Team sports can be viewed as dynamical systems unfolding in time and thus require tools and approaches congruent to the analysis of dynamical systems. The analysis of the pattern-forming dynamics of player interactions can uncover the clues to underlying tactical behaviour. This study aims to propose quantitative measures of a team’s performance derived only using player interactions. Concretely, we segment the data into events ending with a goal attempt, that is, ‘‘Shot’’. Using the acquired sequences of events, we develop a coarse-grain activity model representing a player-to-player interaction network. We derive measures based on information theory and total interaction activity, to demonstrate an association with an attempt to score. In addition, we developed a novel machine learning approach to predict the likelihood of a team making an attempt to score during a segment of the match. Our developed prediction models showed an overall accuracy of 75.2% in predicting the correct segmental outcome from 13 matches in our dataset. The overall predicted winner of a match correlated with the true match outcome in 66.6% of the matches that ended in a result. Furthermore, the algorithm was evaluated on the largest available open collection of soccer logs. The algorithm showed an accuracy of 0.84 in the classification of the 42, 860 segments from 1, 941 matches and correctly predicted the match outcome in 81.9% of matches that ended in a result. The proposed measures of performance offer an insight into the underlying performance characteristics.

With the rise of tracking data, sports analytics can influence the game’s tactical aspects like never before. In football, measuring the quality of the players’ positioning to receive a pass in condition to score has much value. The Off-Ball Scoring Opportunity model was built to do just that. With that, players receive credit for being well-positioned to score, even if a teammate cannot get them the ball. It was originally modeled to consider the game’s snapshot only when a player executes an action. However, football is a continuous sport, where decision-making happens at all times, and actions are not discretized as in sports such as baseball and American football. In this paper, we propose a reinterpretation of the original model, where the Off-Ball Scoring Opportunity is calculated for every timestep an attacking player has the ball at his feet. It makes sense, since as long as a player has control of the ball, he can move it somewhere else on the pitch. Through this new form of applying the Off-Ball Scoring Opportunity model, we can build time-series that represent the scoring probability of the next on-ball event at any given moment in time. Later, we demonstrate how this way of using the model offers a much more in-depth view of attacking creation at an individual and team level.

Competitive team sports are one of the most informative scenarios in the research of team cooperation analysis. However, there is a lack of simple, robust, and accurate key eventbased methods when evaluating the performance of a soccer team. In this paper, we first built a ball-passing network to facilitate teamwork analysis of a soccer team, with the help of which we then proposed a novel model for quality assessment based on highlight moments to evaluate the performance of the team. Further, we develop a third model to identify the rhythm conversion of offensive/defensive tactics so as to quantify them. Using spatiotemporal tracking data of key events in 38 Premier League games, a comprehensive and systematic analysis is formed on the performance of both the team staff and players of the Everton Football Club. Also, the key factors to the match result are quantitatively explored and modeled.

In this paper, we study interaction dynamics in the game of football–soccer in the context of ball possession intervals. To do so, we analyze a database comprising one season of the five major football leagues of Europe. Using this input, we developed a stochastic model based on three agents: two teammates and one defender. Despite its simplicity, the model is able to capture, in good approximation, the statistical behavior of possession times, pass lengths, and number of passes performed. In the last section, we show that the model’s dynamics can be mapped into a Wiener process with drift and an absorbing barrier.

This paper reports an empirical investigation of game interruptions in competitions based on data from the European Big-5 leagues. The results showed that 1) visiting teams and teams facing strong opponents tended to have longer game interruptions; 2) the durations of throw-ins and stoppages due to out-of-bounds balls were significantly longer in the English Premier League than in the other four leagues (p < 0.001), and the durations of free kicks were longer in the English Premier League than in three other leagues (except Ligue 1, p < 0.001); 3) the durations of free kicks (p < 0.001), throwins (p = 0.038) and goal kicks (p < 0.001) increased towards the end of matches, but the durations of interruptions related to free kicks (p < 0.001) in the offensive zone decreased significantly as the end of the game approached; and 4) the durations of many interruptions served strategic purposes, as, for example, the time spent by a one-goal leading team on throw-ins (p < 0.001), goal kicks (p < 0.001), corner kicks (p < 0.01) and free kicks (p < 0.01) increased with game time, while for one-goal trailing teams, the time spent on these interruptions decreased over time (all p < 0.001).

Analysis of large dataset of soccer can be extremely difficult if it is not represented graphically. This paper proposes SoccerMetrics, an interactive, open source data visualization system for predicting the individual and team performance. The effectiveness of SoccerMetrics is demonstrated by the available dataset and its visualization.

After the great success of the Women’s World Cup in 2019, several platforms have started identifying the reasons for gender inequality in European football. Even though these inequalities emerge from a variety of key aspects in the modern sport, we focused on the game and evaluated the main differential features of European male and female football players in match actions data under the assumption of finding significant differences and established patterns between genders. A methodology for unbiased feature extraction and objective analysis is presented based on data integration and machine learning explainability algorithms. Female (n0 = 1511 ) and male (n1 = 2703) data points were collected from event data and categorized by game period and player position. We set up a supervised classification pipeline to predict the gender of each player by looking at their actions in the game. The comparison methodology did not include any qualitative enrichment or subjective analysis to prevent biased data enhancement or gender‑related processing. The pipeline included three representative binary classification models; A logic‑based Decision Trees, a probabilistic Logistic Regression and a multilevel perceptron Neural Network. Each model tried to draw the differences between male and female data points, and we extracted the results using machine learning explainability methods to understand the underlying mechanics of the models implemented. The study was able to determine pivotal factors that differentiate each gender performance as well as disseminate unique patterns by gender involving more than one indicator. Data enhancement and critical variables analysis are essential next steps to support this framework and serve as a baseline for further studies and training developments.

What do football passes and financial transactions have in common? Both are networked walk processes that we can observe, where records take the form of timestamped events that move something tangible from one node to another. Here we propose an approach to analyze this type of data that extracts the actual trajectories taken by the tangible items involved. The main advantage of analyzing the resulting trajectories compared to using, e.g., existing temporal network analysis techniques, is that sequential, temporal, and domain-specific aspects of the process are respected and retained. As a result, the approach lets us produce contextually-relevant insights. Demonstrating the usefulness of this technique, we consider passing play within association football matches (an unweighted process) and e-money transacted within a mobile money system (a weighted process). Proponents and providers of mobile money care to know how these systems are used—using trajectory extraction we find that 73% of e-money was used for stand-alone tasks and only 21.7% of account holders built up substantial savings at some point during a 6-month period. Coaches of football teams and sports analysts are interested in strategies of play that are advantageous. Trajectory extraction allows us to replicate classic results from sports science on data from the 2018 FIFA World Cup. Moreover, we are able to distinguish teams that consistently exhibited complex, multi-player dynamics of play during the 2017–2018 club season using ball passing trajectories, coincidentally identifying the winners of the five most competitive first-tier domestic leagues in Europe.

As one of the most basic techniques of modern football, passing not only involves coordination within the team and different passing routes but is also closely related to the opponent team and the situation of the game. How to evaluate players’ ability to pass in such a complex environment has been a popular issue in the field of sports performance. Based on social network analysis, this paper constructs passing network and calculates ten indicators such as degree centrality. Then, using classification algorithms –The Gradient Boosting Decision Tree where these ten indicators serve as features to train a model to predict the average goal of each player. After adjusting parameters, the accuracy of the model reaches 0.628 while the value of AUC is 0.709, which shows the prediction of the model is relatively high, and social network analysis is an effective way to evaluate players’ passing ability to some extent.

The prediction of the player's positions, or determining which position a player is suitable for based on sports performance and physiological indicators, plays a major role in association football. This research is based on the public dataset provided by Wyscout, from which player-related indicators are extracted and processed. Six indicators, including the accuracy of shot, the accuracy of simple pass, the accuracy of glb (Ground loose ball), the accuracy of defending duel,the accuracy of air duel, the accuracy of attacking duel, are selected according to the ANOVA (analysis of variance) test, and being imported into BP neural network for training. Since the neural network has three hyperparameters: training rate, iterations, and the number of neurons in the hidden layer, it is required to use the k-fold cross-validation to evaluate by which hyperparameter pair the model predict best. It is found that when the learning rate is set to 0.0125 and the hidden layer neuron is set to 6, the average accuracy of the cross-check is the highest, which is 73%. When iterations reach 300, the accuracy curve tends to converge. The final accuracy rate can reach 77%.

Women’s football isgaining supporters and practitioners worldwide, raising questions about what the differences are with men’s football. While the two sports are often compared based on the players’ physical attributes, we analyze the spatio-temporal events during matches in the last World Cups tocompare male and female teams based on their technical performance. We train an artificial intelligence model torecognize ifateam ismale orfemale based on variables that describe amatch’s playing intensity, accuracy, and performance quality. Our model accurately distinguishes between men’s and women’s football, revealing crucial technical differences, which we investigate through the extraction ofexplanations from the classifier’s decisions. The differences between men’s and women’s football are rooted inplay accuracy, the recovery time ofball possession, and the players’ performance quality. Our methodology may help journalists and fans understand what makes women’s football adistinct sport and coaches design tactics tailored tofemale teams.

The aim of this research was to analyze the player’s pass style with enhanced accuracy using the deep learning technique. We proposed Pass2vec, a passing style descriptor that can characterize each player’s passing style by combining detailed information on passes. Pass data was extracted from the ball event data from five European football leagues in the 2017–2018 season, which was divided into training and test set. The information on location, length, and direction of passes was combined using Convolutional Autoencoder. As a result, pass vectors were generated for each player. We verified the method with the player retrieval task, which successfully retrieved 76.5% of all players in the top-20 with the descriptor and the result outperformed previous methods. Also, player similarity analysis confirmed the resemblance of players passes on three representative cases, showing the actual application and practical use of the method. The results prove that this novel method for characterizing player’s styles with improved accuracy will enable us to understand passing better for player training and recruitment.

In elite-level soccer, the ability to take shots with both limbs from different positions in the pitch may be key to success. This research aimed to: 1) analyze footedness of elite-football players in European leagues during shooting by computing frequency of right- and left-foot use and accuracy; and 2) investigate whether an athlete’s distance from the target (goal, penalty, and outside penalty area) and pitch zone (center, left, or right from the goal) can constrain foot selection during shooting. We analyzed 1826 games from the 2017/18 season, divided between: Spanish LaLiga (380 matches); Italian Serie A (380 matches); English Premier League (380 matches); German Bundesliga (306 matches); and French Ligue 1 (380 matches). Results revealed asymmetrical proportions of foot selection, favoring the preferred foot for right- and left-footed athletes. Frequency of preferred foot selection increased as a function of distance from the target (i.e., the farther the athlete, higher the percentage of preferred foot selection). Shots taken from the left side were more often performed with the right foot and vice-versa, for both left- and right-footed athletes. Interestingly, asymmetries were observed only in foot selection, but not in performance, as success rate did not vary between limbs in any position.

We introduce the design and implementation of TimeElide, a visual analysis tool for the novel data abstraction of non-contiguous time series slices, namely time intervals that contain a sequence of timevalue pairs but are not adjacent to each other. This abstraction is relevant for analysis tasks where time periods of interest are known in advance or inferred from the data, rather than discovered through openended visual exploration. We present a visual encoding design space as an underpinning of TimeElide, and the new sparkbox technique for visualizing fine and coarse grained temporal structures within one view. Datasets from different domains and with varying characteristics guided the development and their analysis provides preliminary evidence of TimeElide’s utility.

Association football has been the subject of many research studies. In this work we present a study on player similarity using passing sequences extracted from games from the top-5 European football leagues during the 2017/2018 season. We present two different approaches: first, we only count the motifs a player is involved in; then we also take into consideration the specific position a player occupies in each motif. We also present a new way to objectively judge the quality of the generated models in football analytics. Our results show that the study of passing sequences can be used to study player similarity with relative success.

In this study, we define a new way of representing football player roles based on passing and receiving interactions. We develop a definition of player roles consisting of a linear combination of 12 common and interpretable passing/receiving patterns. Linear combinations are derived from the decomposition of players’ pitch passing and receiving networks using non-negative matrix factorization (NMF). Our model shows that 43% of the 1491 players studied in this paper had a maximum weight of less than 50% in each of the 12 common passing/receiving patterns. This suggests that a substantial percentage of players do not follow the specific passing/receiving patterns typically associated with their conventional role. The model also reveals the underlying differences in passing/receiving patterns amongst players who hold the same conventional role. It shows the intricacies of player patterns optimally when tasked with analyzing the most complex conventional roles such as midfielders, wingers, and forwards. Lastly, we show that the combinations of the 12 common passing/receiving patterns can be used as a footprint to find players with similar passing/receiving styles. For instance, our model found that Shaqiri and Fabinho had the highest similarity in passing/receiving styles to Oxlade-Chamberlain and Henderson. This is consistent with Liverpool FC’s transfers of Shaqiri and Fabinho to replace Oxlade-Chamberlain and Henderson’s positions respectively in the summer of 2018.

This paper analyses how sabotage, measured as the number of fouls and yellow and red cards, is affected by the relative performance of and asymmetries between teams, determined based on score margin and the odds gap between one team and its opponent. By applying detailed controls for within-match dynamics and differences in team quality, we first observe that badly losing favourites are more willing than losing underdogs to increase their unsporting behaviour in heterogeneous contests. There are more yellow cards and fouls as the game progresses. We further find that most sabotage (yellow cards) can be observed when the (absolute) goal difference is equal to 1. Teams decrease their level of sabotage (fouls) by increasing the number of goals of lead or trail. In addition, weaker teams have an incentive to engage in more unfair and destructive strategies, such as committing more fouls, including those penalized with yellow cards. However, the incidence of red cards is not influenced by the score margin or strength gap.

The aim of the current study is to analyze the effects of red and yellow cards on the scoring rate in elite soccer. The sample was composed of 1826 matches in the top five European leagues. All events were structured in 5-min intervals and were analyzed by means of a Generalized Linear Mixed Model with Poisson distribution, considering the presence of correlated data, where the dependent variable is represented by scoring rate. Team strength and home advantage were considered implicitly by means of a transformation of the betting odds for each game. The model also took into account the goal difference and time evolution. Overall, we found that after a sending off, each team’s scoring rate changes significantly, damaging the penalised team and favouring its opponent. When the player who is sent off belongs to the Away team, the impact of a red card is more or less maintained over time intervals. The red card effect, on the other hand, tends to fade over time when the affected team is stronger. The relative difference in scoring rates is also affected by the goal difference and the difference in booked players, being slightly lower for the team going ahead if it has more booked players. Our approach allows estimating the expected cumulative soring rate through time for various red card scenarios. Particularly if a red card is given with 30 min of remaining time, the expected impact is 0.39 goals if the guilty player is on the visiting team and 0.50 if he plays for the home team. Coaches and analysts could use this information to establish objectives for players and teams in training and matches and to be prepared for these very different scenarios of numerical superiority or inferiority.

Football analytics is a field that has been growing incredibly over the years thanks to the improvement of technologies capturing data in sports events. Outcomes of football matches are highly affected by the in-game decisions of football manager such as defending and attacking strategies or substituting particular football players. That is why football player recommendation is an important decision making task to gain the best results from a football match. To assist the football managers in this decision making process, a system that recommends the most suitable football player to replace a certain player is proposed. Our proposed model utilizes passing information during a game to learn feature embeddings of football players. Using the learned feature embeddings, a k-nearest neighbors (k-NN) model, an XGBoost model and an artificial neural network (ANN) model are trained to recommend the most similar and suitable replacement for a football player. The novelty of this recommendation system is that learned embeddings generate high-quality representations of football players which yield high performance for player recommendation when a replacement is needed.

Soccer is a sport characterised by open and dynamic play, with player actions and roles aligned according to team strategies simultaneously and at multiple temporal scales with high spatial freedom. This complexity presents an analytics challenge, which to date has largely been solved by decomposing the game according to specific criteria to analyse specific problems. We propose a more holistic approach, utilising Transformer or RNN components in the novel Seq2Event model, in which the next match event is predicted given prior match events and context. We show metric creation using a general purpose context-aware model as a deployable practical application, and demonstrate development of the poss-util metric using a Seq2Event model. Summarising the expectation of key attacking events (shot, cross) during each possession, our metric is shown to correlate over matches (𝑟 = 0.91, 𝑛 = 190) with the popular xG metric. Example practical application of poss-util to analyse behaviour over possessions and matches is made. Potential in sports with stronger sequentiality, such as rugby union, is discussed.

Lineup selection is an essential and important task in soccer matches. To win a match, coaches must consider various factors and select appropriate players for a planned formation. Computation-based tools have been proposed to help coaches on this complex task, but they are usually based on over-simplified models on player performances, do not support interactive analysis, and overlook the inputs by coaches. In this paper, we propose a method for visual analytics of soccer lineup selection by tackling two challenges: characterizing essential factors involved in generating optimal lineup, and supporting coach-driven visual analytics of lineup selection. We develop a lineup selection model that integrates such important factors, such as spatial regions of player actions and defensive interactions with opponent players. A visualization system, Team-Builder, is developed to help coaches control the process of lineup generation, explanation, and comparison through multiple coordinated views. The usefulness and effectiveness of our system are demonstrated by two case studies on a real-world soccer event dataset.

Association Football is probably the world’s most popular sport. Being able to characterise and compare football players is therefore a very important and impactful task. In this work we introduce spatial flow motifs as an extension of previous work on this problem, by incorporating both temporal and spatial information into the network analysis of football data. Our approach considers passing sequences and the role of the player in those sequences, complemented with the physical position of the field where the passes occurred. We provide experimental results of our proposed methodology on real-life event data from the Italian League, showing we can more accurately identify players when compared to using purely topological data.

Recently, football has seen the creation ofvarious novel, ubiquitous metrics used throughout clubs’ analytics departments. These can influence many oftheir day-to-day operations ranging from financial decisions on player transfers, toevaluation ofteam performance. At the forefront ofthis scientific movement isthe metric expected goals, ameasure which allows analysts toquantify how likely agiven shot istoresult inagoal however, xG models have not until this point considered using important features, e.g., player/team ability and psychological effects, and isnot widely trusted by everyone inthe wider football community. This study aims tosolve both these issues through the implementation ofmachine learning techniques by, modelling expected goals values using previously untested features and comparing the predictive ability oftraditional statistics against this newly developed metric. Error values from the expected goals models built inthis work were shown tobe competitive with optimal values from other papers, and some ofthe features added inthis study were revealed tohave asignificant impact on expected goals model outputs. Secondly, not only was expected goals found tobe asuperior predictor ofafootball team’s future success when compared totraditional statistics, but also our results outperformed those collected from an industry leader inthe same area.

Complex networks have been widely used in studying collective behaviours in soccer sports, such as examining tactical strategies, recognizing team characteristics, and discovering topological determinants for high team performance. The passing network of a team evolves and displays different temporal patterns, that are strongly linked to team status, tactical strategies, attacking/defending transitions, etc. Nevertheless, existing research has not illuminated the state dynamics of team passing networks, whereas similar methods have been extensively used in examining the dynamical brain networks constructed from human brain neuroimage data. This study aims to investigate the state dynamics of team passing networks in soccer sports. The introduced method incorporates multiple techniques, including sliding time window, network modeling, graph distance measure, clustering, and cluster validation. The final match of the FIFA World Cup 2018 was taken as an example, and the state dynamics of teams Croatia and France were analyzed respectively. Additionally, the effects of the time windows and graph distance measures on the results were briefly discussed. This study presents a novel outlook on examining the dynamics of team passing networks, as it facilitates the recognition of important team states or state transitions in soccer and other team ball-passing sports for further analysis.

Nowadays, coaches exploit data analysis in soccer (football) matches to plan their strategies against opponents. Network science, a subdomain of data analytics, is widely used to analyze soccer matches by treating players as nodes and passes between them as edges. However, single-layer methods for analyzing games overlook critical information by aggregating different types of passes into one layer. This paper introduces a new model called the Complex Multiplex Passing Network (CMPN) for analyzing team sports performance, with a focus on soccer matches. We utilized a real-world dataset to construct the multilayer structure of the CMPN. Each layer represents a specific type of pass between players. Using the CMPN, we conducted various analysis tasks at different topological scales. Firstly, we identified the core players of teams by calculating the PageRank versatility of each player. Next, we discovered the types of passes between trios of players based on multilayer motifs. Additionally, we measured similarities between passing tactics using the Pearson interlayer assortativity measure. Finally, we employed a long short-term memory network to predict the outcomes of attacking plays using the CMPN model. The predictions achieved over 90% accuracy and approximately 70% F-measure. These findings offer practical value to coaches and performance analysts, as they enable appropriate planning by predicting playing styles in different competitions and neutralizing the strategies of opposing teams.

It is still largely unclear to what extent bettors update their prior assumptions about the strength and form of competing teams considering the dynamics during the match. This is of interest not only from the psychological perspective, but also as the pricing of live odds ideally should be driven both by the (objective) outcome probabilities and also the bettors’ behaviour. Using state-space models (SSMs) to account for the dynamically evolving latent sentiment of the betting market, we analyse a unique highfrequency data set on stakes placed during the match. We find that stakes in the live-betting market are driven both by perceived pre-game strength and by in-game strength, the latter as measured by the Valuing Actions by Estimating Probabilities (VAEP) approach. Both effects vary over the course of the match.

Within the context of professional football, we examined the impact of the interim game state on risk-taking and performance during a dynamic tournament. This study used 9,256 segments from the top five European football leagues as samples. These segments were derived from 1,826 games played during the 2017–2018 season. Poisson regression was employed to analyze the distinct effects of game state and heterogeneity on performance under pressure. The results indicated that stronger teams tended to increase their attack intensity when facing weaker opponents. However, as their lead expanded, they tended to reduce their attack intensity, particularly in matches with heterogeneous characteristics. Moreover, teams trailing in scores tended to intensify their attacks but achieved little. However, leading teams consistently underperformed in terms of blocked shots and corner kicks. Additionally, tied teams systematically exhibited lower performance in shots on target and free kicks compared to leading teams, despite having a higher motivation to excel. These findings extend our understanding of how risk-taking and performance depend on disclosing information regarding relative performance.

In this work, a novel framework for the evaluation of individual football (soccer) players using event stream data is introduced. Applying a debiased machine learning approach (DML), we estimate the contribution of players to a possession sequence, i.e. a sequence of consecutive on-ball events stopped either by the opponent team gaining the possession or by an action of the referee. The estimates are then used to derive a metric to rate players, which is able to account for team strengths and game context. To show the potential of our novel rating approach we compare it to existing ratings by measuring the quality of match outcome forecasts generated when the ratings are used as predictor variables.

Using automated data analysis to understand what makes a play successful in football can enable teams to make data-driven decisions that may enhance their performance throughout the season. Analyzing different types of plays (e.g., corner, penalty, free kicks) requires different considerations. This work focuses on the analysis of corner kick plays. However, the central ideas apply to analyzing all types of plays. While prior analyses (univariate, bivariate, multivariate) have explored the link between contextual factors (e.g., match period, type of defensive marking) and the level of success of a corner kick (e.g., shot, shot on goal, goal), there has been no attempt to combine spatiotemporal event data (sequences of ball movements through the field) and contextual information to determine when and how (strategy) a particular type of corner kick play (tactic) is more likely to succeed or not. To address this gap, we propose an approach that (1) transforms spatiotemporal data into an alternative representation suitable for mining sequential patterns, (2) identifies and characterizes the sequential patterns used by offensive teams to move the ball toward the scoring zone (tactics), and (3) extracts contrast patterns to identify under what conditions different tactics result in increased chances of success or failure; we call these conditions strategies. Our results suggest that favorable and unfavorable conditions for tactic application are not the same across different tactics, supporting the argument that there is a benefit in performing an analysis that treats different tactics separately, where spatiotemporal information plays a crucial role. Unlike prior works on the corner kick, our approach can capture how the interaction between multiple contextual factors impacts the outcome of a corner kick. At the same time, the results can be explained to others in natural languages.

Momentum has been a consistently studied aspect of sports science for decades. Among the established literature, there has, at times, been a discrepancy between conclusions. However, if momentum is indeed an actual phenomenon, it would affect all aspects of sports, from player evaluation to pre-game prediction and betting. Therefore, using momentum-based features that quantify a team’s linear trend of play, we develop a data pipeline that uses a small sample of recent games to assess teams’ quality of play and measure the predictive power of momentum-based features versus the predictive power of more traditional frequency-based features across several leagues using several machine learning techniques. More precisely, we use our pipeline to determine the differences in the predictive power of momentum-based features and standard statistical features for the National Hockey League (NHL), National Basketball Association (NBA), and five major first-division European football leagues. Our findings show little evidence that momentum has superior predictive power in the NBA. Still, we found some instances of the effects of momentum on the NHL that produced better pre-game predictors, whereas we view a similar trend in European football/soccer. Our results indicate that momentum-based features combined with frequency-based features could improve pre-game prediction models and that, in the future, momentum should be studied more from a feature/performance indicator point-of-view and less from the view of the dependence of sequential outcomes, thus attempting to distance momentum from the binary view of winning and losing.

What determines the outcome of a shot (scored or unscored) in football (soccer)? Numerous studies have investigated various aspects of this question, including the skills and physical/mental state of the shooter or goalkeeper, the positional information of shots, as well as the attacking styles and defensive formations of the opposing team. However, a critical question has received limited attention: How does the passing path affect the outcome of a shot? In other words, does the path of the ball before shooting significantly influence the result when the same player takes two shots from the same location? This study aims to fill the gap in the literature by conducting qualitative studies using a dataset comprising 34,938 shots, along with corresponding passing paths from toptier football leagues and international competitions such as the World Cup. Eighteen path features were extracted and applied to three different machine-learning models. The results indicate that the passing path, whether with or without the positional information of shots, can indeed predict shooting outcomes and reveal influential path features. Moreover, it suggests that taking quick actions to move the ball across areas with a high probability of scoring a goal can significantly increases the chance of a successful shot. Interestingly, certain path features that are commonly considered important for team performance, such as the distribution of passes among players and the overall path length, were found to be less significant for shooting outcomes. These findings enhance our understanding of the effective ball-passing and provide valuable insights into the critical factors for achieving successful shots in football games.

While it is evident that disparities exist across various areas on a football pitch, and numerous studies have investigated spatio-temporal datasets in football for various analyses, there remains a lack of an effective method for quantitatively partitioning the pitch into specific areas with different properties. To address this gap, this article presents a novel approach to partitioning a football pitch into distinct areas based on successful passing paths that lead to goals. Utilizing hierarchical clustering and spatial/temporal features derived from successful passing paths, the study provides multi-level partitions of the football pitch, revealing detailed insights into the relationships between specific areas and scored shots in football games. Empirical analysis of over 4000 successful passing paths from various football leagues and international football events demonstrates the effectiveness of the proposed methodology in identifying and understanding the diverse areas of football pitches. The findings suggest practical applications in football analysis, aiding coaches and specialists in tactics development and informing player positioning and movement strategies.

Spatio-temporal event sequences consist of activities or occurrences involving various interconnected elements in space and time. Exploring these sequences with topic modeling is a relatively new and evolving research area. We use topic modeling to analyze football games, as an example of complex and under-explored spatio-temporal event data. A key challenge in topic modeling is selecting the most suitable number of topics for the downstream application. Selecting too few topics oversimplifies the data, merging distinct patterns, whereas selecting too many can fragment coherent themes into overlapping categories. We propose a visual analytics technique that uses dimensionality reduction on topics derived from multiple topic modeling runs, each with a different number of topics. Our technique organizes all the topics in a hierarchical layout based on their spatial similarity, making it easier to make an informed decision about selecting the most expressive set of topics that represent distinctive spatial patterns. We apply our visual analytics technique to a football dataset, illustrating how it can be used to select an appropriate set of topics for this data. We then use these topics to represent game episodes, which help us summarize game dynamics and uncover insights into the games.

This paper introduces the Large Events Model (LEM) for soccer, a novel deep learning framework for generating and analyzing soccer matches. The framework can simulate games from a given game state, with its primary output being the ensuing probabilities and events from multiple simulations. These can provide insights into match dynamics and underlying mechanisms. We discuss the framework’s design, features, and methodologies, including model optimization, data processing, and evaluation techniques. The models within this framework are developed to predict specific aspects of soccer events, such as event type, success likelihood, and further details. In an applied context, we showcase the estimation of xP+, a metric estimating a player’s contribution to the team’s points earned. This work ultimately enhances the field of sports event prediction and practical applications and emphasizes the potential for this kind of method.

In soccer, player scouting aims to find players suitable for a team to increase the winning chance in future matches. To scout suitable players, coaches and analysts need to consider whether the players will perform well in a new team, which is hard to learn directly from their historical performances. Match simulation methods have been introduced to scout players by estimating their expected contributions to a new team. However, they usually focus on the simulation of match results and hardly support interactive analysis to navigate potential target players and compare them in fine-grained simulated behaviors. In this work, we propose a visual analytics method to assist soccer player scouting based on match simulation. We construct a two-level match simulation framework for estimating both match results and player behaviors when a player comes to a new team. Based on the framework, we develop a visual analytics system, Team-Scouter, to facilitate the simulative-based soccer player scouting process through player navigation, comparison, and investigation. With our system, coaches and analysts can find potential players suitable for the team and compare them on historical and expected performances. For an in-depth investigation of the players’ expected performances, the system provides a visual comparison between the simulated behaviors of the player and the actual ones. The usefulness and effectiveness of the system are demonstrated by two case studies on a real-world dataset and an expert interview.

One promising application of data analytics in soccer is the identification of groups of players with similar playing styles. This can enhance recruitment processes and assist in tactical decisions, such as substitutions during a match. Previous research has explored various methods to measure a player’s contribution, identify key variables for player positions, and predict market value using machine learning techniques. However, these efforts often focus on traditional player positions or a limited set of gameplay characteristics, lacking a comprehensive analysis of player participation. This paper introduces a novel approach to clustering similar soccer players using supervised clustering and dimensionality reduction. Unlike previous studies, our method aims to refine the well-known positions of players and incorporates a wide range of in-game metrics to provide a holistic view of each player’s playing style.

Soccer attracts the attention of many researchers and professionals in the sports industry. Therefore, the incorporation of science into the sport is constantly growing, with increasing investments in performance analysis and sports prediction industries. This study aims to (i) highlight the use of complex networks as an alternative tool for predicting soccer match outcomes, and (ii) show how the combination of structural analysis of passing networks with match statistical data can provide deeper insights into the game patterns and strategies used by teams. In order to do so, complex network metrics and match statistics were used to build machine learning models that predict the wins and losses of soccer teams in different leagues. The results showed that models based on passing networks were as effective as “traditional” models, which use general match statistics. Another finding was that by combining both approaches, more accurate models were obtained than when they were used separately, demonstrating that the fusion of such approaches can offer a deeper understanding of game patterns, allowing the comprehension of tactics employed by teams relationships between players, their positions, and interactions during matches. It is worth mentioning that both network metrics and match statistics were important and impactful for the mixed model. Furthermore, the use of networks with a lower granularity of temporal evolution (such as creating a network for each half of the match) performed better than a single network for the entire game.

Predictive modeling plays a crucial role in machine learning, data analysis, and statistics. In sports, predictive modeling methods have emerged to provide insights and evaluate performances based on key performance metrics. However, most existing models tend to focus on predicting only partial aspects of an event, such as the outcome, action type, or location, while neglecting the temporal factors involved. To address this gap, this study introduces the Transformer-Based Neural Marked Spatio-Temporal Point Process (NMSTPP) model, specifically designed for football event data. The NMSTPP model predicts a comprehensive set of future event components, including inter-event time, zone, and action. Additionally, it features a dependent prediction layers architecture to enhance model performance. The Holistic Possession Utilization Score (HPUS) metric is also proposed to evaluate the effectiveness and efficiency of possession periods in football based on the NMSTPP model. With open-source football event data, the NMSTPP model successfully predicted the aforementioned three components of future events, with an improvement of up to 4% overall and 9% for individual components compared to baseline models. The HPUS demonstrated a 0.9 correlation with existing performance metrics, highlighting its utility in performance evaluation. The NMSTPP and HPUS were applied to the Premier League to demonstrate their practical feasibility.

This paper presents a novel framework for evaluating association football (soccer) players using possession sequences to estimate player strengths. The methodology is similar to classical plus-minus rating models, primarily employing regularized regression techniques. By analyzing possessions, our framework distinguishes on-ball and off-ball contributions of players to the game. The framework explores four different penalization schemes, which exploit soccer-specific structures such as the grouping of players into position groups as well as into common strength groups. These four models lead to four ways to rate players by considering the respective estimate of each model corresponding to the player. The ratings are used to analyze the 2017/18 season of the Spanish La Liga. We compare similarities as well as particular use cases of each of the penalized models and provide guidance for practitioners when using the individual model specifications. Finally, we conclude our analysis by providing a domain-specific statistical evaluation framework, which highlights the potential of the penalized regression approaches for evaluating players.