**DeepMind**

# Playing hard exploration games by watching YouTube

## Content Extraction

1. “Hard exploration” game is caused by sparse environment rewards.
   1. What is sparse environment reward?
   2. Why is RL agent () cannot work well in hard exploration?
      1. Available action trajectory grows exponentially in the # of frame separating rewards. [ possible action sequences in IceHockey and MONTEZUMA’S REVANGE]
      2. 100 frames take long time back-up signal, RL’s learning is not stable
   3. Solutions:
      1. Manually set the informative measurements for state or action trajectory
      2. Imitation learning ()
      3. Inverse reinforcement learning
2. Imitation learning needs videos, however various forms of videos will be used.
   1. Self-supervised learning can help unlabelled videos
   2. Temporal distance classification and cross-modal temporal distance classification can help videos with various of forms (e.g. different video colour, size and etc.) work better in self-supervised learning.
      1. Cross-modal involve audio and visual sequences.(method: )

**s**

# Maximum entropy inverse reinforcement learning

## Content Extraction

1. Inverse/Reinforcement Learning use model-based: MDP
2. Deterministic path distribution
3. Non-deterministic distribution