Playing the game of Maraffone with Reinforcement Learning

Davide Ragazzini

June 13, 2023

Game Environment

- Observation encoding: 780 variables
  - Partially observable state
  - Card encoding: 4 v. suit, 10 v. rank
  - Information about current round
  - Information about past rounds
- Reward: point difference
- Action encoding: 16 actions
  Illegal actions

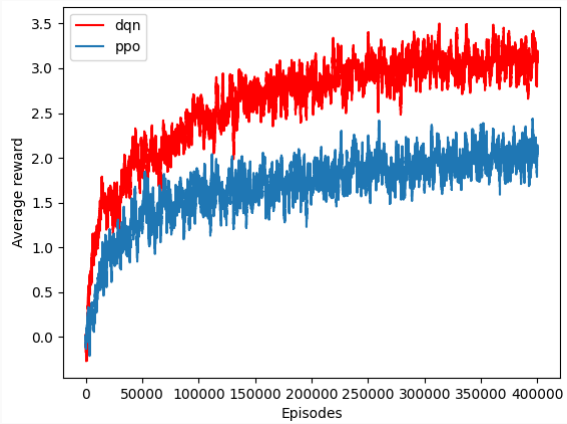
Methods

- Semi Gradient Q-Learning
  Double Q-Network
- Proximal Policy Optimization
  Clip version
  Generalized Advantage Estimation
- Training vs random and self play

Results

Agent 0	Agent 1	A. 0 wr	Avg. diff.
DQN	PPO	0.75	1.2
DQN sp	PPO sp	0.78	1.3
DQN sp	DQN	0.57	0.2
PPO sp	PPO	0.53	0.2
DQN	PPO sp	0.76	1.2
DQN sp	PPO	0.79	1.3

(a) Angent vs Agent



(b) Average reward during training