Fundamental framework for the space evaluation in football games

Takuma Narizuka\textsuperscript{1}, Yoshihiro Yamazaki\textsuperscript{2}, and Kenta Takizawa\textsuperscript{1}

\textsuperscript{1}Department of Physics, Faculty of Science and Engineering, Chuo University, Bunkyo, Tokyo 112-8551, Japan
\textsuperscript{2}Department of Physics, School of Advanced Science and Engineering, Waseda University, Shinjuku, Tokyo 169-8555, Japan

Abstract

In football analysis, space evaluation is one of the major important issues. The present study proposes a fundamental framework for the space evaluation. We employ the motion model and calculate a minimum arrival time $\tau$ of each player to all locations in the football field. In particular, two variables $\tau_{df}$ and $\tau_{of}$ corresponding to the minimum arrival time for defense and offense teams are considered. Using $\tau_{df}$ and $\tau_{of}$, we define new variables $z_1$ and $z_2$, which represent degrees of safety and sparsity of each location. Detailed characterization of $z_1$ and $z_2$ is given in terms of ball passing. As applications of our method, a field division and an evaluation for ball passing are discussed.

Keywords— Space evaluation, Motion model, Minimum arrival time, Football

1. Introduction

Football (soccer) is a complex system in which 22 players in two different teams interact with each other in order to win. In the process of carrying a ball to each goal, vast number of options are allowed for players, and various behaviors emerge from individual to team levels: ball passing, dribble, marking an opponent player, and arranging formation. Such nature of football and the recent development of data acquisition tools have promoted a wide range of analyses for football games from the viewpoint of statistics and physics \cite{1,2}. Examples include stochastic process approach to a scoring event \cite{3}, complex network analysis of the ball passing \cite{4}, fractal analysis of the ball and players’ motions \cite{5}, and characterization of formations using Delaunay network \cite{6}.

Characterization of space in the field is one of the major important issues in football analysis. A well-known approach for space evaluation is a “dominant region”, introduced by Taki et al. \cite{7,8}, in which a certain player can arrive prior to any other players. The typical example is the Voronoi region, which corresponds to the dominant region defined by the Euclidean distance \cite{9,11}. Meanwhile, Fujimura and Sugihara proposed more realistic definition of the dominant region based on the “motion model” \cite{12}. In the motion model, each player is assumed to move according to the following equation of motion:

$$m \frac{d^2 \bar{x}(t)}{dt^2} = F \vec{n} - k \frac{d \bar{x}(t)}{dt}.$$  \hspace{1cm} (1)

Here, $m$ is the mass of the player, and $\bar{x}(t)$ is the player’s position at time $t$. $F$ and $\vec{n}$ are the magnitude and direction of the attractive force. $k$ is the coefficient for viscous resistance. The solution of Eq. (1)

\begin{itemize}
  \item $E-mail address$: pararel@gmail.com (T. Narizuka).
\end{itemize}
under the initial position \( \vec{x}_0 \) and the initial velocity \( \vec{v}_0 \) is expressed as
\[
\vec{x}(t) = \vec{x}_0 + \frac{1 - \exp(-\alpha t)}{\alpha} \vec{v}_0 + V_{\text{max}} \left( t - \frac{1 - \exp(-\alpha t)}{\alpha} \right) \vec{n},
\]
where \( \alpha = k/m \) and \( V_{\text{max}} = F/m \) are arbitrary constants. Given \( \alpha \) and \( V_{\text{max}} \), we obtain arrival times of each player to all locations from Eq. (2). Therefore this equation brings about modification of the dominant region by velocity and acceleration of players. In the study by Fujimura and Sugiura, they set \( \alpha = 1.3 \) and \( V_{\text{max}} = 7.8 \) [m/s], which were empirically estimated as the typical values in the sprint condition. For extension of the dominant region, there are several similar studies [2,13].

The dominant region is one approach for space evaluation when positions and motions of players are given. However, this approach seems to be oversimplified, because all locations in a dominant region have the same evaluation in that one specific player reach them fastest. Namely, in this approach, football field is divided into several dominant regions depending on positions and motions of players.

In our study, instead of such a field division by dominant regions, another framework for space evaluation is proposed. Our space evaluation is based on a “minimum arrival time” \( \tau \) to all locations in the field. \( \tau \) is calculated from the motion model. Specifically, we consider two variables \( \tau_a \) and \( \tau_{df} \) corresponding to the minimum arrival time for defense and offense teams. Using \( \tau_a \) and \( \tau_{df} \), we introduce new variables, \( z_1 \) and \( z_2 \), and apply them to the space evaluation.

2. Framework of space evaluation using minimum arrival time

Let \( \tau_a(\vec{x}, t) \) be the minimum arrival time for a player \( a \) to a location \( \vec{x} \) at time \( t \). \( \tau_a(\vec{x}, t) \) is defined as the time required for the player \( a \) to move from the position at \( t \) to \( \vec{x} \). In order to calculate \( \tau_a(\vec{x}, t) \), we employ the solution of the above motion model (Eq. (2)), where \( \alpha = 1.3 \) and \( V_{\text{max}} = 7.8 \) [m/s]. Meanwhile, we also define the minimum arrival time for a team \( A \) to \( \vec{x} \) at time \( t \) as \( \tau_A(\vec{x}, t) \equiv \min_{a \in A} \tau_a(\vec{x}, t) \). In particular, we denote the minimum arrival time for the defense and offense teams as \( \tau_{df}(\vec{x}, t) \) and \( \tau_{of}(\vec{x}, t) \) (offense team is defined as the ball-possession team). Figure 1 is a visualization of \( \tau_{df} \) at a certain time by contour plot in the range of 2 seconds.

Fig. 1. Visualization of \( \tau_{df} \) at a certain time by contour plot in the range of 2 seconds. The players in offense (ball-possession) and defense teams are shown by blue leftward and red rightward triangles, respectively. The blue open circle shows the position of the ball.
Now we evaluate the location $\vec{x}$ at time $t$ by the two variables $\tau_{df}(\vec{x}, t)$ and $\tau_{of}(\vec{x}, t)$. In Fig. 2, the domain $\tau_{of}(\vec{x}, t) < \tau_{df}(\vec{x}, t)$ corresponds to the safe space for the offense team in that an offense player can arrive there prior to any other defense players. Specifically, the degree of safety of $\vec{x}$ at $t$ for the offense team can be quantified by introducing a new variable $z_1(\vec{x}, t)$ as

$$z_1(\vec{x}, t) = \frac{\tau_{df}(\vec{x}, t) - \tau_{of}(\vec{x}, t)}{\sqrt{2}}.$$  

Similarly, another variable $z_2$, which is perpendicular to $z_1$, can be introduced as

$$z_2(\vec{x}, t) = \frac{\tau_{df}(\vec{x}, t) + \tau_{of}(\vec{x}, t)}{\sqrt{2}}.$$  

Since $z_2(\vec{x}, t)$ is proportional to $\tau_{df} + \tau_{of}$, it roughly quantifies the degree of sparsity of the location $\vec{x}$ at $t$. Relationship between $\tau_{of}, \tau_{df}$ axes and $z_1, z_2$ axes are shown in Fig. 2. In the next section, we give further characterization of $z_1$ and $z_2$ using real ball-passing data.

![Fig. 2. Relationship between $\tau_{of}, \tau_{df}$ axes and $z_1, z_2$ axes.](image)

3. Data analysis

The following analysis was based on datasets composed of 45 football games of J1 League 2018, provided by DataStadium Inc., Japan. There are five games per team and all the games were taken place between Aug. 10 and Sept. 2, 2018. Each dataset contains all player positions every 0.04 seconds and play-by-play data. The datasets are not open; the DataStadium provided us with a permission to use them for this research. All calculations were conducted using python packages on a MacBook pro with a 2 GHz Intel Core i5 processor and 16 GB of memory.

Based on the above datasets, we created time-series data including 34189 ball passes made in 45 football games. A row of the data corresponding to a pass is expressed as $[t_o, \vec{x}_o, t_e, \vec{x}_e, q]$. Here, $t_o$ and $\vec{x}_o$ represent the time and the positional coordinates for the origin of the pass, and $t_e$ and $\vec{x}_e$ are those for the end of the pass. The variable $q \in \{1, 0\}$ shows the success (1) and failure (0) of the pass. From Eqs. (3) and (4), we calculate $z_1(\vec{x}_e, t_o)$ and $z_2(\vec{x}_e, t_o)$ for all successful and failed passes; Fig. 3(a) shows the result.

We first focus on $z_1(\vec{x}_e, t_o)$. Figure 3(b) presents the probability distributions of $z_1(\vec{x}_e, t_o)$ for each of successful and failed passes. We find that both distributions can be fitted well by a normal distribution function, whereas the peak values are located at $z_1 > 0$ for successful passes and at $z_1 < 0$ for failed passes. It appears that the success and failure of pass are strongly correlated with $z_1(\vec{x}_e, t_o)$. Then, we can estimate the success probability of passes as a function of $z_1(\vec{x}_e, t_o)$ by averaging the value of $q$ over
each $z_1$. Figure 3(c) shows the result, and it is found that the success probability can be fitted well by a sigmoid function given by

$$f(z) = \frac{1}{1 + \exp[-a(z + b)]}$$

(5)

where $a = 4.68$ and $b = 0.10$. Thus, with respect to ball passing, $z_1(\vec{x}, t)$ corresponds to the degree of safety for the pass which is made to $\vec{x}$ at $t$.

Next, we focus on $z_2(\vec{x}_e, t_0)$. As we mentioned above, $z_2$ represents the degree of sparsity at $\vec{x}_e$. In order to elucidate the meaning of sparsity, we calculate the following quantity for each pass:

$$\tilde{R} = \frac{||\vec{x}_e - \vec{x}_e(t_0)||}{\sigma(t_0)}.$$  

(6)

Here, $\vec{x}_e$ and $\sigma$ are the centroid position and the standard deviation for all field players excluding goal keepers; they are defined as follows:

$$\vec{x}_e(t) = \frac{1}{N} \sum_{j=1}^{20} \vec{x}_j(t),$$

(7)

$$\sigma(t) = \sqrt{\frac{1}{20} \sum_{j=1}^{20} |\vec{x}_e(t) - \vec{x}_j(t)|^2}.$$  

(8)

Figure 4(a) shows the schematic representation for the definition of $\tilde{R}$. When $\tilde{R}$ is less (more) than one, the end point $\vec{x}_e$ of the pass is roughly located at inside (outside) the formation. From our data, the relationship between $\tilde{R}$ and $z_2$ are obtained as shown in Fig. 4(b). It is found that the value of $z_2$ at which $\tilde{R} \simeq 1$ is obtained as $z_2 \simeq 2$, namely $\tau_{df} + \tau_{of} \simeq 2\sqrt{2}$. $z_2 \simeq 2$ gives the threshold whether the location is inside the formation or not. Owing to this result, the space with $z_2 > 2$ and $z_2 < 2$ can be regarded as sparse and dense, respectively.
\( \sigma(t_o) \) 

\( \| \vec{x}_c - \bar{x}_c(t_o) \| \) 

Fig. 4. (a) Schematic representation for the definition of \( \tilde{R} \) (Eq. (6)). The cross and star markers represent the end points of \( \vec{x}_c(t_o) \) and \( \bar{x}_c \), respectively. The dotted line shows \( \sigma(t_o) \). (b) Relationship between \( \tilde{R} \) and \( z_2(\bar{x}_c, t_o) \).

4. Discussion and Summary

Using our framework, a new field division based on the degrees of safety and sparsity can be introduced according to the ball-passing analysis. Specifically, using the two axes, \( z_1 = 0 \) and \( z_2 = 2 \), the football field can be divided into the following four spaces: (A) safe dense space (\( z_1 > 0 \) and \( z_2 < 2 \)), (B) safe sparse space (\( z_1 > 0 \) and \( z_2 > 2 \)), (C) risky sparse space (\( z_1 < 0 \) and \( z_2 > 2 \)), and (D) risky dense space (\( z_1 < 0 \) and \( z_2 < 2 \)). We show a typical example of the field division (A) ~ (D) in Fig. 5. It is found that the dense spaces (A) and (D) are almost located within the formation.

In the situation shown in Fig. 5, for example, the blue leftward triangle overlapped with the blue open circle shows the player having the ball. In order to make a shoot from this situation, the ball has to be sent to a space in front of the opponent goal; however, such a space is risky because the space corresponds to (C) or (D). We show that such feature is common in football games. Especially for the passes just before the shoots, Figs. 6(a) and (b) present the scatter plot of \( z_1(\bar{x}_c, t_o) \) and \( z_2(\bar{x}_c, t_o) \), and probability distribution of \( z_1 \) for the successful passes. We find from Fig. 6(b) that the peak value \( z_1 \approx 0 \) is smaller than that for the case of all passes (\( z_1 \approx 1 \) from Fig. 3(b)). This feature indicates that passes just before a shoot have a tendency to become risky because the success probability of the pass is the monotonically increasing function of \( z_1 \) as shown in Fig. 3(c).
Fig. 5. Typical example of the field division into (A) ~ (D). The players in offense (ball-possession) and defense teams are shown by blue leftward and red rightward triangles, respectively. The blue open circle shows the position of the ball.

Finally, we explain some future topics of the present study by comparing our framework with relevant studies. First, Fernandez et al. have proposed a similar approach for the space evaluation in football games [14]. In their method, player’s influence on each location is defined by a multivariate normal distribution and is transformed into a single value for evaluating which teams are dominant at the location. In contrast, our method evaluates a location in the field by the two variables, $z_1$ and $z_2$, which are defined by the motion model. It is noteworthy that the physical background and the meaning of $z_1$ and $z_2$ are explicit as shown above; however several extensions are still required for the practical applications to the real game analysis. For example, the information of the distances from the ball and the goal to each location should be considered as in the case of Fernandez et al., because the outcome of passes or shoots generally depends on these factors. Regarding the ball’s motions, Spearman et al.
have proposed another model based on the equation of motion for the ball [15]. Such an extension is also possible for our framework and it enables us to evaluate accuracy of passes. Lastly, estimation of α and $V_{\text{max}}$ and introduction of the motion model for each player probably lead to the assessment of ability of players.

In summary, we have proposed a framework of space evaluation in football games based on the minimum arrival times $\tau_{df}$ and $\tau_{df}$, which are calculated by the motion model. We have introduced the new variables $z_1$ and $z_2$ and obtained the values of them using ball-passing data. We find that $z_1$ represents the degree of safety of a location; the success probability of a pass as a function of $z_1$ follows the sigmoid function. For $z_2$, it is found to be associated with the sparsity of a location. In particular, $z_2 \approx 2$ gives the threshold between sparse and dense spaces. We expect that our framework provides the theoretical foundation for the space evaluation not only for football but also for other team sports such as basketball and hockey.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors are very grateful to DataStadium Inc., Japan for providing the player tracking data. This work was partially supported by the Data Centric Science Research Commons Project of the Research Organization of Information and Systems, Japan, a Grant-in-Aid for Young Scientists (18K18013) from the Japan Society for the Promotion of Science (JSPS).

References

[1] D. Sumpter: Soccermatics: Mathematical adventures in the beautiful game (Bloomsbury Sigma, London, 2017).

[2] J. Gudmundsson and M. Horton: ACM Computing Surveys (CSUR) 50 (2017) 22.

[3] A. Clauset, M. Kogan, and S. Redner: Physical Review E 91 (2015) 062815.

[4] J. Duch, J. S. Waitzman, and L. a. N. Amaral: PloS One 5 (2010) e10937.

[5] A. Kijima, K. Yokoyama, H. Shiima, and Y. Yamamoto: The European Physical Journal B 87 (2014) 41.

[6] T. Narizuka and Y. Yamazaki: Scientific reports 9 (2019) 1.

[7] T. Taki, J. Hasegawa, and T. Fukumura: Proceedings of the 3rd IEEE International Conference on Image Processing 3 (1996) 815.

[8] T. Taki and J. Hasegawa: Proceedings of the Computer Graphics International 2000 (2000) 227.

[9] A. Okabe, B. Boots, K. Sugihara, and S. Nok-Chiu: Spatial tessellations: concepts and applications of Voronoi diagrams (John Wiley & Sons, New York, 2000).

[10] S. Kim: Nonlinear Analysis: Modelling and Control 9 (2004) 233.

[11] S. Fonseca, J. Milho, B. Travassos, and D. Aratijo: Human Movement Science 31 (2012) 1652.

[12] A. Fujimura and K. Sugihara: Systems and Computers in Japan 36 (2005) 49.
[13] U. Brefeld, J. Lasek, and S. Mair: Machine Learning 108 (2019) 127.

[14] J. Fernandez and L. Bornn: Sloan Sports Analytics Conference, 2018.

[15] W. Spearman, A. Basye, G. Dick, R. Hotovy, and P. Pop: Proceeding of the 11th MIT Sloan Sports Analytics Conference, 2017.