Lecture 9: Deep Reinforcement Learning
Object Manipulation and Task Planning

Victor Risager

March 5, 2024

1 Deep Reinforcement Learning

Strarted around 2014/2015. Not the same as *deep* in computervision. In DRL (deep reinforcement learning)they do only have around 3 layers.

2 RL Recap

- Learn how to make good actions by chasing a high reward.
- Trial and error learning.

Types of ML:

- Supervised learning (Labelled data)
- Unsupervised learning (Unlabelled data → clustering algorithms)
- Reinforcement learning (Agent acts with the enviromnent by trial and error.)

Sim2real gap → There will be issues when moving from simulaiton to a real robot.
It is much faster to have multiple agents in the same environment that learns the same policy. By moving the robot a small step makes a very small step in the observation space, gives a very small change in the policy. Therefore it is very difficult to explore the observation space. Therfore having multiple robots, can result in much greater variety in the learning.

2.1 MDP's

- Sequential Decision Making

Rewards for pickup an object:

- Distance to object
- Force sensors on the grippers
- Negative survival rewards.
- Negative reward for unexpected force on the arm.

continuous states are usually bounded between some values.
Similar algorithms

- Behaviour cloning: Tesla runs this in shadow mode.
- Inverse reinforcement learning

2.2 Actions

Continous or discrete actions

2.3 Policy

Map between states to actions

2.4 Value function

how good is it to be in the current state. The far sighted judgement about what rewards i will get if i take the following actions.

2.5 Task

- Episodic: Termination of the environment
- Continuing: Smart thermostat, it will keep accumulating rewards in this episode, which entails that every "episode" will yield infinite rewards.

2.6 Rewards

- Sparse: Reward when episode ends.
- Dense: reward for every step.

3 Practical applications of DRL

Apprentenchip learning. There will be a professional person controlling the helicopter while recording his inputs, and then make an RL algorithm learn from that. You can also use RL to control mobile robots, instead of using MoveBase.

3.1 Space Rover Exploration of planetary surfaces

Train thousands of robots in parallel to adopt path planning behaviour. You can also use multiple robot arms to grasp objects. Take depth and rgb data to extract object position, and use CNN's to extract features of the objects to train the algorihm on.

3.2 Papers

- DQN
- PPO
- DDPG

4 PPO

Talk much more about it next time. Learns the actions directly instead of just learning the Q-value to a fixed action. Has 2 networks.

5 Exercises

Use tensorboard to keep track of training.
MLP → 2 layers.
Use weights and biases page (like tensorboard, but cloud based and good for hyperparameter tuning.)