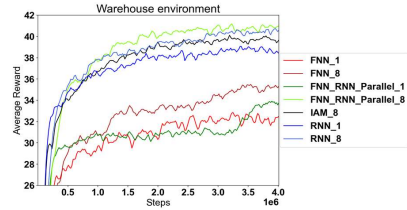
Implementing Influence-aware Memory Architectures for Deep Reinforcement Learning

Virancab · 9 min read · Draft

Influence-aware Memory Architectures for Deep Reinforcement Learning is a deep learning algorithm resembling more popularly known transformer architectures in an actor-critic setting, using a Proximal Policy Optimization algorithm to evolve agents in various environments.







Implementing Influence-aware Memory Architectures for Deep Reinforcement Learning in Pytorch

Viranca Balsingh, Kevin Bislip, Taylan Turan
CS4240 (Deep Learning), April 2021 - Delft University of Technology

Goal of paper

The original paper tries to improve training and convergence difficulties of RNN's with influence aware memory (IAM). IAM achieves this by limiting the input of the recurrent layers to variables that influence the hidden state information, using a so-called D-set.

Architecture and technique

IAM
IAM consists of an FNN and an RNN in parallel, the D-set uses multi-head attention or is manually set.

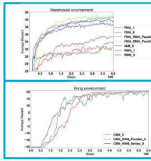
PPO
PPO is a policy gradient method for RL. The models are trained using PPO.

$$L^{CLIP}(\theta) = \hat{E}_t[\min(r_t(\theta)\hat{A}_t, clip(r_t(\theta), 1-\varepsilon, 1+\varepsilon)\hat{A}_t)]$$

Data and Simulation

Warehouse environment
This environment runs an robot (purple) which fetches items(yellow) that appear with a 0.05 probability on the edges of the 7x7 grid. A +1 reward is given per fetched box, and the boxes disappear after 8 timesteps, hence the robot must maintain a time counter and decide which box is best to go for.

OpenAI's Pong-v0 environment
The favorite Atari game everyone knows…



Results

The reproduced algorithms are very close to the original results.

However, as can be seen in the warehouse environment, the IAM does worse due to a different implementation of the d-set. The d-set was not manually set but had to learn automatically which sections of the warehouse was relevant. Also, a GRU was used instead of an LSTM.