Football Pass Prediction using Player Locations

Philippe Fournier-Viger[1], Tianbiao Liu[2], and Jerry Chun-Wei Lin[3]

[1] School of Natural Sciences and Humanities,
Harbin Institute of Technology (Shenzhen), Shenzhen, China
[2] College of Sports and Phys. Education, Beijing Normal University, Beijing, China
[3] Department of Computing, Mathematics and Physics,
Western Norway University of Applied Sciences (HVL), Bergen, Norway
philfv8@yahoo.com, ltb@bnu.edu.cn, jerrylin@ieee.org

Abstract. In many sports, predicting the passing behavior of players is desirable at it provides insights that can help to understand and improve player performance. In this paper, we describe a novel model for football pass prediction, developed to participate in the Prediction Challenge of the 5th Workshop on Machine Learning and Data Mining for Sports Analytics, collocated with ECML PAKDD 2018. The model called Football Pass Predictor (FPP) considers various aspects to generate predictions such as the distance between players, the proximity of players from the opposite team, and the direction of each pass. Experimental results shows that the model can achieve a prediction accuracy of 33.8%, and more than 50% if two guesses are allowed. This is considerably more than the random predictor, which obtains 8.3%.

Keywords: football, pass prediction, prediction challenge

1 Introduction

Passing is a key aspect of the football game. It is performed between players, occurs almost everywhere on the pitch, and creates scoring opportunities. According to statistics from the Union of European Football Association, during the 2017-2018 season, a top European team could perform more than 400 passes in a Champions league match [**?**]. Passing can account for the majority of tactical/technical behaviors in a game. Many researchers have studied passing behavior from the perspective of football tactics and techniques [**?**,**?**]. Nowadays, with the development of data science and computer technology, researchers can analyze and simulate passing using massive databases and complicated models, to acquire a deeper understanding of passing behaviors. For instance, Liu et al. [**?**,**?**] evaluated the effect of passing on creating scoring opportunities using a Markov chain model. In another study, an Apriori-based algorithm was applied to perform a descriptive analysis of passing behavior [**?**,**?**] Apriori-based diagnostical analysis of football passes was also done using a sequential rule discovery approach [**?**] based on the RuleGrowth algorithm [**?**].

Passing behavior is influenced by many factors and a player must quickly react to the sudden changes of situations around him. Considering the playing

context and different playing situations, Stöckl et al. [**?**] provided an approach to describe the tactical difficulty of passes. In another study, Rein et al. [**?**] assessed passing effectiveness in elite soccer using two algorithms considering the number of defenders and players' control of space. Gyarmati et al. [**?**] developed a QPass evaluation system to estimate players role in building up an attack. Another way to identify key players in a team is using network analysis and to consider pass difficulty [**?**]. Generally, passing differs according to the context. Sometimes, passings yield dangerous situations with respect to opponents, while sometimes few risks are involved. Cakmak et al. [**?**] developed a descriptive model to quantify the effectiveness of passes and identify key passes and regular passers in a team. Using Player Trajectory technology, Lida and Mase [**?**] studied the ball passing behavior of players by considering their trajectories. Dhar and Singh [**?**] analyzed video footage to develop a passing strategy. Although, all these studies have been done to analyze football passing behavior, they do not provide a computational model that can be used for pass prediction. But pass prediction could have many applications.

The contribution of this paper is to fill this void by proposing a model named FPP (Football Pass Predictor) to predict the player who will receive a pass initiated by another player. This work is done in the context of the Prediction Challenge of the 5th Workshop on Machine Learning and Data Mining for Sports Analytics, collocated with ECML PAKDD 2018. The proposed model considers various aspects to generate predictions such as the distance between players, the proximity of players from the opposite team and the direction of each pass. Experimental results shows that the model can achieve high prediction accuracy.

The rest of this paper is organized as follows. Section 2 provides a brief description of the provided dataset, and key observations that were made. Section 3 presents the proposed FPP model. Section 4 presents experimental results. Finally, Section 5 draws a conclusion.

2 Observations about the Data

The dataset provided for the prediction challenge contains 12,124 records describing passes from 14 football matches of a Belgian team and opposing teams during the 2014/2015 football season. In a football match, two teams are facing each others, where each team has 14 players (including 3 substitutes). A database record describes a pass. It provides the (1) the location of the 14 football players of each team using 2D coordinates, (2) the time at which the pass started and ended, (3) the player who sent the ball, and (4) the player who received the ball. Coordinates are expressed in the [-5250, 5250] [-3400, 3400] intervals for the X and Y axes, respectively. Note that player names are not indicated in the data as well as the names of the teams. Moreover, it has not been indicated if the positions of the players have been recorded when a pass starts or ends. Besides, although timestamps are provided in hte data, records from all matches were put in a single file and randomly shuffled. Thus, each pass can only be considered individually rather than in the context of a match. The

data was collected by the prediction challenge organizers, and made available at `https://github.com/JanVanHaaren/mlsa18-pass-prediction`

By analyzing the data, the authors of this paper made a few interesting observations. First, out of the 12,124 passes, only 17% of the passes are intercepted by the opposite team. Because unsuccessful passes are much less likely than successful ones, a design principle for the proposed model is to assume that all passes will be successful when making predictions. Second, if was found that the 163th line of the dataset is an invalid record. In that record, the player number 15 who sends the ball has no coordinates. This record has been ignored. Third, although the dataset provides timestamps, it is difficult to use this information for pass prediction since each record is often separated by numerous seconds, and the position of players is given only once for each pass but each record contains two timestamps. For this reason, the trajectories of players are not available, and it is hence difficult to analyze each pass in the context of the overall game. Fourth, a related issue is that records from all matches are stored together in the dataset. Thus, it is unclear which passes belong to which match. Besides, it is not indicated which team is playing on which side of the football field. We have inferred this information by assuming that the left (right) side of the pitch belongs to the team having the leftmost (rightmost) player.

3 The FPP Model

Based on the observations made on the data, the proposed FPP model was developed. To design the model an iterative design approach was used where several versions of the model were successively designed, each adding additional criteria to increase prediction accuracy. Among the multiples versions of the model, four are described in the paper. These versions, sorted by ascending order of complexity, are called M1, M2, M3, and FPP, respectively. They are described in the following paragraphs, and illustrated in Fig. **??**.

M1. The first model is based on the assumption that the sender will pass the ball to the closest player of his team. Assume that a player X has the ball and that we want to predict who will receive the ball. Let P be the set of players from the same team as X (excluding X). For each player $Y \in P$, the Euclidian distance between X and Y is calculated, denoted as $d_{X,Y}$. Then, for each player $Y \in P$ a score is assigned to Y, defined as $score(Y) = d(X, Y)$. The player with the smallest score is chosen as the prediction.

M2. The second model is an improvement of the M1 model. An additional idea is considered, which is that a player may be less likely to receive the ball if a player of the opposite team is close to him. The motivation is that this situation may be considered more risky for the sender, and that the opposite team player may intercept the ball. Formally, let O be the set of players from the opposite team. For each player $Y \in P$, its score is defined as $score(Y) = d(X, Y) + penaltyC(Y, O)$. The term $penaltyC$ is defined as $penaltyC(Y, O) = 900$ if there exists a player $Z \in O$ such that $d(Y, Z) < 700$, and otherwise $penaltyC(Y, O) =$

0. The values 700 and 900 were found empirically (by trial and error) to obtain a high prediction accuracy.

M3. The third model is an improvement of M2, which considers that more than one player from the opposite team may be close to a potential receiver and increase risks. For each player $Y \in P$, its score is defined as $score(Y) = d(X,Y) + penaltyC(Y,O) + penaltyD(Y,O)$. The term $penaltyD$ is defined as $penaltyD(Y,O) = 55$ if there exists two players $Z \in O$ such that $d(Y,Z) < 700$, and otherwise $penaltyD(Y,O) = 0$. The value 55 was found empirically.

FPP. The fourth model is an improvement of the M3 model, which considers the direction of the ball, based on the assumption that a player prefers to send the ball forward. Let the notation $Z.x$ denotes the position of a player Z on the x axis. The score of a player $Y \in P$ is defined as $score(Y) = d(X,Y) + penaltyC(Y,O) + penaltyD(Y,O) + direction(X,Y)$. The term $direction(X,Y)$ is defined as $-0.3 \times |X.x - Y.x|$ if the pass is a forward pass (toward the opposite team goal) or as $0.1 \times |X.x - Y.x|$ if the pass is a backward pass. The values 0.1 and 0.3 were found empirically as providing the best results.
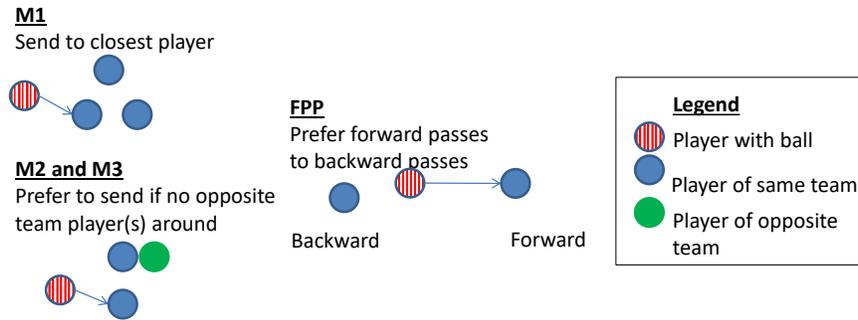

Fig. 1: An illustration of the ideas introduced in the proposed models

4   Experimental Evaluation

An experimental evaluation was performed to evaluate the five versions of the designed FPP model. The models were compared with a random predictor as baseline. Since no performance measure was explicitly specified for the prediction challenge, it was decided to evaluate the models in terms of accuracy (number of correct predictions divided by total number of passes to be predicted). Furthermore, the accuracy when two guesses are allowed was measured. In that situation, a model can make two predictions for each pass, and if one of them is right, the pass is considered as correctly predicted. The source code of the proposed model and evaluation framework can be downloaded from `http://philippe-fournier-viger.com/foot2018/`. It is written in Java.

Table 1: Accuracy of the compared models

Model	Accuracy for one guess (%)	Accuracy for two guesses (%)
Baseline (random)	7.88	16.51
M1	27.44	46.58
M2	32.83	50.92
M3	32.97	50.94
FPP	**33.38**	**51.81**

Results are shown in Table **??**. It is first observed that the heuristic of predicting that the ball is passed to the closest player of the same team achieves a high accuracy (27.44%) compared to the baseline random predictor (7.88%). Then, if we add a penalty if there is a player from the opposite team that is close to a potential receiver, it increases accuracy by more than 5%, from 27.44% (M1) to 32.83% (M2). If this model is further extended for the case of two players from the opposite team, the accuracy increases slightly, from 32.83% (M2) to 32.97% (M3). Moreover, if the direction of the ball is considered, the accuracy increases from 32.97% (M4) to 33.38% (FPP). Finally, if we allows to perform two guesses, the accuracy of the best model (FPP) increases to 51.81%.

Besides, the described models, the authors of this paper have also tried various other ideas including calculating the angle between players to determine if a player from the opposite team may intercept a pass. But these ideas did not improve accurracy, or even decreased it. Other models could also be considered. However, what can be developed remains limited by the data. For example, having more rich data such as the real time locations of players, and data were records are not shuffled, could allow to obtain player trajectories and develop more complex models. Besides, an improvement of the FPP model could be to use a genetic algorithm to tune its parameters instead of tuning them by hand, and to split the data in training and testing sets, or using k-fold cross validations to avoid the potential problem of overfitting.

5   Conclusion

This paper has proposed a model called Football Pass Predictor (FPP) to predict the receivers of passes in football matches. The model considers various aspects such as the distance between players, the proximity of players from the opposite team, the direction of each pass, to generate predictions. The performance of the model was compared with a baseline random predictor and several variations of the proposed models. Results from experiments shows that FPP can achieve a prediction accuracy of 33.8%, and more than 50% if two guesses are allowed. An interesting perspective for future work is to collect richer data, which would allow to develop more complex models. We also plan to evaluate the possibility of using pattern mining approaches for football pass prediction [**?,?**].

References

1. http://www.uefa.com/uefachampionsleague/season=2018/statistics/round=20008 81/matches/kind=passes/index.html, Last accessed on the 15th June (2018)
2. Garratt, K., Murphy, A., Bower, R.: Passing and goal scoring characteristics in Australian A-League football. International Journal of Performance Analysis in Sport, 17(1-2), pp. 77–85 (2017)
3. Plummer, B.T.: Analysis of attacking possessions leading to a goal attempt, and goal scoring patterns within men's elite soccer. Journal of Sports Science and Medicine, 1(1), pp. 1–38 (2013)
4. Barreira J., Vendite C., Vendite L. L.: Analysis of Shots and Passing Sequence of a Soccer Team and Its Opponents during 2014 Brazilian Championship. International Journal of Sports Science, 6(4): 163–167 (2016)
5. Liu, T.: Systematische Spielbeobachtung im internationalen Leistungsfußball. Ph.D. disseration, University of Bayreuth (2014)
6. Liu, T., Hohmann, A.: Apriori-based diagnostical analysis of passings in the football game. In: Proc. IEEE 2016 Int. Conf. Big Data Analysis, pp. 1–4, IEEE (2016)
7. Liu, T., Fournier-Viger, P., Hohmann, A.: Using Diagnostic Analysis to Discover Offensive Patterns in a Football Game. In: Proc. Int. Conf. Data Science and Business Analytics, pp. 381–386, Springer (2018)
8. Fournier-Viger, P., Nkambou, R., Tseng, S. M.: RuleGrowth: mining sequential rules common to several sequences by pattern-growth. In: Proc. 26th Symposium on Applied Computing, pp. 954–959, ACM Press (2011)
9. Fournier-Viger, P., Lin, J. C.-W., Vo, B, Chi, T.T., Zhang, J., Le, H. B.: A Survey of Itemset Mining. WIREs Data Mining and Knowledge Discovery, Wiley, e1207 doi: 10.1002/widm.1207 (2017)
10. Fournier-Viger, P., Lin, J. C.-W., Kiran, R. U., Koh, Y. S., Thomas, R.: A Survey of Sequential Pattern Mining. Data Science and Pattern Recognition (DSPR), 1(1), pp. 54-77 (2017)
11. Stöckl, M., Cruz, D., Duarte, R.: Modelling the Tactical Difficulty of Passes in Soccer. In: Proc. 10th Int. Symp. Comp. Sci. Sports, pp. 139–143, Springer (2016)
12. Rein, R., Raabe, D., Memmert, D.: "which pass is better?" novel approaches to assess passing effectiveness in elite soccer. Human Movement Science, 55, 172–181 (2017)
13. Gyarmati L, Stanojevic R.: QPass: a Merit-based Evaluation of Soccer Passes. Preprint on arXiv:1608.03532 (2016)
14. McHale, I. G., Relton, S. D.: Identifying key players in soccer teams using network analysis and pass difficulty. European Journal of Operational Research. 268(1), 339–347 (2018)
15. Cakmak, A., Uzun, A., Delibas, E.: Computational Modeling of Pass Effectiveness in Soccer. Journal of Advances in Complex Systems, in press (2018)
16. Lida, R., Mase, K.: Ball passing course creating behavior in soccer game detection from player trajectory. IEICE Technical Report, 113(432), 171–176 (2014) vol. 113, no. 432
17. Dhar, J., Singh, A.: Game analysis and prediction of ball positions in a football match from video footages. In: Proc. Int. Conf. Recent Advances and Innovations in Engineering, pp. 1–6, IEEE (2014)