Imagination-Augmented Agents for Deep Reinforcement Learning

Manan Tomar : ED14B023

Abstract

This work (Weber et al., 2017) proposes a model based method for reinforcement learning called I2A which combines concepts of both model based and model free learning by using predictions of a learned model. The results are shown on a relatively less popular puzzle environment called Sokoban and on the game of Minipacman.

1. Approach

This work revolves around combining model based and model free methods in order to benefit from both their advantages. This is done by 3 module setup proposed by the authors. The first is an environment model which is learnt over sample trajectories s_0, a_0... to predict the next state at each time step as well as the reward obtained. This model is used to generate imagination rollouts by feeding model predictions after every step, for a given rollout policy and starting from a sampled state s.

The key idea here is that since there are always model errors present, directly using these imagination rollouts will not work efficiently. Instead the authors propose to encode each imagination trajectory and aggregate them to feed as an added input to the policy module. The policy module simply takes as input the oservation and outputs the value function and the policy using standard model free techniques. The only model based component present in the policy module is the aggregated encoded imagination input. The intuition provided is that this can be considered as an additional context which can help the agent take better decisions.

As mentioned, the rollout policy is used to produce imagination rollouts once an enviornment model is pre trained. This rollout policy is a distillation of the policy used while training the environment model which in turn is that of a partially trained model free agent and not a random policy.

For training, a pre trained model is used. The authors use the RGB sprites directly and learn using A3C (Mnih et al., 2016). The baselines used is a standard model free version and a agent which predicts the same observations for each model prediction step in the above described setup.

2. Experiments

The major part of experiments is run on Sokoban which is a puzzle environment involving an agent which is required to push multiple boxes to desired locations. The agent cannot pull a box at any time, and therefore the environment is a irreversible setup and requires planning. The authors use a shaping reward scheme here which provides a +1 reward for every box pushed to the right location, -1 for pushing off target and +10 for finishing a level. A penalty of -0.1 is also applied at each time step. When compared against the baselines mentioned above, I2A is shown to outperform all. Additional experiments comparing prediction rollout performance and how well I2A does when the predictions are less accurate are also performed.

3. Critique

- The most important issue here is the environment which seems compatible with the methodology proposed. For instance, the environment being deterministic, allows learning accurate models in the first place. Moreover, the case where a mismatch in policy between the one used to learn the model and the one currently being used to learn the task is not apparent as the tasks are short lived and very few changes are observed as each level is being played.

- The reward used is shaping reward and not a sparse one, which begs the question if such a method works in an extremely sparse setting as well. For ex. only rewarding when *all* boxes have been pushed to their target locations. Moreover, the actions used here are discrete for both environments.

References

Mnih, Volodymyr, Badia, Adria Puigdomenech, Mirza, Mehdi, Graves, Alex, Lillicrap, Timothy, Harley, Tim, Silver, David, and Kavukcuoglu, Koray. Asynchronous methods for deep reinforcement learning. In *International conference on machine learning*, pp. 1928–1937, 2016.

Weber, Théophane, Racanière, Sébastien, Reichert, David P, Buesing, Lars, Guez, Arthur, Rezende, Danilo Jimenez, Badia, Adria Puigdomenech, Vinyals, Oriol, Heess, Nicolas, Li, Yujia, et al. Imagination-augmented agents for deep reinforcement learning. *arXiv preprint arXiv:1707.06203*, 2017.