Identification of Distinctive Behavior Patterns of Bots and Human Teams in Soccer

Georgii Mola Bogdan [0000-0001-8527-5895] and Maxim Mozgovoy [0000-0003-0684-4012]

School of Computer Science and Engineering, The University of Aizu, Aizuwakamatsu 965-8580, Japan
{d8212101, mozgovoy}@u-aizu.ac.jp

Abstract. The design of human-like AI agents requires evaluation methods that check both robustness of the system and its believability. In this paper we attempt to examine whether is possible to assess similarity of play styles between different human teams and artificial teams in soccer. We rely on "behavior fingerprints" based on heat maps and their comparison using dot product. Our method shows no distinctive differences between the fingerprints of human teams, however, clearly indicates the difference between human teams and artificial teams. This approach is aimed to assist the design of human-like soccer teams but can also be useful in the domain of sports analytics.

Keywords: Soccer, Game AI, Human-likeness, Machine learning

1 Introduction

Being a worldwide popular sport, soccer extended beyond its original scope and became both a serious media business and cultural factor and a de facto testbed for AI research and competitions due to its relatively complex multiagent environment (as exemplified by RoboCup [1]). In addition, soccer-based video games are found among the most popular electronic entertainment products. As any video game it should possess some basic features or "fun factors" that keep the users engaged. One such feature is suspension of disbelief which is considered as indicator of high quality, contributing to player immersion [2]. Suspension of disbelief is usually associated with believability or human-likeness, which is crucial in case of soccer. Since soccer games are played by humans, their artificial representations in a game world are also supposed to be human-like. This task is partially accomplished by using highly detailed models of real athletes and real team structures/symbolics. However, we think that the ultimate contribution to believability can be achieved by interacting with AI agents that possess human-like behavior patterns. Arguably, the most straightforward approach to obtain such agents is to transfer knowledge from real players.

In recent years the role of information technologies in sports and soccer in particular is steadily growing. The development of video tracking systems provides spatiotemporal data that has numerous applications. For example, media companies became able to improve spectators' experience, sport analytics experts and coaches can extract information to have a bird's eye view of game tactics or locomotion performance [3]. There also exist works trying to address the problem of performance definition and measurement [4] and game situation modeling [5]. Thus, it may be possible to extract athletes' behavior patterns from real match recordings.

Believability assessment and play style analysis is an integral part of our ongoing research efforts [6, 7]. One of the intermediate goals in this work is to construct an automated assessment approach, which would allow to indicate whether the system under development possesses certain required traits. We are primarily focusing on believability rather than on efficiency due to our goal to provide "believable teams", which is reasonable from game development point of view. While reasonable AI skill level is expected, striving for the best performance is not the primary target for game AI, since it must provide entertainment for players of varying skills. Our current research questions include the following:

- (RQ1) How to design a method that would allow to assess team play styles?
- (RQ2) Does team style persist throughout match and/or between the matches?
- (RQ3) Is it possible to distinguish one team from another?
- (RQ4) Is it possible to distinguish human players from AI-controlled players?

The first question addresses the problem of whether is possible to define properties that represents teams' unique behavior patterns in principle, and whether it is easy to automate their analysis.

The second question is related to the idea that behavior patterns of team in different phases of the game may differ. For example, in the second half of a match physical fatigue of athletes may alter their movements. Next, behavioral patterns of a team may not persist between matches. There are circumstances that may influence team tactics, such as the wish to adapt to the next opponent or to meet some specific subgoal in the ongoing competition. These factors are a part of a "meta-game" and typically not reflected in spatiotemporal datasets.

The third and the fourth questions can be derived form the first two questions: if team behavior styles are identifiable, we should be able to distinguish them. In addition, we should be able to assess "human-likeness" by comparing behavior of human teams with their AI-controlled counterparts. Additional challenges emerge due to scant size of most available spatiotemporal data sets. Even if we obtain all game records of a particular team for the past year or two, the resulting collection would be relatively small for convention machine learning.

The core difficulty in team play style identification lies in a choice of properties representing a unique "team fingerprint". Player and ball tracking datasets provide "low-level" knowledge about team activities, while the logic behind them has to be reconstructed. Experts discuss play style identification, but this work still seems to be in its relatively early stage, mostly limited to discussion of possible options or to in-depth analysis of isolated aspects of the game [1, 2]. Still, the properties analyzed in these works constitute a good starting point.

Since the game of soccer consists mostly of manipulation with a ball and passes, we rely on characteristics of these events as basic features that may reflect individuality. Possession time of the ball is usually related to the efficiency of a team. In our case, we consider location-based possession of the ball represented with a heatmap. Similarly, we use a heatmap to represent successful pass and receive points. These actions correspond to the moments of the game when the players can do deliberate decision making and thus showing their tactical intentions [3].

2 Datasets

Tracking systems that gather data have been developed by different companies, and their data format is different. That is why preprocessing is necessary to obtain a unified data source. For instance, not all datasets have information about the third (height) coordinate of the ball, or its precision is low. Some sets, such as STATS [4] consist of small anonymized fragments of gameplay. In our work we rely on three independent data sources, obtained from STATS and DataStadium companies, and from Google Research Football environment.

2.1 A. DataStadium: complete matches

This dataset is collected and provided by DataStadium Inc. [5]. It consists of five full games played by six Japanese J1 League teams in 2011 season. Some statistical data like team names and formations is also available. The dataset has no event markup, so we had to reconstruct player movements, passes and shots using an automated method described in [6]. We analyze only the data of four teams that played twice to be able to compare team behavior in different matches.

2.2 B. Google Research Football: virtual teams

This dataset contain collection of 4800 game sequences recorded during Google Research Football with Manchester City F.C. competition [7]. This dataset contains event markup, which was possible to use after some cleanup.

Google Research's dataset provides a point of reference for the behavior of virtual teams. Unfortunately, in this competition the challenge was to create an AI system controlling the player with the ball (in case of attack) or the player closest to the ball (in case of defense), while the rest of the team is directed by the same built-in rule-based AI engine. Thus, behavioral diversity of virtual teams in the dataset is very limited. Still, the logic of player with the ball has a major impact on the whole team's tactics, so we can start with the presumption that the teams in the dataset are indeed distinct.

For our experiments we took two virtual teams (WeKick and SaltyFish) that played the largest number of matches in the dataset. As result we use 501 game sequences for SaltyFish and 432 sequences for WeKick team (each sequence has a duration of about five minutes).

2.3 C. STATS: Anonymized data

The STATS dataset [4] consists of 7578 short game episodes (from 5 to 150s), taken from 45 matches played in a top European league with total time of 2220 min. An individual episode starts when a certain team gets possession of the ball and ends when the team loses control of the ball.

Since each episode is anonymous, there is no way to extract both attacking and defending patterns of the same team or to obtain episodes where a specific team participates. Thus, we can treat this dataset as a “collective image” of a highly skilled human team. Like the DataStadium set, STATS does not include event markup, so passes, movements and shots on goal have to be reconstructed. One distinctive feature of this dataset is a low number of shots on goal. In most cases an episode ends when the ball is lost due to *any* reason including a shot on goal, so our options for analyzing episode outcomes are limited.

3 Team behavior fingerprinting

Before discussing possible approaches to fingerprint team behavior patterns, we have to make two preliminary notes. First, every team in subsequent experiments is treated as having its defensive zone on the right-hand side of the field. When processing left hand-side team data, we mirror all the coordinates to allow direct comparison between teams. Second, full-length matches from the DataStadium set were divided into halves to make possible to compare team behavior in the first and the second half of the game. We tried to split game recordings into smaller blocks (quarters), but it did not yield any significant changes in the results.

3.1 Ball possession-based fingerprinting

The idea of a ball possession metric is to gather information about locations, where the player (belonging to the team of interest) controlling the ball spends time. We divide the soccer field divided into 64 cells (see Fig. 1).

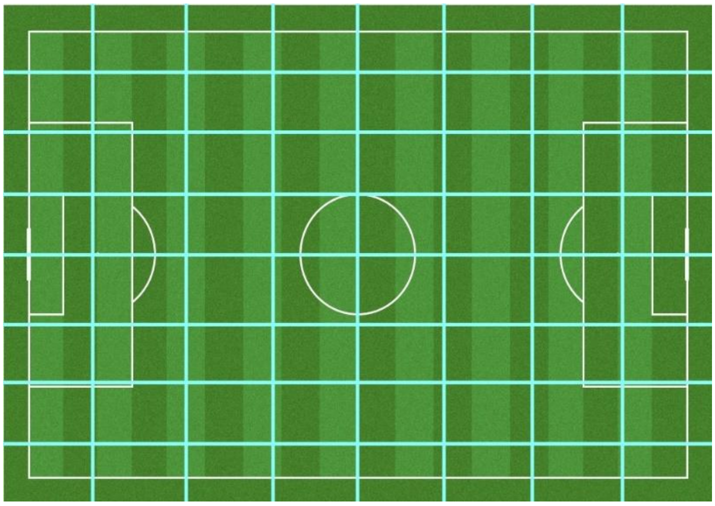

Fig. 1. Cells for ball possession analysis

On every frame of game recording, we increase the counter associated with a cell currently occupied by the player possessing the ball. We do not increase counters during passes (when the ball is not possessed by any particular player) and defensive actions (when the ball is possessed by the opposing team). At the end of this process we convert frequencies into percentages that can be visually represented as a heatmap (see Fig. 2). Such a heatmap can presumably reflect team-specific behavior patterns.

3.47	4.41	5.90	6.64	1.92	0.62	1.02	0.71
0.43	0.07	1.69	0.43	0.87	0.87	0.13	0.12
0.51	0.95	1.52	0.83	0.42	0.48	2.20	0.31
0.05	0.66	1.40	1.98	1.79	0.81	0.62	4.62
0.04	0.78	1.37	0.85	0.58	1.59	0.67	2.16
0.12	0.00	0.75	0.12	1.12	0.94	1.52	2.41
0.07	0.42	0.98	0.47	2.31	0.83	0.00	0.00
5.30	4.84	6.96	5.27	3.98	3.01	2.07	0.03

Fig. 2. Sample heatmap of ball possession

3.2 Pass-based fingerprinting

The idea of a pass/receive metric is similar to ball possession. Whenever a player performs a pass, we increase a counter for the cell occupied by the player. Similarly, we increase a counter for the cell where a pass is received (see Figure 3). The logic behind this metric is a presumption that players have more freedom in choosing the targets of their passes than in their movements. Thus, passes may reflect players' tactical preferences more accurately.

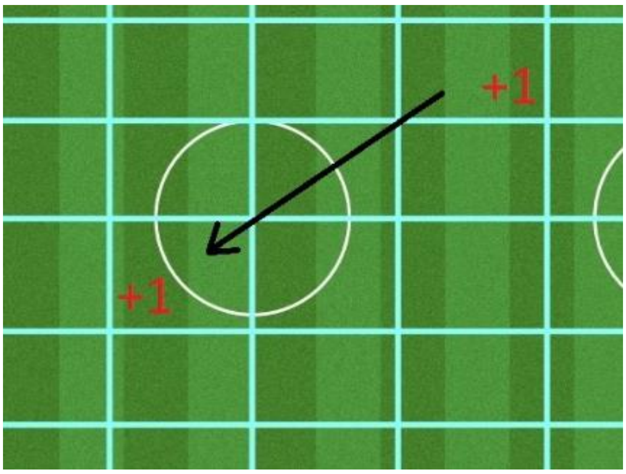

Fig. 3. Pass/receive frequency counting.
Cells marked with +1 will get their counters increased

3.3 Heatmap Comparison

To compare heatmaps, we use a conventional cosine similarity measure, based on a dot product formula:

$$a \cdot b = \sum_{i=1}^{n} a_i b_i \tag{1}$$

Each vector is obtained by placing heatmap rows sequentially. Since the soccer field is divided by 64 cells, we have to compare vectors of 64 elements. The result of such comparison is a number in a range [0, 1], corresponding to similarity estimation.

4 Results

As mentioned in the previous section, we split full matches of real teams into halves. Tables 1 and 2 report aggregated results obtained from the following seven teams:

Vd_1d_2	Human Team 1 (DataStadium set)
Nd_1d_2	Human Team 2 (DataStadium set)
Yd_1d_2	Human Team 3 (DataStadium set)
Sd_1d_2	Human Team 4 (DataStadium set)
ST	“Combined” Human Team (STATS set)
SF	SaltyFish Bot (Google set)
WK	WeKick Bot (Google set)

The first digit in a team abbreviation is used to specify the same team in different matches. The second digit (1 or 2) corresponds to the first or the second half of the match.

Table 1. Similarity of ball possession heatmaps (%)

V12	78																	
N11	77	79																
N12	65	79	73															
Y11	81	80	80	79														
Y12	70	77	69	77	83													
N21	68	77	74	78	83	75												
N22	68	81	61	73	74	82	75											
S11	71	79	82	74	85	79	79	71										
S12	69	78	61	74	78	84	71	85	74									
S21	72	83	75	85	88	80	82	81	84	76								
S22	71	80	71	70	82	79	72	85	77	76	86							
Y21	67	71	84	75	82	71	73	58	82	63	76	72						
Y22	74	82	75	80	83	76	79	83	75	81	82	79	83					
V21	68	73	73	79	83	79	77	71	78	70	89	77	75	74				
V22	70	78	74	70	86	84	74	67	85	75	80	77	79	75	81			
ST	80	89	86	86	93	88	86	82	90	82	92	87	88	89	89	90		
SF	52	38	64	47	43	38	39	28	40	34	29	30	55	48	33	29	46	
WK	54	38	55	43	38	38	34	29	35	37	24	29	50	47	28	29	43	86
	V11	V12	N11	N12	Y11	Y12	N21	N22	S11	S12	S21	S22	Y21	Y22	V21	V22	ST	SF

Table 2. Similarity of pass/receive heatmaps (%)

V12	78																	
N11	74	86																
N12	77	76	72															
Y11	86	81	77	82														
Y12	80	77	68	79	81													
N21	74	80	79	84	83	74												
N22	75	74	65	72	73	73	74											
S11	80	79	79	81	80	87	84	76										
S12	79	85	75	79	79	81	78	83	86									
S21	83	86	82	84	87	83	88	73	89	83								
S22	85	82	74	79	88	85	72	75	78	79	82							
Y21	76	86	82	77	73	80	81	70	85	82	90	73						
Y22	75	87	75	77	78	72	79	83	75	83	83	79	80					
V21	70	68	65	82	79	70	82	66	77	72	77	70	69	67				
V22	80	81	75	77	82	86	77	73	86	86	86	78	82	75	64			
ST	88	92	83	87	90	90	87	82	92	90	94	91	90	88	81	89		
SF	57	50	47	50	43	39	46	49	45	46	39	43	50	43	46	40	54	
WK	61	54	53	59	48	49	52	61	51	55	46	52	55	52	51	47	62	93
	V11	V12	N11	N12	Y11	Y12	N21	N22	S11	S12	S21	S22	Y21	Y22	V21	V22	ST	SF

5 Discussion

Ball possession has been chosen as the most straightforward approach and was supposed to provide a common-sense basic representation of a team style. However, the results show that most human teams follow similar behavior patterns, yielding similarity values in the range 64-90%. It is also seemingly hard to derive “team-specific” patterns from heatmap visualizations (see Fig. 4). Comparison of human teams’ heatmaps obtained with the pass/receive metric yields more diverse values lying in a range 58-89%.

0.1	0.0	3.8	3.3	8.4	3.0	0.9	0.0
0.6	0.0	1.0	2.2	1.0	2.7	1.3	0.5
0.1	1.0	0.1	0.8	0.5	2.3	1.3	1.7
0.2	0.8	1.8	6.8	0.7	1.3	0.0	4.2
0.0	1.0	3.3	2.7	4.0	0.6	1.0	2.8
0.0	0.3	1.2	1.5	0.6	0.6	0.1	0.2
0.0	1.0	1.8	0.8	1.5	0.1	0.1	0.4
2.7	3.1	2.4	7.1	4.6	0.2	1.3	0.9

1.2	3.0	4.5	4.1	6.6	3.2	0.7	0.6
0.0	0.4	0.9	1.9	3.1	1.6	3.0	0.0
0.0	0.7	0.8	1.3	1.1	3.3	1.4	0.6
0.8	0.1	0.8	1.2	1.6	0.6	0.5	2.5
0.4	0.0	1.5	1.0	2.6	2.9	0.3	3.4
0.8	1.3	0.6	1.9	2.6	4.1	0.8	0.9
0.4	0.9	0.4	0.1	3.6	2.7	1.8	0.3
0.4	1.5	1.0	3.1	4.1	2.1	0.0	0.6

Fig. 4. Heatmaps of V11 (up) and N11 (down) (sim: 74%)

Judging from the tables, there are no clear patterns that can be used to distinguish one human team from another or identify the same team in a different half or a match. We can suggest some possible explanations for this result.

The most straightforward theory would be to presume that there are *indeed* no clearly identifiable styles among professional teams in modern soccer. While this idea may sound implausible, some specialists voice similar opinions. For example, a well-known Italian manager and a former player Roberto Mancini argued in 2007 that future advances in football would come from physical preparation of players rather than tactics [8]. The underlying reasoning is that the maturity of football as a game and widespread adoption of fresh ideas drive soccer to a certain global uniform style, where country-specific or team-specific innovations are unlikely *(ibid)*.

A possibly more likely explanation would be to treat the current approach as not nuanced enough to capture subtle differences between the teams rather than patterns typical for most soccer team. In this case we can only hope for more accurate results in follow-up studies.

It is also difficult to obtain reliable results due to relatively small data collections that we have for human teams. Unfortunately, tracking data is not easy to obtain, and even same-team datasets are uneven due to seasonal changes in team lineup, injuries, and diversity of opponents.

At the same time, teams of bots form an isolated subset, clearly distinguishable from human teams using both proposed methods. Similarity values between bot teams and human teams are low, and high between the different bot teams. As noted before, "bot teams" actually consist of the same rule-based players, having player with the ball as the only exception. Thus, we can only confirm that the particular AI system does not possess human-like behavior traits according to our calculations.

As a result, the proposed research questions can be answered as follows.

RQ1. We have limited success with a heatmap/cosine similarity-based evaluation algorithm. It provides reasonable results but fails to distinguish human teams in our dataset. While this method is simple, it could produce reliable player identification in boxing and tennis games [9, 10].

RQ2, RQ3. The answers are "no" under presumption that heatmap comparison is a reliable method to analyze team play style. We see that the same teams behave differently in different matches and/or match halves, while distinct teams often show very similar behavioral patterns. However, a more sophisticated method of evaluation might be required.

RQ4. Yes, at least for the AI system built into Google Research Football. It is hard to say how representative are the two bot teams present in our datasets. Since we mostly rely on passes and movements of the player with the ball, the similarity between SaltyFish and WeKick could have been expected to be lower. On the other hand, both these bots optimize performance and rely on other players' cooperation. Their behavior is described as "simple and reasonable" by the Google Research team [11], so we can suggest that a "typical" AI system should exhibit similar traits. Thus, human-likeness is not a natural property of a team of bots, and even our simple method can easily distinguish virtual teams from real teams.

6 Conclusion

This work is aimed to establish an experimental framework for believability and play style evaluation in the game of soccer. We have integrated several player tracking data sources and obtained a diverse collection consisting of anonymized game fragments, game recordings of known human teams, and game recordings of AI-based teams. We have also reconstructed game events, such as passes, movements, and shots on goal in the datasets where this information was missing.

This setup allows us to evaluate different comparison procedures. We aim to identify team behavior traits, reflecting team individuality, certain tactical and strategic patterns, distinguishing this team from other teams and persistent across matches. Our earlier experiments with other game genres such as tennis and fighting revealed such stable player-specific patterns. However, these studies were focused on computer games rather than real-world events, and were limited to player-vs-player genres where the concept of "team behavior" does not exist.

The identification of such team patterns has application in sports analytics, but our immediate goal is team game AI, where the task of designing a diverse set of virtual opponents, using different attacking strategies, and possessing different skills can be a major challenge. In 2013 Sicart observed that a computer game FIFA'12 is already highly realistic in terms of physics and animation but falls short in the AI department. He considers FIFA's AI system too deterministic, scripted and predictable [12].

Our present study relies on two simple heatmap-based methods of assessing behavior similarity. The first method builds a heatmap of ball possession, and the second method creates a similar heatmap of ball pass/receive events. Heatmaps are compared using a cosine similarity metrics.

This general approach has known limitations. For example, it treats all heatmap elements as independent and does not take into account higher-level factors such as team formation. Still, it proved to be a reliable behavior fingerprinting strategy in simpler games.

While in the present work we could not reveal clearly identifiable differences between individual teams or noticeable similarities in behavior of the same team in different matches, we succeeded in separating real teams from virtual teams. This result is, however, limited to Google Research Football's built-in AI system with individual players controlled by WeKick and SaltyFish AI. We tend to believe that other conventional soccer AI systems (such as the ones used in commercial games) will provide similar results, but this question needs further investigation.

The existence of clearly identifiable team-specific behavior patterns also remains an open question. The conventional understanding of soccer implies the existence of distinctive "national styles", which seem to be much more pronounced in the past [13]. However, the convergence of tactics and strategy in modern soccer is also recognized [8]. Thus, it is not clear whether our current results are caused primarily by the limitations of the methods we use or by the convergence of play styles exhibited by different teams.

References

1. Kempe, M., Vogelbein, M., Memmert, D., Nopp, S.: Possession vs. direct play: evaluating tactical behavior in elite soccer. International Journal of Sports Science, vol. 4, 35–41 (2014)
2. Hewitt, A., Greenham, G., Norton, K.: Game style in soccer: what is it and can we quantify it? international Journal of performance Analysis in sport, vol. 16, 355–372 (2016). doi: 10.1080/24748668.2016.11868892
3. Link, D., Hoernig, M.: Individual ball possession in soccer. PLOS ONE, vol. 12, e0179953 (2017). doi: 10.1371/journal.pone.0179953
4. STATS.com: Soccer Dataset, https://www.stats.com/artificial-intelligence/
5. DataStadium Inc.: Offical website, https://www.datastadium.co.jp
6. Khaustov, V., Mozgovoy, M.: Recognizing events in spatiotemporal soccer data. Applied Sciences, vol. 10, 8046 (2020)
7. Google Research: Google Research Football with Manchester City F.C. Competition, https://www.kaggle.com/c/google-football
8. Wilson, J.: Inverting the pyramid. The history of soccer tactics. Nation Books, New York (2013)
9. Mozgovoy, M., Purgina, M., Umarov, I.: Believable Self-Learning AI for World of Tennis. In: 2016 IEEE Conference on Computational Intelligence and Games, pp. 1–7 (2016)
10. Mozgovoy, M., Umarov, I.: Building a Believable Agent for a 3D Boxing Simulation Game. In: Second International Conference on Computer Research and Development, pp. 46–50 (2010)
11. Kurach, K., Raichuk, A., Stańczyk, P., Zając, M., Bachem, O., Espeholt, L., Riquelme, C., Vincent, D., Michalski, M., Bousquet, O., et al.: Google Research Football: A Novel Reinforcement Learning Environment. AAAI, vol. 34, 4501–4510 (2020). doi: 10.1609/aaai.v34i04.5878
12. Sicart, M.: A tale of two games: football and FIFA 12 Sports Videogames, pp. 40–57. Routledge (2013)
13. Jordan, A.: Soccer Politics. National Playing Styles, https://sites.duke.edu/wcwp/tournament-guides/olympic-football-2016-guide/team-playing-styles-in-soccer/