Learning Believable Player Movement Patterns from Human Data in a Soccer Game*

Victor Khaustov, Maxim Mozgovoy

Active Knowledge Engineering Lab, University of Aizu, Aizuwakamatsu, Japan

d8201106@u-aizu.ac.jp, mozgovoy@u-aizu.ac.jp

***Abstract*—Player movement patterns are one of the behavioral traits immediately visible to an observer. Thus, a soccer AI system striving for believable (human-like) behavior must ensure the believability of player movements. We show how tracking data of real human players in soccer can be used to create a case-based reasoning AI system, able to simulate realistic player movements in a computer soccer game. Our results are confirmed with a direct comparison of actions made by AI-controlled players and professional athletes.**

Keywords—soccer, player tracking, believability, human-like AI, case-based reasoning.

I. Introduction

The last years are marked with the growing availability of sports tracking data. Such datasets are already used in professional sports analytics to provide valuable insights and guidance for players by estimating the effectiveness of particular actions [1]. Several annotated corpora are now freely available for research purposes, which facilitates exploration of their potential use outside the sports analytics area.

A game AI system is a good example of technology that might benefit from human tracking data in sports games. Virtual representations of real-life sports events belong to one of the oldest and most well-established genres of video games. Apparently, the target audience of computer sports games consists largely of sports fans: major PC and game console franchises, such as *FIFA*, *NBA* or *NFL* rely heavily on licensed teams and players and get yearly updates. Thus, an AI system playing a virtual sports game might also get insight and guidance using data obtained from professional players.

In addition to this general consideration, we must note that game AI systems have to satisfy quite specific requirements, stemming from the fact that their principal purpose is to *entertain the player* rather than *to be successful*. Russel and Norvig [2] discuss classic AI technologies mostly using the model of a "rational agent" that *"does the right thing,"* which *"is expected to maximize its [agent's] performance measure"* (p. 37). In the case of a computer game, the overall success of an AI system is determined with its contribution to player enjoyment, even if it means *"playing to lose,"* as Johnson [3] puts it.

The question *"what kind of AI contributes to user enjoyment"* is a subject of a separate discussion. Here it suffices to say that the apparent goal of nearly all mainstream projects is *"to make the games even closer to the actual game, that is, to make the computer game converge with the sport"* [4]. Thus, implementing a "human-like" AI decision-making system to make virtual players act like their real-life counterparts can be a reasonable goal. It can be pursued even further to the level of "virtual stars," imitating particular famous athletes or "virtual teams," playing in style resembling a particular team.

* This paper was accepted at ICACT-2020 and will be presented in February 2020.

The purpose of the present paper is to report our preliminary results on constructing the human-like behavior of a player currently possessing the ball in a game of soccer. We rely on STATS-supplied soccer tracking data [5] to identify movements performed by human players in specific game situations and apply the obtained knowledge to act in previously unseen situations. The human-likeness of our virtual players is confirmed by comparing probability distributions of movement directions made by human and AI-controlled players in similar locations.

II. Previous Work

Our present work can be seen as one of the steps towards the creation of a human data-driven AI system for the game of computer soccer. The idea to use the STATS dataset was inspired by the work [6], focused on realistic coordinated multi-agent defensive behavior in soccer.

One of the most interesting research questions in this work is to find out how well the patterns of human team behavior can be translated into a simplified world of a video game. Our current experimental setup is based on Buckland's SimpleSoccer simulator, serving as a good model of a simple arcade-style soccer game [7] (Ch. 4).

Previously, we used the same environment to achieve human-like passing behavior by learning from human tracking data [8]. Therefore, it was considered reasonable to rely on the same approach to machine learning: we represent virtual agents' knowledge as a graph, having individual game situations as vertices, and actions as weighted edges. This structure represents the fact that situation A turns into situation B as a result of a certain action during the learning phase [9, 10]. During decision making, the system tries to find the best match for the current game situation and acts accordingly. This procedure can be seen as a variation of a Markov decision process [11].

III. Dataset Processing

The "Soccer Dataset," provided by STATS.com [5], consists of 7500 game sequences, represented as series of game situation snapshots, taken at the rate of 10 snapshots per second. Each snapshot (frame) contains the coordinates of all 22 players and the ball. The sequences are taken from actual European league matches and represent around 36 hours of playing time. Every sequence starts when a certain team gets the ball, and ends when the opponent takes control of the ball.

For the current work, we need to identify player movements in the dataset. Unlike real soccer players, player characters of Buckland's SimpleSoccer can only move with constant speed in eight predefined directions. Thus, we must approximate actual human movements with SimpleSoccer actions. This is done with a simple procedure:

1. We identify the movement performed by the player possessing the ball during the last second (10 frames) of the game.
2. We calculate the best matching direction **Dir** for this movement out of 8 possible options.
3. We calculate the duration **D** reflecting the number of frames in SimpleSoccer to cover the required distance.
4. The resulting movement (**Dir**, **D**) is recorded, and the procedure is repeated for the next time interval.

IV. Representing the Game States and Actions

During the learning stage, our system uses game states and actions from the STATS dataset, taken at the one-second interval, to create the game graph. The following attributes are used to represent individual game situations:

- **PwbX**[*range*], **PwbY**[*range*]: the coordinates of the player possessing the ball (in the specified range).
- **DMF**: the "danger to move forward" heuristic estimation (0-5); depends on the distance to the nearest opponent in the forward direction.
- **MoveDir**: the current movement direction of the player with the ball (8 directions are supported).
- **CDir**: the direction (0-7) of the closest opponent, from the perspective of the player possessing the ball.
- **CDist**: the distance to the closest opponent, from the perspective of the player possessing the ball (0-2).
- **SPD**: the "safest pass danger" heuristic estimation on the scale of 0 to 5 (depends on locations of both teammates and opponents).
- **SFD**: the "safest forward pass danger" heuristic estimation on the scale of 0 to 5.
- **PassOpt**: the Boolean attribute indicating that at least one safe pass (with danger estimation of 0-2) is found.
- **FFX**, **SFX**: x-coordinates of two teammates closest to the opponent's goal line, converted to the range [0, 3].

Every game state in the graph is recorded in three different representations, reflecting different approximations of the actual soccer game state:

R_0 PwbX[0-17], PwbY[0-9], DMF, MoveDir, CDir, CDist, SPD, SFD, PassOpt, FFX, SFX

R_1 PwbX[0-17], PwbY[0-9], DMF, FFX, SFX

R_2 PwbX[0-8], PwbY[0-4], DMF

During decision making, the system tries to find a match for the current onscreen game situation using representation R_0. If no matches are found, it proceeds to R_1, and, if necessary, to the "fallback" representation R_2.

V. Experiments and Results

Our primary goal was to confirm that the proposed algorithm is able to control the player possessing the ball and thus can be considered a part of the general AI system, controlling virtual soccer players. Fortunately, SimpleSoccer includes a built-in rule-based AI, so it is possible to set up a match where the player possessing the ball is controlled by our algorithm while the remaining players are controlled by SimpleSoccer AI.

In our test runs, the new data-driven AI player exhibits behavior that looks natural. After acquiring both moving and passing behavior (as described in [8]), the player is able to attack, evade opponents, and make reasonable passes. Its movements are consistent, and there are no erratic changes in the movement direction. The player does not lose the ball unexpectedly, and in general, we are satisfied with the results.

However, we also wanted to perform an objective evaluation of the AI-controlled player's behavior to make sure it satisfies the stated requirement of "human-likeness." In order to do it, we decided to compare how AI-supplied actions correlate with movements performed by real athletes in actual games of the STATS dataset.

We divided the original dataset into a subset of 4000 randomly chosen sequences used to train the AI system, and a test set of the remaining 3500 sequences. The decision-making process was initiated on every 10th frame of each test set sequence, and the results were compared with the actions chosen by the original players possessing the ball.

It is important to note that direct comparison cannot be applied in this scenario: there is no predefined "right" action in the given game situation, since player movements are probabilistic in nature, and the same player might decide to behave differently in similar cases. Furthermore, STATS data is anonymized, so it is not possible to learn actions specific to a particular athlete; instead, we acquire "generalized" behavior exhibited by all ball-possessing players.

Therefore, we decided to compare the probabilities of movement directions of human- and computer-controlled players, observed in the same zones of a soccer field. For the purpose of this task we treat the field as consisting of twelve zones (see Figure 1).

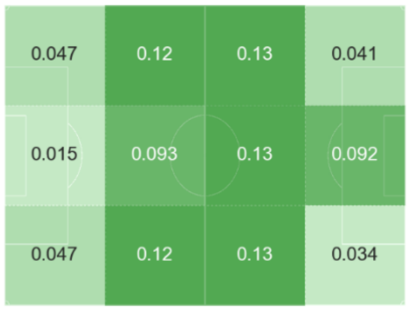

Figure 1. A heatmap of movements of a player controlling the ball. The attacking team's goal line is on the right.

For each zone, we created a radar chart, showing the probabilities of movement in each of 8 possible directions. The resulting visualization is shown in Figure 2. One can note that predominant movement directions of human players indeed depend on their zones: in their own half, players tend to move the ball away from the central zones closer to field flanks and attempt to return the ball closer to the center when approaching the opponent's goal line.

It is also quite clear that the radar charts of human and AI players are very similar. One way to obtain a numerical similarity ratio estimation is to use the cosine similarity. We can represent the complete behavioral fingerprint of a player as a vector V, having elements calculated as:

$$V_{8i+j} = p(Z_i) \times p(D_j),$$

where $p(Z_i)$ is the probability of player to be in the zone $0 \le i \le 11$, and $p(D_j)$ is the probability of choosing the direction $0 \le j \le 7$ for movement action.

In our case, cosine similarity between the "human vector" and the "AI vector" yields a rate of 0.98. This value should be considered a rough estimation of a "true similarity," since it does not take into account fine-grained contextual information about situations where decisions are made, but it still shows that our AI system is indeed able to replicate human behavior to a certain extent.

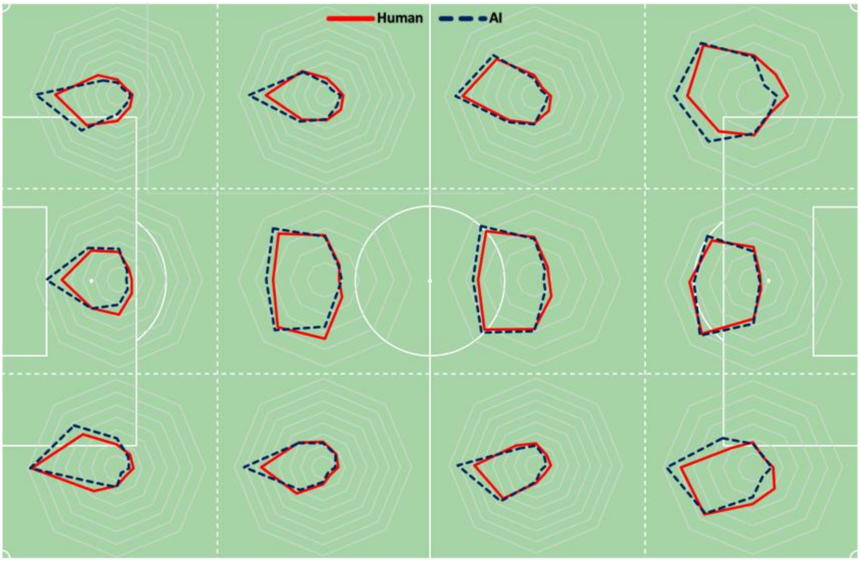

Figure 2. Movements of ball-possessing human- and AI-controlled characters. The attacking team's goal line is on the right.

VI. Discussion and Conclusion

Designing a comprehensive data-driven AI system for soccer is a challenging task, which requires the right choice of methods as well as a good understanding of the game. So far, we have succeeded in using the STATS dataset to obtain a reasonably human-like moving and passing behavior of a player controlling the ball. (Unfortunately, the dataset does not contain any shots on goal, so these actions have to be handled separately for now.) Stable and believable behavior is confirmed both with qualitative evaluation and with numerical analysis, yielding a 0.98 similarity ratio between human and AI players according to our fingerprint metric.

Our algorithm relies on three handcrafted sets of features, reflecting our understanding of soccer situations, as seen by a player controlling the ball. We recognize that this fact makes our approach a somewhat "ad-hoc solution," so one of our future priorities will be to make the process of feature selection more transparent and streamlined. In the current experiments, the AI system is able to find a matching game state using R_0 in approximately 19% of cases, while R_1 is responsible for 63% of cases, and "fallback" R_2 is used in the remaining 18% of game situations.

Naturally, the most interesting directions for future research are related to multi-agent behavior. Controlling both a player possessing the ball and the rest of the team requires a certain degree of player-player coordination, which is outside the scope of our present work.

We should also note that the task of translating real-life actions into a game world can be challenging, too. Real athletes possess different abilities, and they operate in a physical world, only partially replicated in a computer game. However, if game designers strive for realism, they might find studying human tracking data insightful.

References

[1] Z. Lowe, *Lights, Cameras, Revolution.* [Online] Available: https://grantland.com/features/the-toronto-raptors-sportvu-cameras-nba-analytical-revolution/.

[2] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach,* 3rd ed.: Pearson Education UK, 2016.

[3] S. Johnson, "Playing to Lose: AI and "CIVILIZATION" (Lecture)," Game Developers Conference, 2008.

[4] M. Sicart, "A tale of two games: football and FIFA 12," in *Sports Videogames*: Routledge, 2013, pp. 40–57.

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

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

[8] V. Khaustov, G. M. Bogdan, and M. Mozgovoy, "Pass in Human Style: Learning Soccer Game Patterns from Spatiotemporal Data," in *IEEE Conference on Games*, 2019.

[9] 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.

[10] 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.

[11] M. L. Puterman, "Markov decision processes," *Handbooks in operations research and management science*, vol. 2, pp. 331–434, 1990.