Exploration of Sparse and Dense Reward Structures in Google Research Football

Lakshay Dahiya, Steven Grisafi, Melvin Moriniere

Introduction

We used DDQN and PPO algorithm on Google Research Football environments and investigated the efficacy of sparse and dense reward structures for PPO

Methods

- Used reinforcement learning techniques to solve progressively more complex simulations of football (soccer) in Google Research Football Environment.
- Applied DDQN, a value-based algorithm, to solve simpler environments in the series.
- Applied PPO, an actor-critic algorithm, to solve the most complex environment in the series.
- Analyzed performance using sparse (rewards for scoring/missing only) vs dense reward structures (scoring/missing and advancement downfield) in PPO

Data Analysis

A. Determined success rates and learning speeds for DDQN in simpler environments.

B. Determined success rates and learning speeds for PPO in the most complex environment (11 vs 11 lazy opponents).

C. Compared efficacy of sparse vs dense reward structures for PPO.

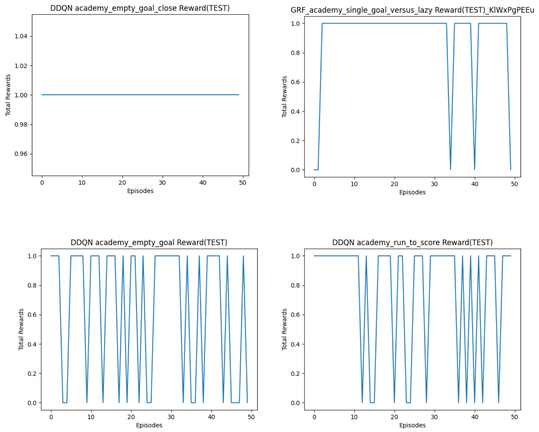

Figure A:
Testing results for trained models in three increasingly complex environments using DDQN and PPO

Training Episodes per Environment			
Environment	Training Episodes	Model	Sprites
Empty Goal (Close)	1,200	DDQN	1 vs 0
Empty Goal	20,000	DDQN	1 vs 0
Run to Score	20,000	DDQN	1 vs 0
Single Goal vs Lazy	30,000	PPO (sparse)	11 vs 11
Single Goal vs Lazy	20,000	PPO (dense)	11 vs 11

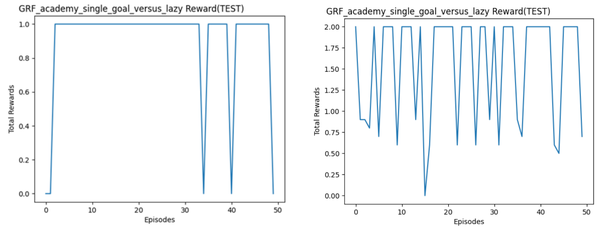

Figure B:
Testing results for sparse (left) vs dense (right) reward structures. Dense reward structures gives additional rewards for advancing downfield. Score of 1.0 in sparse and 2.0 in dense both correspond to scoring a goal.

Results

Using the DDQN algorithm, we were able to train our agent to solve three of the simplest environments in the Football Academy simulation series in Google Research Football. We obtained limited success when using this algorithm in more complex environments in which there was an opposing goaltender. We were, however, able to use the PPO algorithm to successfully solve the most complex environment in this series, in which the agent controlled 11 players against 11 "lazy" opponent players. In this mode all opposing players except the goaltender stand idle unless the ball comes to them.

While it took fewer episodes to train the agent with a dense reward structure and PPO, the agent performed better in testing when a sparse reward structure was used.

Conclusion

- The DDQN algorithm performed well in simpler environments but was not feasible in moderately complex football environments.
- The PPO algorithm performed well when used in the most complex environment in the series.
- Although dense reward structures do speed up training, the agent was sometimes observed behaving in a local maximum in which it ran the ball across the opposing team's goal line (out of bounds – corner kick).
- While slower, using sparse reward structures ultimately led to better performance.

Above: Screen capture of the PPO agent's player running the ball out of bounds instead of attempting to score.

References

1. Kurach, K. et al. (2019). "Google Research Football: A Novel Reinforcement Learning Environment". arXiv: 1907.11180 [cs.LG]
2. Schulman, J. et al. (2017) ."Proximal Policy Optimization Algorithms" arXiv:1707.06347 [cs.LG]

University at Buffalo The State University of New York