Evaluating Passes: Difficulty Factor

Stephanie Tanasia Saputra
40199284

I. Introduction

Passes are the most frequent event in soccer yet there are no fixed metric such as expected goals (xG) to assess them. Identifying a pass is important in player's development and also team's development. In this project, the author would like to expand and improve the example metric that were shown in class: Impact Factor $= \sum_{i_x} n_x$ where n_x is the number of opponents in the x-direction. This metric is simple since we're only calculating the number of opponents in the x-direction.

In this project, the author would like to include more factors to the metric. Let's define the impact factor as the difficulty of the pass. Therefore, if the impact factor is small then the pass was objectively easy to do based on the factors that we included. The factors that the author considered are number of surrounding opponents in a radius of r, the distance of the pass, the field position of the player, and the type of pass.

The objective of this project is:

- To calculate the number of surrounding opponents between the radius of r meters. This factor will be denoted by n_x.
- To calculate the distance of a pass. This factor will be denoted by d_x.
- To determine the location of the pass by looking at the start point, denoted by f_x.
- To assign difficulty level based on the type, denoted by t_x.
- Assign impact factor and rate to each player.
- Compare the impact factor and rate between the 2 teams.

This could be an interesting metric since now we have 4 variables instead of one. In addition, due to the nature of our defined metric, we could assess the player's skill by looking at the impact factor of their passes. This could be useful in the future for player scouting as well as knowing which player is the most valuable in the game. Furthermore, we can use this information to re-design players' current training.

II. Literature Review

Applying data analytic skill to build a model or a metric to evaluate a pass is not as prominent as to evaluating goals. Most of the published works are more focused in a risk-reward assessment for passing decisions, rather than creating a metric to evaluate a pass. However, those works have given the author that don't have any background in soccer to know what influences a pass.

An article written by Brownell in 2013, discussed soccer statistics and key passes [1]. He mentioned how companies like *Opta*, records every single action in a match, as well as where the action took place on the field that is usually observed by breaking it up in thirds. The passing accuracy rate is an interesting statistic, but the location of the pass is where it really matters [1]. It could be argued that passing accuracy in the defensive third of the field is actually more critical than midfield third.

In 2014, Reis et al conducted a research on long distance passes performed throughout 64 matches in the 2014 Brazil FIFA World Cup [2]. Long passes were defined as a pass from the defensive field to the attacking field. A total of 4,512 long passes were examined; teams lost control of the ball in 59% of the attempts, retained control in 28%, recovered control in about 12% of the actions, and only 1% of passes led to shoots on goal.

Goes et al developed a risk-reward assessment in passing decisions where they used tracking data from an entire season of Dutch Eredivisie matches [3]. The risk they're observing is the chance that a pass is intercepted by the opponent. This is not the only study that looks at the risk of a pass- another amazing study was done by Power et al in 2017 [4]. Both Power et al and Goes et al agrees that the probability of interception can be related to the field position, pass length, technical difficulty, pressure on the passer and coverage of the receiver [3] [4].

The novel study done by Anzer and Bauer in 2022 [5] may be too advanced for the author. They managed to determine a player's risk profile, evaluate their passing ability, and calculate the likelihood that hypothetical passes would be completed. Spatio-temporal data of the ball trajectory was used to estimate the intended receiver of the pass since it's impossible to capture that information. This study is promising since it offers a chance to use the model to evaluate player's pass decisions. There is another advanced study is done by Merlin and et al in which they applied multivariate statistical analysis for 32 independent variables [6].

Analogous to the problem discussed by Anzer and Bauer, and Merlin the author was inspired to create a metric from some of the factors used by Goes and Power.

III. The Metric

The impact factor will be calculated for each player, and is formulated as below.

$$\text{Impact Factor} = I_x = n_x + d_x + f_x + t_x \tag{1}$$

where x indicates the player x. The author used the second match of the Metrica data for this project where it has 26 players and 963 passes.

A. Surrounding Opponents

The biggest factor of interception of a pass has to be the opponents. It is trivial that the more opponents a player has around them, the higher the chance the pass is going to be interrupted. The problem now lies on how to measure the surrounding opponents.

To calculate the number of opponents surrounding the player who has the ball, the author created a circle of radius 10 meters and see how many opponents are inside the circle. The 10 meters threshold was chosen based on the median of the length of passes in Metrica data as seen in Figure 1.

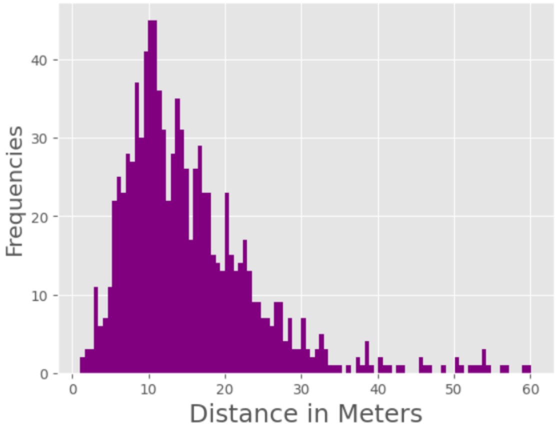

Figure 1: Distance of Passes in Metrica Data

To do this, we need to calculate the Euclidean distance between the start point of the pass and all of the opponents. Then, filter out those who are located outside the circle. The Euclidean distance between Player1:(x_1, y_1) and Player2:(x_2, y_2) is

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{2}$$

After finding the distance, values were assigned to n_x based on Table I.

Number of Surrounding Opponents in 10m radius	n_x
0	0
1,2	1
≥ 3	2

Table I: Values Assigned for Surrounding Opponents

B. The Distance of The Pass

As mentioned in section II, long distance pass is pivotal since it can change the course of the game easily. Looking at the definition used by Reis et al, the author decided that a long distance pass is any pass that is greater than 35 meters (roughly the distance from A to C in Figure 2 or a third of the soccer field's length) [2]. Therefore,

$$d_x = \mathbb{1}_{d \geq 35}$$

where d is the distance of the pass, calculated with equation 2.

C. Field Position

Locations on the pitch influences the passes. As mentioned in Section II, players might have a better accuracy when passing in the midfield compared to defensive-third or attacking-third, since they are experiencing less pressure and more away from the goalkeeper.

Thus, the author divide the pitch into three and assign different values based on the start point of the pass. Only the start point is observed since we're interested in the player who's doing the pass, not the receiver. This also helps with the fact that a pass could start and end in different regions. See below for how the soccer pitch is divided:

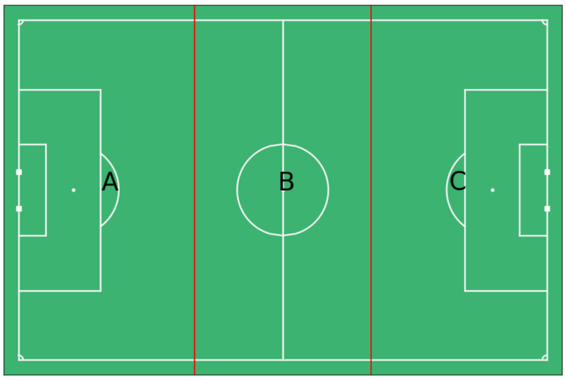

Figure 2: Soccer Field in Thirds

If the player starts in A or C, they are in defensive-/attacking- third. If the player starts in B, they are in the middle third. The values for each cases can be seen in Table II.

Start Position	f_x
A	1
B	0
C	1

Table II: Values Assigned for Field Positions

D. Type of Passes

The author assume that there are some types of passes that are done by only great players which makes them different from others. This skill can make their performance outstanding. Therefore we would like to assign values to the type of passes in Metrica based on their difficulty level. Unfortunately, the author did not find any relevant work to assign grades for each type of the passes.

A common approach in this situation is to ask an expert. We seek advice from Dr. Joshua Wyatt Smith to grade different types of passes (scale from 1 to 5) and his response can be seen in Table III, which were used in this project.

Type of Passes	Grade based on difficulty, t_x
Goal Kick	5
Deep ball	4
Cross	3
Head	2
Nan	1

Table III: Values Assigned for Type of Passes

E. Normalization

Since the variables do not have the same scale, the author normalized the data before calculating the impact factor. The purpose of normalization is to convert the values of the factors to a common scale without distorting the range differences.

Let x be the original factor and x^* be the normalized factor. Then, for $i = 1, 2, 3, \ldots n$

$$x_i^* = \frac{x_i - min(x)}{range(x)} \tag{3}$$

This normalization makes all of the factor be in the range of $[0, 1]$.

Equation (1) becomes

$$\text{Impact Factor} = I_x = n_x^* + d_x^* + f_x^* + t_x^* \tag{4}$$

Furthermore, we can calculate the impact rate of a pass with

$$\text{Rate} = R_x = \frac{I_x}{4} \times 100 \tag{5}$$

We will take a look at the impact rate of the players and compare the rate between the home and away team.

IV. Results and Analysis

After calculating the impact factor I_x based on equation 4, the author sliced the data frame to look at the player's impact rate, R_x. The mean of the impact rate is 30.1%. Player 11 from the home team has the highest impact rate at 53.33%, followed by Player 25 of the away team with 46.43%. This means that they executed the most difficult passes throughout the game. Meanwhile, Player 2 has the lowest impact rate with only 19.67%. See the figures below for each player's rate.

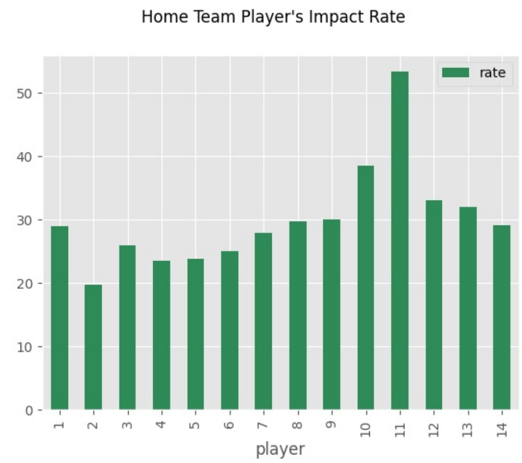

Figure 3: Home Players' Impact Rate

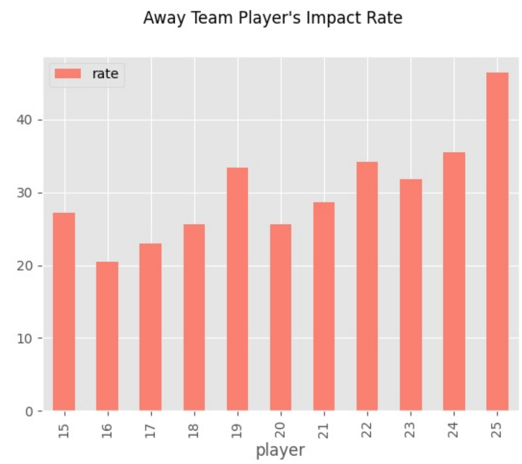

Figure 4: Away Players' Impact Rate

Furthermore, we can compare the impact rate between the home team and the away team. To calculate this, add all of the impact factors of the players based on their team, then divide them with $4 \times$ numbers of passes executed to get the rate. In the end, it was found that the home team has an impact rate of 27.35%: out of 100 passes, around 27 of them were difficult. Meanwhile, the away team has an impact rate of 25%, which is slightly lower than the home team.

Looking at these rates, we can conclude that overall, the home team executed slightly more difficult passes than the away team. This means that the match was at a somewhat equal difficulty level for both team.

V. Discussion

Firstly, this metric falls short since the values are assigned for each factor by personal research. What is hard for one player might not be as hard for the other. However, this is part of the Metrica data limitations as well since the game is anonymous. For further study, we can assign the value better based on the experience study of the team or players' statistics. In addition, maybe adding the factors isn't the optimal way of assessing the difficulty of passes. Further comparison studies should be conducted to see how these factors should be combined or weighted in the best way.

Secondly, since the player's orientation is unknown, we couldn't include the type of passes based on their angle (i.e. backward pass, left/right pass) into the metric. In the future, with enough data and computing power, it is possible to estimate the player's orientation and include this factor as seen in [5].

Another impact of the unknown player's orientation is that we don't know the intended receiver of the pass. The event and tracking data don't capture the failure of a pass- it captures simply where it went. Thus, we do not know if it was successful in terms of the intended receiver. The only thing that can be concluded from the data and our metric is the difficulty level of the executed pass.

VI. Conclusion

For this project, an impact factor that is more complex than what we saw in class was successfully created. We considered 4 factors that influence a pass and found that overall, the home team in the second match of Metrica data, executed a slightly more difficult passes than the away team. Since there's only 2% difference in the impact factor, this match was played at a similar difficulty level for both teams.

It's safe to conclude that Player 11 performed the most difficult passes compared to other players. In practice, this can be used to "level upöther player's passing training difficulty.

References

[1] P. Brownell, "The most important new advanced soccer statistics and why they matter," 2013, last accessed on 1 October 2022. [Online]. Available: https://bleacherreport.com/articles/1597790-the-most-important-new-advanced-soccer-statistics-and-why-they-matter

[2] M. A. Reis, F. V. Vasconcellos, and M. B. Almeida, "Analysis of the effectiveness of long distance passes in 2014 brazil fifa world cup," *Brazilian Journal of Kinanthropometry and Human Performance*, vol. 19, no. 6, p. 676–685, 2017.

[3] F. Goes, E. Schwarz, M. Elferink-Gemser, K. Lemmink, and M. Brink, "A risk-reward assessment of passing decisions: comparison between positional roles using tracking data from professional men's soccer," *Science and Medicine in Football*, vol. 6, no. 3, pp. 372–380, 2022, pMID: 35862167. [Online]. Available: https://doi.org/10.1080/24733938.2021.1944660

[4] P. Power, H. Ruiz, X. Wei, and P. Lucey, "Not all passes are created equal," *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2017.

[5] G. Anzer and P. Bauer, "Expected passes," *Data Mining and Knowledge Discovery*, vol. 36, no. 1, p. 295–317, 2022.

[6] M. Merlin, A. Pinto, A. G. de Almeida, F. A. Moura, R. D. S. Torres, and S. A. Cunha, "Classification and determinants of passing difficulty in soccer: a multivariate approach," *Science and Medicine in Football*, vol. 0, no. 0, pp. 1–11, 2021. [Online]. Available: https://doi.org/10.1080/24733938.2021.1986227