AEteam Soccer Simulation 2D Team Description Paper 2023

Erfan Fathi , Soroush Mazloom , Parham Keyhani , Amir hosein Nikfetrat and Vahid khodabakhshi

Atomic Energy High School, Tehran, Iran

{fathye897, mazloomsoroush , Parham.8484 , amirhosein.nikfetrat79 , khodabakhshi.vahid}@gmail.com

Abstract. This team description paper presents an overview of previous work and recent research topics of Team AEteam. In this article, we aim to enhance our understanding and improve the efficiency of offensive players escaping by exploring and expanding our knowledge on the subject. To achieve this, we employ the mentioned artificial intelligence methods with the hope of improving the trend of players' attacks.

Keywords: Q-Learning, Reinforcement Learning, Mutual Information (MI), Kullback–Leibler divergence, Relative Entropy, I-divergence, Markov Decision Process (MDP), Outer Product, Machine Learning, Soccer Simulation 2D, Robocup.

1 Introduction

In the intricate world of 2D football game strategy, the pursuit of excellence transcends mere player movements; it delves into the realm of cognitive prowess and strategic finesse. Within this arena, our exploration encompasses three pillars of learning methodologies — Q-Learning, Reinforcement Learning, and Mutual Information Learning. These methodologies stand not as isolated entities but as harmonious components orchestrating a symphony of tactical brilliance. As we navigate this terrain, envision Q-Learning as the master strategist, shaping the very fabric of offensive decisions. Reinforcement Learning emerges as the guiding mentor, sculpting players into adaptive tacticians. Introducing Mutual Information Learning, the insightful observer that quantifies the intricate dance between offensive escapades and defensive counterplays. This synergy creates not just agile players but a dynamic strategy that intertwines individual success with the holistic ebb and flow of the 2D football game. Join us on this journey as we dissect, refine, and amalgamate these methodologies, paving the way for a strategic paradigm shift in the realm of 2D football gaming.

1.1 Q-Learning

Q-Learning is a dynamic approach in reinforcement learning, envision it as the master strategist. In a 2D football game, Q-Learning becomes the architect of player decisions. It systematically explores various offensive strategies, learning from each move's success or failure. The system evolves, fine-tuning its approach based on the learned experiences, ultimately crafting agile and effective players.

1.2 Reinforcement Learning

Expanding our scope, reinforcement learning serves as the guiding mentor. In the context of the 2D football game, it empowers offensive players to adapt and improve through continuous feedback. Successful escapes become lessons, shaping the proficiency of players in navigating the complex defensive landscape. Reinforcement learning ensures an adaptive and responsive offensive strategy.

1.3 Mutual Information Learning

Enter Mutual Information (MI) as the insightful observer. This learning method quantifies the statistical relationship between **offensive player escapes** and **defensive responses**. In the 2D football game, Mutual Information assesses not only individual successes but also the strategic harmony between offensive and defensive dynamics (Fig. 1). It provides a nuanced understanding of how player movements and defensive reactions intertwine.

Fig. 1. harmony between offensive and defensive dynamics

1.4 Combination

Now, envision a harmonious blend of Q-Learning, reinforcement learning, and Mutual Information. Q-Learning refines the strategic blueprint, reinforcement learning molds agile players, and Mutual Information evaluates the intricate dance between offense and defense. This combined approach doesn't just yield skilled players; it crafts a dynamic and adaptive strategy in the 2D football game, where each decision is informed by both individual success and the holistic team dynamics.

Note: We are using Cyrus2D Base [1] as our base for Soccer Simulation 2D. Cyrus2D Base is created by merging Helios base (Agent2D) with Glider2D base and applying features from Cyrus2021, the champion of RoboCup2021.

2 Related work

Now, we will mention some articles published by other Soccer simulation 2D teams. HELIOS developed "Player's MatchUp" algorithm for exchanging players' positions during the game for better team performance [2]. CYRUS uses opponent's pass prediction for marking and teammate's pass prediction for unmarking [3]. Persepolis optimized its offensive strategy by randomly placing players in the soccer 2D field and training them for the best attack [4]. MT2022 used HFO (Half Field Offense) for training and testing their shoot algorithm [5]. CYRUS 2023 reduced noise in Soccer Simulation 2D by using different methods including Deep Neural Networks (DNNs) and Long Short-Term Memory (LSTM) networks [6]. HELIOS2023 enhanced ball chasing behavior by using Learning methods such as Learning-to-rank method [7]. YuShan proposed a method based on kernel density negative example learning, which improved the classification effect of different kinds of passing feature data by parsing the game log files [8].

3 Q-Learning in Strategy

As previously mentioned, this method alters the team’s strategic planning by managing switches between three offensive states: central attack, left flank attack, and right flank attack. Utilizing the defined formula, this process involves scoring each state based on the outcomes obtained and conditions tested, ultimately determining the most suitable state for a team.

The overall process is in accordance with the chart shown below (Fig. 2).

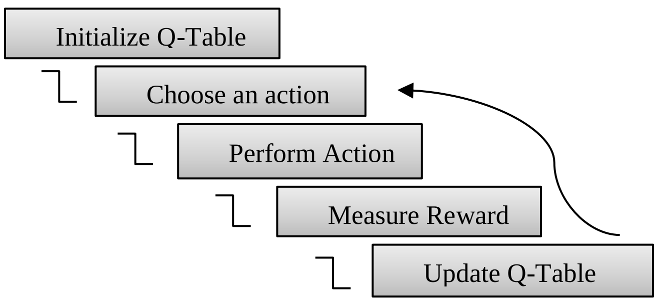

Fig. 2. Q-Learning process

Here is the update formula:

$$Q(S, A) \leftarrow Q(S, A) + \alpha[R + \gamma max_a Q(S', a) - Q(S, A)]$$

Where:

S is the current state.
A is the chosen action.
S′ is the next state.
R is the reward for taking action A in state S.
α is the learning rate (0 < α ≤ 1) - it determines to what extent the newly acquired information will override the old information.
γ is the discount factor (0 ≤ γ < 1) - it quantifies how much importance we want to give to future rewards.
max_a Q(S′, a) is the estimated optimal future value in next state S'.
ε-greedy is a common policy where the agent chooses the best action with 1-ε probability, and a random action with ε probability (exploration).

In the chart below, a tested sample of 500 states during a game against the Cyrus2023 team is presented, showing how the defined values for each state have evolved (Table. 3).

Table. 3. Sample data updated by Q-learning during training

Q-Table		Actions (Modes)		
		left flank	central	right flank
State	0	0	0	0
. . . . .	. . . . .	. . . . .	. . . . .	. . . . .
State	156	-5.62315611	-9.728	8.55648
. . .	. . .	. . .	. . .	. . .
State	500	-2.56562	-11.5	4.927

4 Reinforcement Learning

Reinforcement learning (RL) is an area of machine learning concerned with how software agents ought to take actions in an environment to maximize some notion of cumulative reward. The process is typically modeled as a Markov Decision Process (MDP) and involves learning a policy that maps states of the environment to the actions that the agent should take.

The core formula that underpins many reinforcement learning algorithms is the Bellman equation, which recursively defines the value of a policy.

Bellman equation:

$$V^{\pi}(s) = \sum_{a \in A} \pi(as) \sum_{s\prime \in S} P(ss, a)[R(s, a, s') + \gamma V^{\pi}(s')]$$

This equation expresses the value ($V^{\pi}(s)$) of a state (s) under a policy (π) as the expected sum of rewards ($R(s, a, s')$) obtained by taking an action (a) in state (s), ending up in a new state (s'), and the discounted future rewards from (s'). The discount factor (γ) weighs the importance of future rewards.
Now that we have briefly explained the algorithmic process, let's move on to using it in our composite attack system.

4.1 Exploration and Exploitation

In the realm of reinforcement learning, each attacker player becomes an agent on a quest for mastery. The process begins with a delicate balance between exploration and exploitation. Players explore various offensive strategies, attempting different moves to discover the most effective ones. Simultaneously, they exploit learned successful tactics, refining their arsenal of maneuvers.

4.2 State Representation

Attacker players navigate a dynamic environment where the state of play is constantly evolving. Reinforcement learning involves creating a robust state representation. This representation encapsulates the pertinent information — positions of teammates, opponents, ball location, and contextual cues. It serves as the foundation for informed decision-making.

4.3 Action Selection

Reinforcement learning empowers each attacker to make decisions autonomously. Through continuous interactions with the game environment, players learn the consequences of their actions. The system dynamically adapts, evolving a strategy for each player that aligns with their unique strengths and the ever-changing dynamics of the match.

4.4 Reward Mechanism

At the heart of reinforcement learning lies the reward mechanism. Attacker players receive feedback in the form of rewards or penalties based on the outcomes of their actions. Successful goal attempts yield positive rewards, while ineffective maneuvers or defensive counterplays result in negative feedback. This constant feedback loop refines the players' decision-making over time.

4.5 Temporal Credit Assignment

As attackers engage in the ebb and flow of the game, reinforcement learning employs temporal credit assignment. It traces the consequences of actions back in time, attributing credit or responsibility to decisions that contributed to successful outcomes. This nuanced understanding allows players to learn not just from immediate successes but from the sequence of events leading to success.

4.6 Adaptation and Dynamic Play

Attacker players, under the umbrella of reinforcement learning, embody adaptability. The system learns from each player's unique style, adjusting strategies based on individual strengths and weaknesses. This adaptability translates into dynamic play, where attackers seamlessly adjust their approach in response to the evolving game state.

In essence, reinforcement learning for attacker players orchestrates a symphony of exploration, adaptation, and strategic finesse. Each player becomes an autonomous agent, contributing to the collective intelligence of the team in the pursuit of offensive mastery within the 2D football game.

5 Mutual Information

Mutual information (MI) as a learning algorithm is often used for gauge the dependency between variables in a dataset and is particularly useful in feature selection for machine learning (ML). MI measures the amount of information one can obtain about one random variable by observing another.

Let (x, y) be a pair of random variables with values over the space $x.y$. If their joint distribution is $p(x, y)$ and the marginal distributions are $p(x)$ and $p(y)$, the mutual information is defined as

$$I(X;Y) = D_{KL}(P(X,Y), ||, P(X)P(Y))$$

Where D_{KL} is the kullback-leibler divergence, and $P(X)P(Y)$ is the outer product distribution which assigns probability $P(X)$(x). $P(y)$(y) to each (x,y).

$$MI(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right)$$

In this formula, $(p(x, y))$ is the joint probability distribution function of (X) and (Y), and $(p(x))$ and $(p(y))$are the marginal probability distribution functions of (X) and (Y) respectively.

It seems that everything is apparent,

5.1 Information Gathering

Mutual Information learning begins with the meticulous gathering of information. Each attacker player acts as an information seeker, analyzing the intricate interplay between their offensive maneuvers and the corresponding defensive reactions. This continuous process involves capturing the nuances of how player actions influence the overall game dynamics.

5.2 Statistical Dependency Analysis

At its core, Mutual Information seeks to unveil the statistical dependencies between offensive player movements and defensive responses. Attacker players, through this learning process, become statistical analysts, discerning patterns in the game data. The system identifies not only direct correlations but also subtle relationships that contribute to successful offensive plays.

5.3 Adaptive Decision-Making

Armed with insights from statistical dependencies, attacker players embrace adaptive decision-making. Mutual Information learning equips each player to dynamically adjust their strategies based on the observed dependencies. It's not merely about past success but a real-time responsiveness to the evolving relationship between offense and defense.

5.4 Contextual Evaluation

Mutual Information learning introduces a contextual evaluation layer. Attacker players don't just aim for successful goal attempts; they aim for strategically significant interactions. The system evaluates the context surrounding each offensive move, considering how it contributes to the broader team strategy and disrupts the opposing defense.

5.5 Quantifying Significance

In the pursuit of offensive mastery, Mutual Information provides a metric for quantifying the significance of each attacker's contribution. This metric considers the information gain achieved through specific player actions. It guides the players to prioritize moves that not only lead to personal success but enhance the overall team's understanding of offensive dynamics.

5.6 Feedback-Driven Adjustment

Attacker players engage in a feedback-driven adjustment process. The insights gained from Mutual Information learning become a continuous feedback loop, shaping each player's decision-making. Positive information gain reinforces effective strategies, while negative information gain prompts adaptive shifts in offensive approaches.

5.7 Collaborative Information Sharing

Mutual Information learning fosters a culture of collaborative information sharing among attacker players. Insights gleaned from individual actions contribute to the collective intelligence of the team. It's not just about personal success but the synergy created by aligning the information-gathering efforts of each attacker towards a shared offensive vision.

In essence, Mutual Information learning transforms attacker players into informed strategists. They become adept at not only recognizing successful moves but understanding the intricate web of dependencies that define effective offensive play within the 2D football game.

6 What exactly does a Team AEteam player do?

Simultaneous analysis of game data, utilization of individual experiences, updating these experiences, scoring movements, and paying attention to changes between the attacking and defending movements of opponents result in the integration of these multiple artificial intelligence methods for these players (Fig. 3).

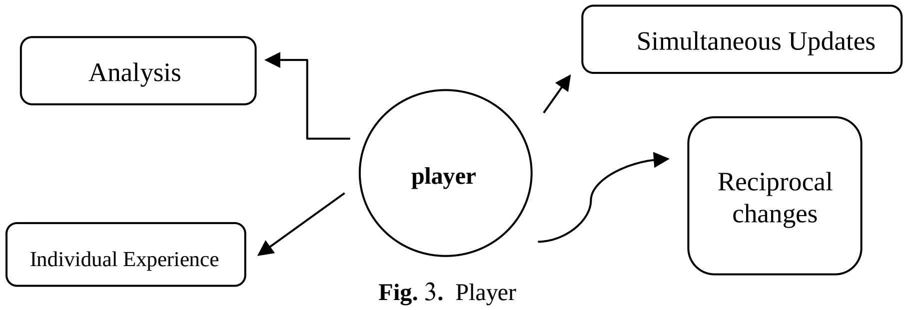

Fig. 3. Player

References

1. CYRUS team. Cyrus2DBase. [Online] Available: https://github.com/Cyrus2D/Cyrus2DBase
2. Yamaguchi, M., Kuga, R., Omori, H., Fukushima, T., Nakashima, T., & Akiyama, H. (2021). *HELIOS2021: Team Description Paper.*
3. Zare, N., Firouzkouhi, A., Amini, O., Sarvmaili, M., Sayareh, A., Ramezani Rad, S., . . . Amilcar, S. (2022). *CYRUS Soccer Simulation 2D Team Description Paper 2022.*
4. Noohpisheh, M., Shekarriz, M., Zaremehrjardi, F., Khademi Ardekani, F., & Khorsand, S. (2021). *Persepolis Soccer 2D Simulation Team Description Paper 2021.*
5. Guan, L., Chen, Q., Wang, J., Xiang, H., Meng, S., Wang, C., . . . Chen, S. (2022). *MT2022: Team Description Paper.*
6. Sayareh, A., Zare, N., Amini, O., Firouzkouhi, A., Sarvmaili, M., Matwin, S., & Soares, A. (2023). *Observation Denoising in CYRUS Soccer.*
7. Akiyama, H., Nakashima, T., Hatakeyama, K., Fujikawa, T., & Hishiki, A. (2023). *HELIOS2023: Team Description Paper.*
8. Cheng, Z., Ren, Y., Liu, C., Huang, J., Wang, J., Zhu, Y., & Zhang, L. (2023). *YuShan2023 Team Description Paper for RoboCup2023.*
9. Akiyama, H., Nakashima, T., & Hatakeyama, K. (2022). *HELIOS2022: Team Description Paper.*
10. Akiyama, H., Nakashima, T., Fukushima, T., Suzuki, Y., & Ohori, A. (2019). *HELIOS2019: Team Description Paper.*
11. Gabel, T., Eren, B., Eren, Y., Sommer, F., & Godehardt, E. (2023). *FRA-UNIted — Team Description 2023.*
12. Gabel, T., Eren, Y., Sommer, F., Vieth, A., & Godehardt, E. (2022). *FRA-UNIted — Team Description 2022.*
13. Gabel, T., Kloppner, P., Eren, Y., Sommer, F., Breuer, S., & Godahardt, E. (2021). *FRA-UNIted — Team Description 2021.*

14. Fathi, E., & Mazloum, S. (2023). *EMPEROR Soccer Simulation 2D Team Description Paper 2023*.
15. Marian, S., Luca, D., Sacuiu, R., Sarac, B., & Cotalea, O. (2023). *OXSY 2023 Team Description.*
16. Marian, S., Luca, D., Sarac, B., & Cotalea, O. (2022). *OXSY 2022 Team Description.*
17. Marian, S., Luca, D., Sarac, B., & Cotarlea, O. (2021). *OXSY 2021 Team Description*
18. Noohpisheh, M., Shekarriz, M., Bordbar, A., Liaghat, M., Salimi, A., Borzoo, D., & Zarei, A. (2019). *Razi Soccer 2D Simualtion Team Description Paper 2019.*
19. Noohpisheh, M., Shekarriz, M., Zaremehrjardi, F., & Karimi, M. (2022). *Persepolis Soccer 2D Simulation Team.*
20. Tomoharu, N., Akiyama, H., Suzuki, Y., Ohori, A., & Fukushima, T. (2018). *HELIOS2018: Team Description Paper.* (Noohpisheh, et al., 2023)
21. Noohpisheh, M., Shekarriz, M., Nematollahi, R., Ghasemi, F., Mohammadi, M., Amiri, N., & Amiri, S. (2023). *The8 Soccer 2D Simulation Team Description Paper 2023*.
22. Wikipedia. Q-learning. [Online] Available: https://en.wikipedia.org/wiki/Q-learning
23. Wikipedia. Reinforcement learning. [Online] Available: https://en.wikipedia.org/wiki/Reinforcement_learning
24. Wikipedia. Matual Information. [Online] Available: https://en.wikipedia.org/wiki/Mutual_information