Multi-Agent Deep RL in Imperfect Information Games: Eric V York (zataomm@stanford.edu)

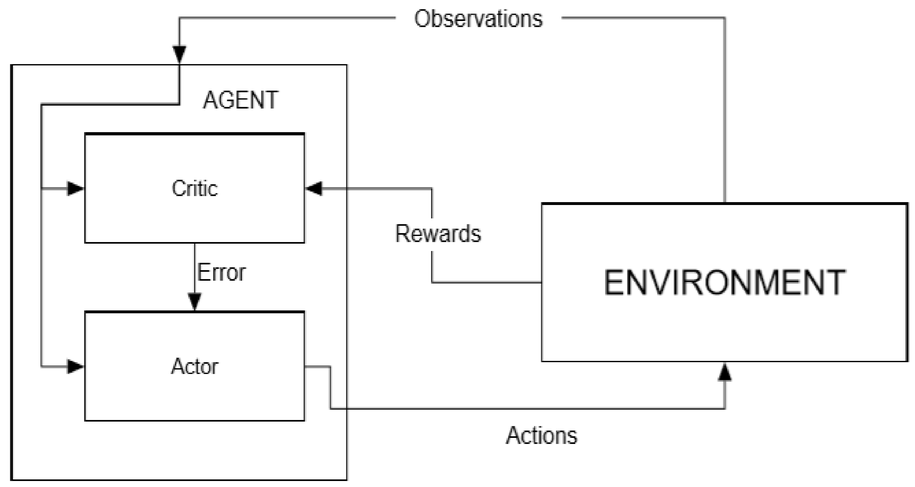
Code: https://github.com/Zataomm/cs230_catch5 : Video: https://www.youtube.com/watch?v=i9Zv2wZZ3eM

Proposed Problem

- *Learn Catch Five (Pitch with Fives)*
- *Four Player Card Game*
- *Bidding*
- *Suit Selection*
- *Strategy (Catching the Five)*
- *State/Action Space > 1.9x10^11*

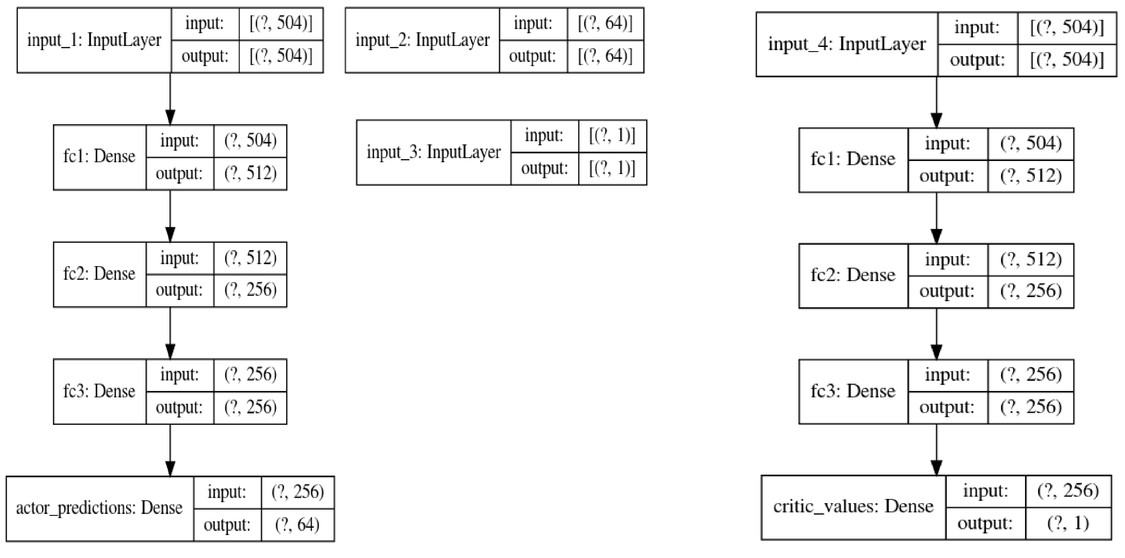
Solution

- *Proximal Policy Optimization*
- *Learns online*
- *Sample Efficient*
- *Easy to Implement*
- *Relatively Robust*
- *Easy to Implement :-)*



State Space Representation & Deep Actor-Critic Neural Net

Bids	Suit	Cards In Play	Players Live Cards	Players Discards	Player 1's Discards	Player 2's Discards	Player 3's Discards
32	4	52x4	52	52	52	52	52
≤ 4	≤ 1	≤ 4	≤ 9	≤ 14	≤ 5	≤ 5	≤ 5



PPO Algorithm see: arXiv:1707.06347v$_2$

$$L_t(\theta) = \min(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta)), 1-\epsilon, 1+\epsilon)\hat{A}_t$$

$$r_t(\theta) = \frac{\pi_\theta(a_t \mid s_t)}{\pi_{\theta_{\text{old}}}(a_t \mid s_t)}$$

$$\hat{A}_t = \delta_t + (\gamma\lambda)\delta_{t+1} + \cdots + \cdots + (\gamma\lambda)^{T-t+1}\delta_{T-1},$$
$$\text{where} \quad \delta_t = r_t + \gamma V(s_{t+1}) - V(s_t)$$