MASTER

Retrieving Episodes of Football Matches using an Information Retrieval Model for Tracking Data

van Rooijen, Bart W.M.

Award date:
2020

EINDHOVEN UNIVERSITY OF TECHNOLOGY

Master Thesis

Department of Mathematics and Computer Science
Data Mining Research Group

Retrieving Episodes of Football Matches using an Information Retrieval Model for Tracking Data

B.W.M. van Rooijen

Supervisors:
Prof.dr. M. Pechenizkiy
Dr. O. Papapetrou
MSc. G. Budziak

version 3.5

Eindhoven, October 2020

"Every trainer talks about movement, about running a lot. I say don't run so much. Football is a game you play with your brain. You have to be in the right place at the right moment, not too early, not too late."

JOHAN CRUIJFF

Abstract

Football clubs can benefit from automated episode retrieval from tracking data. The main benefits are related to match analysis done by analysts. Automated retrieval can reduce the workload and increase the quality of match analysis in comparison to current methods. Interviews and a literature review were conducted to understand the information needs of coaches and analysts. The results indicate that limited attention is given to the retrieval of second-order events. Second-order events explain the collective behaviour of a team. A second-order event consists typically of one or multiple first-order events, such as a pass or shot. Another finding is the potential of temporal aggregation and collective features to capture collective behaviour.

These techniques are included in an information retrieval model to retrieve episodes from tracking data. The model is designed to be understandable, expressive, customizable, and retrieve episodes in a reasonable amount of time. The model consists of a document representation, query representation, framework, and ranking function. The ranking function is based on the retrieval functions Temporal Query and Query by Semantic Example. These functions satisfy the information needs of known-item search and existence search.

A series of experiments was conducted to test the quality and utility of the model. The experiments were designed to cover questions that a coach could ask an analyst. The quantitative results show the ability to replace manual annotation of both first- and second-order events when perfect recall is not essential. The qualitative results based on the method Evaluation of End Products indicate that the model can improve the quality of match analysis.

Acknowledgements

First of all, I wish to express my gratitude to my supervisor Guido Budziak. I am really gratefull for introducing me to Football Analytics and putting me in touch with experts in the field. Moreover, I really appreciate the encouragement, trust and critical reflection he has given on my work throughout the duration of the project.

In addition, I want to express my great appreciation to my other supervisor Mykola Pechenizkiy for his professional guidance and valuable support during this project. His patient guidance, support and useful critiques have been a great help in finalizing this project.

I also want to thank Pascal Bauer for the cooperation with the Deutscher Fußball-Bund. In particular for the data provision.

Finally, I want to give a special thanks to my family and friends. Their support and encouragement was worth more than I can express on paper.

Bart van Rooijen

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Liverpool FC won the premier league for the first time in 30 years after a phenomenal performance during the season 2019/2020. Many people are aware of the contribution of the coach, Jürgen Klopp, to the success of Liverpool FC. But not that many people know the crucial contribution of the Data Science department. Without the Data science department of Liverpool, Jürgen Klopp would have never been hired. The last season before Jürgen Klopp moved to Liverpool, he had a disappointing season with Borussia Dortmund that finished at the seventh place of the Bundesliga. The director of research, Ian Graham, found out that Borussia Dortmund had a lot of bad luck during the season and would have finished second based on true performance. This finding contributed to the decision to hire Jürgen Klopp. Or using the words of Klopp: *"The department [Data science] there in the back of the building? They are the reason I'm here."*[49]

The story indicates how data can be an asset to support the evaluation of performance in football. The field that covers this activity is known as Performance Analysis (PA) in football. PA refers to the analysis of technical, physical, tactical activities, and their interactions [29]. The foundations of PA can be traced back to Charles Reep. Although his findings are criticized, his paper 'Skill and Chance in Association Football' [42] was the start of PA. Today, many elite football clubs have their own department for PA to support coaches with important activities, such as player recruitment, effectiveness measurement of training sessions, and evaluation of playing strategies for two sides [31]. Besides, PA is used by other parties, like journalists, gamblers, and analysts to understand factors that explain success in football [13].

1.1 Business perspective

The main goal of an elite football club is to achieve a desired result [33]. A desired result is prices or a high ranking after a complete season in the league. Clubs want to maximize their performance to increase the chances of achieving their goals. They are restrained by their resources which in most cases is highly affected by their budget. As a consequence, clubs have to make decisions to use their resources as optimal as possible to maximize performance. The remainder of this section explains which aspects are important for decision making with respect to PA.

1.1.1 Player recruitment

A lot of the important decisions are related to the selection of the technical staff and players, as clubs spend 60% of their budget on salaries [58]. For the recruitment of players, clubs use scouts that recommend players to a club. Traditionally, scouts make human observations to estimate the quality of a player and decide if the player can be of added value for the club. Nowadays, according to several articles, more and more clubs are inspired by the moneyball-principle for player recruitment [1][45][47]. The moneyball-principle is based on the work of Billy Bean at the

Oakland Athletics baseball team which is covered in the book 'Moneyball: The Art of Winning an Unfair Game'. By using this principle, players are being judged on their performance based on data instead of human judgement.

In the end, the clubs want to have players that lead to the best performance within the available resources. However, it is not always the case that a team with the best players leads to the best results. Players have to suit each other, the coach, and his playing strategy. Bransen & Van Haaren [10] refer to this as chemistry and argued the importance of it. In addition, they show how PA and in particular chemistry can support scouts with the evaluation of player quality.
In practice, clubs make use of data service companies such as Sci-sports, Wyscout, and Scoutpad to help them with the evaluation of target players. These companies are specialized in estimating the (added) value of players based on performance estimated with data. Furthermore, clubs hire one or multiple analysts to do PA of players. Both data service companies and analysts estimate player quality based on match performance. They decide on the quality of the player based on how he has performed during the season. In order to estimate this correctly, they have to take playing strategies and tactics into account, because the performance of a player is influenced by the choices of his coach.

In short, the use of PA for the recruitment of players saves money, as clubs do not have to send scouts to watch matches [15]. Moreover, they can save money or increase performance, because they make better decisions on the players they recruit. This is the result of better judgements with the help of data instead of completely relying on human opinions of scouts. To achieve this, it is necessary to evaluate player performance and team performance of individual matches.

1.1.2 Match analysis

To optimize the performance of a team, match analysis can play a great role. Typically, an elite football club employs one or multiple analysts to support the technical staff with the analysis of training sessions and matches. An analysis helps to evaluate the performance of a team and the players. A coach can use this analysis to decide who is the best player to play the next game or to optimize the tactics and strategies of the team. Moreover, some individual players make use of external analysts, because a club does not have sufficient resources to analyze their individual performance. The goal is to identify weaknesses to improve their abilities. Even a minor improvement can make a difference in winning or losing the next game.

Analysis of a match is a very labour-intensive task and can take up to several days to complete [57]. The total time spent on post-match analysis was investigated by means of a survey send to 48 match analysts at elite football clubs. 72.9 % indicated that post-match analysis takes more than 2 hours of which 22.9% indicated that it takes 6 hours or more [61].
The base of the analysis is match annotation. Most analysts annotate a match live during a game. This type of annotation is limited to the labeling of first-order events. The annotated events are used for widespread statistical metrics or identification of interesting episodes. The widespread statistical metrics, such as the number of corners or passes do not give explanations to the coach. His interest lies within the episodes that give explanations to certain situations rather than the statistical data. Episodes are fragments of a match that may include information that helps to give these explanations. These explanations have the purpose to find the 'why' in addition to the 'what' and 'where' [56]. These episodes consist of second-order events. The analyst annotates these events after the match, because it takes more time than live annotation. For one match, a likely approach is the following: the analyst watches the game. Then, he has to watch the game again while he does the annotation. For the annotation he has to pause, play forward and backwards continuously to identify events. So, for post-match annotation it takes approximately 4.5 hours.

In addition, video analyses have to be prepared. This work is built upon the work of the annotator. A video analysis consists of important highlights from a match based on the criteria from the coach. Examples are mistakes or deviations from the desired playing style. The analyst has to identify these moments and presents them to the coach. Based on the coach his opinion, the analyst either presents the analysis to the team or updates his work [54]. Besides, the video analyst has to analyze the next opponent to identify their strengths and weaknesses [4]. The routine of an elite football club between two consecutive matches and the work of the analyst to meet the information needs is described in greater detail in Section 2.1.

All the activities mentioned in this section share a similar goal. Identification of episodes that can support the coach in evaluation and explanation of his football philosophy to the team with the goal to improve performance.

1.1.3 Analysis automation

Analysts could benefit greatly from automation of their activities. The three main benefits of automation are: (1) a reduction in workload, (2) a quality improvement of their work and (3) the speed of result delivery.

Workload reduction

Automation of activities can reduce the workload of the analyst. By using the right techniques and algorithms, tasks normally done in hours or days, can be done in several seconds. One of these tasks is the annotation introduced in Section 1.1.2. The most important task of automatic annotation is to detect an event. Automatic annotation is very suitable for first-order events because there are clear rules to follow for detection. But also, annotation of second-order events can be done faster. The effect is even greater, as the annotation of second-order events takes more time in general.

Quality improvement

Automation in the work of analysts could increase the quality of his work. For example, manual annotation suffers from bias, human error, lack of time, or simple lack of knowledge and with automation it can become faster, more reliable, and cost-effective [57]. The first advantage is that the analysis can be done symmetrically. For years, football is seen from the perspective of the team. It is common that two analysts of different clubs annotate the same game. However, the use of software makes it possible to annotate from the home team, away team, or neutral perspective. It is called symmetric annotation. The advantage is that that annotation can be done simultaneously for the perspective of both teams.
The next advantage is higher reliability in contrary to manual annotation, because a computer is objective and consequent. During manual annotation, the annotator is likely to miss or misjudge events, for example due to a lack of concentration.
The last advantage is the possibility to extend the annotation. Currently, manual annotation is limited to the events that are available on the template of the analyst. With the use of a computer, it is possible to increase the number and complexity of events, because there is no risk of cognitive overload.

Offline and online algorithms

The algorithms for automatic annotation can be divided into two categories. Some can analyze a game in real-time, while others need the complete dataset of a match. The former is known as an online algorithm, the latter goes as an offline algorithm. Online algorithms are used when future knowledge is not essential to give a solution. The main benefit of this kind of approach is that fewer resources are necessary in terms of running time and computational power in comparison to offline algorithms. In PA, online algorithms have the advantage that they can provide results during the match or during the break. For example, the coach could use a tablet during the game to receive real-time insights related to the positioning of the players. On the contrary, offline algorithms have the benefit that they perform better than online algorithms and are more likely to provide optimal solutions. For PA, offline algorithms are very suitable for performance evaluation of a team and detecting strategic aspects of the match.

Football clubs could benefit to a great extent from automation of the activities of analysts. The time spend on labour-intensive activities, such as annotation could be reduced from days to a couple of seconds. It gives analysts the opportunity to spend more time on the transfer of knowledge and advancement of their work. Furthermore, the quality of the analysis can improve while the input of the coach is being maintained. Besides, better decisions could be made for player recruitment. All things considered, it could lead to a cost reduction and increase in performance.

1.2 Football Perspective

To analyse a football match it is necessary to have a basic understanding of how the game is played and what is considered important by experts. Besides, particular knowledge on the differences between football philosophies and evaluation by experts is interesting, because it shows there is not one truth or one-size-fits-all in terms of evaluation and judgement. This was already discovered in the fifties for the similar sport American football [21]. Apart from the difference in the evaluation of performance, it was found that the key performance indicators in football differ per expert [14]. That means that different coaches focus on different indicators.
Also, a football match is highly influenced by coincidence due to goals being relatively rare. This observation makes it difficult to draw conclusions in accordance with true performance. Both arguments should be taken into account for the development of tools that support the work of analysts.

This can be illustrated with an extraordinary example. The home team is one goal ahead during the stoppage time of the second half. The coach of the home team decides to substitute a striker for an extra defender. Then, that extra defender wins ball possession close to his own goal and decides to look for a 'footballing solution'. It means that a player tries to keep the ball in possession. In this case, the defender gives a short pass to another defender. The ball is intercepted by the striker of the away team and he makes the equalizer. The game ends in a draw. In this kind of situation, it is very unlikely that experts would agree on these decisions. First of all, the decision for the substitution, there are experts that claim that you should not make a defensive substitution, because they are afraid that the team will be pushed further back and they say it is better to look for a two-goal advantage. On the other hand, there are experts that state the extra defender is necessary to deal with the offensive actions of the opposition and defend the 1-goal advantage. The same applies to the decision of the defender that went for the 'footballing solution'. There are experts that believe that a player should always go for the 'footballing solution', while others would argue that in this particular case it would be better to shoot the ball 'out of the stadium'. These are examples that indicate differences in football philosophy.

ASPOGAMO

Therefore, it is not likely that all tasks of the analyst can be replaced by one algorithm, but instead, automation will change the work of an analyst instead of completely replacing them. To give a quantification of the subjectivity, we can use the ASPOGAMO Abstract Model [12]. The model is aimed to give a better comprehension of the game. It divides the game into smaller segments that form a hierarchical model. Every layer addresses a specific part of the game. An illustration of the model is displayed in Figure 1.1. As a rule of thumb, the level height indicates the level of subjectivity. At the positional layer for example, there can be no misunderstanding of where a player is positioned at the field. On the other hand, at the top level, the strategic layer, it is more likely to have a disagreement over the type of playing style, but also the playing style is open for debate.

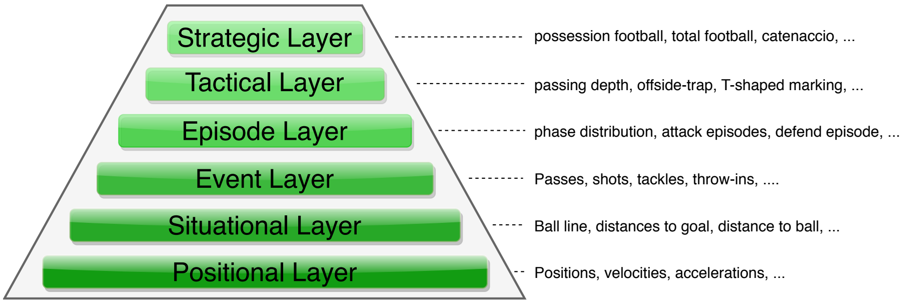


Figure 1.1: The ASPOGAMO Abstract Model [12]

The level of subjectivity is an indicator of the ability to fully automate a task. That means that it is easier to automate the task to detect if a player is in an outside position than the annotation of second-order events. Besides, the differences in matches and players can make it difficult to completely automate a task [53]. A semi-automatic approach or a human-in-the-loop could overcome the issues. It means that the algorithms calculate metrics or generate visualizations automatically, but the analyst can interact with the results by adjusting parameters, selecting metrics, and selecting visualizations [23][52].

Interaction

To deal with the different opinions of experts, it is necessary that the experts can interact with the results. Interaction makes it possible to translate their view and preferences to the information system. It also reduces the complexity of the task and supports the sense-making process [53]. Interaction can only be provided when the expert understands how the information system works. In comparison to sport science studies, computer science studies tend to rely on more advanced methods that are more difficult to interpret by experts [19]. As a consequence, experts do not understand how the information system works which leads to a lack of adoption of research findings.

This emphasizes the need for understandable methods for information systems. On the other hand, to work with patterns from the top 3 layers of the ASPOGAMA model, you need methods that can capture complex behaviours at those levels. Hence, the right balance has to be found between understandability and complexity. Apart from the choice of methods, the same trade-off has to be made for attribute selection. *"An attribute is a data field, representing a characteristic or feature of a data object [20]"*. An attribute is further referred to as feature as these names are interchangeable. These features cover patterns in the data that explain the behaviour observed in the data. Interpretable features or semantic features represent a characteristic with a meaning that can be understood by the expert. So, the trade-off has to be found between understandability and the extent a feature can capture complex behaviour. A great example of the right balance is a collective feature, such as Team Width. Further explanation on collective features can be found in Section 2.2.1.

1.3 Thesis objective

Football clubs can benefit from the automation of the work done by analysts. The improvements in player recruitment and match analysis, give opportunities for a cost reduction and performance increase in comparison to current methods. The most important part of a match analysis is to identify episodes of a match that are useful for the coach. That task is very labour-intensive due to the time spent on manual annotation of first- and second-order events to make episode retrieval possible. Automation of this task can save money and increase the performance of a team. The problem is covered in the following research question:

Main research question: How to retrieve interesting match episodes of football matches from spatial temporal data?

A common problem is the knowledge gap between computer scientists and football experts which led to a lack of application of research to practice [31][48]. One of the causes is the lack of specific football knowledge of scientists [11] which makes it difficult to understand the needs of a football expert. On top of that, coaches have different football philosophies. Therefore, it is necessary to define the information needs to understand which episodes are interesting and relevant for experts. This can be illustrated with a practical example: A coach is interested in the answer to the question: *How does the opposition organizes their build-up?* A computer scientist might not know what is meant with build-up. Moreover, he probably does not know what aspects of the build-up are interesting for the coach because it can be related to positioning, player movements, passing decisions, or something else.
Furthermore, some information needs with respect to episode retrieval are satisfied by the current state-of-the-art literature. We aim to address information needs that are not satisfied to give a novel answer to the main research question. That means that a research gap has to be identified. Hence, we want to limit the definition of the information needs to the scope of the identified research gap. This is covered in the following sub-question:

Sub-question 1: What are the information needs of football experts?

After determining the information needs, we want to investigate how the information needs can be satisfied with the use of spatial temporal data. This is usually done using an information system. An information system can be seen as a system that consists of the components hardware, software, data, people, and process to turn data into information [9]. The process component is a high-level component that can be decoupled from implementation details. This is relevant because analysts use different software applications for their work. Hence, a solution that deals with the process component of an information system is more likely to be adopted by analysts in comparison to the other components. A solution for the process can be provided using a model, as it can serve as a blueprint for implementation [22].

So, the objective is to design a model for episode retrieval to satisfy the information needs. The most important design choices are related to the model type and information retrieval methods. The information needs have to be taken into account for these choices. This is all covered in the following sub-question:

Sub-question 2: How can we model episode retrieval from spatial temporal data?

1.4 Main results

The main contribution is a model that is able to retrieve episodes from spatial temporal data. The model is formally introduced in Chapter 4.
To the best of our knowledge, the model is the first attempt to systematically retrieve episodes with second-order events using a formal model. It means that complex patterns of collective behaviour can be explained using the retrieval methods of the model. These methods, based on Temporal Query [4] and Query by Semantic Example (QBSE) [41], are optimized for episode retrieval. Temporal Query is extended with Temporal Aggregation (TA). The methods are further explained in Chapter 3.
The utility of the model is supported by an experimental study. A series of experiments on real match data shows the practical utility of the model. These experiments can be found in Chapter 5.

1.5 Approach & outline

The problem explained in Section 1.3 has been solved using the Cross-industry standard process for data mining (CRISP-DM) [60]. The CRISP-DM is an iterative process of the steps (1) Business understanding, (2) Data understanding, (3) Data preparation, (4) Modelling, (5) Evaluation, and eventually (6) Deployment. The structure of the thesis follows the steps of the CRISP-DM methodology.

The first step, business understanding, was conducted with the help of experts and scientific literature. Several interviews were held with experts from the second professional football league in the Netherlands. The results of the interviews were used to understand the workflow and to establish the information needs of the experts. In addition, the current state-of-the-art literature was used to identify research gaps and limitations with respect to episode retrieval. This is further discussed in Chapter 2.

The data understanding step has the goal to get familiar with the data and to derive the first insights. During this step it was validated if the research question could be answered using the data. Most work with respect to this step is omitted, because of non-relevance. An explanation of the data itself and the relation between an episode and the data is given in Section 4.1

The next step in the process, data preparation, covers the feature extraction from the spatial temporal data. During this step raw sensor data is converted to useful features that can be used for model development. Section 4.2 covers the relevant information with respect to this step.

The modelling step was divided into two parts. The first part covered the translation of information needs to design principles. This is explained in Chapter 3. The second part was the formalization of the model. The result is introduced in Chapter 4.

The last step was the evaluation. For the evaluation of the model, the relevance was measured during an experimental study. Relevance was measured both quantitatively and qualitatively. Further explanation of the evaluation metrics is given in Chapter 4.5. The experimental study was conducted using an implementation of the model with real spatial temporal data. The results of the evaluation are presented in Chapter 5.

Chapter 2

Background

The background chapter aims to introduce the information needs of the experts with the goal to answer sub-question 1. The information needs are identified by establishing the workflow of the technical staff. This is done by using interviews with domain experts that analyze matches professionally. In addition, relevant literature is studied to see how the current literature relates to the findings of the interviews. The results are explained in the next Section 2.1.

With the knowledge of the workflow, the current literature has been analyzed to identify research gaps with respect to episode retrieval and the information needs of experts. The results are explained in Section 2.2. At the end of the chapter, Section 2.3 contains a short summary that presents the relevant information needs that are used for the IR model.

2.1 Workflow of the technical staff

The technical staff is a set of of employees that prepares the players for a match. The staff focuses on everything that is directly related to the game. Typical members of the technical staff are the coach, assistant-coach, physiotherapist, analyst, and scout. The exact number of people involved and the roles they convey differ per club. For simplicity and the scope of this thesis, the technical staff refers to the coach, the analyst, or in some cases a combination of both roles: an expert. That means that all relevant activities are either performed by the expert, coach, or analyst, while in practice it could be a different member of the technical staff.

The period from one match to the next match can be described using a cycle consisting of four phases. These phases are (1) pre-match briefing, (2) match, (3) after-match briefing (4) practice. The same phases can be derived from the analysis workflow created by Stein et al. [54]. The workflow was designed to show the interaction between coach and analyst. The 'Strategic guidelines' transition would be the 'pre-match briefing, 'Football match' is the same, and the 'team briefing' would be the 'post-match briefing'. During every phase the coach has the opportunity to interact with his players to optimize their performance. The role of the analyst is to support the coach. The coach gives tasks to the analyst. After completion of these tasks, the analyst reports back to the coach.

2.1.1 Pre-match briefing

The briefing is a combination of a session the day before and just before the match.

Coach

The coach instructs the players for the upcoming match. This strategy consists of many factors and is dependent on the football philosophy of the coach. Among other things, he informs the

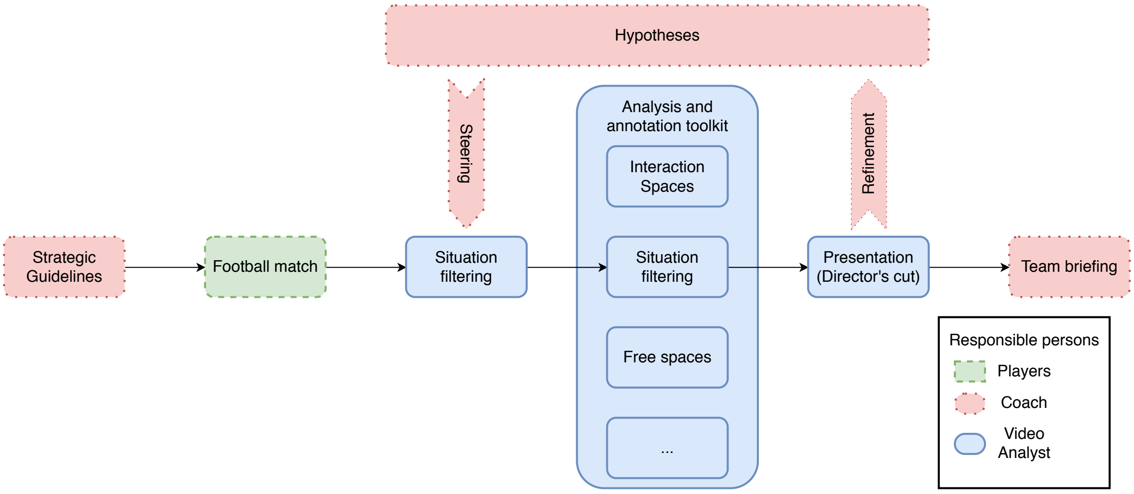


Figure 2.1: Analysis workflow introduced by Stein et al. [54]. The scheme gives an overview of all relevant activities. The different activities are represented with a color to indicate the responsible person

players on who will be playing and which formation will be used. Other aspects that are commonly addressed are playing style, pressure strategy, and standard situation agreements. An important part of the briefing is the strategy of the opponent on all these aspects. Based on the information from the analyst(s), the coach illustrates the strengths and weaknesses of the opponent. He tells the players how to overcome the strengths and how to take opportunity of the weaknesses. During the last part of the session, the coach prepares the players mentally, for example with a motivational speech.

Analyst

The analyst prepares an analysis of the next opponent. Tactical information is conveyed, such as expected formation, playing style, standard situation approaches. Also, he identifies the strengths and weaknesses of the opponent. Typically, the information is supported with video material of the opponent.

2.1.2 Football Match

The coach

When the whistle blows for the start of the match, the coach instructs the team from the sideline. These instructions are directed to individual players with the goal to correct the position or to motivate a player. There is a break after the first half. During the break the coach can instruct the whole team and adjust the strategy. In some cases, the analyst has prepared some episodes that are shown to the players with the goal to support the story of the coach. In most cases this is difficult due to limited time for preparation. After the break, the coach continues instructing the player from the sideline till the end of the match.

Analyst

The analyst records the match with his own equipment, because video material from the television is limited in the sense that it does not always has all the players in the picture. An overview of all the players is essential to extract information with respect to the positioning of the players. Annotation is necessary to derive the metrics and can be used for episode detection. Analysts either receive annotated match data from a data collection company (Opta, Bepro11, etc) or they manually annotate the match themselves during the match. The latter is much cheaper and for that reason clubs choose to do the annotation themselves [56]. It should be noted that manual annotation is a very labour intensive task. There are two approaches for manual annotation, live annotation and post annotation. The former is done during a match and takes 1.5 hours and is mainly used to annotate first-order events. Manual annotation is prone to errors which reduces the quality of the annotations [4][57]. Most errors are the consequence of either under-annotation or over-annotation. Another problem is that the events are annotated asymmetrically from the perspective of the analyst. To illustrate this, we use the example of a cross of team A that was intercepted by a player of team B. The annotator of team A is likely to annotate an event with the description 'a failed pass from player X4 to X5", while the annotator of team B is likely to annotate an event with description 'an interception made by Y6". Both analysts annotated the same event differently. During the interview with the analyst, this problem was confirmed. He indicated that when he received the annotated events from a colleague he had to interpret them reversely to identify events.

2.1.3 After-match briefing

Coach

One or multiple days after the match, there is the after-match briefing. The analyst prepares an analysis for the coach. Based on the report, the coach gives a presentation. He shows his players what went wrong and how it can be improved for next time. In some cases, players receive individual footage on their individual performance.

Analyst

The analyst uses his footage and annotation to make an analysis of the played match. Very similar to the 'Pre-match briefing' phase, he reports the performance of the players to the coach. The focus of this analysis is to assess what went well and what went wrong during the match. At some clubs, the analyst also prepares individual footage to see how they can work on the improvement of their individual abilities.

2.1.4 Practice

Coach

There are a couple of days in between the last match and the next match. During these days, the coach organizes a couple of training sessions. The goal of these sessions is to let the player recover and get fit for the next match. In addition, he wants to improve the strategy. Based on what was discussed during the after-match briefing and the qualities of the next opponent, the coach tailors the exercises to make sure the players are ready for the next match.

The analyst

The analyst starts with the preparation of the pre-match briefing for the next game. The most important part is the analysis of the next opponent.

2.1.5 Interaction between the analyst and the coach

Interaction with the coach is very important for the work of the analyst, because the work of the analyst has to align with the football philosophy of the coach. So, the coach and the analyst have to communicate to make sure that the analyst delivers what the coach wants. Stein et al. [54] illustrated the interaction with the analysis workflow as displayed in Figure 2.1. The interaction takes place during the 'Practice' phase and is illustrated with the 'Refinement' and 'Steering' arrows between hypotheses (Coach) and the analysis (Analyst). Essential during the interaction is that the analyst understands the information needs of the coach. The information need is based on the performance and football philosophy of the coach and can be captured with questions. It is important to realize that there is no one-size-fits all in terms of information needs. However, in most cases the coach is interested in tactical and strategic aspects, such as playing style, pressure strategy, and standard situation agreements.

2.2 Related work

In this section, relevant research to the problem described in Section 1.3 is discussed. The goal is to show what is already done, identify challenges in the field, and introduce research gaps. The scope of the literate review is limited to the field of PA in football and focused on aspects that are relevant for episode retrieval and the IR model. The review covers the most relevant work on Query types, spatial temporal features from tracking data, and software applications related to episode detection.

2.2.1 Query types

There are several papers published on the retrieval of information from spatial temporal data. In this section a selection of relevant query types is presented. At the end of the section, an overview of all papers and their characteristics is given. That overview is used to introduce the research gap.

Basic Query

Mortensen et al. [36] created an application that can be used to retrieve situations using basic queries from tracking data. Examples of basic queries are:

- *all situations in game D where player "X" is in the opponents 18-yard box*
- *a player (X) runs faster than 10 meters per second*
- *a striker (X) is closer than 3 meters to the defender (Z)*

The relevance of the work is difficult to judge, because their work only showed the use of basic queries as a showcase of temporal data. The work of Andrienko et al. [3] and Stein et al. [57] is more relevant, because they used basic queries for episode retrieval. Andrienko et al. [3] used features, such as ball status, possession, absolute or relative spatial, positions of the players and/or the ball, and their movement characteristics to retrieve episodes. In a follow-up paper they extend their work with a feature for pressure [4].
The work of Stein et al. [57] is even more advanced, because it could identify more advanced situations using a combination of basic queries. A Graphical User Interface (GUI) is used to make a combination of simple events (basic queries) using the different options of Allen's interval algebra.

Temporal query

A temporal query allows for the filtering of a time series or temporal database. There are 2 common approaches for a temporal query: (1) by use of a continuous-time interval or (2) setting positions within a time cycle. Andrienko et al. [4] argued that these methods are inadequate for PA, because there is a need for a temporal query based on time-dependent features. To tackle this problem, they introduced a 'time mask' filter that hides intervals in which the (basic) query conditions do not suffice. An impression can be found in Figure A.2.

Visual query

The application of Shao et al. [50] showed the ability to search for situations using a visual query language. Users could sketch a situation of interest that is used to find similar situations based

on a local similarity. To find the situations, they first applied three different filters: (1) start- and endpoint of the candidate trajectories has to be close to the one of the sketches, (2) the length of the candidate trajectories has to be similar to the length of the sketch trajectory, and (3) candidate trajectories should fall in a bounding box with respect to the sketch. For these three filters, the size of the margin of error is a parameter. After the preprocessing, a feature-based similarity score was used to find the minimum distance between the user-sketch and the candidate trajectories. The first feature represents the spatial distribution and the second is a characterization of the general structure. The distance was calculated using the Levenshtein distance. They showed several example queries that could identify corner kicks and side attacks.
Both Belguinha et al. [8] and Richly [43] used something similar to identify episodes. They both developed a graphical query language inspired on the tactics boards to identify episodes. A combination of events can be set to model complex situations. The application of Richly [43] is displayed in Figure A.1a.

A combination of both the sketch-based and tactic board approach was created by Stein & Janetzko [55]. Using the application, an analyst could sketch a trajectory and add the occurrence of certain events and/or a certain amount of pressure. Then, the similarity is calculated using extended Euclidean distance where the 'normal' method for calculating the Euclidean distance is extended with additional features, such as speed. For every feature, a weight is specified to make the query customizable to the preferences of the expert. After the calculation, the application returns the most similar situations with respect to the input. Figure A.1b illustrates the application.

Query by semantic example

All query methods introduced are examples of Semantic Query. Problems with semantic query types are that they can be imprecise or that users have a vague notion of the information they want [28]. These problems can be tackled using Query by Semantic Example (QBSE). The idea of QBSE builds upon Query by Example (QBE). QBE is a query method where the user can enter commands, example elements, and conditions. Afterwards, the query is converted to a database manipulation language form, such as SQL. QBSE makes use of a semantic vocabulary where every element of the vocabulary is linked to an observation. When a query is executed, a similarity score is used to determine the closest neighbors of that observation [41].

Overview

Table 2.1 gives a complete overview of all selected methods for episode retrieval from a spatial temporal database. A couple of observations can be derived from that table.

If you take a look at the column 'Query methods', you can see the different methods used for episode retrieval as earlier introduced in this section. Investigation of the implementation of these methods gives the striking result that most methods rely on queries containing exact positions of players. However, this can be a limitation for the detection of some events, because an expert can be interested in an episode containing events that do not necessarily happen at an exact position. For example, loss of ball possession that leads to a counterattack or moving the ball to an open space. Canales [12] argued the need to recognize events that do not solely dependent on the positional data of the players. Consequently, the motivation behind the use of features without an exact location is introduced.

Furthermore, most papers focus on individual behaviour. Individual behaviour refers to the behaviour of individual players during a match. Individual behaviour can be captured from spatial temporal data with individual features, such as individual trajectories, player speed, and player distance to the ball. First-order events are in general the consequence of individual behaviour.
Collective behaviour on the other hand is referred to as entities that behave in a similar, coordinated, or interdependent way [55]. An example of collective behaviour is the creation of a low block

Paper	Query methods	Individual features	Collective features	Query interaction	Visualization of retrieved episodes	Quanti. eval.* 1st-order events	Evaluation 2nd-order events	Quali. eval.** application
Mortensen et al. [36]	Basic	✓	-	-	-	-	-	-
Belguinha et al. [8]	Basic	✓	-	Filter, sketch	Unknown	-	-	✓
Shao et al. [50]	Basic, trajectory	✓	-	Filter, sketch	Trajectory list	-	-	✓
Richly [43], [44]	Trajectory	✓	-	Sketch	Unknown	✓	Build-up (quali.)	✓
Andrienko et al. [3], [4]	Basic, temporal	✓	Pressure, Team Centroid	Filter, horizontal graph	Unknown	-	Gegenpress. (quali.)	✓
Stein et al. [55] [57]	Basic, temporal, trajectory, QBE	✓	Pressure	Filter, feature weights, advanced sketch	event and trajectory list	✓	-	✓

*Quantitative evaluation of, ** Qualitative evaluation of

Table 2.1: Different characteristics of episode retrieval methods for spatial temporal data

after the loss of ball possession where all players collectively fall-back to the defending third to make the 'field' compact. Collective behaviour can be captured using collective features, such as Team Centroid and Team Length. second-order events are in general the consequence of collective behaviour.
Although the work of Andrienko et al. [3] and Stein et al. [55] used some collective features, the focus on the use of individual features can be derived from the table. Moreover, very limited attention is given to the use of individual or collective features to identify episodes with second-order events. Only Andrienko et al. [3] and Richly [43] did a qualitative evaluation of episodes with respectively Gegenpressing and Build-up. Therefore, the use of collective features can improve and extend episode retrieval.

This observation was confirmed by Goes et al. [19]. He stated that less attention is given to the aggregation of spatial temporal data to interpretable spatial temporal features. These spatial temporal features are able to capture the advanced dynamics of the tactical layer. These features resemble what we call collective features. In addition, he stated that temporal aggregation of collective features is a key challenge that is needed to understand collective behaviour in a temporal context. He suggested that temporal aggregation should be done on smaller time windows or in the best scenario on event-based time windows. It means that the division of ball possession or average Team Width over 90 minutes has limited information gain, while the Team Width during the build-up phase can be interesting. Hence, the use of collective features can help to understand tactics and other forms of collective behaviour during short episodes.

2.2.2 Spatial temporal features from tracking data

Features retrieved from spatial temporal data are called spatial temporal features or time-dependent attributes. They are used for several purposes, such as basic queries or classification of events using Machine Learning. For this project, the spatial temporal features serve the purpose of making temporal queries and enable similarity search. In this section, a selection of spatial temporal features are presented that could serve this purpose. The following criteria were used: (1) Features are time-dependent, (2) features can be retrieved from spatial temporal data, (3) features are numeric, and (4) features can be understood by a football expert with a limited amount of computer science knowledge. An overview of the identified features can be found in Table 2.2.

If you look at these features, it can be derived that most of them are distance based and collective. This seems in contrary to the statement earlier made that the use of collective features is limited, however the use of collective features for episode retrieval is limited. Most papers of Table 2.2 use collective features for episode identification or to explain team performance. The difference between episode retrieval and episode identification is that with the former specific arguments are given as an input to retrieve the desired episode. With the latter, no or limited input is given and the characteristics of the episodes are visualized. In most cases the values are displayed on a horizon graph. An example of a horizon graph can be found in Figure A.2.
It is argued that the variability of these collective features is low in some cases [34]. Ideally, a set of features covers all variability in the data in a way that all episodes can be explained. Hence, a balance is necessary between the number and complexity of features for episode retrieval. The presented features were in most cases used to show the potential of features with respect to performance analysis. Some showed the use of features for manual identification of interesting episodes, but they only compared the values of the different features with each other or between teams [18][37][34]. Other approaches looked at the number of changes [26] or basic statically indicators for variability in the distribution for the whole match or a long period (15 minutes), like standard deviation, coefficient of variation, root mean-square difference [16][39]. Therefore, it can be derived that very limited attention is given to the use of these features for retrieval of episodes with first-order events and second-order events.

Indicator	Method	Description	Ref.
Team centroid*	Distance	The mean position of the players of one team at a point of time.	[16][18][39]
Length, width*	Distance	The size of the team with respect to x, y or both dimensions	[16]
Team area*	Distance	The covered space of one team, also referred to as convex hull.	[16][18][37]
Pressure	Parametrized formula	the movement of the players towards the team in possession and the ball with the goal to take possession.	[4][29]
Team spread*	Distance	Calculated by taking the Frobenius norm for each point of time from the euclidean distance between each player and his team players.	[37]
Stretch index*	Distance	The mean of the euclidean distance of every player to the centroid.	[16] [39]
Speed	Linear formula	The average speed during a certain interval of the ball or the players	[57]
Pitch control	Parametrized formula	Total area of positional control on the field	[51]

*Goalkeeper excluded

Table 2.2: Collective behaviour features

2.2.3 Software applications

A software application serves the purpose to visualize results and let the expert interact with them. The episode retrieval methods presented in Table 2.1 are all (except for Mortensen et al. [36]) implemented in a software application. They used the application to be able to do a qualitative evaluation with experts to measure the quality of their work. Besides that, it shows the practical value. In this section, applications of the episode retrieval applications are presented together with applications that do not facilitate episode retrieval, but serve as a visualisation tool for episode identification. As earlier explained, the interaction is different between both methods. The applications teach us how experts work in practice and how interaction is facilitated. The remainder of the section gives a short introduction to the most relevant functionality of the selected applications. Several screenshots of the software can be found in Appendix A to get an impression of the applications.

Andrienko et al. [4] implemented the Temporal Query earlier mentioned in a visual analytics tool for movement data. The tool was initially developed for other applications of spatial temporal data. The Figure A.2 shows how the tool can be used for episode detection. The qualitative evaluation showed that the use of the application was too complex for experts, as people of the department of visual analytics had to assist to translate the situations described by experts into queries.

An example of an application optimized for experts is the work of Shao et al. [50] earlier introduced. From Figure A.1 it can be derived that the analyst can draw a trajectory of player movement, then a search is executed to find similar player movement trajectories during a match. The query can be extended with a filter for the occurrence of specific movements, events, or time intervals. They showed the results in form of a list that displays the trajectories make a fast selection possible.

Machado et al. [30] showed the use of a heatmap-based approach to summarize player attributes and changes of tactical formation to analyze the evolution of the match. An impression can be found in Figure A.6.

The application named Soccerstories was designed to extract phases by the analyst using a time-interval filter. Afterwards, summary statistics and a node-link diagram were displayed. The node-link diagram used nodes to represent the players and links representing passes or player moves of a selected phase [40]. The main goal of the application has been to visualize event data. A screenshot of the application can be found in Figure A.7.

Stein et al. [53] introduced a flow-like visualization that visualizes semi-automatically-identified interesting moves and their characteristics as displayed in Figure A.3. The rectangles represent features, such as the number of passes and the number of overcome players for a period of time. Interestingly, they used temporal aggregation to calculate the feature values. The features could be ranked by the analyst. Every stack of rectangles represents a move which is a interval of arbitrary length, starting with the seizure of the ball and ending with a final turnover. The evaluation with experts showed the added value of the application for the detection of dangerous situations, but they also indicated the added value of more in-depth, move-related features.

Janetzko et al. [23] developed a tool (Figure A.4) that can be used for episode identification of both individual and collective behaviour. The workflow has been the following. The expert selects input features. Then, a clustering method is executed to cluster all phases for a selected period. Afterwards, all phases are visualized using a horizon graph. The expert uses the horizon graph to identify episodes of interest. An impression of the horizon graph can be found in Figure A.8. Besides that, the tool has been able to classify first-order events, but an elaborate quantitative evaluation is missing to evaluate the quality of the classification model.

The ForVizor application was developed to analyze spatial temporal data with a focus on the analysis of formations. It could automatically detect and visualize formation changes together with implicit reasons for the change and evaluation of the performance of a formation [62]. The main visualization for episode identification was horizon graph. The user could interact with the tool for further information as displayed in Figure A.5

Paper	Episode retrieval	Episode identification	Episode vis.	Indi. features	Collective features
Andrienko et al. [4]	✓	✓	Horizon graph	✓	Pressure
Belguinha & Rodrigues [7]	✓	-	Unknown	✓	-
Shao et al. [50]	✓	-	List of trajectories	✓	-
Stein & Janetzko [55]	✓	-	List of events, trajectories	✓	-
Machado et al. [30]	-	✓	Horizon graph	✓	Formation
Stein et al. [53]	-	✓	flow-like visualization of features	✓	Number of players with touch the ball
Janetzko et al. [23]	-	✓	Horizon graph	✓	Team width, team length, Back-four formation, opposite players around player
Wu et al. [62]	-	✓	Horizon graph	-	Formation

Table 2.3: Overview of software applications features

A complete overview of the discussed applications can be found in Table 2.3. It can be seen that most applications either focus on either episode retrieval or episode identification. While in practice, the task is a combination of both. Experts want to answer their questions, so they give input for a query. Afterwards, they want to analyze the results to adjust the query to their preferences. This vicious circle is underexposed in the current work. Also, there are a lot of different visualization methods used in the applications. As development went in consultation with experts, it shows the differences in the activities and preferences of experts. However, despite of the different visualizations, the end goal was similar in most cases: To retrieve or identify short episodes and visualize a summary of the attributes.

2.3 Information needs of experts

This chapter clarifies that although the information needs of coaches are very different, they are is in general focused on the tactical and strategical aspects of a game. This can vary from repeated situations (e.g. build-up) to a very specific situation (e.g. a free-kick variant). This observation is supported when evaluating the applications developed by scientists to support experts. All these applications make use of different techniques and have different methods for the visualization the results. This has to be taken into account for the adoption of an information retrieval method. However, all applications deal in some way with identification, retrieval, or visualization of episodes. Therefore, it supports the information needs of experts with respect to episodes in contrast to full match summary information.

The literature review showed that limited attention was given to the detection of episodes with second-order events. Some papers showed how certain feature(s) can be used for the detection of a specific second-order event. However, a systematic model for the detection of episodes with second-order events is absent. Also, limited attention is given to aspects that contribute to the detection of these events. This includes the use of temporal aggregation, collective features, and features that are not strictly tied to an exact location on the pitch.

Chapter 3

Design Principles for Episode Retrieval

This chapter explains the design principles for the development of the model that will be presented in Chapter 4. Design principles are fundamental pieces of advice that were used to guide the development. The design principles are based on the information needs that have been established in Chapter 2. The remainder of this chapter introduction explains how the information needs have been translated to design principles.

One goal of the model is to reduce the amount of time necessary for match analysis by an analyst. That means that the analyst spends less time on an analysis task with respect to current methods. One of these tasks is the retrieval of episodes with second-order events. The results of the literature review showed that limited attention was given to this activity. Therefore, the objective is to decrease the amount of time necessary for retrieval of these episodes.

Episodes with second-order events are described by complex patterns in spatial temporal data. These patterns can be retrieved with features that capture complex patterns. Besides that, these features have to be understandable for an expert. To cover both requirements, the model has to incorporate collective features. Moreover, the spatial information within a collective feature has to be limited because the patterns of second-order events are seldom characterized by exact spatial aspects.

Another requirement is that experts can interact with the results. Hence, the expert has to be able to understand how the model works. An understandable model can be explained as a model that can be understood without a computer science background. Besides that, for every design choice of the model the complexity has to be balanced with the added value. For example, a complex method is only favoured for a less complex method when there is a significant increase in utility.

Another requirement for interactivity is the speed of information retrieval. Google developed the Time to Interactive (TTI) metric to indicate the speed of a webpage. TTI represents the time between loading a web page until the moment it can immediately and reliably respond to user input. Google indicates that a TTI below 3.7 seconds is fast, below 7.3 seconds is moderate, and above is slow. As a result, an implementation of our IR model has ideally a TTI of less than 3.8 seconds, but not higher than 7.3 seconds. Although the TTI depends on many implementation details, there are a couple of design choices that affect the TTI. One important factor is the time necessary to retrieve the results. To facilitate a low TTI, the amount of time necessary to retrieve the results has to be as short as possible. From a high-level perspective, the algorithmic complexity of a retrieval method is important. In addition, the nature of the methods plays a role. Online algorithms can provide a solution in real-time and are easier to compute in parallel. Therefore, the query methods have to be categorized as online algorithms with the aim to have the lowest time complexity as possible.

Lastly, to ease adoption it is necessary that the model can be implemented in new or existing applications. That means that the model should focus on high-level components. Moreover, incorporation of low-level aspects has to be avoided, such as implementation details and platform dependencies.

The remainder of this chapter is used to explain which design choices are made to meet these requirements.

3.1 Model selection

The main objective of this project is to create a model for episode retrieval. Episode retrieval is a form of Information Retrieval (IR), because of the following reasoning: *IR is the science of searching for documents or information in documents* [20]. Documents can be any form of multimedia, which are episodes in this study. In addition, IR is characterized by the use of unstructured data and queries that are formed using keywords without a complex structure [20]. The unstructured data source is spatial temporal data. Although, the query methods that will be introduced in Chapter 4 are more complex than a typical IR system, the query methods rely on a combination of domain knowledge and semantic features. Also, the results of these queries can be imprecise, because it does not necessarily return what the intent of the expert was. For example, when the intent of the expert is to find episodes that contain a corner kick, it could occur that it returns episodes where a long thrown-in takes place close from a position close to the corner flag. These episodes are very similar, but could be less relevant for the expert.

To meet the last requirement, The ability to implement the model in new or existing applications, the model has to be decoupled from implementation details and platform dependencies. The IR model of Baeza-Yates & Ribeiro-Neto [6] is a suitable framework to satisfy this requirement. Besides, the model consists of a document representation, query representation, framework, and ranking function which all can be found in a typical IR model.
To be more precise, the model is a boolean retrieval model or vector space model dependent on the query method used. These two types are simple and support the data types. Other classical IR models were considered inadequate because of the following reasons. A probabilistic model can not be used, because no prior knowledge is available and language models are considered too complex and difficult to apply to spatial temporal data. Instead, it was decided to add the possibility for manual weights of the features during ranked retrieval. The model will be formally introduced in Chapter 4.

3.2 Document representation

The first component of the model represents the document collection. This study deals with episodes that have to be retrieved from spatial temporal data. As a consequence, the formal definition of an episode is based on the characteristics of spatial temporal data.

3.3 Feature selection

The second component is the Query Representation. The Query Representation is partly determined by the features. A selection of features has been made based on the results of the literature review. The following conditions are still relevant: (1) Features are time-dependent, (2) features can be retrieved from spatial temporal data, (3) features are numeric, and (4) features can be understood by a football expert with a limited amount of computer science knowledge. In addition, we want features that are able to capture collective behaviour of a team. The spatial information within these features has to be limited.

Based on these criteria, a list of candidate features was made to test whether the features are related to each other. As a consequence, the candidate features Team area, Team spread, and Stretch index were removed, because the correlation with other candidate features was too high. As a rule of thumb, a Pearson correlation above 0.70 was considered too high.

3.4 Choice of query methods

For a query method it was a requirement that it could deal with the Document and feature representation. In addition, the methods have to be understandable for an expert and facilitate interaction. For the latter it is necessary that the expert can adjust a query based on the retrieved results. The results have to be retrieved in a reasonable amount of time.

The most important factor is the ability of the query method to satisfy the information needs of the expert. In general, the information need is either (1) known-item search, (2) Existence search, (3) exploratory search [46]. The last option is out of the scope of the IR model, because it belongs to episode identification. The first two are relevant, as in both cases the expert searches for an episode based on his own input. In case of a known-item search, the user knows clearly what he is looking for. For example, the coach wants to show an episode during the break of a situation of the first half where a mistake was made. In this case, the expert has a particular item in mind. A suitable retrieval method for a known-item search is exact search. Retrieval methods are a form of exact search return results that are either relevant or irrelevant. This is in accordance with the information need. It is important to consider that the particular item does not necessarily have to present in the data [27].
The other information need is existence search or answer seeking. In this case, the user knows what his search intent, but does not know if the answer exists or how to describe it. An example is the question: *"How does the opposition organizes their build-up?"*. In this case the user is looking for episodes that contain the build-up of the opposition, while he does not actually know what the build-up looks like. A suitable method for existence search is ranked retrieval, as it sorts the documents on the criteria that are most relevant according to the user.

The first method, Temporal Query, has been considered for the ability to capture first-order events [57]. This method is a form of exact search and can be used for known-item search. Another motivation has been that the method is comparable to the filter functionality in search engines. Experts are familiar with the technique. Moreover, the possibility of temporal aggregation and the extensiveness of Allen's interval algebra to set relations between conditions makes the method very expressive. Therefore, it has to be able to capture complex patterns that are needed to capture episodes with second-order events. Besides, Temporal Query supports the possibility to retrieve domain knowledge, because it supports categorical variables.
On the other hand, the extensiveness of Allen's interval algebra can be a problem, as it facilitates the possibility to make very complex relations between conditions. This can be computationally expensive to search for a solution, since it takes exponential time to check for consistency. As a consequence, it was decided to limit the problem to singleton labelling. Singleton labelling makes it possible to check consistency in polynomial time. In addition, it tremendously decreases the complexity of the problem. It means that for any two intervals, there is only one basic relation. This can be illustrated with the idea of a timeline. When all singleton relations are consistent, it is possible to place all intervals on a timeline [59].

However, there are a couple of disadvantages to the Temporal Query method. Writing a temporal query can be too difficult or effortful. additionally, the number of retrieved results can be too much or too few. Moreover, the expert is not always able to convert an information need to an exact Temporal Query, because the expert does not always no how to describe what he is looking for.

To overcome these issues, QBSE is introduced. It extends the expressiveness of experts to better cover their information needs, because it is a form of ranked retrieval. The expert can simply provide an example of his information need. Optimization of a query is facilitated using the possibility to set custom weights for the different features.
The method is also very fast, because the results can be retrieved in polynomial time.

Chapter 4

Formal Information Retrieval Model

The chapter introduces the formal information retrieval model to retrieve episodes from tracking data based on the design principles explained in Chapter 3. The IR models follows the definition of Baeza-Yates et al. [6]:

An IR model is a quadruple $[D, Q, F, R(q_i, d_i)]$ where

1. D is a set of composed logical views (or representations) of the documents in a collection.
2. Q is a set of composed of logical views (or representations) of the user information needs. Such representations are called queries.
3. F is a framework for modeling document representations, queries, and their relationships, such as sets and Boolean relations, vectors, and linear algebra operations, sample spaces and probability distributions.
4. $R(q_i, d_i)$ is a ranking function that associates a real number with a query representation $q_i \in Q$ and a document representation $d_j \in D$. Such ranking defines an ordering among the documents with regards to the query q_i.

The remainder of the chapter follows the structure of the quadruple components to introduce the model.

4.1 Episode representation

The term D represents a logical view of the documents in a collection. In this study, it deals with episodes from Tracking data. Experts often refer to an episode as a fragment, situation or phase during a football match that typically last between 5 and 30 seconds. In this section, a formal definition is introduced using a combination of basic concepts retrieved from spatial temporal data.

Spatial temporal data is a collection of snapshots with the 2D-location (X, Y) of all 22 players and the location of the ball (X, Y) with a typical frequency between 10 and 25 Hz. In addition, the exact date and time of an observation is given. So, the smallest unit is a set of x, y coordinates. Both the ball and the players have a set of coordinates for every moment in time from the start of the match. Every player is part of a team. For every match there are two teams. For simplicity reasons it is chosen to refer to the 'Home' team and the 'Away' team. The coordinates are standardized such that the 'Home' team plays towards the negative X-axis during both halves and the 'Away' team plays towards the positive X-axis during both halves. The coordinate system

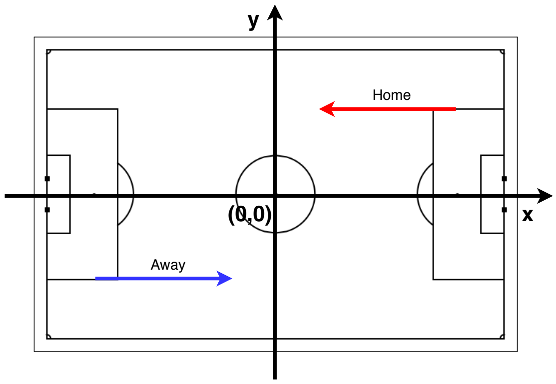


Figure 4.1: The coordinate system with respect to the football field and the point (0,0) at the center spot of the field. The coordinates are corrected in such a way that during both halves the 'Home' team plays from the positive x-axis to the negative x-axis and the 'Away' the other way around.

is illustrated in Figure 4.1. If the coordinates of all players of both team are combined with the coordinates of the ball, it is called a frame. An illustration of all information in a frame can be found in Figure 4.2.

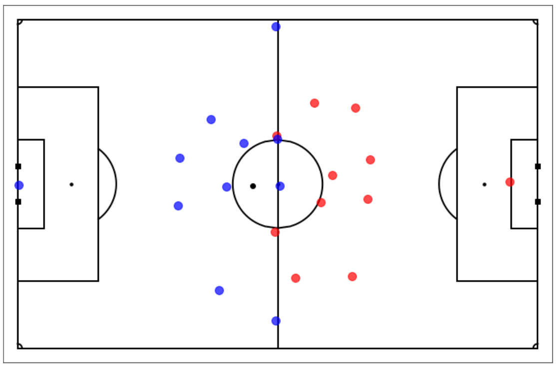


Figure 4.2: A snapshot retrieved from spatial temporal data. The red circles represent the Home team, the blue circles represent the Away team, and the black circle represents the ball.

The combination of the information of two consecutive frames or more is what we call an episode. It is possible to view an episode as a video when you display the separate frames very fast after each other. However, an episode that consists of two frames would take 0.08s which gives limited understanding. For that reason, the semantic episode is introduced. A semantic episode is a episode where at least one event occurred. An overview of all concepts with formal definitions can be found in Appendix B.

4.2 Query representation

Q represents the user information needs. These information needs are different from the (informal) information needs established in Chapter 2. Here, the information needs are formalized using a query representation that approximate the informal information needs as close as possible. A query is a way to request information from a database, which is spatial temporal data in this study. The request is composed using features retrieved from the data. This section is limited to the introduction and explanation of these features. The query methods itself are introduced in the Section 4.4.

4.2.1 Feature extraction

The features can be derived from spatial temporal data. Inspiration for the selection of features was taken from the overview presented in Table 2.2. In addition to the candidate features of Table 2.2, the Relative Occupancy Map is introduced. These features have been adopted from Basketball analytics, as they cover collective behaviour with limited spatial information. For some of the features, there is a separate feature for both the 'Home' team and the 'Away' team. If this is the case a ✓ is set in the column 'Home/Away'. For example, the basic feature 'Team centroid' differs for both 'Home' team and the 'Away' team.

Basic	Window	Keeper excl.	Home/Away
Ball location (Zone)	Sum of differences	-	×
Ball speed	Average	-	×
Ball possession	-	×	×
Players location -	Team Distance (cumulative)	✓	✓
Team centroid (Zone)	Average, sum of differences	✓	✓
Rest defence	Average, sum of differences	✓	✓
Width of the team	Average, sum of differences	✓	✓
Length of the team	Average, sum of differences	✓	✓
Relative Occupancy map	Average, sum of differences	✓	✓
Events	-	-	✓

Table 4.1: Overview of the basic and window features

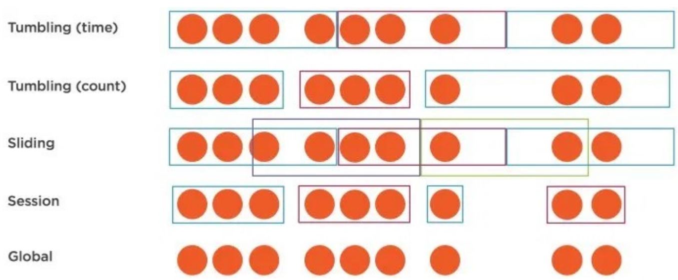
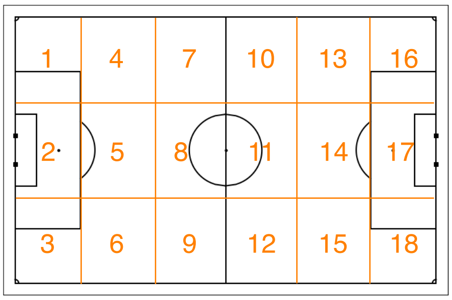


Figure 4.3: The different type of grouping elements (windows) in a stream. [38]

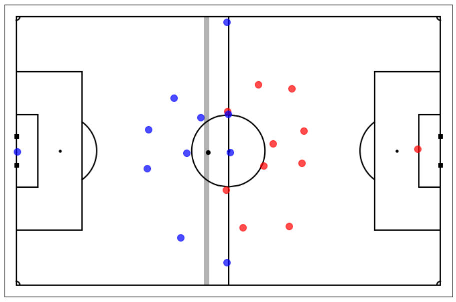
The calculation of features is based on the information of a single frame. Most features are explained in Figure 4.4 using illustrations. This is with the exception of Ball possession, ball speed, trajectory distance, and event. Ball possession was calculated on the assumption that the team is in ball possession when his player is closest to the ball. Possession of a team starts when the ball goes into the field and moves. Then, the team with the player closest to the ball is in ball possession. Transition takes place when a player or multiple players of the other team are closer to the ball then any other player from the other team for at least 2 consecutive seconds. This only holds when the closest distance of a player to the ball is smaller than three meters.
Ball speed was calculated based on the method of Laurie Shaw. He calculated the speed using the positional information of the ball at frame t_{-1} and t. Afterwards, he applied Savitzky-Golay smoothing with window size 7 and polyorder 1. The trajectory distance is calculated similar to the method of ball speed without application of smoothing. That means the distance is calculated by taking the difference between the positions at frame t_{-1} and t.

The last feature, Events, is a bit different from the other features, because it can not be directly derived from the spatial temporal data. Events represents either a first- or second-order event as defined in Section 4.1. There are two methods to add an event to Events. The first method makes use of manual annotated event. This data can be processed and added to the feature storage space. The other way is to derive the events from spatial temporal data using features. That means an event has to be defined using one or multiple features. For example, we have an event 'Out' which means that the ball is not on the pitch. If the x-coordinate of the ball is greater than the boundaries of the pitch, the ball is not on the pitch. Then, we can add the event 'Out' to the corresponding frame. The advantage of this approach is that it improves the IR model with respect to efficiency. Moreover, it gives the analyst the opportunity to generate a vocabulary of events based on personal definitions.

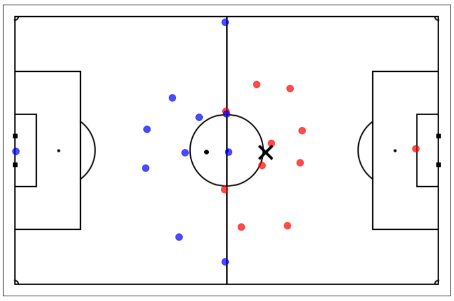
The presented features are so called basic features. It means that they represent information of a single frame. The information of multiple frames can be captured using temporal aggregation on windows. Windows are in most cases used to analyze data streams. For a stream of data, the data is first ordered on a certain characteristic. In most cases this is based on the timestamp that indicates the exact moment an observation occurred. Then, the stream is divided in smaller parts (windows). The different type of windows are illustrated in Figure 4.3. For every window, the features are aggregated. Aggregation is a method to group a series (window) of observations. Grouping can be done using aggregation functions. An overview of the aggregation functions can be found in Table C.1. For example, the average centroid of the Home team can be calculated for all observations within a window. Some basic features are aggregated using the sum of differences. It means that the differences between consecutive frames within window are summed. This is similar to the calculation of the trajectory distance of the ball earlier explained. This has the advantage that positional aspects are removed from the features. Take for example the Team centroid. When taking the sum of differences of team centroid, the value only indicates the direction and distance covered during the window, but it does not explain the exact position.



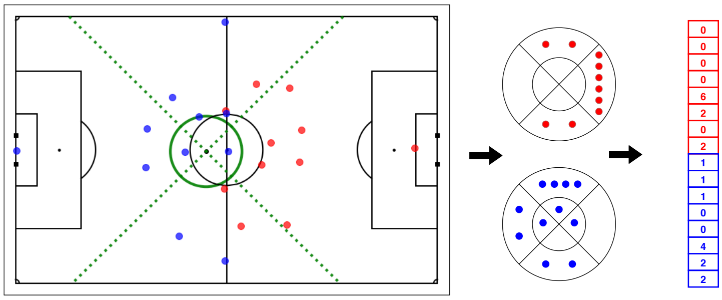
(a) **Zone** - The pitch is divided in 18 different zones. These zones are well-known by experts and help to express themselves. For example Zone 14 is known as the 'Golden Square' and zone 13 till 18 refers to the attacking third (zone 13 till 18) from the perspective of the team playing from left to right.



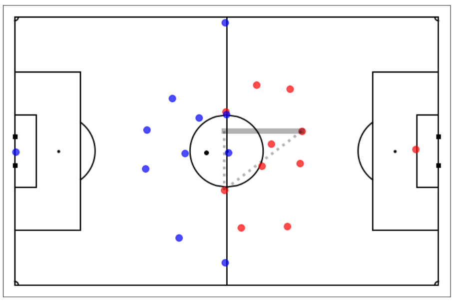
(b) **Rest defence** - The rest defence is determined by the total number of players behind the ball. The grey line illustrates the border. So, for the team playing from left to right, the rest defence is 6 (Goalkeeper excluded). For the other team it is 10.



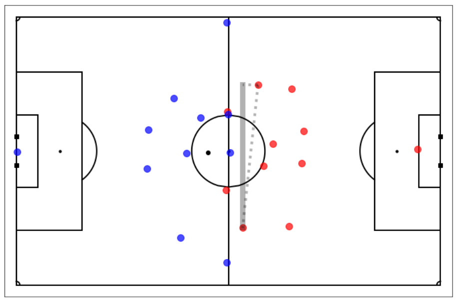
(c) **Team centroid** - Average position of all players of one team. The position of the centroid of the team in red is represented with the black cross.



(d) **Relative Occupancy Map** - The shape context representation is used for the relative distribution of the players. The field is divided into two levels with the boundary set on 9m distance of the ball. Then both levels are divided into 4 directions with a 90 degree angle between. For every part, the number of players are counted, resulting in a feature with 8 values per team.



(e) **Team length** - The team length is the distance with respect to the x-axis between the player with the maximum x-coordinate and with the minimum x-coordinate.



(f) **Team width** - The team width is the distance with respect to the y-axis between the player with the maximum y-coordinate and with the minimum y-coordinate with respect to the y-axis.

Figure 4.4: Explanation of several features using sketches.

4.3 Framework

The F is a framework that shows the relationship between the components of the model. A schematic overview of the model can be found in Figure 4.5. The process can be explained as follows:
At the start, the expert has a question, for example: *How does the opposition organizes their build-up?* The question is converted to a query that should return episodes that answer the question. This can be either a Temporal query, Query by Example, or a combination of both. A query requests data from the features storage space that consists of both the episode collection and features. These are processed at the ranking function. In this study it can be an exact match, similarity ranking, or a combination of both, because it depends on the query type. The ranking function determines which episodes are returned to the expert.

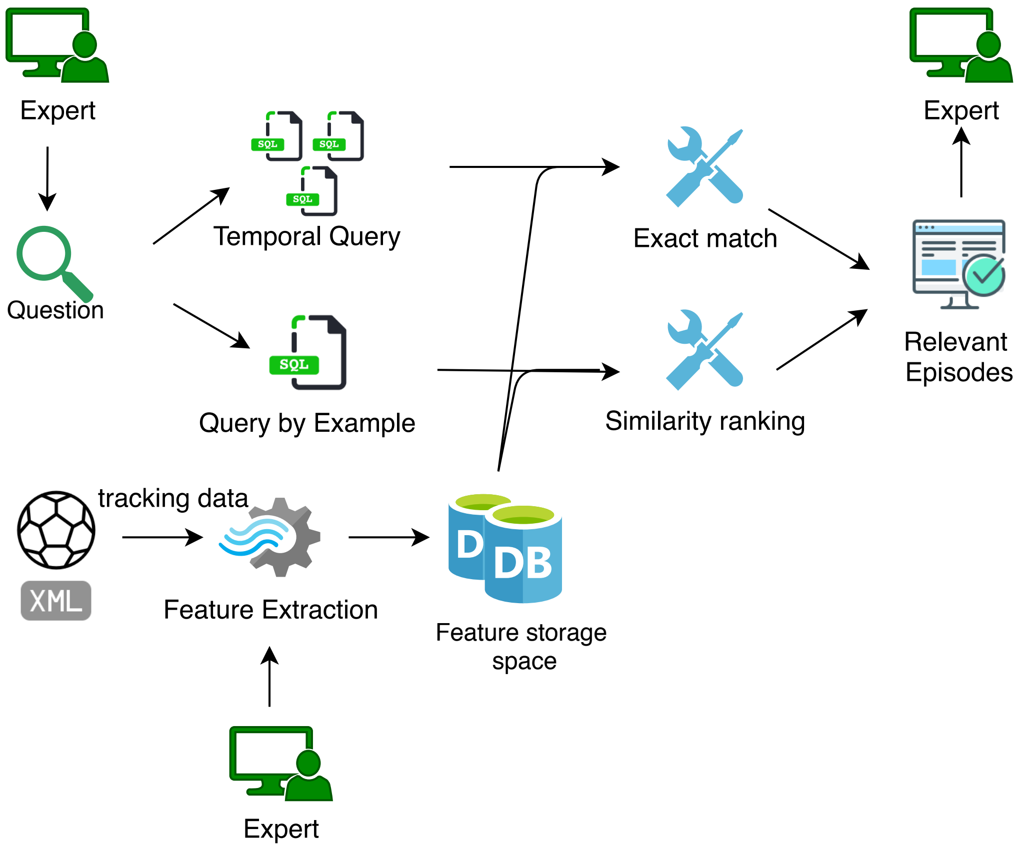


Figure 4.5: The IR model for episode retrieval. Every icon represents an element of the model, the flow between the elements is illustrated with arrows.

4.4 Retrieval functions

The last part of the IR model is the ranking function $R(q_i, d_i)$. The ranking function assigns a rank to an episode based on the query. The term ranking function is somewhat inadequate for this case, since the model is partly a Boolean IR model. For that reason, we will refer to as retrieval function. The model consists of two different retrieval functions dependent on the two query methods, Temporal Query and Query by Example. The former is a way to conduct an exact search and the latter is a method to execute a similarity search that ranks all episodes with respect to the given example. So, an exact search is a form of unranked retrieval and the similarity search

of ranked retrieval. In the case of unranked retrieval, all results of a query are considered equally important because the results are either relevant or irrelevant. With ranked retrieval, the results are ordered on a similarity score. The higher the score, the higher the rank.
Both methods make use of a set of features presented in Table 4.1. Nevertheless, the proposed IR model can be used with other spatial temporal features.

4.4.1 Temporal Query

Temporal query was introduced in Section 2.2.1. It was chosen to use the method of Andrienko et al. [4] where a basic query can be executed using time-dependent features. It was chosen to extend this method with temporal aggregation. A query is composed as follows.

A complete query consists of one or multiple interval expressions. An interval expression consist of one or multiple conditions. Apart from setting the conditions, the expert has to make two choices with respect to time. The first choice is to select a period of interest or a combination of periods. This does not necessarily have to be from a single match. Examples of a period are the first half or the first five minutes after the break. The second choice is related to the TA. The expert has to choose a window method. This can be either an event-based window or a sliding window. An example of an event-based window is possession where a single episode covers the time between gaining and losing possession. In the case of a sliding window, the window size and the sliding gap have to be specified in seconds. For example, an episode with a size of 15 seconds with a sliding gap of 5 seconds.

Afterwards, the conditions have to be set. The expert selects one or multiple features. For every feature he can choose from the available aggregate functions presented in Table C.1 with exception of the Events feature. Only the function COUNT can be set for the Events feature. Then, the expert sets a value using the available operators presented in Table C.2. If there are multiple conditions, the expert has to choose a relation between the conditions. Only one relation can be chosen for every two conditions. This completes the interval expression.

If the expert completes composing all interval expressions, then the complete query can be executed. The query returns all episodes that meet the conditions. A query is executed in the following sequence. The first condition of the first interval expression is checked for all observations in the period of interest. This is followed by a replication of the same step for the next condition. Afterwards, the intersection of intervals is calculated for the chosen relation of the two conditions. For every next condition, both steps are replicated on the remaining observations after the interval intersection calculation.

In Appendix D a couple of examples are given to indicate the expressiveness of the query language.

4.4.2 Query by Semantic Example

The QBSE is a form of QBE where the example has meaning. The expert determines what the meaning of an episode is. In this study, this is formally defined as a semantic episode. The expert wants to retrieve episodes based on a given episode that serves as the example. This can be illustrated with an example. The expert identified a moment where the ball was given to the opposition that led to a counterattack. He wants to know if comparable situations occurred more often during the match. Then, he has to select that situation as a representative episode to search for similar episodes. To perform the query, the similarity between the representative episode and candidate episode is calculated.

Similarity

There are different ways to calculate the similarity between two episodes. Some of them are mentioned in Chapter 2. Among other things, the choice for a similarity calculation method depends on the features used. We chose to take the difference between values or take the Euclidean distance in case of multidimensional features. The choice for the Euclidean distance was taken for several reasons. First of all, during data exploration it was found that the clustering methods based on the Euclidean distance were able to generate meaningful clusters. Secondly, the Euclidean distance is easy to calculate and is therefore understandable for experts. To correct for the different value ranges of the features, the values have to be normalized. Min-max feature scaling is a method to normalize a feature to make sure all values are between zero and one. The formula is displayed in Formula 4.1. If possible, a manual maximum and minimum were taken to generalize the normalization. For example, the world record on the 100 meter sprint was measured with a top speed of 12.27 m/s. Hence, the maximum speed of a player is set to 12.27m/s and a minimum of 0 m/s

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{4.1}$$

The formula to calculate the distance between two episodes p and q is displayed in Formula 4.2 with the feature weight w, n the number of features, and the vectors p_i and q_i. The feature weight is a parameter than can be adjusted by the expert. That means that he can set the relevance per feature. The feature weight is a value between zero and one for every feature and the sum of all weights is equal to 1.

$$D(p, q) = \sum_{i=1}^{n} w_i \cdot ||p_i - q_i|| \tag{4.2}$$

Ranking

After calculation of the similarity score for all candidate episodes with respect to the representative episode, the episodes have to be sorted. To retrieve the results, either a cut-off value can be set for the similarity score or the top N results are returned. In the case of a cut-off value x, only results with a similarity score above x are returned. In the case of the top N results, a value for N is set, then the N results with the highest similarity are returned.

4.5 Evaluation Metrics

Evaluation serves the purpose of measuring the quality of the work and to be able to compare the work of others. As explained in Section 1.3, we want to measure the quality of the retrieved episodes, and thus the quality of the IR model. There are many ways to evaluate an IR model, but in general the key utility for IR systems is user happiness [32]. User happiness is based on many factors, such as relevance, satisfaction, efficiency, usefulness, and interface design.
The scope of the project is limited to show an IR model as a proof of concept. As a consequence, the implementation is not optimized for speed and a graphical user interface is absent. Therefore, it was decided to limit the evaluation to the measurement of relevance. Relevance can be measured quantitatively and qualitatively.

4.5.1 Quantitative relevance

Measuring relevance quantitatively is done differently for unranked retrieval than ranked retrieval. That means that a Temporal Query is evaluated differently than QBSE.

Unranked retrieval

Using the classical way to evaluate relevance for unranked retrieval, the following three requirements are needed according to the 'Introduction to Information Retrieval' book by Manning et al. [32]:

1. A document collection
2. A test suite of information needs, expressible as queries
3. A set of relevance judgments, standardly a binary assessment of either relevant or non relevant for each query-document pair.

These three requirements correspond to the components D and Q of the introduced IR model. The last item, a set of relevance judgements, has to be established. It is important to consider that relevance is measured with respect to the information need and not the query. Relevance is in most cases considered binary which means that an item is either relevant or non-relevant. In an optimal case, a ground truth is available that consists of a collection of evaluations that can be considered trustworthy. Precision and Recall are the most basic and most frequently used methods for unranked retrieval effectiveness [32]. In many cases, the F-measure is also calculated to combine Precision and Recall in one measure. The F-measure is the weighted harmonic mean of Precision and Recall. The formulas used to calculate Precision, Recall, and F are displayed below:

$$Precision = \frac{\#(relevant\ items\ retrieved)}{\#(retrieved\ items)} \tag{4.3}$$

$$Recall = \frac{\#(relevant\ items\ retrieved)}{\#(relevant\ items)} \tag{4.4}$$

$$F = \frac{2 \cdot Precision \cdot Recall}{Precision + recall} \tag{4.5}$$

In the case that there is no ground truth available, the evaluation has to be done manually. Manual evaluation of second-order events can be difficult because of the subjective nature of the items. Therefore, it was decided to change the evaluation from binary to a common four-point likert scale. The following evaluations are distinguished (1) highly relevant, (2) somewhat relevant, (3) somewhat irrelevant, and (4) highly irrelevant. Ideally, a group of judges estimate the items with overlap. That means that multiple judges estimate the same item. Afterwards, the kappa statistic can be calculated. The kappa statistic is a common measure for agreement between judges and can be used for categorical judgements. With the use of the kappa statistic you can derive an impression of the quality of the estimations.
The four-level relevance score makes it possible to calculate Precision and Recall by converting the four-level score to a binary score. This can be either done by setting the threshold after 'highly relevant' or 'somewhat relevant'. In the case of the former, all items that are judged 'somewhat relevant', 'somewhat irrelevant', and 'highly irrelevant are set to 'irrelevant'.

Ranked retrieval

Evaluation for ranked retrieval differs from unranked retrieval, since all candidate episodes are returned when additional steps are absent. In that case, Precision and Recall give no insights. In general, the top-k-retrieved documents are returned for a query [32]. For this set, Precision and Recall can be calculated. This metric is called 'Precision at k'. However, it should be noted that the evaluation metric is less stable, because Precision is highly affected by the total number of relevant documents available [32]. Also, the number of items retrieved plays an important role. A precision-recall curve can be created to show the trade-off. A precision-recall curve shows the average precision for different recall levels. The values of the precision-recall are used to calculate the Average Precision.

Alternatively, the R-precision can be calculated. R-precision is based on the number of relevant documents present in the data. This number determines the number of retrieved items. If there are 20 relevant items in a collection, then the top-20 items for a query are returned. Afterwards, Precision can be calculated which is identical to Recall in this case. However, if the total number of relevant documents is unknown, the R-precision can not be calculated or an estimation of the number of relevant documents has to be made.

4.5.2 Qualitative relevance

Relevance can be evaluated qualitatively. In that case, an expert discusses the results and indicates how useful they are with respect to the intended query. A qualitative evaluation is very valuable in addition to quantitative measures, because it gives a better indication of the practical utility of the IR model. Besides that, it can be used for comparison with other models. Although, it is very difficult to compare qualitative evaluations with each other, a better comparison can be made than one that is solely based on quantitative evaluations. Table 2.1 shows that all episode retrieval methods were evaluated on this matter.

There are several ways to do a qualitative evaluation. Most of them involve users. If no GUI is present, it is difficult to do tests with users. An alternative is the Evaluation of End Products. Evaluation of End products focuses on the outcome or product of the search by verifying if the information needs match the user model behind the search task [25].

Chapter 5

Experimental Study

An IR model is proposed to retrieve episodes from spatial temporal data. In order to evaluate the quality of the model, an experimental study was conducted. As a guideline, the evaluation methods proposed in Section 4.5 were used. The experiments are designed to show the contribution of collective features for the detection of episodes with second-order events. Six different questions are proposed based on the results of the interviews with the experts and the literature review. All questions are formulated in a way a coach could have formulated them. They cover all states of a football match (attack, defence, and transition) with the goal to show the extensiveness of the model. These are the questions:

1. How is the build-up organized during the first 15 seconds after a goal kick?
2. How long does it take before a team switches play?
3. Which locations of ball possession loss lead to a counterattack?
4. How often does Gegenpressing leads to reconquering of the ball?
5. How is my defence organized during an attacking corner kick?
6. What are the weaknesses of a team to break down a low block?

The chapter is organized as follows. First, an introduction is given to the data sources that have been used for the experiments. Afterwards, relevant information with respect to the implementation of the model is given. This is followed by a description of the experiments. The description includes an explanation of the general steps that are taken for all experiments to compose and evaluate queries. The implementation details concerning individual experiments are explained in subsections corresponding to individual experiments. Finally, the results of the experiments are explained and discussed.

5.1 Data

The data used for the development and verification consisted of two different data sources. The first data source was a public dataset provided by the video analysis platform Metrica Sports [35]. The dataset consists of two matches that are completely anonymized. The positions are collected with a temporal resolution of 25Hz. The position of the ball has the same resolution, but the positions could be unreliable. Manual inspection indicated that the location of the ball is unnatural. This could be the consequence of manual annotation. To these matches is referred with 'MET1' and 'MET2'. The second data source consists of two matches provided by the Deutscher Fußball-Bund (DFB). These matches have the same temporal resolution of 25 Hz and are also anonymous. In contrary to the Metrica data source, the location of the ball seems trustworthy. For each game, there are over 3,000,000 observations that represent either player or ball location.

To these matches is referred with 'DFB1' and 'DFB2'.

All experiments were executed on computer using a python implementation of the proposed model. To speed up the feature extraction process, it was decided to apply a date reduction technique. The data was aggregated by taking the average position on a tumbling window with a size of ten frames. That means that for every ten frames, the positions of a player or the ball were aggregated to 1 value (mean). This was applied to all players and the ball.

5.2 Experiments

The experiments are designed to retrieve an answer to the question proposed in the introduction of this chapter. Every experiment will be introduced and explained in an individual subsection. A video has been made for every experiment to support the explanation. A complete overview of all videos can be found here: `https://www.youtube.com/playlist?list=PLK19aH1iN-06jZk-vZOT7bJc8W1KYj0Qy`. Individual links to the videos can be found in Table F.1.

For every experiment it was investigated which events had to be retrieved from the spatial temporal data. Afterwards, the queries had to be designed as both a Temporal Query and QBSE. A description of the Temporal Query can be found in the subsection corresponding to an experiment. Besides that, the parameters for for QBSE with TA were determined. An overview of these parameters can be found in Table 5.1. For all experiments an example was provided for both the "Home" variant and the "Away" variant. An impression of these examples can be gotten from the provided videos. For the variant of QBSE using DTW, the weights were equal because the software library did not support unequal weights for the features. A custom implementation of DTW was developed to overcome that problem, but could not be used because it gave different results during testing in comparison to the trusted library.

Evaluation of the retrieved episodes with first-order events was done using the supplied event data. All events were manually annotated by an analyst. As manual annotation is prone to errors, the retrieved intervals of a query were expanded with 5 seconds before the start and 10 seconds after the end. A retrieved episode was considered relevant if the start of the manually annotated event falls within the range of the episode.
The review of episodes with second-order events was done manually using the procedure explained in Section 4.5. For the ranked retrieval results, the number of retrieved results was based on the number of Temporal Query results. On top of that, it was aimed to have a similar amount of results for the different experiments. Hence, either 20 or 40 results were retrieved for a ranked query.
The quantitative results of the experiments can be found in Table 5.2 and 5.3. Table 5.2 shows the results for the different query methods for episodes with first-order events and Table 5.3 shows the results for episodes with second-order events.

The qualitative evaluation of the experiments is discussed in the corresponding subsection.

5.2.1 Experiment 1: Build-up

During the interview with the head coach, the coach indicated that he always analyzes the build-up of the opponent. Build-up refers to the playing style of a team during the phase after a goal kick or when the ball is passed back to the goalkeeper. For evaluation purposes it was decided to only look at the case of goal kick. It is generally considered that the build-up phase ends when the team reaches the attacking third. The most interesting part for the coach is to analyze the first 15 seconds after the goal kick, because that part is often orchestrated by the coach and thus likely to be executed systematically. While the phase is developing, the choices of the players become

Feature	Build-up	Switch	Counter	Gegen	Corner	Low
AVG Ball velocity	0.052	0.25	0.125	052	0	0
AVG Team Centroid X Home	0.052	0	0	052	0.111	0.166
AVG Team Centroid Y Home	0.052	0	0	052	0.111	0.166
AVG Team Width Home	0.052	0	0	052	0.111	0
AVG Team Width Away	0.052	0	0	052	0.111	0
AVG Team Length Home	0.052	0	0	052	0.111	0.166
AVG Team Length Away	0.052	0	0	052	0.111	0.166
AVG Rest Defence Away	0.052	0	0.125	052	0	0
AVG Inside	0.052	0.125	0.125	052	0.111	0.166
SUM Ball Distance X	0.052	0	0.125	052	0	0
SUM Ball Distance Y	0.052	0.25	0.125	052	0	0
SUM Team Centroid X Home	0.052	0	0	052	0.111	0
SUM Team Centroid Y Home	0.052	0.25	0.25	052	0.111	0
SUM Team Width Home	0.052	0	0	052	0	0
SUM Team Width Away	0.052	0	0	052	0	0
SUM Team Length Home	0.052	0	0	052	0	0
SUM Team Length Away	0.052	0	0	052	0	0
SUM Rest Defence Away	0.052	0	0.125	052	0	0
Roms	0.052	0.125	0	052	0	0.166
Window size [s]	5.2	10.0	10.0	5.2	10.0	10.0
Absolute difference	✓	×	✓	✓	×	×

Table 5.1: Parameters for different queries of QBSE with TA

Temporal query					**QBSE** R-precision	
First order event	No. of events	Precision	Recall	F-Score	Eucledian	DTW
Corner	43	81.3%	90.7%	85.7%	51.2%	-
Goal Kick	76	82.1%	90.8%	86.3%	34.8%	13.6%

Table 5.2: Results of Temporal queries and QBSE for episodes with first-order events

more dynamic and context-oriented. Therefore, a coach wants to inform their players on how to anticipate on those first 15 seconds. So, the following question is composed:

How is the build-up organized during the first 15 seconds after a goal kick?.

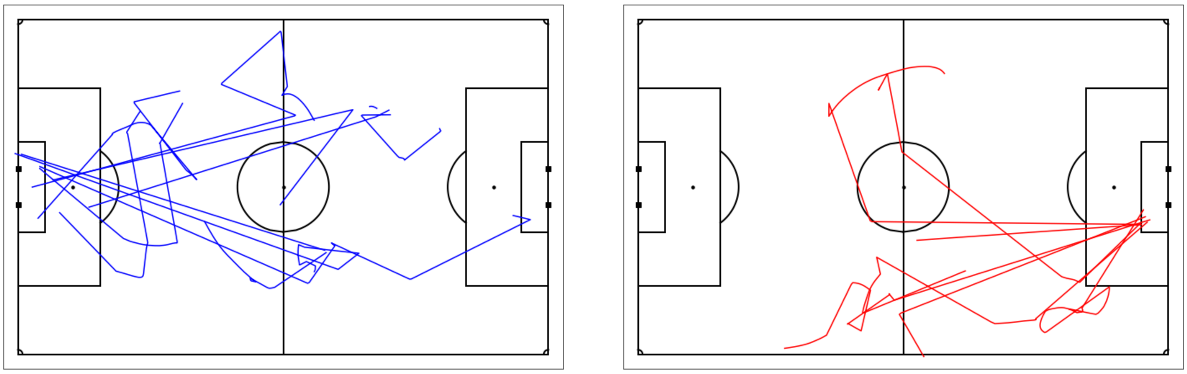
To retrieve episodes that convey an answer to that question, it was necessary to build a query that returns episodes with goal kicks. To see the differences between the methods, it was decided to query goal kicks with both Temporal Query and QBSE. In addition, QBSE was also implemented with Dynamic Time Warping (DTW). DTW was used by Stein et al. [55] to compare trajectory combined events with each other. So, the comparison of our implementation of QBSE with DTW gives a better indication of the quality.

The interval expression that identifies goal kicks from the Home team can be explained as follows. First the ball has to be out of the field. This event is followed by an event where the ball moves in position 17 and a maximum of one player of the Home team is behind the ball, all players of the Away team are in front of the ball, and the Home team is in possession. For the Away goal kick, the query conditions were mirrored over the y-axis. A formal description can be found in Figure E.1.

The second part of the evaluation is to check whether the retrieved episodes can answer the expert question. To do so, the results of the Temporal Query were used. For every result, an

Method	Second order event	No. of events	H. rel.	S. rel.	S. irrel.	H. irrel.	Precision
Temporal	Switching play	50	36	7	5	2	72.0%
QBSE TA	Switching play	40	27	5	8	0	65.0%
QBSE DTA	Switching Play	40	17	5	4	14	42.5%
Temporal	Counterattack	39	20	6	3	10	51.3%
QBSE TA.	Counterattack	40	19	6	0	15	47.5%
QBSE DTW	Counterattack	40	17	7	5	11	42.5 %
Temporal	Gegenpressing	18	13	1	0	4	72.2%
QBSE TA	Gegenpressing	20	6	4	4	6	30%
QBSE DTW	Gegenpressing	20	8	4	2	6	40%
Temporal	Low block	104	101	3	0	0	97.1%
QBSE TA	Low block	40	31	2	2	5	77.5%
QBSE DTW	Low block	40	36	1	1	2	90.0%

Table 5.3: Results of Temporal queries and QBSE for episodes with second-order events



(a) Away

(b) Home

Figure 5.1: Ball trajectories during 15 second episodes after a goalkick.

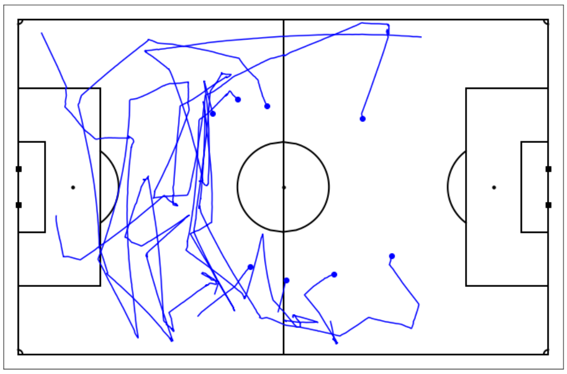
episode was generated with the start at the moment that the ball enters the field and the end 15 seconds later. The ball trajectory during these episodes was visualized. Examples can be found in Figure 5.1. From the figure can be derived that the Away team alternates short passing with a long ball to both sides of the team. On the contrary, the Home has a strong preference for a long ball to the negative Y-axis of the field.

5.2.2 Experiment 2: Switching play

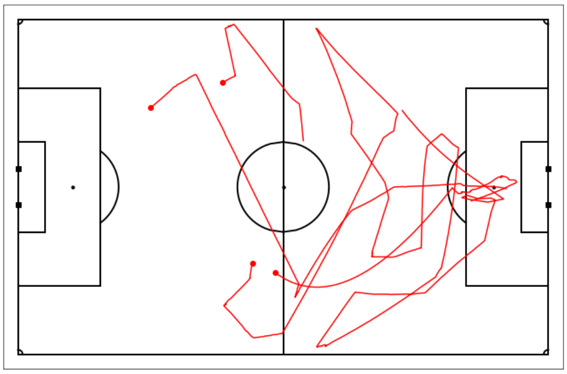
After the build-up, a team wants to penetrate zone 14 or penalty box to create a scoring chance. Every team has a different playing style and tactics to achieve this goal. One of these is 'Switching play'. During switching play, a team moves the ball from one side to the other side with high speed. This can be done either with short passing or a cross. Switching play can be applied for various reasons, like to play out of a high- pressure zone and exploit space on the opposite side of the pitch or to keep the opposition moving to create space behind the opposition. For a coach it can be interesting to know how long it takes before a team switches play, because he can prepare his players for the event. Hence, the question:

How long does it take before a team switches play?

The Temporal Query consists of four interval expressions and can be explained as follows. The first interval expression represents switching play from the negative y-axis to the positive y-axis for the Away team. The first condition is that the Away team is in possession in zone 9, 12 or 15. The second condition is that the Away team is still in possession, but in zone 7, 10, 13, or 16. The time between the end of the first condition and the start of the second condition can be a maximum of 2.04s. A formal description can be found in Figure E.2. This interval expression is mirrored over the x-axis to get switching play from the positive to the negative x-axis. Both interval expressions were mirrored over the y-axis to get switching play for the Home team.



(a) Away



(b) Home

Figure 5.2: Trajectories of the ball for episodes for the period between ball possession gain and Switching play

The qualitative evaluation is done by checking if we can answer the experiment question. Unfortunately, the number of detected episodes with switching play is low. Hence, calculating the average duration of the period between gaining ball possession and switching play does not provide a useful answer, because of the big deviation between the episodes. Instead, the trajectories of the ball were visualization for that same period. Two examples are displayed in Figure 5.2 and two different styles with respect to switching play can be derived. The Away team had a short passing game on the own half and after a while switched to the other side to continue build-up. On the contrary the Home team aims to get the ball around the middle line with a long ball and proceeds immediately with a switch to the other side.

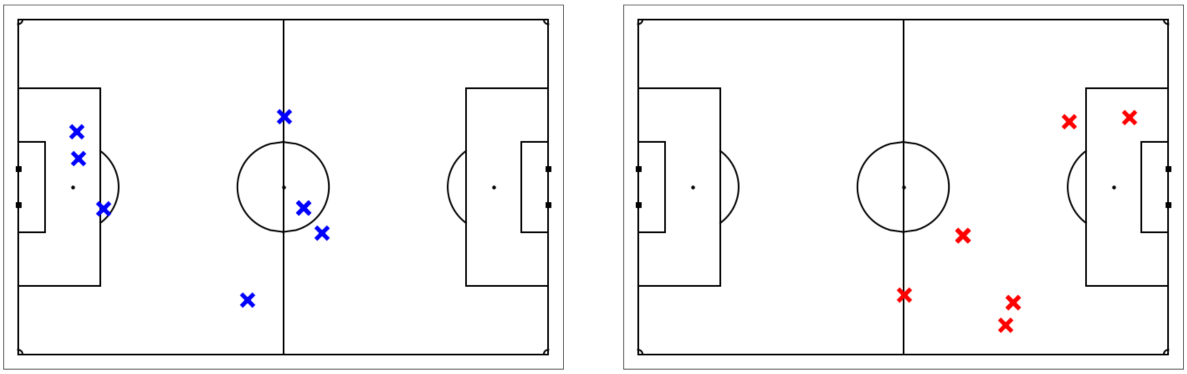
5.2.3 Experiment 3: Counterattack

A counterattack is characterized by three aspects: transition, a pass or dribble that cuts out a number of defenders, and high speed. The first aspect, transition, is the moment that a team loses ball possession. The transition before a counterattack is often unexpected and takes place at a critical position at the wrong moment. In most cases, a critical position is on a team his own half and a wrong moment is often the case when rest defence is low at the moment of transition. After transition, the team in possession gives a pass or makes a dribble that reduces the rest defence of the opposition while reducing the distance to the opposite goal. This has to be done with a high speed to be effective, otherwise the opposition can bring their defence back in position. A counterattack leads often to a scoring opportunity with a high chance of scoring. For that reason, a coach wants to intercept the ball as soon as possible or prevent the counterattack. Hence, the coach is interested in where ball possession loss leads to a counterattack to see if there are systematic problems with the positioning of his players. This is covered in the following question:

Which locations of ball possession loss leads to a counterattack?.

The Temporal query consists of two intervals expressions for a counterattack, one for the Home team and one for the Away team. A counterattack of the Away team has the following conditions. Team Home is in ball possession. This event is directly followed by ball possession of the Away team. During ball possession of the Away team, the ball is in zone 7,8,9,10,11 or 12 while the Home team has four or less players behind the ball. This event is eventually followed within 6 seconds of the ball reaching zone 14 or 17 while maintaining ball possession of the Away team. A formal description of the interval expression can be found in Figure E.3. For a counterattack of the Home team, the conditions of the interval expression were mirrored over the y-axis.

To answer the relevant question, the locations of ball possession loss were visualized. Examples can be found in Figure 5.3. A counterattack for both teams can start from their own penalty box. A possible explanation can be the goalkeeper that throws the ball far away or kicks the ball to the attackers to perform a counterattack. Another remarkable observation is that the Home team only looses ball possession at the side of the negative x-axis. This can be the consequence of a full back positioning too 'high' on the field.



(a) Away

(b) Home

Figure 5.3: The location of Ball possession loss that leads to a counterattack are marked with a X

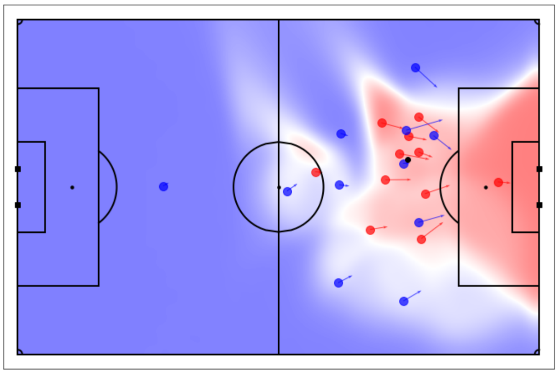
5.2.4 Experiment 4: Gegenpressing

A method to prevent a counterattack is gegenpressing. Coaches have different goals with gegenpressing. Some coaches use it as a defensive mechanism to prevent the occurrence of counterattacks. On the other hand, some coaches use gegenpressing as an attacking style, because they say that succesful gegenpressing leads to the best scoring opportunities. A crucial aspect for a coach is the execution of Gegenpressing. A failed attempt of gegenpressing can lead to a counterattack, because there is a lot of space behind the defenders. A coach can measure the effectivity by checking how often the ball is reconquered as a consequence of applying Gegenpressing:

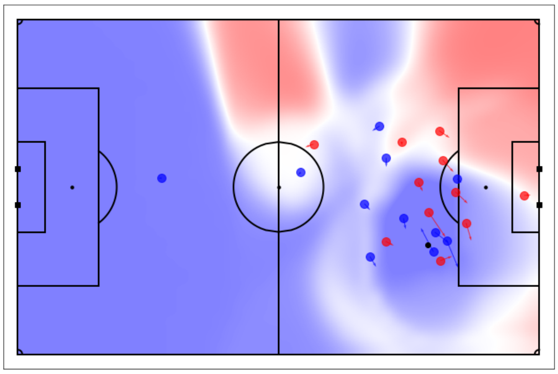
How often does Gegenpressing leads to reconquering of the ball?.

To answer this question, it is necessary to retrieve episodes where gegenpressing occurs. The followings conditions were used to check for Gegepressing applied by the Away team using a Temporal Query. First the Away team has to be in ball possession on their own half and then they loose the ball. The 5 seconds after ball possession loss, the X centroid of the Away team has to go towards the opposite goal with an average speed of more than 0.75m/s (0.03m per frame). For gegenpressing of the Home team, the query conditions were mirrored over the y-axis. A formal description of the interval expression can be found in Figure E.4.

It was challenging to answer the relevant question as there were not so many relevant cases of episodes with gegenpressing identified. Hence, the evaluation was done on individual cases to show the value of the results. Figure 5.4 and 5.5 show pitch control maps. Pitch control maps are introduced by Spearman and he defines pitch control as follows: *The probability that a player can control the ball, assuming it was at this location"*[51]. The pitch control can be used to see whether gegenpressing was executed successfully. In Figure 5.4, gegenpressing is succesfully applied by the Away team represented with the color blue. The left figure shows the situation at the moment of ball possession lost. At that moment the area around the ball is red, but after 5 seconds the area around the ball is covered with blue and the ball is reconquered. This is the consequence of the players pressuring the ball. In the other Figure 5.5, the attempt was failed, because after 5 seconds the area around the ball is not covered by the Away team in blue. As a consequence, the Home team can pass the ball to a fellow player with a lot of free space.

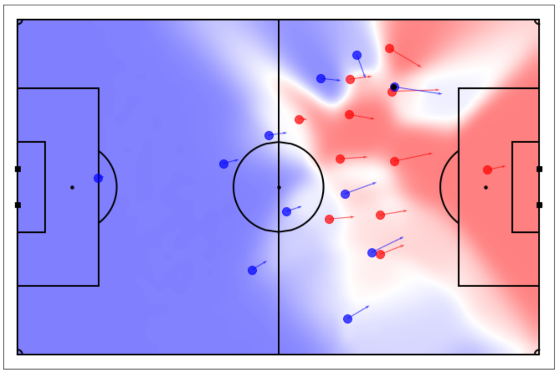


(a) Team Away (blue) loses the ball

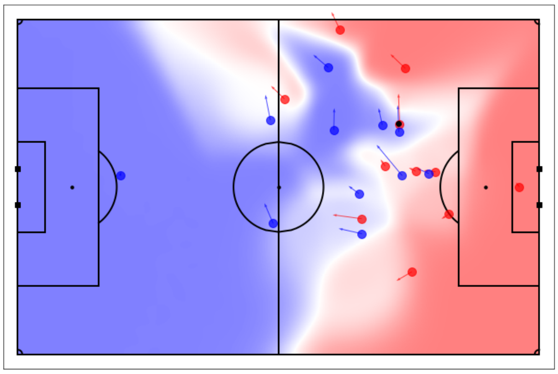


(b) 5 seconds after ball possession loss

Figure 5.4: A successful execution of gegenpressing



(a) Team Away (blue) loses the ball



(b) 5 seconds after ball possession loss

Figure 5.5: An unsuccessful execution of gegenpressing

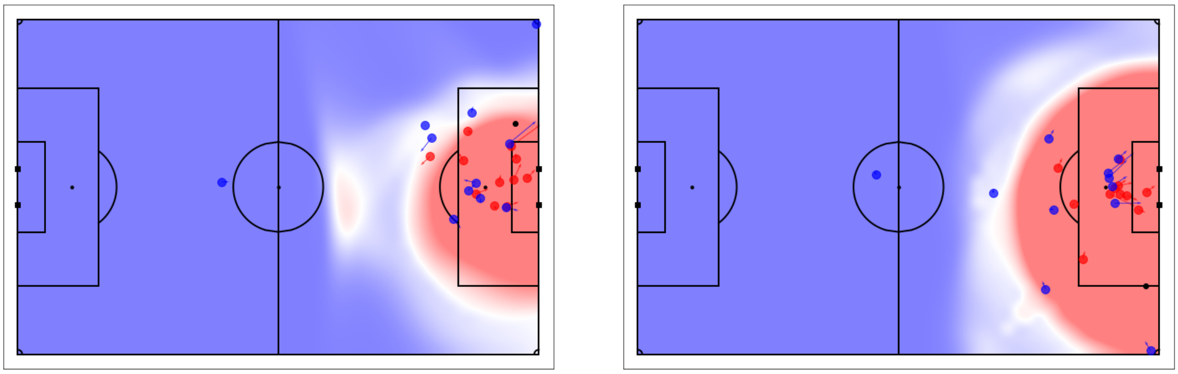
5.2.5 Experiment 5: Corner kick

During the interview, the head coach explained the importance of set pieces. A set piece is an event where the ball is returned to open play, such as a free kick and corner kick. He indicated that he always looked at the set pieces of the last five matches of the coming opponent to see how the opponent approach these set pieces. He used the information to prepare his team and minimize the scoring opportunity.
On the other hand, a set piece can be an opportunity for the defending team for a counter attack. The performance analyst of Ajax, Vosse de Boode, indicated that this was a huge threat for Ajax, because the opponents created many high chance scoring opportunities after Ajax performed a

corner kick. So, for a coach it is interesting to see if the players are well organized during a corner kick:

How is my defence organized during a corner kick?.

The first step is to detect episodes with corner kicks. The Temporal Query consists of four interval expressions, one for every corner of the pitch. To detect a corner kick at the lower-left corner, it was first checked if the ball did not move for 0.8 seconds. Afterwards, the ball has to move in zone 3 with the rest defence of the Home team greater or equal than 8 and the Away team less or equal than 2. During that moment the Team Centroid of the Home and Away team has to be in zone 2 or 5. This is formalized with the interval expression displayed in Figure E.5. An interval expression was mirrored over the x-axis to cover the other corner of the Home team. Both these interval expressions were mirrored over the y-axis to get corner kicks of the Away team.



(a) Corner A of the Away team

(b) Corner B of the Away team

Figure 5.6: Pitch control map just after the corner kick

To answer the relevant question, we used the earlier introduced pitch control maps to study the corner kicks. Two examples can be found in Figure 5.6. The first figure shows a potentially dangerous situation where the Away team is poorly positioned. As can be seen, there is a red surface close to the middle circle. If the Home team is able to gain ball possession, then it can move the ball to that area to exploit the possibility of a counterattack. The second example shows an example where the Away team is well positioned, because there is no space uncontrolled that could lead to a counterattack.

5.2.6 Experiment 6: Low block

During a game, it often happens that one team has possession for a longer amount of time. In most cases, the other team falls back to decrease the size of the 'field' (Convex hull of all players). Currently, there is a trend to use a low block. When a team forms a low block, it is very difficult to score. A coach is interested in finding out how he could break down a low block to create a high scoring chance. Therefore, he has to identify weaknesses when the opponent forms a low block:

What are the weaknesses of a team to break down a low block?

A low block was detected using the following Temporal Query. In case of a low block formed by the Away team, the Home team has to be in possession. In addition, the Team Length of the Away team has to be smaller or equal than 25 meters and the Team X Centroid of the Away team has to be positioned 7 meters or more away from the middle line at the side of the negative x-axis. It was decided that only episodes where these conditions are met for at least 10 seconds are returned to limit the number of results. for a low block formed by the Home team, the conditions were mirrored over the y-axis. A formal description of the interval expression can be found in Figure E.6.

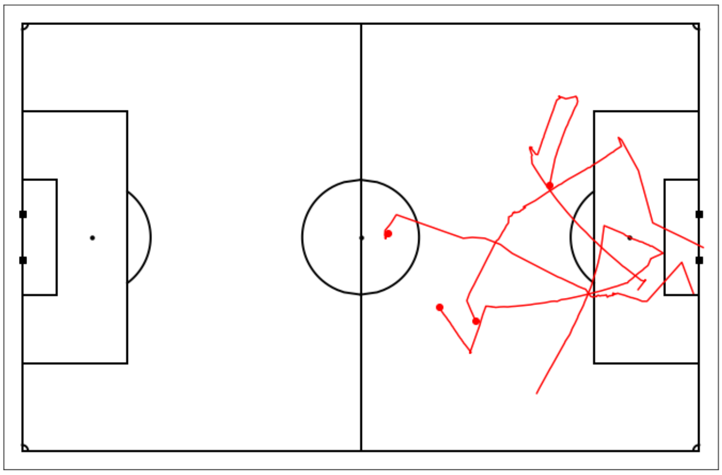
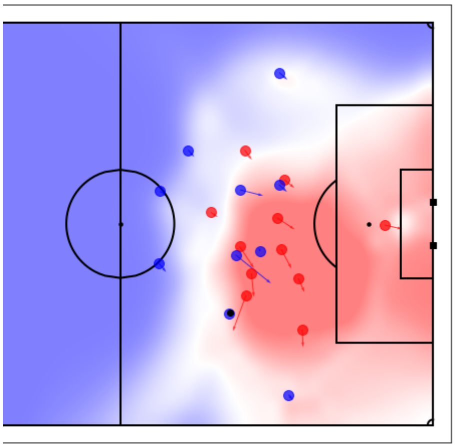
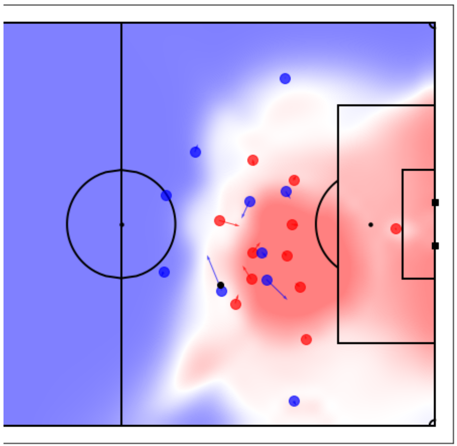


Figure 5.7: Ball trajectories where the box was penetrated after the occurrence of a low block

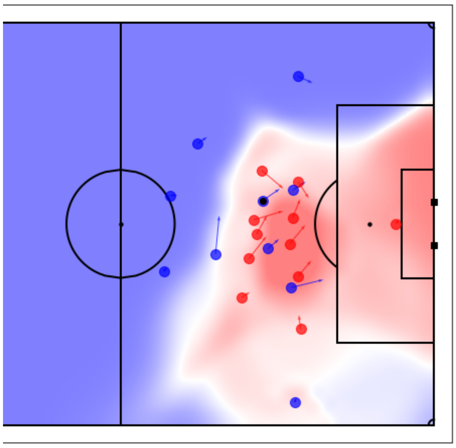
To answer the relevant question, the ball possession phases where a low block occurred were analyzed. An extra condition was added that checks whether the ball reaches the zone of the penalty box (Zone 2 or 17) and the attacking team could keep ball possession. The results of that query for one match is displayed in Figure 5.7. There were not enough cases to find a systematic way to break down the low block. To show the potential to analyze how a low block can be broken down, one case was highlighted. In Figure 5.8, the case is visualized. The figure consists of four snapshots of a ball possession phase where a low block is circumvented and the box is penetrated. The snapshots are strategically chosen to see what events were important.



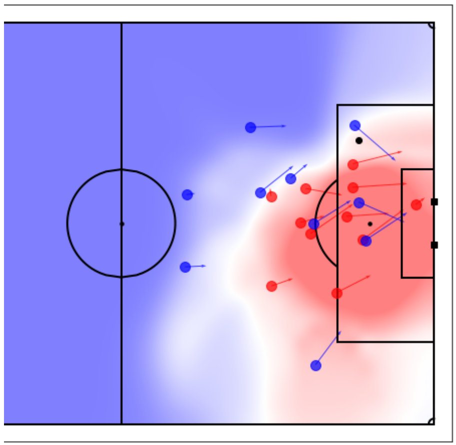
(a) A player moves away from the center to make space for the player in possession



(b) The player in possession uses the available space to pass the ball to a fellow player in zone 14.



(c) The home team pressures the player in possession. As a result, space becomes available at the side of the positive y-axis



(d) The player at the side receives the ball and can dribble into the box.

Figure 5.8: An attack where the attacking team penetrates the box during a low block. The evolution of the attack is displayed with four snapshots of relevant moments with order a-b-c-d.

5.3 Results & Discussion

This section covers the main results of the six experiments. The first category of discussion is the quantitative results. The quantitative results are divided into episodes with first-order events and second-order events. Afterwards, the qualitative results are discussed.

5.3.1 Results

The results of the experiments related to the retrieval of first-order events are displayed in Table 5.2. What stands out of the table is the F-score of 86.3% for the detection of goal kicks and 85.7% for the detection of corners kicks. It indicates that the majority of events are correctly retrieved. This is remarkable, as the Information Retrieval (IR) model was not optimized for the retrieval of episodes with first-order events. A comparison with earlier work reveals that our results are similar to the work of Stein et al. [57] with an F-score varying from 62% to 94% on the detection of first-order events. The Temporal Queries were optimized using the scores for precision and recall on a single match. After testing with different test queries, the final query was chosen to have a balance between precision and recall. However, the results can be changed. If you add more or stronger conditions to the interval expression, the precision will increase. If you do the opposite, then the recall will improve.
The results of both implementations of QBSE were less promising. An R-precision of 51.2 % was achieved for corner kicks and 34.8% for Goal kicks using the TA implementation. The other implementation with DTW was worse than TA with an R-precision of 13.6% for goal kicks and no result could be retrieved for corner kicks. The main problem with DTW is that it could not handle examples were the position of the ball was missing. For both versions of QBSE it should be noted that R-precision cannot be directly compared to regular precision. Another note is that the results might have been better if another example was chosen.

The second part of the quantitative results is related to the experiments with second-order events. A complete overview can be found in Table 5.3. What is interesting about the data in this table is that the results differ greatly per query method and experiment. Interestingly, the Temporal Query method has the best scores on all experiments. The precision ranges from 51.3% for retrieval of counterattacks to 97.1% for retrieval of low blocks. These results indicate the potential of Temporal Query to capture second-order events, because the scores do not deviate much from first-order events. It also shows that some events are easier to retrieve than others. A note with respect to the results is that the precision depends on the Temporal Query. This is similar as explained with the first-order events, the interval expressions can be strengthened or weakened to increase or decrease precision. this can be illustrated with an example: In the case of Switching play, the time necessary to move the ball from one side to the other side is a parameter. If you set a shorter period, then fewer items will be retrieved. As a result the Precision will be higher.

The results for the QBSE method were unexpected, because the performance was not comparable to the results of Temporal Query. It might be that QBSE was deployed for the wrong task. Most second-order events are defined by the occurrence of certain events. These events could not be incorporated in the query, while the query was evaluated on the ability to detect these event. For example, prior to a counterattack, the event 'transition' takes place.
To get a better idea of the value of the new method TA, the scores are compared to implementation with DTW. It reveals that TA scores better on the retrieval of counterattacks and Switching play, while DTW scores better on retrieval of occurrences with gegenpressing and low blocks. It is difficult to drawn conclusions from this comparison.
A more remarkable observation is the performance difference of a method for the experiments. DTW ranges from 40.0% to 90.0% and TA ranges from 30.0% to 77.5%. It again shows that some events are easier to retrieve than others. It is important to notice that all these results are affected by the number of retrieved results and by the provides examples.

The last part, the qualitative evaluation, is the most important part because it gives an indication for the utility of the IR model in practice. The qualitative evaluation showed that all questions were answered using the query methods. This is a rather interesting outcome and indicates that potential value for experts. However, a remark has to be made with respect to this observation. The number of retrieved episodes was insufficient for the experiments Switching play, Gegenpressing and low block to give a complete answer. Instead, an answer was given for an individual episode to show the potential to give a complete answer when more episodes would be available.

5.3.2 Discussion

The results of the experiments have potential limitations. The first limitation is related to the data. The quality of both data sources is unknown and the ball position of the Metric data seems manually annotated. The same holds for the provided event-data. In addition, a data reduction technique is applied. As a consequence, it is uncertain how reproducible the results are with other data sources of spatial temporal data. Another reason for the uncertainty is the amount of data. The experiments were run on four different matches of presumably eight different teams. In comparable studies, they used match data of a complete season for one or multiple teams.

Another limitation is related to the evaluation. First of all, the data was used for both the development of the IR model and the evaluation of the experiments.
Secondly, there was no ground-truth available for the experiments with second-order events. As a consequence, the recall is unknown and the results must be interpreted with caution. It is likely that the number of events differs per event type, match, and team, while the number of items retrieved was constant for these event type and team. It means that the precision at k would be higher if this could be better balanced. Another source of caution is the possible bias that can occur with manual evaluation. Specifically because the manual evaluation was done without the involvement of an expert.

These limitations could be overcome with the availability of more data and elaborate evaluations. It gives also the opportunity to improve the results because the features, query methods, and queries can be optimized.
Firstly, the results of the experiments could be improved by changing and expanding the features. The current features could be standardized with respect to pitch size and playing direction. Moreover, the number of features could be expanded as the current features are limited to several criteria. During the development it became clear that the patterns of some second-order events can be explained to a greater extent with a specific feature. For instance, the Team Length was very suitable to detect a low block. Consequently, future studies could include the use of other features, such as features that are related to Pressure and Formation. Future studies on the QBSE method can investigate automatic feature extraction methods, such as an autoencoder.

Secondly, there is room for improvement concerning the query methods. The use of more domain-knowledge could improve the query design, as a better understanding of the characteristics of the events helps to approximate queries with the events. For example, an ontology could be created for a second-order event that shows how a second-order event is a combination of second-order events.
The QBSE method was tested with different feature weights to retrieve a specific event. During the tests it became clear that it was difficult to estimate the effect of the weights. One explanation was that it is labour-intensive to manually check episodes. The other explanation was that the importance of a feature could not be directly derived. This can be illustrated with an example: There are two episodes with identical values for Team Length, but different values for Team Width. The similarity is calculated for an example episode. If we change the feature weights for both Team Length and Team Width, then the similarity scores for both episodes will change. The problem is that it cannot be derived which change of feature weights is responsible for the differences in similarity score. A solution could be to indicate the feature importance.

The results of the QBSE method could be improved with another ranking method. DTW was not tested to full potential. Also, the use of different similarity scores has to be investigated, for instance, the use of psychometric testing [63]. Another improvement could be the use of a different IR, such as the use of language models.

Chapter 6

Conclusions

In this study is investigated how interesting episodes can be retrieved from spatial temporal data. An IR IR model is presented that is able to retrieve episodes that satisfy the information needs of experts. The model is implemented and evaluated using a series of experiments.

6.1 Contribution

The results of the interviews with experts and literature review indicated the need for an automated episode retrieval method in order to reduce the time spend on match analysis. Previous work did not address the systematical retrieval of episodes with second-order events from spatial temporal data. Furthermore, it was found that a requirement for the solution is that experts can interact with the results to deal with the different workflows and football philosophies. The study demonstrates how a classical IR model can be applied to overcome these issues. In addition, it shows how the information retrieval methods Temporal Query and QBSE can be applied to spatial temporal data.

The model for episode retrieval has a high-level design to be compatible with different work environments of the analysts. It consists of a document representation, query representation, framework, and ranking function. The query methods of the model are selected to satisfy the information needs of known-item search and existence search. The model can be categorized as both a Boolean Retrieval Model and Vector Space Model.

An important requirement for the model was interactivity. To be interactive, the model has to be understandable, expressive, customizable, and the results have to be retrieved in a reasonable amount of time. As a consequence, the following design choices were made. All components of the model are selected to be comprehensible for an expert without computer science knowledge. The features are semantic and capture the complex patterns of second-order events. The query methods can be customized by an expert and run in polynomial time.

An experimental study was conducted to test the quality and utility of the model. The quantitative results show the ability to replace the manual annotation of both first- and second-order events when perfect recall is not essential. An analyst can save a lot of time using an implementation of this model. In particular for the retrieval of second-order events, it was found that an analyst can save hours of work. Even if we take into account the manual removal of false positives, a lot of time can be saved. Moreover, the quality of analysis increases because an analyst can study a lot of matches simultaneously.

The results of the qualitative evaluation of the experiments support the stated practical utility of the model. The Evaluation of End Products method was used to show how the model can be used to solve realistic questions of experts.

6.2 Limitations

The results of this study have to be viewed in light of some limitations. First of all, the model is designed to satisfy a set of requirements, but not all requirements could be tested. These are the ability to implement the model in a variety of software applications, the understandability of the model, and the interactivity.

These requirements are a limitation of the study. A model designed without these requirements could perform better in terms of relevance, as different type of data sources and more advanced models could be applied.

Another limitation is related to practical utility. The experiments were designed to show the practical utility, but they were not tested in real-life situations. Moreover, it would be interesting to see how to model would perform if has to answer questions related to a specific football philosophy.

During the experiments, the novel 'Relative Occupancy Map' feature is introduced. It can be interesting for practitioners and other researchers to know the added value of this and the other features. Although, during the experiments, a selection of features are used for Temporal Query and the weights are set for QBSE, it is difficult to draw conclusions about the contribution of the individual features. The same holds for the contribution of the event index function.

The last limitation is related to the data. There is some concern with the general reproducibility, as the number of matches used for analysis is relatively low. In addition, the data quality of other spatial temporal data sources could deviate to an extent that it affects the general reproducibility.

6.3 Future work

To develop a full picture of the quality of the model, we recommend to test all requirements extensively in a real-life situation with experts. The development of a software application can be a great tool for this purpose. It is advisable to involve an analyst during the development to optimize usability and to ensure that the application suits the current workflow of the analyst. The applications can be used to derive insights on the understandability, interactivity, and practical utility. In particular, it would be interesting to measure the time saved compared to the manual method. Moreover, the effect on the quality of a match analysis (in real-time) is of great interest, because the model has the potential to derive completely new insights on the strategical and tactical aspects of the game.

Furthermore, there is abundant room for further progress in the study of information retrieval with spatial temporal data. The index function could be improved and expanded. Specifically, the incorporation of domain-knowledge could improve performance. Domain-knowledge could also be used to improve the query and ranking methods. Aside from that, future studies could answer if different configurations of those methods or different methods are more suitable for episode retrieval.

Bibliography

[1] Dale vince: Forest green rovers using 'moneyball' model. *bbc.com*, Dec 2015. `https://www.bbc.com/sport/football/34186514`, last accessed on 26/03/2020. 1

[2] Jamie Adams. Tactical terms. `https://jadamsftbl.wordpress.com/tactical-terms/`, last accessed on 21/04/2020. 70

[3] Gennady Andrienko, Natalia Andrienko, Gabriel Anzer, Pascal Bauer, Guido Budziak, Georg Fuchs, Dirk Hecker, Hendrik Weber, and Stefan Wrobel. Constructing Spaces and Times for Tactical Analysis in Football. *IEEE Transactions on Visualization and Computer Graphics*, 14(8):1–1, 2019. 12, 14, 15

[4] Gennady Andrienko, Natalia Andrienko, Guido Budziak, Jason Dykes, Georg Fuchs, Tatiana von Landesberger, and Hendrik Weber. *Visual analysis of pressure in football*, volume 31. Springer US, 2017. ix, 3, 6, 11, 12, 14, 16, 17, 29, 56

[5] Gennady Andrienko, Natalia Andrienko, and Stefan Wrobel. Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations Newsletter*, 9(2):38–46, 2007. ix, 56

[6] Ricardo Baeza-Yates, Berthier Ribeiro-Neto, et al. *Modern information retrieval*, volume 463. ACM press New York, 1999. 20, 23

[7] António Belguinha, Pedro Rodrigues, P. J.S. Cardoso, J. M.F. Rodrigues, and Domingos Paciência. A visual programming language for soccer. *ICSOFT-PT 2014 - Proceedings of the 9th International Conference on Software Paradigm Trends*, pages 121–127, 2014. 17

[8] António Belguinha, Pedro Rodrigues, Pedro JS Cardoso, João MF Rodrigues, and Domingos Paciência. A visual programming language for soccer. In *2014 9th International Conference on Software Paradigm Trends (ICSOFT-PT)*, pages 121–127. IEEE, 2014. 13, 14

[9] David T Bourgeois. Information systems for business and beyond. 2018. 6

[10] Lotte Bransen and Jan Van Haaren. Player chemistry: Striving for a perfectly balanced soccer team. *arXiv preprint arXiv:2003.01712*, 2020. 2

[11] Michel S Brink, Jurian P Kuyvenhoven, Tynke Toering, Geir Jordet, and Wouter GP Frencken. What do football coaches want from sport science? *Kinesiology*, 50(1):150–154, 2018. 6

[12] Francisco Siles Canales. Automated Semantic Annotation of Football Games from TV Broadcast. 2014. ix, 4, 5, 13

[13] Christopher Carling, Tom Reilly, and A Mark Williams. *Performance assessment for field sports*. Routledge, 2008. 1

[14] Christopher Carling, Craig Wright, Lee John Nelson, and Paul S. Bradley. Comment on 'Performance analysis in football: A critical review and implications for future research'. *Journal of Sports Sciences*, 32(1):2–7, 2014. 4

[15] Michiel de Hoog. How data, not people, call the shots in denmark. `https://thecorrespondent.com/2607/how-data-not-people-call-the-shots-in-denmark/230219386155-d2948861`, last accessed on 10/04/2020. 2

[16] Ricardo Duarte, Duarte Araújo, Hugo Folgado, Pedro Esteves, Pedro Marques, and Keith Davids. Capturing complex, non-linear team behaviours during competitive football performance. *Journal of Systems Science and Complexity*, 26(1):62–72, 2013. 15, 16

[17] Karën Fort and Vincent Claveau. Annotating football matches: Influence of the source medium on manual annotation. *Proceedings of the 8th International Conference on Language Resources and Evaluation, LREC 2012*, pages 2567–2572, 2012. 70

[18] Wouter Frencken, Koen Lemmink, Nico Delleman, and Chris Visscher. Oscillations of centroid position and surface area of soccer teams in small-sided games. *European Journal of Sport Science*, 11(4):215–223, 2011. 15, 16

[19] FR Goes, LA Meerhoff, MJO Bueno, DM Rodrigues, FA Moura, MS Brink, MT Elferink-Gemser, AJ Knobbe, SA Cunha, RS Torres, et al. Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review. *European Journal of Sport Science*, pages 1–16, 2020. 5, 15

[20] Jiawei Han, Micheline Kamber, and Jian Pei. Data mining concepts and techniques third edition. *Morgan Kaufmann*, 2011. 5, 20

[21] Albert H Hastorf and Hadley Cantril. They saw a game; a case study. *The Journal of Abnormal and Social Psychology*, 49(1):129, 1954. 4

[22] Djoerd Hiemstra. Information retrieval models., 2009. 6

[23] Halld'Or Janetzko, Dominik Sacha, Manuel Stein, Tobias Schreck, Daniel A. Keim, and Oliver Deussen. Feature-driven visual analytics of soccer data. *2014 IEEE Conference on Visual Analytics Science and Technology, VAST 2014 - Proceedings*, pages 13–22, 2015. ix, x, 5, 17, 57, 59

[24] Jules Boyle. Football hipster terms explained : what is 'the low block'? *FootballScotland*, juli 2019. `https://www.footballscotland.co.uk/features/long-reads/football-hipster-terms-explained--16523865`, last accessed on 09/04/2020. 70

[25] Diane Kelly. Methods for evaluating interactive information retrieval systems with users. *Foundations and trends in Information Retrieval*, 3(1—2):1–224, 2009. 32

[26] Ho Chul Kim, Oje Kwon, and Ki Joune Li. Spatial and spatiotemporal analysis of soccer. *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*, pages 385–388, 2011. 15

[27] Jin Ha Lee, Allen Renear, and Linda C Smith. Known-item search: Variations on a concept. *Proceedings of the american society for information science and technology*, 43(1):1–17, 2006. 21

[28] Lipyeow Lim, Haixun Wang, and Min Wang. Semantic queries by example. *ACM International Conference Proceeding Series*, pages 347–358, 2013. 13

[29] Hongyou Liu, Miguel-Angel Gómez, Bruno Gonçalves, and Jaime Sampaio. Technical performance and match-to-match variation in elite football teams. *Journal of sports sciences*, 34(6):509–518, 2016. 1, 16

[30] Vinicius Machado, Roger Leite, Felipe Moura, Sergio Cunha, Filip Sadlo, and João L.D. Comba. Visual soccer match analysis using spatiotemporal positions of players. *Computers and Graphics (Pergamon)*, 68:84–95, 2017. x, 16, 17, 58

[31] Rob Mackenzie and Chris Cushion. Performance analysis in football: A critical review and implications for future research. *Journal of sports sciences*, 31(6):639–676, 2013. 1, 6

[32] Christopher Manning, Prabhakar Raghavan, and Hinrich Schütze. Introduction to information retrieval. *Natural Language Engineering*, 16(1):100–103, 2010. 30, 31

[33] Scott McLean, Paul M. Salmon, Adam D. Gorman, Gemma J.M. Read, and Colin Solomon. What's in a game? A systems approach to enhancing performance analysis in football. *PLoS ONE*, 12(2):1–15, 2017. 1

[34] Daniel Memmert, Koen A.P.M. Lemmink, and Jaime Sampaio. Current Approaches to Tactical Performance Analyses in Soccer Using Position Data. *Sports Medicine*, 47(1):1–10, 2017. 15

[35] Metrica. Metrica sports sample data, 2020. data retrieved from Github, `https://github.com/metrica-sports/sample-data`,. 33

[36] Asgeir Mortensen, Vamsidhar Reddy Gaddam, Håkon Kvale Stensland, Carsten Griwodz, Dag Johansen, and Pål Halvorsen. Automatic event extraction and video summaries from soccer games. In *Proceedings of the 5th ACM Multimedia Systems Conference*, pages 176–179, 2014. 12, 14, 16

[37] Felipe Arruda Moura, Luiz Eduardo Barreto Martins, Ricardo de Oliveira Anido, Ricardo Machado Leite de Barrosr, and Sergio Augusto Cunha. Quantitative analysis of Brazilian football players' organisation on the pitch. *Sports Biomechanics*, 11(1):85–96, 2012. 15, 16

[38] Ivan Mushketyk. Stream processing with apache flink. *Dzone*, october 2017. `https://dzone.com/articles/stream-processing-with-apache-flink`, last accessed on 24/04/2020. ix, 25

[39] Sigrid B.H. Olthof, Wouter G.P. Frencken, and Koen A.P.M. Lemmink. The older, the wider: On-field tactical behavior of elite-standard youth soccer players in small-sided games. *Human Movement Science*, 41:92–102, 2015. 15, 16

[40] Charles Perin, Romain Vuillemot, and Jean Daniel Fekete. SoccerStories: A kick-off for visual soccer analysis. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2506–2515, 2013. x, 17, 58

[41] Nikhil Rasiwasia, Pedro J Moreno, and Nuno Vasconcelos. Bridging the gap: Query by semantic example. *IEEE Transactions on Multimedia*, 9(5):923–938, 2007. 6, 13

[42] Charles Reep and Bernard Benjamin. Skill and chance in association football. *Journal of the Royal Statistical Society. Series A (General)*, 131(4):581–585, 1968. 1

[43] Keven Richly. Leveraging Spatio-Temporal Soccer Data to Define a Graphical Query Language for Game Recordings. *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, pages 3456–3463, 2019. 13, 14, 15, 55

[44] Keven Richly, Max Bothe, Tobias Rohloff, and Christian Schwarz. Recognizing compound events in spatio-temporal football data. *IoTBD 2016 - Proceedings of the International Conference on Internet of Things and Big Data*, (IoTBD):27–35, 2016. 14

[45] Dan Robinson. Fc midtjylland chairman rasmus ankersen on the importance of football analytics – and what ceos can learn. *NS Business*, Jul 2019. `https://www.ns-businesshub.com/technology/fc-midtjylland-rasmus-ankersen-football-analytics/`, last accessed on 26/03/2020. 1

[46] Louis Rosenfeld and Peter Morville. *Information architecture for the world wide web.* " O'Reilly Media, Inc.", 2002. 21

[47] Emma Sanders. Summer transfer window: Who carried out the best business? *bbc.com*, Aug 2019. `https://www.bbc.com/sport/football/49319480`, last accessed on 26/03/2020. 1

[48] Hugo Sarmento, Rui Marcelino, Jorge Campanico, Nuno Matos, and José Leitão. Match analysis in football: a systematic review. *Journal of sports sciences*, 32:1831–1843, 12 2014. 6

[49] Bruce Schoenfeld. How data (and some breathtaking soccer) brought liverpool to the cusp of glory. `https://www.nytimes.com/2019/05/22/magazine/soccer-data-liverpool.html`, last accessed on 02/07/2020. 1

[50] Lin Shao, Dominik Sacha, Benjamin Neldner, Manuel Stein, and Tobias Schreck. Visual-interactive search for soccer trajectories to identify interesting game situations. *IS and T International Symposium on Electronic Imaging Science and Technology*, (February), 2016. 12, 14, 16, 17

[51] William Spearman. Quantifying pitch control, 02 2016. 16, 39

[52] Manuel Stein, Thorsten Breitkreutz, Johannes Haussler, Daniel Seebacher, Christoph Niederberger, Tobias Schreck, Michael Grossniklaus, Daniel Keim, and Halldor Janetzko. Revealing the Invisible: Visual Analytics and Explanatory Storytelling for Advanced Team Sport Analysis. *2018 International Symposium on Big Data Visual and Immersive Analytics, BDVA 2018*, 2018. 5

[53] Manuel Stein, Johannes Häußler, Dominik Jäckle, Halldór Janetzko, Tobias Schreck, and Daniel A. Keim. Visual soccer analytics: Understanding the characteristics of collective team movement based on feature-driven analysis and abstraction. *ISPRS International Journal of Geo-Information*, 4(4):2159–2184, 2015. 5, 17

[54] Manuel Stein, Halldor Janetzko, Thorsten Breitkreutz, Daniel Seebacher, Tobias Schreck, Michael Grossniklaus, Iain D. Couzin, and Daniel A. Keim. Director's Cut: Analysis and Annotation of Soccer Matches. *IEEE Computer Graphics and Applications*, 36(5):50–60, 2016. ix, 3, 9, 10, 11

[55] Manuel Stein, Halldór Janetzko, Tobias Schreck, and Daniel A. Keim. Tackling Similarity Search for Soccer Match Analysis: Multimodal Distance Measure and Interactive Query Definition. *IEEE Computer Graphics and Applications*, 39(5):60–71, 2019. 13, 14, 15, 17, 35, 69

[56] Manuel Stein, Halldór Janetzko, Daniel Seebacher, Alexander Jäger, Manuel Nagel, Jürgen Hölsch, Sven Kosub, Tobias Schreck, Daniel A. Keim, and Michael Grossniklaus. How to make sense of team sport data: From acquisition to data modeling and research aspects. *Data*, 2(1), 2017. 2, 11, 69

[57] Manuel Stein, Daniel Seebacher, Tassilo Karge, Tom Polk, Michael Grossniklaus, and Daniel A Keim. From movement to events: Improving soccer match annotations. In *International Conference on Multimedia Modeling*, pages 130–142. Springer, 2019. 2, 3, 11, 12, 14, 16, 21, 43

[58] UEFA. The european club footballing landscape; club licensing benchmark report financial year 2018, 2018. 1

[59] Peter Van Beek and Robin Cohen. *Approximation algorithms for temporal reasoning*. University of Waterloo. Department of Computer Science, 1989. 21

[60] Rüdiger Wirth and Jochen Hipp. Crisp-dm: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, pages 29–39. Springer-Verlag London, UK, 2000. 7

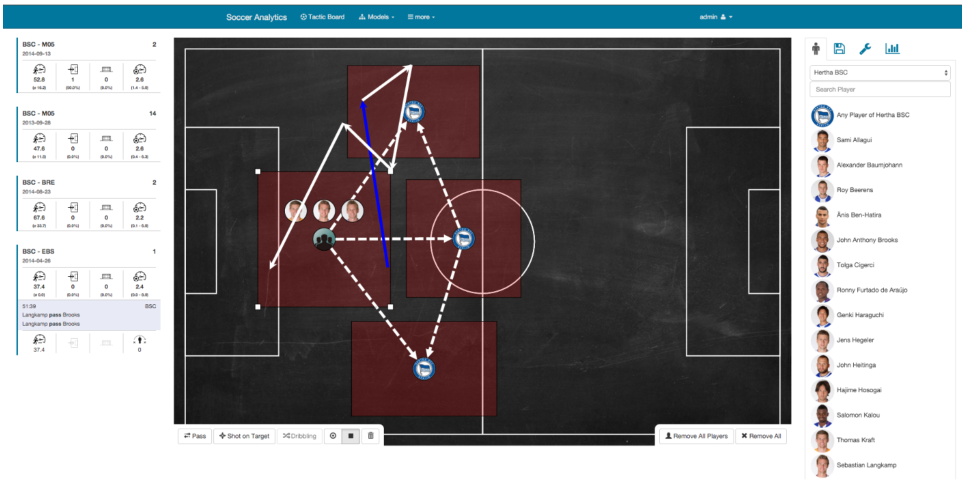
[61] Craig Wright, Steve Atkins, Bryan Jones, and Jason Todd. The role of performance analysts within the coaching process: Performance analysts survey 'the role of performance analysts in elite football club settings.'. *International journal of performance analysis in sport*, 13(1):240–261, 2013. 2

[62] Yingcai Wu, Xiao Xie, Jiachen Wang, Dazhen Deng, Hongye Liang, Hui Zhang, Shoubin Cheng, and Wei Chen. ForVizor: Visualizing Spatio-Temporal Team Formations in Soccer. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):65–75, 2019. x, 17, 57

[63] Lu Yin, Vlado Menkovski, and Mykola Pechenizkiy. Knowledge elicitation using deep metric learning and psychometric testing. *arXiv preprint arXiv:2004.06353*, 2020. 45

Appendix A

Impression of applications



(a) A visual query language inspired by the tactic board developed by Richly [43]

(b) A visual query language combined with the possibility to add event information.

Figure A.1: Two implementations of a visual query language

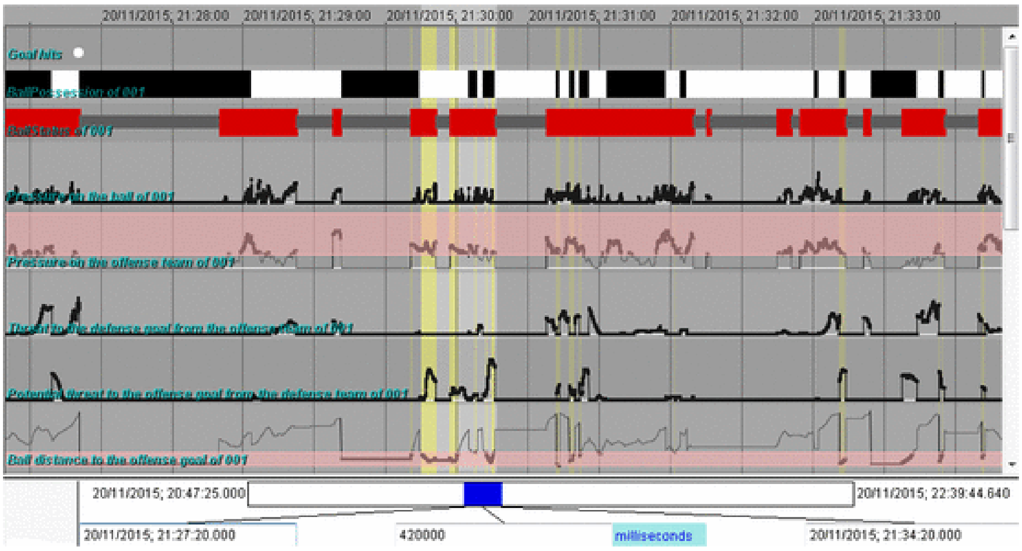


Figure A.2: The screenshot is taken from a V-Analytics software [5]. It displays the values for different features which gives the expert the ability to set query conditions and select interesting game episodes [4].

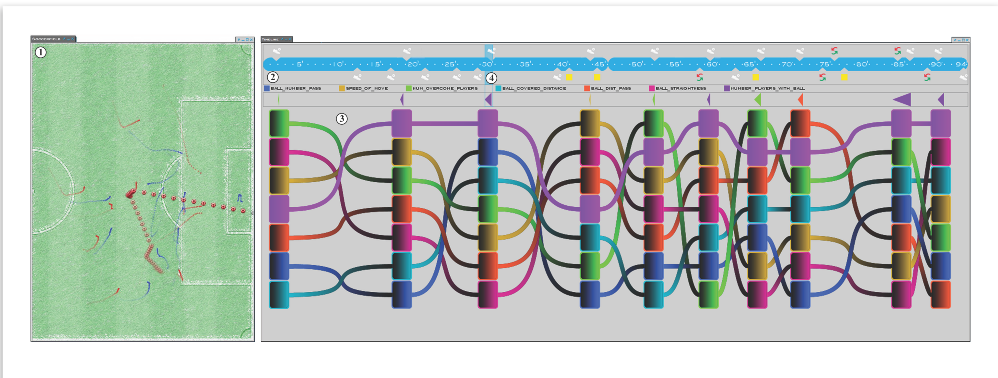


Figure A.3: A combination of three views. (1) A soccer pitch with visualization of player and ball movement. (2) A timeline with icons for quick navigation. (3) A flow-like visualization of semi-automatically-identified interesting moves and their characteristics.

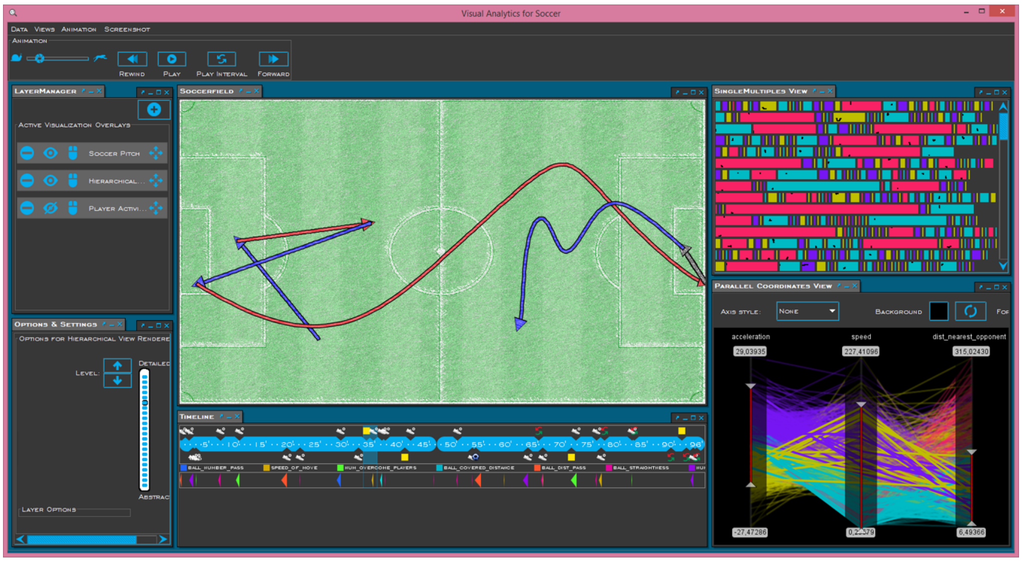


Figure A.4: The tool developed by Janetzko et al. [23] for explorative analyses of football matches.

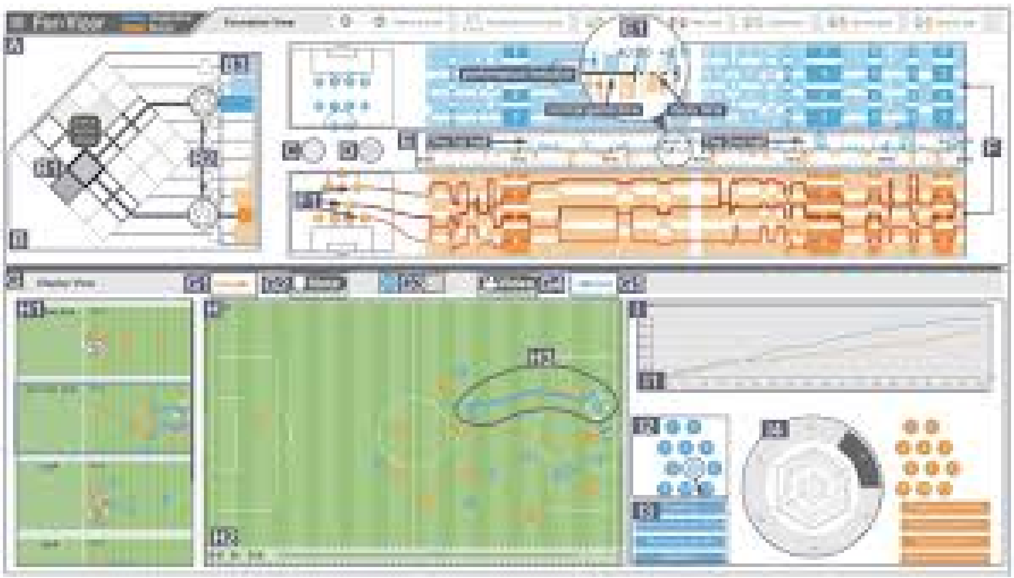


Figure A.5: The tool developed by Wu et al. [62] for analyses of formation. It consists of a a formation view (A) and a display view (G). Formation view contains a confrontation matrix (B), a narrative timeline (E), and two formation flows (F). Display view contains a pitch (H) and a statistical dashboard (I)

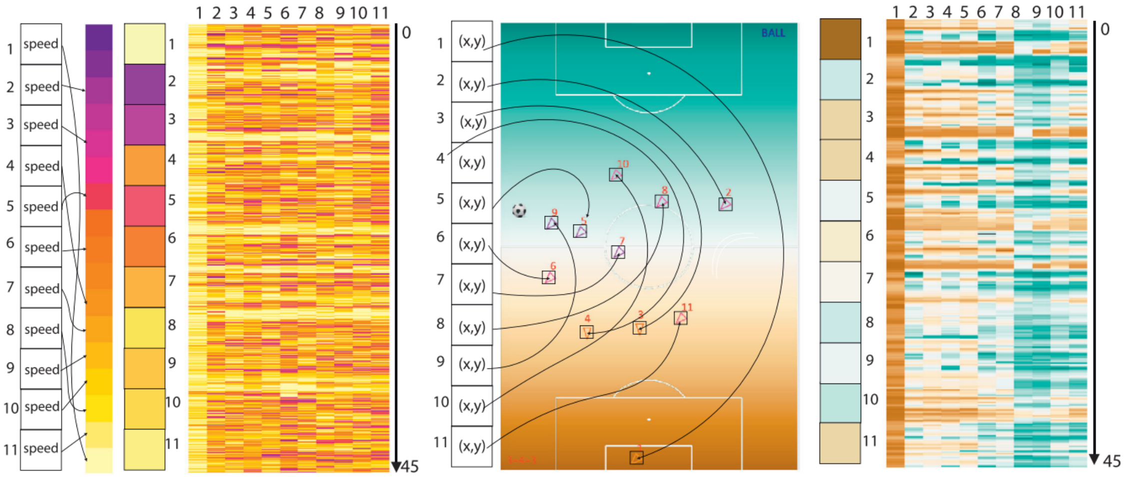


Figure A.6: The left picture shows the player attribute heatmap for speed. The color represents the evaluation of the attribute over time. On the left a 2d mapping is shown to visualize the player positions [30].

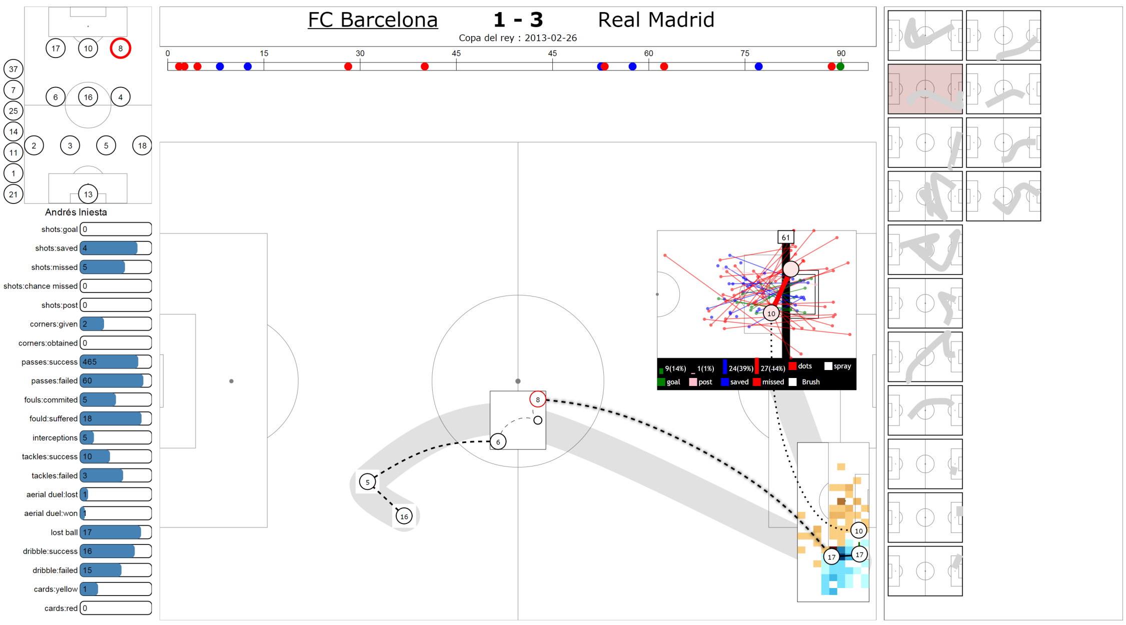


Figure A.7: Interface of SoccerStories with summary statistics on the left hand side, timeline on top, visualization of a phase in the middle, and thumbnails of different phases are displayed on the right hand side [40].

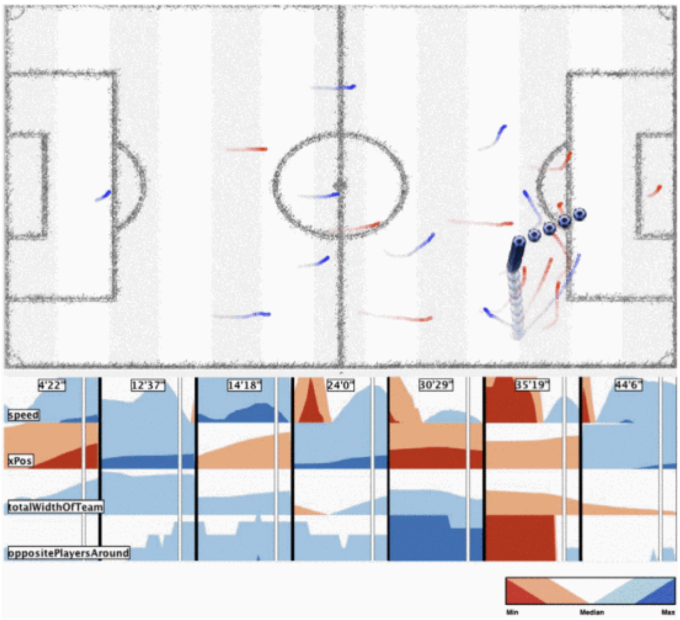


Figure A.8: The horizon graph to identify interesting situations. All situations in the picture show shots on goal [23].

Appendix B

Formal episode representation

coordinates - A tuple (x_t, y_t) where:

- x_t and y_t are real numbers that represent the coordinates with respect to the football field with the center spot of the field (0,0);
- t is a real number that represents the time with the start of the match $t_0 = 0.00s$;

ball - $< ball\_id,\ coordinates,\ t >$ where:

- ball_id is a positive integer as ball identifier;

player - $< player\_id,\ coordinates,\ t >$ where:

- player_id is a positive integer as player identifier;

team - $< team\_id,\ \{player_1, player_2, ..., player_n\},\ t >$ where:

- team_id is a boolean as team identifier with either value ”Home” or ”Away”;
- n is a positive integer that represents the number of players in the team;

frame - $< frame\_id,\ ball,\ team_{home}, team_{away},\ t, >$ where:

- frame_id is a positive integer as frame identifier with $frame_1 < frame_2 < frame_3\ ...\ frame_n$ and n the number of frames;

episode - $< episode\_id, \{frame_i, ..., frame_j\} >$ where:

- episode_id is a positive integer as frame identifier;
- for i, j it holds that $1 \leq i < j \leq n$;
- For all i with $frame_i$ an element of an arbitrary $episode_k$ and k a real number, it holds that $frame_{i+1}$ is an element of $episode_k$;

event - $< event\_id,\ type,\ t_{start},\ t_{end}\ >$ where:

- event_id is a positive integer as event identifier;
- type is a string representing the meaning of the event
- t_{start} is a real number that represents the time when the event started with the start of the match with $t_0 = 0.00s$;
- t_{end} is a real number that represents the time when the event finished with the start of the match with $t_0 = 0.00s$;

semantic episode - $< semantic\_episode\_id,\ episode, \{event_1, event_2, ..., event_n\} >$ where:

- semantic_episode_id is a positive integer as semantic episode identifier;
- the length of $\{event_1, event_2, ..., event_n\} >= 1$;

Appendix C

Temporal Query attributes

Aggregate	Explanation
MIN	Minimum
MAX	Maximum
SUM	Sum of values
AVG	Average of values
COUNT	Number of non-null observations
DIFF	Difference between first and last element of the window

Table C.1: Overview of aggregate functions

Comparison	Explanation
$=$	is equal to
$>$	is greater than
$<$	is less than
$\geq$	is greater or equal than
$\leq$	is less or equal than
$\neq$	not equal to

Table C.2: Value comparison operators

Relation	Explanation
$End(X) < Begin(Y)$	X before Y
$Begin(X) = Begin(Y)$ & $End(X) = End(Y)$	X is equal to Y
$End(X) = Begin(Y)$	X meets Y
$Begin(X) < Begin(Y)$ & $End(X) < End(Y)$ & $End(X) > Begin(Y)$	X overlaps Y
$Begin(Y) \leq Begin(X)$ & $End(Y) \geq End(X)$ & $X \neq Y$	X during Y
$Begin(X) = Begin(Y$ & $End(X) < End(Y)$	X starts Y
$Begin(X) > Begin(Y)$ & $End(X) = End(Y)$	X finishes Y

Table C.3: Allen's interval comparisons

Appendix D

Temporal Query examples

Example 1: Basic query on collective feature

The first example is a basic query. It can be seen as an interval expression with the duration of the window set to the minimum frame size of 1 (0.04s). So, in practise no aggregation takes place. The following interval expression returns all frames where the Team Width of the Home team is greater than 20 meter:

- Match interval: A[0:00 - 90:00]
- Interval: 0:90 minutes
- Window size: 1 frame
- Sliding gap: -
- Condition A: Team Width Home > 20 meter

Example 2: Basic query with events

The next example is very similar to the Example 1. The difference with example 1 is that the condition instead on the events features. In this case, the expert wants to retrieve frames that are labelled with Kick-off during the first half.

- Match interval: B[0:00 - 45:00]
- Window size: 1 frame
- Sliding gap: -
- condition A: Events = 'Kick-off'

Example 3: Combination of Example 1 and 2

Example 3 shows how a interval expression can be executed with two conditions. The interval expression returns the results of the inner join of the condition A in Example 1 and condition A in Example 2.

- Match interval: A[0:00 - 90:00]
- Window size: 1 frame
- Sliding gap: -
- condition A: Team Width Home > 20 meter
- condition B: Events = 'Corner'
- Relation: A overlaps with B

Example 4: Temporal aggregation with fixed window on collective feature

The following example serves the purpose to indicate the use of temporal aggregation. The window size is set to 15 seconds with a sliding gap of 5 seconds. The feature of interest is set to Ball Speed with aggregation type AVG. The interval expression returns all episodes where the average ball speed is greater than 5 meter per seconds.

- Match interval: A[0:00 - 90:00]
- Window size: 15 seconds
- Sliding gap: 5 seconds
- condition A: AVG (players speed) > 5 meter per second

Example 5: Temporal aggregation with fixed window on event

Example 5 is similar to Example 4. The example shows how the expert can look for episodes where one or multiple counterattacks took place. In this example, the window size is set to 10 seconds without sliding. Then, it is counted how many frames are labelled with the event 'Counteratack'. If the count is greater than 125 frames (5 seconds), the episode is returned.

- Match interval: A[0:00 - 90:00]
- Window size: 10 seconds
- Sliding gap: -
- Condition A: COUNT (Events = 'Counterattack') > 125 observations

Example 6: Temporal aggregation with event-based window on event

Instead of a fixed window, an event-based window can be used. Frames are aggregated to episodes based on a condition. In this example the window condition is set to the event 'Counterattack'. In addition, one or multiple conditions can be set, such a the duration of the episode. The interval expression returns all episodes that have a duration over 5 seconds of which all frames are labelled with Counterattack.

- Match interval: A[0:00 - 90:00]
- Window size: event-based
- Window condition: Counterattack
- Sliding gap: -
- Condition A: SUM (Duration) > 5 seconds

Example 7: Temporal aggregation with event-based window on event with multiple conditions

Example 7 is a more advanced version of Example 6. It makes it possible to search for a counter-attack of the Home team, because it aggregates on phases of ball possession of the Home team. These episodes are filtered with the condition A en B using two different aggregation types. The interval expressions returns all episodes where the Home team has ball possession that take more than 5 seconds and 125 observations with the label Counterattack are present.

- Match interval: A[0:00 - 90:00]
- Window size: event-based
- Window condition: Possession Home
- Sliding gap: -
- Condition A: SUM (Duration) > 5 seconds

- Condition B: COUNT (Events = 'Counterattack') $>$ 125 observations
- Relation: A is equal to B

Appendix E

Temporal Query experiments

- Match interval: MET1 [0:00 - end], MET2 [0:00 - end], DFB1 [0:00 - end], DFB2 [0:00 - end]
- Window size: -
- Sliding gap: -
- Condition A: Possession = 'Out'
- Condition B: Ball velocity $\neq 0 \wedge$ Ball zone $= 17 \wedge$ Rest defence home $\leq 1 \wedge$ Rest defence away $= 10 \wedge$ Possession = 'Home'
- Relation: A meets B

Figure E.1: Experiment 1: Build-up

- Match interval: MET1 [0:00 - end], MET2 [0:00 - end], DFB1 [0:00 - end], DFB2 [0:00 - end]
- Window size: -
- Sliding gap: -
- Condition A: Possession = 'Away' $\wedge$ Ball zone $= 9 \vee 12 \vee 15$
- Condition B: Possession = 'Away' $\wedge$ Ball zone $= 7 \vee 10 \vee 13 \vee 16$
- Relation: A before B with a maximum gap of 2.04 seconds in between.

Figure E.2: Experiment 2: Switching play

- Match interval: MET1 [0:00 - end], MET2 [0:00 - end], DFB1 [0:00 - end], DFB2 [0:00 - end]
- Window size: -
- Sliding gap: -
- Condition A: Possession = 'Home'
- Condition B: Possession = 'Away'
- Condition C: Possession = 'Away' $\wedge$ Ball zone = $7 \vee 8 \vee 9 \vee 10 \vee 11 \vee 12$ $\wedge$ Rest defence Home ≤ 4
- Condition D: Possession = 'Away' $\wedge$ Ball zone = $14 \vee 17$
- Relation: A meets B, B overlaps C, C before D with a maximum gap of 6 seconds

Figure E.3: Experiment 3: Counterattack

- Match interval: MET1 [0:00 - end], MET2 [0:00 - end], DFB1 [0:00 - end], DFB2 [0:00 - end]
- Window size: event-based on condition B
- Sliding gap: -
- Condition A: Possession = 'Away' $\wedge$ Ball zone = $1 \vee 2 \vee 3 \vee 4 \vee 5 \vee 6 \vee 7 \vee 8 \vee 9$ 4
- Condition B: Possession = 'Home'
- Window condition C: DIFF (Team Centroid X) > 0.03*
- Relation: A meets B, B filtered on C

* Only the first 5 seconds are considered for the condition

Figure E.4: Experiment 4: Gegenpressing

- Match interval: MET1 [0:00 - end], MET2 [0:00 - end], DFB1 [0:00 - end], DFB2 [0:00 - end]
- Window size: -
- Sliding gap: -
- Condition A: Ball velocity = 0 $\vee$ Null
- Condition B: Inside = True $\wedge$ Rest defence Home ≥ 8 $\wedge$ Rest defence Away ≤ 2 $\wedge$ Centroid Zone Home = $2 \vee 5$ $\wedge$ Centroid Zone Away = $2 \vee 5$
- relation A meets B

Figure E.5: Experiment 5: Corner kick

- Match interval: MET1 [0:00 - end], MET2 [0:00 - end], DFB1 [0:00 - end], DFB2 [0:00 - end]
- Window size: event-based on condition A
- Sliding gap: -
- Condition A: Possession = 'Home' $\wedge$ Team Length Away ≤ 25 $\wedge$ Team Centroid X Away $\leq$ -7
- Window condition B: Duration > 250
- Relation: A filtered on B

Figure E.6: Experiment 6: Low block

Appendix F

Video library

Event	URL
Cornerkick Home	https://www.youtube.com/watch?v=k5BxkB8i1zw
Cornerkick Away	https://www.youtube.com/watch?v=e4UDqIoATUY
Switch play Home	https://www.youtube.com/watch?v=3P-flW45ngY
Switch play Away	https://www.youtube.com/watch?v=nJIppvVfuH4
Low block Home	https://www.youtube.com/watch?v=-L_2WH3-1pg
Low block Away	https://www.youtube.com/watch?v=_x8sh-4kogc
Goal kick Home	https://www.youtube.com/watch?v=B-ak4NDXDU8
Goal kick Away	https://www.youtube.com/watch?v=wfRmKFmWit0
Gegenpressing Home	https://www.youtube.com/watch?v=5IJYNOKIdEw
Gegenpressing Away	https://www.youtube.com/watch?v=2t0jzAZRvTU
Counterattack Home	https://www.youtube.com/watch?v=Uobv4Wxf2t4
Counterattack Away	https://www.youtube.com/watch?v=lA_w5VncAwk

Table F.1: Overview of the web addresses of the videos. Every video represents an event.

Acronyms

CRISP-DM Cross-industry standard process for data mining. 7

DFB Deutscher Fußball-Bund. 33

DTW Dynamic Time Warping. 34, 35, 43, 45

GUI Graphical User Interface. 12, 32

IR Information Retrieval. 9, 12, 19–21, 23, 26, 28–33, 43–45, 47

PA Performance Analysis. 1, 2, 12

QBE Query by Example. 13, 14, 29

QBSE Query by Semantic Example. 6, 13, 22, 29, 30, 34–36, 43–45, 47, 48

TA Temporal Aggregation. 6, 29, 34, 35, 43

TTI Time to Interactive. 19

Glossary

annotation The task of labelling events that occur during a match. For the annotation, it is necessary to have the position, time, and event-specific information.. 2, 47

attacking third A football pitch can be divided horizontally into three equal parts. The third that is the furthest away from the own goal and closest to the opposition goal is referred to as the attacking third.. 34

collective behaviour Collective behaviour is referred to as entities that behave in a similar, coordinated, or interdependent way [55]. An example of collective behaviour is the creation of a low block after loss of ball possession where all players collectively fall-back to the defending third to make the 'field' compact.. 69

collective feature A feature that characterizes collective behaviour. An example of a collective feature is the Team Centroid . 19

event The event data consists of events that happen during a match. Examples of events are shots, passes, and fouls with information about position and time. These events are in most cases manually annotated by an analyst. As a result, all analysts and data service providers have different types of event data. Some of these events could be detected automatically from spatial temporal data, but for events like cautions, fouls, or substitutions it is very difficult. . 26

event An event is an action that happened during a match and is relevant, such as a pass, interception, corner, tackle, etc. [56]. 69

first-order event Events that are rule-based and can be objectively identified, such as foul, throw-in, or a substitution. 2, 3, 11, 13, 15, 17, 21, 34, 43, 70

football philosophy The opinion about answers on questions like *Where on the pitch do players position themselves?*, *What do the players do when the team is in ball possession?*, *What do the players do when a team loses the ball to the other team?*. Famous examples are *Total Football* played by the Dutch national team during the world cup of 1974 and *Tiki-taka* of Barcelona introduced by Johan Cruyff in the period 1988 to 1996.. 4, 9, 11, 12, 48

formation A formation is part of the strategy for a match. For every position there is an unique role relative to the other roles [**?**]. A formation reflects the relative spatial arrangement of the players. A typical way of coaches to describe a formation is 4-3-3 or 4-2-2, but in practice the positioning of the players is more dynamic. .. 10

gegenpressing gegenpressing or Counter pressing is a term introduced by the coach Ralf Rangnick. The idea behind Gegenpressing is to immediately try to reconquer ball possession after loss of ball possession. That means that the players have to run forward to pressure the ball instead of running backwards. Gegenpressing can also be referred to with the five-second rule. The coach Josep Guardiola is famous for applying the five-second rule. It means that players have five seconds to reconquer the ball after loss of ball possession. If the team is not able to regain ball possession within five seconds, the team fall backs to a defensive formation. . 38, 43

low block *"The low block is a defensive tactic in which a team defends from a very deep position on the pitch, meaning they then restrict the amount of space the opposition has to exploit. As a result, opposing teams tend to find it much more difficult to score due to the lack of room to operate in central areas of the pitch."*[24] . 13, 33, 40, 41, 43, 44, 69

over-annotation Annotated events that are wrong or unnecessary due to a wrong judgement. This happens mostly during live annotation when decisions have to be made in a very short period. For example, When a the analyst annotates a short on target, but resulted in a foul because the player was in 'outside position'. That means that the event 'shot on target' should be deleted. If this does not happen, we speak of over-annotation. . 11

rest defence Positioning of the players of the team in ball possession that are not actively involved in the attack [2]. A good rest defense helps to prevent counterattacks and regain possession.. 37

second-order event Events that include collective behaviour. These events are more advanced than first-order events and are prone to subjectivity. These events are selected by the analyst based on his own interpretation and preference of the coach. It is possible that an second-order event is a combination of several first-order events and forms an episode. Examples of events are the build-up of a team and a counterattack. . 2, 3, 5, 6, 15, 18, 19, 21, 31, 33, 34, 43, 44, 47

under-annotation The lack of annotated events. These are mainly mainly caused by lapse of concentration [17]. For example, When a the analyst missed to register that a player passed the ball to another player. . 11