Report from Dagstuhl Seminar 24081

Computational Approaches to Strategy and Tactics in Sports

Ulf Brefeld[*1], **Jesse Davis**[*2], **Laura de Jong**[*3], **and Stephanie Kovalchik**[*4]

1 **Leuphana Universität Lüneburg, DE.** ulf.brefeld@leuphana.de
2 **KU Leuven, BE.** jesse.davis@kuleuven.be
3 **Deakin University – Melbourne, AU.** mail@lmsdejong.nl
4 **Zelus Analytics – Austin, US.** skovalchik@zelusanalytics.com

Abstract

One of the most challenging and interesting aspects in sports are *Strategy* and *Tactics*. In this interdisciplinary Dagstuhl Seminar, we aimed to develop a computational understanding of these concepts in an interdisciplinary setting with researchers and practitioners from Machine Learning, Statistics, and Sports. The seminar was organized around the themes "Discovery", "Evaluation", and "Communication" that were introduced with tutorial and overview style talks about the key concepts to facilitate a common ground among researchers with different backgrounds. These were augmented by more in-depth presentations on specific problems or techniques. Besides several topical discussions in larger groups, there were two panel discussions dealing with differences between individual and team sports and bringing computational analytics into practice, respectively.

Seminar February 18–23, 2024 – https://www.dagstuhl.de/24081
2012 ACM Subject Classification Computing methodologies → Artificial intelligence; Computing methodologies → Machine learning
Keywords and phrases AI, machine learning, sports, team, athletes, strategy, tactics
Digital Object Identifier 10.4230/DagRep.14.2.164

1 Executive Summary

Ulf Brefeld
Jesse Davis
Laura de Jong
Stephanie Kovalchik

The rapid growth in spatio-temporal data in sport over the past decade has generated numerous methodological developments from the statistical and machine learning communities. The richness of modern sports data is enabling sports researchers to analyze every action and decision during a competitive event in increasing detail. Two central topics that have emerged from this new phase of methodological research in sport are data-driven approaches to *strategy* & *tactics*. In a nutshell, *strategy* & *tactics* allow weaker teams or athletes to win over stronger ones. Therefore, they are one of the most interesting and challenging aspects in sports.

[*] Editor / Organizer

Computational Approaches to Strategy and Tactics in Sports, *Dagstuhl Reports*, Vol. 14, Issue 2, pp. 164–181
Editors: Ulf Brefeld, Jesse Davis, Laura de Jong, and Stephanie Kovalchik
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Although both terms describe similar aspects and are even sometimes used interchangeably, they range on different time scales. A *strategy* serves as an overarching umbrella to reach long-term goals. Hence, strategic decisions involve long-term training plans, signing players and coaches, as well as deciding on team formation, pacing, equipment, rotations, or playing philosophy. On a shorter time scale a match/race strategy is the plan made by coaches and athletes before the start of the match or race.

Tactics, on the other hand, is rather short-term. Tactics are the execution and adaptations to the planned strategy to have an edge over the opponent during the match or race. Tactics are therefore often broken down into building blocks or patterns that can be easily communicated to athletes. Note that communication is of utmost importance as tactics are invented by the coaching staff while their implementation is the task of the players/athletes. A tactical pattern may thus involve only subgroups of athletes, or subsections of a race and assign concrete tasks for predefined, context-sensitive situations.

The goal of this Dagstuhl Seminar was to bring together a diverse set of researchers from both academia and industry working on these topics. The seminar drew from people with various backgrounds in terms of area of specialization (Artificial Intelligence, Operations Research, Sport Science, Statistics), role (Academic, Data Provider, Federation, Sports Club) and sport (Australia Rules Football, Baseball, Basketball, Darts, Ice Hockey, Soccer, Speed Skating, Tennis, Wheelchair Rugby).

The seminar was structured around three themes:

Discovery The goal of this theme was to discuss different methods that can automatically identify tactical and strategic patterns from spatio-temporal data. Examples were given for problems such as detecting formations, identifying commonly occurring sequences of actions (e.g., passing sequences), discovering player movement trajectories, and deciding where players should aim a tennis serve.

Evaluation This theme focused on the challenges and pitfalls associated with trying to evaluate the finding of computational approaches to identifying strategies and tactics. This theme focused on highlighting a number of methodological issues and describing ways to assess the validity of discoveries. There were a number of interesting examples given about how causal analysis could be used to evaluate the efficacy of certain tactics. Finally, the potential and risks for using large language models in sports were also discussed.

Communication This theme tackled the problem of how to communicate the findings from tactical studies to an interdisciplinary audience. The emphasis was on how to marry finding from the research literature to things that could be translated into practice. A key point that was made is that it is crucial to think about what types of information will be useful and actionable for practitioners.

The first three days of the seminar focused on one theme, which was introduced with a longer tutorial and then shorter presentations. The final full day of the seminar was open to all topics under the themes and there was a greater focus on presentations from early-career researchers in attendance. The seminar also featured two panels and (small) group discussions about five different topics.

Results

During the seminar, we identified and agreed upon the following action points aimed at trying to continue integrating the various different communities (Sports Science, Operations Research, Statistics, Artificial Intelligence) working on computational approaches to tactics in sports:

1. We will collect a list of venues where computational approaches to tactics in sports are often published. We will host this on the web: `https://dtai.cs.kuleuven.be/sports/venues/`
2. We will explore setting up a slack or discord channel to facilitate more continuous interaction and the ability to quickly get answers to questions. Joris Bekkers and Jan Van Haaren will take the lead on this point.
3. We have setup a document that contains the biographies, contact details, and topics of interest for all seminar participants that are willing to share their information. That will help people stay in touch.
4. We will strive to setup some basic tutorials that illustrate how to implement standard, concepts that reoccur across sports. For example, many team sports have variants of plus-minus, expected possession value metrics, and expected statistics such as expected goals (soccer, ice hockey) or expected rush yard gained (American Football).
5. We will continue to promote the mailing list for disseminating computational sports-related information (job ads, conference call for papers, etc.) and we will use this list to distribute the report on the seminar to reinforce our thanks to the attendees and excitement about the seminar's outcomes: ml-ai-4sports@googlegroups.com

2 Table of Contents

3 Overview of Talks

3.1 The Secrets of Competition: Using AVATARS to Better Understand Exercise Regulation and Competition

Florentina Hettinga (Northumbria University, GB)

Competition between athletes is central to sport. Athletes need to determine how and when to invest their limited energy resources to optimise and self-regulate their pace depending on their physiological and biomechanical capacity as well as environmental factors, such as the presence and behaviour of another athlete. Remote technology can be used to explore mechanisms involved in exercise regulation and competition, for example via the use of avatars, which can be a graphical representation of an athlete's own performance or that of another athlete. I will overview a series of studies using avatar scenario's to better understand decision-making and pacing in sport and competition.

3.2 Using a Markov Chain Model to Identify Optimal Football Match Tactics

Benjamin Holmes (University of Liverpool, GB)

In this paper, we develop a Markov chain representation of football. States consist of different locations on the pitch and set-pieces, and different actions, such as passes, shots, and tackles, move the chain between these states. Novel variables which describe the abilities of the opposing teams in different aspects of football over different zones of the pitch drive the transition probabilities. By simulating a match numerous times using different scenarios, we can identify the optimal choice of tactics: who should play and in what formation, or what style of football to implement. A case-study using a recent match between Everton and Chelsea demonstrates the usefulness of the model, as well as the detailed predictions one can obtain.

3.3 Unveiling Tactical Advantage in Baseball: Insights from Kinematic Analysis and Predictive Modelling

Mamiko Kato (Toyo University, JP)

A remarkable development of data measurement and collection technology in baseball has provided us with opportunities to gather comprehensive performance data. Our research team conducted detailed analyses of batter and fielder performances in the official professional baseball games. Specifically, the kinematic characteristics of batted balls were analysed collectively to evaluate how fast, how high, and in which direction the batters should aim

to hit the ball to increase the chance of making base hits and home runs. Additionally, we introduced a novel method for fielder performance evaluation, addressing the challenge of comparing abilities fairly due to asymmetrical distributions of batted ball characteristics across the baseball field. A machine learning algorithm was used to predict the probability of flyout from the kinematic characteristics of fly balls and compared the probability score for a systematically constructed set of fly balls, uniformly distributed across the field. We discovered that hitting towards the same side of the field increased the chance of a base hit and argued that the advantage was attributable to the large variance in both the direction and magnitude of deflection seen in the trajectories of the pulled fly balls. Based on these studies, using classical statistical methods as well as simulation with predictive modelling with vast amounts of batted ball data will help us conduct fair and rigorous comparisons of players' performance and provide target values with a given aim in baseball. The future development of this methodology in baseball and its potential applications in other sports will be discussed to provide findings and implications in practical situations effectively.

3.4 Expected Thread Models – Overview and Ideas of Validation

Matthias Kempe (University of Groningen, NL)

In this talk, I will give a short intro one a new line of research in 1 vs 1 actions. To study them and to quantify action in football in general, one needs a valid success measure. Expected Thread is one of the suggested measures. However, models that circulate right now might have different short comings and are not validated properly. I like to give an overview on the different models in the literature and propose some standards and guidelines to validate them.

3.5 SoccerCPD: Formation and Role Change-Point Detection in Soccer Matches Using Spatiotemporal Tracking Data

Hyunsung Kim (Seoul National University, KR)

In fluid team sports such as soccer and basketball, analyzing team formation is one of the most intuitive ways to understand tactics from domain participants' point of view. However, existing approaches either assume that the team formation is consistent throughout a match or assign formations frame-by-frame, which disagree with real situations. To tackle this issue, we propose a change-point detection framework named SoccerCPD that distinguishes tactically intended formation and role changes from temporary changes in soccer matches. We first assign roles to players frame-by-frame and perform two-step change-point detections: (1) formation change-point detection based on the sequence of role-adjacency matrices and (2) role change-point detection based on the sequence of role permutations. The evaluation of SoccerCPD using the ground truth annotated by domain experts shows that our method accurately detects the points of tactical changes and estimates the formation and role assignment per segment. Lastly, we introduce practical use-cases that domain participants can easily interpret and utilize.

3.6 What Is Supportive for Coaching Practice and What Is Not (So Much)?

Martin Lames (TU München, DE)

I try to classify computational approaches to strategy and tactics according to their usefulness for sports practice. Basic research in informatics based on sports data, Messi-and-Ronaldo-look-good-in-my-data or I-found-a-new-performance-indicator typed studies may be criticized in this respect. In the narrow sense of support for coaches and teams (not in the wider sense of sports analytics), practical hints for training should be the ultimate aim. The analysis process in sports practice is presented and more concrete requirements for support are derived.

3.7 Match Analysis in German Beachvolleyball – An Ecosystem for Data Collection, Data Analysis, Communication of Results and Training

Daniel Link (TU München, DE)

This talk introduces the game observation concept used by the German national teams in beach volleyball. First, the talk discusses the type of statements in match analysis from a conceptual perspective and explores questions that relevant for beach volleyball coaches. The methods section outlines the logical work steps of match analysis, including: i) querying game scenes according to a classifier; ii) the quantitative preliminary analysis supported by descriptive statistics and graphical reports; and iii) the main qualitative analysis based on video-recordings. It also shows how results are communicated to the German athletes by using a custom made presentation tool and how they use the data for anticipation training.

3.8 Scaling Soccer Data and Analysis using Multimodal Language Modeling

Patrick Lucey (Stats Perform – Chicago, US)

The integration of data and artificial intelligence (AI) has significantly enhanced performance measurement in sports, particularly in soccer. Event data enables the assessment of on-ball performance, encompassing metrics such as xG (expected goals), Possession Value, and win probability. Complementary tracking data captures off-the-ball actions, informing fitness assessments, team tactics, defensive strategies, and passing options. Despite advancements, the scalability of these measurements remains constrained by the lack of "complete" tracking data. For 25 years, tracking data in soccer has relied on in-venue systems, established since 1998. However, the necessity for being in-venue for collection has limited it wide use. Recent years have witnessed efforts to utilize tracking data from broadcast video as a

scalable alternative. Yet, challenges persist, with occlusions caused by players out of view (or complete segments of play being missed completely) resulting in incomplete data and therefore incomplete downstream analysis. In this presentation, I will explore the application of Multimodal LLM approaches to address this limitation. By leveraging these methodologies, I will discuss how complete tracking data can be extracted from broadcast video, ensuring accuracy and completeness akin to in-venue systems. This advancement holds promise for reliable and trustworthy analysis and strategic decision-making in soccer and beyond.

3.9 Team Tactical Performance in Small-sided Games in Football

Sigrid Olthof (John Moores University – Liverpool, GB)

The purpose of this presentation is three-fold: i) measuring team tactical performance, ii) using small-sided games, and iii) highlighting applications to the football practice. Team tactical performance in football (soccer) refers to the cooperation of players within a team and the competition between teams. Positional data from tracking technology is used for team tactical performance metrics representing the positioning and dispersion of players on the pitch. Knowledge of match performance allows for developing and improving performance through training. Small-sided games (SSGs) are training formats representing the match (phases) and therefore a popular training drill. Usually, changing the pitch size and number of players in SSGs affects the individual performance (physical and technical) and team tactical performance compared to match performance, compromising the usefulness and representativeness of SSGs. In this presentation, I shared insights of optimising SSGs by using a similar relative pitch area (total pitch surface / number of players) as the match. This leads to similar team tactical performance in SSGs and the match. These insights can be applied to the football practice with examples of new designs of grassroots competitions, coach dashboards, and a range of games for training programs. Involving coaches in this process is crucial for successful applications.

3.10 Presenting Multiagent Challenges in Team Sports Analytics

David Radke (Chicago Blackhawks, US)

This talk will present several challenges and opportunities within the area of team sports analytics and key research areas within multiagent systems (MAS). We specifically consider invasion games, where players invade the opposing team's territory and can interact anywhere on a playing surface (ice hockey or soccer). We discuss how MAS is well-equipped to study invasion games and will benefit both MAS and sports analytics fields. We highlight topics along two axes: short-term strategy (coaching) and long-term team planning (management).

3.11 Unlocking Insights: Interpretable Models in Soccer Analytics

Pegah Ramihian (Twelve Football – Stockholm, SE)

In the fast-evolving landscape of soccer analytics, leveraging the power of deep learning has become imperative for gaining a competitive edge. However, the opacity of these models often leaves coaches and players in the dark, hindering the practical application of insights. We need to delve into the significance of using interpretable models in soccer analytics, shedding light on the "why" behind every "what."

3.12 Masked Autoencoder for Multiagent Trajectories

Yannick Rudolph (Leuphana Universität Lüneburg, DE)

Automatically labeling trajectories of multiple agents is key to behavioral analyses but usually requires a large amount of manual annotations. This also applies to the domain of team sport analyses. In this paper, we specifically show how pretraining transformer models improve the classification performance on tracking data from professional soccer. For this purpose, we propose a novel self-supervised masked autoencoder for multiagent trajectories to effectively learn from only a few labeled sequences. Our approach employs a masking scheme on the level of individual agent trajectories and makes novel use of a factorized transformer architecture for multiagent trajectory data. As a result, our model allows for a reconstruction of masked trajectory segments while being permutation equivariant with respect to the agent trajectories. In contrast to related work, our approach is conceptually much simpler, does not require handcrafted features and naturally allows for permutation invariance in downstream tasks.

3.13 Practical Implications and Solutions to Foster Innovation in Sports Data Analytics

Martin Rumo (OYM AG – Cham, CH)

The transformation into a data-driven culture presents significant challenges for sports organizations, notably the issue of data silos that impede the free flow of information necessary for innovation in sports data analytics. This talk aims to tackle these barriers by offering practical solutions and discussing their implications. Firstly, we will introduce a data-centric approach to break down internal data silos within sports clubs, ensuring optimized access to data for analytics. Secondly, for industry-wide data silos, we will present the blockchain-based protocol OCEAN, illustrated through a case study from Swiss ice hockey, to privacy conserving data sharing across different clubs. Furthermore, the talk will address the necessity of seamless integration of academic research and solutions into the sports industry's

existing technological frameworks. Strategies to facilitate this integration, enhancing the flow of innovative ideas from universities to the field, will be explored. Additionally, the growing complexity of sports information systems is identified as a challenge to innovation; we propose the adoption of microservices architecture as a scalable and flexible solution to this problem. Conclusively, the presentation will outline actionable steps that organizations can take to overcome these innovation barriers, paving the way for a more integrated, efficient, and innovative future in sports data analytics.

3.14 An Experiment to Investigate the Spatial Component of Serving Strategy in Tennis

Nathan Sandholtz (Brigham Young University, US)

We conducted an experiment with the Brigham Young University Men's Tennis Team to investigate the spatial element of serving strategies in tennis. Serve data, including precise spatial coordinates, were collected for 12 players, with known targets for each serve. Leveraging this data, we estimate player-specific optimal aim locations, accounting for factors such as first vs second serve, speed, and the distribution of their serves around the intended targets, termed "execution error". Our experiment also provides insights on the interplay between conscious beliefs and on-court performance. Our preliminary results reveal apparent differences between players' subconscious behavior and their explicit articulation of optimal aiming locations.

3.15 Towards full automation and scalability when collecting spatiotemporal data in tennis

Joshua Smith (Concordia University – Montreal, CA)

Data analytics in tennis is a fast growing topic of interest, often related to fan engagement at tournaments. Higher ranked players are also able to benefit from the statistics and data collected by these tournaments. However, the collection of spatiotemporal (and even event) data has historically been complicated and expensive, and so developing players at all levels still do not have access to these types of resources. Recent advances in computing power and AI algorithms have allowed for cheaper and more efficient data collection for individual tennis matches. This promises to expose more players to the idea of analytics, with the hope that it can level the playing field. But there are still challenges to consider when building for even more scalability and versatility. This talk will explore some of the issues that arise when trying to fully automate the data collection process of a tennis match. This includes many facets of the approach from court detection, player identification, bounce detection etc. and includes the edge cases that one needs to consider when aiming for fully automatic data collection.

3.16 Tactical Problems in Football using Tracking Data and Causal Methods

Tim Swartz (Simon Fraser University – Burnaby, CA)

A frequent impediment to applied causal analysis is the identification and quantification of confounding variables. With the advent of tracking data in sport, there is often a realistic chance of dealing with the confounding variable problem. In this presentation, we consider three questions involving soccer tactics that are each approached using causal methods (a) what is the benefit of crossing the ball? (b) what is the benefit of playing with pace? and (c) what is the benefit associated with throw-in decisions? The problems each have a common structure, and we provide a template for approaching such problems. The differences between the problems lie in the nature of the independent variables, the dependent variables and the confounding variables.

3.17 Evaluating Sports Analytics Models

Jan Van Haaren (Club Brugge & KU Leuven, BE)

Joint work of Jan Van Haaren, Maaike Van Roy, Pieter Robberechts, Jesse Davis

There has been an explosion of data collected about sports. Because such data is extremely rich and complicated, machine learning is increasingly being used to extract actionable insights from it. Typically, machine learning is used to build models and indicators that capture the skills, capabilities, and tendencies of athletes and teams. Such indicators and models are in turn used to inform decision-making at professional clubs. Unfortunately, how to evaluate the use of machine learning in the context of sports remains extremely challenging. On the one hand, it is necessary to evaluate the developed indicators themselves, where one is confronted by a lack of labels and small sample sizes. On the other hand, it is necessary to evaluate the models themselves, which is complicated by the noisy and non-stationary nature of sports data. The goal of this presentation is three-fold. First, we detail some aspects about how analytics are used within a club environment. Second, we discuss pitfalls and best practices for evaluating models learned from sports data. Third, we overview various ways to validate developed indicators, which requires assessing if they can provide value to the workflow of practitioners.

3.18 A Markov Framework for Learning and Reasoning About Strategies in Professional Soccer

Maaike Van Roy (KU Leuven, BE)

Strategy-optimization is a fundamental element of dynamic and complex team sports such as soccer, American football, and basketball. As the amount of data that is collected from matches in these sports has increased, so has the demand for data-driven decision-making

support. If alternative strategies need to be balanced, a data-driven approach can uncover insights that are not available from qualitative analysis. This could tremendously aid teams in their match preparations. In this talk, I present a novel Markov model-based framework for soccer that allows reasoning about the specific strategies teams use in order to gain insights into the efficiency of each strategy. The framework consists of two components: (1) a learning component, which entails modeling a team's offensive behavior by learning a Markov decision process (MDP) from event data that is collected from the team's matches, and (2) a reasoning component, which involves a novel application of probabilistic model checking to reason about the efficacy of the learned strategies of each team. I will illustrate the framework on one use case, namely that it can be used to optimise a team's defensive strategies when playing against a particular team.

3.19 Discovering Tactics from Team Sports Data

Albrecht Zimmermann (Caen University, FR)

Team sports tactics are about who (which player and/or player role) does what (in terms of legal actions, which includes moving around) where on the pitch, court or whatever the playing field is called, potentially modified by when (in the shot clock, game clock etc).I used my presentation to quickly touch on those different aspects (excluding time) by summarizing a number of works from the literature. The lines are not always clearly drawn: some papers mix what athletes do with where it happens or with who does it. The "what" part is arguably the most complex one and the papers I touched on all use what I've called the "vocabulary" of possible actions/movements to describe what teams and individual athletes do. Some do so rather explicitly, using topic models originally developed for text document analysis, others learn or predefine patterns that occur in certain situations or for certain teams. Some learn the vocabulary from the available data, others predefine it and "only" learn the "phrases" that are being formed. When it comes to the "who", finally, network-based modeling proves to be very powerful and allows to explore "what if" scenarios that would occur if one put different line-ups in the field.

3.20 Shape Descriptors Applied to Tactical Analysis in Football

Felipe Arruda Moura (State University of Londrina, BR)

The idea of the presentation is to introduce and discuss the use of shape descriptors related to team behaviour during football matches. In the literature, the surface area is usually represented by the area of a polygon obtained from the position of the teammates, and represents the total space covered by a given team on the pitch. Although the absolute values of the surface area represent good parameters for the characterization of teams' organization during a match, one can argue that different distributions of players on the pitch can provide equal surface areas or, for similar organization shapes, it is possible that the teams present different areas. Thus, the shape description of the polygon provides more

in-depth information on the complexity of organization during the matches. Some shape descriptors will be presented and discussed about their validity. Also, Multiscale Fractal Dimension will be presented as an alternative for shape description. Finally, some applied results associating shape description with performance indicators will be presented.

4 Panel discussions

We organized two panels during the seminar. The topic of the first panel was "Team versus Individual Sports" and the panelists were Gabriel Anzer (RB Leipzig), Tim Chan (University of Toronto), Florentina Hettinga (University of Northumbria), and Stephanie Kovalchik (Zelus Analytics). The topics discussed during the panel included:

- What do tactics encompass in team and individual sports?
- In team sports, is there tension around salaries in terms of an athlete doing something that is for the good of the team (e.g., playing multiple different positions, playing out of position) that may hurt them individually?
- Is there a trend of team sports reaching out to experts in individual sports to get help with a specific skill or problem?
- At what point is a sport individual and what point is it team? Does this distinction matter?
- Is the type of analytics different between team and individual sports?
- What impact does gamification have in sport? Is this a positive or negative impact? Could playing a game translate to becoming a professional in some sports? How close does a game need to be to mimic real sport?
- Major League Baseball has a revenue of around $10 Billion and its teams employ around 500 analysts. For professional tennis, the annual revenue is around $1 Billion: Why are there not 50 tennis analysts?

The topic of the second panel was "Putting Analytics into Practice" and the panelists were Joris Bekkers (Freelance Data Analyst), Max Goldsmith (Royal Belgian Football Association), Sigrid Olthof (Liverpool John Moores University), and Darren O'Shaughnessy (St. Kilda Football Club). The topics discussed during the panel included:

- How do you handle communicating with domain experts in a practical setting? Do you explicitly undertake initiatives to raise data literacy in the sports organizations?
- How do you manage expectations, e.g., avoid over-promising?
- How do you deal with probabilistic outputs and uncertainty and communicating this to experts? Or how do you overcome the fact that people are bad at thinking statistically, understanding sample sizes, etc.?
- Working at a club can be very tenuous and uncertain as there can be significant turnover based on how a team performs. Strategically, how can you set yourself for the next job, particularly when much of the work you do is protected?
- Earlier on, there was often a mismatch between the types of problems considered by researchers and what the practitioners actually need? Does this gap still exist, or are there more researchers tackling problems that are directly applicable to practitioners?
- How has data analytics influenced coach/practitioners behaviour in the teams you've worked with?

5 Discussion Topics

5.1 Tactics vs Strategy

There was no final agreement on the definition of tactics and strategy after this break-out session and we also realised that the vocabulary is used differently across the world. For example, in North America the word tactics is barely used while in Europe the word tactics is more common. Most participants agreed that tactics have a shorter time-line than strategies. The outcome of the groups is as follows.

Group 1 defined tactics as a set of unconstrained actions and reactions that are anchored in a respective strategy and take the environment into account. In contrast, strategy is planned out and ranges on a longer term. This group also questioned whether it is possible to quantify the quality of a strategy.

Group 2 viewed strategy also as a long term, pre-planned idea of how a team aims to play. Tactics are decisions made on the field, where it makes a difference who takes the decision. While the strategy is decided by coaches, the tactics is rather a joint effort by players and coaches. The difference is also dependent on the actual sport and there may be fluent transitions but all strategies are pre-planned (like load management and player rotation) while this does not hold for all tactics.

Group 3 focused on athletics and skill action plans as tactical knowledge. The group also agreed that tactical decisions are very short term and considered a continuum that ranges from one player, via groups of players, and the team to a game and finally to the entire season and the philosophy of the coach. There are certainly action plans with various time scales, however, strategy can be viewed as long term tactics.

Group 4 also aimed to draw a line between tactics and strategy. The long term strategy was placed at a business level, however discussions went about where to draw the line for in-game decisions? The group agreed that strategy is decided on before the game while tactics are adjustments of the strategy during the game.

5.2 Common Data Format

Event data or play-by-play data is a common type of data that is collected about many different team sports such as soccer, ice hockey, and rugby among others. This type of data records specific semantically meaningful events that occur during a match. For example, relevant events in soccer include passes, tackles, and shots. Each event is annotated with information such as the players involved and the location where the event took place. Such data forms the backbone of many different analysis tasks.

However, event data can be challenging to work with because it is collected by multiple different providers in each sport and each provider often records the data in a different format. This makes working with similar data from different providers tedious for practitioners because each data source usually has to be converted or mapped into a unified format. An even bigger challenge is that each provider may use different definitions for certain events, particularly since some events are inherently subjective.

Less critical, but also challenging in the domain of soccer is the comparison of different tracking data sources. While the format of the data-set itself is somehow restricted to x/y/z-coordinates of either the center of mass or different body-points, different vendors share the data also in different formats, and use different preprocessing algorithms.

Gabriel Anzer, Pascal Bauer and Joshua Smith are involved in a project called the **Common Data Format** (initiated by FIFA and the DFB) that will attempt to address this challenge for soccer. They presented the key ideas underlying the format, which strives for a touch-based model that incorporates both event and positional data. Moreover, they are committed to providing clear definitions of events that companies can use to collect data in the proposed format.

Most participants had experienced some of the aforementioned issues. From an academic and club perspective, there is a clear need for such a unification as academia and clubs cannot reproduce or repeat their experiments on other data formats and comparability of approaches and results is an issue. However, from a provider perspective there is an appreciation that changing definitions is difficult from an operational perspective and creates legacy issues for data that was collected differently. Moreover, the data format also has to adhere to the needs of other user, such as the broadcasting media, which makes reaching an agreement on a joint format difficult.

5.3 Communication and Visualization

Communication between analysts and coaches can be difficult as there is a gap between the terms used by analysts and coaches, which can make it challenging for analysts to prepare what the coaches want. Martin Rumo calls this the "semantic gap". To help bridge this gap, he advocates for a co-creation approach between the analyst and user/coach.

More generally, several challenges exist that hinder the adoption of a common terminology. First, Jan Van Haaren noted that sometimes no term exists and people simply operate by describing the concept. In these cases one may need to invent a term. Second, terminology is often specific to a club and may stay consistent even with changes to personnel. Third, cultural factors affect terminology. For example, players come from different places and players from South American may use different terminology than those from Europe. Sigrid Olthof noted that this an issue in academic work too as papers from the computational literature may use, e.g., artificial intelligence specific terms instead of ones used in sports science. Fourth, misalignment of goals can lead to different terminology. For example, people may build their career around specific terminology and hence are not incentivized to change. One initiative in football (soccer) to help mitigate this problem is FIFA's common football language that will be used in coach & analyst courses.[1] As more people follow these training courses, it will help introduce a more standard vocabulary.

Beyond this, several pieces of advice that were offered included:

- It can be useful to provide a clip library to illustrate good / bad behavior.
- When making visualizations, avoid using colors associated with a rival.
- When discussing probabilities, it can be useful to explicitly state the chances of each open. For example, instead of just saying "There is an 80% chance that we reach the playoffs" also add "and there is a 20% chance that we do not reach the playoffs".
- The Pysport website contains a list of visualization packages: `https://opensource.pysport.org/?categories=Visualization`
- The following article summarizes some good advice about visualizations: `https://knowablemagazine.org/content/article/mind/2019/science-data-visualization`

[1] `https://www.fifatrainingcentre.com/en/resources-tools/football-language/`

5.4 Longitudinal Data

Currently most analysis is done on data from a few matches, or a few seasons. However, very interesting questions can be posed when looking at data that has been collected over many years. For example women's soccer has gone through a relatively quick development and is considered to have a very different style than the men's game. With longitudinal analysis it could be investigated if the women's game is developing to get closer to the men's game over time and if this development for example is going faster than how the men's game developed over the last 20 years. Other questions that could be answered with longitudinal analysis are around variability across seasons, changes in coaching styles and finally talent development. Having longitudinal data from developmental players starting when they join the academy at a very young age until they are retired can answers questions about talent development programs but also about how to maximize a players value/time at the club.

5.5 Context

It may be important when doing analysis to normalize data or exclude outliers because it does not help finding large trends in the data. However, if data on the context was collected, such as players injured or weather data new insights might be created from these outliers. Context also matters on a match level: For example, using event data alone allows track whereabouts of the ball, but not whether the ball possessing player is facing one, two, or even more opponents. The context of how many defenders the player faces influences where he/she will pass the ball to.

Participants

- Gabriel Anzer
Hertha BSC – Berlin, DE
- Felipe Arruda Moura
State University of Londrina, BR
- Pascal Bauer
Frankfurt am Main, DE
- Joris Bekkers
Breda, NL
- Luke Bornn
Zelus Analytics – QSacramento, US
- Timothy Chan
University of Toronto, CA
- Jesse Davis
KU Leuven, BE
- Laura de Jong
Deakin University – Melbourne, AU
- Uwe Dick
Sportec Solutions – Unterföhring, DE
- Max Goldsmith
RBFA – Tubize, BE
- Florentina Hettinga
University of Northumbria – Newcastle, GB
- Benjamin Holmes
University of Liverpool, GB
- Mamiko Kato
Toyo University – Tokyo, JP
- Matthias Kempe
University of Groningen, NL
- Hyunsung Kim
Seoul National University, KR
- Stephanie Kovalchik
Zelus Analytics – Austin, US
- Martin Lames
TU München, DE
- Daniel Link
TU München, DE
- Jim Little
University of British Columbia – Vancouver, CA
- Patrick Lucey
Stats Perform – Chicago, US
- Jakub Michalczyk
Sportec Solutions – Unterföhring, DE
- Darren O'Shaughnessy
St Kilda Football Club – Moorabbin, AU
- Sigrid Olthof
John Moores University – Liverpool, GB
- David Radke
Chicago Blackhawks, US
- Pegah Rahimian
Twelve Football – Stockholm, SE
- Yannick Rudolph
Leuphana Universität Lüneburg, DE
- Martin Rumo
OYM AG – Cham, CH
- Nathan Sandholtz
Brigham Young University, US
- Raimund Seidel
Universität des Saarlandes – Saarbrücken, DE
- Joshua Smith
Concordia University – Montreal, CA
- Tim Swartz
Simon Fraser University – Burnaby, CA
- Jan Van Haaren
FC Bruges, BE
- Maaike Van Roy
KU Leuven, BE
- Christoph Weber
DBS – Frechen-Buschbell, DE
- Hendrik Weber
DFL – Frankfurt, DE
- Albrecht Zimmermann
Caen University, FR