Smart Data Scouting in Professional Soccer: Evaluating Passing Performance based on Position Tracking Data

Matthias Kempe
University of Groningen, UMCG,
Center for Human Movement Sciences
Groningen, The Netherlands Country
m.kempe@umcg.nl

Floris R. Goes
University of Groningen, UMCG,
Center for Human Movement Sciences
Groningen, The Netherlands
f.r.goes@umcg.nl

Koen A. P. M. Lemmink
University of Groningen, UMCG,
Center for Human Movement Sciences
Groningen, The Netherlands
k.a.p.m.lemmink@umcg.nl

***Abstract*—Sports analytics in general and soccer analytics, in particular, have evolved in recent years due to the increased availability of large data amounts of (tracking) data. Especially in terms of evaluating tactical behavior, data science could change the way we think about soccer. In this study, we evaluate passing performance in soccer to prove the hypothesis that tactical behavior in team sports can be analyzed based exclusively on tracking data. To prove this point, we explore the relationship between changes in spatiotemporal variables in relation to passing and key performance indicators. Based on our results that demonstrate the ability of spatiotemporal variables to predict pass accuracy and key performances indicators on an individual level, we confirmed our hypothesis. Furthermore, we calculated a simple composite performance indicator to evaluate passes and players based on tracking data. In conclusion, our results can be used as an approach for real-time evaluation of tactical behavior and as a new method to scout and evaluate players in soccer and team sports in general.**

Keywords—football, performance indicators, hypothesis testing, machine learning, big data

I. INTRODUCTION

In recent years, soccer has become an attractive field for data and computer scientists, as demonstrated by the fact that the number of published contributions has been increasing annually since 1995 [1]. This increased interest could be due to two factors. On the one hand, the interactions of twenty-two players and the ball form a highly complex and dynamic system. On the other hand, the use of (semi-) automated tracking systems that capture the coordinates of all players, referees and the ball has rapidly increased over the recent years.

Recently, several approaches have been developed using tracking data to assess passing (tactical) performance more objectively [2] with the help of data science approaches. Although these approaches are very promising, they are all based on the probability of scoring a goal or creating a goal-scoring opportunity as an assessment factor. This is problematic as these factors 1) are usually evaluated by human observers and 2) do not allow a continuous, real-time (in-game) evaluation. The aim of this study is to prove the hypothesis that new approaches are able to overcome these two deficits with the help of data science methods. Therefore, it will be evaluated if spatiotemporal variables based exclusively on – or computed directly from – tracking data are able to predict goal scoring opportunities and other key performance indicators in soccer. In addition, these variables are used to determine the key attributes of an effective and successful pass and to develop a prediction model for successful passing.

II. DATA-PROCESSING

A. Player tracking data

For this work, we utilized player tracking data from 82 professional soccer matches collected with a SportsVU optical tracking system (SportsVU, STATS LLC, Chicago, IL, USA). This system tracks the X and Y coordinates, the velocity and acceleration of all 22 players, the referee and the ball at 16 Hz, resulting in over 4 million data points per match.

B. Match event data

Notational event data is annotated manually based on observational analysis of match broadcasts and is comprised of data on common match events like passes, yellow cards, assists, tackles, etc. We collected notational event data for the past 4 seasons of the Dutch Eredivisie and determined the number of key passes (passes leading to a shot) and the number of assists per game for every individual player.

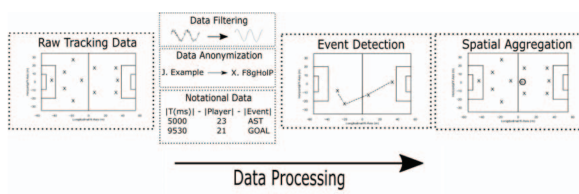

Figure 1 Data-processing steps used to pre-process position tracking data collected in soccer

C. Pass attributes

Based on the location of the pass and the subsequent location of the reception or interception we computed the pass attributes length (m), angle (degrees), and velocity (m/s). A pass was deemed accurate when it was received by a teammate, while it as deemed inaccurate when it was intercepted by an opponent.

To measure pass effectiveness two separate variables were computed. The first (I-Mov) measures the absolute displacement of all players of the defending team during a 3-second window, while the second (C-Mov) measures the absolute displacement of the defending team centroid during a 3-second window. Both variables were computed as the sum of an X (I-MovX & C-MovX) and Y (I-MovX & C-MovY) component representing the absolute movement in the X- and Y-axis. Displacement was computed as the absolute difference in X and Y positions of all defending players and the defending team centroid between the moment of pass initiation and 3 seconds after the pass.

978-1-5386-9156-4/18/$31.00 ©2018 IEEE
DOI 10.1109/eScience.2018.00126

III. Evaluating player pass effectiveness

To test our hypotheses, we pre-selected players with at least 100 accurate passes (N = 24,334 accurate passes distributed over 86 players) from our dataset. For these players, we computed the mean I-Mov and C-Mov effectiveness variables, the mean pass accuracy, the number of key passes, and assists per game. Subsequently, we fitted a Ridge Regression model with generalized cross validation to study the relationship between total I-Mov, total C-Mov, and the X and Y components of both variables with pass accuracy. In addition, we fitted the same model to study the relationship between those variables and key passes and assists per game. Total data was randomly split between a training set (70% of the data) and a test set (30%).

Table 1 - Descriptive statistics of average pass performance

Statistic	Mean ± SD (N = 86 players)
Pass accuracy (%)	69.26 ± 8.51
I-Mov (m)	82.16 ± 4.15
C-Mov (m)	6.13 ± 0.39
Key passes per Game	0.85 ± 0.65
Assists per game	0.10 ± 0.10

A. Pass effectiveness vs. pass accuracy

Based solely on pass effectiveness variables computed from the tracking data, our model predicted average pass accuracy with an accuracy of 75.6%. We found a trade-off between average pass effectiveness and average pass accuracy. These results indicate, performing a pass that mainly causes movement along the X-axis of the field corresponds to a lower pass accuracy while passes that mainly cause movement along Y-axis of the field correspond to a higher pass accuracy (see Figure 2).

B. Pass effectiveness vs. traditional performance indicators

Based solely on pass effectiveness variables, we were able to predict key passes per game with a 37.2% accuracy and assists per game with a 30.2% accuracy. However, those results are still promising as previous prediction models using 'traditional' notation data showed lower prediction rates. These results demonstrated that individual movement as a result of a pass, especially along the X-axis, contributes to the number of key passes and assists per game. In contrast, collective movement, especially along the Y-axis, seems to have a negative relationship with the number of key passes and assists per game.

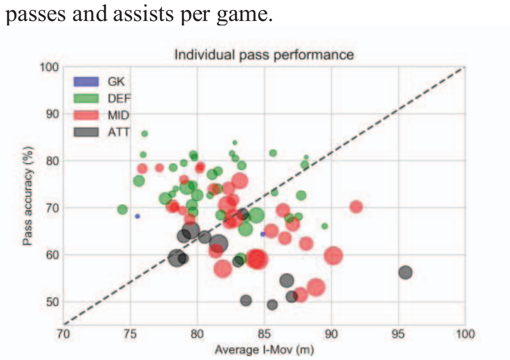

Figure 2 - Average pass accuracy vs. average pass effectiveness (I-Mov) per player. Color represents field position, size of the marker represents the number of key passes per game (more key passes = bigger marker).

IV. Predicting & evaluating pass performance

We demonstrated in the previous section that our model provides a valid approach to evaluate the passing performance of individual players. This model might be used by coaches to scout, select, and evaluate players. However, our model is currently comprised of multiple variables that again consist of multiple components. Yet, most coaches and performance analysts prefer single composite indicators to evaluate performance. To construct a composite indicator of pass performance (CMP), we took the data of all 86 players with >100 accurate passes also used in section III. We then applied a simple mathematical model based on inter-item correlations and standard deviations first proposed by Horst [3] and shown in equation 4. The composite indicator CMP could then be computed from the raw data using the following formula (Eq. 1).

(1) CMP = -0.009 [Accuracy] + 0.345 [I-Mov] + 0.252 [I-MovX]

As a final step in our analysis, we applied the composite performance indicator (CMP) to our dataset of players. We randomly split the data of 86 players in a training set of 70% of the data (n = 60 players), and a test set of 30% (n = 26 players). In the next step, we fitted a Ridge Regression model with generalized cross validation to the training set and evaluated its' performance on the test set.

Our regression model was able to explain 27.5% of the variance in CMP based solely on the average pass location, length, angle, and velocity. The regression equation indicated that especially X location of the pass and velocity had a positive relationship with pass performance while a more forward directed angle seemed to have a slightly negative impact on pass performance.

V. Summary and Future Work

To sum up, this paper provides empirical evidence that tactical performance in soccer can be evaluated using tracking data without relying on human observations. This finding is important to the field as it demonstrates the merits of smart data scouting. This approach is the first one to be applicable in real-time game analysis and therefore, provide coaches with more reliable data on their players. This would be helpful for in-game coaching. In addition, this approach could serve as a new way to evaluate and scout players and, therefore, provide clubs with substantial advantages on the transfer market. From a data science perspective, this study demonstrates the opportunity that data science methods provide to investigate complex human behavior.

References

[1] J. Gudmundsson and M. Horton, "Spatio-Temporal Analysis of Team Sports -- A Survey," 2016.

[2] R. Rein, D. Raabe, and D. Memmert, "'Which pass is better?' Novel approaches to assess passing effectiveness in elite soccer," *Hum. Mov. Sci.*, vol. 55, pp. 172–181, Oct. 2017.

[3] P. Horst, "Obtaining a Composite Measre from a Number of differnt Measures of the same Attribute," *Psychometrika*, vol. 1, no. 1, pp. 53–60, 1932.