Deep V-Learning for Pool-Based Active Learning

March 4, 2022

1 Introduction

Even though pool-based Active Learning (P-AL) is the more popular setting for Active Learning, it creates significant challenges when reinforcement learning is applied to it. The most prominent one is that both Q-Learning and Policy Gradient Methods require a fixed action space. In the P-AL setting the action space has the same size as the pool of unlabeled images to choose from (or a subsample thereof). Even for simple datasets P-AL methods require a samplesize of >100 to work effectively, creating a large action space for the reinforcement learning agent.

1.1 Contribution

- First V-Learning Approach for P-AL
- Variable sample sizes to fit every dataset
- (Dataset and model agnostic state space) *Probably not the first*

2 Related Work

3 Background

3.1 V-Learning

V-Learning ([1] p.119) is the predecessor of Q-Learning. It estimates the value of states rather that the value of all actions given a state (Q-Learning [1] p.131) This requires the environment to provide possible future states $s_{t+1} \in S_{t+1}$ given any state s_t. A policy π_v typically chooses the most promising future state $argmax\ V(s_{t+1})$
Temporal difference (TD) learning for a neural network parametrized by θ gives us a nearly identical update formula compared to Q-Learning

$$\theta \leftarrow \theta + \eta \left(r_t + \gamma \hat{V}_\theta(s_{t+1}) \right) \tag{1}$$

3.2 Active Learning as a MDP

References

[1] Andrew G. Barto Richard S. Sutton. *Reinforcement Learning: An Introduction.* MIT Press, Massachusetts, 2 edition, 2020.