Team-Scouter: Simulative Visual Analytics of Soccer Player Scouting

Anqi Cao, Xiao Xie, Runjin Zhang, Yuxin Tian, Mu Fan, Hui Zhang, and Yingcai Wu

Abstract—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.

Index Terms—Soccer Visualization, Player Scouting, Design Study

1 Introduction

In soccer, player scouting is essential for recruiting high-quality players to strengthen the team [3]. Especially in professional clubs, successful player scouting can help obtain suitable and well-performed players at low costs, while failed ones will lead to expensive but ineffective player transfers [22]. Various cases among top soccer clubs have witnessed the importance of player scouting. During the pre-half season of 2023/24, several teams without abundant transfer budgets and multiple star players, such as Leverkusen in *German Bundesliga* and Girona in *Spanish La Liga*, have ranked at the top of the leagues mainly because of their wise player scouting strategies and appropriate team tactics [55].

Scouting proper players for a team is a complicated task. Coaches and analysts need to compare huge amounts of players on diverse performance indicators to assess whether the players will be suitable for the team [3]. Numerous studies have been proposed to evaluate player performance in different aspects [11, 28, 44]. Although useful, it is difficult to know how the players will perform in a new team directly from historical match data, which is one of the most significant reasons for unsuccessful player recruitments [3, 22]. Thus, simulative analysis has been introduced to the process of soccer player scouting to estimate the players' performance in a new team [22, 33]. However, those methods solely provide the results of the expected increment of team performance indicators with complex mathematical models. Lacking interactive navigation tools, experts often experience difficulties in thoroughly exploring the large space of potential target players. Moreover, without the knowledge of the computational process of the models, experts encounter challenges in understanding how the players can coordinate with team tactics to release their potential. Although various interactive visualization systems have been developed to support an in-depth investigation of soccer player performance [10, 21, 47], visual tools supporting simulative-based player scouting are still rare.

In this study, we have worked closely with professional soccer analysts and coaches to provide a simulative visual analytics method for soccer player scouting. During the cooperation, we tackled two major challenges. The first challenge is to provide fine-grained simulation for player behaviors in a new team. When scouting players, experts need to know how the players will behave in matches within a new team to evaluate whether they are suitable for the team's tactics and understand how to effectively arrange them in future matches. Existing studies mostly simulate the summarized statistical indicators if the players join the team [22, 33], which is insufficient for experts to obtain a comprehensive view of players from their possible behaviors. Player behaviors in matches are related to multiple heterogeneous features, such as the player decision styles and the current team tactics [52]. It is difficult to extract and integrate those complex features into the simulation of player behaviors. The second challenge is to develop effective visual tools to support player navigation and comparison. To identify target players for further investigation, experts need to explore the whole player space from various aspects of their performances. However, the target players might come from any other club and even league, which usually overwhelms experts in this navigation process. Furthermore, it is non-trivial to compare potential target players over multiple spatio-temporal dimensions, such as the simulated behaviors and their results under diverse team tactics, to decide who is the most suitable for the team. Effective tools should also be provided for visual comparison of target players on such various complex aspects.

To solve the first challenge, we construct a soccer player action simulation model with the consideration of player and team styles. Based on the model, we develop a match simulation framework incorporated with match result simulation and player behavior evaluation, to estimate the expected contributions of the players on both match results and player behaviors. To solve the second challenge, we design a visual analytics system, Team-Scouter, to support systematic soccer player scouting by simulative analysis. Users can navigate the possible space of target players and compare their expected contributions to match results by the navigation view. The investigation view supports comparing players on simulated behaviors to decide the most suitable player.

The main contributions of this work are as follows:

- A **characterization of domain problems** of soccer player scouting, including the summarization of requirements on player navigation, comparison, and investigation.
- A **match simulation framework** that supports the simulation of match results and player behaviors to help with player scouting.
- A **visual analytics system** to support player scouting through navigating players and comparing them on simulated behaviors.

- *A. Cao, R. Zhang, Y. Tian, and Y. Wu are with the State Key Lab of CAD&CG, Zhejiang University. E-mail: {caoanqi, runjinzhang, yuxintian, ycwu}@zju.edu.cn.*
- *X. Xie, M. Fan, and H. Zhang are with the Department of Sports Science, Zhejiang University. E-mail: {xxie, fanmu_032, zhang_hui}@zju.edu.cn.*
- *X. Xie is the corresponding author.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

2 Related Work

2.1 Soccer Player Scouting

Player scouting is significant for soccer teams and has attracted extensive research attention during recent years [3]. Traditional methods for soccer player scouting mostly quantify player performance on multiple aspects and focus on those highly-ranked players. For instance, Brooks et al. [7] and Pappalardo et al. [34] presented player ranking methods based on the weighted sum of player performance features to discover the players who contribute most to their teams. Decroos et al. [11, 13] and Bransen et al. [4–6] concentrated on player contributions through each fine-grained action and ranked players by aggregated action values on different types. Besides, various criteria have also been proposed for evaluating player performance, such as threat prevention [28] and passing creativity [44]. However, the practical player scouting process is more complicated than simply selecting the most well-performed players, usually with constraints such as player roles and team transfer budgets [3]. Thus, a series of work formulated soccer player scouting as a multi-criteria decision-making (MCDM) problem and solved the most suitable players for a team by optimizing the summarization of the players' performance indicators under multiple constraints [31, 36, 49].

Soccer is a highly dynamic sport, and a player will perform differently when coming to a new team because of the different match contexts related to team tactics [3, 22]. Such uncertainty of player performance often leads to unsuccessful player recruitment decisions, like expensive and well-performed players being unable to adapt to unfamiliar team tactics in their new teams. However, the difference between the player's performance in the current and new teams was largely ignored by the studies mentioned above. To estimate player performance in a new team, Jarvandi et al. [22] pioneered a study to introduce simulative analysis into soccer player scouting. They modeled soccer matches as a semi-Markov decision process to simulate the goal differential after a new player has been added to the team, and selected the players who would contribute high goal differentials as the most compatible with the team. Similarly, Pantuso [33] formalized soccer player scouting as a stochastic programming problem and simulated the total market value of the team after adding new players to discover the players with potential in the financial aspects.

These studies provided valuable insights for scouting well-performed players for a soccer team. Nevertheless, they seldom integrate visual tools to help navigate potential target players and compare the players' fine-grained simulated behaviors under the complex diverse team tactics. Our work is positioned as a simulative visual analytics method to support systematic and comprehensive soccer player scouting.

2.2 Visualization of Soccer Data

Visualization tools have been widely integrated into the in-depth analysis of various sports data [16, 40], including basketball [19, 20, 26, 60, 62], baseball [14, 23, 30, 32], racket sports [24, 42, 59, 61, 63, 67], and other events [25, 41]. In particular, numerous studies focus on visualizing soccer match data for different analysis tasks. For team ranking analysis, Perin et al. designed visualizations called A table! [39] and Gap Charts [37] to illustrate the temporal evolution of the ranking table and team scores during the tournaments. For team tactic analysis, current studies mainly concentrate on the discovery of player passing patterns and player movement patterns [40]. In the category of analyzing player passing patterns, SoccerStories [38] provided a group of diagrams to visualize player passing patterns within similar possessions. PassVizor [65] revealed the dynamics of player passing patterns by visualizing the occurrence of patterns throughout a match with a pattern flow. On the aspect of player movement patterns, one of the most popular methods is aggregating player trajectories and detecting the coordination among them, as seen in the studies of Shao et al. [51], Sacha et al. [48], Andrienko et al. [1], and Seebacher et al. [50]. Besides, the player movement patterns can also be directly discovered from match videos with the help of embedded visualizations on the players and their trajectories [53, 54]. Several studies also focused on high-level movement patterns with tactical semantics like team formation [27, 64] and defensive pressure [2]. For tactical decision-making, interactive visual analytics systems have been proposed for important decision-making tasks in soccer, such as lineup selection [8] and player migration [9].

Although these visualization studies can help analyze soccer match data effectively, they can hardly be applied to soccer player scouting due to the different analysis tasks. To be mentioned, player performance evaluation is one of the most widely adopted methods for soccer player scouting, which has also attracted the attention of recent visualization studies [40]. For instance, Rusu et al. [45, 46] and Ryoo et al. [47] provided metaphor-based and pixel-based visualizations, respectively, to compare statistical performance indicators among multiple players. Janetzko et al. [21] developed a visual analytics system to evaluate players by their performance features under different match contexts. Action-Evaluator [10] visualized the players' performance based on their detailed actions conducted in matches. However, it is difficult to utilize these studies directly for soccer player scouting because the navigation of target players and the comparison of players' behaviors under different team tactics have not been supported. Thus, we provide a visual analytics system incorporated with a match simulation framework to help navigation and comparison in soccer player scouting.

3 Background

In this section, we introduce the background of soccer player scouting and characterize the problem through interviews with domain experts.

3.1 Background and Domain Concepts

Soccer is a team sport involving eleven players for both teams, and each team aims to score more goals than its opponent to win the match. The eleven players appearing in a match are selected by coaches from the squad and contain different roles. Our work focuses on scouting suitable new players to strengthen the squad of a soccer team.

Soccer players. A *squad* includes the players who can be selected by coaches to appear in a match. Besides, coaches usually need to find new players to join the squad to extend their player choices in future matches [3]. Each player in or is a candidate for the squad is usually described by the following basic attributes: the role, the birthdate, the market value, the current club, the height, and the weight (Table 1).

Soccer match structure. An *action* is a match event that records when a player interacts with the ball. It is the fundamental analysis element of players' on-ball decisions [11]. Such an aspect is mostly considered by coaches in player scouting because it directly influences their tactical arrangements for the team [3]. Each action is defined by a tuple with six event attributes: the acting player, the time, the start location, the end location, the action type, and the result (Table 1).

A *possession* is a sequence of consecutive actions in which the ball is controlled by the same team [12]. It describes the interactions with the ball at the team level, such as regaining the ball, building up, and finishing the attack. A possession is denoted by a list of actions and can be further divided into defensive and offensive phases based on the intentions of the actions. We use the definition of the defensive phase as the actions from the start of possession to the completion of the first pass and the offensive phase as the following actions [43].

A *tactic* is a player behavior pattern frequently occurring in matches. We focus on *on-ball tactic*, one of the most analyzed tactics, because it significantly affects coaches' preferences in selecting players [3]. An on-ball tactic is an action pattern reflecting how the team tends to deal with the ball. It is indicated by a representative frequent sequential pattern of similar possessions [12]. Following the division of defensive and offensive phases, the on-ball tactics can also be divided into defensive and offensive ones as the sequential pattern in each phase.

The problem of soccer player scouting. Player scouting means finding potential players from outside a team to provide more player choices for coaches. Based on the scouting results, the staff members would design strategies for player recruitment and transfer. We work on player scouting for professional soccer clubs, which is the most complicated scenario due to the large scope of candidate players [3]. In the player scouting process, analysts and coaches need to consider players' performance and physical attributes as well as other important factors such as financial conditions, player personality, and player reliability. We mainly focus on the preliminary phase of player scouting, which usually considers players' in-match performance and basic attributes.

3.2 Interview

We worked closely with three experts for one year to develop a visual analytics system for soccer player scouting. The experts were treated as collaborators, including a coach with years of experience in professional soccer clubs (EA), a senior sports data analyst familiar with soccer player performance analysis (EB), and a doctoral student in sports science who was a professional player of a top soccer league (EC).

Domain problem characterization. To understand the problem of soccer player scouting, we surveyed relevant literature and held a series of meetings with our experts. The meetings focused on summarizing the workflow of player scouting from the experts' experiences. We recorded the discussion process and summarized the analysis workflow after the meetings. According to the experts, they usually find players by analyzing their performance on various criteria, such as statistical indicators [7, 34] and fine-grained action scores [11], and select potential target players based on their tactical preferences. Then, we interviewed the experts individually to learn the challenges they encountered in player scouting. We also recorded the answers of each expert during the interviews. All the experts mentioned that the most significant challenge they faced is the uncertainty of player performance in a new team. For instance, a player who performed well in the original team might not achieve the coaches' expectations in a new team, which leads to large numbers of transfer failures with high costs. EC explained that such performance changes on players mainly because of the difficulty of adapting to new tactics. Besides, EA and EB pointed out that it is laborious to analyze and compare large-scale players from hundreds of clubs in numerous leagues to find suitable ones for the team. Based on the challenges, we confirmed and revised the analysis workflow and summarized the initial design requirements for our system.

Model development and visual design iteration. We held weekly meetings with our experts during system development and iteration. We recorded the discussion process in the meetings and iterated the model and visual design according to the experts' comments after each meeting. In model development, we surveyed models for soccer player scouting [11, 22, 33, 44] and discussed with the experts which model would be effective in finding suitable players. The experts mentioned that match simulation methods [22, 33] are helpful because they support estimating team performance with new players, therefore reducing transfer failures. Besides, EB stated that a fine-grained simulation of player behaviors under different tactics would help coaches arrange appropriate tasks for new players. Thus, we decided to extend the match simulation framework to support simulation for both team performance and player behaviors. After model development, we designed a system prototype that supported recommending players by simulated match results to efficiently find suitable players from a large scale of candidates and displaying players' simulated action decisions by on-ball tactics of a new team. EA suggested that finding new players similar to the players in the squad is important because coaches usually tend to maintain the consistency of the tactical system for the team. EB also commented that comparing simulated action decisions among the players is also required to decide the most suitable players based on tactical preferences. According to the experts' comments, the design requirements were also updated along with the visual design iteration.

3.3 Requirement Analysis

Based on the interview results, we have developed six design requirements for soccer player scouting at three levels.

Team-level navigation aims to help experts find and identify the part of the team squad that needs to be strengthened by new players.

N1 *Supporting the overview of the team squad.* The squad of a soccer team usually contains several players for each role to ensure adequate player choice for coaches. Thus, experts tend to focus on the thin part of the squad and find new players to strengthen it. The overview should present important players for each role in the squad to help identify the roles of players that need to be scouted.

N2 *Supporting the illustration of player characteristics in the team squad.* A new player is expected to replace the original players in the squad in future matches. Player characteristics, such as their importance in the team and market values, are useful for experts to indicate which player might be replaced by a new player. Based on this information, experts can find potential target players by the change in team performance if they replace a certain player.

Player-level comparison can help experts obtain information on recommended new players to distinguish the players suitable for the team.

C1 *Supporting the comparison of essential personal information of potential target players.* After indicating the player to be replaced in the squad, the match simulation framework supports recommending a series of players based on coaches' requirements, such as player performance or similarity. Comparing the essential personal information of these players, including their ages and market values, is important in the decision of target players because experts prefer players with proper ages and reasonable market values.

C2 *Supporting the comparison of the historical and expected performances of potential target players.* Players' performances in historical matches are vital references for experts to scout the target players for the team. Besides, experts also need to compare their expected contributions to the new team to estimate whether they are suitable for the team's tactical system. Effective tools should be provided for such comparison of players among multiple criteria.

Action-level investigation can assist in exploring detailed simulated actions to decide the most suitable player based on tactical preferences.

I1 *Supporting the exploration of player's simulated actions by on-ball tactics.* Experts might select several potential target players by different criteria and decide the most suitable one according to their tactical preferences. For instance, a coach will likely choose a player who can perform well in the most adopted tactics. It is necessary to support the exploration of players' simulated action decisions and expected action scores under on-ball tactics of interest.

I2 *Supporting the comparison of player's simulated actions to actual ones.* When experts decide on the most suitable player, they demand to inspect the differences in action decisions between the new player and the player who might be replaced. It can help coaches understand how the player will integrate into the new tactics to arrange proper tasks. Such differences also reveal the effect of the new player and provide guidance on adjusting tactics accordingly.

3.4 Data Processing

The data used in our analysis is from an open-sourced soccer event dataset published by Pappalardo et al. [35], including all matches in the 2017/18 season of the five top European leagues: *English Premier League*, *Spainish La Liga*, *German Bundesliga*, *Italian Serie A*, and *French Ligue 1*. The dataset contains both player information and match events. In detail, each player is described by several physical attributes, and each event is characterized by spatio-temporal attributes and categorical attributes. As market values and fine-grained roles of players are also important in player scouting, we collected them from Transfermarkt.com [55], one of the most reliable websites for soccer player transfer information. The essential player attributes and event attributes are presented in Table 1. In total, the dataset contains 98 teams, 1,224 matches, 3,074 players, and 3,071,395 events.

We preprocessed the data in two steps: on-ball tactic identification and feature construction. In on-ball tactic identification, we referred to the previous work [12] to divide the event data stream into possessions, calculate distances among the possessions to cluster them into different groups, and identify the frequent sequential pattern to represent the on-ball tactic. In feature construction, we calculated essential features for each action by their attributes to train the match simulation models.

Table 1: Player and event attributes

Player Attributes	
Role	The role of the player (*i.e., goalkeeper, left-back, central midfield, center-forward, etc.*).
Current Club	Current club that the player belongs to.
Market Value	The estimated market value of the player.
Physical Attributes	Height, weight, and birthdate of the player.
Event Attributes	
Player	Player who interacts with the ball.
Event Type	Technique that the player used to deal with the ball (*i.e., pass, dribble, cross, shot, etc.*).
Event Result	Result of the event (*i.e., succeeded or failed*).
Time	The time when the event occurred in match.
Event Locations	The start and end coordinates of the event.

4 System Overview

Team-Scouter is a web-based application that consists of three components: data processing, match simulation framework, and visualization interface (Fig. 1). The data processing constructs action features and labels for training models in the match simulation framework and discovers the on-ball tactics used by each team. The match simulation framework supports simulating both match results with new players and the detailed action decisions of the new players when coming to the new team. The visualization interface includes a navigation view and an investigation view to facilitate simulative analysis for player scouting. The data processing, match simulation framework, and visualization interface are implemented by Python, PyTorch, and React, respectively.

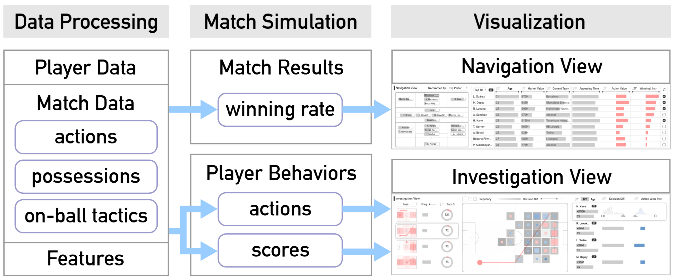

Fig. 1: The system overview. The system consists of three components: data processing, match simulation framework, and visualization interface.

5 The Match Simulation Framework

In this section, we define the task of soccer match simulation and provide a framework for both match results and player behavior simulation.

5.1 Task Definition and Framework Overview

The main task of soccer player scouting is to find players who can improve the team's performance. Specifically, for a set of candidate players $P = \{p_1, \ldots, p_n\}$, the scouting process aims to rank the players in P by estimating their contributions to the team. A series of methods have been proposed to assist player scouting, including analyzing historical performance [11, 44] and simulating performance in future matches [22, 33]. We choose the match simulation method because it directly predicts whether a player will perform well in a new team, which is practical to reduce transfer failures caused by unsuitable players.

The match simulation method estimates the change of a team performance indicator I if a new player p_i replaces an original player p_j due to the fixed amounts of players in a soccer match. Existing models mainly focus on estimating match-level performance indicators such as match results [22, 33]. However, fine-grained simulation of players' behaviors is also demanded by coaches because it is vital to understand how the players will take part in the tactics of a new team to arrange proper tasks for them in future matches. Thus, we extend the match simulation method to behavior-level performance indicators.

We build a two-level match simulation framework for estimating both match results and player behaviors (Fig. 2). The framework includes three components: player action prediction (Fig. 2(A)), match result prediction (Fig. 2(B)), and action score estimation (Fig. 2(C)). The player action prediction component predicts the player's action given the match state, and provides the learned player features to the match result prediction component to simulate the match result. The action score estimation component converts the player's action into team performance indicators to value the player's expected performance.

5.2 Player Action Prediction

The player action prediction component supports the behavior-level simulation and extracts player features for the match-level simulation. We focus on simulating player action because it is one of the most critical player behaviors and directly affects match development [11]. The component takes the match state, the acting player, and the player's team as input and provides the player's action as output (Fig. 3).

Match state construction. The notation of match state S_i at action a_i is a subsequence of all actions in the match $S_i = \{a_1, \ldots, a_i\}$. As a player's action is more affected by nearly occurred actions than previous ones [11], we consider the most near k actions. Thus, match state can be simplified as $S_i = \{a_{i-k+1}, \ldots, a_i\}$, and k is a hyperparameter for the number of previous actions that would affect the action. We set $k = 10$ based on the previous work [52] and domain knowledge. The previous work on player action prediction [52] provides several options for k, including 100, 40, 10, and 5. We focus on values no larger than 10 because players' actions are usually related to the previous actions within the same possession, and most of them are shorter than 10 [12]. However, the experiment results provided by the previous work indicate that models with $k = 5$ will perform poorly [52]. Thus, we choose $k = 10$ in our model. We apply a fixed hyperparameter setting in our system to reduce the effort of users in hyperparameter searching.

For each action a_j in the state S_i, we construct action feature vector AF_j by basic action features, relevant action features, and match context features to include the most essential features that affect player actions [11]. The basic action features are the start and end locations, the type, the occurring time, and the acting team of the action. The relevant action features combine features among consecutive actions, consisting of the distances and angles to the goal and from the former action and the time elapsed from the former action. The match context feature is the score difference to the opponent team when the action occurred.

Player and team representations. We consider the acting player and team in action prediction because both the players' playing style and the teams' tactic style will influence the players' action decisions [66]. To extract player and team features, we integrate representation learning [29] into the prediction of player actions. It transforms the one-hot encoding of player p_i and team t_i to embedding vectors EP_i and ET_i, respectively. The embedding vectors will be learned through the backpropagation of the loss in the model training process to capture the similarities among players and teams in predicting action decisions. In this method, the playing style of players and the tactic style of teams can be represented by these learned embedding vectors.

The sequential learning model. We use a sequential learning model to predict the player's action a_{i+1} by the match state S_i constructed by a sequence of previous actions [52]. In detail, the model contains the representation learning of the player and the team to extract their embedding vectors EP_i and ET_i and a Transformer network [56] to extract the feature vector of the match state SF_i. We adopt the Transformer network for sequential learning because of the less training time on the large soccer event dataset. The concatenation of these vectors will be used in the prediction of the player decision on action a_{i+1}, including the coordinates of the end location end_{i+1} and the type $type_{i+1}$.

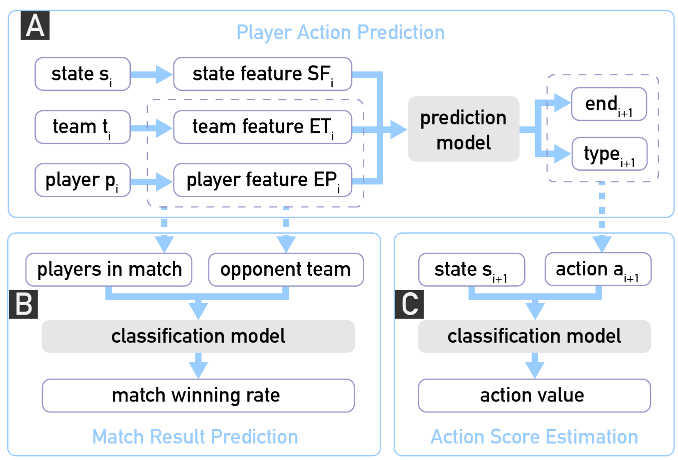

Fig. 2: The structure of the match simulation framework. The match simulation framework contains three components: player action prediction (A), match result prediction (B), and action score estimation (C).

5.3 Match Result Prediction

This component estimates the match results with the players in the match and the opponent team for match-level simulation based on the player and team vectors. We measure the expected match result by winning probability and predict it by a random forest classifier with the embedding vectors of the players and the opponent team, as well as the player attributes and total action values as input. Thus, the expected contribution of the new player p_j on the match level can be estimated by the increment of winning probability:

$$C_{match}(p_j) = I_{match}(P_j) - I_{match}(P_i). \tag{1}$$

where P_j denotes the player set in a match including the new player p_j and P_i refers to that containing the original player p_i, and $I_{match}(P_j)$ and $I_{match}(P_i)$ indicate the winning probability with P_j or P_i, respectively.

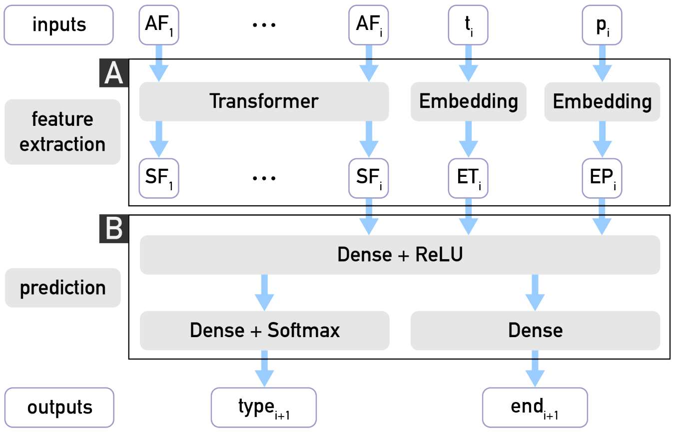

Fig. 3: The architecture of the player action prediction model. The model includes two steps: feature extraction (A) and result prediction (B).

5.4 Action Score Estimation

With the predicted player actions, the action score estimation component provides a score for each action to evaluate the player's expected contribution to the team on the behavior level. We use the definition of the action score as the sum of the increment of goal probability for the own team and the decrement of goal probability for the opponent team [11]. We utilize the previous work [11] and train two CatBoost classifiers to predict the goal probabilities of the two teams in the near future with the current match state as input. Based on the action score estimation, the expected contribution of the new player p_j on the behavior level can be calculated as follows:

$$C_{behavior}(p_j) = I_{action}(a(p_j,S)) - I_{action}(a(p_i,S)), \tag{2}$$

where $a(p_j,S)$ and $a(p_i,S)$ are the simulated action choices of the new player p_j and the original player p_i on the current match state S, and $I_{action}(a(p_j,S))$ and $I_{action}(a(p_i,S))$ refer to the action scores.

5.5 Model Evaluation

We evaluate the two newly constructed components in our match simulation framework, including player action prediction and match result prediction, with two experiments, respectively. In the experiments, we randomly select 30% of the data as test data and others as training data.

The first experiment is to evaluate the prediction performance of player action prediction. We use standard metrics for evaluating prediction performance, including root-mean-square errors (RMSE) for the action's end location, and precision and F1-score for the action's type. We compare our model to a baseline model without the representation learning component. The results (Table 2) show that all three metrics of our model are better than the baseline model. It indicates that the representation learning component can improve the prediction performance of the model and extract the action decision features for the players and teams through the learning process. **The second experiment** is to evaluate match result prediction with precision and F1-score (Table 2). The results indicate that the embedding vectors for players and teams are effective in predicting match results for each league in our dataset.

Table 2: The model evaluation results

Player Action Prediction			
	RMSE_end	precision_type	F1-score_type
baseline	0.17	0.83	0.51
our model	**0.16**	**0.86**	**0.69**

Match Result Prediction					
	England	France	Germany	Italy	Spain
precision	0.69	0.71	0.85	0.71	0.75
F1-score	0.66	0.68	0.84	0.63	0.68

6 Visual Design

Team-Scouter contains two views: a navigation view for team-level navigation (**N1**, **N2**) and player-level comparison (**C1**, **C2**) (Fig. 4(A)), and an investigation view for action-level investigation (**I1**, **I2**) (Fig. 4(B)). In the navigation view, users can observe the role of players that need to be scouted (**N1**) and identify the player who might be replaced by a new player (**N2**) with the squad board (Fig. 4(A1)). After specifying the scouting metric, the player ranking list will present a series of automatically recommended players (Fig. 4(A2)). Users can compare those players on multiple criteria, such as essential personal information (**C1**), historical performance, and expected match result (**C2**). During the comparison, users can add potential target players to the investigation view. In the investigation view, users can find the on-ball tactic of interest by the on-ball tactic list (Fig. 4(B1)), and compare the expected action scores of the players in the player record list to decide the most suitable player (**I1**) (Fig. 4(B2)). They can further investigate the difference between the simulated actions and historical ones to understand the effect of the player on the team's tactical system (**I2**) (Fig. 4(B3)). We use red and blue to encode the positive and negative values of player performance throughout the whole user interface.

6.1 Navigation View

The navigation view (Fig. 4(A)) consists of a squad board to indicate the player will be replaced as the reference of player recommendation (**N1**, **N2**) and a player ranking list to compare the recommended players on multiple criteria to scout the potential target players (**C1**, **C2**).

Squad board. The squad board presents all players in the squad by their roles (**N1**) (Fig. 4(A1)). The board is half of a soccer pitch and is divided into ten regions based on the spatial locations of different player roles on the pitch (i.e., goalkeeper, left-back, right-back, center-back, defensive midfielder, central midfielder, offensive midfielder, left winger, right winger, and center-forward). We place players with the same role together on the corresponding region of the board. Coaches and analysts usually identify the player who might be replaced by a new one according to their importance in the team and market values. Specifically, the new player might be scouted as a rotation or strengthen for an important player in the team, or as a replacement for a player who will leave the team with high market values or cannot appear in future matches. Thus, we rank the players with the same role descending by their appearance time to illustrate their importance and encode their market values as bar charts for effective comparison (**N2**).

Justification. We use a pitch-based board that places all players in the squad together rather than a ranking list because it directly shows the distribution of players by roles to help users decide the role of players that need to be scouted. Besides, such a representation is familiar to soccer coaches and analysts because it is widely used in the player scouting process as an overview of the squad.

Player ranking list. The player ranking list facilitates scouting players from large-scale candidates by player recommendation and presents the recommended players with a sortable list to assist the comparison among the players (Fig. 4(A2)). Each row of the list indicates a player and consists of a player name label, essential personal information (**C1**), and historical and expected performances (**C2**) (Fig. 4(A5)). The essential personal information includes the players' age, market value, and current team with its similarity to the analyzing team, the three most important player attributes influencing player scouting (Fig. 4(A3)). In detail, the players' ages represent their potential in the future, and their market values are constrained by the team's budget. The similarity of the player's current team to the analyzing team is useful for estimating the difficulty of the players in adapting to the new team. The historical performance contains the total action values per 90 minutes and the appearing time of the players, and the expected performance is the expected increment of the winning rate calculated by the match result prediction model (Fig. 4(A4)). All the metrics are encoded by bar charts to facilitate the comparison under multiple criteria. We also provide texts for the exact values of essential player information because they are familiar to users in describing players on these aspects.

Interaction. The interaction of the navigation view is as follows.

- *Selecting replaced players.* Users can click players in the squad board to select the player as the one to be replaced.
- *Selecting scouting metric.* Users can select the scouting metric based on their requirements with the drop list at the top of

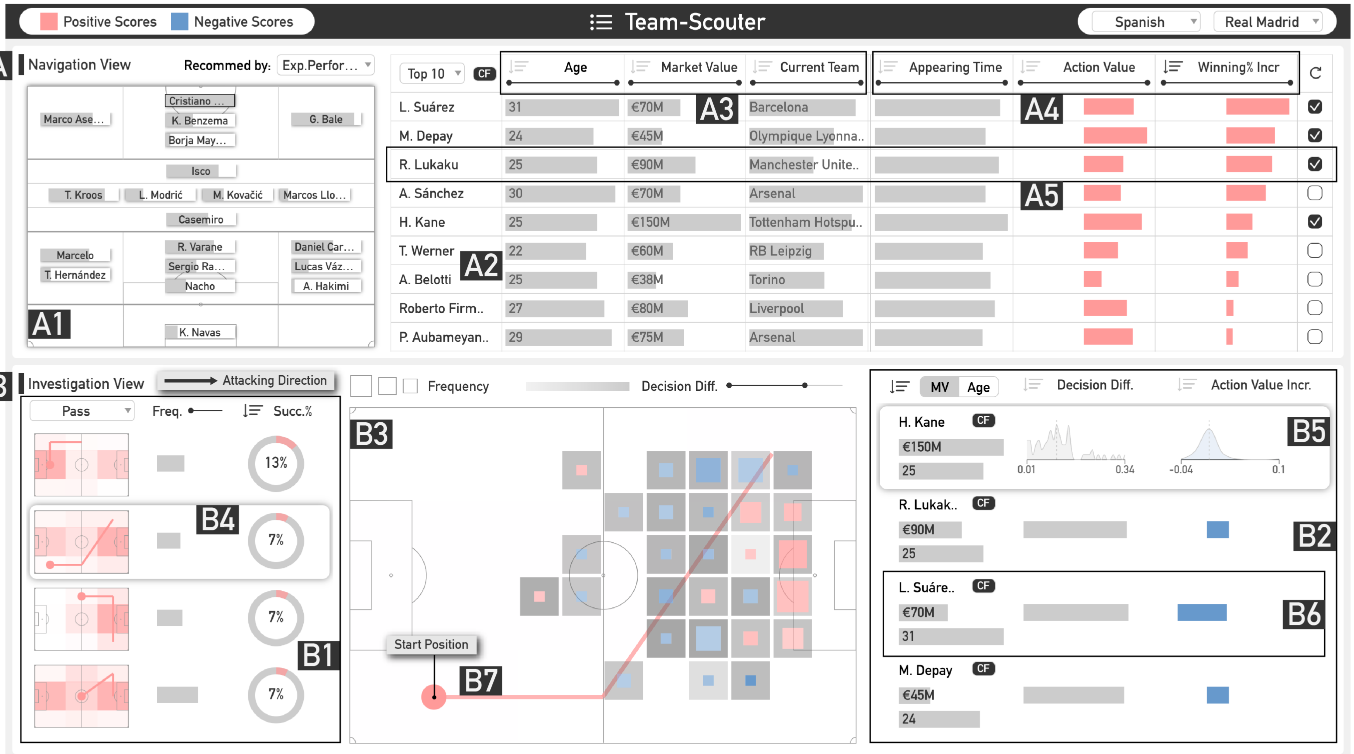

Fig. 4: System user interface. The interface contains two views: a navigation view (A) and an investigation view (B). The navigation view consists of a squad board (A1) to navigate players will be replaced and a player ranking list (A2) to compare players by personal information and performances. The investigation view includes an on-ball tactic list (B1) for exploring essential on-ball tactics, a player record list (B2) to compare players' simulated actions under a certain on-ball tactic, and a simulated action map (B3) to display players' detailed simulated actions.

the squad board. The metrics include similarity, historical performance, and expected performance. After the selection, the recommended players will be shown in the player ranking list.

- *Sorting and filtering.* Users can click the sorting buttons or drag the sliders at the top of the player ranking list to sort or filter the players by different player comparison criteria.
- *Selecting potential target players.* Users can click the checkbox at the right of a player item in the player ranking list to add the player to the investigation view as a potential target player.

6.2 Investigation View

The investigation view (Fig. 4(B)) includes an on-ball tactic list to select on-ball tactic of interest and a player record list to observe players' expected action-level contributions under the selected on-ball tactic (**I1**). The detailed spatial distribution and expected contributions of simulated actions for a player are presented in the simulated action map (**I2**). The pitch orientation in this view is attacking from left to right.

On-ball tactic list. The on-ball tactic list is a sortable list of all on-ball tactics of the analyzing team (Fig. 4(B1)). With the list, users can select an on-ball tactic of interest to analyze players' simulated actions within it (**I1**). Each row of the list represents an on-ball tactic, including an on-ball tactic map that indicates the on-ball tactic and its frequency and success rate (Fig. 4(B4)). The on-ball tactics can be filtered by categories because coaches and analysts often need to identify a certain category of tactics for analysis based on their tactical preferences. The categories contain offensive ones, such as simple pass, corner, and freekick, and defensive ones, such as tackle and interception.

The on-ball tactic map is a pitch-based visualization of on-ball tactic and consists of an abstract spatio-temporal trajectory and a heatmap (Fig. 5). An on-ball tactic is defined as a representative frequent sequential pattern of similar possessions [12]. In detail, each action in possession can be discretized by pitch regions, and the frequent sequential pattern is obtained by sequential pattern mining of the possessions after action discretization. Thus, an on-ball tactic can be represented by a sequence of actions discretized by pitch regions. We link the pitch regions as the abstract spatio-temporal trajectory to illustrate the on-ball tactic. The dot in the trajectory means the starting position. We use a heatmap to encode the total action values in each region within the on-ball tactic to help understand its effectiveness on the spatial aspect.

We use frequency and success rate to explore on-ball tactics because coaches and analysts need to scout players based on the strength and weakness of a team. The success rate is calculated by the number of possessions that end with a shot dividing the total number of possessions in the on-ball tactic. With the two metrics, users can identify the on-ball tactic to be focused on in the analysis of potential target players.

Justification. We choose a pitch-based representation of an on-ball tactic rather than a sequence of regions [65] because it keeps the spatio-temporal context of the development of the on-ball tactic. In the on-ball tactic map, we divide the pitch into nine regions, including the defensive third, the middle third, and the attacking third, based on the widely used pitch division method in soccer tactic analysis [18]. We use such a division method because it maintains the basic spatial features of on-ball tactics, such as defense, control, or attack, and left-wing, midway, or right-wing. We place the abstract spatio-temporal trajectory and the heatmap of an on-ball tactic together on the pitch to help users perceive the on-ball tactic and its effectiveness in the same concise view.

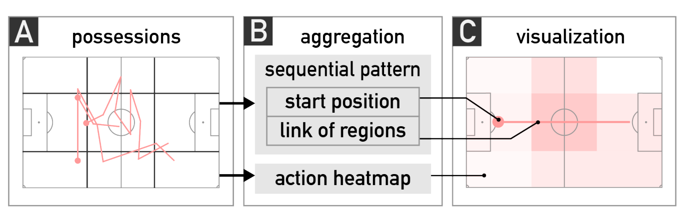

Fig. 5: The visualization of the on-ball tactic map. (A) presents a group of similar possessions. (B) shows the aggregation of the possessions. (C) is the visualization of an on-ball tactic with the aggregation results.

Player record list. The player record list records the potential target players to analyze their expected action-level contribution under certain on-ball tactics (**I1**) (Fig. 4(B2)). Each row in the list contains a player label, the average difference between the player's simulated actions and original ones, and the average action value increment (Fig. 4(B6)). These two metrics indicate a player's compatibility with on-ball tactics and the performance under them. The player label also contains the player's age and market value to assist users in deciding the target player. All the metrics are encoded by bar charts to facilitate comparison. The bar charts of the average difference and action value increment can also be unfolded into area charts to present their distribution (Fig. 4(B5)).

Simulated action map. The simulated action map illustrates the spatial distribution of simulated actions under a certain on-ball tactic (**I2**) (Fig. 4(B3)). We divide the pitch into 96 rectangular grids with 12 columns and 8 rows, and each grid is 8.75m in width and 8.5m in height. The simulated actions ended in the same grid are aggregated and visualized by a glyph. The glyph presents the comparison results between the simulated and the original actions, including the frequency of the simulated actions, the average action value increment, and the average difference compared to the original actions (Fig. 6(A)). The average difference is encoded by the saturation of the glyph background, where the lighter means the closer. The inner rectangle of the glyph represents the spatial distribution and effectiveness of the simulated actions. The frequency is encoded by the rectangle's area, and the average value increment is encoded by its saturation.

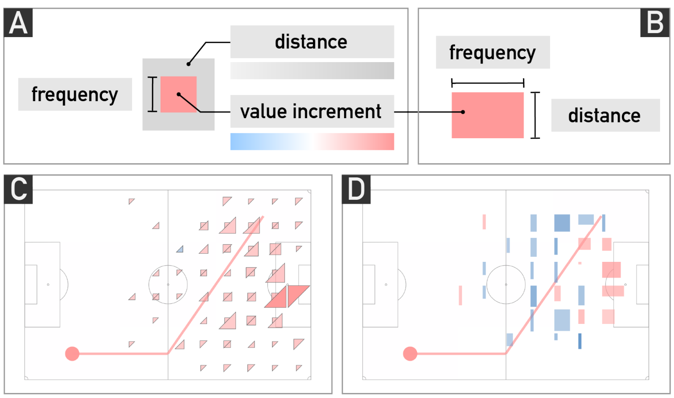

Fig. 6: The visualization of the simulated action map. (A) is the glyph encoding of the simulated action map. (B) is an alternative glyph encoding. (C) and (D) are two alternative maps with the individual display of actions and the alternative glyph encoding, respectively.

Justification. We illustrate the simulated and original actions in the same pitch rather than separately with other widely used visualizations for soccer action spatial distribution, such as heatmaps [17], for efficient comparison. We also provide two alternative visualizations for action comparison. The first is to display the simulated and original actions individually in the glyphs (Fig. 6(C)). It remains the information of the two kinds of actions and enables the comparison in the same grid. However, it falls short in the overview of the spatial distribution of the simulated actions and requires further calculation to obtain the comparison results. According to our experts, coaches and analysts focus more on the simulated actions of new players than the original ones for analyzing the new players' characteristics. Thus, we choose to directly visualize the spatial distribution of the simulated actions with their comparison results to the original ones. The second is to encode the frequency of the simulated actions and the comparison results in the same rectangle (Fig. 6(B, D)). However, the frequency and difference encoded by the height and width of the rectangle will affect each other. We thus choose the current design of the simulated action map.

Interaction. The interaction of the investigation view is as follows.

- *Sorting.* Users can click the sorting buttons at the top of on-ball tactic list and player record list to sort the items in the lists.
- *Filtering.* Users can filter the on-ball tactics by category through the drop list and by frequency with the slider at the top of the list. In the simulated action map, users can filter the simulated actions by the difference from the original ones by dragging the slider.
- *Selecting on-ball tactic.* Users can click the item in the on-ball tactic list to select the on-ball tactic, and the actions under it will be shown in the player record list and the simulated action map.
- *Selecting player.* Users can click the item in the player record list to inspect the player's actions in the simulated action map.

7 System Evaluation

In this section, we evaluate the system usability with two case studies conducted by two experts, respectively. After each case study, we also interview the expert in the same session to collect feedback.

7.1 Case Study

We invited two soccer experts who did not participate in the requirement analysis to conduct two case studies for our system. The first expert is a senior sports data analyst (EA), and the second is a professional soccer coach (EB). The case studies are based on the soccer event dataset used in our analysis [35] with all players and matches in the five top European leagues in the 2017/18 season. At the beginning of the case studies, we introduced the visual design and interactions in our system to the experts. Then, we demonstrated a typical analysis workflow and helped the experts get familiar with the system through free exploration. Afterward, the experts respectively tried to scout suitable players for a team of interest. During the case studies, we recorded the analysis process as well as the comments and valuable insights of each expert.

7.1.1 Case 1: Scouting Suitable Replacements of Cristiano Ronaldo for Real Madrid

The first case study is about exploring suitable players for Real Madrid who can replace Cristiano Ronaldo. We invited EA to conduct this case study. EA mentioned that C. Ronaldo, one of the most excellent center-forwards, left Real Madrid at the end of the 2017/18 season, which had attracted significant attention among soccer analysts. Thus, EA was interested in scouting players under such a scenario.

At first, EA clicked C. Ronaldo on the squad board and selected expected performance as the scouting metric to find players to improve the team's performance (Fig. 4(A1)). Then, EA moved to the player ranking list to inspect the top 10 players recommended by the system and found that most of the players met expectations, including outstanding center-forwards such as R. Lukaku from Manchester United and H. Kane from Tottenham Hotspur (Fig. 7(A)). EA noticed that L. Suárez, a center-forward from Barcelona, ranked at the top on winning rate increment (Fig. 7(A1)). Meanwhile, the current team of L. Suárez, Barcelona, was similar to Real Madrid, and the historical performance of L. Suárez, indicated by his action value, was also relatively high. Based on these findings, EA summarized that **L. Suárez might be the best replacement for C. Ronaldo in terms of increasing the team's performance** and added him to the investigation view for further analysis. Besides, EA found that M. Depay from Olympique Lyonnais, a center-forward with potential, ranked second in winning rate increment (Fig. 7(A3)). EA also mentioned that H. Kane and R. Lukaku ranked top on market value and might also be the target players of Real Madrid (Fig. 7(A2)). Thus, EA also added them to the investigation view.

After selecting the potential target players, EA turned to the investigation view to observe their detailed simulated actions by on-ball tactics. EA first analyzed the on-ball tactics of Real Madrid in the on-ball tactic list. Specifically, EA chose the simple pass tactics because they are one of the most important ways for center-forwards to create goal chances. Then, EA filtered out the on-ball tactics whose frequency was less than 0.05 and sorted them by success rate to focus on successful on-ball tactics of Real Madrid. EA noticed that the left-wing organization tactic was the most successful one of Real Madrid (Fig. 4(B1)). Thus, EA selected this on-ball tactic and moved to the player record list to compare the players' simulated actions. In the player record list, EA sorted the list by the average increment of action value and found that L. Suárez ranked at the top and his value was larger than zero, meaning that L. Suárez could conduct more valuable actions under the left-wing organization compared with C. Ronaldo and the other three potential target players (Fig. 7(B1)). EA also sorted the list by the difference in the action decision and found that the simulated actions of L. Suárez were the most different from those of C. Ronaldo (Fig. 7(B2)). It indicated that L. Suárez would improve the left-wing organization by different action decisions. EA further selected those simulated actions of L. Suárez and turned to the simulation action map to inspect the details. In the simulated action map, EA found that the simulated actions with higher values mainly ended in the midway (Fig. 7(C1)). Through this process, EA concluded that **L. Suárez could increase the winning rate for Real Madrid by improving the action choices under the left-wing organization tactic to transition to midway**.

However, the selection of L. Suárez might be unlikely to happen due to the historical rivalry between the two teams. Thus, EA was also interested in how the players would perform under other frequently used on-ball tactics to explore suitable players under different on-ball tactics. EA selected the other three frequently used on-ball tactics in the on-ball tactic list, respectively, and inspected the simulated actions of the players in the player record list. EA found that H. Kane would perform the most valuable actions under the right-wing attack and the left-wing attack tactics (Fig. 7(D1, D2)), while M. Depay would perform best under the left-wing transition from midfield. EA commented that it was reasonable because H. Kane is good at finishing attacks by creating goal chances, and M. Depay performs well in organizing attacks from the left-wing. EA further investigated the simulated actions of H. Kane under the left-wing attack because the average action value increment is larger than zero. In the simulation action map, EA noticed that the simulated actions with increased values were concentrated on the area near the goal and generally similar to the actions conducted in the historical matches (Fig. 7(E2)). It was also verified by the knowledge of EA that H. Kane could improve the attacking on-ball tactics by increasing the effectiveness of similar action choices in assists and goals due to his more excellent skills. Through this process, EA concluded that **H. Kane might also be a suitable replacement for C. Ronaldo if Real Madrid needs to find a center-forward for finishing attacking on-ball tactics such as left-wing attack and right-wing attack**.

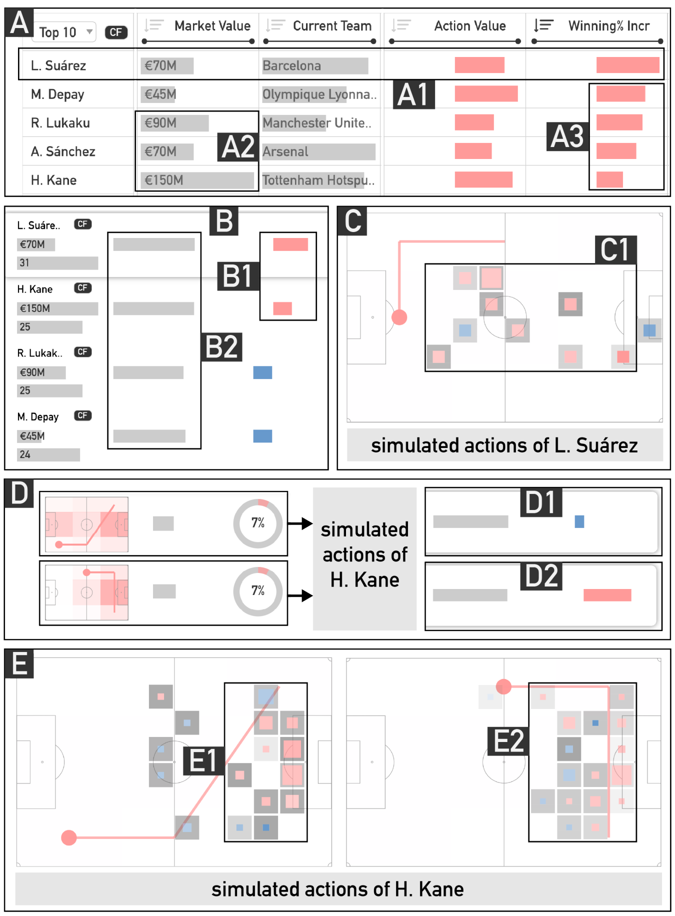

Fig. 7: The analysis process of case 1. (A) presents the recommended players based on C. Ronaldo. (B) presents simulated actions of the potential target players under the most successful on-ball tactic of Real Madrid. (C) presents the detailed simulated actions of L. Suárez. (D) and (E) are simulated actions of H. Kane under two different on-ball tactics.

7.1.2 Case 2: Scouting Suitable Improvements of Right-Wingers for Manchester City

The second case study is about finding suitable right-wingers as improvements for Manchester City. We invited EB to conduct this case study. As mentioned by EB, Manchester City won the champion of *English Premier League* in the 2017/18 season. Thus, EB was interested in scouting suitable players to strengthen the squad of Manchester City.

EB began with the squad board in the navigation view to identify the role of players to be scouted. EB focused on the forwards and found that the squad only contained one left-winger and two right-wingers (Fig. 8(A1)). EB commented that considering the tight match schedule of Manchester City with multiple tournaments, both roles required more players as the rotation for the original players. EB further concentrated on right-wingers and selected L. Sané as the player to be replaced because of his longer appearing time than P. Foden. EB also selected historical performance as the scouting metric to scout well-performed players. Then, EB switched to the player ranking list and filtered out the players who appeared less than 1000 minutes and whose market value was less than 10 million Euros to focus on the players with stable performances. EB sorted the list by expected performance increment and found three well-performed right-wingers ranked at the top of the list, including Á. di María from PSG, L. Messi from Barcelona, and R. Mahrez from Leicester City (Fig. 8(B1)). Based on these observations, EB summarized that **Á. di María, L. Messi, and R. Mahrez might be suitable right-wingers for Manchester City, according to both historical and expected performances**. Thus, EB added the three players to the investigation view for a detailed comparison.

In the investigation view, EB started from the on-ball tactic list and selected the two frequently used on-ball tactics of Manchester City on the right-wing, which right-wingers usually participate in (Fig. 8(C, D)). Based on the selection, EB observed the simulated actions of the three players in the player record list. EB found that under both on-ball tactics, the average action value increment of R. Mahrez was higher than the other two players, indicating that R. Mahrez would perform better than Á. di María and L. Messi under the tactics on the right-wing (Fig. 8(C2, D2)). Meanwhile, his difference in the action decision was also the highest among the three players under the two on-ball tactics. However, the action value increments of all players in the list were less than zero (Fig. 8(C3, D3)). EB explained that this might be because L. Sané was an excellent right-winger and more suitable to the tactics of Manchester City than any other right-winger from other teams. Based on the comparison results, EB summarized that R. Mahrez would perform better in the on-ball tactics on the right-wing than the other two players but might not be the most suitable to the tactics compared with the original player. To explain the detailed action choices of R. Mahrez, EB unfolded the simulated actions under the two on-ball tactics, respectively, to the simulated action map. In the simulated action map, EB noticed that the simulated actions of R. Mahrez were distributed both on the right-wing and the left-wing (Fig. 8(E)). EB explained that R. Mahrez could also play as a left-winger and conduct the on-ball tactics by switching between the right-wing and the left-wing to increase the defense difficulty for the opponent players. Besides, EB mentioned that, actually, R. Mahrez had transferred to Manchester City at the end of the 2017/18 season. Thus, EB compared age and market value among the three players and found that R. Mahrez was the youngest, and his market value was relatively reasonable, which could also explain the actual transfer decision. Through this process, EB concluded that **R. Mahrez could be the best target right-winger for Manchester City, and would perform flexible actions on both right-wing and left-wing in the right-wing attack tactics**.

7.2 Expert Interview

After each case study, we interviewed the expert who conducted this case to collect feedback. In each interview, we asked the expert several questions on system usability, including whether the system has fulfilled the requirements for soccer player scouting, whether the system improved the traditional workflow, and whether they had any suggestions to improve the system. The answers are summarized as follows.

System usability. Both experts are satisfied with our system and agree that it could fulfill their requirements for player scouting. EA thought highly of the simulative-based workflow, *"We usually scout players based on historical performance, which is tedious and often leads to players unsuitable for the team. Such a simulation method could help estimate whether the players are suitable for a team and reduce transfer failures"*. EA also appreciated our match simulation framework, *"Most of the players recommended by the system meet my*

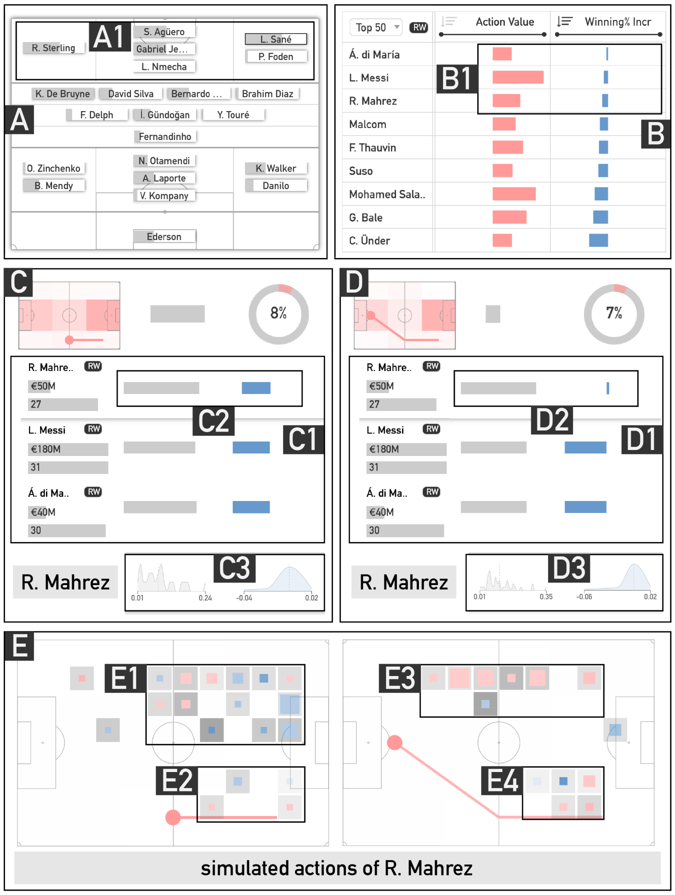

Fig. 8: The analysis process of case 2. (A) is the squad board of Manchester City. (B) presents the recommended players based on L. Sané. (C) and (D) are the simulated actions of the potential target players. (E) presents the detailed simulated actions of R. Mahrez.

expectations. The simulated actions and their values could also provide new insights, such as how the players will act in a certain tactic and whether the action choice is good or bad". As for the visual design, EA favored the on-ball tactic list and the player record list, *"It is efficient to compare the players' expected performance under a certain tactic, which helps a lot in deciding players suitable for key tactics"*. EB liked the squad board and the player ranking list, *"The squad board is familiar to me, and the player ranking list is also useful in finding players according to my preference"*. EB was also impressed by the simulated action map, *"It could help me know how the new players will behave in a tactic and inspire me to arrange proper tasks for them or adjust the tactic based on the player's competence and style"*.

Suggestion. The experts also provided several suggestions for improving our system. EA hoped to improve the accuracy of the match simulation framework with fine-grained soccer match data such as player trajectories, *"Integrating fine-grained features of player behaviors such as their trajectories could enable a more accurate simulation for both match results and player behaviors"*. EB suggested supporting the navigation of multiple roles of new players, *"The navigation view only supports finding players based on a certain player in the squad. However, in some scenarios, coaches would reconstruct the tactical system of a team with multiple roles of new players, which might be a different workflow that needs to be supported"*.

8 Discussion

Significance. Player scouting is essential in team sports to obtain suitable players to strengthen the team [3]. Match simulation methods have been applied to scouting players by estimating their expected performance in a new team to avoid unsuitable players. However, they fall short of interactive navigation of potential target players and investigation of their expected performance through fine-grained simulated behaviors. In this work, we develop a visual analytics approach to address these challenges in soccer through a match simulation framework that supports simulating both match results and player behaviors and a visual analytics system enables player navigation, comparison, and investigation. With our approach, coaches and analysts can scout players who will improve team performance and understand how they can be integrated into the tactics of a new team in future matches.

Generalizability. In this work, we focus on soccer player scouting in professional clubs because it is the most complicated scenario with the largest scope of candidate players and the most factors to consider. Thus, it can be directly extended to other simpler scenarios, such as in national teams. Our approach can also be generalized to other team sports with similar match structures, such as basketball and ice hockey. The match simulation framework is intrinsically based on a sequential prediction model for player behaviors and can be easily adapted to these sports according to their definitions of match states and player behaviors. The analysis workflow and visual design of our system can also be easily modified based on their specific domain concepts.

Lessons learned. We have learned two lessons through this work. First, visualizations related to specific domain concepts can be effective for expert users. During visual design iteration, we proposed both pitch-based and general visualizations to represent the squad and on-ball tactics. Our experts indicated that although the general visualizations, such as a ranking list for players in the squad and sequential-based visualizations for on-ball tactics, are also effective in exploration and comparison, they are difficult to relate to soccer domain knowledge for understanding. Thus, we choose pitch-based visualizations for them. Such a lesson can help design or generate visualizations highly related to specific domain concepts [58, 68]. Second, directly presenting the comparison results can facilitate visual comparison for spatio-temporal events. Compared with placing different players' actions on split views, presenting the comparison results in the same pitch can reduce switching between two views and additional calculation of their differences. Moreover, we encode the comparison results by glyphs on the pitch grids to maintain the spatial distribution of the action choices. The visualization can also be generalized to other visual comparison tasks for spatio-temporal events with multiple attributes [15, 57].

Limitations. We have observed two limitations in this work. First, the player attributes involved in our system are inadequate. In soccer, scouting players not only needs to consider their in-match performance and physical attributes but also other various important factors. We plan to collect more player attributes and integrate them into our workflow. In detail, the attributes that can be summarized as player comparison criteria, such as financial factors and player reliability, can be integrated into the player ranking list for player comparison, and those related to action decisions, such as player personality, can be added to the match simulation framework to obtain more precise simulation results. Second, the scalability of the simulated action visualization is limited. In our system, the player's actions are simulated and visualized after being selected from the player ranking list due to the limited computational resources. In the future, we will support parallel simulating actions for all players in the player ranking list and directly illustrate players expected to perform well under each on-ball tactic. We will also design summarized visualizations for detailed simulated actions to support comparisons among the players in the player record list.

9 Conclusion

In this work, we explore the problem of soccer player scouting based on match simulation methods. At first, we establish a two-level soccer match simulation framework to support estimating players' expected contribution to a new team on both match results and their behaviors. Based on the framework, we develop a visual analytics system, Team-Scouter, to facilitate simulative-based player scouting through team-level navigation, player-level comparison, and action-level investigation, for finding and deciding the players suitable for the team.

In the future, we plan to improve our work in two aspects. First, we will collect fine-grained soccer match data, such as player trajectories, and incorporate such data into the match simulation framework. It can provide a more accurate simulation for soccer matches, thus enhancing the estimation of players' expected contribution to a new team. Second, we plan to generalize our simulative-based player scouting method to other team sports, such as basketball and ice hockey, after summarizing the general player scouting pipeline for multiple team sports.

Acknowledgments

The work was supported by Key "Pioneer" R&D Projects of Zhejiang Province (2023C01120), NSFC (U22A2032, 62072400), and the Collaborative Innovation Center of Artificial Intelligence by MOE and Zhejiang Provincial Government (ZJU).

References

[1] G. L. Andrienko, N. V. Andrienko, G. Anzer, P. Bauer, G. Budziak, G. Fuchs, D. Hecker, H. Weber, and S. Wrobel. Constructing Spaces and Times for Tactical Analysis in Football. *IEEE Transactions on Visualization and Computer Graphics*, 27(4):2280–2297, 2021. doi: 10.1109/TVCG.2019.2952129 2

[2] G. L. Andrienko, N. V. Andrienko, G. Budziak, J. Dykes, G. Fuchs, T. von Landesberger, and H. Weber. Visual Analysis of Pressure in Football. *Data Mining and Knowledge Discovery*, 31(6):1793–1839, 2017. doi: 10.1007/s10618-017-0513-2 2

[3] R. Beal, T. J. Norman, and S. D. Ramchurn. Artificial Intelligence for Team Sports: A Survey. *The Knowledge Engineering Review*, 34, 2019. doi: 10.1017/S0269888919000225 1, 2, 9

[4] L. Bransen, P. Robberechts, J. Van Haaren, and J. Davis. Choke or Shine? Quantifying Soccer Players' Abilities to Perform Under Mental Pressure. In *2019 MIT Sloan Sports Analytics Conference*, 2019. 2

[5] L. Bransen and J. Van Haaren. Measuring Football Players' On-the-Ball Contributions from Passes During Games. In *International Workshop on Machine Learning and Data Mining for Sports Analytics*, pp. 3–15, 2019. doi: 10.1007/978-3-030-17274-9_1 2

[6] L. Bransen, J. Van Haaren, and M. van de Velden. Measuring Soccer Players' Contributions to Chance Creation by Valuing Their Passes. *Journal of Quantitative Analysis in Sports*, 15(2):97–116, 2019. doi: 10.1515/jqas-2018-0020 2

[7] J. Brooks, M. Kerr, and J. Guttag. Developing a Data-Driven Player Ranking in Soccer using Predictive Model Weights. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 49–55, 2016. doi: 10.1145/2939672.2939695 2, 3

[8] A. Cao, J. Lan, X. Xie, H. Chen, X. Zhang, H. Zhang, and Y. Wu. Team-Builder: Toward More Effective Lineup Selection in Soccer. *IEEE Transactions on Visualization and Computer Graphics*, 29(12):5178–5193, 2023. doi: 10.1109/TVCG.2022.3207147 2

[9] A. Cao, X. Xie, J. Lan, H. Lu, X. Hou, J. Wang, H. Zhang, D. Liu, and Y. Wu. MIG-Viewer: Visual Analytics of Soccer Player Migration. *Visual Informatics*, 5(3):102–113, 2021. doi: 10.1016/j.visinf.2021.09.002 2

[10] A. Cao, X. Xie, M. Zhou, H. Zhang, M. Xu, and Y. Wu. Action-Evaluator: A Visualization Approach for Player Action Evaluation in Soccer. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):880–890, 2024. doi: 10.1109/TVCG.2023.3326524 1, 2

[11] T. Decroos, L. Bransen, J. Van Haaren, and J. Davis. Actions Speak Louder than Goals: Valuing Player Actions in Soccer. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1851–1861, 2019. doi: 10.1145/3292500.3330758 1, 2, 3, 4, 5

[12] T. Decroos, J. Van Haaren, and J. Davis. Automatic Discovery of Tactics in Spatio-Temporal Soccer Match Data. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 223–232, 2018. doi: 10.1145/3219819.3219832 2, 3, 4, 6

[13] T. Decroos, J. Van Haaren, V. Dzyuba, and J. Davis. STARSS: A Spatio-Temporal Action Rating System for Soccer. In *Proceedings of the 4th Workshop on Machine Learning and Data Mining for Sports Analytics co-located with 2017 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2017)*, pp. 11–20, 2017. 2

[14] C. A. Dietrich, D. Koop, H. T. Vo, and C. T. Silva. Baseball4D: A Tool for Baseball Game Reconstruction and Visualization. In *Proceedings of IEEE Conference on Visual Analytics Science and Technology*, pp. 23–32, 2014. doi: 10.1109/VAST.2014.7042478 2

[15] J. Dong, H. Zhang, M. Cui, Y. Lin, H.-Y. Wu, and C. Bi. TCEVis: Visual Analytics of Traffic Congestion Influencing Factors Based on Explainable Machine Learning. *Visual Informatics*, 8(1):56–66, 2024. doi: 10.1016/j.visinf.2023.11.003 9

[16] M. Du and X. Yuan. A Survey of Competitive Sports Data Visualization and Visual Analysis. *Journal of Visualization*, 24(1):47–67, 2021. doi: 10.1007/s12650-020-00687-2 2

[17] J. Fernández, L. Bornn, and D. Cervone. Decomposing the Immeasurable Sport: A Deep Learning Expected Possession Value Framework for Soccer. In *2019 MIT Sloan Sports Analytics Conference*, 2019. 7

[18] J. Fernandez-Navarro, L. Fradua, A. Zubillaga, P. R. Ford, and A. P. McRobert. Attacking and Defensive Styles of Play in Soccer: Analysis of Spanish and English Elite Teams. *Journal of Sports Science*, 34(24):2195–2204, 2016. doi: 10.1080/02640414.2016.1169309 6

[19] Y. Fu and J. Stasko. HoopInSight: Analyzing and Comparing Basketball Shooting Performance Through Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):858–868, 2024. doi: 10.1109/TVCG.2023.3326910 2

[20] K. Goldsberry. Courtvision: New Visual and Spatial Analytics for the NBA. In *Proceedings of MIT Sloan Sports Analytics Conference*, 2012. 2

[21] H. Janetzko, D. Sacha, M. Stein, T. Schreck, D. A. Keim, and O. Deussen. Feature-Driven Visual Analytics of Soccer Data. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 13–22, 2014. doi: 10.1109/VAST.2014.7042477 1, 2

[22] A. Jarvandi, S. Sarkani, and T. Mazzuchi. Modeling Team Compatibility Factors Using a Semi-Markov Decision Process: A Data-Driven Approach to Player Selection in Soccer. *Journal of Quantitative Analysis in Sports*, 9(4):347–366, 2013. doi: 10.1515/jqas-2012-0054 1, 2, 3, 4

[23] M. Lage, J. P. Ono, D. Cervone, J. Chiang, C. A. Dietrich, and C. T. Silva. StatCast Dashboard: Exploration of Spatiotemporal Baseball Data. *IEEE Computer Graphics and Applications*, 36(5):28–37, 2016. doi: 10.1109/MCG.2016.101 2

[24] J. Lan, J. Wang, X. Shu, Z. Zhou, H. Zhang, and Y. Wu. RallyComparator: Visual Comparison of the Multivariate and Spatial Stroke Sequence in Table Tennis Rally. *Journal of Visualization*, 25(1):143–158, 2022. doi: 10.1007/s12650-021-00772-0 2

[25] P. A. Legg, D. H. S. Chung, M. L. Parry, M. W. Jones, R. Long, I. W. Griffiths, and M. Chen. MatchPad: Interactive Glyph-Based Visualization for Real-Time Sports Performance Analysis. *Computer Graphics Forum*, 31(3):1255–1264, 2012. doi: 10.1111/j.1467-8659.2012.03118.x 2

[26] A. G. Losada, R. Therón, and A. Benito. BKViz: A Basketball Visual Analysis Tool. *IEEE Computer Graphics and Applications*, 36(6):58–68, 2016. doi: 10.1109/MCG.2016.124 2

[27] V. Machado, R. Leite, F. Moura, S. Cunha, F. Sadlo, and J. L. Comba. Visual Soccer Match Analysis Using Spatiotemporal Positions of Players. *Computers and Graphics*, 68:84–95, 2017. doi: 10.1016/j.cag.2017.08.006 2

[28] C. Merhej, R. J. Beal, T. Matthews, and S. Ramchurn. What Happened Next? Using Deep Learning to Value Defensive Actions in Football Event-Data. In *Proceedings of the 27th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 3394–3403, 2021. doi: 10.1145/3447548.3467090 1, 2

[29] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient Estimation of Word Representations in Vector Space. *arXiv Preprint*, arXiv:1301.3781, 2013. doi: 10.48550/arXiv.1301.3781 4

[30] B. Moon and R. Brath. Bloomberg Sports Visualization for Pitch Analysis. In *Workshop on Sports Data Visualization*, 2013. 2

[31] M. M. Nasiri, M. Ranjbar, M. Tavana, F. J. S. Arteaga, and R. Yazdanparast. A Novel Hybrid Method for Selecting Soccer Players During the Transfer Season. *Expert Systems*, 36(1):e12342, 2019. doi: 10.1111/exsy.12342 2

[32] J. P. Ono, C. A. Dietrich, and C. T. Silva. Baseball Timeline: Summarizing Baseball Plays Into a Static Visualization. *Computer Graphics Forum*, 37(3):491–501, 2018. doi: 10.1111/cgf.13436 2

[33] G. Pantuso. The Football Team Composition Problem: A Stochastic Programming Approach. *Journal of Quantitative Analysis in Sports*, 13(3):113–129, 2017. doi: 10.1515/jqas-2017-0030 1, 2, 3, 4

[34] L. Pappalardo, P. Cintia, P. Ferragina, E. Massucco, D. Pedreschi, and F. Giannotti. PlayeRank: Data-Driven Performance Evaluation and Player Ranking in Soccer via a Machine Learning Approach. *ACM Transactions on Intelligent Systems and Technology*, 10(5):1–27, 2019. doi: 10.1145/3343172 2, 3

[35] L. Pappalardo, P. Cintia, A. Rossi, E. Massucco, P. Ferragina, D. Pedreschi, and F. Giannotti. A Public Data Set of Spatio-temporal Match Events in Soccer Competitions. *Scientific Data*, 6(1):1–15, 2019. doi: 10.1038/s41597-019-0247-7 3, 7

[36] V. M. Payyappalli and J. Zhuang. A Data-Driven Integer Programming Model for Soccer Clubs' Decision Making on Player Transfers. *Environment Systems and Decisions*, 39:466–481, 2019. doi: 10.1007/s10669-019-09721-7 2

[37] C. Perin, J. Boy, and F. Vernier. Using Gap Charts to Visualize the

Temporal Evolution of Ranks and Scores. *IEEE Computer Graphics and Applications*, 36(5):38–49, 2016. doi: 10.1109/MCG.2016.100 2

[38] C. Perin, R. Vuillemot, and J.-D. Fekete. SoccerStories: A Kick-off for Visual Soccer Analysis. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2506–2515, 2013. doi: 10.1109/TVCG.2013.192 2

[39] C. Perin, R. Vuillemot, and J.-D. Fekete. A table!: Improving Temporal Navigation in Soccer Ranking Tables. In *CHI Conference on Human Factors in Computing Systems*, pp. 887–896, 2014. doi: 10.1145/2556288.2557379 2

[40] C. Perin, R. Vuillemot, C. D. Stolper, J. T. Stasko, J. Wood, and S. Carpendale. State of the Art of Sports Data Visualization. *Computer Graphics Forum*, 37(3):663–686, 2018. doi: 10.1111/cgf.13447 2

[41] H. Pileggi, C. D. Stolper, J. M. Boyle, and J. T. Stasko. SnapShot: Visualization to Propel Ice Hockey Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2819–2828, 2012. doi: 10.1109/TVCG.2012.263 2

[42] T. Polk, D. Jäckle, J. Häußler, and J. Yang. CourtTime: Generating Actionable Insights into Tennis Matches Using Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):397–406, 2020. doi: 10.1109/TVCG.2019.2934243 2

[43] P. Rahimian, J. Van Haaren, T. Abzhanova, and L. Toka. Beyond Action Valuation: A Deep Reinforcement Learning Framework for Optimizing Player Decisions in Soccer. In *2022 MIT Sloan Sports Analytics Conference*, 2022. 2

[44] P. Robberechts, M. Van Roy, and J. Davis. un-xPass: Measuring Soccer Player's Creativity. In *Proceedings of the 29th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 4768–4777, 2023. doi: 10.1145/3580305.3599924 1, 2, 3, 4

[45] A. Rusu, D. Stoica, and E. Burns. Analyzing Soccer Goalkeeper Performance Using a Metaphor-Based Visualization. In *2011 15th International Conference on Information Visualisation*, pp. 194–199, 2011. doi: 10.1109/IV.2011.57 2

[46] A. Rusu, D. Stoica, E. Burns, B. Hample, K. McGarry, and R. Russell. Dynamic Visualizations for Soccer Statistical Analysis. In *2010 14th International Conference on Information Visualisation*, pp. 207–212, 2010. doi: 10.1109/IV.2010.39 2

[47] M. Ryoo, N. Kim, and K. Park. Visual Analysis of Soccer Players and a Team. *Multimedia Tools and Applications*, 77(12):15603–15623, 2018. doi: 10.1007/s11042-017-5137-4 1, 2

[48] D. Sacha, F. Al-Masoudi, M. Stein, T. Schreck, D. A. Keim, G. L. Andrienko, and H. Janetzko. Dynamic Visual Abstraction of Soccer Movement. *Computer Graphics Forum*, 36(3):305–315, 2017. doi: 10.1111/cgf.13189 2

[49] S. A. F. Salles, H. R. M. da Hora, M. E. Júnior, A. S. Velasco, and P. R. Croce. Multiple Choice Method with Genetic Algorithm for the Formation of Soccer Teams. *Pesquisa Operacional*, 42:e243537, 2022. doi: 10.1590/0101-7438.2022.042.00243537 2

[50] D. Seebacher, T. Polk, H. Janetzko, D. A. Keim, T. Schreck, and M. Stein. Investigating the Sketchplan: A Novel Way of Identifying Tactical Behavior in Massive Soccer Datasets. *IEEE Transactions on Visualization and Computer Graphics*, 29(4):1920–1936, 2023. doi: 10.1109/TVCG.2021.3134814 2

[51] L. Shao, D. Sacha, B. Neldner, M. Stein, and T. Schreck. Visual-Interactive Search for Soccer Trajectories to Identify Interesting Game Situations. *Electronic Imaging*, (1):1–10, 2016. doi: 10.2352/ISSN.2470-1173.2016.1.VDA-510 2

[52] I. Simpson, R. J. Beal, D. Locke, and T. J. Norman. Seq2Event: Learning the Language of Soccer Using Transformer-based Match Event Prediction. In *Proceedings of the 28th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 3898–3908, 2022. doi: 10.1145/3534678.3539138 1, 4

[53] M. Stein, H. Janetzko, T. Breitkreutz, D. Seebacher, T. Schreck, M. Grossniklaus, I. D. Couzin, and D. A. Keim. Director's Cut: Analysis and Annotation of Soccer Matches. *IEEE Computer Graphics and Applications*, 36(5):50–60, 2016. doi: 10.1109/MCG.2016.102 2

[54] M. Stein, H. Janetzko, A. Lamprecht, T. Breitkreutz, P. Zimmermann, B. Goldlücke, T. Schreck, G. L. Andrienko, M. Grossniklaus, and D. A. Keim. Bring It to the Pitch: Combining Video and Movement Data to Enhance Team Sport Analysis. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):13–22, 2018. doi: 10.1109/TVCG.2017.2745181 2

[55] Transfermarkt. Football Transfers, Rumours, Market Values, News and Statistics | Transfermarkt. `https://www.transfermarkt.com/`, 2024. 1, 3

[56] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is All you Need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 6000–6010, 2017. doi: 10.5555/3295222.3295349 4

[57] H. Wang, Y. Ni, L. Sun, Y. Chen, T. Xu, X. Chen, W. Su, and Z. Zhou. Hierarchical Visualization of Geographical Areal Data with Spatial Attribute Association. *Visual Informatics*, 5(3):82–91, 2021. doi: 10.1016/j.visinf.2021.09.001 9

[58] J. Wang, X. Li, C. Li, D. Peng, A. Z. Wang, Y. Gu, X. Lai, H. Zhang, X. Xu, X. Dong, Z. Lin, J. Zhou, X. Liu, and W. Chen. AVA: An Automated and AI-Driven Intelligent Visual Analytics Framework. *Visual Informatics*, 8(2):106–114, 2024. doi: 10.1016/j.visinf.2024.06.002 9

[59] J. Wang, J. Ma, K. Hu, Z. Zhou, H. Zhang, X. Xie, and Y. Wu. Tac-Trainer: A Visual Analytics System for IoT-based Racket Sports Training. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):951–961, 2023. doi: 10.1109/TVCG.2022.3209352 2

[60] W. Wang, J. Zhang, X. Yuan, and S. Liu. MatchOrchestra: A Generalized Visual Analytics for Competitive Team Sports. *Journal of Visualization*, 19(3):515–528, 2016. doi: 10.1007/s12650-015-0337-3 2

[61] J. Wu, D. Liu, Z. Guo, Q. Xu, and Y. Wu. TacticFlow: Visual Analytics of Ever-Changing Tactics in Racket Sports. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):835–845, 2022. doi: 10.1109/TVCG.2021.3114832 2

[62] Y. Wu, D. Deng, X. Xie, M. He, J. Xu, H. Zhang, H. Zhang, and Y. Wu. OBTracker: Visual Analytics of Off-ball Movements in Basketball. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):929–939, 2023. doi: 10.1109/TVCG.2022.3209373 2

[63] Y. Wu, J. Lan, X. Shu, C. Ji, K. Zhao, J. Wang, and H. Zhang. iTTVis: Interactive Visualization of Table Tennis Data. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):709–718, 2018. doi: 10.1109/TVCG.2017.2744218 2

[64] Y. Wu, X. Xie, J. Wang, D. Deng, H. Liang, H. Zhang, S. Cheng, and W. Chen. ForVizor: Visualizing Spatio-Temporal Team Formations in Soccer. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):65–75, 2019. doi: 10.1109/TVCG.2018.2865041 2

[65] X. Xie, J. Wang, H. Liang, D. Deng, S. Cheng, H. Zhang, W. Chen, and Y. Wu. PassVizor: Toward Better Understanding of the Dynamics of Soccer Passes. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1322–1331, 2021. doi: 10.1109/TVCG.2020.3030359 2, 6

[66] C. Yanai, A. Solomon, G. Katz, B. Shapira, and L. Rokach. Q-Ball: Modeling Basketball Games Using Deep Reinforcement Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, pp. 8806–8813, 2022. doi: 10.1609/aaai.v36i8.20861 4

[67] S. Ye, Z. Chen, X. Chu, Y. Wang, S. Fu, L. Shen, K. Zhou, and Y. Wu. ShuttleSpace: Exploring and Analyzing Movement Trajectory in Immersive Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):860–869, 2021. doi: 10.1109/TVCG.2020.3030392 2

[68] Y. Ye, J. Hao, Y. Hou, Z. Wang, S. Xiao, Y. Luo, and W. Zeng. Generative AI for Visualization: State of the Art and Future Directions. *Visual Informatics*, 8(2):43–66, 2024. doi: 10.1016/j.visinf.2024.04.003 9