Learning attacking strategies for the game of soccer from real-world data

Atsuto Hiraide s1230100 Supervised by Prof. Maxim Mozgovoy

Abstract

The purpose of this study is to learn attack strategies for the soccer game from real world soccer data. Previous studies have shown that it is possible to learn reliable AI agents from human behavior data. We show that by learning passing behavior from a real soccer team, it is possible to obtain the same passing patterns as exist in human tracking data. We also shows that typical rule-based soccer AI teams have significant differences in passing behavior compared to real teams.

1.Introduction

Recently, efficient game AI systems that can defeat human players have already been developed for many genres of computer games [2]. However, skill is not the only requirement for game AI. Its original purpose is to contribute to user engagement. Studies show that people generally prefer playing with other people rather than with bots [3], Therefore, the goal of creating a realistic human-like AI agent is important for actual game development [4, 5].

The importance of AI believability is especially high in computer simulations of real games such as sports video games. Computer soccer and basketball players are hoping to faithfully reproduce their favorite sport, such as expressing athletes' behavior realistically. This task can be approached by learning behavioral patterns from real human tracking data, as demonstrated in several recent studies [6, 7].

Our ultimate goal is to create a team that behaves like a human in a soccer video game by learning from data obtained from real soccer games. Since this is a complex task involving numerous different objectives, we are trying to address some of them independently. Our experimental setup is based on the SimpleSoccer simulator with an AI engine developed by Mat Buckland [8]. It has a built-in rule based AI, which we call a default algorithm [1].

2.Method

2.1 Passing action

Passing is one of the key elements of soccer team strategy. Characteristic passing patterns can also be observed at the individual team and player level, so they are probably related to specific human-like behavioral characteristics that the player can recognize. In addition, passes are abundant in every soccer game, making them easy to classify and compare [1].

One of the main assumptions of this project is the availability of a limited dataset of player tracking data. As the adoption of tracking systems expands, the availability of such data is generally expected to increase. However, typical professional soccer teams only play dozens of games per season, so relying on small data samples to learn team-specific behavior patterns, especially to handle relatively rare game events is needed. Because an average professional team makes about 365-370 passes per game [9], it is possible to expect to learn its passing behavior that will be distinguishable from other teams.

Given these considerations, we decided to conduct a preliminary experiment using the method used in our previous projects [10, 11]. They rely on a combination of case-based reasoning decisions and a database of human behavior, such as Markov chains. In brief, agent knowledge is represented as a graph, having individual game situations as vertices, and actions as weighted edges. it represents the fact that the situation A turns into the situation B as a result of a certain action during the learning phase. A decision making algorithm tries to find the best match for the current game situation and acts accordingly.

For learning passing behavior, each game situation is represented as a set of the following attributes:

- The coordinates of the player with the ball (the passer) in the 18×10 grid.
- The “danger to move forward” heuristic estimation on the scale of 0 to 5 (depends on the distance to the nearest opponent in the forward direction).
- The current movement direction of the passer (8 directions are supported).
- The direction (0-7) of the closest opponent (from the passer’s perspective).
- The distance to the closest opponent (from the passer’s perspective), on the scale from 0 to 2.
- The “safest pass danger” heuristic estimation on the scale of 0 to 5 (depends on locations of both teammates and opponents).
- The “safest forward pass danger” heuristic estimation on the scale of 0 to 5.
- The Boolean attribute indicating that at least one safe pass (with danger estimation of 0-2) is found.

Each action we learn is a pass action, characterized by the coordinates of both sending and receiving players. STATS.com "Soccer Dataset [12]" is used as a training set for the learning algorithm. It consists of 7,500 game sequences, represented with series of game situation snapshots taken at a rate of 10 snapshots per second. Each snapshot (frame) contains the coordinates of all 22 players and the ball. The sequences are taken from real European League matches and represents approximately 36 hours of play time. Since there is no event markup in this data, we use a simple rule -based algorithm to detect passes using closeness of the ball to individual players as a criterion of ball possession.

2.2 Shot action

Our set does not contain shooting actions, we needed to implement a feature of AI simulation of the shots on goal. We made an algorithm determining whether PWB (a player with ball) can shoot to a goal or not. If there are enemy players between the goal and PWB, it is difficult for PWB to shoot. The following method was used to determine if the enemy players(except a goalkeeper) were inside or outside a triangle consisting of the coordinates of the PWB and the coordinates of the edge of the goal [Figure 1].

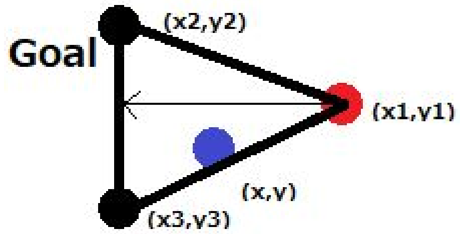

Figure 1. Calculate area of triangle formed.

Each coordinate is defined as follows.

- PWB: A (x1, y1)
- Left side of goal: B (x2, y2)
- Right side of goal: C (x3, y3)
- Enemy players: P (x, y)

The area of triangle ABC is calculated as follows.

$$\frac{1}{2}\left|x1(y2-y3)+x2(y3-y1)+x3(y1-y2)\right|$$

Using this method, the areas of triangles ABC, PBC, PAC, and PAB are calculated. We check if total area of triangles PBC, PAC, and PAB is same as triangle ABC. If this calculation is correct, a second player is inside a triangle surrounded by the coordinates of the PWB and the goal. Next we check if we can use a function to find a second player around the ball (sometimes two players are close to each other and it is impossible to shoot). It is very hard to shoot from an acute angle when the player is not in front of the goals, and it is very difficult to shoot far from the goals. If there are any of these problems, PWB will determine that it is impossible to shoot, otherwise it will determine that it is possible to shoot.

3.Result

3.1. Pass algorithm

To evaluate the performance of a new passing algorithm, we compared the characteristics of passes made by the new algorithm, the old (default) algorithm, and by the real-life teams. We simulated a number of AI vs. AI matches using old and new algorithms until 500 passes are made in each case, and extracted 500 random passes from the STATS.com dataset. The passes were classified according to their length and direction (see Table 1 and Table 2) .

TABLE I. PASS LENGTH DISTRIBUTION
Pass length, m (%)

Team	0-10	10-20	20-30	30-40	40+
Default AI	0.00	11.74	52.84	27.84	7.58
New AI	17.44	47.29	20.74	7.36	7.17
Real Teams	28.40	49.63	16.89	3.71	1.37

TABLE II. PASS DIRECTION DISTRIBUTION
Pass direction (%)

Team	FW	FR	R	BR	B	BL	L	FL
Defailt AI	8	11	21	11	8	9	17	14
New AI	16	14	17	10	4	7	16	16
Real Teams	10	14	16	12	9	12	15	12

To compare passing patterns, we represented pass length and pass direction values for different teams as vectors and calculated their cosine similarity ratios (see Table 3).

TABLE III. SIMILARITY OF PASSING PATTERNS
Distance Similarity / Direction Similarity (%)

	Default AI	New AI	Real Team
Default AI		95.64	97.89
New AI	51.15		95.87
Real Team	43.35	97.45	

The results showed that the default AI system did not show a human-like passing pattern in terms of distance, but the distribution in the path direction was similar to the distribution found in the actual data. Our system is based on learning by observation and exhibits human-like passing behavior according to these criteria [1].

3.2. Shot recognition

Because of the limitation of the dataset it is difficult to compare the results, since we do not have the tracking data of how the episode finished. However by counting the number of recognized shots in files and number of lines in the recordings (the frame rate is 10 frames (lines) per second of game), it is possible to calculate approximate value of shots per game. The available dataset contains 36 hours of active game data (The redundant state set to has been removed.). According to FIFA statistics, an average game consists of 57.6 minutes of action [13]. Estimating based on them, the dataset has a total of about 38 games. The number of shots recognized from the data using the algorithm was 676 (17.7 shots per game). To compare the actual number of shots, we looked at the average number of shots in the English Premier League [14]. According to Premier League website, the average league team shots 12.4 times per game, which gives 24.8 shots per game for 2 competing teams (see Table 4). Sunderland, the lowest average number(10.7) of shots in the English Premier League, is also included in the table to compare the result.

TABLE Ⅳ. AVERAGE SHOTS

Average recognized shots	17.7
Average shots in EPL	24.8
Average Least shots in EPL	21.4

From these results, the number of recognized shots are smaller than ideal number. This algorithm assumes that the player with the ball will shoot to the goal in front, however real soccer players shoot in various situation.

Our next challenge about the shot recognition are to develop a more sophisticated algorithm (For example, considering where is the player with the ball and adding some randomization to the shot to imitate real life, or recognizing shots in consideration of the shot angle, and speed as [15] , etc) or to try learning from data (machine learning) to do the shot recognition instead.

4.Conclusion

Soccer is a complex multi-agent game that challenges AI technology. The diversity and sophistication of the technologies used by modern RoboCup AI teams show how difficult it is to design a skilled AI system for soccer [16]. However, with the advancements in practical AI development, the importance of other factors such as believability of AI-controlled teammates and opponents will grow. These factors contribute to the overall enjoyment of the computer game, and the main purpose of the game is to entertain players.

We analyzed the passing behavior of human teams and compared it with a rule-based AI system according to two features: pass length and pass duration. Our tests show that real teams have very different passing patterns with AI. In addition, we created learning using an observation-based system, obtained passing behavior from actual human passing data, and demonstrated that it showed a more human-like passing pattern.

Interestingly, both AI systems are fairly human, according to the passing direction criteria. The distribution of passing directions in real matches actually reflects the specific general logic of soccer matches and may be less dependent on individual team tactics. Also, the direction / distance classification does not fully reflect the complexity of the actual soccer path, and as suggested in [17], the path is classified according to other criteria such as risk estimation and path quality assessment needs to be classified [1].

Reference

[1] V. Khaustov, G. Mola Bogdan, M. Mozgovoy. “Pass in Human Style: Learning Soccer Game Patterns from Spatiotemporal Data”, Proceedings of the 2019 IEEE Conference on Games, London, UK, 2019, pp. 777-778.

[2] K. Wiggers, OpenAI Five defeats professional Dota 2 team, twice. [Online] Available: https://venturebeat.com/2019/04/13/openai-five-defeats-a-team-of-professional-dota-2-players/.

[3] P. Sweetser, D. Johnson, J. Sweetser, and J. Wiles, “Creating engaging artificial characters for games,” in International Conference on Entertainment Computing.

[4] B. Soni and P. Hingston, “Bots Trained to Play Like a Human are More Fun,” in IEEE International Joint Conference on Neural Networks, 2008, pp. 363–369.

[5] S. Temsiririrkkul, N. Sato, K. Nakagawa, and K. Ikeda, “Survey of how human players divert in-game actions for other purposes: Towards human-like computer players,” Lecture Notes in Computer Science, vol. 10507, https://link.springer.com/chapter/10.1007/978-3319-66715-7_27 , 2017.

[6] H. M. Le, Y. Yue, P. Carr, and P. Lucey, “Coordinated multi-agent imitation learning,” in Proceedings of the 34th International Conference on Machine Learning-Volume 70, 2017, pp. 1995–2003.

[7] B. Banerjee and M. E. Taylor, “Coordination Confidence based Human-Multi-Agent Transfer Learning for Collaborative Teams,” in Proceedings of the Adaptive Learning Agents (ALA-18) Workshop, 2018.

[8] M. Buckland, AI game programming by example. Plano, Tex, Lancaster: Wordware; Gazelle, 2004.

[9] P. Luhtanen, A. Belinskij, M. Häyrinen, and T. Vänttinen, “A comparative tournament analysis between the EURO 1996 and 2000 in soccer,” international Journal of performance Analysis in sport, vol. 1, no. 1, pp. 74–82, 2001.

[10] M. Mozgovoy and I. Umarov, “Behavior Capture with Acting Graph: A Knowledgebase for a Game AI System,” in Databases in Networked Information Systems (DNIS): 7th International Workshop, S. Kikuchi, A. Madaan, S. Sachdeva, and S. Bhalla, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 68–77.

[11] M. Mozgovoy, M. Purgina, and I. Umarov, “Believable Self-Learning AI for World of Tennis,” in 2016 IEEE Conference on Computational Intelligence and Games, 2016, pp. 1–7.

[12] STATS.com, Soccer Dataset. [Online] Available: https://www.stats.com/artificial-intelligence/.

[13] World Cup stats: Looking at the numbers from a record-breaking tournament - ProSoccerTalk | NBC Sports https://soccer.nbcsports.com/2014/07/14/world-cup-stats-looking-at-the-numbers-from-a-record-breaking-tournament/

[14] WhoScored.com, England-Premier-League-2014-2015 https://www.whoscored.com/Regions/252/Tournaments/2/Seasons/4311/Stages/9155/TeamStatistics/England-Premier-League-2014-2015

[15] Khashayar NikiMaleki, Mohammad HadiValipour, RoohollahYeylaghiAshrafi, SadeghMokari, M. R.Jamali, CaroLucas, “A simple method for decision making in RoboCup soccer simulation 3D environment”, 2008

[16] M. Prokopenko and P. Wang, “Disruptive innovations in RoboCup 2D soccer simulation league: from Cyberoos’ 98 to Gliders2016,” in Robot World Cup, 2016, pp. 529–541.

[17] S. Chawla, J. Estephan, J. Gudmundsson, and M. Horton, “Classification of passes in football matches using spatiotemporal data,” ACM Transactions on Spatial Algorithms and Systems (TSAS), vol. 3, no. 2, p. 6, 2017.