Progressive Deep Reinforcement Learning for Strategic Set-Plays in Robotic Football

Tomás Azevedo[1,2], Francisco Silva[1,2], Nuno Lau[3], Luís Paulo Reis[2]

[1] FCUP - Faculty of Sciences, University of Porto
[2] FEUP - Faculty of Engineering, University of Porto
[3] IEETA - Insitute of Eletronics and Informatics Engineering of Aveiro

FC Portugal's research has always been focused on high-level coordination. Recent development include 2012 MsC thesis[1] that tries a RL approach to Set-Plays involving Two-Player scenarios. Mota[2], in his 2023 MsC thesis, presents a novel programming language TSAL, designed to implement team strategies. The work developed was inspired in concepts presented in the referred approaches and extends these.

The Hierarchical Learning Framework has been adopted by FC Portugal in the team's learning environments. This framework leverages the already acquired lower-level skills to learn new higher-level ones. The stratification of the learning process follows the principles in [3].On top of this, a progressive method for learning was followed, consisting in starting in the simplest two-player corner scenario and slowly unlocking agents' capabilities without changing any of the environments' state space or the reward function.

A corner kick set-play `InCorner` creation, the main focus of this work, uses two previously developed, already implemented behaviours. The action space consists of a τ and θ for a pass/kick controller and x, y, z coordinates for a walking behaviour. The observation space consists in both players positions, robot's joint information and ball position. The first scenario considered, $Phase1$, is a two-player corner-kick. One player R_1 starts at $x = 16$ and $y = 6, -6$, randomly selected, and the other player R_2 starts at the $(13, 0)$ coordinates with random orientation. $Phase2$ has the same initial conditions, differing only in the freedom of each player's action. $Phase3$ implements different initial positions for R_2. These vary uniformly between $x = [9, 15]$ and $y = [-5, 5]$. The permanent reward signal follows:

$$R = 10 - ((t_s/t_m)^{0.9}) * 0.5 - (t_s/t_m)^{0.1} - 0.001 * t \tag{1}$$

A custom single network simulates individual networks by only connecting through output concatenation. Each player has its own actor and critic networks. The actor and critic networks differ only in the last layers' activation function for the policy and value heads. The inputs are split at start, fed through linear fully connected feature extractors with $softplus$ activation, one for each actor/critic network, then connected to a policy/value head with $softplus$ and $tanh$ activations respectively. The dimensions of the hidden layers are of 64 units. The resulting model was trained through 14M steps across the various phases, learning good transitions. Further training will focus in unlocking the full potential of the set-play, including adding reactive defenders.

References

[1] Santiago, M.: FC Portugal - High-Level Skills Within A Multi-Agent Environment. Master's thesis, Faculdade de Engenharia da Universidade do Porto, Porto, Portugal (2021)

[2] Mota, P.: A Team Strategy Programming Language Applied to Robotic Soccer. Master's thesis, Faculty of Sciences of University of Porto (2023)

[3] Stone, P., Veloso, M.M.: Layered learning and flexible teamwork in robocup simulation agents. In: Robot Soccer World Cup (1999), https://api.semanticscholar.org/CorpusID:43688658