Quantifying Pitch Control in Soccer

Balázs Imre
Supervisor: Balázs Csanád Csáji

***Abstract*—This report, created as part of the Eötvös Loránd University Applied Mathematics Project Work I course, presents the initial phase of the project, focusing on quantifying and visualizing pitch control in soccer. The analysis utilizes Voronoi tessellation and Javier Fernandez and Luke Bornn's pitch control model. The report provides a brief discussion of the implementation of these methods, supported by visualizations. In addition, the final section includes a match analysis to demonstrate the practical application of the models.**

***Index Terms*—Sports data analysis, Soccer, Pitch control, Voronoi tessellation, Javier Fernandez, Luke Bornn, Data visualization**

I. Introduction and Motivation

Soccer analytics has traditionally focused on on-ball events, such as pass and shot efficiency or the dribbling success rate. However, the importance of off-ball events has grown significantly in modern soccer. As Johan Cruyff said, "It is statistically proven that players actually have the ball 3 minutes on average. So, the most important thing is what you do during those 87 minutes when you do not have the ball. That is what determines whether you're a good player or not." [1] This perspective emphasizes the value of analyzing off-ball dynamics. While the methods discussed in this report were developed for soccer analytics, they also have potential applications in other fields as well, such as traffic management, marketing, and healthcare.

A. Background

Pitch control refers to the ownership of space by teams. In regions controlled by Team A, the players of that team can act quickly and occupy positions earlier than their opponents. Quantifying the pitch control ratio is crucial for analyzing teams' tactical approaches and evaluating players' abilities to gain an advantage in different areas of the field through effective positioning.

B. Project Goals

In this initial phase, we aim to implement two approaches to pitch control: a basic method using Voronoi tessellation and a more advanced method based on Javier Fernandez and Luke Bornn's concept of player influence area.

II. Data Description

One of the main challenges in this project is the limited availability of high-quality publicly accessible tracking data. However, for this initial phase, the dataset provided by Metrica Sports is sufficient. The data can be accessed through this [2] GitHub repository.

A. Dataset Characteristic

At the time of writing, the GitHub repository contains three anonymized soccer matches. The anonymization ensures that there are no references to the names of players, teams and competitions. The dataset includes approximately 25 data points per second, recording player positions with x and y coordinates in a range from 0 to 1, with the kickoff point at $(0.5, 0.5)$. Additionally, synchronized events, such as passes, shots, and interceptions are included with details about the start and end times, and about the involved players.

B. Basic Visualization

Data visualization plays a crucial role in understanding and communicating results effectively. For this purpose, I used Python's *Matplotlib* and *mplsoccer* libraries. Visualization provides insights into strategies and tactics. For instance, plotting the average positions of players allows straightforward categorization into player roles, such as goalkeepers, defenders, midfielders and attackers. (Fig. 1.)

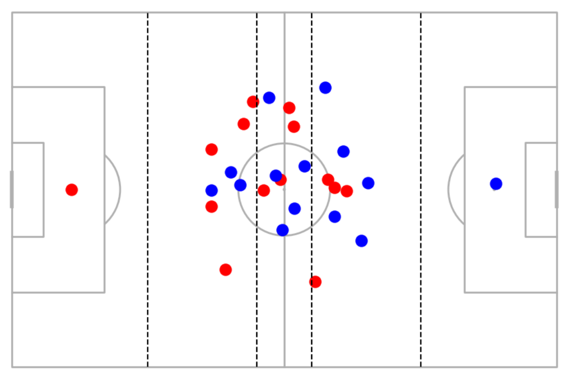

Fig. 1. Player's Average Positions

III. Pitch Control Quantification with Voronoi Tessellation

The Voronoi tessellation method is our first approach to quantifying pitch control. This technique is intuitive and straightforward. For each player, there is a corresponding region, called Voronoi region, which consists of all points on the pitch closer to that player than to any other.

A. Formal Definition and Alternative Approaches

The definition and computation of Voronoi tessellation are from Mark de Berg, Marc van Kreveld, Mark Overmars and Otfried Schwarzkopf's Computational Geometry book. [3]

Let $P = \{p_1, p_2, ..., p_n\}$ be a set of points, where each p_k represents a pair of real numbers. The Voronoi region of $p_k \in P$ is defined as:

$$V(p_k) \doteq \{x \in \mathbb{R}^2 \colon d(x, p_k) < d(x, p_l),\ \forall l \in \{1, ..., n\},\ l \neq k\},$$

where $d(x, p_k)$ denotes the Euclidean distance.

In our case, the points $h_1, ..., h_{11}$ and $a_1, ..., a_{11}$ correspond to the players of the home and away teams. Thus, the Voronoi region of player p_k is defined as:

$$V(p_k) \doteq \{x \in F \colon d(x, p_k) < d(x, p_l),\ \forall p_l \in \text{players},\ l \neq k\},$$

where F refers to the set of all points on the field.

We further define H and A as the unions of the Voronoi regions controlled by the home and away players, respectively. These regions can be interpreted as the areas of the pitch dominated by the team.

A more generalized approach for defining influence regions is presented in the work of Tsuyoshi Taki and Jun-ichi Hasegawa. [4] In their method, the Euclidean distance is replaced by the function $t(x, p_k)$, representing the shortest time necessary for player p_k to move from their current position to the point x.

B. Implementation and Visualization

For the implementation, it was crucial to ensure that the Voronoi regions remained within the boundaries of the pitch. To address this, the players' positions were reflected across the four boundaries of the pitch. This ensured that the bisectors of the original and reflected points lay on the pitch edges, thereby constraining the Voronoi regions to the field.

In the following visualization (Fig. 2.), we can see a scenario where the away team (in blue) is attacking and the home team (in red) is defending. The majority of the pitch is controlled by the away team, as indicated by their larger Voronoi regions. The ball is represented by the yellow point.

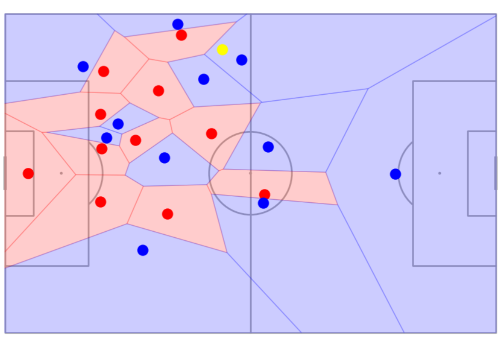

Fig. 2. Voronoi Regions Visualization

IV. Pitch Control Quantification with Player Influence Areas by Fernandez and Bornn

Javier Fernandez and Luke Bornn introduced an advanced method to measure space generation and occupation in this [1] article, enabling to evaluate player's off-ball movements. This relies on the concept of player influence areas.

A. Model Description

A player's influence on nearby areas depends on several factors, such as their location, velocity, and distance to the ball. Players farther from the ball have influence over larger areas, as they have more time to reach the ball within a wider region. Furthermore, player's velocity is also a crucial factor, players sprinting at full speed in a particular direction exert greater influence on the corresponding area compare to those walking or jogging.

The influence of player k at a given location x and time t is defined as

$$I_k(x, t) \doteq \frac{f_k(x, t)}{f_k(x_k(t), t)},$$

where $x_k(t)$ refers to the position of player k at time t, and $f_k(x, t)$ is the density function of a bivariate normal distribution. The covariance matrix and expected value dynamically change based on the player's velocity, direction, and distace from the ball. The exact mathematical formulation can be found in the article's appendix.

B. Implementation and Visualization

I implemented their approach by creating a mesh grid representing points on the pitch, and aggregated the influence values for each player at each grid point.

Now, we revisit the scenario from the previous section. (Fig. 3.) The red points represents the players of the home team, while the blue points denote the away team's players. The yellow point indicates the ball's position, and the green arrows represent the players' velocities and directions. Due to visualization constraints, these arrows are illustrative and do not represent the exact velocity values. The resulted plot highlights the controlled areas; while home team controlled areas are represented by higher values, away team controlled areas are shown with lower values.

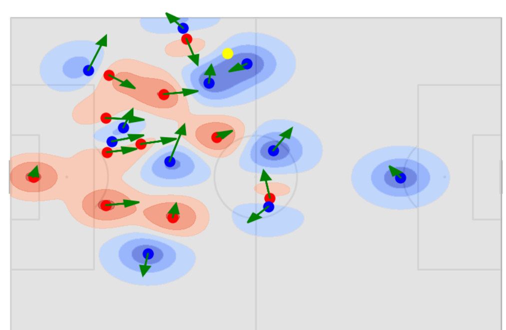

Fig. 3. Player's Influence Regions

V. Match Analysis and Future Work

A. Match Analysis

Using the Voronoi tessellation approach, it was straightforward to quantify the area of controlled regions. For the player's influence areas model, I introduced a threshold to determine

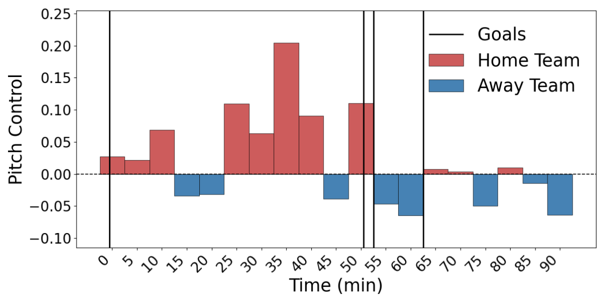

Fig. 4. Pitch Control Ratio - Match Analysis

the area of controlled regions. A region was considered controlled by the home team if the aggregated value of influences was higher than the threshold and by the home team if it fell below the negative threshold.

The x-axis represents time, while the y-axis represents the average pitch control ratio in 5-minute intervals in (Fig. 2.). The match resulted a 4-0 victory for the home team, with goals scored in the 2nd, 55th, 58th and 65th minutes. The data reveal that the home team dominated pitch control during the first half and the early stages of the second half, which correlates with the periods when they scored goals. After the home team secured victory, they shifted to a more defensive strategy, allowing the away team to dominate possession and pitch control. Toward the end, the away team more aggressively attempted to score at least one goal, led to increased pitch control in the final minutes. This analysis demonstrates how pitch control can correlate with strategy and match outcomes.

B. Limitations and Next Steps

As I mentioned, the bottleneck of this project could be the limited availability of high-quality, publicly accessible tracking data. A goal for the next semester could be to train a machine learning model to generate tracking data from matches.

An alternative direction is to quantify the value the points of the pitch. For instance, space occupied near the opponent's goal is far more valuable than space occupied by a team's goalkeeper. Developing metrics to evaluate the values of regions could be beneficial for further soccer analytic methods.

REFERENCES

[1] Javier Fernandez and Luke Bornn. Wide open spaces: A statisticial technique for measuring space creation in professional soccer. *MIT Sloan Sports Analytics Conference*, 2018.

[2] Metrica Sports. Metrica sports sample data. https://github.com/metrica-sports/sample-data. [Online; accessed 04-Dec-2024].

[3] Mark de Berg, Marc van Kreveld, Mark Overmars, and Otfried Schwarzkopf. *Computational Geometry, Algorithms and Applications*. Springer, 2000.

[4] Tsuyoshi Taki and Jun-ichi Hasegawa. Quantitative measurement of teamwork in ball games using dominant region. *International Archives of Photogrammetry and Remote Sensing*, 5:125–131, 2000.