Training a Cooperating Team in GFootball Environment using Deep RL

Witalis Domitrz, Zuzanna Opała, Mateusz Sieniawski, Konrad Staniszewski

Faculty of Mathematics, Informatics, and Mechanics, University of Warsaw; Stefana Banacha 2, 02-097 Warsaw, Poland

witekdomitrz@gmail.com, opala.zuzanna@gmail.com, msieniawski98@gmail.com, staniszewskiconrad@gmail.com

University of Warsaw
Faculty of Mathematics, Informatics and Mechanics

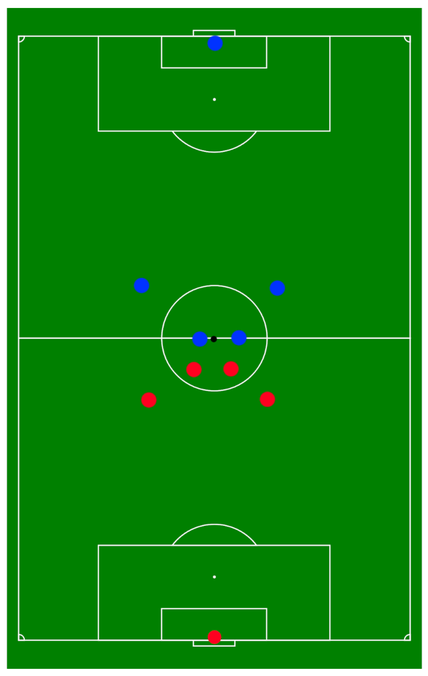
Introduction

Training multiple agents at once poses significant challenges and is a vibrant part of nowadays RL research. We train teams of agents to play football looking for emergence of coordination and strategy. In this work we present agents trained using different methods. We show two different curricula – one manipulating the opponents number, and the other with custom initial positions – that were able to resolve the sparse reward problem. We also compare two approaches to multi-agent training – centralized and decentralized.

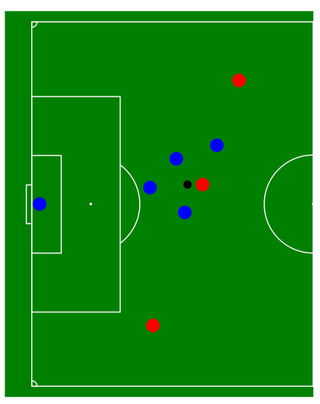
Environment

We use Google Research Football Environment (GRF)[1]. Our agents:

- play 5 vs 5 matches,
- control 4 players, while goalkeeper is controlled by built-in AI,
- use mini-map observation with marked positions of players and the ball,
- perform high level actions, like running in specific direction, passing, or scoring.



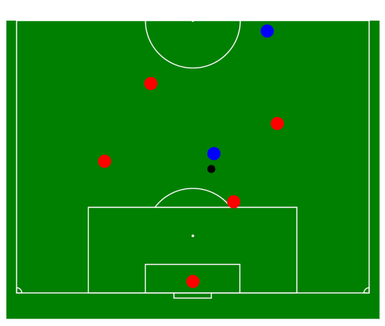
Emerging behaviours



Defensive play. The agents trained against the built-in AI did not express any organized defensive strategy[a] which emerged only in the self-play training[2]. Notably, inspired by [3] exploiter that was trained only against one frozen opponent to exploit its weaknesses, turned out to be great at defence also against previously unseen opponents[b].

Offside war. In a naive self-play, where policy trained against the current version of self, it engaged in something which could be described as an offside war, where the team focused mainly on not being present on their side of the pitch as it can be seen in the videos[c].

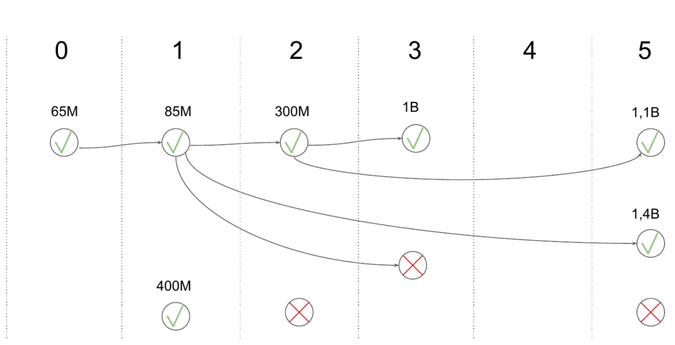
One-man army and the horde attack. Multi-head network preferred one-man army[d] while single-head sometimes played as a horde[e], that is running towards the opponent goal in a large group.

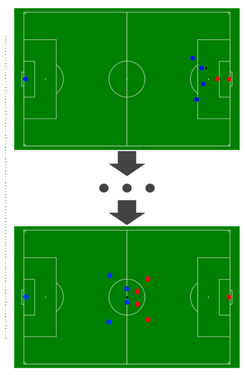


[a] https://youtu.be/w1hODs8QFQM?t=83
[b] https://youtu.be/iRXpLARkvJk?t=48
[c] https://sites.google.com/view/rl-football/multiagent-team/offside-strategy
[d] https://youtu.be/w1hODs8QFQM
[e] https://youtu.be/6elrWEHQuFk

Sparse Reward Problem and Curricula

Using only the goals as the reward makes it really sparse and our's 'pure' RL policies was not able to learn using it.





Left: opponents number curriculum process. Ticks mark successful training with accompanying number of steps.

Right: example scenarios from the second curriculum. In easy scenario we can see high number of agent's players near the opponent's goal.

We created two successful curricula to overcome this issue: the first gradually increased **number of opponents** while the second used **gradually harder scenarios** starting from the extremely simple ones up to the full-scale matches.

The first was slightly better against our internal pool of players but was loosing in the direct comparison with the second.

Centralized and Decentralized Approach

We compared the two approaches after training them in self-play manner using the same amount of compute.

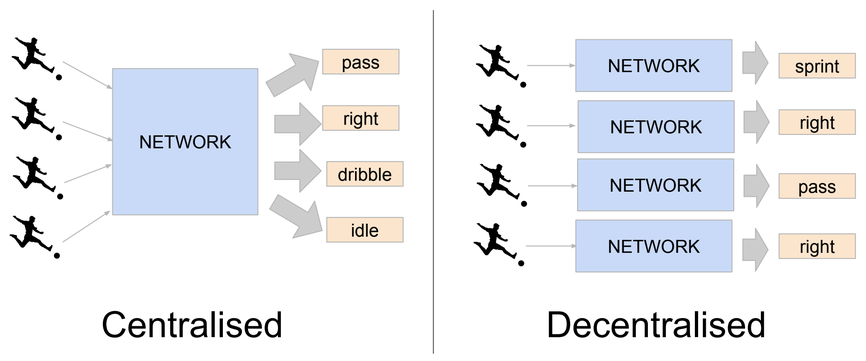
The decentralised policy was winning with the centralised one with an average advantage of 2.8 (± 1.8) goals.

It is surprising as we would expect better coordination and strategy from a centralised one, though similar findings were also presented by [4].

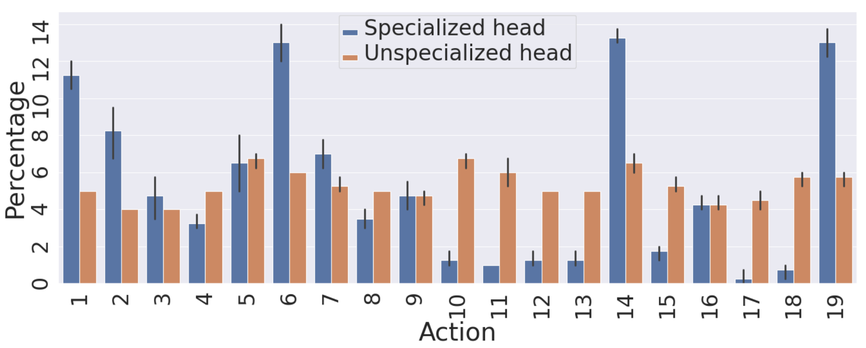
We also tried both architectures on a simpler scenario with three attackers and just one defender.

Unlike in previously presented results, in this scenario the centralised architecture learned faster and had greater accuracy.

We suspect this easier scenario required players to specialize which was done easier by a centralized setup, whilst adapting to a full 5 vs 5 was simpler for the decentralized policy as roles of the players are more universal in a full 5 vs 5 match.



In centralised setup we have one, central policy which controls all the players. In decentralised setup each controlled agent performs actions separately. In our case we have one network - that separately controls all the players.



Action distribution of two players controlled by a centralized architecture. Out of four controlled players two had more specialized action set *(blue)*, while the other two output rather random actions *(orange)*.

6. References

[1] Karol Kurach, Anton Raichuk, Piotr Stańczyk, Michał Zając, Olivier Bachem, Lasse Espeholt, Carlos Riquelme, Damien Vincent, Marcin Michalski, Olivier Bousquet, and Sylvain Gelly. Google Research Football: A Novel Reinforcement Learning Environment. *arXiv e-prints*, page arXiv:1907.11180, July 2019.

[2] Joel Z. Leibo, Edward Hughes, Marc Lanctot, and Thore Graepel. Autocurricula and the emergence of innovation from social interaction: A manifesto for multi-agent intelligence research, 2019.

[3] Oriol Vinyals, Igor Babuschkin, Wojciech M. Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H. Choi, Richard Powell, Timo Ewalds, Petko Georgiev, Junhyuk Oh, Dan Horgan, Manuel Kroiss, Ivo Danihelka, Aja Huang, Laurent Sifre, Trevor Cai, John P. Agapiou, Max Jaderberg, Alexander S. Vezhnevets, Rémi Leblond, Tobias Pohlen, Valentin Dalibard, David Budden, Yury Sulsky, James Molloy, Tom L. Paine, Caglar Gulcehre, Ziyu Wang, Tobias Pfaff, Yuhuai Wu, Roman Ring, Dani Yogatama, Dario Wünsch, Katrina McKinney, Oliver Smith, Tom Schaul, Timothy Lillicrap, Koray Kavukcuoglu, Demis Hassabis, Chris Apps, and David Silver. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, 575(7782):350–354, October 2019.

[4] David L. Leottau, Javier Ruiz del Solar, and Robert Babuška. Decentralized reinforcement learning of robot behaviors. *Artificial Intelligence*, 256:130–159, March 2018.