Abstract

1. Introduction
   - DRL shows that agents learn
   - Rarely shown what they learn
   - CPSU as example → A3C end2end
   - Detailed investigation of learned behavior
     - → what's the reason for success?
     - → what's the feature from model-free learning
   - Extract feature from
   - Reflect on learned behavior
   - incorporate reflection (relevant/identified features) into reward
   - Self-reflecting Agent !?

RM

2. Related Work
   - 2013 Mnih, Atari, DQN (replay memory, ...)
   - 2015 Lillicrap, DDPG, Continuous Control
   - A3C, IMPALA, Sac-X, Reptile, etc...
   - Evaluation of these Algos → all in terms of simple metric, learning is not flexible

   ! ⇒ check intrinsic Motivation Papers !

   ⇒ Learnable Reward functions !

RM+DS

Inverse RL
Alex Irpan

IRL, A. Ng + S. Russel → find R(s,a)
policy given

3. Background / Methods
   - Double MA
   - A3C, end2end DRL
   - AWS GPU Cloud
   - etc...

DS

4. Results
   - Reproduced previous work (Reproducability Crisis in DRL ?) !
   - Presentation of Results (Gradcam and Reward)

DS

   - Asking, what the agent learned to be successful
   - Investigated $C_x(t) \to \tilde{C}_x(\omega) \to$ freq. components
   - Correlation with incr. reward → relevant and important feature for last few % !

RM

   - Retrain Agent with incorporation of new feature
     - → important to note: feature is humanly recognized but not built on a-priori information ! → agent developed the feature !
   - Show that new feature speeds up training !

DS

ToDo

5. Summary and Discussion
   - Reproduced prev. Work.
   - investigated in detail „What learns agent?"
   - found relevant feature not trivially accessible as it develops over time

RM

   - Concept of Self-Reflecting agent
     - → self-recognition and ident. of relevant features
     - → incorporate them in reward and verify their relevance
     - → Basically following the human way of hypothesizing and verifying / falsifying

ToDo