Advances in Complex Systems
Vol. 21, Nos. 3 & 4 (2018) 1850010 (28 pages)

DOI: 10.1142/S0219525918500108

COMPUTATIONAL MODELING OF PASS EFFECTIVENESS IN SOCCER

ALI CAKMAK*, ALI UZUN† and EMRULLAH DELIBAS‡

Department of Computer Science,
Istanbul Sehir University, Istanbul, Turkey
**alicakmak@sehir.edu.tr*
†aliuzun@std.sehir.edu.tr
‡emrullahdelibas@std.sehir.edu.tr

Received 16 May 2017
Revised 8 February 2018
Accepted 20 June 2018
Published 17 July 2018

The emerging data explosion in sports field has created new opportunities to practice data science and analytics for deeper and larger scale analysis of games. With collaborating and competing 22 players on the field, soccer is often considered as a complex system. More specifically, each game is usually modeled as a network with players as nodes and passes between them as the edges. The number of passes usually define the weight of each edge, and these weights are employed to identify the key players using network modeling theory. However, the number of passes metric considers each pass the same and cannot differentiate players who are making ordinary passes, usually in their own pitch to a close teammate, from those who make key passes that start or improve an attack. As a solution, in this paper, we present a descriptive model to quantify the effectiveness of passes in soccer to differentiate between key passes and regular passes with not much contribution to the play of a team. Our model captures the perception of domain experts with a careful combination of risk and gain assessments. We have implemented our model in a soccer data analytics software. We performed a user study with domain experts, and the results show that our model captures domain expert evaluations of a number of example scenarios with 94% accuracy. The proposed model is not computationally demanding which allows real-time pass assessment during games on commodity hardware as demonstrated by our software prototype.

Keywords: Data analytics; information modeling; genetic algorithms; sport analytics; complex systems.

1. Introduction

With the recent advancements in sensor and camera technologies, it is now possible to automatically track the movement of the ball and players on the field with high precision [3]. These tracking technologies generate tremendous amounts of raw data, and the computational analysis of such data may lead to new insights about teams and individual players [8, 9, 28, 14, 29]. The derived insights may later be used to design game tactics, develop personalized training programs, discover young talents, analytically compare teams/players, etc. [30].

The game of soccer may be considered as a complex system where each player interacts with other players leading to an emerging behavior for each team [24, 25, 18, 19]. In this regard, complex systems theory is often employed to reveal the governing dynamics of the game [20]. As an example, Stöckl *et al.* [22] employ recurrence plots to visually characterize different phases of soccer games based on the distinct play patterns over the time. Moreover, Gréhaigne *et al.* [23] introduce the notion of effective play-space and study different shapes of effective play-space to identify a comprehensive set of configurations that a game may evolve into. Besides, Vilar *et al.* [26] study the distributions of home and away teams' players at different regions of the field and analyze the emerging patterns of dominance during defense and offense at key areas of the pitch. As a predictive approach, Pappalardo and Cintia [27] employ the numbers of different events, e.g., pass, goalkeeping, interception, etc., as features to build classification and regression models which predict the outcome of a game. They simulate a complete season and show that the simulated rankings are highly correlated with the actual rankings.

A popular trend in the complex systems view of soccer is the network modeling approach [31–33], as network models facilitate understanding the structural configurations of a soccer game as a system [35]. In such approaches, individual players represent the nodes and various game interactions between them represent the edges. In particular, passes are commonly used to model the edges between nodes in such networks [36, 37], as passes are among the most essential game elements that often determine the fate of a game. As an example, Cintia *et al.* [40] report high correlation between passing activity and team success. Similarly, Redwood-Brown [34] shows that teams have significantly higher pass accuracy within 5 min before a scoring takes place than the game average. Moreover, players with decent pass skills are usually rated much higher than others [43]. One analysis aspect of such network models is to identify the key players that contribute most to their teams [33]. Among the employed measures, the most common ones are the variations of centrality measures [39]. One gap in such network modeling approaches is that they consider all passes between players as the same and weigh an edge between two players as the total number of passes between them. Such an equalitarian view undervalues the key players who may not necessarily exchange the most number of passes but may make the most key passes that contribute to the play of a team significantly. Those key players may be considered to represent key actors controlling the gates of various flows in the complex system view of a team [35]. Hence, it is important to identify them effectively to understand the underlying structure of the system, i.e., a soccer team.

In this paper, in order to fill this gap, we propose a computational framework that quantifies pass effectiveness in a robust and scalable manner. To this end, we craft the desired properties of an effective pass and, accordingly, create a computational model. More specifically, our model considers (i) the *risk* of intervention by opponent team players, (ii) the *gain* in terms of the number of eliminated opponent team players from the attacking region, (iii) the goal chance of the player who receives the pass, (iv) the availability of well-positioned teammates to whom the pass-receiving player can pass the ball in the next step, and (v) the amount of time that the pass-receiving player has to make decision on how to use the ball before s/he (i.e., she or he) is pressured by the nearest opponent players. In addition, our model is recursive in its nature in the sense that it evaluates a pass not by itself individually but as part of a sequence of passes that it belongs to. That is, an effective pass is assumed to lead to another effective pass (or a goal shot).

Coaches and players may benefit from such a model (as implemented in a software tool) in many different ways. As an example, coaches may design their own pass scenarios (maybe by modifying passes from past games) and show players how the effectiveness of a pass is changing when some players take different positions in defense or offense [4]. Players may also be given feedback during exercises about their pass choices and missed opportunities (if any).

In order to evaluate the effectiveness and the accuracy of our model, we perform a user study with domain experts (i.e., professional coaches and players) on a set of many possible passes. Experimental results show that the agreement between the evaluation of pass effectiveness by the proposed model and those by the domain experts on a number of example scenarios is quite high (over 94% accuracy).

Finally, we have incorporated the proposed model into an interactive software tool [4] with a graphical interface. This tool allows to (i) replay the game visually at different speeds with real-time effectiveness scores reported for each pass, (ii) perform "what-if" scenarios with support for dragging and dropping player objects and defining artificial passes between them, (iii) create heatmap-style visualizations which are overlaid on the field and show risk and gain map of a particular game configuration, (iv) save any snapshot of the game on disk in a file and load it later (possibly on a different computer), (v) perform dominant region analysis with real-time visualization support, and so on.

Contributions: Our contributions in this work are as follows:

- Definition of the desired properties of an effective pass,
- A computational model and associated algorithms to quantify pass effectiveness,
- Incorporation of a genetic algorithm to automatically learn model parameters from expert knowledge,
- Preparation and conduction of a survey study with domain experts,
- Implementation of a fully functional software tool with graphical user interface that packs the proposed models along with other useful features.

Organization: This paper is organized as follows. In the next section, we discuss the related work. In Sec. 3, we present our pass effectiveness model. Section 4 provides an overview of the software implementation. In Sec. 5, we present the experimental evaluation the proposed model. Section 6 concludes with a discussion on possible impact of presented results, the limitations of the current study, and pointers for future work.

2. Related Work

Even though soccer has a long history, soccer data analytics using computational means has recently gained traction in the last decade. Gudmundsson and Horton [38] provide an extensive survey on the recent research works in the field with pointers to open research questions. In this section, we particularly focus on algorithmic studies that aim to assess the value of a pass in soccer.

With the increasing popularity of machine learning, an intuitive direction is building a supervised model that automatically predicts the quality of a pass. Horton *et al.* [8] employ a classifier which assesses a given pass by assigning one of the three labels: "good", "OK", or "bad". The model uses the spatial features of the game at the time of the pass and achieves an accuracy of 85.8%. While the reported results are promising, this approach does not allow to differentiate between "good" passes and provide an assessment of how good a pass is. In comparison, our approach does not only rank passes, it also provides a quantitative assessment of each pass individually, independent of other compared passes. That is, we can now differentiate between two "good" passes, and even without a comparison, based on pass effectiveness value, one may have an idea about how good a pass is.

Similarly, Beetz *et al.* [2] employ statistical learning models for the analysis of passes as part of their ASPOGAMO project. As one difference from Horton *et al.*'s approach, ASPOGAMO approach is an interactive one that combines both supervised and unsupervised machine learning models. More specifically, users may manually mark different classes of passes, and then the system automatically creates the descriptions of these classes by turning the trained decision trees into rules in the form of conjunctions and disjunctions. Then, these rules are presented as descriptions of the corresponding pass groups. Moreover, the proposed system also features unsupervised learning capabilities in that it can automatically categorize a given set of passes of a team from a game. The employed features are mostly positional ones, e.g., starting and ending points of passes. Pass initiating and receiving players' qualities may also be used as features. All in all, the proposed tool may be very helpful in exploratory post-game analysis to see the pass characteristics of different teams. Nevertheless, pass descriptions are totally structural referring to their length (e.g., long or short passes), the region of the pitch that they are initiated from and/or targeted to, etc. Thus, it does not provide any assessment regarding whether a pass was a good one or not.

Another example of using passes to differentiate between game styles of different teams or players may be seen in Gyarmati *et al.*'s work [7]. In particular, the authors construct the pass network of a game and analyze pass sequence patterns, called "flow motifs". Such motifs may be useful to compare different teams or reveal their passing characteristics. On the other hand, the proposed framework does not provide any way to evaluate the value of individual (or even a sequence of) pass(es).

As an alternative to above *quantitative* learning models, Vercruyssen *et al.* [41] follow a slightly different route and propose a *qualitative* learning model to predict the receiver and likelihood of a pass. The authors employ a number of qualitative features in the form of binary relationships, such as player A is located to the north of player B during a successful pass between players A and C and train an inductive learning model to learn a set of rules that describe successful passes. As training data, they employ all successful passes in a number of matches. One drawback of this approach is the assumption that, in all past passes, the players made the best decision as to whom to pass the ball, which may not be the case in many passes. Moreover, the likelihoods provided by the model are relative/local and not comparable between passes. That is, having a higher likelihood for a pass between players A and B than that of a pass between C and D does not necessarily show that first is a better pass than the latter. This limits the applicability of the model in practice to a large extent.

While the above studies focus on passes themselves, Maheswaran *et al.* [15] are interested in the after effects of missing a particular pass, i.e., a shot, in basketball. The authors employ a binary classifier model to predict what team will get the rebound if a particular shot is missed. Among its features does the model use player locations and ball height. Although it is interesting, it does not directly apply to soccer, as only an ignorably small percentage of shots hits the goal post and comes back to the pitch in any game.

From the basketball domain, the expected point value (EPV) approach of Cervone *et al.* [16] is particularly interesting. The authors propose to compute EPV for each movement in basketball games. To some extent, it may be applicable to soccer data as well, but obviously, more research is needed. The immediate concerns (hence, research questions) on the applicability of EPV approach on soccer data would be as follows:

(i) EPV approach employ's a kind of Markov Model with states and transitions between these states. The field is divided into discrete regions. The number of states is proportional to [number of players] $\times$ [number of field regions] $\times$ [whether the player is being defended]. Given the number of players and the size of the soccer field, the number of states would be significantly larger. Training such a model would require an enormous amount of training data to have a sufficiently expressive model for each player. It may be challenging to obtain such a data given the size of probable path space.

(ii) Basketball is a game where a large number of points are scored in relatively smaller sequence of movements. Hence, connecting each movement to a score/point value is more intuitive. On the other hand, in soccer, a lot of games end with no goals scored, and in many others, one or two goals would be scored, despite hundreds of passes made during the entire game. Hence, it may be challenging to compute expected point values for soccer.

(iii) As related to item (i), the computational complexity of coming up with the expected point values for passes in soccer may be too high, and real-time pass assessment may not be possible. In our current setting, real-time evaluation is critical, as one of the main motivations of our proposed model is to give instant feedback to players and coaches during training sessions.

(iv) The whole focus of EPV is how each movement will contribute to a team's goal of scoring. Inherent to this approach is the assumption that the movement is successful. On the other hand, our assessment of a pass directly considers the likelihood that whether such a pass can be made successfully. The goal here is to value high gain low risk movements over high risk high gain movements to guide players and coaches on possible pass alternatives. Nevertheless, the EPV approach ignores this aspect.

A much simpler version of the EPV approach is studied by Link *et al.* [42] within soccer domain. In particular, the authors propose a custom model that quantifies the likelihood of the current ball owner to pose a serious threat (i.e., scoring a goal) for the opponent team. Even though the proposed measure, "dangerousity", is somewhat relevant to the concept of *goal chance* that we use in our model, it is designed for particularly attacking scenarios and becomes relevant only when the player is in the final third of the pitch. On the other hand, our pass effectiveness model does not assume particular game setting and provides evaluations for all passes in each point of the field.

The above study is complemented by Lucey *et al.*'s work [29] in that the authors focus on the quality assessment of shots on goal. In particular, the authors employ spatiotemporal features, such as location, game context (e.g., open play, counter attack, etc.) to predict the expected probability that the shot would result in scoring a goal. The authors consider only the last 10 s before a shot. As different from Lucey *et al.*'s work, our work focuses on the more general problem of assessing the quality of passes. Moreover, we consider the entire game period, rather than a small fraction (i.e., 10 s) before a particular event, e.g., a shot as in the above study.

Szczepański and McHale [43] argue that the number of completed passes metric fails to assess the players' passing ability, as it ignores many intrinsic details, such as the difficulty of passing against strong teams, or in regions close to the goal. As an alternative, the authors propose a statistical model to predict the probability that a pass would complete successfully and assess the players based on the completion probability of their passes. Although their results show that assessment of players by the probability measure is better than by the count of successful passes, the proposed model does not provide any measure regarding the quality of a pass.

Similar to the above approaches, as input, we use ball and player position data that are generated by sensors and cameras installed on the field. However, what makes our work novel is its output, i.e., our model provides a quantitative evaluation of each pass individually as well as with respect to the pass network of a team. In addition, instead of predictive models adopted by most of the above related work, we employ more of a descriptive approach due to the lack of labeled pass evaluation data that includes a large number of passes in different settings with their true effectiveness values as determined by experts. Presently, such data is not available, and is costly to generate, which limits the applicability of predictive approaches.

More recently, somewhat more directly related to our approach, the following two studies are presented. The first work [44] is a summarization-based approach that (i) looks at the history of matches in the past, (ii) scores zones based on the number of successful passes and/or goal shots made on them, and (iii) then, on this scored zone map, assesses a pass that transfers the ball from a lower value zone to higher value zone as a more valuable pass than a pass that transfers the ball between two nearly equally scored zones. The authors ignore the current game configuration such as defending players' positions and base their assessment mostly on the summarized values of pass starting and ending zones. In comparison, our model captures the current game context in detail and accordingly decides on the value of a pass in a dynamic way. The second work [45] is more involved and enriched with game context to some extent. Similar to our proposal, the authors assign risk and reward scores to each pass. However, the employed measures are over-focused on immediate goal scoring potential (represents pass reward) and successfully passing the ball to another teammate (represents pass risk). Such an approach is greedy, i.e., it focuses on immediate benefit and undervalues "effective" passes that start or contribute to building up an attack and are far from the goal region. In this study, we fill these gaps by considering the likelihood that the current pass may lead to other *effective* passes in a recursive manner with the inclusion of *pass advantage* component.

3. Pass Effectiveness Model

In this section, we present basic elements that constitute our pass effectiveness model. Many of the notions that motivate the components of our model build on the factors influencing pass success as discussed by Szczepański and McHale [43].

We first define the risk assessment of a pass which is based on the likelihood of an intervention by opponent players. Our employed motion model in the below risk area discussion is partly inspired by the *dominant region* concept proposed by Taki and Hasegawa [14, 17].

Definition (Pass Risk with respect to an Opponent Player): *Pass risk* represents the probability that a pass between two players will be intervened by an opponent player. More specifically, we define the *risk of a pass between teammates P_1 and P_3 with respect to an opponent player P_2* as follows:

$$\mathrm{Risk}(P_2, \mathrm{pass}(P_1, P_3)) = \mathrm{intervention\_prob}(P_2,\ \mathrm{pass}(P_1, P_3)),$$

where intervention_prob is the probability that a player P_2 can intervene a pass between players P_1 and P_3 (Fig. 1). The value of intervention_prob is learnt from the past game data. More specifically, let $V_{\min}$ be the minimum speed that player P_2 should run to intervene the pass from player P_1 to P_3. Then, intervention_prob is the

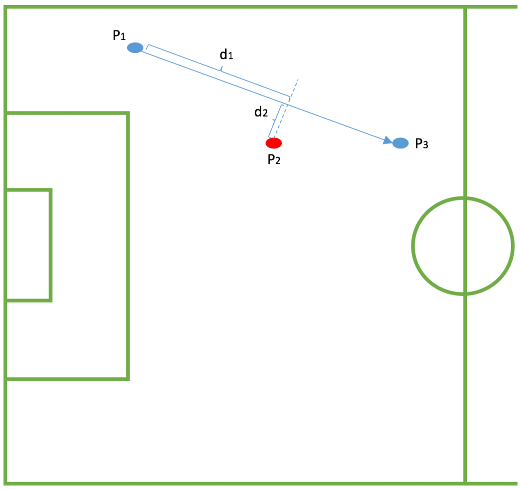

Fig. 1. Pass risk evaluation with respect to an opponent player — A pass between players P_1 and P_3 where a player P_2 can intervene. d_2 is the distance of player P_2 to the path between P_1 and P_3. d_1 is the distance of P_1 to the point on the pass path that P_2 is closest to.

ratio of player runs with speed V_{min} or higher among all player runs in past game data. V_{min} may be computed as d_2/t_{pass} where d_2 is the distance of the player P_2 to the pass line as shown in Fig. 1, and t_{pass} is the time that the ball takes to travel distance d_1 (Fig. 1). That is, $t_{\text{pass}} = d_1/V_{\text{ball}}$ where V_{ball} may be available as part of game data. Otherwise, an average ball speed in a pass is assumed (i.e., reported as 23.28 m/s in Asai *et al.*'s study [1]).

Definition (Risk Area): The risk area of a certain pass is a region in which any opponent team player might intervene the pass.

The size of the risk area depends on the pass distance. The risk area consists of three components, namely, (i) the region around the player who initiates the pass, (ii) the region around the player who receives the pass, and (iii) the area between these two regions (see Fig. 2 for an example). The regions around players are modeled as circles with the passing players at the center, respectively. The radius of these circles are learnt from training data via a genetic algorithm [6, 13] as described in the next section.

Definition (Overall Pass Risk): The *overall risk of a pass* is defined as the cumulative risk of the pass with respect to all the opponent players that are located in the risk region of a pass. More formally, we define the *overall risk of a pass between teammates* P_1 *and* P_2 as follows:

$$\text{Overall risk}(\text{pass}(P_1, P_2)) = \sum_{P_i} \text{Risk}(P_i, \text{pass}(P_1, P_2)),$$

where P_i is defined over the set of all opponent players that are located in the risk region of the pass. As an example, for the pass in Fig. 2, players 8 and 26 from the

Fig. 2. (Color online) Risk region with respect to a pass — A pass is defined between player 2 and 16. The red lines indicate the borders of the risk area in which an opponent player may pose a threat to the pass.

opponent team are the two players who are included in the risk region of the pass; hence, they contribute to the overall risk of the pass.

Definition (Shooting Line): Shooting line conceptually represents a set of points that are as far away from the goal as possible to the extent that a ball-possessing player would still consider that shooting from one of these points will result in a goal with high likelihood.

In this work, we represent the shooting line as a semi-circle which is centered at the midpoint of the goal line. Figure 2 shows an example shooting line with dashes around the goal on the left-hand side. We choose the circle shape as it allows placing the goal region in the center with equal distance from all sides. The radius of the semi-circle is learnt from training data through a genetic optimization algorithm [13] as described in the next section.

Definition (Threat-Posing Player): Let P_1 be the ball-possessing player at a time point t_i and s be the closest point to P_1 located on the shooting line. An opponent player P_2 is called a threat-posing player for P_1, if P_2 can reach s no later than P_1. We assume that both P_1 and P_2 start running at the same time with their maximum speeds (we consider maximum speed of 8.97 m/s as reported in Rampinini *et al.*'s work [11], if player specific speed data is not available).

As an example, in Fig. 2, for player 3 from the red team, player 22 from the opponent blue team is the only threat-posing player.

Definition (Pass Gain): Let $T(P)$ denote the number of threat-posing players for a player P. Given a pass from player P_1 to player P_2, *pass gain* is defined as follows:

$$\text{Gain}(\text{pass}(P_1, P_2)) = T(P_1) - T(P_2).$$

Pass gain corresponds to the measure of how offensive or neutral the pass is as discussed by Reiner [46] who in turn adapts the concept from Gréhaigne *et al.*'s work [10]. Moreover, pass gain is also similar to "packing-rate" of Impect (a soccer analytics startup based in Germany) [21]. While packing-rate simply counts the decrease in the number of opponent players between the ball and the opponent goalkeeper, pass gain measure differentiates between *threat-posing opponent players* and those are who have no chance of tackling or intervention regarding the current pass. This way, passive opponent players who are being bypassed by a pass but are located at far distances (e.g., on the opposite side) are automatically eliminated to prevent artificially high pass gains.

Definition (Pass Advantage): Pass advantage is an attribute of a player position in relation to his teammates. It quantifies the appropriateness of a player's position at a time point t for making an effective pass to another teammate. We consider that when a player receives a pass, regarding to whom s/he would pass the ball, s/he would choose the teammate for whom the gain/risk ratio of the possible pass would be the maximum. More specifically, the *pass advantage of a player P* is defined as follows:

$$\text{Pass advantage}(P) = \text{argmax}_{P_i \in \text{TeamMates}(P)} \left\{ \frac{10 + \text{Gain}(\text{pass}(P, P_i))}{10 + \text{Overall risk}(\text{pass}(P, P_i))} \right\},$$

where 10 is an additive term that makes sure that pass advantage is always a positive value so that maximization operator would produce accurate results even for back pass scenarios. In order to prevent this additive term in the numerator from distorting Gain/Risk ratio considerably, we add the same amount into the denominator as well.

Based on the fact that it combines often conflicting gain and risk, pass advantage may be considered as similar to reward/risk combination in Power *et al.*'s work [45]. One difference of our work is that Power *et al.*'s reward concept is only relevant if there is a chance of goal shot within 10 s, which usually happens in goal area. On the other hand, in this work, we consider a more general "reward" that stays relevant even in a team's own half of the pitch.

Definition (Goal Chance): Goal chance represents the probability that a player P will score a goal if s/he chooses to make a shot to the goal from his current position. More specifically, we consider that:

- Goal chance increases as the distance of the player to the goal decreases [5, 24].
- Furthermore, goal chance increases as the angle between the player and the goal area corners increases [5]. Here, we assume that penalty point has the highest goal

chance. Hence, we use the angle of the player at the penalty point as the best angle that one could get.

- Finally, to account for defenders between the player and goal [5, 24], we take advantage of our existing risk model in that we consider a shot to the goal as a pass from the shooting player to the opponent goal keeper. Lower the risk of such a pass is higher the goal chance.

Translating the above considerations, we mathematically define the goal chance as follows:

$$\text{Goal chance}(P) = \frac{\text{goal\_width}}{d} * \frac{\min(\alpha, \text{penalty\_angle})}{\text{penalty\_angle}} * \frac{1}{1 + \text{Overall risk}(\text{pass}(P, \text{GoalKeeper}))}$$

where goal_width is the width of the goal area, d is the distance of the shooting player P from the goal, α is the angle between the lines drawn from P's location to the two endpoints of the goal area, penalty_angle is the angle between the lines drawn from the penalty point to the two endpoints of the goal area. penalty_angle is computed as 59 based on standard field size. Angles larger than penalty angle imply that the shooting player is even closer to the goal keeper than the penalty point, and we still consider that angle the same as the penalty angle. GoalKeeper is the opponent team's goal keeper. These parameters are illustrated in Fig. 3.

The goal chance concept employs similar features, such as the number and location of defenders between the player and goal, location of the shot, etc., which are

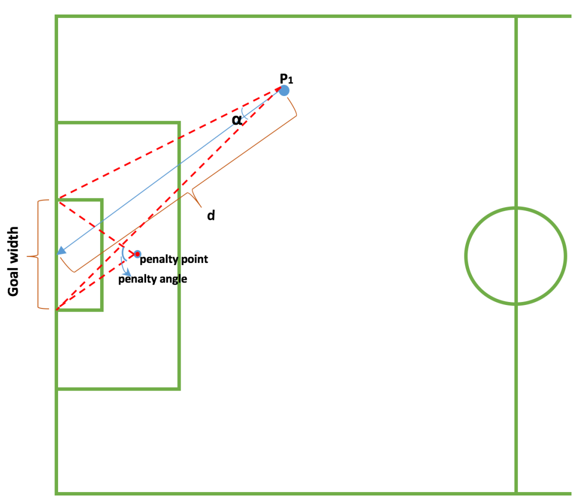

Fig. 3. Goal chance evaluation elements — α is the angle between the lines that connect the shooting player to the goal area poles. d is the distance of the shooting player to the center of the goal area. Penalty angle is the value of α, when the shooting player is located at the penalty point, i.e., the best angle.

also incorporated into the computation of the expected goal value (EGV) proposed by Lucey *et al.* [29]. EGV is a predictive model of a shot success, while goal chance is a descriptive model. Although Lucey *et al.* does not explicitly discuss it, an earlier work by Pollard *et al.* [5] considers the angle of the shot as well, in addition to position and defenders' distance to the player.

Definition (Decision Time): Decision time represents the length of duration that a pass receiving player has to decide how to use the ball before the closest opponent player challenges him/her. We assume that at least the closest player to a pass receiving player P runs towards P to defend against P. The decision time for a pass receiving player P is computed as follows:

$$\text{Decision time}(P) = d/\text{max\_speed}(\text{closest player to } P),$$

where, if not defined in a player specific way, we assume the same maximum speed for all players (8.97 m/s based on Rampinini *et al.*'s study [11]). As an example, in Fig. 2, player 22 has better decision time than player 77, if they were to receive a pass from any blue team player.

Decision time corresponds to the *pressure* concept in Link *et al.*'s work [42], where the authors put more focus on defenders that are between the player and goal, as they also consider the *risk* concept as part of *pressure*. In comparison, we treat all defenders in any direction around the player the same in decision time component, as the closest player, from any direction, would be reaching to the player first to put pressure.

Definition (Successful Pass): A *successful pass* is the one for which the players who initiate and receive the pass belong to the same team.

Definition (Mis-Pass): A *mis-pass* is the one for which the players who initiate and receive the pass belong to different teams.

Definition (Effective Pass): An *effective pass* is the one that eventually leads to a goal scoring opportunity (possibly through a number of additional passes) for the team of the player who initiates the pass. More specifically, we list the following properties that an *effective pass* should have.

- **Properties of an effective pass:** An effective pass:
  - is a successful pass.
  - eliminates several opponent team players off of the active game (high *pass gain*).
  - is targeted to a player with high *pass advantage*.
  - leads to a high *goal chance* situation.
  - is targeted to a player with sufficient *decision time* to make the next move.
  - leads to an *effective* pass (recursive definition), i.e., the next pass in a sequence of passes should be an *effective pass* as well.

Based on the above properties, we define *pass effectiveness score* as follows.

Definition (Pass Effectiveness Score): Given a pass(P_1, P_2) from player P_1 to player P_2, assume that pass(P_1, P_2) is part of a pass sequence S in which it is followed by another pass pass(P_2, P_3) from player P_2 to player P_3. Then, *pass effectiveness score* is defined as follows:,

$$\begin{aligned}\text{Effectiveness}(\text{pass}(P_1, P_2))_{\text{NextPass: pass}(P_2,P_3)} = {} & w_1 \times \text{Gain}(\text{pass}(P_1, P_2)) + \\ & w_2 \times \text{Pass advantage}(P_2) + \\ & w_3 \times \text{Goal chance}(P_2) + \\ & w_4 \times \text{Decision time}(P_2) + \\ & w_5 \times \text{Effectiveness}(pass(P_2, P_3))\end{aligned}$$

where w_i are weights which may be tuned based on insights from domain experts (e.g., team coaches). However, even for experts, coming up with precise weights may be challenging. In this paper, as a more robust solution to this problem, we employ a genetic optimization algorithm [13] to automatically find the best set of weights that match domain experts' evaluation. For cases where a pass is evaluated individually (i.e., not as part of a pass sequence), the last term in the above equation (i.e., the effectiveness of the next pass) is ignored.

The above effectiveness scores are defined for successful passes. As for mis-passes, we score them in the same way but with a negative sign. That is, a mis-pass that would normally be very effective if it were successful pass (e.g., starting a counter attack, for instance) with a computed score s will get a score of $-s$.

4. An Overview of the Software Implementation

We have implemented the proposed model into a standalone software tool [4] in Python. Users can load game data from files organized in common formats, such as XML, JSON, etc. At a high level, the tool provides a variety of features that enable a user to analyze, replay, and visualize game data. First of all, it allows users to define a new pass between any players. Users may also change the position of any player by dragging and dropping the player on the field in any direction. Besides, users can visually replay the game at different speeds. During game replay, effectiveness of each pass is displayed on the screen real-time with details on subcomponents of the effectiveness presented on the left pane (Fig. 4). Paths of last three passes are shown for visual game analysis.

When a user defines a pass between any two players, pass effectiveness values for that pass is shown again on the left pane. In addition, pass risk area is also shown to visually identify the opponent players who pose a risk for the current pass (Fig. 5).

4.1. *Visual pass evaluation analytics*

Our tool also provides visual pass analytics from different aspects in the form of heatmaps (Fig. 6). This allows to analyze *what if scenarios* in a holistic manner.

Fig. 4. Live pass effectiveness display during game replay — Effectiveness value is displayed for the most recent pass which is between players 22 and 7 in the figure. Solid arrows show the preceding two passes, and dashed lines show the path that two previous ball owning players moved the ball before passing it to the next player. On the left-hand side of the tool, the subcomponents of each pass is shown with their numeric values.

Fig. 5. (Color online) Pass risk area visualization for manually defined passes — For each pass, the red lines indicate the borders of the risk area in which an opponent player may pose a threat to the pass.

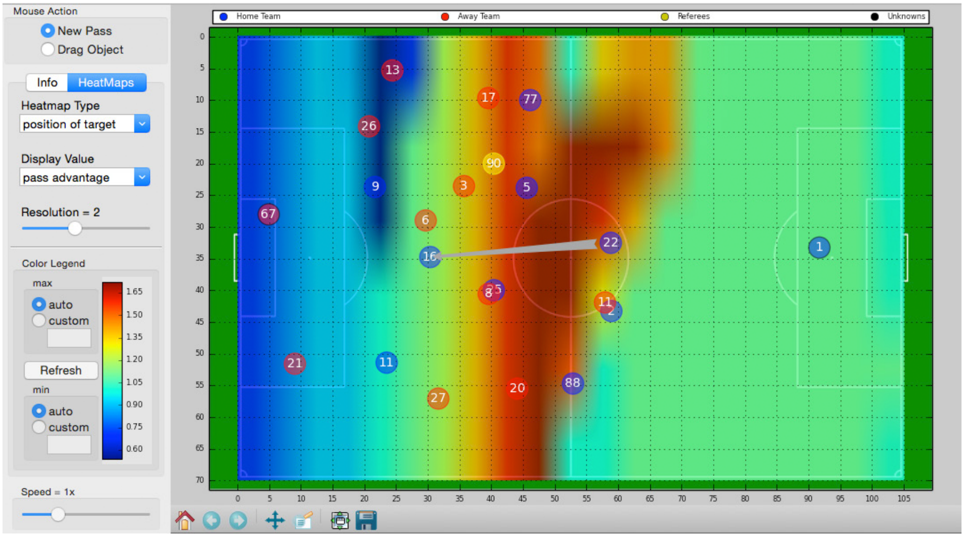

Fig. 6. Pass target position taking analysis through a heatmap — A pass is defined between players 22 and 16. The heatmap shows how the pass advantage subcomponent of the pass effectiveness would change if the pass target, player 16, received the pass at other points on the field.

Once a pass is defined by the user, the tool allows three different kinds of pass evaluation analytics. As an example, consider the example pass defined in Fig. 6 between players 22 and 16. The first analysis scenario evaluates the appropriateness of pass target's (player 16) positioning. To this end, the locations of all the players on the field except player 16 is assumed to be fixed. Then, assuming that player 16 may have chosen to be positioned at any point on the field, the target of the pass in Fig. 6 is set to each point on the field, and the effectiveness score is recomputed for the pass. Then, according to the resulting pass effectiveness value, that particular point is colored based on the used color spectrum (see the left side of Fig. 6). After this is repeated for all coordinates on the field, the resulting coloring of points is visualized in the form of a heatmap. The second one is very similar to the first one with the only difference that instead of the position of the pass target, the position of pass source (player 22) is varied over all points on the field, and the effectiveness score is recomputed for each point. Finally, in the third one, we consider that there is an imaginary defending player from the opposite team, whose goal is to prevent this pass. This imaginary defending player may be located anywhere on the field. We introduce this imaginary player and vary his position over all points on the field. For each point that this defender is located at, the effectiveness score of the pass is recomputed and registered for that particular point. The results are visualized again in the form of a heatmap. Besides effectiveness scores, users may choose to visualize the change in any subcomponents of pass effectiveness (e.g., pass advantage, gain, risk, etc.).

5. Evaluation and Validation

In order to evaluate the proposed pass effectiveness model, we performed a user study with a number of domain experts who are mostly active professional trainers or players. To this end, we prepared a survey which asks the participants to evaluate and rank different pass alternatives from different perspectives. We next present a summary of the survey and participant profile. Survey questions and responses are provided in the supplementary material.

5.1. *Survey summary*

The survey consists of 15 questions. First 3 questions are about the participant's background such as experience in the domain, certifications, etc. In each of the remaining 12 questions, several alternative pass scenarios were shown to a participant, and s/he was asked to sort the given scenarios (often, in terms of how good each pass is, and sometimes, how risky each pass is). For instance, if the participants were asked to sort given three pass alternatives initiated by P_1, P_2, and P_3 in terms of how good they were, the answer is expected to be in the form of an ordering like P_2–P_1–P_3 where the first one is the best pass, while the last one is the worst among the provided scenarios.

5.2. *Survey participants profile*

In total, 34 people participated in the survey. We eliminated those participants who did not answer some of the questions in the survey, i.e., 17 participants are left after this filtering. Among 17 participants, several of them provided responses to some questions in a format different than expected (e.g., listing only two passes, such as P_3–P_1, while all three passes were expected in the ordering). We ignored such vague responses, as we could not interpret them fully. However, we kept the proper responses from those participants.

Table 1 provides the experience level of the participants as a coach/trainer. In summary, the majority of the participants (88%) have 1+ years of coaching/trainer experience.

Table 2 presents the experience level of the participants as a soccer player. In summary, a substantial number of the participants (82%) have 3+ years of soccer playing experience.

Table 1. Experience level of participants as a coach/trainer.

Experience as a coach/trainer?	# of participants
3+ Years	11
1–3 Years	4
No experience	2

Table 2. Experience level of participants as a soccer player.

Experience as soccer player?	# of participants
3+ Years	14
No experience	3

Table 3. Formal certification profile of survey participants.

Which of the following license do you have?	# of participants
UEFA-licensed coach	5
TFF-licensed coach	5
TFF course-completed coach	2
Other	3
No official course	2

Finally, Table 3 provides the figures for the types of official certification/training that the participants had. In summary, 88% of participants have completed some sort of formal training, while 59% of participants hold coaching licenses from UEFA (the Union of European Football Associations) and/or TFF (Turkish Football Federation).

5.3. *Brief overview of surveyed pass scenarios*

The first pass scenario ($Q1$ and $Q2$) focuses on evaluating different choices while moving the ball from the end of one's own field to the opponent's field during an attack initiation. Considered options involve (i) a greedy-type choice of making a long deep pass to a teammate close to the edge of the opponent penalty region, (ii) a short pass to the center with relatively open space ahead, and (iii) a diagonal long pass to the opposite side which is defended by less opponents. The next scenario ($Q3$ and $Q4$) studies different passes to improve already a mature attack. More specifically, it explores the options of (i) passing the ball further deeper into the opponent's side with a short pass to a teammate who is closely defended by an opponent, (ii) passing the ball back away from the opponent's goal towards the center to move the game to a more open space with no "very close" defenders, (iii) making a horizontal pass to the opposite open side close to the edge of penalty region. Another scenario ($Q5$) focuses on the most appropriate position taking decision for a forward player to increase the likelihood of receiving a highly effective pass. In different alternatives, only the position of the pass receiving player is changed, all the other players keep their position. In particular, this scenario considers that a forward player may wait for a pass (i) inside a heavily defended penalty region, (ii) outside of the penalty region but with a blocked shooting angle by a defendant, and (iii) outside of the penalty region with somewhat clear shooting angle. The next scenario ($Q6$) evaluates a back pass inside a team's own half pitch. Since often a back pass leads to seriously dangerous situations for the back passing team, this scenario investigates the potential risk that is posed by each back pass. An important aspect that is given a special emphasis is the effect of potential threat of pressure that the opponent team may apply right after the back pass, which may stress the back passing team and force them to make a serious mistake. In this scenario, (i) an almost horizontal back pass where the opponent team may immediately pose pressure on the receiving player, (ii) a deep back pass with some pressure possibility by the opponent team, and (iii) a similar deep back pass with minimal pressure risk by the opponent team. Another scenario ($Q7$) investigates the case where a team has just moved to the opponent's half pitch and is building up an attack. It studies different decisions that an attacking midfielder may make, i.e., (i) long horizontal pass to a wing with no immediate blocking defenders around the path of the pass, (ii) short vertical pass to an area crowded with opponent players, (iii) long horizontal pass to another wing through or over several opponent players that are on or around the pass path. The next scenario ($Q8$) illustrates passing the ball from own half of the pitch to the opponent's half. However, the center area is crowded with opponent players. Hence, an intermediate pass is required to pass the ball in two steps to the other side of the pitch. In this context, this scenario assesses the position taking of the intermediate player who usually acts as the playmaker. The investigated positioning options of the playmaker include (i) a position that is at a distance from opponent players and no opponents on the path between him and the next pass receiving player, (ii) similar to the previous one, but this time, there are some threat posing defenders between the playmaker and his passing partner, (iii) the playmaker is in the middle of a group of opponent players and relatively closer to the player to whom s/he is going to pass the ball. In another scenario ($Q9$), we focus on a case where a team makes a mistake and passes the ball to an opponent player (i.e., mis-pass) during a developing attack around the middle region of the opponent. Each considered mis-pass has a different degree of likelihood to start a counter attack before the ball-losing team gets organized in the defense. The considered options include (i) a short mis-pass to a point with a crowd of teammates who may prevent a counter attack, (ii) a medium-length mis-pass to an opponent with a teammate in front around the center without any players from own team, who may pose a threat to a possible quick pass to this next player, and (iii) a long mis-pass to an opponent player who is defended by at least one player, and the only passing option for the ball-receiving player is a back pass which is not likely to create a counter attack. The next scenario ($Q10$) investigates the position-taking decision of a defense player on the effectiveness of a pass in a mature attack. In particular, the explored options include (i) staying close to the center of the pass path for possible intervention, (ii) staying close to the ball receiver to either steal the ball or tackle the pass-receiving player, and (iii) stay inside the penalty region to prevent any opponent player entering the region or shooting from that particular angle. Another scenario ($Q11$) focuses on post-ball-stealing pass decision to quickly initiate a counter attack. The ball stealing player considers (i) passing the ball with a short pass to a midfielder who is close to the center of the field and has opportunity to pass the ball further to a forward player who is not defended well, (ii) passing the ball with a long pass to the right wing which has open space, but there are several threat-posing players who may intervene the pass, and (iii) passing the ball with a medium range risky pass to the left wing which is somewhat defended better than the right wing. Finally, the last scenario evaluates the initiation of an attack in a set play by a defense player from the middle of his team's half of the pitch. There are three midfielders that the defense player may consider to pass the ball. One is at the center of field and closely defended by an opponent. Another one is on the near-right of the field center with some open space ahead and some opponent players at mid-range distance. The third midfielder is on the far-left of the field center with large open space that s/he may possibly run with the ball along the field border on the left.

5.4. *Survey evaluation and scoring metrics*

In order to calculate the accuracy of the proposed pass evaluation model, we compare the survey results with those provided by our tool. To compute an accuracy score, we use the following metric. Assume that for a particular survey question on comparative evaluation of n different pass scenarios, the correct (as voted by the majority of the domain experts) ordering of different pass scenarios is $S = P_1, P_2, \ldots, P_n$. Then, we create a set of all possible pairs (P_i, P_j) of these pass scenarios where the relative ordering of P_i, P_j is consistent with their order in S. That is, P_i should come before P_j in S. As an example, for $S = P_1, P_2, P_3$, the set of valid pairs would be $\{(P_1, P_2), (P_1, P_3), (P_2, P_3)\}$. A similar pairs set is computed for the ordering created by our pass evaluation model. Then, the number of common pairs between these two sets is used as the score of our model for that question. For instance, for the above example, assume that the ordering of pass scenarios is $S' = P_2, P_1, P_3$ based on our pass evaluation model. Then, the accuracy score would be 2, as two pairs (i.e., $\{(P_1, P_3), (P_2, P_3)\}$) are shared between S and S'. Such a scoring scheme allows for the quantitative evaluation of partial matches between expert evaluations and that of our model.

In order to treat all questions equally, in our survey, we include 3 pass scenarios in each question, which makes sure that the maximum score that can be contributed by each question is the same (i.e., 3, as there are three possible pairs). Therefore, we do not further normalize with the possible number of pairs in each question.

Table 4 summarizes the responses to the questions in the survey by the participants where columns are questions, rows are all possible permutations of pass alternatives, and a number in cell (i, j) represents the number of participants who provided permutation i as answer to question j (e.g., frequency of each permutation to appear under each question in participant responses). For each question, the most common answer is chosen as the consensus answer. In two questions (6 and 12), there were 2 alternative responses that shared the highest number of votes (5 and 6, respectively). In those questions, instead of opting for one of the answers, we considered both answers as correct. Note that the sum of the frequencies under each

Table 4. Responses of participants to survey questions.

	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8	Q_9	Q_{10}	Q_{11}	Q_{12}
P_1–P_2–P_3	1	0	2	1	0	**5**	1	2	2	1	4	1
P_1–P_3–P_2	0	3	0	0	0	2	3	0	1	0	2	**6**
P_2–P_1–P_3	1	3	**7**	0	**6**	1	**11**	**13**	4	1	3	1
P_2–P_3–P_1	4	**6**	2	2	4	2	1	1	**9**	**9**	**7**	0
P_3–P_1–P_2	0	1	2	4	0	2	0	1	0	0	0	**6**
P_3–P_2–P_1	**9**	3	2	**9**	3	**5**	0	0	0	6	1	2

question may be slightly different, as some participants did not respond to some questions in the way we expected (e.g., provided permutation may not include one or two passes). Those incomplete responses are eliminated. Therefore, some questions do not have a response from all participants.

5.5. *Results*

In this section, we report our accuracy evaluation results with respect to the above survey data.

5.5.1. *Learning the values of model parameters*

Our model includes several parameters (Table 5). One set of parameters are the weights associated with each subcomponent of the pass effectiveness model. Besides, we have radius parameters for the risk area and shooting line definitions as discussed in Sec. 3. Hence, we need to set these parameters first before we can get results. In order to avoid arbitrary setting of these parameter values, and compute their optimum values, we turn the problem into an optimization problem with an objective function that maximizes the similarity of the model-produced survey results to those answers provided by the domain experts. We employ and test three established optimization algorithms, namely, genetic optimization, hill-climbing, and simulated annealing [13]. Pseudocodes of these algorithms are provided in Appendix A. The accuracy results that we obtained from these alternative algorithms are provided in Table 6. Since the genetic optimization algorithm provides the best accuracy,

Table 5. Model parameters learnt through training (w_5 is not part of the model for this specific study).

Symbol	Description
w_1	Weight of gain
w_2	Weight of pass advantage
w_3	Weight of goal chance
w_4	Weight of decision time
r_1	Radius of pass source
r_2	Pass target radius coefficient
r_3	Shooting line radius

Table 6. Evaluating alternative optimization algorithms to learn the values of weights and parameters.

Optimization algorithm	Input details	Number of iterations	Max score (out of 36)	Accuracy (%)
Genetic	popsize = 50, mutprob = 0.2, step = 1, elite = 0.2, maxiter = 100 weights range: (1,10), (1,5), (1,20), (1,100), (1,100), (1,100), (1,100)	100	34	94.4
Simulated annealing	$T = 10,000$, cool = 0.95, step = 1 weights range: (1,10), (1,5), (1,20), (1,100), (1,100), (1,100), (1,100)	1000	25	69.4
Hill-climbing	weights range: (1,10), (1,5), (1,20), (1,100), (1,100), (1,100), (1,100)	N/A	25	69.4

we employ this algorithm while learning the values of model parameters. We adopted the default parameters as suggested by Segaran [13] (population size = 50, mutation probability = 0.2, step = 1, elite ratio = 0.2), and run 100 iterations.

5.5.2. *Accuracy evaluation*

In order to evaluate the success of the resulting model, we performed leave-one-out cross validation approach [12]. That is, we trained the model with all survey response data except for one question which is used for testing. We repeated this 12 times by making sure that each time a distinct question is chosen for testing. We computed score and accuracy in each iteration where scores are computed as discussed in the previous section (Sec. 5.4). During the computation of pass effectiveness scores, we employ z-scored values of individual pass effectiveness components to extract meaningful weight values which are normalized to make sure that they sum up to 1. Table 7 provides the detailed results for each iteration.

6. Discussion

Our experimental results point to three major observations. First, our model is able to capture the expert evaluations of a number of pass scenarios with relatively high accuracy, 94.7% (std: 2.3%). Second, based on the average weight values, pass advantage and goal chance components contribute most to the pass effectiveness score, while the contribution of gain and decision time is limited. Third, over the iterations, Table 7 reports very diverse values of the weights, while 10 out of 11 training questions remain the same in any pair of table lines. This suggests that the model may be too general, or rather, that the validation dataset may be too scarce.

As opposed to the fact that many of the reviewed techniques in the literature are offense oriented (e.g., [42, 45]), our model can also accurately evaluate passes that are not in offense region. As an example, in the case of Q_{12}, we compare three alternative passes that defense players are making to the midfielders within their own half of the pitch, i.e., almost no goal possibility. Even at such an early stage of the set play,

Table 7. Leave-one-out cross validation results.

	Score	Learnt parameter values (r_1, r_2, r_3, w_1, w_2, w_3, w_4) — z score							Overall accuracy
Test quest	(Out of 3)	r_1	r_2	r_3	w_1	w_2	w_3	w_4	(Out of 33) (%)
Q_1	3.0/ 100%	10	3	4	0.03	0.55	0.36	0.06	31–93.9
Q_2	2.0/ 66.60%	10	5	9	0.09	0.58	0.20	0.13	31–93.9
Q_3	3.0/ 100%	10	5	4	0.03	0.63	0.24	0.10	31–93.9
Q_4	3.0/ 100%	10	3	9	0.08	0.44	0.44	0.04	30–90.9
Q_5	3.0/ 100%	10	5	9	0.05	0.64	0.24	0.06	32–96.9
Q_6	3.0/ 0%	10	5	9	0.18	0.55	0.26	0.01	31–93.9
Q_7	2.0/ 66.60%	10	3	4	0.10	0.47	0.42	0.02	32–96.9
Q_8	3.0/ 100%	10	5	9	0.13	0.39	0.47	0.02	31–93.9
Q_9	2.0/ 66.60%	10	5	11	0.16	0.42	0.38	0.05	31–93.9
Q_{10}	2.0/ 66.60%	10	5	9	0.24	0.59	0.16	0.01	33–100
Q_{11}	3.0/ 100%	10	3	4	0.06	0.46	0.47	0.01	31–93.9
Q_{12}	3.0/ 100%	10	3	9	0.02	0.62	0.23	0.13	31–93.9
Mean		10.0	4.2	7.5	0.10	0.53	0.32	0.05	31.25–94.7
Std		0.0	1.0	2.6	0.07	0.09	0.11	0.05	0.75–2.3

our model predicts the ordering of passes in terms of their effectiveness scores fully consistent with expert evaluations. As another example, in Q_{11}, there is a possibility of counterattack after stealing the ball from an opponent. Even though the event is taking place in the middle of the passers' own half of the pitch, the proposed model's evaluation of alternative passes for the best counterattack completely overlaps with that of experts. Finally, in the case of Q_4, although pass P_1 is a vertical pass that moves the ball closer to the opponent's goal area, it is evaluated as less effective than P_2 which is actually a back pass moving the ball away from the goal area. In such a tricky case, many models in the literature would value P_1 over P_2, but our model successfully makes the right assessment which is consistent with expert evaluations.

This work may contribute to complex systems view understanding of soccer in two distinct ways. First, network models of soccer games [31–33] may be constructed at a finer granularity with edge weights that are computed based on pass effectiveness scores. On such models, key players that are at the center of most key passes may be detected in a more realistic manner. This may guide coaches to re-assign players in their team based on their effective pass making skills. Besides, they may implement additional measures to prevent players with high pass effectiveness scores in the opponent team.

Another potential contribution of the proposed model in understanding soccer game is that now a particular type of simulation studies may also become possible. More specifically, over past games, effective pass making skill levels for each player may be determined. Based on this data, one may perform what-if scenario analysis (e.g., what would happen if a certain player with high effective-passing skills is prevented by the opponent) on the network structure of a team. To this end, the passing network of a team may be considered as a flow network, and flux balance analysis [48] from the systems biology domain [49] may be adapted to soccer, where effective passing skill values of players will be set as upper bound constraints for the amount of flow that they can handle. This way, one may simulate and analyze potential players who may make up for the prevented/blocked/sold player in the game. In a way, this will allow to perform robustness analysis of teams, in that, teams that depend on a single star player may be shown not to be robust, while those that rely on team play may stand out as teams that are more robust to opponent teams with such tactical measures on their players with high effective passing skills.

6.1. *Limitations*

We note the following limitations of the presented study. First, our model does not directly incorporate game context into its consideration. For instance, being under press may have a different consideration of what a good pass is than a counter-attack setting. Second, our motion model is a simple one which does not consider many parameters, such as the acceleration of players, ball speed changes, etc. Third, the user survey dataset used for validating our model is small, which may lead to overfitting during parameter learning. Moreover, survey questions cover only a small subset of the space for all possible pass types that may take place in practice. Hence, the accuracy values presented in the previous section may not fully reflect the general performance of the model. Therefore, one may consider the presented results as a first step for the validity of the model, and more experiments with a larger dataset are required to further validate the descriptive power of the proposed pass effectiveness model.

7. Conclusion and Future Work

To sum up, in this paper, we propose a quantitative pass evaluation model for computational soccer game data analytics. Our model combines a variety of factors such as risk, gain, goal chance, etc. to mimic expert evaluation. We employ machine learning to learn some of the model parameters from expert knowledge. In order to evaluate the effectiveness of our model, we conduct a user survey with certified and experienced professionals with domain knowledge. We demonstrate that our model captures domain expert evaluations of a number of example scenarios with 94% accuracy. We have also implemented and incorporated our model into a soccer data analytics tool which is open-source and freely available to the researchers and practitioners in the field.

As for future work directions, our work may be extended in several ways. First, one may take the game context into consideration. More specifically, instead of a static flat model, one may create a multilevel-model similar to a decision-tree structure. Based on the current game context parameters, different branches of the model may be executed. In practice, such a hierarchical organization may be at full model level with differing definitions of pass effectiveness at different branches, or it may also be at a level, where only the weights of pass effectiveness components change based on the game context. For the latter, the weight parameters may be learned in a game context specific way from past games or domain experts via enriched questionnaires that also describe the game setting for each particular pass. Second, one may adopt a more involved motion model that considers, for instance, how the ball moves in three-dimensional space on a particular trajectory following a dynamic speed function. Besides, player specific acceleration, maximum speed, passing skill quality parameters may be learnt from pass matches, and incorporated into the model. This will allow to create player specific personalized game models that will assign different pass effectiveness scores depending on who the passing, receiving, and defending players are. Finally, our results rely on a relatively small set of data points. In order to validate the proposed model and assess the value of the pass effectiveness scores, more experiments with larger datasets may be carried out.

Acknowledgment

Authors would like to thank Dr. Onur Guzey for fruitful discussions on the material presented in this paper.

Appendix A. Optimization Algorithms

```
CostFunction (weights)
    Get answers from tool based on the given weights (eg.[p1,p3, p2])
    i = 0
    score = 0
    While i < 12: (12 is # number of question)
      get Qis (Answer for Qi from survey) (Sample answer: . [p1,p3, p2])
      get Qia (Answer for Qi from tool)
      get possible pairs of two from Qia and Qis and keep order the same
      if pair1 of Qia == pair1 of Qis:
             score++
    return score
```

```
Mutation (w)
    generate a random int i between 0 to length of domain
    t = generate a random number between 0–1
    if t <0.5
      w=w[0:i]+[w[i]-step]+w[i+1]
```

```
else if w[i]<domain[i]-1
  w=w[0:i]+[w[i]+step]+w[i+1]
else
  w=w
return w
```

```
Crossover (w1,w2)
    generate a random int i between 0 to length of domain
    w=w1[0:i]+w2[i:] # get the firt I item from list w1
                     # and the last item from w2
                     # and create a new list
    return w
```

```
Genetic Optimization
    domanin = [(1, 10), (1, 5), (1, 20), (1, 100), (1, 100), (1, 100), (1, 100)]
    Initialize random weights, i = 0
    for i < maxiter
         evaluate the weights via cost function
         choose top elites
         while newpopulation < maxpopulation
              i = a random number between 0 to 1
              if i<mutation probability
                   generate new weight population with mutation
              else
                   generate new weight population with crossover
         i++
    return best weight
```

```
HillClimbing Optimization Pseudo Code
    Initialize random weights
    while True
          for i < length of domain
                generate the list of neighbor weight
                while newpopulation < maxpopulation
          get current best weight via cost function
          for i < length of neighbor list
                compare each neighbor weight
                if new weight has more more than current weight
                        current weight = new weight
          if cost of current weight == cost of new weight
          break
    return current weight
```

```
Simulated Annealing Optimization Pseudo Code
    Initialize random weights
    while T > 0.1
          get a random index between (1 and range-1)
          dir = get a random direction between (-step size and step size)
          generate a new weight via the dir to value at selected index
          get the cost value of original and new weight
          if new weight cost > original weight cost
                current weight= new weight
```

```
        cool down T
return current weight
```

References

[1] Asai, T., Akatsuka, T., Nasako, M. and Murakami, O., Computer simulation of curve-ball kicking in soccer, in *The Engineering of Sport* (Blackwell Science, Oxford, UK, 1998), pp. 433–440.

[2] Beetz, M., von Hoyningen-Huene, N., Kirchlechner, B., Gedikli, S., Siles, F., Durus, M. and Lames, M., Aspogamo: Automated sports game analysis models, *Int. J. Comput. Sci. Sport* **8**(1) (2009) 1–21.

[3] Castellano, J., Alvarez-Pastor, D. and Bradley, P. S., Evaluation of research using computerized tracking systems (Amisco® and Prozone®) to analyse physical performance in elite soccer: A systematic review, *Sports Med.* **44**(5) (2014) 701–712.

[4] Delibas, E., Uzun, A., Inan, M. F., Guzey, O. and Cakmak, A., Interactive exploratory soccer data analytics, *J. Inf. Syst. Oper. Res.* (under review) (2018).

[5] Pollard, R., Ensum, J. and Taylor, S., Estimating the probability of a shot resulting in a goal: The effects of distance, angle and space, *Int. J. Soccer Sci.* **2**(1) (2004) 50–55.

[6] Goldberg, D. E., *Genetic Algorithms in Search Optimization and Machine Learning*, Vol. 412 (Addison-Wesley, Reading, 1989).

[7] Gyarmati, L., Kwak, H. and Rodriguez, P., Searching for a unique style in soccer, in *KDD Workshop on Large-Scale Sports Analytics* (2014).

[8] Horton, M., Gudmundsson, J., Chawla, S. and Estephan, J., Classification of passes in football matches using spatiotemporal data, arXiv:1407.5093.

[9] Lucey, P., Oliver, D., Carr, P., Roth, J. and Matthews, I., Assessing team strategy using spatiotemporal data, in *Proc. 19th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining* (Chicago, IL, USA, August 2013), pp. 1366–1374.

[10] Gréhaigne, J. F., Godbout, P. and Bouthier, D., Performance assessment in team sports, *J. Teach. Phys. Educ.* **16**(4) (1997) 500–516.

[11] Rampinini, E., Coutts, A. J., Castagna, C., Sassi, R. and Impellizzeri, F. M., Variation in top level soccer match performance, *Int. J. Sports Med.* **28**(12) (2007) 1018–1024.

[12] Refaeilzadeh, P., Tang, L. and Liu, H., Cross-validation, in *Encyclopedia of Database Systems* (Springer, New York, 2009), pp. 532–538.

[13] Segaran, T., *Programming Collective Intelligence* (O'Reilly Media, Inc., Sebastopol, 2007).

[14] Taki, T. and Hasegawa, J. I., Visualization of dominant region in team games and its application to teamwork analysis, in *IEEE Proceedings: Computer Graphics International*, 2000, pp. 227–235.

[15] Maheswaran, R., Chang, Y.-H., Henehan, A. and Danesis, S., Deconstructing the rebound with optical tracking data, in *MIT Sloan Sports Analytics Conf.*, 2012.

[16] Cervone, D., D'Amour, A., Bornn, L. and Goldsberry, K., A multiresolution stochastic process model for predicting basketball possession outcomes, *J. Am. Stat. Assoc.* **111**(514) (2016) 585–599.

[17] Taki, T. and Hasegawa, J. I., Dominant region: A basic feature for group motion analysis and its application to teamwork evaluation in soccer games, in *Electronic Imaging'99, International Society for Optics and Photonics* (San Jose, CA, USA, 1999), pp. 48–57.

[18] Grehaigne, J. F., Bouthier, D. and David, B., Dynamic-system analysis of opponent relationships in collective actions in soccer, *J. Sports Sci.* **15**(2) (1997) 137–149.

[19] McGarry, T., Anderson, D. I., Wallace, S. A., Hughes, M. D. and Franks, I. M., Sport competition as a dynamical self-organizing system, *J. Sports Sci.* **20**(10) (2002) 771–781.

[20] Folgado, H., Duarte, R., Fernandes, O. and Sampaio, J., Competing with lower level opponents decreases intra-team movement synchronization and time-motion demands during pre-season soccer matches, *PloS One* **9**(5) (2014) e97145.
[21] Packing-rate. Available at http://www.impect.com/en/#idea, accessed on February 1, 2018.
[22] Stöckl, M., Plück, D. and Lames, M., Modelling game sports as complex systems — application of recurrence analysis to golf and soccer, *Math. Comput. Model. Dyn. Syst.* **23**(4) (2017) 399–415.
[23] Gréhaigne, J. F., Godbout, P. and Zerai, Z., How the "rapport de forces" evolves in a soccer match: The dynamics of collective decisions in a complex system, *Revista de Psicología del Deporte* **20**(2) (2011) 747–764.
[24] Davids, K., Araújo, D. and Shuttleworth, R., Applications of dynamical systems theory to football, *Sci. Footb. V* **537** (2005) 550.
[25] Reed, D. and Hughes, M., An exploration of team sport as a dynamical system, *Int. J. Perform. Anal. Sport* **6**(2) (2006) 114–125.
[26] Vilar, L., Araújo, D., Davids, K. and Bar-Yam, Y., Science of winning soccer: Emergent pattern-forming dynamics in association football, *J. Syst. Sci. Complex.* **26**(1) (2013) 73–84.
[27] Pappalardo, L. and Cintia, P., Quantifying the relation between performance and success in soccer, *Adv. Complex Syst.* (2017) 1750014.
[28] Kang, C. H., Hwang, J. R. and Li, K. J., Trajectory analysis for soccer players, in *Sixth IEEE International Conference on Data Mining Workshops, ICDM Workshops* (Hong Kong, December 2006), pp. 377–381.
[29] Lucey, P., Bialkowski, A., Monfort, M., Carr, P. and Matthews, I., Quality vs quantity: Improved shot prediction in soccer using strategic features from spatiotemporal data, in *Proc. 8th Annual MIT Sloan Sports Analytics Conf.* (Boston, MA, USA, February 2014), pp. 1–9.
[30] Brooks, J., Kerr, M. and Guttag, J., Developing a data-driven player ranking in soccer using predictive model weights, in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining* (San Francisco, CA, USA, August 2016), pp. 49–55.
[31] Cintia, P., Rinzivillo, S. and Pappalardo, L., A network-based approach to evaluate the performance of football teams, in *Machine Learning and Data Mining for Sports Analytics Workshop*, Porto, Portugal, September 2015.
[32] Clemente, F. M., Couceiro, M. S., Martins, F. M. L. and Mendes, R. S., Using network metrics in soccer: A macro-analysis, *J. Human Kinet.* **45**(1) (2015) 123–134.
[33] Peña, J. L. and Touchette, H., A network theory analysis of football strategies, in *Sports Physics: Proc. 2012 Euromech Physics of Sports Conf.*, Palaiseau, 2012, pp. 517–528 (ISBN 978-2-7302-1615-9).
[34] Redwood-Brown, A., Passing patterns before and after goal scoring in FA Premier League Soccer, *Int. J. Perform. Anal. Sport* **8**(3) (2008) 172–182.
[35] Grehaigne, J. F., Bouthier, D. and David, B., Dynamic-system analysis of opponent relationships in collective actions in soccer, *J. Sports Sci.* **15**(2) (1997) 137–149.
[36] Gonçalves, B., Coutinho, D., Santos, S., Lago-Penas, C., Jiménez, S. and Sampaio, J., Exploring team passing networks and player movement dynamics in youth association football, *PloS one* **12**(1) (2017) e0171156.
[37] Cotta, C., Mora, A. M., Merelo-Molina, C. and Merelo, J. J., Fifa world cup 2010: A network analysis of the champion team play, arXiv:1108.0261.
[38] Gudmundsson, J. and Horton, M., Spatio-temporal analysis of team sports, *ACM Comput. Surv.* **50**(2) (2017) 22.
[39] Borgatti, S. P., Centrality and network flow, *Soc. Netw.* **27**(1) (2005) 55–71.

[40] Cintia, P., Giannotti, F., Pappalardo, L., Pedreschi, D. and Malvaldi, M., The harsh rule of the goals: Data-driven performance indicators for football teams, in *IEEE Int. Conf. Data Science and Advanced Analytics (DSAA)* (Paris, France, 2015), pp. 1–10.

[41] Vercruyssen, V., De Raedt, L. and Davis, J., Qualitative spatial reasoning for soccer pass prediction, in *Proc. Workshop on Machine Learning and Data Mining for Sports Analytics*, Riva del Garda, Italy, September 2016.

[42] Link, D., Lang, S. and Seidenschwarz, P., Real time quantification of dangerousity in football using spatiotemporal tracking data, *PloS one* **11**(12) (2016) e0168768.

[43] Szczepański, Ł. and McHale, I., Beyond completion rate: Evaluating the passing ability of footballers, *J. R. Stat. Soc.: Ser. A (Stat. Soc.)* **179**(2) (2016) 513–533.

[44] Gyarmati, L. and Stanojevic, R., Qpass: A merit-based evaluation of soccer passes, in *KDD Workshop on Large-Scale Sports Analytics*, 2016.

[45] Power, P., Ruiz, H., Wei, X. and Lucey, P., Not all passes are created equal: Objectively measuring the risk and reward of passes in soccer from tracking data, in *Proc. 23rd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining* (Halifax, Nova Scotia, Canada, August 2017), pp. 1605–1613.

[46] Reiner, M., Task constraints of large-sided soccer games: Effects of manipulating rewarding rules. Doctoral dissertation, School of Human Kinetics, University of Lisbon (2017).

[47] Walpole, R. E., Myers, R. H., Myers, S. L. and Ye, K., *Probability and Statistics for Engineers and Scientists*, Vol. 5 (Macmillan, New York, 1993).

[48] Orth, J. D., Thiele, I. and Palsson, B. Ø., What is flux balance analysis? *Nat. Biotechnol.* **28**(3) (2010) 245.

[49] Kitano, H., Systems biology: A brief overview, *Science* **295**(5560) (2002) 1662–1664.